Population genetic structure and demographical history of South African abalone, *Haliotis midae,* in a conservation context.

by

Aletta Elizabeth van der Merwe (neè Bester)

Submitted in partial fulfilment for the degree

Ph.D. Agriculture

at

Stellenbosch University

Genetics Department

AgriSciences

Supervisors: Dr Rouvay Roodt-Wilding
Dr Eugenia D'Amato
Prof Filip Volckaert

Date: Maart 2009

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 25 February 2009

Abstract

South African abalone, *Haliotis midae*, has been the subject of major concern regarding its survival and conservation over the last decade or more. Being the only one of five endemic species with commercial value, there is considerable interest and urgency in genetic management and improvement of this species. Limited genetic information and the increasing conservation concern of this species are considered the key motivations for generating information on the micro- and macro-evolutionary processes of *H. midae*, the overall objective of this study.

This study reported the first microsatellite and Single Nucleotide Polymorphism (SNP) markers developed specifically for *Haliotis midae.* Both these marker types were applied to elucidate the degree of gene flow in nine natural abalone populations whilst testing for two contrasting hypotheses; panmixia versus restricted gene flow. Data was analysed using a series of methodological approaches ranging from traditional summary statistics to more advanced MCMC based Bayesian clustering methods with and without including spatial information. Using only microsatellite data, the historical demography of the species was also examined in terms of effective population size and population size fluctuations. Finally, the evolutionary positioning and origin of *Haliotis midae* with regards to other *Haliotis* species was investigated based on mitochondrial and nuclear sequence data.

Both microsatellite and SNP data gave evidence for subtle differentiation between West and East coast populations that correlates with a hydrogeographic barrier in the vicinity of Cape Agulhas. Population substructure was supported by AMOVA, FCA and Bayesian clustering analysis. Clustering utilizing spatial information further indicated clinal variation on both sides of the proposed barrier with a region in the middle coinciding with a secondary contact zone, indicating possible historical

isolation during glacial periods. Overall, the similar degree of substructure observed with both microsatellites and SNPs supported the existence of contemporary and/or historical factors with genome-wide effect on gene flow. The population expansion measured with the microsatellites was inconsistent with the known recent decline but taking the species' life cycle and large effective population size into account, a shrinkage in population size will probably only be apparent in a few generations time.

On a macro-evolutionary scale, this study presents the first classification of South African abalone as a monophyletic group within the Haliotidae family. The topology based on the combined mitochondrial and nuclear dataset is highly suggestive of a relatively recent radiation of the SA species from the Indo-Pacific basin.

The study concludes by describing the most likely factors that could have affected overall population structure and makes suggestions on how the given genetic information should be incorporated into strategies aimed towards the effective management and conservation of *Haliotis midae*.

Opsomming

Die Suid-Afrikaanse perlemoen, *Haliotis midae*, is oor die laaste dekade of meer die onderwerp van groot bekommernis betreffende die spesie se oorlewing en bewaring. Aangesien dit die enigste van vyf endemiese SA spesies is met kommersiële waarde, is daar besonderse belang en erns in die genetiese beheer en verbetering van die spesie. Beperkte genetiese inligting en ‘n toenemende behoefte om die spesie te bewaar is die hoof motivering agter die generering van informasie rakende mikro- en makro-evolusionêre prosesse in *Haliotis midae* en is die oorhoofse doel van hierdie studie.

Hierdie studie beskryf die eerste mikrosatelliete en enkel basispaar polimorfismes wat ontwikkel is spesifiek vir *Haliotis midae.* Beide tipe merkers is aangewend om die mate van gene vloei in nege wilde perlemoen populasies te ondersoek terwyl twee hipoteses ondersoek is; panmiksie versus beperkte gene vloei. Data is geanaliseer deur gebruik te maak van ‘n reeks metodieke benaderings wat wissel van tradisionele opsommings statistieke tot meer gevorderde MCMC gebasseerde groeperings metodes met of sonder die gebruik van geografiese data. Mikrosatelliet data is ook aangewend om die historiese demografie van die spesie te bepaal in terme van effektiewe populasie grootte asook veranderinge in populasie groottes. Laastens is die evolusionêre posisionering en oorsprong van *Haliotis midae* teenoor ander *Haliotis* spesies ondersoek deur gebruik te maak van mitokondriale en nukleêre DNA volgorde data.

Beide mikrosatelliet en enkel basispaar polimorfisme data lewer bewys van ‘n subtiele genetiese verskil tussen wes en ooskus populasies wat verband hou met ‘n hidrografiese skeiding in die omgewing van Kaap Agulhas. Populasie struktuur is ondersteun deur die analise van molekulêre variansie (AMOVA), faktoriale

komponente analise asook Bayesiese groeperings analise. Groeperings analise wat geografiese informasie insluit dui klinale genetiese variasie aan beide kante van die skeiding aan met ‘n area in die middel wat ooreenstem met ‘n sekondêre kontak gebied. In totaal, ondersteun die soortgelyke mate van struktuur verkry met beide die mikrosatelliete en enkel basispaar polimorfismes die bestaan van hedendaagse en/of historiese faktore met genoom wye invloed op gene vloei. Die toename in populasie grootte vasgestel deur die mikrosatelliet data stem nie ooreen met die onlangse afname waargeneem in die spesie nie, maar met inagneming van *Haliotis midae* se lewenssiklus en groot effektiewe populasie grootte, sal die afname in populasie grootte moontlik eers oor ‘n paar generasies na vore kom.

Op ‘n makro-evolusionêre skaal lewer hierdie studie die eerste klassifikasie van Suid-Afrikaanse perlemoen as ‘n monofiletiese groep binne die Haliotidae familie. Die topologie gebaseer op ‘n gesamentlike mitkondriale en nukleêre datastel is hoogs aanduidend van ‘n relatiewe onlangse verspreiding van die Suid-Afrikaanse spesies uit die Stille-Indiese Oseaan.

Die studie sluit af deur die mees algemene faktore te bespreek wat populasie struktuur kon beïnvloed het en maak voorstelle op watter wyse hierdie genetiese inligting aangewend kan word vir die effekiewe beheer en bewaring van *Haliotis midae*.

Acknowledgements

I sincerely thank my supervisors, Dr. Rouvay Roodt-Wilding, Dr. Maria Eugenia D'Amato and Professor Filip Volckaert for all their support, encouragement and inspiration throughout this project. I'd especially like to thank Rouvay for her input, ideas and logistical precision during the study and preparation of this manuscript. My appreciation goes out to Eugenia for her major contribution to advancing my knowledge in population genetics and most importantly, the interpretation thereof. I would like to thank Filip for his contributions and comments during the final stages of this thesis.

My gratitude also goes out to several people from the Genetics Department at the University of Stellenbosch in particular my colleagues Sam Orchard, Mandi Engelbrecht, Louise van der Merwe, Ruhan Slabbert and Professor Danie Brink for informal discussions related and unrelated to work. Carel van Heerden and his team at the Central Sequencing Facility are thanked for their services and giving me the opportunity to analyse my microsatellite data according to my own time and preference. Especially Renè and Gloudi are thanked for never failing to provide me with sequences of a very high standard. From the Molecular Aquatic Research Group, I would like to express thanks to Lise Sandenbergh and Belinda Swart for assistance during the SNP genotyping and to Julie Hepple for language revision of my introductory chapter.

To my family (especially my mother) and friends who I could not always explain to what all this was about, thank you for the unconditional support and belief throughout. To my husband, Gerhard, thank you for your sincere love and encouragement when

at some stages I doubted completion of my Ph.D. was possible. Lastly I would like to extend my gratitude to my father Mr M. J. de B. Bester for whom this thesis is dedicated for all his love, inspiration and encouragement even in his absence.

Thanks be to God.

Table of Contents

List of Figures

[a] http://web.uct.ac.za/depts/zoology/abnet/

List of Tables

Chapter I

Overview of evolutionary genetic processes in marine species with an emphasis on the conservation of *Haliotis midae*

This introductory chapter aims to give an overview of the latest developments in the genetics and conservation of natural populations of marine invertebrates. It also discusses the genetic and conservation status of the South African abalone, *Haliotis midae* (Fig 1.1), in relation to other abalone species.

Fig 1.1 The South African abalone, perlemoen, *Haliotis midae* (photo: Gert le Roux)

The purpose of the first section is to discuss classical as well as more recently developed approaches towards the inference of population structure in marine organisms which should in turn provide a framework for the marker development and data analysis performed in the chapters that follow. A number of aspects concerning population and conservation genetics of species within a marine environment will be addressed in terms of the following important questions frequently raised in literature:

- What defines a natural population and how is gene flow effectively measured between them?
- What are the main factors determining levels of gene flow in marine species?
- How does information on population structure and demographical history assist in the management and conservation of natural marine species?
- Where does *Haliotis midae* stand with regards to the conservation of natural abalone populations in general?

1.1 What defines a natural population?

Despite various definitions given for populations in literature, two ideas remain consistent when referring to a population of individuals. Firstly, a population always refers to a group of individuals of the same species co-occurring in space and time and secondly, this co-occurrence is usually driven by either demographical or genetic interactions between such individuals. Recent work by Waples and Gaggiotti (2006) places the concept of a population either within an ecological or an evolutionary context. Within the ecological framework, the cohesive driving force behind a population is social and behavioral interactions, while from an evolutionary perspective, the concept of a population is primarily based on genetic distinctiveness. A natural population is therefore considered to be a group of individuals that are predominantly defined by ecological or genetic boundaries laid down by nature only and conforms to the expectations of Hardy (1908) and Weinberg (1908) if the allele frequencies remain unchanged over time. In any natural environment however, populations are usually out of Hardy-Weinberg equilibrium because of environmental and genetic instabilities such as non-random

mating, new mutations, migration, selection, random genetic drift and gene flow. Sweepstake recruitment, population size fluctuations and hidden population structure, for example, are known factors that cause disequilibrium in marine populations (Hedgecock, 2000).

Over the last few decades, management and conservation of marine species have become so important that the identification of isolated populations of fish or marine invertebrates (stock assessment) became the focus point of many marine conservation projects. The term “stock” is generally used in fisheries management, and refers to a demographically independent population that maintains and sustains itself over a reasonable period of time. A stock may also represent a particular genetic unit that needs to be managed and preserved separately (Carvalho and Hauser, 1994; Palsbøll *et al*., 2007). The identification of both management (MU) and evolutionary significant units (ESU), as described in Moritz (1994), has been well recognized as critical to the short and long-term conservation of a species. Such units or stocks can only be identified once the patterns of gene flow and historical associations within a particular species have been assessed. Despite the ongoing debate about the definition of these units (Crandall, 2000; Fraser and Bernatchez, 2001), an ESU is generally regarded as a reproductively isolated group of individuals or evolutionary lineage that shows significant genetic divergence from other conspecific groups over a sufficient period of time. MUs are described as populations currently showing significant differences in allele frequencies and levels of gene flow between them (Paetkau, 1999).

1.2 How is gene flow measured among natural populations?

1.2.1 Molecular markers

Genetic structuring among natural populations is mainly a consequence of the interaction between gene flow, mutations, genetic drift and natural selection specific to a particular species (Amos and Harwood, 1998). In a marine environment, gene flow is highly correlated with the dispersal ability of a species (Bohonak, 1999) while the oceanic surroundings can directly affect the patterns of dispersal owing to the influence of climate, hydrodynamics and biogeographic barriers. In such a complex setting, the connectivity among natural populations is best investigated through comparing the genetic compositions of several geographically separated samples taken throughout the species' distribution range.

Various genetic markers and analytical methods have thus far been used to study population dynamics amongst marine populations. The most commonly used markers include allozymes, mitochondrial DNA genes, microsatellites, amplified fragment length polymorphisms (AFLPs) and more recently, single nucleotide polymorphisms (SNPs). It is well known from literature, that different molecular markers are not equally able to reveal the same level of genetic structuring which highlights the risk of drawing conclusions based on only a single marker type (Olsen *et al*., 2002; Chappell *et al.*, 2004; D'Amato and Carvalho, 2005; Miller *et al.*, 2005). In contrast to anadromous and freshwater species, marine species are generally characterised by weak or low levels of population structure and therefore high resolution molecular markers together with a well planned sampling strategy is needed to reveal hidden structure (Bohonak, 1999).

1.2.1.1 Microsatellites

Commonly, molecular markers are divided into two categories: type I markers which are associated with known genes or coding regions and type II markers which are situated within non-coding regions (O'Brien, 1991). The type II markers most frequently used in population genetics are microsatellites and several studies have confirmed their superiority over other markers, especially with regards to marine populations (Corujo *et al.*, 2004; Ramstad *et al.*, 2004; Kenchington *et al*., 2006; Rose *et al*., 2006; Sønstebø, 2007). What makes microsatellites so attractive for diversity studies is their high level of polymorphism exhibited, particularly at di-nucleotide repeats and their well-documented abundance throughout most genomes. In addition, microsatellite isolation methods have become increasingly effective whilst less expensive and time-consuming. For example, a double-enrichment protocol recently described in Diniz *et al.* (2007) allowed for a 100% recovery of repeat-containing clones together with tailed microsatellite primers that provided an inexpensive approach to high resolution genotyping of compound repeats. The ability for automated high-throughput genotyping has also contributed to the popularity of these markers. The information revealed by microsatellite markers lies within the number of polymorphic loci used in combination with the allelic diversity exhibited at each of these loci and a critical balance between these two criteria is often necessary in order to disclose cryptic population structure (O'Reilly *et al*., 2004).

1.2.1.2 Single Nucleotide Polymorphisms (SNPs)

Microsatellite markers are however not devoid of limitations and the occurrence of size homoplasy in particular, are reason enough to predict that SNPs will become the marker of choice as genomic sequence information accumulates (Seddon *et al.*, 2005). Although SNPs are far less variable than microsatellites, their higher

coverage of the genome together with a simpler mutational model that minimizes the potential for homoplasy, places these markers at the forefront of large scale and cost effective genotyping. The application of SNPs in population studies of marine organisms is relatively new but their potential in estimating genetic diversity as well as addressing issues regarding history, ecology and evolution of populations has been widely investigated in model and non-model organisms (Kuhner *et al.*, 2000; Nielsen, 2000; Brumfield *et al.*, 2003; Morin *et al.*, 2004; Seddon *et al.*, 2005). A recent comparison of molecular markers used for population inference of wild and farmed Atlantic salmon stocks confirmed the smaller amount of polymorphic information gained from SNPs when compared to microsatellites. The study also emphasized the ease of increasing the number of SNPs and showed that in most routine investigations the power of approximately 12 SNPs should be equal to that of eight microsatellite markers (Rengmark *et al.*, 2006; Artamonova, 2007). As the two aforementioned marker types are the only ones implemented in this study, other molecular markers (e.g., allozymes, AFLPs and RFLPs) used for population structure inference will not be discussed further.

1.2.2 Methodological approaches to population structure inference

The underlying idea behind the study of gene flow throughout the years has been to identify differential allele frequencies at a particular set of loci using statistical methods relying on different stochastic migration models. Formerly these methods were primarily based on models formulated by summary statistics e.g. the island model (Wright, 1940), the stepping stone model (Kimura and Weiss, 1964; Nagylaki, 1982) and the isolation by distance model (Wright, 1943). However, as the power in computation increased over time, more modern approaches started to make exhaustive use of simulation methods; better known as model-based approaches.

Simulation in model-based approaches either refers to a straight-forward replication of datasets under the same model or the statistical simulation of a given dataset under a variety of parameter values in order to infer the highest likelihood of a particular parameter set (Marjoram and Tavaré, 2006).

1.2.2.1 Summary statistics

The most widely used summary statistic approach to date is Wright's (1951) F-statistic, F_{ST} or fixation index, which is based on the island model of migration where distance plays no role in genetic differentiation between groups. In this context, fixation refers to the increase in homozygosity as a result of inbreeding which in turn could be a result of population subdivision. F_{ST} assumes the infinite allele model (IAM) of mutation and is a direct measure of genetic divergence among populations calculated in terms of the variance in allele frequencies across populations standardized by the mean allele frequency. Consequently, F_{ST} values range from zero when all populations have equal allele frequencies, to one, when different alleles are fixed within every population. Although F_{ST} values below 0.05 are generally considered to be negligible, literature has increasingly shown that depending on the type of molecular marker used and the species under investigation, possible substructure should not be disregarded based on low fixation indices alone (Waples, 1998; Balloux *et al.*, 2000; Ruvinsky and Graves, 2005). The significance of low F_{ST} values is not as straightforward as it appears to be, especially with hypervariable loci such as microsatellites. In an empirical study by O'Reilly *et al.* (2004) a negative relationship was found between F_{ST} and the variability at microsatellite loci of the marine fish, walleye pollock (*Theragra chalcogramma*). This phenomenon was attributed to size homoplasy and implied that the reduced ability to detect genetic differences among weakly structured

marine populations were more likely because of the step-wise mutational mode of microsatellites rather than the effect of locus variability on F_{ST} *per se*. Similar results were reported in the endangered dusky grouper (*Epinephelus marginatus*) and sockeye salmon (*Oncorhynchus nerka*) where less polymorphic allozymes and microsatellites provided higher estimates of F_{ST} and therefore greater power to resolve weak population structure than highly variable microsatellites (De Innocentiis, 2001; Olsen *et al*., 2004a).

The analog to F-statistics, R_{ST}-statistics have been developed to account for loci undergoing stepwise mutations and provides another summary statistic from which population differentiation can be measured (Slatkin, 1995). Similar to the F_{ST} estimator (θ) of Weir and Cockerham (1984), the measure of genetic differentiation (R_{ST}) of Goodman (1997) is unbiased with respect to differences in sample size and the variance between loci. The simulations of Balloux and Goudet (2002) showed that neither F_{ST} nor R_{ST} can be claimed as the better statistic overall and, in agreement with O'Reilly *et al*. (2004), they proposed that F_{ST} is more reliable in cases of high levels of gene flow while R_{ST} better reflects population differentiation in an environment characterized by low levels of gene exchange.

1.2.2.2 Model-based approaches

While summary statistics approaches have been extensively used for population structure and demographic inferences, model-based analyses in which the performance of a given model is monitored by a set of parameter values, currently provides a more realistic and accurate approach towards studying the processes underlying population dynamics. The model-based approach that has undergone the most important theoretical and computational advances over the past few years is Bayesian computation analysis (Beaumont, 2002; Beaumont and Rannala, 2004).

Bayesian analysis allows for the incorporation of varying parameter values as prior information into the model which, combined with the properties of an observed dataset, are used to estimate certain population parameters (Marjoram and Tavaré, 2006). Without using prior knowledge, Bayesian methods are similar to maximum-likelihood inferences (Edwards, 1972) which aim to evaluate the likelihood of a particular parameter given the observed data. Thus Bayes' theorem postulates that the posterior probability is proportional to the product of the likelihood of observing data given the prior (conditional probability) P(data|ψ) and the probability distribution of a given prior P(ψ); the prior is scaled by the marginal probability of the data:

$$P(\psi \mid \text{data}) = \frac{P(\text{data} \mid \psi)P(\psi)}{P(\text{data})}$$

Bayes (1763), Edwards (1972)

Several computer programs have been developed for the estimation of Bayesian posterior probabilities of population parameters and most of them use the computationally intensive simulation-based Markov Chain Monte Carlo (MCMC) method. The MCMC method allows for the estimation of a joint posterior probability without having to test all the possible combinations of the interdependent parameters (Excoffier and Heckel, 2006). The Markov chain generates a number of random variables that determine the probability distribution of future variable states. The aim of the MCMC methods is to obtain chains that reach stationary distribution, usually a joint distribution of interest, and then sample from these chains to make the necessary inferences. In the field of population genetics, Bayesian inference is not restricted to the detection of underlying genetic structure of populations but also used to estimate admixture and assign individuals to particular populations. There

are several software packages that apply Bayesian inferences for population structure, admixture and assignment analyses e.g. STRUCTURE (Pritchard *et al.*, 2000), PARTITION (Dawson and Belkir, 2001), BAPS (Corander *et al.*, 2003), GENELAND (Guillot *et al.*, 2005), GENECLUST (François *et al.*, 2006) and TESS (Chen *et al.*, 2007). The last three programs utilize geographical information for the spatial detection and location of genetic discontinuities between populations. In GENECLUST and TESS, a new Bayesian algorithm is introduced using Hidden Markov Random Fields (HMRFs) where cluster membership is best explained by comparing allele frequencies at different geographical sites. The concept is the same as in photographic imaging where adjacent pixels are more likely to have the same colour than that of more distant ones (François *et al.*, 2006).

In general, it is expected that by incorporating more parameters into a model, the outcome should be more reliable and discriminative but simulation studies have shown that design variables such as sample size, number of loci, and correlation of allele frequencies can also have a substantial effect on clustering patterns (Rosenberg *et al.*, 2005). François *et al.* (2006) provided evidence that issues regarding clustering (e.g. unrepresentative geographical sampling) could be solved through incorporating the hierarchical Bayes' algorithm of HMRF. This mathematical model accounts for the continuity of a discrete random variable on a graph. Within the context of population genetics, a continuous distribution is related to the concept of isolation by distance (IBD), where individuals are more likely to be more similar to their immediate neighbours. This concept is incorporated as a new prior in the Bayesian algorithm of François *et al*. (2006) and is called the interaction parameter, Ψ. This parameter controls the amount of importance given to the spatial interaction between individuals where $\Psi = 1$ signifies complete and unrestricted interaction

between populations while Ψ = 0 indicates no spatial prior and can therefore be considered as similar to the classical Bayesian clustering model of STRUCTURE (Pritchard *et al*., 2000). By contrast to the latter program, TESS also incorporates departures from HWE in its model. Since most Bayesian clustering programs describe a reduced ability in the detection of distinct clusters as population differentiation decreases, it seems sensible to analyse data with the best combination of software especially in an environment where relatively low F_{ST} values are expected. A few studies have made comparisons on the performance of Bayesian computer programs of which the most relevant to low levels of population differentiation is the assessments made by Latch *et al.* (2006) and Chen *et al.* (2007). Latch *et al.* (2006) determined that both STRUCTURE and BAPS performed extremely well for inferring the number of clusters under low levels of differentiation (F_{ST} = 0.02-0.03) while Chen *et al.* (2007) went further by claiming that STRUCTURE in combination with TESS provides the best possible means to resolve spatial population structure. The simulated data of Chen *et al.* (2007) also supported the hypothesis that TESS is the most effective program for identifying genetic discontinuities between weakly differentiated populations, particularly in a very recent contact zone or clinal scenario. Furthermore, these studies showed that assignment performance varied according to the level of geographical admixture assumed within species and confirmed that assignment methods were well suited for the detection of genetic partitioning or geographical barriers.

Overall spatial Bayesian clustering methods seem to be as consistent as non-spatial programs and since this is especially true when the number of available polymorphic loci is low (Chen *et al*., 2007), these evaluation studies confirm the importance of fine-tuning population structure analysis according to the molecular

and geographical data and the sophisticated computational methods currently available.

1.3 What are the main factors determining levels of gene flow in marine species?

Gene flow is a powerful cohesive force holding populations together and in the marine environment where exchange of migrants are expected to be mostly unrestricted, the understanding of the underlying factors that determine varying levels of population structure has become a central point of discussion in many marine species. Although the relation between larval dispersal, gene flow and population structure has been well recognized for many years (Bohonak, 1999), it is the changes in ecosystems and need for protected marine areas that have called for a better understanding of the origin and maintenance of population structure. For example in marine invertebrate species, which are mostly characterized by large populations and high genetic variability, genetic divergence among populations is thought to be determined by a number of factors associated with spatial and temporal heterogeneity. For intertidal species in particular, the rate of gene flow and subsequent level of structure are dependant on the species' life cycle (Johnson and Black, 2005), length of larval stage (Bohonak, 1999; Lambert *et al.*, 2003) and physical features unique to their particular environment. Detection of substructure in invertebrate species has previously been associated with features such as low dispersal capability relative to the species' distribution range (Hellberg, 1994; Palumbi *et al.*, 1997; Palumbi, 2003; José and Solferini, 2007), physical or oceanographic barriers to larval dispersal (Apte and Gardner, 2002; Baus *et al.*,

2005; Kenchington *et al.*, 2006) and other environmental conditions that might favor isolation (Johnson and Black, 1998).

Genetic structure observed among marine populations generally varies between the extremes of panmixia and completely isolated populations. Here the intermediate levels of gene flow are discussed in terms of the major conditions responsible for genetic divergence in a marine environment e.g. IBD, clinal variation, abrupt genetic discontinuity and genetic patchiness. A more exhaustive study of these patterns was reviewed by Hellberg *et al.* (2002).

1.3.1 Isolation by distance

A positive correlation between genetic distance and geographical distance among populations is commonly known as the stepping stone model of gene flow and is often referred to as isolation-by-distance. Wright (1943) used this term to describe a pattern in which gene flow is more likely to occur among neighboring populations than distant ones so that the outlying populations are connected via a series of ‘stepping stones’. Although the primary factor responsible for this step-wise differentiation is geographical distance *per se*, the magnitude of the relationship between genetic and geographical distance is greatly dependant on migration and mutation rates in addition to the time it takes to reach equilibrium between genetic drift and migration (Hutchinson and Templeton, 1999). As migration rate is influenced by dispersal ability, individuals will only move limited distances because of restricted dispersal. In some species, individuals do however travel very large distances and genes can cover any distance within the species’ distribution range (Dingle, 1996). Nonetheless, simulations by Palumbi (2003) show that in species with continuous distribution, isolation by distance is fairly robust and occurs in populations with a variety of dispersal patterns and oceanographic settings. In

species with discontinuous habitats however, isolation by distance is found on different geographical scales ranging from local to transoceanic, and is usually associated with some sort of limitation to larval dispersal (Pogson *et al.*, 2001; Buonaccorsi *et al.*, 2004; Purcell *et al.*, 2006). An individual assignment test in salmonids demonstrated how dispersal can be a function of geographical distance, despite no apparent isolation by distance pattern observed (Castric and Bernatchez, 2004). Although isolation by distance plays a prevalent role in the structuring of a number of marine species, the correlation between genetic distance and other factors such as temporal (Hendry and Day, 2005) or longitudinal and latitudinal distance has also been described recently. Hemmer-Hansen *et al.* (2006) showed that the pattern of structuring in European flounder (*Platichthys flesus*) reflected isolation by latitude rather than with geographical distance. Maes *et al.* (2006) reported significant isolation by time (IBT) in yearly samples of European eel (*Anguilla anguilla*) where gene flow is restricted because of differences in reproduction times between individuals.

1.3.2 Clinal variation

In contrast to phylogeographical breaks, which are an abrupt change in genetic variation, clinal variation is a gradual change in allele frequencies along a geographical or other environmental gradient (Storz, 2002). It is relatively common along oceanographic coastlines where clines can develop as a result of differential adaptation to conditions such as temperature, pH, salinity or depth (Sokolova *et al*., 1997). On the other hand, a cline may also develop between two genetically divergent populations that experience secondary contact after a period of isolation (Barton and Hewitt, 1985; Durrett *et al.*, 2000). As with the previous patterns of population differentiation, clinal variation also suggests that there are several factors

influencing the pattern of allelic variation besides gene flow, and once again, challenges the perception that allelic frequency distributions automatically become homogenized within a marine system. Gene flow does however counteract these additional acting forces and it is precisely the balance between these opposing pressures that may result in allelic clines (Maes and Volckaert, 2002). Clinal variation has been reported in several marine species of which the most frequent observations were made at allozyme loci (Koehn and Williams, 1978; Ropson *et al.*, 1990; Chan *et al.*, 1997; Cimmaruta *et al*., 2005). Since allozymes code for proteins or enzymes that are more likely to be affected by differential selection pressures than neutral markers, the frequency of the alleles can vary according to the strength of the selection in populations at a particular environmental site. An interesting example is that of the leucine aminopeptidase (Lap) locus in blue mussel, *Mytilus edulis* that exhibits an allele which frequency is strongly correlated with water salinity (Koehn *et al.*, 1980). This allele, which progressively declines in low salinity environments, is a good representation of clinal variation driven by habitat-associated selection. Furthermore, clinal variation on a latitudinal gradient is a common occurrence among marine and freshwater fishes and has been confirmed by microsatellite data for species such as Atlantic cod, *Gadus morhua* (Nielsen *et al.,* 2003), turbot, *Scophthalmus maximus* (Nielsen *et al.,* 2004), and killifish, *Fundulus heteroclitus* (Adams *et al.*, 2006). Variation at allozyme loci, mtDNA haplotypes and microsatellites suggest that these clines originally formed as a result of secondary contact associated with post-Pleistocene colonization history of the species' habitat (Ropson *et al.*, 1990; González-Villaseńor and Powers, 1990; Adams *et al.*, 2006).

1.3.3 Abrupt genetic discontinuity

Genetic discontinuity is the sudden change in genetic composition between two neighbouring populations due to restricted gene flow across a biogeographic barrier. While most genetic breaks are associated with present-day physical barriers such as ocean currents, thermal fronts and land barriers, genetic discontinuity can also indicate historical events where barriers could have resulted from the direct and indirect consequences of glacial episodes (Uthicke and Benzie, 2003).

Genetic discontinuity, whether due to historical or contemporary factors, is mostly described in species with pelagic larvae with limited dispersal capacity. Depending on a species' dispersal ability, a phylogeographic break can exist across a wide geographical range and as proposed by Longhurst (1998) is often consistent with the regional oceanography and suggested biogeographic provinces of a given coastline (Waters and Roy, 2003; Teske *et al.*, 2006; Zardi *et al.*, 2007). In South Africa for example the three biogeographic provinces are the cool-temperate West coast, the warm-temperate South coast, and the subtropical East coast (Emanuel *et al*., 1992). The precise location of the boundaries between them is uncertain but the biogeographic barriers near Cape Agulhas and just south of Port St Johns are generally accepted as the main areas of transition (Figure 1.2).

If the barrier to larval dispersal is strong enough to block gene flow entirely, the populations on either side may be completely isolated from each other while populations elsewhere in the species' range are still open to genetic exchange Hellberg *et al.* (2002). The physical features generally associated with restricted gene flow along continuous coastlines are ocean currents, upwelling systems, temperature gradients/fronts and habitat instability. Deep trenches and circular currents surrounding islands may also limit adult migration through the retention of early juvenile stages and act as a sufficient barrier to gene flow (Hansen and

Østerhus, 2000). For instance, the genetic break between cod populations from Iceland or the Faeroes and that of the European continent has been widely documented as examples of the physical isolation of island populations through a deep-water barrier (Joensen *et al.*, 2000; Roman and Palumbi, 2004; Hoarau *et al.*, 2004).

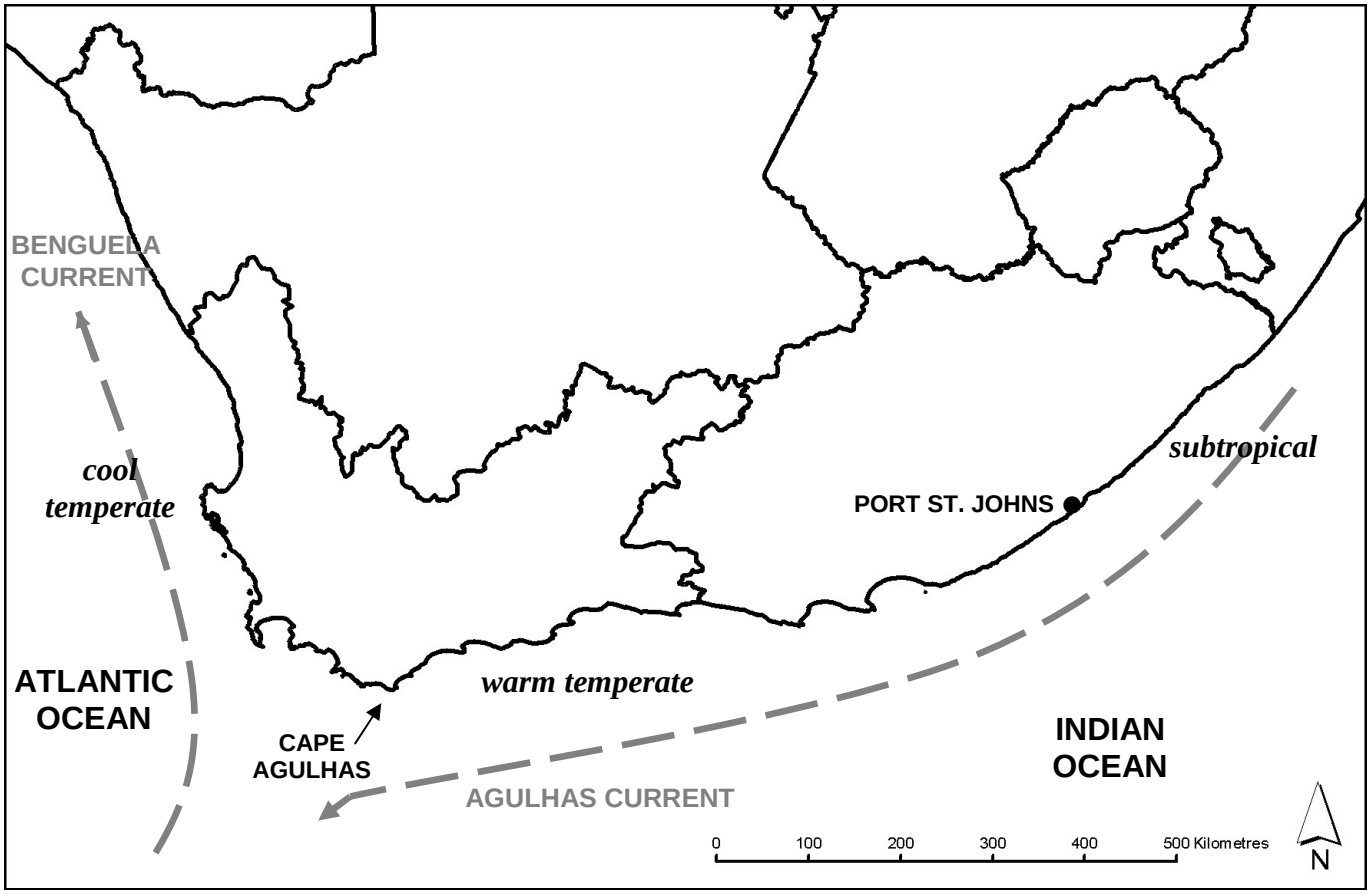

Figure 1.2 Map of South Africa showing the major currents (the cool Benguela current and the warm Agulhas current) dominating the region and the biogeographic boundaries suggested in the literature (Cape Agulhas and Port St. Johns)

From a historical perspective, examples of major barriers that could have affected dispersal and migration of marine biota includes the Pliocene-Pleistocene glaciations, subsequent climatic changes (Widmer and Lexer, 2001; Olsen *et al.,* 2004b; Fraser and Bernatchez, 2005; Rabassa *et al*., 2005) and the closure of former seaways such as the Isthmus of Panama (3 MYA) (Knowlton *et al*., 1993;

Knowlton and Weigt, 1998; Marko, 2002), the Tethyan Sea (~14 MYA) (Rögl and Steininger, 1983) and the Indonesian Seaway (~13 MYA) (White, 1994). Interglacial periods also often resulted in geographical range expansion and secondary contact of formerly fragmented populations among isolated refugia in a number of marine biota (Flores *et al.*, 1997). Such range expansion is usually identified by a significantly lower heterozygosity level of populations in the recolonized areas (Hellberg *et al.*, 2002). All together, the reason behind any genetic break is rarely singular and what is most often reflected in the existing population structure of a species is the combined effect of historical obstructions to gene flow along with contemporary restricted gene flow related to present-day physical barriers.

1.3.4 Genetic patchiness

The term genetic patchiness was first introduced by Johnson and Black (1982) and refers to the genetic heterogeneity observed in marine species over a relatively small temporal or spatial scale. Various studies have shown that such genetic heterogeneity is likely to be a result of temporal variation in the genetic composition of recruits (David *et al.*, 1997; Li and Hedgecock, 1998; Planes and Lenfant, 2002; Gilbert-Horvath *et al.*, 2006) and could therefore be directly linked to the population dynamics of the pelagic larvae themselves (Johnson and Black, 2005). A few mechanisms have been proposed to explain the temporal variation observed in different recruit samples and the most frequently reported is the variation in reproductive success of adult populations, especially in free-spawning marine invertebrates with limited opportunity for survival during the larval stage. Only a few adults may contribute to the next generation and this random variance among seasons or cohorts leads to the genetic patchiness, characteristic of many recruits. Under such a theory of ‘sweepstakes reproductive success’, variation in parental

contributions are attributed to spatial as well as temporal changes in oceanographic conditions within and between seasons (Hedgecock, 1994). Chaotic genetic patchiness due to sweepstakes reproduction is most common in marine invertebrates such as limpets (Johnson and Black, 1982, 1984), urchins (Watts *et al.*, 1990; Edmands *et al.*, 1996; Moberg and Burton, 2000), soft coral (Burnett *et al.*, 1994), bivalves (David *et al.*, 1997) and oyster (Li and Hedgecock, 1998) but have also been described in marine fishes for example anchovy (Hedgecock *et al.*, 1994), shortbelly rockfish (Larson and Julian, 1999), sea bream (Planes and Lenfant, 2002) and European eel (Pujolar *et al.*, 2006). Another possible reason for temporal variation among recruits could be selection during early life stages which could lead to differential survival of genotypes during settlement (Johnson and Black, 1984; Singh and Green, 1984). As with clinal variation, different recruit samples often exhibit varying allele frequencies along a gradient of environmental change (Koehn *et al.*, 1980). Hence, larvae coming from different source populations such as formerly isolated refugia with variable genetic composition may be responsible for the genetic patchiness of recruits (Kordos and Burton, 1983; Hare and Avise, 1996; Ruzzante *et al.*, 1996; Larson and Julian, 1999). Overall, genetic patchiness is expected to have the highest occurrence in species where recruitment is to a larger extent dependant on oceanic conditions.

In light of the factors affecting patterns of distribution and amount of genetic diversity discussed above, restriction to gene flow in the marine environment seems to be more common than previously assumed and hopefully explains the degree to which marine populations may actually be closed to migrants. The interpretation of the mechanisms shaping structure is just as important as estimating the magnitude

and scale of gene flow and provides a solid platform towards the management and subsequent conservation of marine stocks, the main objective of most biodiversity studies.

1.4 How does information on population structure and demographical history assist in the management and conservation of natural marine species?

The previous sections emphasized the importance of detecting levels of gene flow as understanding patterns of connectivity among marine populations is the first step towards effective management and subsequent conservation of individual stocks (Waples *et al.*, 1998; Paetkau, 1999; Palumbi, 2003). The stock concept which is reviewed by Carvalho and Hauser (1994) is mainly linked to the management of populations or units that are ecologically separated through space and time, while long-term conservation of such units is more concerned with the evolutionary or reproductive interaction among individuals.

When significant structure is found between populations, it is suggested to consider and manage them as independent stocks (Frankham, 1995; Avise, 1995; 2000). In cases where the population structure of a species is well defined, marine protected areas (MPAs) or reserves can be set up in order to preserve as much of the species' genetic diversity as possible (Gerber *et al.*, 2003; Buonaccorsi *et al.*, 2005). The identification of stock structure is especially important in species that have recently shown a dramatic decline in numbers or have become increasingly threatened by human and other interferences. Not only is it fundamental to the preservation of genetic diversity (Kenchington and Heino, 2002; Kenchington *et al.*, 2003) but it also has important implications for brood stock collection and breeding

programs within associated stock enhancement efforts. The genetic structure of natural populations forms the basis of identification of representative populations most suitable for stock enhancement or recovery (Ward, 2006). Stock assessment has played a vital role in the successful implementation of management strategies for a diverse range of endangered marine species e.g. corals (Van Oppen and Gates, 2006); loggerhead turtles, *Caretta caretta* (Bowen *et al.*, 2005); sockeye salmon (Kozfkay *et al.*, *in press*) and white abalone, *H. sorenseni* (Gruenthal and Burton *et al.*, 2005).

Most of these management efforts also included the potential increase in stock sizes through ranching; the release and introduction of hatchery-reared individuals into the wild (Mustafa, 2003). To assess whether wild stocks have been disrupted with genetically underrepresented ones, stock evaluation of natural populations is important prior to as well as after the release of cultured animals (Leary *et al.*, 1995; Cross, 1999). The same factors affecting genetic diversity in natural populations can lead to a strong reduction in genetic variation of cultured stocks, especially since the latter is often characterized by smaller population sizes. Whether in a natural or a hatchery environment, small populations are more likely to be negatively affected by genetic effects such as reduction in effective population size (N_e), natural selection, genetic drift and inbreeding (Lynch *et al.*, 1995; Frankam *et al.*, 2003). Hence, small populations will often exhibit lower fitness (Reed and Frankam, 2003) and face increased risk to extinction (Saccheri *et al.*, 1998). Populations that are currently large may have small effective population sizes due to a bottleneck or population size decline in the past (Crandall *et al.*, 1999; Turner *et al.*, 2002) and the identification of present populations with reduced N_e is therefore crucial for conserving a species' genetic diversity. As it is necessary to distinguish between the

populations that naturally exhibit reduced genetic variation and those that recently experienced a reduction in population size (Pearse *et al.*, 2004), the latest developments in estimation of N_e and other historical processes are discussed in the following sections.

1.4.1 Effective population size

Effective population size is estimated through the composite parameter θ ($4N_e\mu$) where N_e is the effective population size and μ is the mutation rate (Kimmel and Chakraborty, 1996). N_e is one of the primary indicators of populations at risk in that it affects the degree to which populations respond to genetic drift or selection and thus evolve over time. Evolutionary and conservation studies are increasingly relying on the estimation of N_e and obtaining reliable estimates of this parameter depends heavily on the type of data available. In general N_e estimates can be based on genetic or demographic data, but since demographic information (e.g. sex ratio, fluctuation in *N* and pedigree data) is often unavailable for natural populations, current approaches to estimating effective population sizes are mainly based on indirect methods using genetic information only.

As reviewed by Wang (2005) the theoretical bases for N_e estimation can rely on 1) heterozygote excess, 2) linkage disequilibrium, 3) temporal changes in allele frequencies or 4) genetic variation within and between populations. Genetic data can be obtained from either a single sampling event where N_e is inferred from linkage disequilibrium between pairs of loci (Hill, 1981; Bartley, 1992; Vitalis and Couvet, 2001) or from multiple samplings where N_e is based on allele frequency change between samples taken at different points in time (Nei and Tajima, 1981; Waples *et al*., 1989; Berthier *et al.*, 2002; Wang and Whitlock, 2002). For both single and multiple sampling methods, several advances have been made towards

their computation using summary statistics (Tallmon *et al*., 2004), likelihood methods (Anderson et al., 2000; Wang, 2001; Berthier *et al.*, 2002), Bayesian calculation (Tallmon *et al*., 2004, 2008), coalescent approaches (Beaumont, 2003; Laval *et al*., 2003) or a combination thereof. Although the likelihood and other probabilistic models are considered to be more accurate and precise, some form of bias has been reported for most N_e estimators, irrespective of theoretical and computational bases. England *et al*. (2006) showed that single sampling estimators are severely biased especially when the sample size is small. Several studies have noted bias in temporal N_e estimates where the effects of migration or natural selection were ignored (Luikart *et al.*, 1999; Wang, 2001; Berthier *et al.*, 2002). Araki *et al.* (2007) in particular found strong downward biases in temporal N_e estimates of admixed populations in which sampling is unknowingly taken from two or more gene pools with potential differences in reproductive success. So far, the likelihood based estimator of Berthier *et al.* (2002) showed superiority to standard moment estimators but only when genetic drift is strong and the sample size is small relative to N_e. A more recent comparative study evaluated the performance of a Bayesian-based summary statistic estimator (SummStat) relative to moment- and likelihood-based methods and concluded that it was the least biased over all parameters tested (Tallmon, 2004).

The coalescent approach to estimating N_e is also based on the composite parameter θ (Xu and Fu, 2004) but even though N_e can therefore be estimated without knowledge of the mutation rate, calculations of θ are not devoid of constraints. Simulation studies have showed that the error rates can be reduced through increasing the number of independent loci used (Kuhner *et al.*, 2000; Nielsen, 2000; Carling and Brumfield, 2007). Also because of the computational

demand, most coalescent-based methods such as of Beerli and Felsenstein (2001) will continue to reveal some sort of bias associated with running only subsets of the data (Austin *et al.*, 2004; Hanfling and Weetman, 2006; Hemmer-Hansen *et al*., 2007).

1.4.2 Historical demographic parameters

Historical events such as bottlenecks, range expansion, and colonization following a founder event can have profound effects on the current genetic variability within species. Of particular importance to the management and conservation of existing populations is the identification of populations that recently went through genetic bottlenecks or population declines. Before molecular markers such as microsatellites and SNPs became readily available, the most commonly used approach to infer population size changes were to examine the shape and distribution of observed pair-wise differences in mtDNA sequences (Rogers and Harpending, 1992; Harpending and Rogers, 2000). This mismatch distribution approach was initially based only on an infinite sites model but further developments incorporated the finite-sites model of evolution as well as a bootstrap procedure to define confidence intervals (Schneider and Excoffier, 1999). Still, interpretation of results are not considered straight-forward given that the level of gene flow is known to affect the distribution of the mismatch in such a way that the time of coalescence can sometimes hide a true historical event (Ray *et al.*, 2003).

Fortunately as molecular markers became more accessible, a variety of methods were developed for the inference of historical demographic events from microsatellite data for example. Based on the unique mutational properties of microsatellites, most of these approaches were modeled to detect recent bottlenecked populations. These include reduced allelic diversity from loss of rare

alleles (Allendorf, 1986); non-random association of alleles at different loci (Waples, 1991, 2002); change in allele frequency distribution (Luikart *et al.*, 1998; Reich and Goldstein, 1998; Reich *et al*., 1999; Beaumont, 1999); increased HWE heterozygosity compared to that expected at equilibrium from number of alleles when the population is at mutation-drift equilibrium (Cornuet and Luikart, 1996; Luikart and Cornuet, 1998; Piry *et al*., 1999) and the rare allele ratio *M* of Garza and Williams (2001). Also, coalescent-based methods formerly developed for the Bayesian inference of demographic parameters such as mutation rate, time to the most recent common ancestor and growth rates from mtDNA sequence data (Beerli and Felsenstein, 2001; Kuhner *et al.*, 1995, 1998; Wilson and Ranalla, 2003) has been modified to accommodate Bayesian based coalescent inference of demographic parameters from microsatellite or SNP data. The most widely used software packages includes BATWING (Wilson *et al*., 2003); MIGRATE (Beerli, 2006) and LAMARC (Kuhner, 2006). The growth parameter in LAMARC for example does not only determine if the population has been through a recent period of growth or decline but also estimates the rate at which such an event is taking place. Despite the computational demand of these programs and bias introduced through sub-sampling, relative values of growth, migration and mutation rate can effectively be compared between populations (Hemmer-Hansen *et al.,* 2007; Palstra *et al.,* 2007).

The methods described above by no means cover all the different statistical approaches or software programs available today, but rather give an overview of the parameters that are considered most essential for management and conservation purposes. Although important conservation issues such as inbreeding versus outbreeding depression have not been discussed in any way, this section stresses

how a better understanding of population structure and historical demography within a given species has profound implications for the effective management and conservation of its genetic diversity as a whole.

1. 5 *Haliotis midae-* an introduction to the species

Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda; Haliotidae; *Haliotis*

Haliotis midae is one of the first seven abalone species described by Linnaeus (1758) and belongs to the marine gastropod family Haliotidae (Gastropoda). It is one of the largest abalone species and because of size, meat quality and abundance has long reached commercial sustainability on the overseas market. This makes *H. midae* one of the most sought-after and valuable marine mollusks of the African continent and as the demand for it is increasing, so is the potential for overexploitation and the need for effective management of natural stocks. Knowledge of the species' evolutionary relation to other abalone, life cycle, environmental surroundings, intra- and inter-specific genetic variation and external threats all plays a vital role in its potential for survival and genetic restoration.

1.5.1 Evolutionary history and phylogeny of abalone

The marine gastropod family Haliotidae (Gastropoda) includes about 56 recognised species of abalone distributed in tropical and temperate waters around the world. The evolutionary history of the family dates back to at least the Upper Cretaceous while fossil records place the majority of the southern temperate species in the Pleistocene era (Geiger and Groves, 1999). Three hypotheses have thus far been proposed for the origin of abalone, the Pacific Rim model (Talmadge,

1963), an Indo- Pacific origin (Lindberg, 1992; Briggs, 1999) and an origin in the Tethys Sea (Geiger and Groves, 1999). The most recent studies on the phylogeny of abalone indicated a radiation of the genus into the southern and northern hemispheres from the Indo-Pacific area (Coleman and Vacquier, 2002; Estes *et al.,* 2005; Streit *et al.,* 2006; Degnan *et al.,* 2006). As *H. midae* is the only SA species to be included in these studies, the precise origin of extant abalone in South Africa has not been determined. Colonisation of South African waters can either entail a relatively recent monophyletic radiation or represent several independent lineages that colonised the area at different times.

1.5.2 Distribution and Habitat

In South Africa, *H. midae* has a wide coastal distribution of approximately 1500 km ranging from St Helena Bay on the West coast to Port St. Johns on the East coast (Lindberg, 1992; Branch *et al.*, 2002). Only one other South African species, *H. spadicea*, exceeds the distribution range of *H. midae* and appears up to the northern coast of KwaZulu-Natal. Although three more species (*H. parva; H. speciosa* and *H. queketti)* are found along the coast of South Africa, the name abalone is mostly associated with *H. midae* and is locally referred to as 'perlemoen'. The global distribution of abalone species is concentrated in temperate and tropical oceans but some species are found as far north as Alaska (*H. kamtschatkana*) and as far south as New Zealand (*H. iris, H. australis*, *H. virginea*) or the South Pacific Ocean (Lindberg, 1992; Wood, 1993). Typically, the habitat of abalone comprises nearshore rocky reefs, crevices, boulders and other oceanic substrates. While *H. midae* lives in water temperatures between 7 and 20℃, globally abalone can be found in temperatures ranging from 2 to 30℃ (Leighton, 2000). Adult *H. midae* do not usually occur below depths of 50 m while some species such as white abalone

(*H. sorenseni*) have been sited in locations as deep as 140 m off the Californian coast. On average, the depth distribution of abalone ranges between one and 35 m. Both *H. midae* and *H. spadicea* are intertidal species found between 10 and 50 m, while *H. parva* has a subtidal depth range (10-20 m) and is found in more sheltered areas. *Haliotis queketti* and *H. speciosa* are extremely rare and found in deeper waters.

1.5.3 Life history

Similar to all other abalone, *H. midae* is a dioecious broadcast spawner with a seasonal reproductive cycle that only reaches sexual maturity after eight to ten years (Newman, 1968; Branch *et al.*, 2002). They are benthic invertebrates with a non-feeding pelagic larval phase of approximately seven days while the exact duration depends primarily on the temperature and not so much on the availability of nutrients, as with feeding larvae. Although the larval phase is relatively short and dispersal is not expected to happen across oceans, larvae can still travel hundreds of kilometers before settlement. The near-shore rocky reefs habitat of *H. midae* and the complex hydrodynamics around the South African coast-line makes a direct correlation of larval duration with its dispersal potential inappropriate. An experimental study by De Waal *et al.* (2003) owed a significant difference in dispersal ability of *H. midae* juveniles, released at two geographically distinct locations, to a more intense wave action at one of the sites. Several studies have shown that marine dispersal is often more complex than it appears to be and that a higher proportion of pelagic larvae actually settle near source populations in contrast to their apparent dispersal potential (Cowen *et al.*, 2000; Barber *et al.*, 2002). Nonetheless, larvae settle once they have found a suitable substrate and develop into the crawling stage or juvenile phase. Juveniles feed on benthic

microflora while larger animals feed mainly on marine algae (Leighton, 2000). Growth is slow and once they have reached sexual maturity at approximately eight years old, adult abalone remains sedentary and aggregate in groups to assure efficient fertilization. Female abalone can produce up to 10 million eggs but the actual amount being spawned may depend on temperature or length of the day. Once fertilized, with only about 1% success rate, eggs hatch into microscopic free living larvae to complete the cycle of reproduction.

1.5.4 Applied genetics in H. midae and other abalone

Whether it is for the successful management of natural stocks or the genetic enhancement of commercial stocks, DNA markers has been realized as indispensable for the effective sustainability of genetic resources in abalone. DNA markers such as allozymes, mitochondrial DNA, microsatellites, AFLPs and SNPs have been widely applied to address various aspects related to management of natural and commercial abalone stocks.

Besides assessment of genetic diversity and population demographic parameters, further applications of DNA markers include species identification, parentage assignment, linkage mapping, quantitative trait loci (QTL) identification and marker-assisted selection. As highlighted in previous sections of this chapter, the effectiveness thereof depends mostly on the number of informative markers available and the series of methods used to analyze them. In the last few years, the development of species-specific molecular markers has been a major focus point of a collaborative effort concerning the genetic enhancement of *H. midae* in South Africa. Seventy-four microsatellite markers (Bester *et al.,* 2004; Slabbert *et al.,* 2008) and 32 SNP markers (Bester *et al.,* 2008; Rhode *et al.,* 2008) have thus far been reported for *H. midae* while a number of potential AFLP markers have been

positioned on a preliminary linkage map of a F1 parental cross (Badenhorst, 2008). A cDNA library constructed from *H. midae* gill tissue enabled the submission of 73 expressed sequence tags (ESTs) to Genbank (Bester *et al*., 2008) and currently provides sequence data for the isolation of additional SNPs. Furthermore, *H. midae* specific primers have been developed for the amplification of several mitochondrial as well as nuclear genes of which all have been or could be implemented in population structure analysis, phylogenetics or species identification (Sweijd *et al.* 1998a; Sweijd 2002; Evans *et al.* 2004b). These include primers for cytochrome oxidase I and III, NADH dehydrogenase subunit I and 3, the 16S rRNA gene, haemocyanin and the lysin precursor gene.

In general, the advances made in abalone genetics has mainly concentrated around species with either commercial potential, those becoming increasingly threatened or both. Several studies have focused on the study of population subdivision within abalone species and the genetic differences between commercial and natural stocks. Specifically in aquaculture, an increasing number of molecular markers are currently employed in the assignment of offspring to parents following the investigation of the mode of inheritance of these markers. As a consequence, first-generation linkage maps have become available for a number of species which in turn could lead to the identification of genes that control important quantitative traits in abalone. For example, Baranski (2006) found putative QTLs linked to growth rate in *H. rubra* and *H. laevigata* using microsatellite markers and a selective DNA pooling strategy. Table 1.1 provides a summary of the marker-related genetic studies performed over the last decade on species considered to be of high commercial value.

Table 1.1 Summary of marker-related genetic studies performed on commercially important abalone species.

Species	Molecular Markers	Genetic Application	Reference
H. asinina	microsatellites	parentage assignment	(Selvamani *et al.*, 2001)
	mtDNA, RAPDs, RFLPs	population structure	(Klinbunga *et al.*, 2003)
	RAPDs, SCARs	species identification	(Klinbunga *et al.*, 2004)
	microsatellites	population structure	(Tang *et al.*, 2004)
	RAPDs, microsatellites	population structure	(Tang *et al.*, 2005)
H. cracherodii (black abalone)	allozymes, mtDNA	population structure	(Hamm and Burton, 2000)
	allozymes	population structure	(Chambers *et al.*, 2006)
H. discus discus (Pacific abalone)	microsatellites	parentage assignment	(Hara *et al.*, 2004)
	microsatellites	species identification	(Sekino and Hara, 2007a)
	microsatellites	commercial stock assessment	(Hara and Sekino, 2007a)
H. discus hannai (Pacific abalone)	microsatellites	parentage assignment	(Li *et al.,* 2003a, b)
	microsatellites	species identification	(Sekino and Hara, 2007a)
	microsatellites	commercial stock assessment	(Li *et al.*, 2004, 2007)
	AFLPs, RAPDs, microsatellites	linkage mapping	(Liu *et al.*, 2006)
	microsatellites	linkage mapping	(Sekino and Hara, 2007b)
	microsatellites	linkage analysis	(Sekino *et al.*, 2006)
	microsatellites	parentage assignment	(Hara and Sekino, 2007b)
H. fulgens (green abalone)	allozymes	population structure	(Zúñiga *et al.*, 2000)
	microsatellites	population structure	(Gutiérrez-Gonzales *et al.*, 2007)
H. laevigata (greenlip abalone)	microsatellites	population structure	(Maynard *et al.*, 2004)
H. midae (perlemoen)	mtDNA, lysin	species identification	(Sweijd *et al.*, 1998a, 2002)
	allozymes, mtDNA	population structure	(Evans *et al.*, 2004b)
	microsatellites	commercial stock assessment	(Evans *et al.*, 2004a)
H. roei (Roe's abalone)	allozymes	population structure	(Hancock, 2000)
H. rubra (blacklip abalone)	RAPD, minisatellite, microsatellites	population structure	(Huang *et al.*, 2000)
	RFLPs, microsatellites	population structure	(Conod *et al.*, 2002)
	microsatellites	commercial stock assessment	(Evans *et al.*, 2004a)
	microsatellites	linkage mapping	(Baranski, 2006; Baranski *et al.*, 2006)
	microsatellites	population structure	(Temby *et al.*, 2007)

H. rufescens (red abalone)	microsatellites allozymes, mtDNA mtDNA, microsatellites, AFLPs	population structure stock enhancement population structure	(Kirby *et al.*, 1998) (Burton and Tegner, 2000) (Gruenthal *et al.*, 2007)
H. sorenseni (white abalone)	VERL gene, mtDNA, microsatellites	Species identification	Gruenthal and Burton, 2005

1.5.5 Conservation status of H. midae and other abalone

Over the past two decades, a severe decline of abalone stocks has been witnessed along the South African coast. It is believed that both anthropogenic and natural processes have contributed to the decline, however, over-exploitation through illegal harvesting and commercial fishing has by far had the greatest effect on dwindling numbers. The exponential increase in the number of abalone confiscated by enforcement authorities every year gives a sound indication of the severity of the problem South Africa has with poaching of perlemoen. In 2002, more abalone had been confiscated from poaching activities than was harvested through commercial fishing that year and since then illegal harvesting has escalated to over 500 tons per year (Tarr, 2003). Due to the slow growth rate of *H. midae* and other abalone, reaching sexual maturity only at the age of eight years old in the wild, these species are particularly vulnerable to poaching and removal of a large fraction of undersized animals from the wild can lead to sudden extinction. Fortunately, if not too late, the South African government has recently (February 2008) prohibited all commercial fishing of wild abalone as well as imposed restrictions on diving in certain areas along the Western Cape coastline. These bans were considered necessary since poaching escalated, despite the compliance efforts of anti-poaching units such as Operations Neptune and Trident. Even though the environmental courts recorded an increase in the conviction of perlemoen-related

offences, it is of personal opinion that the illicit trade of perlemoen will continue as long as a lucrative international trade exists.

Forensic DNA testing can without doubt assist in the prosecution of perlemoen poachers on an individual level (Roodt-Wilding and Bester, 2006) but the highly efficient organized crime networks and well established illicit trade routes between South Africa and East-Asia are conditions that could only be dealt with on an international scale. A hopeful attempt was the listing of *H. midae* in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) on February 5, 2007. Illegal trade is therefore expected to be reduced since all international consignments have to be accompanied by CITES documentation.

In addition to the law enforcement efforts to protect *H. midae* from extinction, the South African aquaculture industry has investigated artificial seeding of juvenile *H. midae* in the wild (ranching) for its potential to alleviate pressures on heavily exploited abalone populations (Sweijd *et al.*, 1998b; Godfrey, 2003; De Waal *et al.*, 2003). The seeding experiments of both Sweijd *et al.* (1998b) and De Waal *et al.* (2003) reported low to average survival rates but confirmed that if the appropriate seeding sites are identified based on ecological conditions, ranching of *H. midae* is a feasible prospect for the enhancement or recovery of depleted stocks.

A few studies have also focused on better understanding the biology of the species which in turn could assist or reinforce conservation of existing abalone populations (Tarr, 1995, 2000; De Waal *et al.*, 2003; Evans *et al.*, 2004b; Proudfoot *et al.*, 2006). Although Evans *et al.* (2004b) suggested the presence of two potentially independent reproductive stocks based on molecular data analysis, none of these former studies resulted in the implementation of obtainable management strategies relevant to the conservation of natural *H. midae* stocks. Such measures

are critical in that even when commercial fishing is banned, the recovery of depleted stocks can not be guaranteed unless remaining stocks are managed in the best way possible (Withler *et al.*, 2000). In particular with species that are commercially produced, the unregulated transfer of larvae and juveniles between farms or the unintentional release of domesticated animals into the wild should be monitored on a regular basis. The potential decrease in genetic variability and fitness of wild populations through the introduction of domesticated animals has been a major concern for the aquaculture industry in general. A number of studies have found a loss of genetic variation in hatchery produced abalone compared to their native stocks (Evans *et al.*, 2004a; Li *et al.*, 2004; Slabbert, 2004; Hara and Sekino, 2007a).

The exploitation and decline in abundance of abalone is a worldwide occurrence (Karpov *et al.*, 2000) and several species have been recognized as vulnerable or endangered (see Table 1.2). *Haliotis sorenseni* or white abalone became the first abalone species as well as the first marine invertebrate species to be listed as an endangered species by the United States Fish and Wildlife Service (USFWS) (Hobday *et al.*, 2001). Commercial and recreational fishing of abalone in South California was banned as early as 1996 and soon thereafter the United States National Marine Fishery Service started investigation into recovery efforts for white abalone specifically. By 2001, two recovery programs were operational in southern California raising thousands of young white abalone from only 20 wild-caught brood stock animals (Gruenthal and Burton, 2005).

Table 1.2 *Haliotis* species with endangered status and their primary threats identified

Species	Distribution	Threats
H. corrugata (pink abalone)	Central California to central Baja California	Illegal harvesting, Predation, Competition, Disease, Climate change
H. cracherodii (black abalone)	Northern California to central Baja California	Disease (Withering Syndrome), Overfishing, Climate change
H. fulgens (green abalone)	Central California to central Baja California	Disease (Withering Syndrome)
H. iris (paua abalone)	New Zealand	Illegal harvesting, High ornamental value
H. kamtschatkana (pinto abalone)	Alaska, Canada to Point Conception, Southern California	Overfishing, Illegal harvesting, Predation, Disease
H. midae (perlemoen)	South Africa	Overfishing, Illegal harvesting, Habitat change, Predation
H. rufescens (red abalone)	Northern California - Central Baja California	Pollution, Increased market value, Predation (sea otters)
H. sorenseni (white abalone)	Central and Southern California	Overfishing, Illegal harvesting

Reseeding of hatchery-reared juveniles has been achieved in several other abalone species (see also review by Roodt-Wilding, 2007) but the overall success in replenishment of wild stocks remains dependent on the availability of suitable habitat, release size, biology of the species, natural predators, and the connectivity among the existing populations (Tegner and Butler, 1989; Lapota *et al.,* 2000; Bell *et al.,* 2005). A few authors have even claimed that despite high survival rates the reseeding of abalone has not contributed to stock rebuilding and that recovery will more likely be attained through maintaining optimal reproduction densities in wild

adult communities (Kawamura *et al.*, 2006; Horii and Kawamura, 2006). This suggests that the maintenance and recovery of wild stocks through genetic enhancement programs has a better chance of success if it is carried out within the framework of a species natural ability to survive.

1.6 Research objectives and dissertation outline

The primary objective of this study was to generate basic information of the micro- and macro-evolutionary processes of the local species, *Haliotis midae,* with the ultimate purpose of its application to the conservation of this species.

For this, two main topics were addressed:

1. The degree of connectivity or gene flow between *H. midae* populations along the species' geographical distribution range. Limited genetic information on this species and both ecological and phylogeographic information of other coastal South African species suggested the possibility of two contrasting scenarios of panmixia or restricted gene flow. The alternative hypotheses of panmixia vs restricted gene flow was investigated by the application of species-specific molecular markers and the most updated and advanced computational methods in population genetics.

2. The phylogenetic position of *H. midae*. The scarce literature on the South African species indicated close affinity to other southern temperate species without resolving the origin of the species from either the Tethyan Sea or the Indo-Pacific. The alternative hypotheses of recent founder dispersal

and ancient vicariance were contrasted by means of phylogenetic reconstruction using DNA sequence information.

The thesis has been organized in six chapters covering literature review, the aims stated above addressed by different experimental approaches and the final conclusions and implications of this study.

Chapter1: Overview of evolutionary genetic processes in marine species with an emphasis on the conservation of *Haliotis midae*

An overview on literature regarding the inference of population genetic structure within marine species while describing the current status of conservation genetics in the South African abalone, *H. midae* and other *Haliotis* species is given.

Chapter 2: Development and characterization of novel molecular markers specifically for *Haliotis midae*

The isolation and characterization of eleven species-specific microsatellite and twenty single nucleotide polymorphic markers for *H. midae* is described. These results have been published in *Molecular Ecology Notes* and *Animal Genetics,* respectively.

Chapter 3: Population structure analysis and demographics of *Haliotis midae* evidenced by microsatellite DNA genotyping

Population structure analysis and demographical history of 428 wild-caught *H. midae* individuals sampled at nine geographically distinct locations is described. The analysis is focused on estimating the degree of connectivity among the sampling sites as well as detecting population size fluctuations using eight microsatellite markers and a combination of traditional and more contemporary analysis methods.

These results together with the results obtained in the next chapter will be prepared for submission to an international peered journal that specializes in conservation genetics.

Chapter 4: Verification of population structure within *H. midae* using Single Nucleotide Polymorphism (SNP) markers.

Twelve SNPs are used to analyse population genetic structure within the same nine *H. midae* populations as in Chapter 3. Data analysis is not as comprehensive as in the previous chapter but informative enough to verify the number of clusters obtained and barriers to gene flow identified using the microsatellite markers.

Chapter 5: Phylogenetic relationships of *Haliotis midae* with other abalone species based on mitochondrial and nuclear DNA sequences.

The phylogenetic reconstruction of eight abalone species is performed, based on DNA fragments amplified from the ND1 mitochondrial gene and the nuclear haemocyanin gene. The previous placement of *H. midae* within the Southern hemisphere group of *Haliotis* species is confirmed while the phylogenetic positioning of the two South African species, *H. spadiceae* and *H. parva* are reported for the first time. Results of this chapter will be submitted for publication to an international peered journal that specializes in molecular phylogenetics.

Chapter 6: Concluding comments on the population dynamics of *Haliotis midae*: implications on future conservation and management strategies

This chapter does not only make final comments about the degree of connectivity among nine *H. midae* wild populations but also makes a proposal towards how this information can be used to better manage and conserve the natural stocks of *H. midae* in the future.

References

Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology* **15:** 1109–1123.

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* **5:** 181–190.

Allendorf FW and Luikart G (2007) *Conservation and the Genetics of Populations*. Blackwell Publishing, Oxford, UK.

Amos W and Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **353:** 177–186.

Anderson EC, Williamson EG, Thompson EA (2000) Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. *Genetics* **156:** 2109–2118.

Apte S and Gardner JPA (2002) Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology* **11:** 1617–1628.

Araki H, Ardren WR, Blouin MS (2007) A potential bias in the temporal method for estimating N_e in admixed populations under natural selection. *Molecular Ecology* **16:** 2261–2271.

Artamonova VS (2007) Genetic markers in population studies of Atlantic Salmon *Salmo salar* L.: Analysis of DNA sequences. *Russian Journal of Genetics* **43:** 341–353.

Austin JD, Lougheed SC, Boag PT (2004) Controlling for the effects of history and nonequilibrium conditions in gene flow estimates in northern bullfrog (*Rana catesbeiana*) populations. *Genetics* **168:** 1491–1506.

Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology* **9:** 686–690.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.

Badenhorst D (2008) Development of AFLP markers for *Haliotis midae* for linkage mapping. M.Sc. thesis, University of Stellenbosch, South Africa, 168pp.

Balloux F and Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology* **11:** 771–783.

Balloux F, Brunner H, Lugon-Moulin N, Hausser J, Goudet J. (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution* **54:** 1414–1422.

Baranski M (2006) Molecular genetic markers for abalone aquaculture. PhD dissertation, Deakin University, Australia, 205pp.

Baranski M, Loughnan S, Austin CM, Robinson N (2006) A microsatellite linkage map of the blacklip abalone, *Haliotis rubra* *Animal Genetics* **37:** 563–570.

Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2002) Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology* **11:** 659–674.

Bartley D, Bagley M, Gall G, Bentley B (1992) Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. *Conservation Biology* **6:** 365–375.

Barton NH and Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113–148.

Baus E, Darrock DJ, Bruford MW (2005) Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star *Asterina gibbosa. Molecular Ecology* **14**: 3373–3382.

Bayes T (1763) Studies in the history of probability and statistics: IX. Thomas Bayes' Essay Towards Solving a Problem in the Doctrine of Chances. *Biometrika* **45:** 296–315.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* **153:** 2013–2029.

Beaumont MA (2003) Estimation of population growth or decline in genetically monitored populations. *Genetics* **164:** 1139–1160.

Beaumont MA and Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* **5:** 251–261.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162:** 2025–2035.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22:** 341–345.

Beerli P and Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent

approach. *Proceedings of the National Academy of Sciences, USA* **98:** 4563–4568.

Bell JD, Rothlisberg PC, Munro JL, Loneragan NR, Nash WJ, Ward RD, Andrew NL (2005) Advances in Marine Biology: Restocking and stock enhancement of marine invertebrate fisheries. San Diego California, Elsevier Academic Press, 375pp.

Berthier P, Beaumont MA, Cornuet JM, Luikart G (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* **160:** 741–751.

Bester AE, Roodt-Wilding R, Whitaker HA (2008) Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae*: a targeted EST approach. *Animal Genetics* **39:** 321–324.

Bester AE, Slabbert R, D'Amato ME (2004) Isolation and characterization of microsatellite markers in the South African abalone (*Haliotis midae*). *Molecular Ecology Notes* **4:** 618–619.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology* **74:** 21–45.

Bowen BW, Bass LE, Soares L, Toonen RJ (2005) Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Molecular Ecology* **14:** 2389–2402.

Branch GM, Branch ML, Griffiths CL, Beckley LE (2002) *Two Oceans: A Guide to the Marine Life of Southern Africa.* David Philip Publishers, Cape Town, 368pp.

Briggs JC (1999) Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution* **53:** 326–335.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18:** 249–256.

Buonaccorsi VP, Kimbrell CA, Lynn EA, Vetter RD (2005) Limited realized dispersal and introgressive hybridization influence genetic structure and conservation strategies for brown rockfish, *Sebastes auriculatus*. *Conservation Genetics* **6:** 697–713.

Buonaccorsi VP, Westerman M, Stannard J, Kimbrell C, Lynn E, Vetter RD (2004) Molecular genetic structure suggests limited larval dispersal in grass rockfish, *Sebastes rastrelliger*. *Marine Biology* **145:** 779–788.

Burnett WJ, Benzie JAH, Beardmore JA, Ryland JS (1994) High genetic variability and patchiness in a common Great Barrier Reef zoanthid (*Palythoa caesia*). *Marine Biology* **121:** 153–160.

Burton RS and Tegner MJ (2000) Enhancement of red abalone *Haliotis rufescens* stocks at San Miguel Island: reassessing a success story. *Marine Ecology Progress Series* **202:** 303–308.

Carling MD and Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLoS ONE* **2:** e160.

Carvalho GR and Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* **4:** 326–350.

Castric V and Bernatchez L (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology* **13:** 1299–1312.

Chambers MD, Van Blaricon GR, Hauser L, Utter F, Friedman CS (2006) Genetic structure of black abalone (*Haliotis cracherodii*) populations in the California islands and central California coast: impacts of larval dispersal and decimation from withering syndrome. *Journal of Experimental Marine Biology and Ecology* **331:** 173–185.

Chan IKK, Chan DKO, Lee SC, Tsukamoto K (1997) Genetic variability of the Japanese eel *Anguilla japonica* related to latitude. *Ecology of Freshwater Fish* **6:** 45–49.

Chappell DE, Van Den Bussche RA, Krizan J, Patterson B (2004) Contrasting levels of genetic differentiation among populations of wolverines (*Gulo gulo*) from northern Canada revealed by nuclear and mitochondrial loci. *Conservation Genetics* **5:** 759–767.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Resources* **7:** 747–756.

Cimmaruta R, Bondanelli P, Nascetti G (2005) Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*). *Molecular Ecology* **14:** 2577–2591.

Coleman AW and Vacquier VD (2002) Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*). *Journal of Molecular Evolution* **54:** 246–257.

Conod NJ, Bartlett P, Evans BS, Elliott N (2002) Comparison of mitochondrial and nuclear DNA analyses of population structure in the blacklip abalone *Haliotis rubra* Leach. *Marine and Freshwater Research* **53:** 711–718.

Corander J, Waldmann P, Sillanpää M (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163:** 367–374.

Cornuet JM and Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144:** 2001–2014.

Corujo M, Blanco G, Vazquez E, Sanchez JA (2004) Genetic structure of northwestern Spanish brown trout (*Salmo trutta* L.) populations, differences between microsatellite and allozyme loci. *Hereditas* **141:** 258–271.

Cowen RK, Lwiza KMM, Sponaugle S, Paris CB, Olson DB (2000) Connectivity of marine populations: open or closed? *Science* **287:** 857–859.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **17:** 390–395.

Crandall KA, Posada D, Vasco D (1999) Effective population sizes: missing measures and missing concepts. *Animal Conservation* **2:** 317–319.

Cross TF (1999) Genetic considerations in enhancement and ranching of marine and anadromous species. In: *Stock Enhancement and Sea Ranching* (ed. Howell BR, Moksness E and Svasand T), Fishing News Books, Oxford, pp37–48.

D'Amato ME and Carvalho GR (2005) Population genetic structure and history of the long-tailed hake, *Macruronus magellanicus*, in the SW Atlantic as revealed by mtDNA RFLP analysis. *ICES Journal of Marine Science* **62:** 247–265.

David P, Perdieu MA, Pernot AF, Jarne P (1997) Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution* **51:** 1318–1322.

Dawson K and Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research* **78:** 59–77.

De Innocentiis S, Sola L, Cataudella S, Bentzen P (2001) Allozyme and microsatellite loci provided discordant estimates of population differentiation in the endangered dusky grouper (*Epinephelus marginatus*) within the Mediterranean Sea. *Molecular Ecology* **10:** 2163–2175.

De Waal SWP, Branch GM, Navarro R (2003) Interpreting evidence of dispersal by *Haliotis midae* juveniles seeded in the wild. *Aquaculture* **221:** 299–310.

Degnan SM, Imron, Geiger DL, Degnan BM (2006) Evolution in temperate and tropical seas: disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae*). Molecular Phylogenetics and Evolution* **41:** 249–256.

Dingle H (1996) *Migration: The Biology of Life on the Move*. Oxford University Press, New York, 486pp.

Diniz FM, Iyengar A, Da Costa Lima PS, Maclean N, Bentzen P (2007) Application of a double-enrichment procedure for microsatellite isolation and the use of tailed primers for high throughput genotyping. *Genetics and Molecular Biology* **30:** 380–384.

Durrett R, Buttel L, Harrison R (2000) Spatial models for hybrid zones. *Heredity* **84:** 9–19.

Edmands S, Moberg PE, Burton RS (1996) Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. *Marine Biology* **126:** 443–450.

Edwards AWF (1974) The history of likelihood. *International Statistical Reviews* **42:** 9–15.

Emanuel BP, Bustamante RH, Branch GM, Eekhout S, Odendaal FJ (1992) A zoogeographic and functional approach to the selection of marine reserves on the West coast of South Africa. *South African Journal of Marine Science* **12:** 341–354.

England PR, Cornuet JM, Berthier P, Tallmon DA, Luikart G (2006) Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conservation Genetics* **7:** 303–308.

Estes JA, Lindberg DR, Wray C (2005) Evolution of large body size in abalone (*Haliotis*): patterns and implications*. Paleobiology* **31:** 591–606.

Evans BS, Bartlett J, Sweijd N, Cook P, Elliott NG (2004a) Loss of genetic variation at microsatellite loci in hatchery produced abalone in Australia (*Haliotis rubra*) and South Africa (*Haliotis midae*). *Aquaculture* **233:** 109–127.

Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliott NG (2004b) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology Progress Series* **270:** 163–172.

Excoffier l and Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Review Genetics* **7:** 745–757.

Flores JA, Sierro FJ, Francés G, Vásquez A, Zamarreño I (1997) The last 100 000 years in the western Mediterranean: sea surface water and frontal dynamics as revealed by coccolithophores. *Marine Micropaleontology* **29:** 351–366.

François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random field in spatial population genetics. *Genetics* **174**: 805–816.

Frankham R (1995) Conservation genetics. *Annual Review of Genetics* **29:** 305–327.

Frankham R, Ballou JD, Briscoe DA (2003) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK, 642pp.

Fraser DJ and Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology* **10:** 2741–2752.

Fraser DJ and Bernatchez L (2005) Allopatric origins of sympatric brook charr populations: colonization history and admixture. *Molecular Ecology* **14:** 1497–1509.

Garza C and Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10:** 305–318.

Geiger DL and Groves LT (1999) Review of fossil abalone (Gastropoda: Vetigastropoda: Haliotidae) with comparison to recent species. *Journal of Paleontology* **73:** 872–885.

Gerber LR, Botsford LW, Hastings A, Possingham HP, Gaines SD, Palumbi SR, Andelman S (2003) Population models for marine reserve design: A retrospective and prospective synthesis. *Ecological Applications* **13:** 47–65.

Gilbert-Horvath EA, Larson RJ, Garza JC (2006) Temporal recruitment patterns and gene flow in kelp rockfish (*Sebastes atrovirens*) *Molecular Ecology* **5:** 3801–3815.

Godfrey B (2003) The potential of abalone stock enhancement in the Eastern Cape province of South Africa. M.Sc. thesis, Rhodes University, South Africa, 163pp.

González-Villaseñor LI and Powers DA (1990) Mitochondrial-DNA restriction-site polymorphisms in the teleost *Fundulus heteroclitus* support secondary intergradation. *Evolution* **44:** 27–37.

Goodman SJ (1997) A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* **6:** 881–885.

Gruenthal KM and Burton RS (2005) Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conservation Genetics* **6:** 929–939.

Gruenthal KM, Acheson LK, Burton RS (2007) Genetic structure of natural populations of California red abalone (*Haliotis rufescens*) using multiple genetic markers. *Marine Biology* **152:** 1237–1248.

Guillot G, Estoup A, Mortier F, Cosson J (2005) A spatial statistical model for landscape genetics. *Genetics* **170:** 1261–1280.

Gutiérrez-Gonzales JL, Cruz P, del Rio-Portilla MA, Perez-Enriquez R (2007) Genetic structure of green abalone *Haliotis fulgens* population off Baja California, Mexico. *Journal of Shellfish Research* **26:** 839–846.

Hamm DE and Burton RS (2000) Population genetics of black abalone, *Haliotis cracherodii*, along the central California coast. *Journal of Experimental Marine Biology and Ecology* **254:** 235–247.

Hancock B (2000) Genetic subdivision of Roe's abalone, *Haliotis roei* Grey (Mollusca: Gastropoda), in South-Western Australia. *Marine and Freshwater Research* **51:** 679–687.

Hanfling B and Weetman D (2006) Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics* **173:** 1487–1501.

Hansen B and Østerhus S (2000) North Atlantic-Nordic Seas exchanges. *Progress in Oceanography* **45:** 109–208.

Hara M and Sekino M (2007a) Genetic differences between hatchery stocks and natural populations in Pacific abalone (*Haliotis discus*) estimated using microsatellite DNA markers. *Marine Biotechnology* **9:** 74–81.

Hara M and Sekino M (2007b) Parentage testing for hatchery-produced abalone *Haliotis discus hannai* based on microsatellite markers: preliminary evaluation of early growth of selected strains in mixed family farming. *Fisheries Science* **73:** 831–836.

Hara M, Sekino M, Kumagai A, Yoshinaga T (2004). The identification of genetic resistance to amyotrophia in Japanese abalone, *Haliotis discus discus. Journal of Shellfish Research* **23:** 1157–1162.

Hardy G (1908) Mendelian proportions in a mixed population. *Science* **28**: 49–50.

Hare MP and Avise JC (1996) Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution* **50:** 2305–2315.

Harpending H and Rogers AR (2000) Genetic perspectives on human origins and differentiation. *Annual Reviews of Genomics and Human Genetics* **1:** 361–385.

Hedgecock D (2000) Population genetics of marine organisms. In: *Encyclopedia of Ocean Sciences* (ed. Steele JH, Thorpe S and Turekian K), Academic Press, London, pp2256–2263.

Hedgecock D, Hutchinson ES, Li G, Sly FL, Nelson K (1994) The central stock of northern anchovy (*Engraulis mordax*) is not a randomly mating population. *California Cooperative Oceanic Fisheries Investigations Reports* **35:** 121–136.

Hellberg ME (1994) Relationships between inferred levels of gene flow and geographic distance in a philopatric coral *Balanophyllia elegans. Evolution* **48:** 1829–1854.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70:** 273–290.

Hemmer-Hansen J, Nielsen EEG, Gronkjaer P, Loeschcke V (2007) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Molecular Ecology* **16:** 3104–3118.

Hendry AP and Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology* **14:** 901–916.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* **38:** 209–216.

Hoarau G, Piquet AMT, van der Veer HW, Stam WT, Olsen JL (2004) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: a comparison of resolving power between microsatellites and mitochondrial DNA data. *Journal of Sea Research* **51:** 183–190.

Hobday AJ, Tegner MJ, Haaker PL (2001) Over-exploitation of a broadcast spawning marine invertebrate: decline of the white abalone. *Reviews in Fish Biology and Fisheries* **10:** 493–514.

Horii T and Kawamura T (2006) How do we recover abalone resources successfully by reseedings? Paper presented during the Third International Symposium on Stock Enhancement and Sea Ranching, Seattle, WA, 18-21 September 2006.

Huang BX, Peakall R, Hanna PJ (2000) Analysis of genetic structure of blacklip abalone (*Haliotis rubra*) populations using RAPD, minisatellite and microsatellite markers. *Marine Biology* **136:** 207–216.

Hutchinson DW and Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53:** 1898–1914.

Joensen H, Steingrund P, Fjallstein I, Grahl-Nielsen O (2000) Discrimination between two reared stocks of cod (*Gadus morhua*) from the Faroe Islands by chemometry of the fatty acid composition in the heart tissue. *Marine Biology* **136:** 573–580.

Johnson MS and Black R (1982) Chaotic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* **70:** 157–164.

Johnson MS and Black R (1984) Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* **38:** 1371–1383.

Johnson MS and Black R (1998) Increased genetic divergence and reduced genetic variation in populations of the snail *Bembicium vittatum* in isolated tidal ponds. *Heredity* **80:** 163–172.

Johnson MS and Black R (2005) Effects of a larval stage on isolation of populations of the trochid *Austrocochlea constricta* in tidal ponds. *Marine Biology* **147:** 189–195.

José J and Solferini VN (2007) Population genetics of *Collisella subrugosa* (Patellogastropoda: Acmaeidae): evidence of two scales of population structure. *Genetica* **130:** 73–83.

Karpov KA, Haaker PL, Tanguchi IK, Rogers-Bennett L (2000) Serial depletion and the collapse of the California abalone (*Haliotis* spp.) fishery. *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 11–24.

Kawamura T, Horii T, Takami H (2006) Have reseedings contributed to the stock enhancement of abalone in Japan? Paper Presented During the Third International Symposium on Stock Enhancement and Sea Ranching, Seattle, WA, 18-21 September 2006.

Kenchington E and Heino M (2002) Maintenance of genetic diversity: challenges for management of marine resources. International Council for the Exploration of the Sea (ICES) CM 2002/Y: 13.

Kenchington E, Heino M, Nielsen EE (2003) Managing marine genetic diversity: time for action? *ICES Journal of Marine Science* **60:** 1172–1176.

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology* **15:** 1781–1796.

Kimmel M and Chakraborty R (1996) Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theoretical and Population Biology* **50:** 345–367.

Kimura M and Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49:** 561–576.

Kirby VL, Villa R, Powers DA (1998) Identification of microsatellites in the California red abalone, *Haliotis rufescens*. *Journal of Shellfish Research* **17:** 801–804.

Klinbunga S, Amparyup P, Leelatanawit R, Tassanakajon A, Hirono I, Aoki T, Jarayabhand A, Menasveta P (2004) Species-identification of the tropical abalone *Haliotis asinina*, *Haliotis ovina* and *Haliotis varia* in Thailand using RAPD and SCAR markers. *Journal of Biochemistry and Molecular Biology* **37:** 213–222.

Klinbunga S, Pripue SP, Khamnamtong N, Puanglarp N, Tassanakajon A, Jarayabhand P, Hirono I, Aoki T, Menasveta P (2003) Genetic diversity and molecular markers of the tropical abalone (*Haliotis asinina*) in Thailand. *Marine Biotechnology* **5:** 493–504.

Knowlton N and Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B, Biological Science* **265:** 2257–2263.

Knowlton N, Weigt LA, Solorzano D, Mills K, Bermingham E (1993) Divergence in proteins, mitochondrial DNA and reproductive compatibility across the Isthmus of Panama. *Science* **260:** 1629–1632.

Koehn RK and Williams GC (1978) Genetic differentiation without isolation in American Eel, *Anguilla rostrata*. Temporal stability of geographic patterns. *Evolution* **32:** 624–637.

Koehn RK, Hall JG, Zera AJ (1980) Parallel variation of genotype-dependent aminopeptidase-I activity between *Mytilus edulis* and *Mercenaria mercenaria*. *Marine Biology Letters* **1:** 245–254.

Kordos LM and Burton RS (1993) Genetic differentiation of Texas Gulf coast populations of the blue crab *Callinectes sapidus*. *Marine Biology* **49:** 363–375.

Kozfkay CC, Campbell MR, Heindel JA, Baker DJ, Kline P, Powell MS, Flagg T (2008) A genetic evaluation of relatedness for broodstock management of captive, endangered Snake River sockeye salmon, *Oncorhynchus nerka. Conservation Genetics (in press)*

Kuhner MK (2006) Lamarc 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22:** 768–770.

Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* **156:** 439–447.

Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140:** 1421–1430.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149:** 429–434.

Lambert WJ, Todd CD, Thorpe JP (2003) Genetic population structure of two intertidal nudibranch molluscs with contrasting larval types: Temporal variation and transplant experiments. *Marine Biology* **142:** 461–471.

Lapota D, Rosen G, Chock J, Liu CH (2000) Red and green abalone seed grow out for reseeding activities off Point Loma, California. *Journal of Shellfish Research* **19:** 431–438.

Larson RJ and Julian RM (1999) Spatial and temporal genetic patchiness in marine populations and their implications for fisheries management. *California Cooperative Oceanic Fisheries Investigations Reports* **40:** 94–99.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE Jr (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7:** 295–302.

Laval G, SanCristobal M, Chevalet C (2003) Maximum-Likelihood and Markov chain Monte Carlo approaches to estimate inbreeding and effective size from allele frequency changes. *Genetics* **164:** 1189–1204.

Leary RF, Allendorf FW, Sage GK (1995) Hybridization and introgression between introduced and native fish. In: *Uses and Effects of Cultured Fishes in Aquatic Ecosystems*. American Fisheries Society Symposium **15:** 91–101.

Leighton DL (2000) *The Biology and Culture of the California Abalones*. Pittsburgh, PA, Dorrance Publishing Co Inc., 216pp.

Li G and Hedgecock D (1998) Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. *Canadian Journal of Fisheries and Aquatic Sciences* **55:** 1025–1033.

Li Q, Park C, Endo T, Kijima A (2004) Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (*Haliotis discus hannai*). *Aquaculture* **235:** 207–222.

Li Q, Park C, Kijima A (2003a). Allelic transmission of microsatellites and application to kinship analysis in newly hatched abalone larvae. *Fisheries Science* **69:** 883–889.

Li Q, Park C, Kobayashi T, Kijima A (2003b) Inheritance of microsatellite DNA markers in the Pacific abalone *Haliotis discus hannai*. *Marine Biotechnology* **5:** 331–338.

Li Q, Shu J, Yu R, Tian C (2007) Genetic variability of cultured populations of the Pacific abalone (*Haliotis discus hannai* Ino) in China based on microsatellites. *Aquaculture Research* **38:** 981–990.

Lindberg DR (1992) Evolution, distribution and systematics of Haliotidae. In: *Abalone of the world: biology, fisheries and culture* (ed. Shepherd SA, Tegner M, and Guzmán del Próo SA), Blackwell Scientific, Oxford, pp3–18.

Linnaeus C (1758) *Systema Naturae. Edition Decima, Reformata* (The System of Nature, Tenth Revised Edition), 824pp.

Liu X, Liu X, Guo X, Gao G, Zhang G (2006) A preliminary genetic linkage map of the Pacific abalone *Haliotis discus hannai* Ino. *Marine Biotechnology* **8:** 386–397.

Longhurst A (1998) Ecological geography of the sea. Academic Press, London, 398pp.

Luikart G and Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12:** 228–233.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89:** 238–247.

Luikart GH, Cornuet JM, Allendorf FW (1999) Temporal changes in allele frequencies provide estimators of population bottleneck size. *Conservation Biology* **13:** 523–530.

Lynch M, Conery J, Burger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist* **146:** 489–518.

Maes GE and Volckaert FAM (2002) Clinal genetic variation and isolation by distance in the European eel *Anguilla anguilla*. *Biological Journal of the Linnean Society* **77:** 509–521.

Maes GE, Pujolar JM, Hellemans B, Volckaert FAM (2006) Evidence for isolation by time in the European eel (*Anquilla anguilla* L.). *Molecular Ecology* **15:** 2095–2107.

Marjoram P and Tavaré S (2006) Modern computational approaches for analyzing molecular genetic variation data. *Nature Reviews Genetics* **7:** 759–769.

Marko PB (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution* **19:** 2005–2021.

Maynard BJ, Benzie H, Hanna PJ (2004) Microsatellite DNA analysis of southeast Australian *Haliotis laevigata* (Donovan) populations: implications for ranching in Port Phillip Bay. *Journal of Shellfish Research* **23:** 1195–2000.

Miller MP, Haig SM, Wagner RS (2005) Conflicting patterns of genetic structure produced by nuclear and mitochondrial markers in the Oregon slender salamander (*Batrachoseps wrighti*): Implications for conservation efforts and species management. *Conservation Genetics* **6:** 275–287.

Moberg PE and Burton RS (2000) Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. *Marine Biology* **136:** 773–784.

Morin PA, Luikart G, Wayne RK and the SNP workshop group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19:** 208–216.

Moritz C (1994) Defining “Evolutionarily Significant Units” for conservation. *Trends in Ecology and Evolution* **9:** 373–375.

Mustafa S (2003) Stock enhancement and sea ranching: objectives and potential. *Reviews in Fish Biology and Fisheries* **13:** 141–149.

Nagylaki T (1982) Geographical invariance in population genetics. *Journal of Theoretical Biology* **99:** 159–172.

Nei M and Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics* **98:** 625–640.

Newman GG (1968) Growth of the South African abalone, *Haliotis midae*. *Investigational Report Division of Sea Fisheries South Africa* **67:** 1–24.

Nielsen EE and Kenchington E (2001) Prioritising marine fish and shellfish populations for conservation: a useful concept? *Fish and Fisheries* **2:** 328–343.

Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Gronkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish belt sea revealed by individual admixture analysis. *Molecular Ecology* **12:** 1497–1508.

Nielsen EE, Nielsen PH, Meldrup D, Hansen MM (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. *Molecular Ecology* **13:** 585–595.

Nielsen R (1997) A likelihood approach to population samples of microsatellite alleles. *Genetics* **146:** 711–716.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154:** 931–942.

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye Pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology* **13:** 1799–1814.

O'Brien SJ (1991) Molecular genome mapping lessons and prospects. *Current Opinion in Genetics and Development* **1:** 105–111.

Olsen JB, Merkouris SE, Seeb JE (2002) An examination of spatial and temporal genetic variation in walleye Pollock (*Theragra chalcogramma*) using allozyme, mitochondrial DNA, and microsatellite data. *Fishery Bulletin US* **100:** 752–764.

Olsen JB, Habicht C, Reynolds J, Seeb JE (2004a) Moderate and highly polymorphic microsatellites provide discordant estimates of population divergence in sockeye salmon (*Oncorhynchus nerka*). *Environmental Biology of Fishes* **69:** 261–273.

Olsen JL, Stam WT, Coyer JA *et al.,* (2004b) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology* **13:** 1923–1941.

Paetkau D (1999) Using genetics to identify intraspecific conservation units: a critique of current methods. *Conservation Biology* **13:** 1507–1509.

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in Ecology and Evolution* **22:** 11–16.

Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history *Molecular Ecology* **16:** 4504–4522.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13:** 146–158.

Palumbi SR, Grabowsky G, Duda T, Geyer L, Tachino N (1997) Speciation and population genetic structure in tropical Pacific Sea urchins. *Evolution* **51:** 1506–1517.

Pearse DE and Crandall KA (2004) Beyond F_{ST}: Analysis of population genetic data for conservation. *Conservation Genetics* **5:** 585–602.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90:** 502–503.

Planes S and 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.

Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* **55:** 131–146.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Proudfoot L, Kaehler S, Mcgarry DK, Uppink PA, Aereboe M, Morris KM (2006) Exploitation status of infralittoral abalone (*Haliotis midae*) and alikreukel (*Turbo*

sarmaticus) in the southern section of the Eastern Cape coast, South Africa, *South African Journal of Science* **102:** 162–168.

Pujolar JM, Maes GE, Volckaert FAM (2006) Genetic patchiness among recruits in the European eel, *Anguilla anguilla*. *Marine Ecology Progress Series* **307:** 209–217.

Purcell JFH, Cowen RK, Hughes CR, Williams DA (2006) Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **273:** 1483–1490.

Rabassa J, Coronato AM, Salemme M (2005) Chronology of the Late Cenozoic Patagonian glaciations and their correlation with biostratigraphic units of the Pampean region (Argentina). *Journal of South American Earth Sciences* **20:** 81–103.

Ramstad KM, Woody CA, Sage GK, Allendorf FW (2004) Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Molecular Ecology* **13:** 277–290.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* **20:** 76–86.

Reed DH and Frankham R (2003) The correlation between population fitness and genetic diversity. *Conservation Biology* **17:** 230–237.

Reich DE and Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences, USA* **95:** 8119–8123.

Reich DE, Feldman MW, Goldstein DB (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions. *Molecular Biology and Evolution* **16:** 453–466.

Rengmark AH, Sletten A, Skaala O, Lie O, Lingaas F (2006) Genetic variability in wild and farmed Atlantic Salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture* **253:** 229–237.

Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite flanking regions: A SNP mine in South African abalone (*Haliotis midae*). *Animal Genetics* **39:** 329.

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

Rögl F and Steininger FF (1983) Vom Zerfall der Tethys zu Mediterran und Paratethys. *Annalen des Naturhistorischen Museums in Wien* **85:** 135–163.

Roman J and Palumbi SR (2004) A global invader at home: population structure of the green crab, *Carcinus maenas. Europe Molecular Ecology* **13:** 2891–2898.

Roodt-Wilding R (2007) Abalone ranching: a review on genetic considerations. *Aquaculture Research* **38**: 1229–1241.

Roodt-Wilding R and Bester AE (2006) The forensic fight against Perlemoen poaching. *Fisheries News* **31:** 214–215.

Ropson IJ, Brown DC, Powers DA (1990) Biochemical genetics of *Fundulus heteroclitus* (L.). Geographical variation in the gene frequencies of 15 loci. *Evolution* **44:** 16–26.

Rose CG, Paynter KT, Hare MP (2006) Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity* **97:** 158–170.

Rosenberg N, Mahajan S, Ramachandran S, Zhao C, Pritchard J, Feldman M (2005) Clines, clusters, and the effect of study design on the influence of human population structure. *PLoS Genetics* **1:** e70.

Ruvinsky A and Graves JMA (2005) *Mammalian Genomics* (ed. Ruvinsky A and Graves JMA), CAB International, Wallingford, UK, 600pp.

Ruzzante DE, Taggart CT, Cook C, Goddard S (1996) Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: Microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Sciences* **53:** 634–645.

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

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* **14:** 503–511.

Sekino M and Hara M (2007a) Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus *Haliotis*). *Conservation Genetics* **8:** 823–841.

Sekino M and Hara M (2007b) Linkage maps for the Pacific abalone (genus *Haliotis*) based on microsatellite DNA markers. *Genetics* **175:** 945–958.

Sekino M, Kobayashi T, Hara M (2006) Segregation and linkage analysis of 75 novel microsatellite DNA markers in pair crosses of Japanese abalone (*Haliotis*

discus hannai) using the 5'-tailed primer method. *Marine Biotechnology* **8:** 453–466.

Selvamani MJP, Degnan SM, Degnan BM (2001). Microsatellite genotyping of individual abalone larvae: parentage assignment in aquaculture. *Marine Biotechnology* **3:** 478–485.

Serre D and Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research* **14:** 1679–1685.

Singh SM and Green RH (1984) Excess of allozyme homozygosity in marine molluscs and its possible biological significance. *Malacologia* **25:** 569–581.

Slabbert R (2004) Molecular analysis of genetic variation and relationships within the population of abalone (*Haliotis midae*) at the Sea Plant Products abalone hatchery, Hermanus, RSA. M.Sc. thesis, University of Stellenbosch, South Africa, 162pp.

Slabbert R, Ruivo NR, Van den Berg NC, Lizamore DL, Roodt-Wilding R (2008) Isolation and characterisation of 63 microsatellite loci for the abalone, *Haliotis midae. Journal of the World Aquaculture Society* **39:** 429–435.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457–462.

Sokolova IM, Granovitch AI, Sergievsky SO (1997) Shell colour polymorphism in *Littorina obtusata* and *L. saxatilis* populations along a salinity gradient in the White Sea estuary. In: *The responses of marine organisms to their environment* (ed. Hawkins LE and Hutchinson S), University of Southampton, pp57–65.

Sønstebø JH, Borgstrøm R, Heun MM (2007) A comparison of AFLPs and microsatellites to identify the population structure of brown trout (*Salmo trutta* L.) populations from Hardangervidda, Norway. *Molecular Ecology* **16:** 1427–1438.

Storz JF (2002) Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Molecular Ecology* **11:** 2537–2551.

Streit K, Geiger DL, Lieb B (2006) Molecular phylogeny and the geographic origin of Haliotidae traced by haemocyanin sequences. *Journal of Molluscan Studies* **72:** 111–116.

Sweijd NA (2000) Molecular markers and Abalone seeding as tools for the conservation and management of the South African Abalone (Perlemoen), *Haliotis midae* (Linn.) resource. Ph.D. thesis, University of Cape Town, South Africa, 265pp.

Sweijd NA, Bowie RCK, Lopata AL, Marinaki AM, Harley E, Cook, PA (1998a) A PCR technique for forensic, species-level identification of abalone tissue. *Journal of Shellfish Research* **17:** 889–895.

Sweijd NA, Snethlage Q, Harvey D, Cook P (1998b) Experimental abalone (*Haliotis midae*) seeding in South Africa. *Journal of Shellfish Research* **17:** 897–904.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) ONESAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources* **8:** 299–301.

Tallmon DA, Luikart G, Beaumont MA (2004) Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics* **167:** 977–988.

Talmadge RR (1963) Insular haliotids in the western Pacific. *The Veliger* **5:** 129–139.

Tang S, Popongviwat A, Klinbunga S, Tassanakajon A, Jarayabhand P, Menasveta P (2005) Genetic heterogeneity of the tropical abalone (*Haliotis asinina*) revealed by RAPD and microsatellite analysis. *Journal of Biochemistry and Molecular Biology* **38:** 182–190.

Tang S, Tassanakajon A, Klinbunga S, Jarayabhand P, Menasveta P (2004) Population structure of tropical Abalone (*Haliotis asinina*) in coastal waters of Thailand determined using microsatellite markers. *Marine Biotechnology* **6**: 604–611.

Tarr RJQ (1995) Growth and movement of the South African abalone, *Haliotis midae*: a reassessment. *Marine Freshwater Research* **46:** 583–590.

Tarr RJQ (2000) The South African abalone (*Haliotis midae*) fishery: A decade of challenges and change. In: *Workshop on Rebuilding Abalone Stocks in British Colombia* (ed. Cambell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 32–40.

R Tarr (2003) *Perlemoen: The South African Abalone,* http://www.environment.gov.za, viewed in November 2007.

Tegner MJ and Butler RA (1989) Abalone seeding. In: *Handbook of Culture of Abalone and other Marine Gastropods* (ed. Hahn KO), CRC Press, Boca Raton, FL, USA, pp157–182.

Teske PR, McQuaid CD, Froneman CD, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology Progress Series* **314:** 283–293.

Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics* **162:** 1329–1339.

Uthicke JA and Benzie H (2003) Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology* **12:** 2635–2648.

Van Oppen MJH and Gates RD (2006) Conservation genetics and the resilience of reef-building corals. *Molecular Ecology* **15:** 3863–3883.

Vitalis R and Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* **157:** 911–925.

Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetical Research* **78:** 243–257.

Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society Biological Sciences* **360:** 1395–1409.

Wang J and Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163:** 429–446.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121:** 379–391.

Waples RS (1991) Genetic methods for estimating the effective size of cetacean populations. *Report of the International Whaling Commission (Special Issue)* **13:** 279–300.

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 (2002) Definition and estimation of effective population size in the conservation of endangered species. In: *Population Viability Analysis* (ed. Beissinger SR and McCullough DR), University of Chicago Press, Chicago, pp147–168.

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

Ward RD (2006) The importance of identifying spatial population structure in restocking and stock enhancement programmes. *Fisheries Research* **80:** 9–18.

Waters JM and Roy MS (2003) Global phylogeography of the fissiparous sea-star *Coscina sterias. Marine Biology* **142:** 185–191.

Watts RJ, Johnson MS, Black R (1990) Effects of recruitment on genetic patchiness in the urchin *Echinometra mathaei* in western Australia. *Marine Biology* **105:** 145–151.

Weinberg W (1908) On the demonstration of heredity in man. *Naturkunde in Wurttemberg, Stuttgart* **64:** 368–382.

White BN (1994) Vicariance biogeography of the open-ocean Pacific. *Progress in Oceanography* **34:** 257–284.

Widmer A and Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology and Evolution* **16:** 267–269.

Wilson GA and Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163:** 1177–1191.

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society. Series A, Statistics in Society* **166:** 155–188.

Withler RE (2000) Genetic tools for identification and conservation of exploited abalone (*Haliotis* spp.) species. In: *Workshop on Rebuilding Abalone Stocks in British Colombia* (ed. Cambell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 101–110.

Wood AD (1993) Aspects of the biology and ecology of the South African abalone *Haliotis midae* Linnaeus, 1758 (Mollusca: Gastropoda) along the Eastern Cape and Ciskei coast. M.Sc. Rhodes University, South Africa, 161pp.

Wright S (1940) Breeding structure of populations in relation to speciation. *American Naturalist* **74:** 232–248.

Wright S (1943) Isolation by distance. *Genetics* **28:** 114–138.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **15:** 323–354.

Xu H and Fu YX (2004). Estimating effective population size or mutation rate with microsatellites. *Genetics* **166:** 555–563.

Zardi GI, McQuaid CD, Teske PR, Barker NP (2007) Unexpected genetic structure of indigenous (*Perna perna*) and invasive (*Mytilus galloprovincialis*) mussel populations in South Africa. *Marine Ecology Progress Series* **337:** 135–144.

Zuñiga G, Guzmán del Próo SA, Cisneros R, Rodríguez G (2000). Population genetic analysis of the abalone *Haliotis fulgens* (Mollusk: Gastropoda) in Baja California, Mexico. *Journal of Shellfish Research* **19:** 853–859.

Chapter II

Development and characterization of novel molecular markers specifically for *Haliotis midae*

2.1 Introduction

Molecular markers, revealing various forms of polymorphisms at the DNA level, have become indispensable for the assessment of genetic diversity within and between natural populations. As a number of different marker types exist, the choice of molecular marker in a particular study depends mostly on the availability, polymorphic content and ease or cost-effectiveness of genotyping. Ultimately, the type of information it provides at a single locus will determine the effectiveness and reliability of the data obtained from that particular marker (Vignal *et al.,* 2002). In population and conservation genetic analysis, both microsatellites and single nucleotide polymorphisms (SNPs) have successfully been used for the estimation of gene flow, population size fluctuations, effective population sizes and assignment of unknown individuals to populations of origin (see reviews by Christiakov *et al.,* 2006 and Morin *et al.,* 2004). Microsatellites are characterised as multi-allelic co-dominant markers and consist of short tandem repeats of one to six nucleotide bases while SNPs are single base pair changes in a DNA sequence representing bi-allelic co-dominant markers.

Despite certain advantages and disadvantages of both these marker systems (see chapter I), their usefulness and efficiency are well portrayed in literature and the only real hindrance is the number and quality of such markers available for a

particular species at the beginning of a genetic study. Most of the time microsatellites have to be isolated *de novo* due to lack of conserved flanking sequences and poor cross-amplification between species (Zane *et al.,* 2002) and therefore several methods have been developed and described for the isolation of microsatellite markers. Protocols range from the traditional colony hybridisation with repeat containing probes to enriched methods where only a selected number of clones are screened for microsatellites. Since the implementation of enrichment protocols, which enabled time- and cost-efficient isolation of a large number of repeat containing sequences, microsatellites have been developed for a number of abalone species including *H. asinina* (Selvamani *et al.,* 2000); *H. kamtschatkana* (Miller *et al.,* 2001); *H. rubra* (Evans *et al.,* 2000; Baranski *et al.,* 2006b); *H. discus discus* (Sekino and Hara, 2001) and *H. fulgens* (Cruz *et al.,* 2005).

For SNPs numerous isolation strategies have also been reported with the most traditional method being the comparison of targeted sequences from multiple individuals. While sometimes costly and time-consuming, the advantage of this approach lies within the direct SNP identification through comparing already available sequence data from EST construction projects or overlapping BAC clones (Sachidanandam *et al.,* 2001; Heaton *et al.,* 2002). With a lack of available sequence data for non-model organisms, methods such as whole-genome shotgun sequencing (Weber and Myers, 1997), reduced representation shotgun (Altshuler *et al.,* 2000) and more recently SNPs by AFLPs (Nicod and Largiadèr, 2003) have been developed for the identification of a high number of SNPs.

Given that neither microsatellites nor SNPs were available specifically for *H. midae* at the onset of the project, this section of the dissertation was aimed at isolating and characterising novel molecular markers for the species. Results have

been published in Molecular Ecology Notes (Appendix 1) and Animal Genetics (Appendix 2) respectively. A selection of these markers was employed to characterise genetic diversity within and between populations of *H. midae*.

2.2 Materials and Methods

2.2.1 Isolation of Microsatellites

Microsatellite repeat sequences were isolated following the Fast Isolation by AFLP of Sequences Containing Repeats (FIASCO) technique described by Zane *et al.* (2002). This method is based on a digestion-ligation reaction of the amplified fragment length polymorphism (AFLP) procedure of Vos *et al.* (1995). An enriched partial genomic library was constructed using genomic DNA from a single mature animal. Genomic DNA was isolated from mantle tissue following a standard CTAB extraction protocol (Saghai-Maroof *et al.,* 1984). Tissue was homogenized in 700µl of CTAB lysis buffer containing 2% CTAB, 1.4M NaCl, 0.2% β-mercapto-ethanol, 20mM Ethylene Diamine Tetra-Acetate (EDTA, pH 8) and 100mM Tris-HCl, pH 8. A final concentration of 0.5mg/ml Proteinase K was added and the mixture was incubated overnight at 60℃. Following two phenol-chloroform :isoamyl alcohol (25:24:1) extractions, the supernatant was precipitated with two volumes of 100% cold ethanol. DNA was redissolved in 100µl of distilled water and stored at -20℃.

Approximately 250ng of genomic DNA was simultaneously digested with *MseI* restriction enzyme and ligated to *MseI* AFLP adaptors (5’ TAC TCA GGA CTC AT 3’/ 5’ GAC GAT GAG TCC TGA G 3’). Digestion-ligation was performed in a heating block at 37℃ for three hours in a total volume of 25µl containing 250ng of DNA, 1x OnePhorAll buffer, 5mM DTT, 50µg/ml BSA, 1µM *MseI* adaptor, 200µM ATP, 2.5

units MseI and 1 unit T4 DNA ligase. The digestion-ligation mixture was diluted 10-fold and selectively amplified using a mixture of four adaptor specific primers (5' CAT GAG TCC TGA GTA AN 3' or *MseI*-N) and PCR conditions exactly as described in Zane *et al.,* 2002). Amplification products were assessed on a 1.5% (m/v) agarose gel. After optimization of the number of cycles, PCR products were used as a template for hybridization to the biotinylated $(AC)_{12}$ and $(GATC)_6$ probes respectively. Hybridisation was performed according to the Travis Glen protocol (http://129.252.89.20/Msats/Microsatellites.html). Repeat-containing fragments were captured by streptavidin magnetic particles while the non-specific DNA was removed by three stringency and three non-stringency washes using TEN_{1000} buffer (10mM Tris-HCI, 1mM EDTA, 1M NaCl pH 7.5). Finally the captured DNA was separated from the beads with two denaturation steps at 95℃ for 5 min and 0.15M NaOH consecutively, precipitated with isopropanol and resuspended in 50µl of distilled water. Two microliters of the recovered DNA were amplified using the same *MseI*-N primers and conditions as before.

Cloning of the PCR amplicons was carried out using the TOPO-TA cloning vector kit (Invitrogen) after which colonies were plated out and grown on Ampicillin-containing Luria Bertani (LB) plates. Single colonies were screened for inserts using colony-PCR with the M13 forward (5' GGT TTT CCC AGT CAC GAC 3') and reverse (5' GGA AAC AGC TAT GAC CAT G 3') vector-specific primers. Colony PCR products were electrophoresed on a 2% (m/v) agarose gel to identify clones ≥ 500bp. The ABI PRISM BigDye Terminator Cycle Sequencing kit version 3.1 and M13 primers were used for sequencing. Sequence data obtained from the 3100 Genetic DNA Analyser (Applied Biosystems) was tested against the NCBI database to verify the origin of the DNA, followed by screening for tandem repeats.

Oligonucleotide primers were designed for repeat-containing sequences using the program OLIGO ^{TM} version 4.0 (National Biosciences Inc.). Primers were designed to have annealing temperatures of approximately 60℃. Loci that resulted in products of the expected size and showed polymorphism in a subset of eight random individuals were characterized further in 32 wild-caught individuals of an East coast population. All polymerase chain reactions were conducted in a Geneamp 2700 thermo cycler (Applied Biosystems) in 10µl reactions containing 20ng DNA, 0.3µM of each primer, 200µM of each dNTP, 1 unit of *Taq* polymerase (Promega), 1x PCR Buffer A (Promega) and 2mM $MgCl_2$. PCR consisted of an initial denaturing step at 94℃ for 5 min followed by 35 cycles of 30s at 94℃, 30s at 59 -61℃, and 1min at 72℃, and a final extension for 10 min at 72℃. PCR products we re separated on an ABI 3100 Automated Sequencer and analysed using the GENESCAN version 3.1 software program (Applied Biosystems). The number of alleles, observed and expected heterozygosities and departure from Hardy-Weinberg equilibrium were all calculated using the program GENETIX version 4.02 (Belkhir *et al.* 2000).

2.2.2 Isolation of SNPs

Six previously developed EST clones (Accession numbers EU135914–EU135918, EX534490) that showed significant similarity to known proteins were selected for the discovery of novel SNPs for *H. midae*. For each EST, a cDNA fragment of approximately 300 base pairs was targeted for amplification. PCR primers were designed using the Primer3 software program (Rozen and Skaletsky, 2000) and care was taken to avoid possible exon/intron border sequences (AG/GT; He *et al.,* 2003) when positioning primers. Primers were used to amplify genomic DNA from a test panel of eight random *H. midae* individuals, selected from disparate sampling sites.

Each PCR reaction was carried out in a total volume of 25µl containing 20ng genomic template DNA, 200µM of each dNTP, 0.4 µM of each primer, 1.5mM $MgCl_2$ and two units of GoTaq® Flexi DNA polymerase (Promega). Amplification was performed in a GeneAmp System 2700 thermal cycler (Applied Biosystems). The PCR cycling conditions were as follows: an initial denaturation of 10 min at 95℃ followed by 30 cycles of: 94℃ for 1min, 55℃ for 1 min, 72℃ for 1 min and a final elongation step of 10 min at 72℃. PCR products were assessed on a 2% agarose gel after which SigmaSpin post-reaction clean-up columns were used for purification of the remaining PCR product. Sequencing was performed in both directions using the ABI PRISM BigDye Terminator version 3.1 Cycle Sequencing kit and the 3100 Genetic DNA Analyser (Applied Biosystems). Alignment of sequences was carried out in BioEdit version 7.0.0 (Hall, 1999) and sites that appeared to exhibit sequence variation, either in heterozygous or homozygous forms, were evaluated by manual inspection of the chromatograms. Heterozygous individuals were detected as multiple peaks at the same sequence position and considered as putative SNPs when both alleles were observed at a frequency of greater than 0.05. To confirm the presence of the SNPs, internal primers sets were developed for the fragments ≥ 300bp while the original primer sets were used for the remaining clones. Amplification was repeated on eight more individuals using the same PCR amplification conditions as before, varying only the annealing temperature. Sequence variations that could still be identified in the heterozygous state with the second amplification were considered true SNP loci.

2.3 Results

2.3.1 Microsatellite enrichment, sequencing and genotyping

Abalone DNA was successfully amplified using the *MseI*-N primer during the first step of the Zane *et al.* (2002) enrichment procedure (Figure 2.1a). Enrichment of repeat containing fragments following, hybridization and washing seemed to be satisfactory as smears in the expected size range (≥200bp) were obtained with the 2nd AFLP amplification after both the denaturation steps (Figure 2.1b).

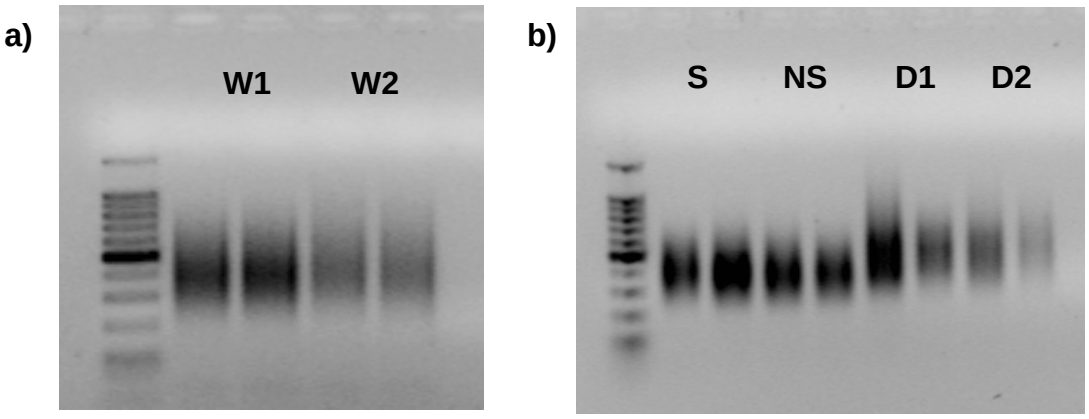

Figure 2.1 Electrophoreses of a) DNA of 2 individuals (W1 and W2) amplified with MseI-N primers and b) microsatellite enriched DNA (S= stringency wash; NS= non-stringency wash; D1= 1st denaturation; D2= 2nd denaturation).

Cloning of the enriched DNA into the TOPO-TA vector delivered approximately 1200 recombinant clones per reaction. Colony screening confirmed a wide range of size inserts and 250 PCR fragments (≥500bp) were selected for sequencing. The presence of repeat sequences was verified in nearly forty-five percent of the clones while only a small number had sufficient flanking regions for primer design. Microsatellite repeats included mostly dinucleotides and tetranucleotides with the most common repeat being the AC/TG dinucleotide repeat followed by the GTCT

tetranucleotide repeat (Figure 2.2 a and b). Many of the repeats were compound or interrupted and ultimately primers were designed for only twenty of these repeat-containing loci. Using the template DNA from which these loci were isolated, amplification products of expected size were obtained for 16 of the 20 microsatellite loci isolated.

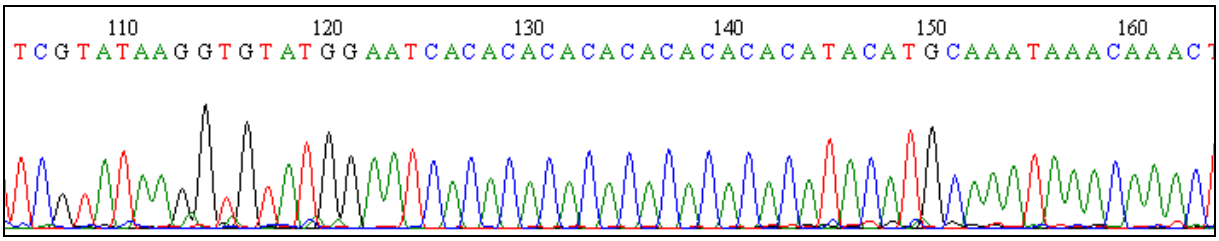

Figure 2.2a Sequence of clone D14 showing a $(CA)_{10}$ repeat

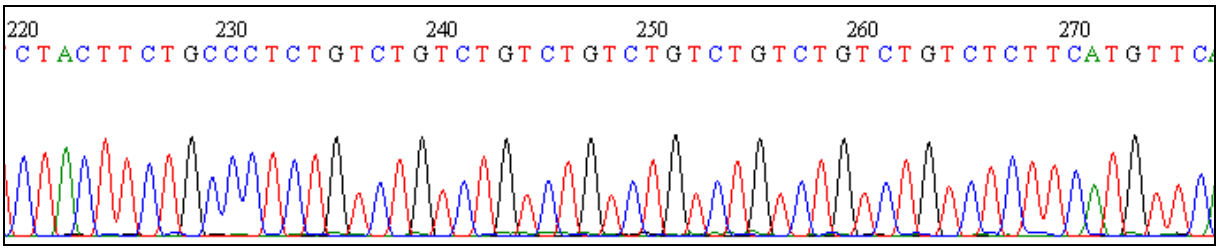

Figure 2.2b Sequence of clone A11 showing a $(GTCT)_8$ repeat

Primers for the 11 loci that showed polymorphism in the test panel were successfully labeled with FAM, NED, VIC or PET fluorescent dyes and genotyped in 32 individuals. Primer sequences and characteristics of these loci are shown in Table 2.1. The number of alleles ranged from five to 21 while observed heterozygosity varied widely between loci ($0.14 \leq H_O \leq 0.78$). Besides natural variation, the large variation could be due to scoring errors, null alleles or the allele drop-out effect observed at loci such as *Hm*D14 and *Hm*SP5. The presence of null alleles is

suspected for loci *HmD33*, *HmD36*, *HmD60* and *HmSP1* as they show significant departure from HWE ($P < 0.01$).

Table 2.1 Characteristics of the 11 microsatellite loci developed for *Haliotis midae* during this study

Locus	Repeat Sequence	Primer Sequence (5'-3')	Size Range	T_a (°C)	H_O	H_E	N_a	Acc Nr
HmD14	$(CA)_{10}$	F TAAGGCAAGTGAATGTCTAG R ATTGCAAGAATCACAACTGC	142-180	60	0.67	0.76	16	AY303333
HmD33	$(GAGT)_{12}$ $AAGT(GAGT)_6$	F TTGAAAGTGAACCAAAATCTG R CATGGGTACAATGTGTAAAGC	129-205	59	0.32	0.87	11	AY303334
HmD36	$(GTGA)_{14}$	F AGATCGAATGACATCAGCTTC R CATATAGCAAGCCTGAAACC	220-304	60	0.43	0.89	15	AY303335
HmD55	$(GTGA)_{12}$	F ATCAAGATAAAACGAGGCG R ACCACTGTGAAAACGTCCA	183-211	60	0.68	0.8	9	AY303337
HmD59	$(CA)_{15}$	F TATACTGCCATTTCCGTCTG R TCTGTATTCTGGTCCTGTCG	106-150	60	0.78	0.84	15	AY303338
HmD60	$(CA)_{16}$	F AAGTTGTTCTCCATAAAGTCGTA R GAAGATCCGGGTTAGAACTG	155-171	60	0.14	0.86	8	AY303339
HmD61	$(CA)_{24}$	F GATATCCAACCCCTGATCAC R GAACATCAACATCTCCATGG	234-298	60	0.61	0.82	11	AY303340
HmA11	$(TCTG)_8$	F AGCTCAGAAAAGTGGTGTACG R TTACCTAGCTAAAGTTGACAACG	292-352	61	0.32	0.66	5	AY303341
HmA30	$(AGTC)_2GGTC$ $(AGTC)_{11}$	F TGATGTTGCTGGAATATTGC R CAATTTCATTTTCAACAGTTCA	124-150	60	0.7	0.8	11	AY303342
HmSP1	$(CA)_{10}CGCA$ $(CA)_4$	F ATAGTGGTCATACAGTCATCACCT R TAGGCATGTTTGAGTTCGTGT	192-276	61	0.48	0.93	21	AY303346
HmSP5	$(AC)_{13}$	F TTCGGCAAGTGAATGTCTAG R ATGCGACACTTACTACACCG	185-219	60	0.63	0.74	14	AY303344

T_a(℃)= annealing temperature; H_O =observed heterozygosity; H_E = expected heterozygosity; N_a = number of alleles

2.3.2 SNP discovery and evaluation

Genomic DNA fragments ranging from 240 bp to 1,2 kb was effectively amplified using the primer sets developed for the six EST clones selected for SNP discovery. Alignments with BioEdit showed several potential sequence variations between the

eight individuals sequenced but fewer were considered as potential SNPs after closer inspection of the chromatograms and allelic states. After verification with internal primers a total of 20 candidate SNPs were identified (Table 2.2). The most common polymorphism was the A/T followed by the G/A sequence variation.

Table 2.2 The origin and positioning of 20 SNPs identified in *Haliotis midae*

EST clone	Homology	GenBank Accession	Primer Sequences (5'-3')	Product Size	SNP Identity	Base Position
1A1	Perlucin	EU135915	TTTGTTAGCCTCGGTCCATC CGATCACAGGGGACATCATT	1250bp	A/C G/A G/T	122 148 155
C12	Cellulase	EU135914	ATTTTTGTCGGTCACCTGGA GTAGGGCTTCCCAGAAGGAC	300bp	T/C T/A	225 260
3B4	Ribosomal protein L8e	EU135916	GAAAGCCATTCCCTCAACAA TGCTTTGTACCACGAAGACG	1230bp	T/A C/T A/T C/T C/G A/G T/A T/C T/A	57 148 272 291 320 325 492 561 563
3D10	Haemocyanin	EU135917	TGAGTGCGGCGTAAAAATAA AGTGGAACGCCAAATGTTGT	240bp	G/A	122
2C3	Fibroblast growth factor receptor	EU135918	GGTGGAAACCCTTACGACAA AACCACGTTCGCTGTGGTAT	375bp	A/G T/A A/C	114 145 280
2H9	Ribosomal protein L22	EX534490	TCCGACCTTCTTGATCTTGG AGGAAATGCTGGTGGACAAC	300bp	A/T A/T	116 177

Within the total of 3,7kb of DNA screened, this represented a rate of approximately one SNP every 185 base positions. Polymorphisms were present at a 1:1 ratio of transitions to transversions. This ratio is slightly lower than expected since transitions involve only two types of base substitutions, whereas transversions involve four substitutions. Polymorphisms that revealed extremely low rare allele frequencies or close proximity to other SNPs were excluded from further analysis. Characterization

of the remaining SNPs showed moderate levels of heterozygosity with frequencies of the least common allele ranging from seven to 49%.

2.4 Discussion

As microsatellites are found in eukaryotic genomes at varying frequencies with respect to absolute number of loci and repeats, the number and type of repeat containing clones obtained with a single enrichment was roughly as expected when compared to other *Haliotis* species (Miller *et al.,* 2001; Sekino and Hara, 2001; Cruz *et al.,* 2005; Baranski *et al.,* 2006b). Although very few technical problems were encountered following the enrichment method, the final number of microsatellites isolated in this study was not necessarily ideal taking into account that a relatively high enrichment level of 45% was achieved. This was for the most part attributable to a lack of flanking regions for primer design in many of the loci sequenced. This phenomenon has previously been observed in other enriched libraries and the main reason seems to be the truncation of cloned sequences related to the restriction enzymes used during the restriction-ligation step (Jones *et al.,* 2001; Baranski *et al.,* 2006a). Another restriction was the large percentage of clones containing interrupted repeats. Even though they can still be applied in, for example, population studies, linkage analysis and parentage assignment, scoring of microsatellites with single or multiple interruptions are known to be difficult because of a more complex mutation process and the high variance in number of repeats (Matsumoto *et al.,* 2004).

Not obtaining desired products for some of the selected loci could be the result of highly variable flanking sequences and changing the priming site when sufficient flanking sequence is available could in many cases solve the problem (Evans *et al.,*

2001; Reece *et al.,* 2004). Alternatively, existing primers can be optimised through varying the DNA and $MgCI_2$ concentrations or PCR cycling conditions.

As for genotyping, the most common difficulties with scoring were experienced with loci exhibiting large-allele dropout or possible null alleles. Both are frequent occurrences in microsatellites and where the preferential amplification of smaller alleles can be minimised through extensive optimisation (Goossens *et al.,* 1998), null alleles are mostly due to the highly variable flanking regions and can only be overcome through the redesign of primers (Hedgecock *et al.,* 2004). Another interesting feature during genotyping was the occasional presence of a third allele of equal intensity to the first two at one of the loci (*Hm*D33) analysed. The incidence of more that two alleles at the same locus has been related to transposable elements (Meglécz *et al.,* 2004), genome duplication, polyploidy or aneuploidy (David *et al.,* 2003; Hubert *et al.,* 2000) and has previously been described in abalone (Evans *et al.,* 2001; Baranski *et al.,* 2006a).

The strategy for SNP discovery used in this study allowed for the rapid and trouble-free identification of an additional set of moderate to highly informative markers for *H. midae*. Single base variations are generally found every 200 to 500 base positions in the genome while Sauvage *et al*. (2007) reported an average of one SNP every 60 bp in coding regions and one every 40 bp in non-coding regions of Pacific oyster, *Crassostrea gigas*. The rate observed in *H. midae* was comparable to or even higher than that found in EST sequences of other species (Fahrenkrug *et al.,* 2002; He *et al.,* 2003; Feau *et al.,* 2007). However, an obvious disadvantage was the small number of EST clones screened allowing for the discovery of only a restricted number of SNPs. As for the type of polymorphism examined, the slightly lower ratio of transitions to transversions obtained compared to other species was

not a major concern as the total length of DNA screened was too small to be representative of the whole genome. Data of SNPs in invertebrate species is still rare but Wang *et al.* (1998) found that in the human genome transitions occur more frequently than transversions due to the highly mutable site present within the frequently occurring CpG islands.

Furthermore, the linkage disequilibrium found for SNPs within the same EST automatically reduced the total amount of polymorphic information content gained. As sequencing and genotyping of SNPs was mainly performed manually, time and cost-effectiveness could be improved upon through screening rapidly evolving gene classes (Nielsen *et al.,* 2005) or targeting only one SNP per exonic region. Besides ESTs, other available sequence data for *H. midae* can also provide a valuable source for the quick and easy identification of SNPs. Rhode *et al.* (2008) for example identified 12 novel SNPs for *H. midae* by screening flanking regions of only four microsatellite sequences generated during the same enrichment procedure as performed in this study. But while SNP identification will increase as sequence information accumulates for *H. midae*, the primary aim should still be to generate the highest possible number of SNPs without introducing ascertainment bias because of the number and source of individuals used in the initial isolation process. If for example the SNP frequency within such a sampling panel is unrepresentative of the true frequency within the species, the estimation of population genetic parameters based on allele frequencies will be unreliable (Morin *et al.,* 2004).

2.5 Conclusions

In conclusion, this chapter reported the isolation and development of the first microsatellite markers and SNP loci specifically for *H. midae.* As both types of markers have proven to be excellent for population structure analysis, the advantage of having such a set of markers outweighs the effort, cost and minor problems experienced during the respective isolation procedures. The fact that five of the microsatellites developed here have successfully been used in the genetic diversity analysis and parentage assignment of commercial *H. midae* cohorts (Slabbert *et al.,* submitted), provides sufficient evidence that these markers could be applied in genetic studies related to parentage, linkage mapping and QTL analysis. However, as both marker types are not devoid of limitations, it is expected that not all of them will be equally useful in some of the applications. The markers presented here are nonetheless fundamental to genetic characterisation and future mapping of the *H. midae* genome.

References

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407:** 513–516.

Baranski M (2006a) Molecular genetic markers for abalone aquaculture. PhD dissertation, Deakin University, Australia, 205pp.

Baranski M, Rourke M, Loughnan S, Austin CM, Robinson N (2006b) Isolation and characterization of 125 microsatellite DNA markers in the blacklip abalone, *Haliotis rubra*. *Molecular Ecology Notes* **6:** 740–746.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (2000) GENETIX, logiciel sous Window TM pour la génétique des populations. Laboratoire Génome et Populations, Montpellier, France.

Christiakov 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.

Cruz P, Ibarra AM, Fiore-Amaral G, Galindo-Sanchez CE, Mendoza-Carrion G (2005) Isolation of microsatellite loci in green abalone (*Haliotis fulgens*) and cross-species amplification in two other North American red (*Haliotis rufescens*) and pink (*Haliotis corrugata*) abalones. *Molecular Ecology Notes* **5:** 857–859.

David L, Blum S, Feldman M, Lavi U, Hillel J (2003) Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution* **20:** 1425–1434.

Evans B, Conod N, Elliott NG (2001) Evaluation of microsatellite primer conservation in abalone. *Journal of Shellfish Research* **20:** 1065–1070.

Evans B, White WR, Elliott NG (2000) Characterization of microsatellite loci in the Australian blacklip abalone (*Haliotis rubra* Leach). *Molecular Ecology* **9:** 1183–1184.

Fahrenkrug SC, Freking BA, Smith TP, Rohrer GA, Keele JW (2002) Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Animal Genetics* **33:** 186–195.

Feau N, Bergeron M-J, Joly DL, Roussel F, Hamelin RC (2007) Detection and validation of EST-derived SNPs for poplar leaf rust *Melampsora medusae* f. sp. *deltoidae*. *Molecular Ecology Notes* **7:** 1222–1228.

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

He C, Chen L, Simmons M, Li P, Kim S, Liu ZJ (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics* **34:** 445–448.

Heaton MP, Harhay GP, Bennett GL, Stone RT, Grosse MW, Casas E, Keele JW, Smith TPL, Chitko-McKown CG, Laegreid WW (2002) Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome* **13:** 272–281.

Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research* **23:** 379–385.

Hubert S, English LJ, Landau BJ, Guo X, Hedgecock D (2000) Microsatellite analysis of trisomic families in the Pacific oyster, *Crassostrea gigas,* Thunberg. *Journal of Shellfish Research* **19:** 615.

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics* **102:** 405–415.

Matsumoto T, Yukawa W, Nozaki Y, Nakashige R, Shinya M, Makino S, Yagura M, Ikuta T, Imanishi T, Inoko H, Tamiya G, Gojobori T (2004) Novel algorithm for automated genotyping of microsatellites. *Nucleic Acids Research* **32:** 6069–6077.

Meglécz E, Petenian F, Danchin E, Coeur D'Acier A, Rasplus J-Y, Faure E (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* **13:** 1693–1700.

Miller KM, Laberee K, Kaukinen KH, Withler RE (2001) Development of microsatellite loci in pinto abalone (*Haliotis kamtschatkana*). *Molecular Ecology Notes* **1:** 315–317.

Morin PA, Luikart G, Wayne RK and SNP-Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19:** 208–216.

Nicod J-C and Largiadèr (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Research* **31:** e19

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Ciuvello D, White TJ, Sninsky JJ, Adams MD, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3:** e170

Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite Flanking Regions: A SNP Mine in South African Abalone (*Haliotis midae*). *Animal Genetics* **39:** 329.

Rozen S and Skaletsky HJ (2000) PRIMER3 on the WWW for general users and for biologist programmers. In: Bioinformatics Methods and Protocols: Methods in Molecular Biology (ed. Krawetz S and Misener S), Humana Press, Totowa, NJ, pp365–386.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Atshuler D (2001) A map of human genome sequence variation containing 142 million single nucleotide polymorphisms. *Nature* **409:** 928–933.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences* **81:** 8014–8019.

Sauvage C, Bierne N, Lapègue S, Boudry P (2007) Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* **405:** 13–22.

Sekino M and Hara M (2001) Microsatellite DNA loci in Pacific abalone *Haliotis discus discus* (Mollusca, Gastropoda, Haliotidae). *Molecular Ecology Notes* **1:** 8–10.

Selvamani MJP, Degnan SM, Paetkau D, Degnan BM (2000) Highly polymorphic microsatellite loci in the Heron Reef population of the tropical abalone *Haliotis asinina*. *Molecular Ecology* **9:** 1184–1186.

Slabbert R, Bester AE, D'Amato ME (2008) Analyses of genetic diversity and parentage within a South African hatchery of the abalone *Haliotis midae* Linnaeus resolved using microsatellite markers. *Journal of Shellfish Research* (submitted)

Vignal A, Milan D, San Cristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34:** 275–305.

Vos P, Hogers R, Bleceker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* **23:** 4407–4414.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:** 1077–1082.

Weber JL and Myers EW (1997) Human whole-genome shotgun sequencing. *Genome Research* **7:** 401–409.

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

Chapter III

Population structure analysis and demographics of *Haliotis midae* evidenced by microsatellite DNA genotyping

3.1 Introduction

The high dispersal capacity of marine fish and invertebrate species in general fuel the expectation of little to no genetic subdivision of these species into stocks or populations (Waples, 1998; DeWoody and Avise, 2000). Especially in environments without major geographical barriers, the obvious expectation is high levels of gene flow and subsequent lack of population structuring. In spite of the apparent potential for long-distance dispersal and absence of barriers to gene flow, some marine species do however exhibit significant population structure over small geographic regions (Bentzen *et al.,* 1998; Hutchinson *et al.,* 2001; Knutsen *et al.,* 2003; Taylor and Hellberg, 2003). In abalone species mostly low levels of genetic structuring have been reported: *Haliotis rubra* (Huang *et al.,* 2000; Conod *et al.,* 2002), *H. cracherodii* (Hamm and Burton, 2000; Chambers *et al.,* 2006), *H. kamtschatkana* (Withler *et al.,* 2001; 2003), *H. midae* (Evans *et al.,* 2004); *H. asinina* (Tang *et al.,* 2005) and *H. rufescens* (Gruenthal *et al.,* 2007).

The higher potential of microsatellite markers over mtDNA and allozymes in stock identification is well known in fisheries genetics (Carvalho and Hauser, 1994; 1998).

Although it is largely as a result of their greater level of polymorphism (Withler *et al.,* 2000), contrasting results obtained with different marker types (Evans *et al.,* 2004; Nielsen *et al.,* 2006; Fauvelot *et al.,* 2007) can also be due to a biological effect such

as natural selection (Silva and Skibinski, 2009). In *Haliotis rubra* for example both allozyme and microsatellites loci exhibited population differentiation but the scale of differentiation differed depending on the type of marker applied (Conod *et al.,* 2002).

Moritz *et al.* (1994) stated that an understanding of the genetic structure of any species is vital to the identification of both evolutionary significant units (ESUs) and management units (MUs). Evolutionary significant units are important phylogeographic subdivisions whose maintenance might be recognized as critical for long-term conservation of biodiversity whereas management units could be described as local populations, whose preservation is critical for short-term maintenance of the species' range and abundance (Withler, 2000). Because of various factors threatening the persistence of this species in certain areas of its geographic distribution (discussed in chapter 1), local authorities are in desperate need to implement and enforce management strategies for the successful conservation of *Haliotis midae*.

The ultimate objective of this chapter was to investigate the patterns of gene flow among local sampling populations of *Haliotis midae* and the potential processes underlying the detected structuring. The main alternative hypotheses in consideration are panmixia vs restriction to gene flow. The expectation of panmixia is based on the large capacity of dispersal at the larval stage, from which the absence of partition in the geographic distribution of genetic variation can be predicted. On the other hand, the restriction of gene flow along the SA coast has been observed in many other organisms (Ridgeway *et al*., 1998; Teske *et al.,* 2006; Zardi *et al.,* 2007; Teske *et al.,* 2007), with high correspondence to the biogeographic regions along the SA coastline. Complex oceanography and coastal habitats along the South African coastline are the main biogeographic factors to consider (Turpie *et al.,* 2000). The

main oceanographic feature that could influence larval dispersal of *Haliotis midae* is the occurrence of major currents along the South African coastline: the cold Benguela current on the West coast and the warm southward flowing Agulhas current on the East coast. The region between Cape Point and Agulhas experiences large temperature fluctuations associated with upwelling of the Benguela current whereas a persistent thermal front has been found in the region just south of Port Elizabeth as a result of the inshore edge of the Agulhas current (Beckley and van Ballegooyen, 1992). For this study, the two areas around Cape Agulhas and Port Elizabeth have been identified as possible biogeographic barriers causing restricted gene flow between the western, southern and eastern coastal areas. Although Evans *et al.* (2004) found significant variation between *Haliotis midae* populations west and east of Cape Agulhas based on mtDNA RFLP polymorphisms, their results using seven allozyme loci and only three non-specific microsatellites genotyped at six sampling sites seem inconclusive and no statistical support was provided .

3.2 Materials and Methods

3.2.1 Sample collection

Samples of *Haliotis midae* were collected between 2000 and 2005 using commercial as well as government scientific divers. In total, 428 animals were collected throughout the range of the species from Saldanha Bay on the West coast to the mouth of the Kei River on the East coast. The distribution range covered approximately 1600km of the South African coastline. The sample size ranged between 33 and 54 individuals per sampling location indicated in Figure 3.1. Because of unfavourable habitat for abalone, permit restrictions and difficult diving conditions, the area along the South coast was not sampled optimally. Most of the collection

occurred up to 1km offshore and samples were received either as live animals or as muscle and gill tissue preserved in absolute ethanol. Although size and sex of the animals were not taken into consideration, only adult animals over the age of 4 years were collected for this study. Sampling locations will hereafter be referred to as populations.

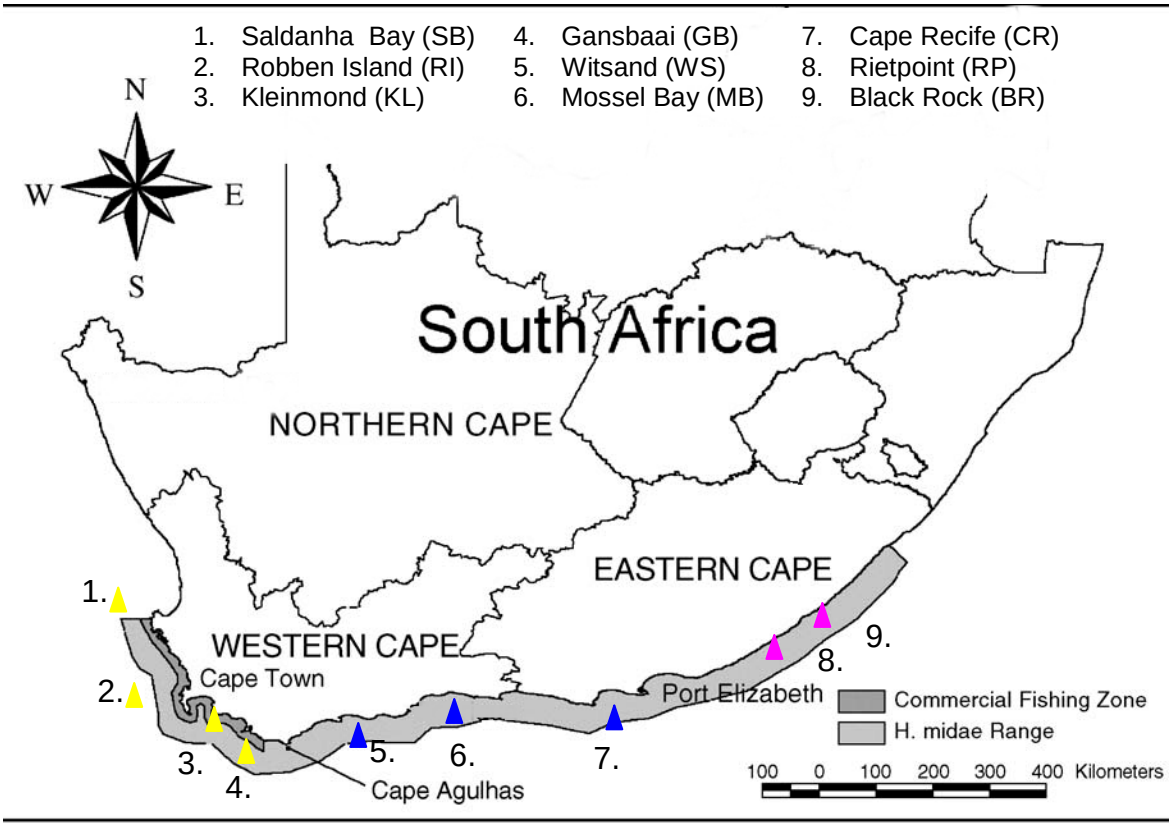

Figure 3.1 Sample collection locations of *Haliotis midae* along its natural range of distribution with yellow, blue and pink triangles representing the West, South and East coast sampling populations respectively.

3.2.2 Nucleic acid isolation

DNA was extracted from gill or muscle tissue using a standard cetyltrimethylammonium bromide (CTAB) method described by Saghai-Maroof *et al.* (1984). Approximately 0.1 g of tissue was placed in 700µl of lysis buffer containing 0.5mg/ml proteinase K and incubated overnight at 60℃. Following incubation, the samples were extracted with chloroform: isoamyl alcohol (24:1) and the nucleic acids

were precipitated with absolute ethanol. The extracts were air dried at room temperature and resuspended in distilled water depending on the size of the pellet. DNA was loaded on a 1% (m/v) agarose gel together with a lambda DNA series to assess the concentration and quality of the DNA. The samples were stored at -20℃ until further use or diluted to a final working concentration of 20ng/µl and stored at 4℃. DNA samples from Kleinmond and Mossel Bay were kindly provided by Dr Raurie Bowie from the Department of Botany and Zoology at the University of Stellenbosch.

3.2.3 Microsatellite genotyping

Allelic variation was assayed at eight microsatellite loci described either in the previous chapter or selected based on the degree of polymorphism and scoring ability (Table 3.1). Isolation and characterization of the six markers, *HmD14*, *HmD36*, *HmD55*, *HmD59*, *HmA11*, *HmSP5* were reported by Bester *et al.* (2004) while *HmAD102* was a newly isolated locus and *CmrHr2.15*, a *Haliotis rubra* microsatellite locus previously found to be polymorphic in *H. midae (*B van Zyl, unpublished). A total of 9 populations were assayed with a sampling size ranging from 33 to 54 individuals per population.

Table 3.1 Characteristics of the eight microsatellite loci used in this study

Locus	Repeat Motif	Size Range	T_a ℃	H_O	H_E	Allele number	Acc Nr
HmD14	$(CA)_{10}$	142-180	60	0.67	0.76	16	AY303333
HmD36	$(GTGA)_{14}$	220-304	60	0.43	0.89	15	AY303335
HmD55	$(GTGA)_{12}$	183-211	60	0.68	0.8	9	AY303337
HmD59	$(CA)_{15}$	106-150	60	0.78	0.84	15	AY303338
*Hm*A11	$(TCTG)_8$	292-352	61	0.32	0.66	5	AY303341

HmSP5	$(AC)_{13}$	185-219	60	0.63	0.74	14	AY303344
CmrHr2.15	$(CA)_{27}$	231-279	55	0.43	0.71	19	AF195956
HmAD102	$(ACTC)_{15}$	154-256	60	0.53	0.82	24	DQ785747

T_a ℃ = annealing temperature; H_O= observed heterozygosity; H_E = expected heterozygosity

Polymerase chain reactions (PCR) were carried out with fluorescently labeled primers in a total volume of 10µl in a GeneAmp 2700 thermocycler (Applied Biosystems). Amplification conditions consisted either of a basic cycle profile of denaturation at 95ºC, annealing at T_a (see Table 3.1) and extension at 72ºC or a touch-down method where the annealing temperature is lowered by 1ºC with each consecutive cycle. PCR reactions contained 20ng genomic template DNA, 200µM of each dNTP, 3-5 pmol of each primer, 2-3mM $MgCl_2$ and 0.5 units of *Taq* DNA polymerase. PCR products were assessed on a 2% (m/v) agarose gel and the intensity of the bands was used as a reference for making dilutions prior to loading it on an ABI 3100® automated sequencer. Diluted PCR products were mixed with GeneScan 500 Liz size standard (Applied Biosystems) and denatured at 95ºC for 5 minutes before being loaded onto an ABI 3100® automated sequencer. Loci were multiplexed according to size and fluorescent labeling of the primers. Allele sizes were determined using Genotyper® software version 3.7 of Applied Biosystems.

3.2.4 Statistical analysis

Genetic diversity. Basic estimates for genetic diversity such as allele frequencies, observed (H_O) and unbiased expected heterozygosity (H_E) were estimated in GENETIX version 4.03 (Belkhir *et al.,* 2000). The number of alleles per locus and allele richness based on a minimum sample size of 32 was calculated in FSTAT version 2.9.3 (Goudet, 2001). Deviations from Hardy-Weinberg equilibrium (HWE)

were tested using GENEPOP version 3.3 (Raymond and Rousset, 1995). The Markov chain randomization method (Guo and Thompson, 1992) was used to calculate the unbiased estimates of the exact *P*-values per locus and per population. Where observed genotype frequencies deviated significantly from HWE expectations, the software MICRO-CHECKER version 2.2.3 (Shipley, 2003) was used to examine the reason for heterozygote deficiency. MICRO-CHECKER implements the method of Brookfield (1996) for detection of null alleles and aids in the identification of scoring errors due to other causes for HWE deviations such as stuttering and large allele dropout.

Population structure. Genetic differentiation between populations was investigated using the following analytical methods: (1) F_{ST} (2) R_{ST} (3) Exact Test of Genic Differentiation (4) AMOVA (5) FCA and (6) Non-Spatial and Spatial Bayesian clustering. Individuals were also tested for relatedness using IDENTIX.

Heterogeneity in allele frequency distribution among populations was tested by analysis of variance-based methods assuming both the Infinite Allele Model (IAM) and the Stepwise Mutation Model (SMM). Following the IAM, Wright's single locus F-statistics were calculated according to Weir and Cockerham (1984) in GENETIX (θ) while the analogue, *Rho* or R_{ST} of Slatkin (1995) were calculated in RSTCALC (Goodman, 1997) following the SMM. Significant departure from zero was tested with a permutation approach (1000 replicates) for both multilocus F_{ST} and R_{ST} values followed by Bonferonni adjustments (Rice, 1989). Fisher's exact test for genic differentiation (ETGD) as described by Raymond and Rousset (1997) was performed to detect weak differentiation because of the higher power of this method to detect differentiation in the presence of rare alleles. Mantel tests (with 1000 permutations)

were performed with the program MANTEL within the package GENETIX to estimate possible associations between genetic and geographic distances of populations (isolation-by-distance). The average genetic relatedness among individuals within populations was further investigated with the R_{xy} estimator of Queller and Goodnight (1989) available in the program IDENTIX (Belkhir *et al.,* 2002). Using 1000 permutations of alleles among individuals, the significance of the values of relatedness were tested to know the proportions of pairwise relatedness values not attributable to the random sharing of alleles. Furthermore, population allele frequency data were subjected to factorial component analysis (FCA) also available in GENETIX. With this analysis the genetic relationship among populations can be resolved by visualizing the ordination of populations along the factorial axes in a three-dimensional space. FCA analysis was also conducted at the individual level to test whether geographical distribution could be related to the relative distribution of individual genotypes in the multivariate plot. Analysis of molecular variance (AMOVA) was carried out in ARLEQUIN version 2.0 (Schneider *et al.,* 2000). The percentage of the total genetic variation explained by allele frequency variation within populations, among populations within groups and among groups was calculated under different grouping hypotheses. Genetic division was tested (1) overall, (2) between populations west and east of Cape Agulhas, and (3) between three groups assuming barriers at Cape Agulhas and Port Elizabeth.

For the Bayesian inference of population structure, the clustering method implemented in STRUCTURE version 2.1 was applied (Pritchard *et al.,* 2000). This program uses a Bayesian clustering approach to infer the most likely number of populations (*K*) and the proportion of each *K* within individual genomes. Ten runs consisting of 10^6 Markov Chain Monte Carlo (MCMC) iterations and a burn-in of 10%

were performed for each *K*. The analysis was conducted assuming admixture and correlated allele frequencies while *K* was set between 1 and 13. The true number of *K* was then estimated using the statistic delta *K* (Δ*K*), the second-order rate of change of the likelihood function with respect to *K* (Evanno *et al.,* 2005). In brief, the mean difference between successive likelihood values, L'(*K*), was calculated after which the mean of the absolute values of L'(*K*) were averaged over ten runs and divided by its standard deviation. The value of *K* that showed the highest Δ*K* was chosen to evaluate the individual membership coefficient (q_{ind}) to the inferred number of clusters. Thereafter, the two-population model that was chosen based on the Δ*K* results was tested first by forcing *a priori* structure (individuals assigned to West or East coast group according to geographical knowledge) and then without forcing any population structure (no prior population information included). The efficiency of STRUCTURE to assign 'unknown' individuals to the hypothetical West and East coast clusters were tested using the USEPOPINFO option. After identification of a subset of pure West and East coast individuals with a membership coefficient of q≥ 0.75, STRUCTURE was run using this prior information (*K*=2) by treating the rest of the individuals as having unknown origin (USEPOPINFO=0). For every individual the proportion of membership to each cluster was assessed while the level of admixture between the two clusters was given by the average membership coefficient of each population (q_{group}) to each cluster. In addition, cluster membership per population was summarized using the 20% highest likelihood runs for *K*=2 processed in the cluster matching and permutation program, CLUMPP (Jakobsson and Rosenberg, 2007). Because STRUCTURE assumes HWE within populations, analysis was performed excluding the three loci with highest null allele frequencies (*Hm*D36; *Hm*AD102 and *Hr*2.15).

Geographical information can be added to the above analysis as an extra parameter in the inference of clustering. For this, the software TESS version 1.1 (Chen, 2007) was implemented to infer the number of populations and individual assignment probabilities. TESS implements a Bayesian clustering algorithm for spatial population genetics and is based on a hierarchical mixture model where the prior geographic distribution is defined as a Hidden Markov Random Field (HMRF) on tessellations (François *et al.,* 2006). Given individual geographical locations, the program seeks population structure from multilocus genotypes without assuming predefined populations. It further allows making prior assumptions about the level of interaction between populations through controlling the interaction parameter, Ψ. In this study, exploratory runs at four values of Ψ ranging from 0.3 (weak interaction) to 1.0 (strong interaction) were initially tested after which it was fixed at the value most suitable for observing the coexistence of several clusters. TESS runs were performed for 120 000 sweeps and a burn-in period of 50 000 with the maximum number of clusters (K_{max}) ranging between 2 and 5 while the interaction parameter (Ψ) was finally fixed at 0.5. For every combination of parameters, 30 independent runs were performed and the 20% highest likelihood runs were utilized for output post-processing in CLUMPP version 1.1 (Jakobsson and Rosenberg, 2007). The cluster visualization program DISTRUCT (Rosenberg, 2004) was used to display the membership coefficients of the clusters within each population.

Finally, the program BARRIER version 2.2 (Manni *et al.,* 2004) was used to define areas where genetic discontinuity was most pronounced. Barrier uses Monmonier's (1973) maximum difference algorithm together with spatial coordinates to identify genetic barriers. The same geographical coordinates were used as in the program TESS and connected by Delauney triangulation such that each connection has an

associated distance. Analyses were performed based on the genetic distance of Nei (1987). To prevent barriers being identified due to strong differentiation at one locus only, analyses were performed for each locus separately. The program was set to identify only one barrier after which additional barriers were identified according to the number of loci supporting them. In order to assess the robustness of barriers, two more matrices based on Weir and Cockerham's (1984) θ and R_{ST} values were included in the analysis. The robustness was measured according to the number of times each barrier was supported by the different matrix sets.

3.2.5 Demographic changes

Long term effective population size (N_e) was calculated from the estimates of the unbiased expected heterozygosity under both the infinite-alleles model (IAM) and the stepwise mutation model (SMM) according to the formulae of Nei (1987):

IAM: $N_e = H_e/(1-H_e)/4\mu$

SMM: $N_e = \{[1/(1- H_e)]^2-1\}/8\mu$

where H_e is the average expected heterozygosity and μ is the average mutation rate of the microsatellites used. Single locus maximum likelihood estimates of the parameter θ were obtained with the program MISAT version 1.0 (Nielsen, 1997). Different grid values and number of MCMC runs were tested in order to optimize likelihood estimates. The contemporary effective population sizes were estimated from the expression $\theta = 4N_e\mu$ using a conservative mutation rate of 10^{-4} as the mutation rate for microsatellite loci in abalone is not yet known. The proportion of multi-step mutations that were estimated for each locus in MISAT was used as an

average estimate of the percentage of SMM in the two-phase model (TPM) implemented in BOTTLENECK version 1.2.02 (Piry *et al.,* 1999). Sample heterozygosity excess and deficiency were tested for each locus separately and then overall loci assuming strict 1) SMM, 2) IAM, and 3) TPM with 50% of mutations following SMM. The Wilcoxon signed rank test was used to test for mutation-drift equilibrium.

To further test for signs of demographic change, the two parameters developed by Reich and Goldstein (1998) were applied on all eight microsatellite loci. The *k* and *g* test are implemented in the Excel Macro application program, Kgtests (Bilgin, 2007). The *k*-test is based on the intralocus variance of the alleles. A history of population expansion is inferred if a sufficiently high proportion of the loci return negative scores for this parameter. The *g*-test compares the variance across different loci assuming a step-wise mutational model and constant population size. The 5% significance levels of the g statistic are available in Reich *et al.* (1999).

The coalescent-based program MIGRATE version 2.0.6 (Beerli and Felsenstein, 2001) was used to calculate population specific θ and long-term gene flow (migration rate). The program is based on a likelihood approach and calculates both effective population size θ ($4N_e\mu$) and migration rate M (m/μ) where m is the immigration rate and μ the mutation rate. Due to computational constraints, the program was run with only a subset of samples (20 randomly chosen individuals of each population) using mainly default settings. Runs consisted of 10 short and 3 long chains that were run for 500 and 5000 sampled genealogies respectively following an adaptive heating scheme of four temperatures (1.0, 1.2, 1.5, 3.0).

3.3 Results

3.3.1 Sampling and nucleic acid isolation

Table 3.2 Number and origin of samples

Sampling Location	*N*	*Year*
Saldanha (SD)	57	2004
Robben Island (RI)	48	2004
Kleinmond (KL)	48	2000
Gansbaai (GB)	48	2004
Witsand (WS)	33	2004
Mossel Bay (MB)	54	2000
Cape Recife (CR)	48	2003
Riet Point (RP)	53	2003
Black Rock (BR)	48	2003

In a period of four years a total of 428 individuals of *H. midae* were sampled from nine geographically isolated sampling locations. The number of individuals per population as well as the year of collection is indicated in Table 3.2. Genomic DNA was successfully isolated from either gill or muscle tissue for most individuals. DNA concentrations ranged between 20ng/µl and 500ng/µl with a slightly higher yield obtained from gill tissue. In general, DNA was of high quality and no difficulty was encountered with polymerase chain reaction amplification of any of the microsatellite loci.

3.3.2 Microsatellite genotyping

All microsatellite loci were successfully amplified in all sampling populations. When alleles were scored with GENOTYPER software, strict criteria were followed to classify the peaks as true alleles. In case of two peaks, it was only considered as a heterozygote if the two peaks did not differ with more than 2/3 in height. A common occurrence was allele dropout where amplification of the smaller allele is favored and the second allele exhibits a smaller peak height than the first (Gagneux *et al.,* 1997). When the first peak had a considerably smaller height than the second, the first peak was considered to be a stutter band. Stutter bands, or n+x peaks as referred to in GENOTYPER are caused by slippage of the *Taq* polymerase during PCR

amplification (Hancock, 1999). A slippage event is characterized by a loss of one or more repeat units. Data were successfully processed in Microsoft® Excel 2000 and converted to the input format for GENEPOP. The average proportion of missing genotypes per locus was 3.4% and varied from 1.2% (*HmSP5*) to 5.4% (*HmD55*).

3.3.3 Genetic diversity

Except for locus *HmD11* which showed moderate levels of polymorphism, the rest of the microsatellite markers exhibited moderate to relatively high levels of polymorphism. The total number of alleles per locus ranged from 12 to 43 with the highest number of alleles detected at the tetranucleotide locus *HmAD102*. The level of polymorphism increased proportionally to the number of repeats exhibited at each locus. Except for loci *HmD14* and *HmSP5* which exhibited one dominant allele and several rare alleles, most of the loci showed allelic frequency distributions consistent with a stepwise mutation model. Although private alleles (i.e. alleles observed in only one population) were observed at most of the loci, the average number and effective number of alleles did not differ significantly between populations. Genetic variation within the populations was moderate to high with observed and expected heterozygosity, averaged over all populations, ranging from 0.35 to 0.848 and 0.583 to 0.939 respectively (Appendix 3). No significant reduction of expected heterozygosity was detected between populations (Mann-Whitney test, $P>0.05$). Global tests (*P*-values over all loci and populations) for Hardy-Weinberg equilibrium were highly significant ($P<0.001$) while only four of the loci (*HmD14*, *HmD55*, *HmD59* and *HmSP5*) were in equilibrium over all populations. According to MICRO-CHECKER all the remaining loci showed evidence of null alleles or genotyping errors associated with large allele-dropout or stuttering. The null allele frequencies were

highly significant (P<0.001) for loci *HmD36*, *HmAD102* and *Hr2.15* while for loci *HmD55* and *HmD11* miscoring of stutters was suggested as the probable cause of heterozygote deficiency. The overall excess of homozygotes for all loci combined (F_{IS}) was 0.298 and based on a 1000 permutations, the F_{IS} values were significant for four loci (0.001<*P*<0.01).

3.3.4 Population differentiation

Pairwise F_{ST} and R_{ST} values were relatively low ranging from -0.0005 to 0.0549 and -0.0145 to 0.1323 respectively (Table 3.3) and negative R_{ST} values were indicative of greater allele size variance within rather than between populations. Overall populations the tests were not significant (F_{ST} = 0.0145; R_{ST} =0.0509). Only a few F_{ST} pairwise values were significant after sequential Bonferroni adjustment and did not relate to geographical distances. An interesting observation is that most of the significant differences involve either Gansbaai or Kleinmond. The Mantel test also failed to reveal any correlation between genetic and geographical distances (*r*=0.18; *P*>0.05). Tests for genic differentiation (ETGD) however resulted in highly significant values for all but one population comparison (P<0.001) and for five of the loci overall populations (see Appendix 3).

Table 3.3 Pairwise F_{ST} (θ) (above diagonal) and R_{ST} (below diagonal) values between populations depicted in Table 3.2. *Significant values after Bonferroni adjustment are highlighted

	SD	RI	KL	GB	WS	MB	CR	RP	BR
SD		-0.0005	0.0082	0.0229*	0.0070	0.0096	0.0140	0.0086	0.0095
RI	-0.0154		0.0082	0.0264*	0.0098	0.0038	0.0119	0.0087	0.0073
KL	0.0339	0.0063		0.0182*	0.0079	0.0129*	0.0331*	0.0187*	0.0106
GB	-0.0039	-0.0145	0.0171		0.0276*	0.0304*	0.0549*	0.0427*	0.0365*
WS	0.0661	0.0700	0.1101	0.0505		0.0004	0.0148	0.0080	-0.0007
MB	-0.0019	-0.0203	0.0121	-0.0117	0.0581		0.0083	0.0092	0.0142*
CR	0.0135	0.0397	0.1323	0.0420	0.0698	0.0567		0.0072	0.0023
RP	0.0397	0.0193	0.0104	0.0654	0.1207	0.0192	0.0986		0.0023
BR	0.0070	-0.0097	0.0433	-0.0079	0.0444	-0.0116	0.0324	0.0335	

Factorial component analysis (FCA) enabled a three-dimensional view of the genetic relationship among populations with the first three factors explaining 48.02% of the population differentiation. An important observation was the separation of the western populations from the rest when viewed along the x- axis or first factor (Figure 3.2). FCA analysis of the individual genotypes does show substantial overlapping of West, South and East coast individuals, but the distribution from left to right along the first axis agrees with a separation of the West coast individuals from the rest (Figure 3.3).

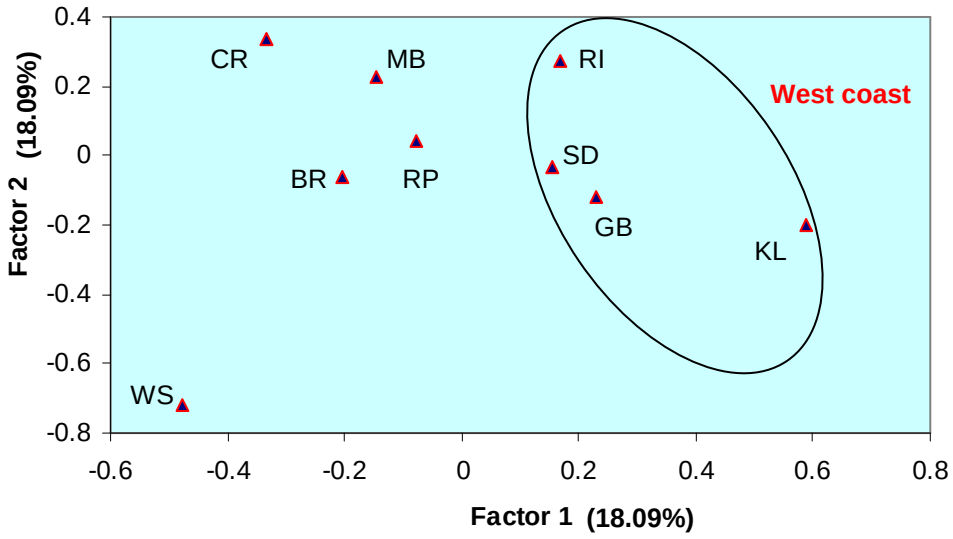

Figure 3.2 Factorial component analysis showing scatter plots of nine populations of *Haliotis midae* along factors 1 and 2. The West coast populations are encircled.

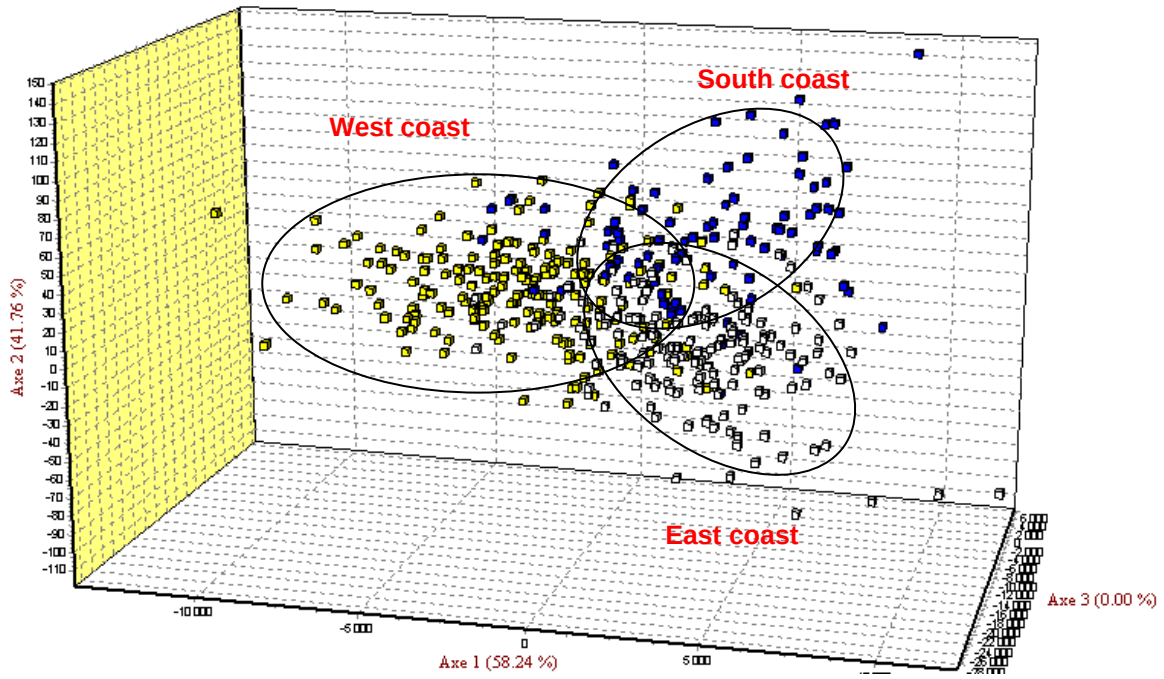

Figure 3.3 Factorial component analysis showing scatter plots of the individual genotypes obtained with eight microsatellite loci

The average values of relatedness (r_{xy}) among individuals within each population were mostly negative and relatively low indicating that sampling within populations was probably at random and could not have biased results significantly.

The AMOVA analysis corroborated the weak differentiation of the West coast populations from the rest. A significant value (F_{CT} = 0.009, *P* = 0.007, 0.93% of the variation) was obtained only when the populations were divided into groups east and west of Cape Agulhas. Most of the variance was distributed between individuals within populations (97.70%) while very little variance was distributed among population within groups (1.37%). Analysis assuming three groups did not support the existence of three groups postulated by an additional barrier at Port Elizabeth (F_{CT} = 0.007, *P* = 0.033, 0.71% of the variation).

With STRUCTURE, averaged over replicates, the highest likelihood values was observed at *K*=8 but when the rate of variation in likelihood values between successive *K's* (Δ*K* statistic of Evanno *et al.,* 2005) was analysed, the number of *K*

was subsequently reduced to two (Figure 3.4). When Δ*K* was plotted against possible *K*'s, the pattern was reflective of the contact zone situation depicted in the simulation study of Evanno *et al.* (2005). His study investigated the effect on Δ*K* under three different migration models, the island, the hierarchical island and the contact zone model and found that in the latter model *K*=2 corresponds to the highest level of structuring.

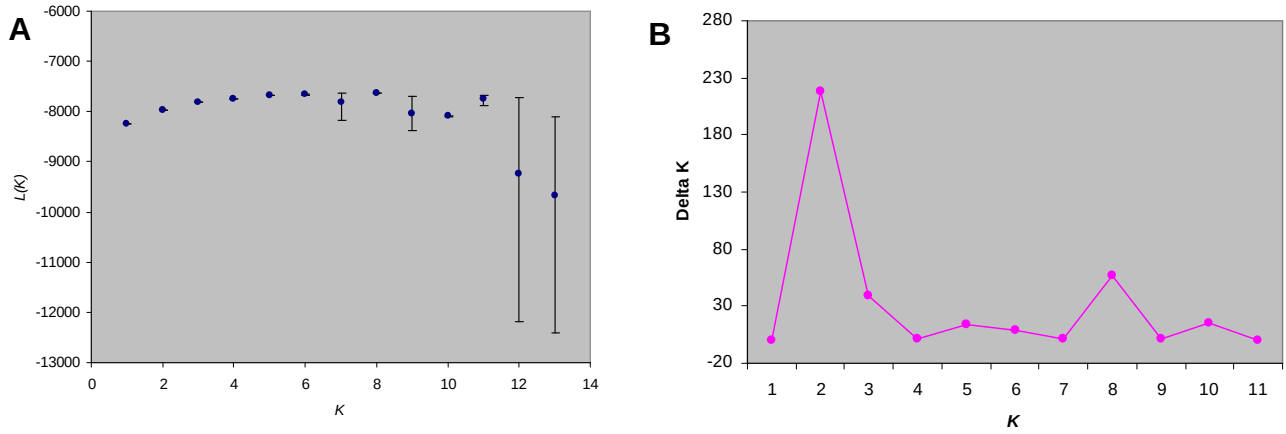

Figure 3.4 (A) The posterior probability of the data, L(*K*) for each *K* and (B) Δ*K* as a function of *K* following Evanno *et al.* (2005)

While the estimated proportion of admixture between the two clusters supported the FCA and AMOVA results during the forced structure, the non-forced model showed a high level of admixture between the two groups (Table 3.4). Similarly, assignment of individuals based on prior information of pure West and East coast individuals indicated high levels of admixture and only 60% of each of the West and East coast population's genomes could be assigned to the correct cluster of origin. The genetic composition and the subtle break between the western and eastern populations were more evident from the population summary plot based on 20% of the highest likelihood runs permutated in CLUMPP (Figure 3.5).

Table 3.4 Estimation of the percentage of individuals clustered in the proposed West (1) and East (2) coast populations using the forced, non-forced and popinfo options in STRUCTURE (Pritchard *et al.,* 2000)

	Forced structure		Non-forced structure		Prior Popinfo	
	1	*2*	*1*	*2*	*1*	*2*
West coast (1)	0.957	0.043	0.518	0.482	0.597	0.403
East coast (2)	0.04	0.96	0.459	0.541	0.418	0.582

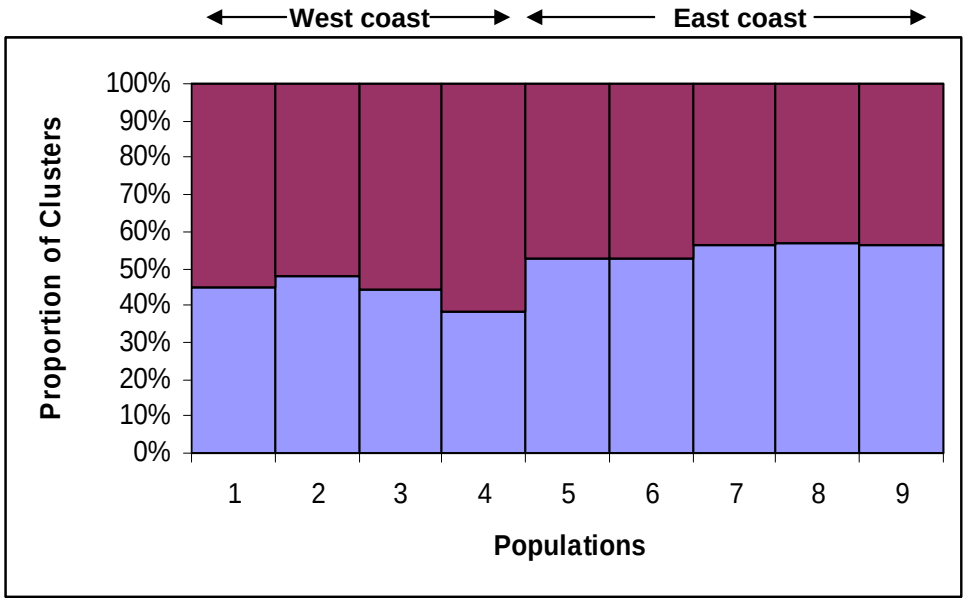

Figure 3.5 Proportion of the model-based clusters (*K*=2) in the ancestry of nine populations. Populations SD, RI, KL, GB, WS, MB, CR, RP and BR are represented by numbers one to nine.

Clustering results with TESS varied quite significantly between the different interaction parameters tested. The best level of clustering was found for Ψ=0.5 and Ψ=0.6 while values higher than 0.7 resulted mostly in single clusters. Maximum *K* however did not have a substantial effect on the likelihood values and respective number of clusters obtained. For each combination of interaction parameter and maximum *K*, data of the 20% highest likelihood runs were combined in CLUMPP to obtain an average closer to the true Bayesian estimate than the highest likelihood run. With DISTRUCT, the permutation effect on the membership coefficients obtained with CLUMPP could be visualized for each cluster. All the combinations tested

(Ψ=0.5 and 2≤K_{max}≤5) resulted in a bidirectional distribution of clusters along the geographical range (Figure 3.6B). In most cases, cluster membership was consistent with clinal variation of allele frequency and the gradients were observed on both sides of Cape Agulhas. This was in full agreement with the FCA and STRUCTURE analysis, showing genetic discontinuity between populations west and east of Cape Agulhas and the possible presence of two metapopulations.

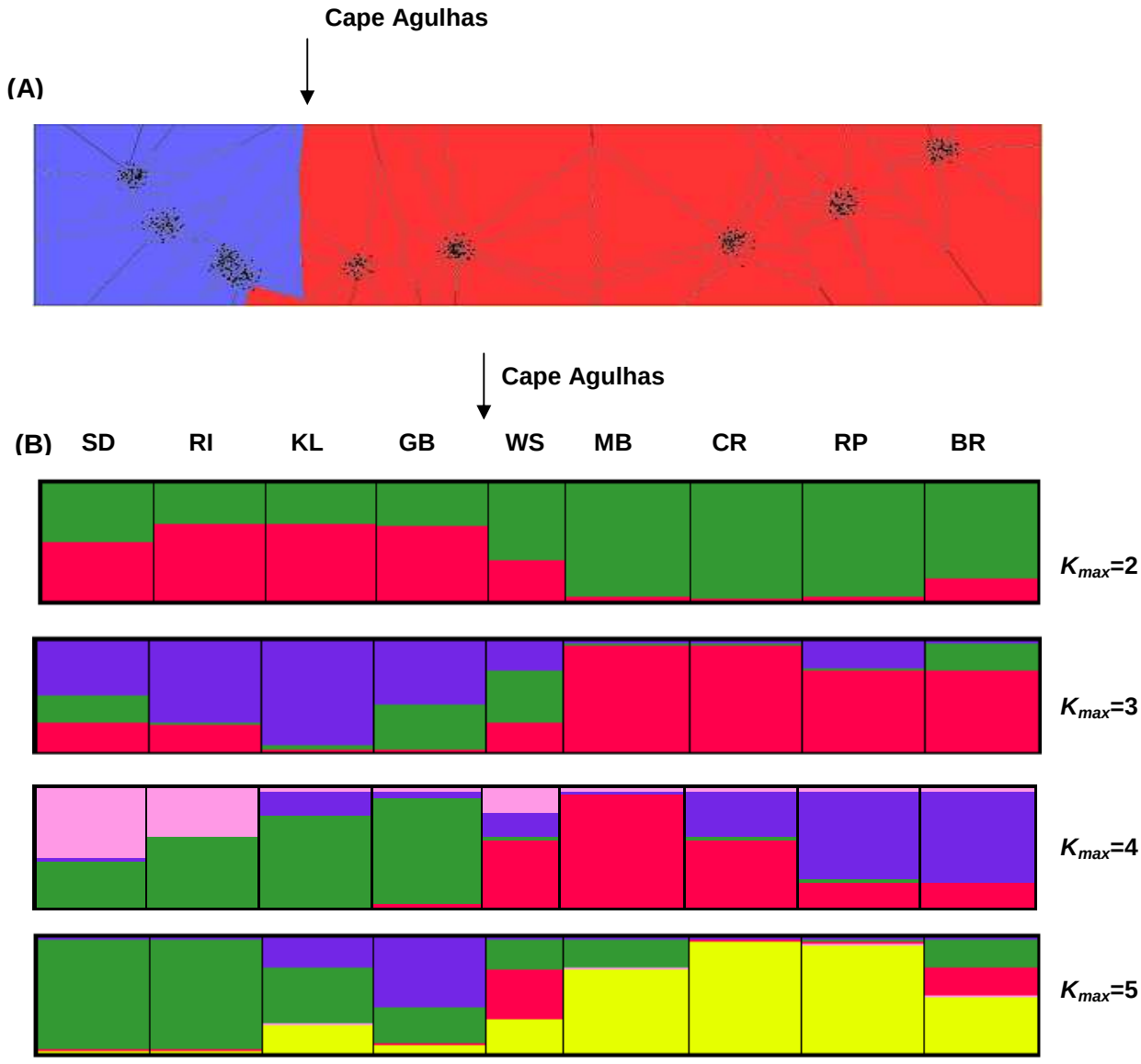

Figure 3.6 Estimated cluster configuration for nine *H. midae* populations based on A) a single maximum likelihood run (Ψ=0.5 and K_{max}=2) and B) CLUMPP permutation of the 20% highest likelihood runs (Ψ=0.5 and 2≤ K_{max}≤ 5)

Furthermore, the Witsand population consistently showed a mixture of the major clusters present within either metapopulation, possibly representing a hybrid admixture zone. Most importantly, TESS results support the hypothesis of restriction to gene flow and the existence of a biogeographic barrier between the West and East coast populations.

BARRIER identified four barriers with decreasing importance (A to D). The first barrier provided additional support for a change in genetic composition around the area of Cape Agulhas as it was located between the populations of Gansbaai and Witsand. Based on the genetic distance of Nei (1987), this barrier was confirmed at six of the eight loci when data was analysed separately for each microsatellite. Two additional barriers were identified within the West coast while another one was identified between Mossel Bay and Cape Recife (Figure 3.7). Additional barriers supported by less than three loci were not considered as significant restriction to gene flow. When the robustness of the four barriers was tested by including θ and R_{ST} matrices, all data sets indicated the west-east coast barrier as the most important one. Interestingly, the second highest support was given to a barrier identified between the South and East coast (Mossel Bay and Cape Recife) while the two west coast barriers were only supported by one of the matrices.

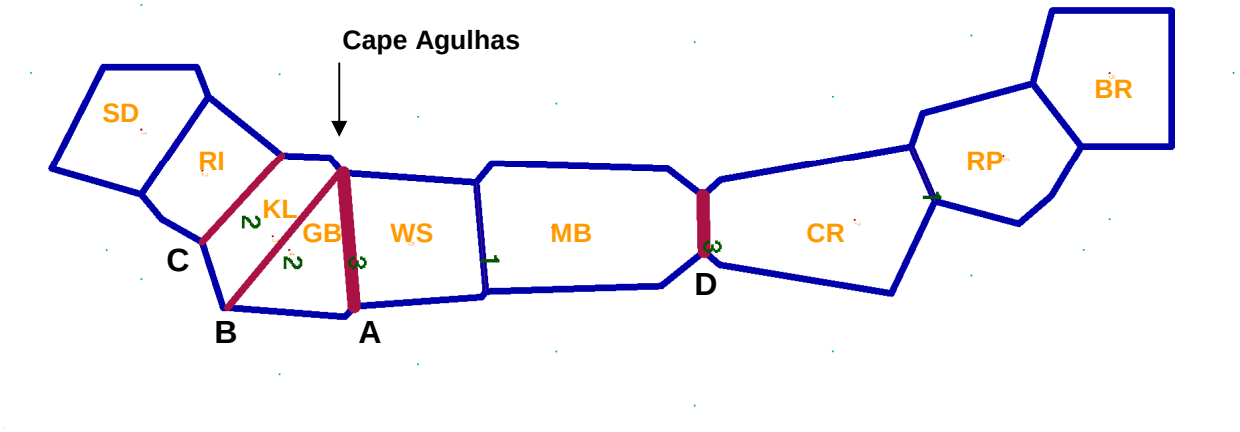

Figure 3.7 Location of the four barriers to gene flow identified with BARRIER (indicated in red). The barrier order reflects the level of importance (A-D) while the barrier width indicates the number of matrices supporting the data (1, 2 or 3).

3.3.5 Effective Population Size and Historical Demographic Changes

The effective population size of the population prior to any demographic change was estimated from the expected heterozygosity in HWE overall and for each locus. It is evident that very different values were obtained for each locus (Table 3.5) which is likely an indication of the stochasticity of the mutation process involved. For the two different mutational models, SMM and IAM, the mean N_e over all loci was 120,286 and 17,313 respectively. Contemporary N_e was estimated using the maximum likelihood estimations of θ. Because of the very high allele number for loci *HmD36* and *HmD102*, MISAT failed to estimate a realistic maximum likelihood value of theta for either of them. A ratio between the maximum SMM long-term and contemporary estimations of N_e was therefore estimated excluding these loci and indicated an estimated two-fold increase in population size (see Mean 2 of Table 3.5). According to the IAM, the population size increased approximately eight times. The method of Cornuet and Luikart (1996, 1998) in BOTTLENECK showed an imbalance between the two heterozygosity (H) estimates based on HW heterozygosity excess and

heterozygosity expected from the number of alleles per locus, thereby also indicating population expansion. Overall loci, the Wilcoxon signed rank test indicated departure from mutation-drift equilibrium under TPM (P one tail for H deficiency= 0.01563) and SMM (P one tail for H deficiency= 0.00781) and equilibrium under the IAM. These results indicate either departure from the tested mutational models or demographic expansion.

Table 3.5 Long-term and contemporary effective population size (N_e) based on heterozygosity and maximum likelihood values of theta. Mean 2 for N_e(IAM) was calculated excluding the loci highlighted in yellow.

Loci	H n.b.	Ne(SMM)	Ne(IAM)	θML	Contemp Ne
HmD14	0.583	5932.368	3492.649	24.83	62075
HmD36	0.940	345972.222	39166.667	67	167500
HmD55	0.872	75043.945	17031.250	45.9	114750
HmD59	0.898	118530.476	21972.482	43.95	109875
HmA11	0.622	7500.927	4114.729	5.98	14950
HmSP5	0.599	6514.545	3730.788	48.83	122075
HmD102	0.943	383483.764	41359.649	107.2	268000
CmrHr2.15	0.753	19310.860	7639	30.5	76250
Mean		120286.1386	17313.432		116934.375
Mean 2		29104.14	7247.642		62496.875

H n.b. = unbiased expected heterozygosity; θML= maximum likelihood estimate of theta

The *k* and *g* statistics calculated with Kgtests (Bilgin, 2007) are shown in Table 3.6. For the *k* test, five of the nine populations had a significant high number of loci with negative values, indicative of population growth. For the interlocus *g* test, the low levels of the *g* statistic indicated signatures of expansion but all the values were higher than the 5% cutoff and therefore not significant. A global *k* test that was performed on all populations resulted in a significant *P*-value (0.013) while the *g* test gave g-ratio values for all populations that were not within but close to the 5% significance level (see Reich 1999). Therefore, the results of both tests support the

hypothesis that this species has undergone some level of population size expansion rather than contraction.

Table 3.6 Tests of population expansion applying the *k*-test and *g*-test (Reich *et al.*, 1999). For the k test, the number of loci that returned significant *P* values is listed in the first column. The *P* value overall loci is shown in the second column. *Significant *P*-values

	k-test		*g*-test	
Population	Neg Value	*P*-value	*g*-ratio	*P*-value
Saldanha	7	0.029*	2.30	*NS*
Robben Island	6	0.126	2.39	*NS*
Kleinmond	6	0.126	1.71	*NS*
Gansbaai	7	0.029*	2.30	*NS*
Witsand	6	0.126	1.17	*NS*
Mossel Bay	6	0.126	1.86	*NS*
Cape Recife	8	0.003*	0.82	*NS*
Riet Point	7	0.029*	1.80	*NS*
Black Rock	7	0.029*	1.30	*NS*

In MIGRATE, the maximum-likelihood estimates of theta overall loci ranged from 0.155 for Cape-Recife to 0.353 for Black Rock (Table 3.7). Although these values give only an indirect estimation of contemporary effective population size, they did not show any significantly small or large population to be associated with a particular region. Except for the populations at the extreme locations, populations within the different coastal regions (West, South and East coast) showed comparable theta values which could indicate that long-term population sizes are related to conditions within the biogeographic provinces. In contrast to theta estimates, the long-term migration values did vary somewhat between the three different runs but were generally within the same range (7.5 to 31) and mostly symmetrical. Of relevance to this study is the central population of Witsand exchanging considerably more migrants with populations on the same side of the proposed barrier at Cape Agulhas

than with populations across the barrier (Table 3.7). It should be stressed that in this study MIGRATE estimates were not taken at face value but rather used to compare values of scaled historic effective population size and migration rate between populations.

Table 3.7 Average estimates of theta (ɵ) and migration (*M*) based on three independent MIGRATE runs. The number of migrants Witsand exchanges with the West vs the East coast populations is indicated within the circles.

Receiving population	ɵ	*M* Donor population								
		SD	RI	KL	GB	WS	MB	CR	RP	BR
Saldanha	0.213	-------	17.13	15.52	17.92	16.71	13.61	15.74	13.77	12.92
Robben Island	0.314	18.61	-------	23.87	17.43	14.93	16.29	22.37	12.59	15.05
Kleinmond	0.348	13.13	22.94	-------	14.13	11.15	22.91	21.29	19.34	11.28
Gansbaai	0.329	10.41	14.37	9.97	-------	8.57	23.23	16.87	13.63	11.29
Witsand	0.298	12.60	15.24	13.43	11.08	-------	26.06	28.15	16.49	19.58
Mossel Bay	0.291	12.64	13.42	22.96	18.92	28.36	-------	14.53	18.98	19.66
Cape Recife	0.155	15.51	26.28	31.00	21.82	21.78	18.25	-------	20.18	22.21
Riet Point	0.178	15.49	17.98	28.80	17.77	21.99	20.67	18.66	-------	15.65
Black Rock	0.353	10.25	12.75	12.76	12.86	20.21	24.21	12.79	7.84	-------

3.4 Discussion

3.4.1 Genetic diversity

Moderate to high levels of genetic diversity characterized all nine populations including the Robben Island population, the only population expected to show signs of a genetic isolation. Genetic diversity, as represented by number of alleles, allelic richness and expected heterozygosity, of South African abalone populations was comparable to that previously observed in other *Haliotis* species (Evans *et al.,* 2004; Li *et al.,* 2004; Tang *et al.,* 2005; Hara and Sekino, 2007; Imron *et al.,* 2007). Li *et al.* (2004), for example, surveyed six microsatellite loci in wild populations of Pacific

abalone and reported an average number of alleles per locus and observed heterozygosity of 22.4 and 0.711 respectively. In this study, the number of alleles was slightly higher than that of a farmed population genotyped with five of the same microsatellite loci. The loss of allelic diversity in the farmed population was mostly attributable to the differential contribution of broodstock to the next generation (Slabbert, 2004). A distinct reduction in allelic variation has been observed in cultured populations of many other abalone species (Li *et al.,* 2004; Evans *et al.,* 2004; Hara and Sekino, 2007). The high microsatellite variation observed in this study is reflective of the larger effective population sizes expected of species inhabiting a marine environment. DeWoody and Avise (2000) reported a progressive increase in microsatellite variation from freshwater to anadromous to marine fishes. In their comparative study that included non-piscine taxa, the marine piscine group exhibited the highest average number of alleles per locus (19.9) as well as the highest heterozygosity averaged across all loci (0.79 ± 0.26). A variety of molecular, ecological, and evolutionary factors are known to impact levels of variation at microsatellite loci (Goldstein and Schlötterer, 1999) and the high genetic variability observed in *Haliotis midae* could therefore be ascribed to a combination of these factors.

Significant deviations from Hardy-Weinberg expectations were observed in almost all the locus-population samples tested for disequilibrium. This departure was attributed to homozygote excess at four of the eight microsatellites. While null alleles are the most likely cause for Hardy-Weinberg deviations, biological factors such as inbreeding, nonrandom mating, natural selection (e.g. sweepstakes in marine organisms) and population substructuring (Wahlund effect) are known to reduce heterozygosity levels significantly (Hedgecock, 1994; Castric *et al.,* 2001; Hoarau *et*

al., 2002; So *et al.,* 2006). Heterozygote deficits can also result from genotyping errors due to large-allele dropout or stuttering. Large-allele dropout occurs due to preferential amplification of smaller alleles (Wattier *et al.,* 1998) and slippage during PCR amplification produces stutter bands that differ in repeat unit lengths from the original amplification product (Shinde *et al.,* 2003). The excess of homozygosity in other species of abalone has previously been explained by inbreeding, artificial selection and cryptic population structure (Huang *et al.,* 2000, Withler *et al.,* 2003). For this data however, heterozygote deficiency varied between loci and since the levels of differentiation between populations (F_{ST}) were far less than the variation obtained within populations (F_{IS}), it was reasonable to reject for example the Wahlund effect (Garcia de Leon, 1997). In addition, MICRO-CHECKER analysis statistically inferred the presence of null alleles at four of the eight loci and therefore partially explains the disequilibrium present over all populations. Although mis-scoring due to stutter bands could result in significant deviations, three of the four loci with suspected null alleles were tetranucleotide repeats in which stutter bands are much weaker and not easily misinterpreted. Rose *et al.* (2006) suggested null alleles as the most plausible explanation for heterozygote deficiency in a study of Eastern oyster (*Crassostrea virginica*), supported by the high level of polymorphism surrounding the PCR priming sites for those loci (Reece *et al.,* 2004) while Li *et al.* (2007) reported over 80% of the loci used in a study on Pacific abalone (*Haliotis discus hannai*) to be prone to carrying null alleles. Although primer redesign can correct for heterozygosity deficiency, the highly polymorphic nature of this species will demand a relatively high number of alternative sequences to be considered for primer design. Alternatively, null allele frequencies can be adjusted using the methods of Brookfield (1996) or Chakraborty *et al.* (1992) implemented in MICRO-CHECKER. This was however not

applied to the current dataset since both methods assume single null alleles at each locus and multiple null alleles may be present at different frequencies for some of these loci. Where possible, data analysis was rather repeated excluding all or most of the microsatellite loci with putative null alleles.

3.4.2 Population differentiation

Despite a few statistically significant pairwise F_{ST} values after Bonferonni adjustment, the variance-based methods indicated no overall population differentiation along the species geographical range (F_{ST} = 0.014). Accepting the hypothesis of panmixia however needed further investigation with higher sensitivity analytical methods as various studies have recorded weak but significant structuring in marine animals despite population dynamics in a marine environment being best explained by low structure-high-gene-flow scenarios (Waples, 1998; DeWoody and Avise, 2000). One reason why low levels of genetic differentiation could have remained undetected with analysis of variance-based (F_{ST} / R_{ST}) methods is the high levels of polymorphism exhibited by the microsatellite loci used for this study. An empirical study by O'Reilly *et al.* (2004) revealed that the probability of estimating weak levels of divergence among populations diminishes when allelic richness and locus heterozygosity are high. This supports previous theoretical and simulation studies (Estoup *et al.,* 2002; Kalinowski, 2002) that indicated an advantage of low to moderate polymorphic loci over highly polymorphic markers in the detection of low-level population structure. This phenomenon is best explained by size homoplasy, an inherent characteristic of microsatellites that could make allelic distributions of loci seem more similar between samples than what it is in reality (Estoup and Cornuet, 1999; Estoup *et al.,* 2002). Size homoplasy occurs when different copies of a locus

are identical in state but not by descent and is dependant on the mutation model and type of repeat. In microsatellite loci, homoplasy is most probable under the step-wise mutation model (SMM) whereas the infinite allele model (IAM) is expected to always give rise to a new allelic state. Simulation studies also demonstrated that homoplasy is most likely to occur in the presence of high mutation rates (μ) and large effective population sizes (Balloux *et al.,* 2000; Estoup *et al.,* 2002). Since the size-based R_{ST} estimator is unaffected by mutation rate, R_{ST} *v*alues usually appear to be higher and more variable than frequency-based F_{ST} estimates in the presence of homoplasy which was indeed the case for *H. midae*. The high variance of R_{ST} in general together with the elevated possibility of size homoplasy in this study may explain why subtle population differences remained undetected by the variance-based methods.

In contrast to the homogeneity indicated among populations using variance-based methods, the exact test of genic differentiation (ETGD) showed significant differentiation in most of the pairwise comparisons. A higher level of significance associated with the ETGD than that assessed by F_{ST} values has been reported before (Garoia *et al.,* 2004; D'Amato *et al.,* 2006). Garoia *et al.* (2004) for instance obtained more than double the number of significant pairwise comparisons in red mullet when using the exact test instead of the F-statistic analyses. This discrepancy is a direct result of the different way in which these two estimators exploit genetic data. While F-statistics compare populations evenly across allelic ranges, ETGD gives more weight to rare alleles and thus is more effective in detecting subtle heterogeneity among populations. The majority of the microsatellite loci used in this study exhibited a high number of low frequency alleles, thereby explaining the higher number of significant pairwise tests obtained with the exact test.

The factorial component analysis (FCA) allowed for a more visual interpretation of association between populations. A separation of the West coast populations from the rest was evident on a population as well as an individual level while the latter was less accentuated due to the high proportion of overlapping individuals. The FCA plot representing individuals from three different regions (west vs south vs east) allowed for a more distinct association pattern between groups with some degree of overlapping between their individuals. However, none of the FCA plots mirrored isolation-by-distance which indicates that historical fragmentation rather than geographic distance was responsible for the separation of the two groups. The hierarchical analysis of variance (AMOVA) supported multivariate analysis by revealing a weak but statistically significant interregional (east vs west) structure between groups. At the same time it corroborated the AMOVA results obtained by Evans *et al.* (2004) which compared a West and East coast group containing three populations each. Using only three non-species-specific microsatellite loci, they observed significant differentiation between the western and eastern localities with an overall F_{CT} of 0.033. In both studies, AMOVA failed to show evidence of further structuring within either group.

The traditional population genetic analyses performed so far assumed *a priori* grouping of individuals into populations that could lead to underestimation of gene flow or the underlying biological processes involved. Hence data were subjected to more specialized programs that rely on new and advanced methodologies with increased ability to detect cryptic population structure. The Bayesian clustering method of Pritchard *et al.* (2000) is able to assign individuals to populations based on their genotypic frequencies without making *a priori* characterization on groups. By including the correlated allele frequency model, the software also has greater power

in detecting distinct but similar populations (Falush *et al.,* 2003). While the log probability of data allows for the estimation of the more likely number of genetically homogeneous groups present, the authors warned that this is only an *ad hoc* guide and not necessarily the true number of clusters (*K*) detected. Evanno *et al.* (2005) suggested that the rate at which the log probability of data changes between successive *K*s provides a more reliable estimator of the number of clusters. In this study, STRUCTURE indicated the presence of two clusters following the method of Evanno *et al.* (2005). While individual assignment analysis failed to assign the majority of individuals to their true population of origin, the summary plot based on the permutated data showed some degree of difference in the genomic composition of individuals west and east of Cape Agulhas (*K*=2). Under a non-forced model with no prior information on populations, admixture was very high and when the ability of STRUCTURE to assign individuals to pre-defined East and West coast groups was tested, only 60% of the individuals were assigned correctly. According to Waples and Gaggiotti (2006), the ability of STRUCTURE to estimate the true number of clusters or correctly assign individuals to populations of origin decreases with higher gene flow, fewer loci and smaller number of individuals tested. For example, Mäkinen *et al.* (2006) reported the insensitivity of STRUCTURE analysis in marine populations of three-spined stickleback (*Gasterosteus aculeatus*) when allele frequency differentiation was low (F_{ST}<0.03). Furthermore, a comparison study by Latch *et al.* (2006), evaluating the performance of three Bayesian clustering methods, found that STRUCTURE worked well for inferring the number of clusters at low levels of differentiation but that its optimal percentages of assignment accuracy was only reached when F_{ST} was at least 0.05. Given the low F_{ST} obtained for *H. midae*, the weak assignment performance obtained with STRUCTURE was not particularly

unexpected but did necessitate further confirmation of the two-population model proposed by Δ*K*.

In recent years, more emphasis has been placed on the addition of spatial or geographic coordinate data to define areas of genetic discontinuity. This followed a debate on whether Bayesian clustering algorithms are repeatable and robust in the detection of well differentiated groups regardless of experimental design (Serre and Pääbo, 2004; Rosenberg *et al.,* 2005; François *et al*., 2006). A number of studies arose in which the inclusion of spatial priors proved to solve cluster assignment dilemmas where results were previously inconsistent (Wasser *et al.,* 2004; Guillot *et al.,* 2005). The program TESS incorporates spatial information through the hidden Markov random field (HMRF) approach (Chen *et al.,* 2007). HMRFs represent the prior geographical distribution of each individual and assume a model where populations are more likely to share cluster membership with nearby sites than with geographically distant populations. TESS furthermore assumes spatial interaction between individuals and allows control over the importance given to this interaction by means of the interaction parameter (Ψ). François *et al.* (2006) for example suggested a value of $0,5 < \Psi < 1$ for datasets of less than 10 populations. In this study, TESS results were interpreted as the average proportion of an individual genome assigned to each of the maximum number of clusters predefined in the admixture version. Assuming that *K* was correctly estimated, the pattern that emerged was highly representative of a natural break in continuous populations between the West and East coast sampling locations and once again indicative of a barrier separating the two regions. In contrast to the STRUCTURE results, a high number of individuals (>90%) were assigned to the population of origin. A recent study by Chen *et al.* (2007) evaluated the performance of four Bayesian clustering

programs including TESS, STRUCTURE, GENELAND and GENECLUST and reported that TESS performed extremely well in assignment estimates at low levels of genetic differentiation. Of particular importance was the hypothesis that TESS is superior to the other Bayesian clustering methods in the detection of recent contact zones between weakly differentiated populations. As a result, greater validity could be given to the maximum number of clusters detected with STRUCTURE in this study, also implying synonymy to a contact zone scenario following the method of Evanno *et al*. (2005). The comparison study of Chen *et al.* (2007) also provided evidence that spatial Bayesian clustering methods are as efficient as nonspatial Bayesian clustering methods in the identification of genetic discontinuities, especially when the number of polymorphic loci is limited. As most of the population genetic analysis methods discussed above indicated subdivision between the western (SD, RI, KL, GB) and eastern (WS, MB, CR, RP, BR) sampling sites and the molecular marker applied in the next chapter could confirm this pattern of gene flow, the evolutionary mechanisms which potentially shaped this structure in *H. midae* will be further discussed in the final chapter of this thesis.

3.4.3 Population history

The recognition of recent bottlenecks or relatively rapid decline in populations has become increasingly important in population genetics, conservation biology and evolutionary theory of many species. Effective population size (N_e) has been the key parameter in monitoring fluctuations in population demographics and by comparing historical N_e with contemporary N_e it is possible to infer the type and magnitude of demographic change. Identifying populations that have recently suffered severe reduction in size are of particular importance to endangered species or species that

have recently become threatened. As in *H. midae*, populations decreasing in size are expected to hold the highest risk of extinction and appropriate management of such populations are necessary to keep the entire species out of danger. In this study, estimating N_e is not only important for the conservation of *H. midae* but the demographic history of populations might assist in verifying the founder event proposed by Evans *et al.* (2004) while the migration rate could confirm levels of gene flow between adjacent populations.

The Wilcoxon signed rank and the Kg tests resulted in no significant outcome for any particular population and the overall data suggested an increase in population size. A two-fold increase in population size was confirmed when the N_e based on expected heterozygosity and the maximum likelihood method of Nielsen (1997) were compared. The effective population sizes might be unrealistic as the same mutation rate was assumed for all loci, but when the loci with null alleles or potentially divergent mutation rates were excluded, the results were still indicative of an increase in population size. In the case of *H. midae*, where certain areas have been highly exploited in the last couple of years, it would be expected to detect population decline rather than expansion. Because microsatellites are supposed to be more informative for recent demographic events due to their high mutation rate, it is interesting that this study is more consistent with results based on mtDNA which also detected population expansion in species that are at present in severe population decline (Lavery *et al*., 1996; Koskinen *et al.,* 2002; Saillant, 2004). However, irrespective of the type of molecular marker used, taking the species life cycle and relatively large population sizes into account, it is perhaps unlikely that the genetic signature of such a recent population decline will already be detectable in the present-day populations. The identification of single populations with particularly small N_e is probably of more value

to the conservation of *H. midae.* For that reason, population-specific theta of MIGRATE was used to investigate size variation between populations. By comparing theta values instead of actual effective population size, there was also no need to choose a single arbitrary mutation rate for all eight loci. Although theta values did not vary significantly between populations, values remained consistent between runs and were indicative of the reliability of the results. Due to computational constraints, MIGRATE was run with only a subset of the data and even though this is commonly accepted (Austin *et al.,* 2004; Hänfling and Weetman, 2006), it may not be representative of the populations as a whole. Nevertheless, the fact that no populations with significantly smaller theta values were detected suggests that historical effective population sizes were relatively stable throughout the species' distribution range and is in direct conflict to the founder event proposed by Evans *et al.* (2004). If the East coast populations resulted from founder dispersal into waters east of Cape Agulhas, it would be expected to see significantly lower historical population sizes in at least samples from the extremes of the distribution range. The estimates of contemporary migration implied that all the populations exchanged migrants with each other while the only pattern consistent with the projected restriction to gene flow, was the significantly lower number of migrants that Witsand received from Kleinmond, the population just west of Cape Agulhas. Since migration estimates are affected by θ which in turn could be biased because of analytical constraints, it is difficult to make conclusive comments on the magnitude of migration between populations, but given the consistency of the results between runs, the observed pattern of migration could in fact reflect the true nature of gene flow between *H. midae* populations.

3.5 Conclusions

In summary, this chapter highlights two of the most controversial issues regarding the population genetic structure of species within a marine environment. Firstly, the results add to the growing evidence that, although subtle, significant population differentiation does exist in species with high gene flow potential and no apparent ecological or hydrographical barriers to dispersal. Secondly, it becomes evident throughout the study that no single particular factor can be responsible for the present-day structuring of populations but that most patterns of gene flow are explained by interplay of both historical and contemporary mechanisms. The single most important feature detected in this study is the genetic discontinuity between West and East coast *H. midae* populations which coincides with the biogeographic barrier at Cape Agulhas. Moreover, the pattern of genetic structure probably indicates a secondary contact between historically isolated western and eastern populations. The hypothesis-testing indicated that although geographical distance *per se* had no effect on structuring, the evidence against panmixia is substantial enough to reject the expected model of unrestricted gene flow between *H. midae* populations. Conversely, evidence for a genetic break in the area of the East coast thermal front could not be supported by a majority of the analyses. From a technical perspective the study advocated for the many ways in which data should be utilized to explain a number of potential scenarios. By including recently developed methods, it was possible to justify the observed pattern of gene flow, which could otherwise have been unresolved or remained undetected. Finally, demographic data provided no evidence for recent population decline but rather suggest that *H. midae* experienced population size expansion in the past. Secondary contact following historical isolation along with present day restriction to gene flow is probably more relevant to

contemporary populations than the founder event proposed by previous structure analysis of *H. midae*.

References

Austin JD, Lougheed SC, Boag PT (2004) Controlling for the effects of history and nonequilibrium conditions in gene flow estimates in northern bullfrog (*Rana catesbeiana*) populations. *Genetics* **168:** 1491–1506.

Beckley LE and Van Ballegooyen RC (1992) Oceonographic conditions during three ichtyhoplankton surveys of the Agulhas Current in 1990/1991. *South African Journal of Marine Sciences* **12:** 83–93.

Beerli P and Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA.* **98:** 4563–4568.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1996-2004) GENETIX, logiciel sous Window TM pour la génétique des populations. Laboratoire Génome et Populations, Montpellier, France.

Bentzen P, Taggart CT, Ruzzante DE, Cook D (1998) Microsatellite polymorphism and the population structure of atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences* **53:** 2706–2721.

Bester AE, Slabbert R, D'Amato ME (2004) Isolation and characterization of microsatellite markers in the abalone *Haliotis midae. Molecular Ecology Notes* **4:** 618–620.

Bilgin R (2007) Kgtests: a simple Excel macro program to detect signatures of population expansion using microsatellites. *Molecular Ecology Notes* **7:** 416–417.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5:** 4534–4555.

Carvalho GR and Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* **4:** 326-350.

Carvalho GR and Hauser L (1998) Advances in molecular analysis of fish population structure. *Italian Journal of Zoology* **65:** 21–33 Suppl S.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierachical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* **55:** 1016–1028.

Chakraborty R, Deandrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* **56:** 45–57.

Chambers MD, VanBlaricon GR, Hauser L, Utter F, Friedman CS (2006) Genetic structure of black abalone (*Haliotis cracherodii*) populations in the California islands and central California coast: impacts of larval dispersal and decimation from Withering syndrome. *Journal of Experimental Marine Biology and Ecology*: **331:** 173–185.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Resources* **7:** 747–756.

Corander J, Marttinen, P, Mäntyniemi, S (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163:** 367–374.

Cornuet JM and Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144:** 2001–2014.

Cornuet JM and Luikart G (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12:** 228–237.

D'Amato ME (2006) Demographic expansion and subtle differentiation in the long-tailed hake *Macruronus magellanicus*: Evidence from microsatellite data. *Marine Biotechnology* **8:** 2228–2236.

DeWoody JA and Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* **56:** 461– 473.

Elliott NG, Evans B, Conod N, Bartlett J, Officer R, Sweijd N (2000) Application of molecular genetics to the understanding of abalone population structure - Australian and South African case studies. *Journal of Shellfish Research* **19:** 510–524.

Estoup A and Cornuet JM (1999) Microsatellite evolution: inferences from population data. In: *Microsatellites: Evolution and Applications*. (ed. Goldstein DB and Schlötterer C), Oxford University Press, Oxford, pp49–64.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11:** 1591–1604.

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

Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliot NG (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology Progress Series* **270:** 163–172.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164:** 1567–1587.

Fauvelot C, Lamaire C, Planes S, Bonhomme F (2007) Inferring gene flow in coral reef fishes from different molecular markers: which loci to trust? *Heredity* **99:** 331–339.

François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random field in spatial population genetics. *Genetics* **174:** 805–816.

Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with non-invasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6:** 861–868.

Garcia de Leon FJ, Chikhi L, Bonhomme F (1997) Microsatellite polymorphism and population subdivision in natural populations of European Sea Bass *Dicentrarchus labrax* Linnaeus, 1758). *Molecular Ecology* **6:** 51–62.

Garoia F, Guarniero I, Piccinetti C, Tinti F (2004) First microsatellite loci of red mullet (*Mullus barbatus*) and their application to genetic structure analysis of Adriatic shared stock. *Marine Biotechnology* **6:** 446–452.

Goldstein DB and Sclötterer C (1999) Microsatellites: Evolution and Applications. Oxford: Oxford University Press, 368pp.

Goodman SJ (1997). RSTCALC: a collection of computer programs for calculating unbiased estimate of genetic differentiation and gene flow from microsatellite data and determining their significance. *Molecular Ecology* **6:** 881–885.

Gruenthal KM, Acheson LK, Burton RS (2007) Genetic structure of natural populations of California red abalone (*Haliotis rufescens*) using multiple genetic markers. *Marine Biology* **152:** 1237–1248.

Guillot G, Estoup A, Mortier F, Cosson J (2005) A spatial statistical model for landscape genetics. *Genetics* **170:** 1261–1280.

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48:** 361–372.

Hamm DE and Burton RS (2000) Population genetics of black abalone, *Haliotis cracherodii*, along the central California coast. *Journal of Experimental Marine Biology and Ecology* **254:** 235–247.

Hancock JM (1999) Microsatellite and other simple sequences: genomic context and mutational mechanisms. In: *Microsatellites, Evolution and Application*. (ed. Goldstein DB and Schlötterer C), Oxford, Oxford University Press, pp1–9.

Hänfling B and Weetman D (2006) Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the River Sculpin, *Cottus gobio*. *Genetics* **173:** 1487–1501.

Hara M and Sekino M (2007) Genetic differences between hatchery stocks and natural populations in pacific abalone (*Haliotis discus*) estimated using microsatellite DNA markers. *Marine Biotechnology* **9:** 74–81.

Hedgecock D (1994) Does variance in reproductive success limit effective population size of marine organisms? In: *Genetics and evolution of aquatic organisms*. (ed. Beaumont A), Chapman and Hall, London, pp122–134.

Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research* **23:** 379–385.

Hoarau G, Rijnsdorp AD, Van der Veer HW, Stam WT, Olsen JL (2002) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. *Molecular Ecology* **11:** 1165–1176.

Huang BX, Peakall R, Hanna PJ (2000) Analysis of genetic structure of blacklip abalone (*Haliotis rubra*) populations using RAPD, minisatellite and microsatellite markers. *Marine Biology* **136:** 207–216.

Hutchinson WF, Carvalho GR, Rogers SI (2001) Marked genetic structuring in localized spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series* **223:** 251–260.

Imron, Jeffrey B, Hale P, Degnan BM, Degnan SM (2007) Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. *Molecular Ecology* **16:** 289–304.

Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity* **88:** 62–65.

Knutsen H, Jorde PE, André C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology* **12:** 385–394.

Koskinen M, Knizhin I, Primmer CP, Schlötterer C, Weiss S (2002) Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal. *Molecular Ecology* **11:** 2599–2611.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE Jr (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7:** 295–302.

Lavery S, Moritz C, Fielder DR (1996) Indo-Pacific population structure and evolutionary history of the coconut crab *Birgus latro*. *Molecular Ecology* **5:** 557–570.

Li Q, Shu J, Yu R, Tian C (2007) Genetic variability of cultured populations of the Pacific abalone (*Haliotis discus hannai* Ino) in China based on microsatellites. *Aquaculture Research* **1:** 1–10.

Mäkinen HS, Cano JM, Merilä J (2006) Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. *Molecular Ecology* **15:** 1519–1534.

Mileikovsky SA (1971) Types of larval development in marine bottom invertebrates, their distribution and ecological significance: a re-evaluation. *Marine Biology* **10:** 193–213.

Monmonier MS (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* **5:** 245–261.

Moritz C (1994) Defining “evolutionary significant units” for conservation. *Trends in Ecology and Evolution* **9:** 373–375.

Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press, 356pp.

Nielsen EE, Hansen MM, Meldrup D (2006). Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L): implications for inferring population structure in nonmodel organisms. *Molecular Ecology* **15:** 3219–3229.

Nielsen R (1997) A maximum likelihood approach to population samples of microsatellite alleles. *Genetics* **146:** 711–716.

O’Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye Pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology* **13:** 1799–1814.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90:** 502–503.

Poissant J, Knight TW, Ferguson MM (2005) Nonequilibrium conditions following landscape rearrangements: the relative contribution of past and current

hydrological landscapes on the genetic structure of a stream-dwelling fish. *Molecular Ecology* **14:** 1321–1331.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Raymond M and Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86:** 248–249.

Reece KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK (2004) Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity* **95:** 346–352.

Reich DE and Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences, USA* **95:** 8119–8123.

Reich DE, Feldman MW, Goldstein DB (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions. *Molecular Biology and Evolution* **16:** 453–466.

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

Ridgeway TM, Stewart BA, Branch GM, Hodgson AM (1998) Morphological and genetic differentiation of *Patella granularis* (Gastropoda: Patellidae) along the coast of South Africa. *Journal of Zoology, London* **245:** 317–333.

Rose CG, Paynter KT, Hare MP (2006) Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity* **97:** 158–170.

Rosenberg N, Mahajan S, Ramachandran S, Zhao C, Pritchard J, Feldman M (2005) Clines, clusters, and the effect of study design on the influence of human population structure. *PLoS Genetics* **1:** e70

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137–138.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences* **81:** 8014–8019.

Saillant E, Patton JC, Ross KE, Gold JR (2004) Conservation genetics and demographic history of the endangered Cape Fear shiner (*Notropis mekistocholas*) *Molecular Ecology* **13:** 2947–2958.

Serre D and Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research* **14:** 1679–1685.

Shinde D, Lai YL, Sun FZ, Arnheim N (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites. *Nucleic Acid Research* **31:** 974–980.

Shipley P (2003) MICRO-CHECKER, Version 2.2.3 Department of Biological Sciences, University of Hull, Hull, UK.

Silva EP and Skibinski DOF (2009) Allozymes and nDNA markers show different levels of population differentiation in the mussel *Mytilus edulis* on British coasts. *Hydrobiologia* **620:** 25–33

Slabbert R (2004) Molecular analysis of genetic variation and relationships within the population of abalone (*Haliotis midae*) at the Sea Plant Products abalone hatchery, Hermanus, RSA. M.Sc. thesis, University of Stellenbosch, South Africa, 162pp.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457–462.

So N, Maes GE, Volckaert FAM (2006) Intra-annual genetic variation in the downstream larval drift of sutchi catfish (*Pangasianodon hypophthalmus*) in the Mekong River *Biological Journal of the Linnean Society* **89:** 719–728.

Tang S, Tassanakajon A, Klinbunga S, Jarayabhand P and Menasveta P (2005) Population structure of tropical abalone (*Haliotis asinina*) in coastal waters of Thailand determined using microsatellite markers. *Marine Biotechnology* **6:** 604–611.

Taylor MS and Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* **299:** 107–109.

Teske PR, Froneman CD, Barker NP, McQuaid CD (2007) Phylogeographic structure of the caridean shrimp *Palaemon peringueyi* in South Africa: further evidence for intraspecific genetic units associated with marine biogeographic provinces. *African Journal of Marine Science* **29:** 253–258.

Teske PR, McQuaid CD, Froneman CD, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology Progress Series* **314:** 283–293.

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.

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

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

Wasser S, Shedlock A, Comstock K, Ostrander E, Mutayoba B, Stephens M (2004) Assigning African elephants DNA to geographic region of origin: applications to the ivory trade. *Proceedings of the National Academy of Sciences, USA* **101:** 14847–14852.

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.

Weir BS and Cockerham CC (1984). Estimation *F*-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Widmer A and Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology and Evolution* **16:** 267–269.

Withler RE (2000) Genetic tools for identification and conservation of exploited abalone (*Haliotis* spp.) species. In: Workshop on rebuilding abalone stocks in British Colombia. (ed. Campbell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 101–110.

Withler RE, Campbell A, Li S, Miller KM, Brouwer D, Lucas BG (2001) High levels of genetic variation in northern abalone *Haliotis kamtschatkana* of British Columbia. Canadian Science Advisory Secretariat (CSAS) Research Document 2001/097, Fisheries and Oceans Science, Ottawa. 27pp. http://www.dfo-mpo.gc.ca/csas/

Withler RE, Campbell A, Li SR, Brouwer D, Supernault KJ, Miller KM (2003) Implications of high levels of genetic diversity and weak population structure for the rebuilding of northern abalone in British Columbia, Canada. *Journal of Shellfish Research* **22:** 839–847.

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

Zardi GI, McQuaid CD, Teske PR, Barker NP (2007) Unexpected genetic structure of indigenous (*Perna perna*) and invasive (*Mytilus galloprovincialis*) mussel populations in South Africa. *Marine Ecology Progress Series* **337:** 135–144.

Chapter IV

Verification of population structure within *Haliotis midae* using Single Nucleotide Polymorphism (SNP) markers

4.1 Introduction

Single variable sites, better known as single nucleotide polymorphisms (SNPs), have become increasingly important for population structure inferences relating to admixture, evolutionary studies, forensics, management and conservation of wildlife (Vignal *et al.,* 2002; Brumfield *et al.,* 2003; Morin *et al.,* 2004). Their functionality regarding a wide range of population studies in non-model organisms has highlighted the benefits of using SNPs relative to other nuclear markers such as microsatellites (Seddon *et al.,* 2005; Elfstrom *et al.,* 2006; Pariset *et al.,* 2006; Morin and Hancock, 2007). The advantages of SNPs over microsatellites include their higher abundance, easy and cost-effective discovery, simpler mode of mutation minimising homoplasy and finally a higher possibility of association with coding regions of the genome. The latter is especially true if SNPs are isolated from expressed sequence tag (EST) clones, especially those which relate to annotated genes. Morin *et al.* (2004) also summarised how a collection of SNPs, despite reduced individual statistical power, can be just as effective as microsatellites in the genetic analysis of population structure and history. In a direct comparison between SNPs and microsatellites, Liu *et al.* (2005) found that SNPs represented the majority of the most informative markers included in their study and that at the right quantity these markers can provide an even better inference of population structure than microsatellites. However, SNPs are not without their limitations and especially in population studies

‘ascertainment bias’ introduced during the discovery phase can have substantial effects on standard populations estimators (Kuhner *et al.,* 2000; Wakeley *et al.,* 2001). Ascertainment bias depends on the number and source of individuals used to identify these variable sites in the first place and if the SNP frequency within such a panel is unrepresentative of the true frequency within the species, the estimation of any parameter based on allele frequencies will be inaccurate (Wakeley *et al.,* 2001; Brumfield *et al.,* 2003; Morin *et al.,* 2004). Besides several methods being available for correcting ascertainment bias in SNP data (Kuhner *et al.,* 2000; Nielsen, 2000; Nielsen and Signorovitch, 2003), a panel of individuals that are representative of the species’ entire geographical distribution range may to some extent reduce bias (Morin *et al.,* 2004).

In this chapter twelve SNP loci previously isolated from EST clones were applied to infer patterns of gene flow within the same nine *H. midae* sampling populations as described in the previous chapter. Since the number of SNP loci used was far less than what is normally required to identify distinct geographically separated populations, results were mainly validated in comparison to the level of structuring obtained with the eight microsatellite markers applied in Chapter III. The same hypotheses testing of panmixia versus restriction of gene flow presented in the previous chapter was tested with SNPs markers applying the same analytical procedures. The primary objective was therefore to corroborate the subtle genetic differentiation implied by the microsatellite data.

4.2 Materials and Methods

4.2.1 Sampling design and SNP genotyping

Thirty-two *H. midae* individuals were randomly chosen from each of the nine sampling populations described in chapter III (refer to Figure 3.1). In total, DNA of 288 individuals was sequenced and genotyped at 12 SNP loci derived from *H. midae* EST clones (Table 4.1). PCR amplification reactions contained 20 ng genomic template DNA, 200 µM of each dNTP, 10 pmol primers, 1.5 mM $MgCl_2$ and two units of GoTaq® Flexi DNA polymerase. Amplification was performed in a GeneAmp System 2700 thermal cycler (Applied Biosystems) as follows: an initial denaturation of 10 min at 95℃ followed by 30 cycles of: 94℃ for 1 min, 55℃ for 1 min, 72℃ for 1 min and a final elongation step of 10 min at 72℃. Assessment of PCR products were conducted on a 2% agarose gel followed by purification through SigmaSpin post-reaction purification columns. Automated sequencing was performed in both directions using the ABI PRISM BigDye Terminator version 3.1 Cycle Sequencing kit and the 3100 Genetic DNA Analyser (Applied Biosystems). Alignment of sequences was carried out in the sequence alignment editor program BioEdit version 7.0.9.0 and manually genotyped at each of the 12 SNP loci. The genotypes of each individual were also confirmed through manual inspection of the sequence chromatograms.

Table 4.1 Origin and PCR amplification conditions of the 12 SNP loci of *Haliotis midae* analysed in this study

EST clone	GenBank Accession	Primer Sequences (5'-3')	SNP	T_a(℃)
1A1 *Perlucin*	EU135915	TTTCATGTTTTGCATCAAAC	A>C	57℃
		AAGAAGGAACTGTATGGCTG	G>A	
			G>T	
C12 *Cellulase*	EU135914	ATTTTTGTCGGTCACCTGGA	T>C	55℃
		GTAGGGCTTCCCAGAAGGAC	T>A	
3B4-1 *Ribosomal protein L8e*	EU135916	AAACATCTGCAACATTTAGG	C>T	57℃
		GACAGCAAAACAAACATCAG	A>T	
			C>T	
			A>G	
3B4-2 *Ribosomal protein L8e*	EU135916	TAAGAATCCACAAGTTGGTG	T>A	57℃
		ATGTATCATCACGGACAGG	T>C	
			T>A	

T_a(℃)= annealing temperature

4.2.2 Statistical analysis

Genetic diversity. The exact test in GENEPOP version 3.2 (Raymond and Rousset, 1997) was used to test for departure from Hardy-Weinberg equilibrium at each locus and the significance (*P*-values) determined by the Markov chain randomization method of Guo and Thompson (1992). Allelic frequencies, F_{is} (Weir and Cockerham 1984) and observed and expected heterozygosities were calculated with GENETIX version 4.02 (Belkhir *et al.,* 1999). To identify possible association among pairs of SNPs, genotypic disequilibrium was assessed according to Fisher's method available in GENEPOP based on a 1000 permutations.

Population differentiation. The following analytical methods were used to test for genetic differentiation between populations: (1) F_{ST} (2) Factorial Component Analysis (FCA) (3) Non-spatial Bayesian clustering and (4) Spatial Bayesian clustering.

Pairwise F_{ST} values between populations were calculated according to Weir and Cockerham, (1984) in GENETIX (θ) and the significance tested using a permutation approach (1000 replicates). Significance levels were adjusted for multiple tests using the Bonferroni correction (Rice, 1989). To view the populations in a multi-dimensional format, allele frequency data were subjected to factorial component analysis (FCA) available in GENETIX.

Further inference of population structure from the SNP data was conducted with the model-based clustering software STRUCTURE version 2.1 (Pritchard *et al.,* 2000) using no prior population information. The same priors and parameters used in the analysis of microsatellite data were applied to the analysis of the SNP data. A burn-in period of 50 000 followed by 10^6 Markov chain Monte Carlo (MCMC) iterations were conducted assuming a model of admixture with correlated allele frequencies among populations. The number of genetic clusters (*K*) tested ranged from 1 to 13 and was repeated five times. The method of Evanno *et al.* (2005) was followed to estimate the delta K (Δ*K*) statistic in an attempt to detect the uppermost level of structuring and, therefore, true number of populations present. To allow for easy comparison with the microsatellite data, cluster membership per population was summarized using the 20% highest likelihood runs for *K*=2 processed in CLUMPP (Jakobsson and Rosenberg, 2007). Admixture assessment was performed first by testing *a priori* forced structure and then without forcing any population structure. By means of the USEPOPINFO option, the assignment accuracy was tested using selected west and east coast individuals with a membership coefficient of q≥ 0.75.

The analysis of population structure was extended to the HMRF method of François *et al*. (2006) with tessellations. The assignment of clusters west and east of Cape Agulhas was investigated using the spatial Bayesian program TESS version

1.1 (Chen *et al.,* 2007). The same priors and parameters applied to microsatellite data were utilized in the analysis of SNPs. Independent Tess runs were performed for 120 000 sweeps and a burn-in period of 50 000 with the maximum number of clusters (K_{max}) ranging between 2 and 5 while the interaction parameter (Ψ) was fixed at 0.5. The program was run 30 times for each K_{max} and the 20% highest likelihood runs were utilized for output post-processing in the cluster matching and permutation program CLUMPP version 1.1 (Jakobsson and Rosenberg, 2007). The cluster visualization program DISTRUCT (Rosenberg, 2004) was used to display the membership coefficients of the clusters within each population.

Finally, the program BARRIER version 2.2 (Manni *et al.,* 2004) was used to define areas of genetic discontinuity and possibly to test for the hypothesis of a hydrogeographic barrier at Cape Agulhas indicated with the microsatellite loci. A single matrix based on the genetic distance (Nei, 1987) was applied to define the original barriers after which a second matrix based on pairwise theta values was included to test the robustness of the barriers.

4.3 Results

4.3.1 Sampling and SNP genotyping

Table 4.2 Number and origin of samples

Sampling Location	*N*	*Coastal Origin*
Saldanha (SD)	32	West coast
Robben Island (RI)	32	West coast
Kleinmond (KL)	32	West coast
Gansbaai (GB)	32	West coast
Witsand (WS)	32	South coast
Mossel Bay (MB)	32	South coast
Cape Recife (CR)	32	South coast
Riet Point (RP)	32	East coast
Black Rock (BR)	32	East coast

Sequences of all 288 individuals (Table 4.2) were aligned to score the sequence variation at the 12 positions previously identified as true SNPs. The presence of a single base indicated a homozygote state, while heterozygous individuals were detected as multiple bases or peaks when referring to the chromatograms at the same sequence position. As a result 32 individuals from each of the nine *H. midae* populations were successfully genotyped at the 12 SNP loci.

4.3.2 Genetic diversity

Observed and expected heterozygosity, averaged over all populations, ranged from 0.121 to 0.870 and 0.232 to 0.489 respectively while the minor allele frequency ranged from 0.143 to 0.489 (Table 4.3). More detailed estimates of genetic diversity including observed and expected heterozygosities, allele frequencies, F_{is} values and probability of Hardy-Weinberg equilibrium are summarised in Appendix 4. The exact test performed in GENEPOP showed that three of the SNP loci (SNP-2; SNP-4 and SNP-5) deviated significantly from Hardy-Weinberg proportions (P<0.05) while none of the populations were in equilibrium over all loci. Significant linkage disequilibrium was observed for polymorphisms within the same EST sequence while linkage

disequilibrium was not detected among SNPs derived from different EST sequences. This restricted the total amount of independent polymorphic information obtained to the number of EST sequence fragments screened in this study.

Table 4.3 Characteristics of the 12 SNP loci from *Haliotis midae* used in this study

EST clone	GenBank Accession	SNP ID	SNP	Minor allele freq	H_O	H_E	F_{ST}
1A1 *Perlucin*	EU135915	*SNP-1*	A>C	C: 0.19	0.33	0.31	0.022
		SNP-2	G>A	A: 0.45	0.87	0.49	0.023
		SNP-3	G>T	T: 0.21	0.33	0.32	0.021
C12 *Cellulase*	EU135914	*SNP-4*	T>C	C: 0.14	0.06	0.23	0.019
		SNP-5	T>A	A: 0.23	0.12	0.31	0.016
3B4-1 *Ribosomal protein L8e*	EU135916	*SNP-6*	C>T	T: 0.48	0.51	0.48	0.022
		SNP-7	A>T	T: 0.48	0.44	0.48	0.021
		SNP-8	C>T	C: 0.50	0.45	0.49	0.023
		SNP-9	A>G	G: 0.39	0.48	0.49	0.023
3B4-2 *Ribosomal protein L8e*	EU135916	*SNP-10*	T>A	A: 0.23	0.40	0.35	0.023
		SNP-11	T>C	C: 0.32	0.45	0.43	0.023
		SNP-12	T>A	A: 0.19	0.32	0.31	0.023

H_O =observed heterozygosity; H_E = expected heterozygosity; F_{ST}= Wright's F-statistics

4.3.3 Population differentiation

Locus-specific F_{ST} estimates ranged from 0.016 to 0.023 (Table 4.3) and overall populations the test was not significant (F_{ST} =0.021). The pairwise multilocus F_{ST} values ranged between -0.0080 and 0.0647 with the most interesting observation being the Witsand population showing comparatively high significant values with all the West coast populations (Table 4.4). Only one of the pairwise values (WS vs GB) was significant after Bonferonni adjustment (P<0.05).

Table 4.4 Pairwise F_{ST} (θ) (above diagonal) values between populations of *Haliotis midae*. *Significant values after Bonferroni adjustment are highlighted.

	RI	KL	GB	WS	MB	CR	RP	BR
SD	-0.0029	-0.0012	0.0067	0.0234	0.0039	-0.0018	0.0208	0.0102
RI		-0.0030	0.0271	0.0558	0.0213	0.0095	0.0647	0.0134
KL			0.0178	0.0502	0.0230	0.0105	0.0488	0.0087
GB				0.0552*	0.0407	0.0188	0.0141	0.0367
WS					-0.0080	0.0137	0.0087	0.0458
MB						-0.0025	0.0180	0.0223
CR							0.0158	0.0132
RP								0.0531

FCA analysis allowed for a visual interpretation of the genetic relationships among populations and the first factor accounted for over 50% of the genetic differentiation observed. Although not as pronounced as with the microsatellite data (see chapter III, Figure 3.2), the three- dimensional FCA plot showed partial separation of the west coast individuals from the rest. When the three loci deviating from HWE were excluded, a clear division between west and east coast individuals was observed (Figure 4.1).

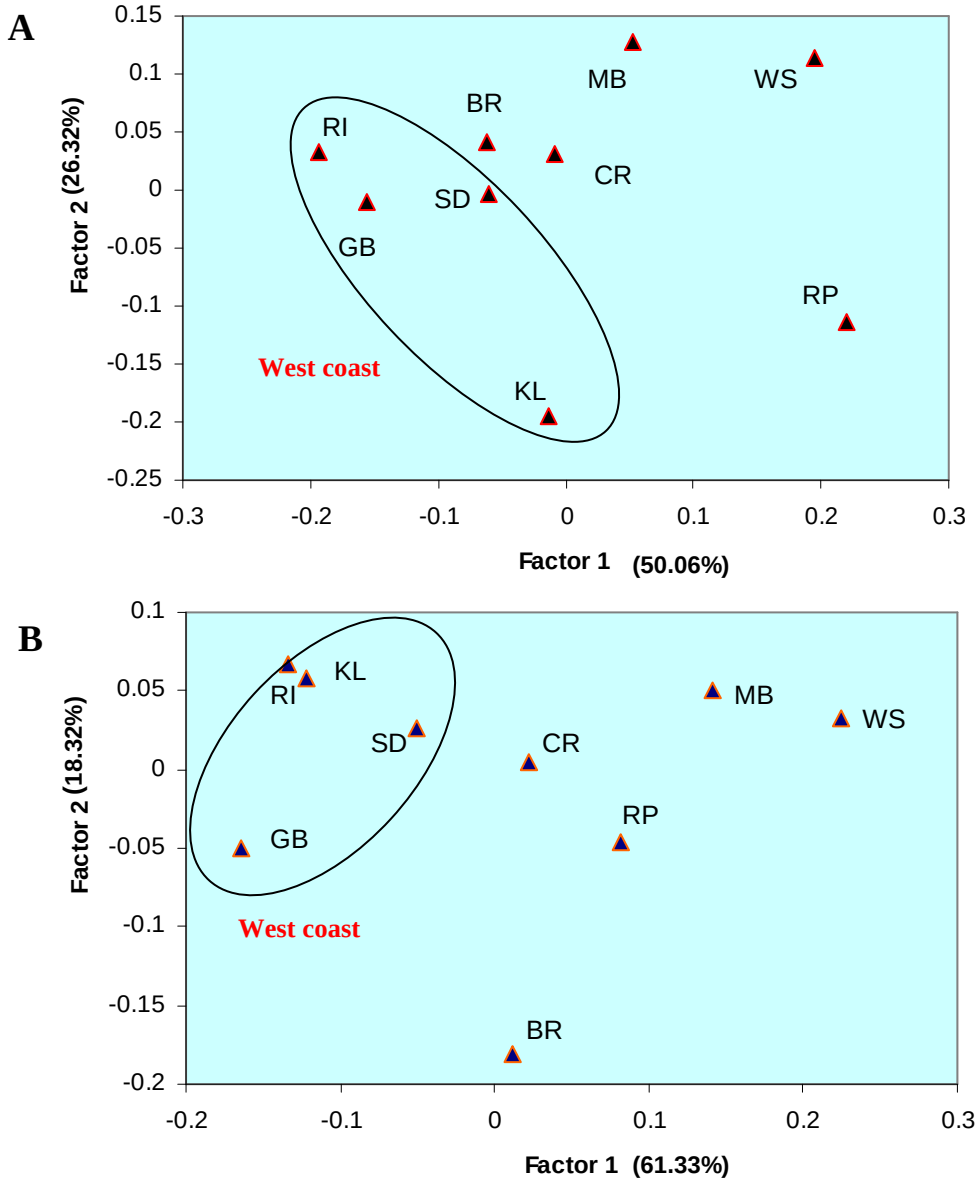

Figure 4.1 Factorial component analysis showing scatter plots of nine populations of *Haliotis midae* along factors 1 and 2 based on A) all 12 SNP loci and B) the nine SNP loci in HWE. The West coast populations are encircled.

The cluster analysis performed in STRUCTURE gave consistent likelihood values between the ten runs performed for each number of populations (*K*) tested. When the likelihood values and the variance of the bootstrap samples were plotted against *K,* likelihood values reached a maximum at *K*=7; the variance across runs showed a sharp increase for K>7 (Figure 4.2). However, when the rate of variation in likelihood values between successive *K's* (Δ*K* statistic of Evanno *et al.,* 2005) was analysed,

the number of *K* was reduced to two. Similar to the microsatellite data, the plot represented a subdivision comparable to a contact zone situation where *K*=2 corresponds to the highest level of structuring in this migration model. The same scenario was reflected in a simulation study of Evanno *et al.* (2005) where the effect on Δ*K* was investigated for three different migration models: the island, the hierarchical island and the contact zone model.

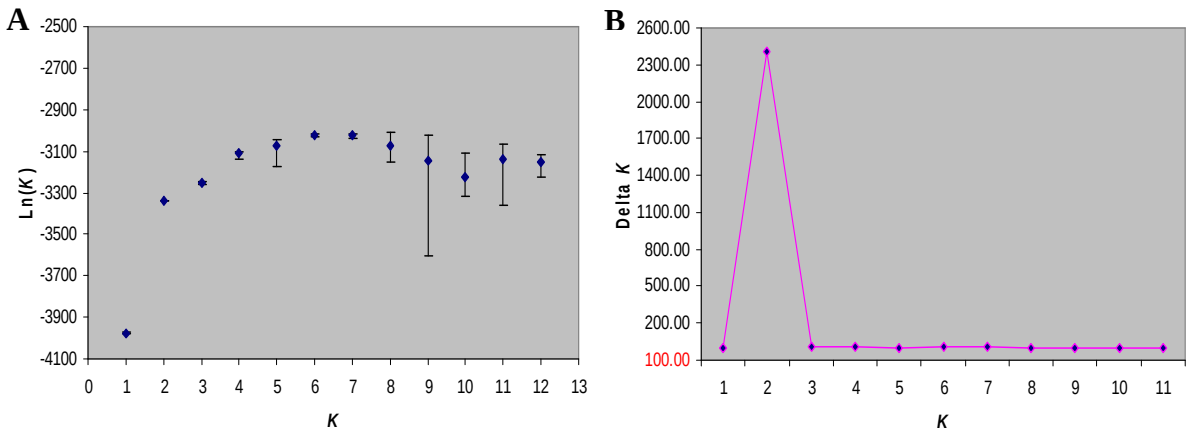

Figure 4.2 (A) The posterior probability of the data, L(*K*) for each *K* and (B) Δ*K* as a function of *K* following Evanno *et al.* (2005)

Assuming *K*=2, cluster assignment for each sampling site showed the most abrupt difference between the samples (populations 4 and 5) on each side of Cape Agulhas whereas the rest of the populations showed high levels of admixture for the two projected clusters (Figure 4.3). Estimated admixture proportions with the forced structure were significantly higher compared to that observed with the microsatellite analyses while slightly lower assignment accuracy (58%) was obtained when the prior population information was used (Table 4.5).

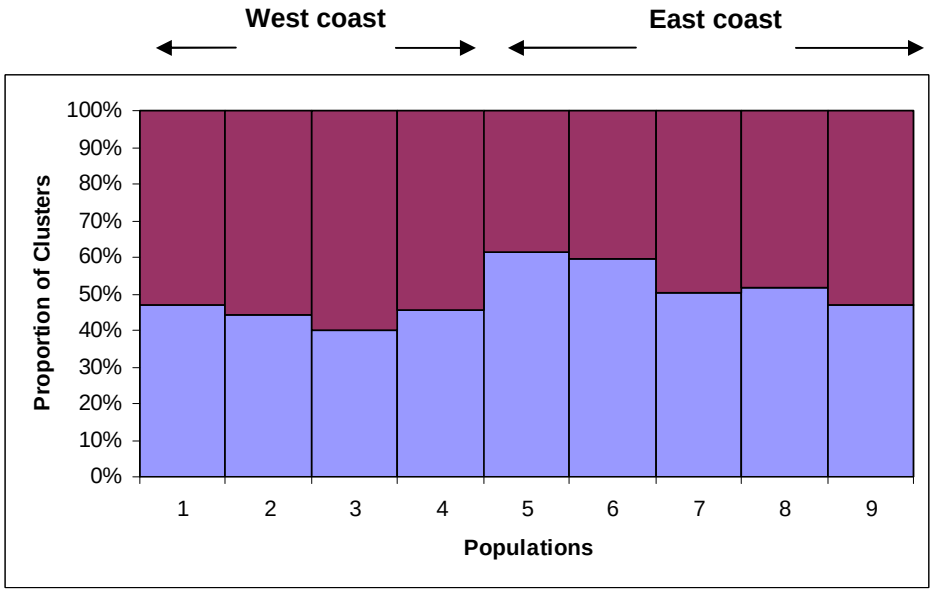

Figure 4.3 Proportion of the model-based clusters (*K*=2) in the ancestry of nine *Haliotis midae* populations

Table 4.5 Estimation of the percentage of individuals clustered in the proposed West (1) and East (2) coast populations using the forced, non-forced and popinfo options in STRUCTURE (Pritchard *et al.*, 2000)

	Forced structure		Non-forced structure		Prior Popinfo	
	1	*2*	*1*	*2*	*1*	*2*
West coast (1)	0.617	0.383	0.558	0.442	0.587	0.413
East coast (2)	0.451	0.549	0.460	0.540	0.422	0.578

The model-based clustering method of TESS allowed for the utilization of spatial data as well as the assumption on the level of interaction between populations. As the interaction parameter was fixed at 0.5, only the effect of maximum *K* could be compared between different runs. For every maximum *K*, combining the 20% highest probability runs in the permutation method of CLUMPP resulted in an average closer to the true Bayesian estimate than the highest likelihood runs. Cluster assignment within sampling sites showed a more abrupt break between western and eastern *H. midae* populations and, contrary to the microsatellite data, the Witsand population failed to represent a transitional area (Figure 4.4). Although there was no evidence

for a consistent gradient in allele frequencies and therefore less support for clinal variation, differential assignment of primary clusters still confirmed the existence of a genetic barrier around the area of Cape Agulhas.

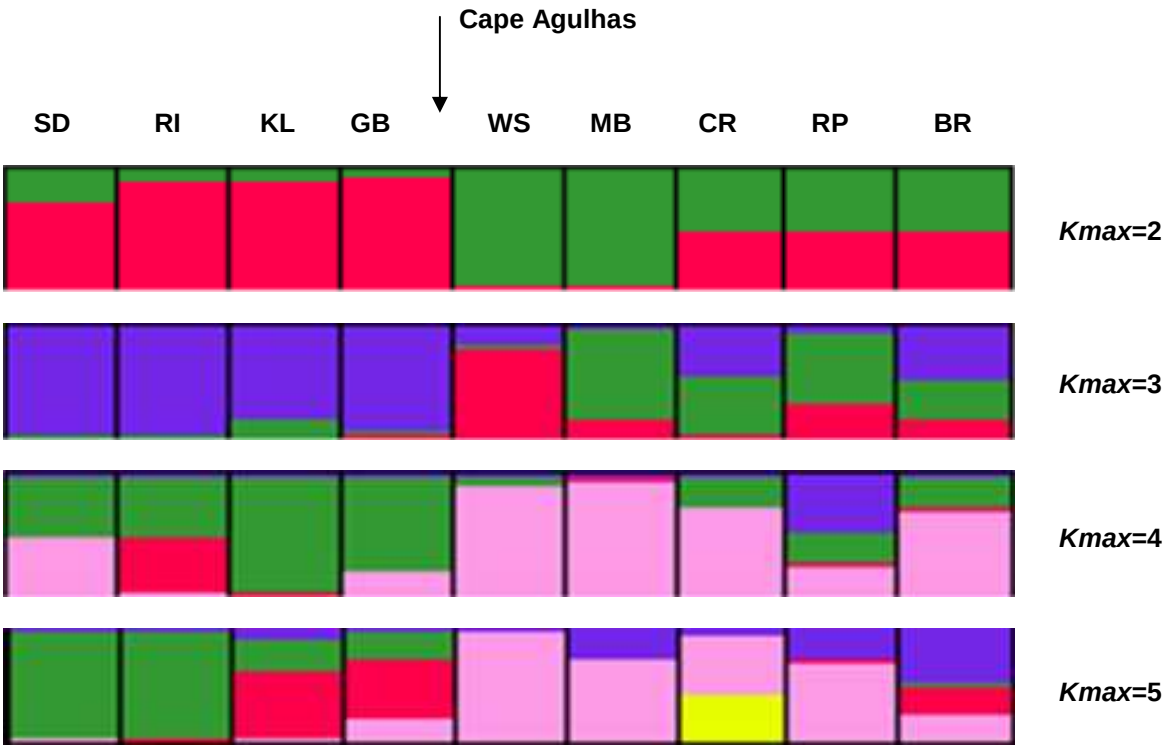

Figure 4.4 Estimated cluster configuration based on CLUMPP permutation of the 20% highest likelihood runs performed in TESS (Ψ=0.5 and $2 \leq K_{max} \leq 5$)

The areas with restricted genetic interchange between populations identified by BARRIER were rated according to level of importance and the number of matrices supporting each of them. Based on only the genetic distance of Nei (1987), a barrier between Gansbaai and Witsand was recognised as the most important, followed by one between Riet Point and Black Rock and two more along the West and East coast (Figure 4.5). When the analysis was performed including an additional matrix based on pairwise θ values, three of the four barriers (A, B and D) were supported by both matrices while a new barrier was identified between Kleinmond and Gansbaai. The only barrier that therefore acquired the same order of importance as the ones

identified with the microsatellite data is the Gansbaai-Witsand barrier and again confirmed the highest level of restricted gene flow around the Cape Agulhas area.

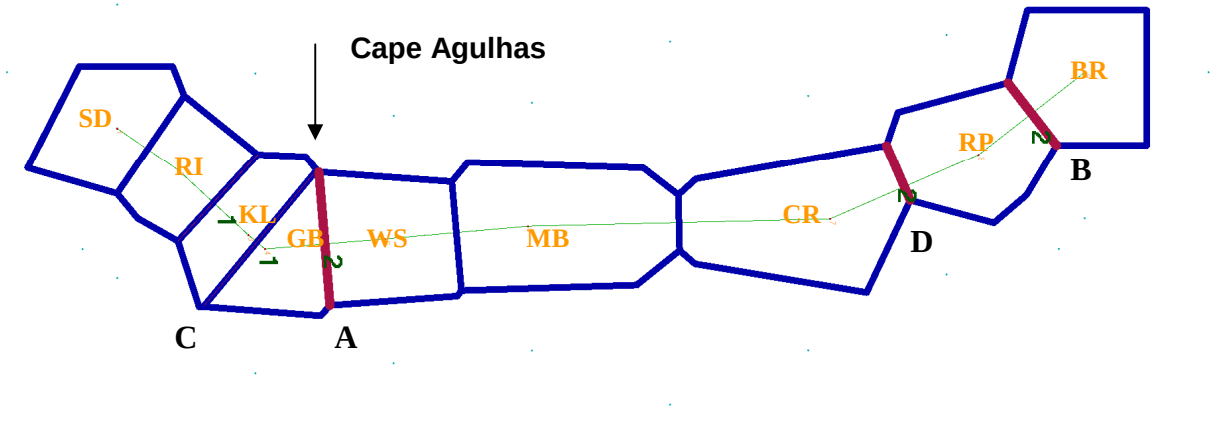

Figure 4.5 Location of the barriers to gene flow identified with BARRIER (indicated in red). The barrier order reflects the level of importance (A-D) while the barrier width indicates the number of matrices supporting the data (1 or 2).

4.4 Discussion

4.4.1 Genetic diversity

As expected, the genetic diversity in terms of number of alleles, observed and expected heterozygosity over all loci were low in comparison to what was reflected with the microsatellite data but similar to estimates of SNP diversity in other marine organisms (Elfstrom *et al.,* 2005, 2007; Cenadelli *et al.,* 2007; Morin *et al.,* 2007; Aguilar and Garza, 2008). The mean minor allele frequency and observed heterozygosity was consistent between populations and indicate that none of the nine *H. midae* populations exhibited significantly lower genetic diversity or signs of genetic isolation compared to the rest. In theory, the precise estimation of genetic diversity requires a much larger number of SNPs relative to microsatellite loci but if the SNPs are chosen based on heterozygosity and not at random, the same level of information content can be obtained with as little as one and a half times the number of

microsatellite markers (Rengmark *et al.,* 2006; Artamonova, 2007). As the selection criteria in this study were purely based on choosing the most informative loci within a selected number of EST clones, it was not expected that the statistical power of the microsatellites used in the previous chapter would necessarily be matched. Still the usefulness of SNP loci in population-based analysis are highly dependant on heterozygosity levels and frequency of the minor alleles (Anderson and Garza, 2006) and since respective averages of 0.395 and 0.326 were obtained in this study, the 12 SNPs showed promise in their ability to detect or otherwise confirm population structure within *H. midae*.

Although only a small number of samples were used for SNP discovery in this study, the wide range in heterozygosity levels as well as the minor allele frequencies across loci is arguing for an unbiased selection of SNPs. In addition, the test panel of individuals represented a wide geographical range, supposedly reducing ascertainment bias (Morin *et al.,* 2004). The significant linkage disequilibrium observed for many of the SNPs examined in this study diminished the overall statistical power and placed further emphasis on the importance of the initial SNP discovery protocol. To increase chances of detecting unlinked SNPs, Nielsen *et al.* (2005) suggested screening rapidly evolving rather than conserved gene classes. In this particular study, the total amount of polymorphic information gained could have been increased through genotyping a single SNP for a larger number of EST sequences.

4.4.2 Population differentiation

As for summary statistics the SNP data, although preliminary, has proven effective for analysis of genetic differentiation between *H. midae* populations. The pairwise F_{ST}

values were relatively comparable to that of the microsatellite data and despite only one significant pairwise value after Bonferonni adjustment, the highest values were obtained between the populations directly separated by the proposed Agulhas barrier. The first indication of sharp genetic discontinuity caused by the barrier is reflected in the considerably higher pairwise values obtained between the population of Witsand and all of the West coast populations. Such equivalent results obtained for the genetic distance measure of F_{ST} is not what normally would be expected between a set of bi-allelic markers versus a set of highly polymorphic microsatellite markers (Kalinowski, 2002). If indeed the SNP loci are giving the same level of precision in this study, the advantage of adding more SNPs rather than microsatellites would be to obtain a higher representation of the *H. midae* genome while minimising unreliable estimates of gene flow due to the homoplasy inherent to microsatellite loci. Conversely, ascertainment bias will then have to be carefully considered given the fact that summary statistics are highly sensitive to the bias introduced during SNP discovery especially in spatially structured populations (Rosenblum and Novembre, 2007).

At a first glance it appeared that the SNP data was less discriminative than the microsatellites in the Factorial Component Analysis. No apparent separation of the East and West coastal regions could be observed and especially the plotting of individual genotypes showed considerably more overlap and dispersion compared to the microsatellites. This result could be accounted for by the much smaller number of SNP alleles contributing to the main factorial components, while in turn explaining why the first component accounts for over 50% of the overall variability. Seeb *et al.* (2005) also found that microsatellites performed slightly better with individual assignments of sockeye salmon (*Oncorhynchus nerka*) using Factorial Component

Analysis, although on a broader scale stock assessment results were concordant between different molecular data types. Another interesting observation was the appearance of the West-East coast structure only when the three SNPs that deviated from Hardy-Weinberg equilibrium were excluded from analysis. If a higher number of SNPs were available an even better inference of structure could have been attained by including only the FCA-correlated loci. FCA-correlated loci can be selected based on a simple Monte-Carlo algorithm developed by Paschou *et al.* (2007) and identifies the SNP markers that represent the major components of genetic variation. Although no further structure or correlation of genetic distance with geographical distance was revealed with the FCA analysis, grouping of populations along the major axis can serve as a realistic default for the number of clusters to be tested in STRUCTURE or any other cluster-based program (Patterson *et al.,* 2006).

The model-based clustering results of STRUCTURE suggested that the SNP loci were just as capable of uncovering the number of *K* present, or in this case the broad-scale west-east structure, as with the microsatellites. The much higher magnitude of the Δ*K* statistic at *K*=2 could possibly be explained by the increased sensitivity of a small number of alleles to variation in likelihood values (and the rate in change thereof) in comparison to a large number of microsatellite alleles. The small number of alleles per locus is also a likely reason for the poor assignment ability obtained with the SNP loci in STRUCTURE, simply because of the limited amount of polymorphic information it represents. In a study comparing markers in the context of population structure inference, Liu *et al.* (2005) demonstrated that only when a considerably large number of SNPs (≥100) are available, they can match or possibly succeed the assignment accuracy of microsatellites. While in this study, the small

number of SNPs unexpectedly confirmed the microsatellite-inferred structure, it is generally expected that the precise clustering obtained is likely to depend on the number and type of loci used (Ben-Ari *et al.,* 2005). Ostrowski *et al.* (2006) for example showed how neither microsatellites nor SNPs were able to infer population structure in accessions of *Arabidopsis thaliana* when used alone but when the two sets of data were combined, a highly robust clustering pattern was observed with the individual membership proportions of STRUCTURE.

With the spatial Bayesian approach the general trend was that the degree of clustering or allelic variation across the populations was considerably smaller for the SNPs than for the microsatellites. This was particularly true for the contact zone area, where in contrast to the microsatellite data the abrupt genetic change was far more noticeable because of mostly a single cluster representing the population of Witsand. As the allele frequency distribution obtained with SNPs is more likely to be shaped by local selection especially if the polymorphism lies within coding regions under selection (Lao *et al*., 2006), the alternative clustering pattern obtained with SNPs is not entirely unexpected. In addition, clinal variation at single SNPs has been proposed to result from adaptive response to local conditions (Sezgin *et al*., 2004; Ingvarsson *et al.,* 2006) which in a marine environment can entail local adaptation to habitat, temperature, pH or altitude.

The differential power of the SNP markers to identify areas of genetic discontinuity is finally manifested with the program BARRIER which identified two further barriers not reflected with the microsatellite data. More important however is the West-East coast barrier that was once again indicated as the most significant one despite the limited number of SNPs and matrices tested. In the future, analysis of SNPs in gene regions known to be under selection could possibly increase the chance of identifying

barriers between geographical areas revealing local adaptation to particular traits (Larsen *et al*., 2007).

The similarity of results in population structure analysis obtained from different markers in this study advocates for a common factor of genome-wide influence that, as discussed in the final chapter, could either be a contemporary disruption of gene flow (ocean currents; thermal fronts; habitat) or a historical factor (e.g. glaciations and sea level fluctuations) that left its signature in both coding and non-coding regions. Although the presence of artefacts like null alleles and allelic dropout cannot be disregarded in microsatellite loci, the observation of an overall deviation from equilibrium in both SNPs and microsatellites furthermore favours the interpretation of an underlying biological factor (e.g. related to the life cycle, sweepstake recruitment, etc) of genome-wide effect.

4.5 Conclusions

In summary, this study has shown that 12 SNPs detected similar level of geographic population structure to that of the eight microsatellite loci described in the previous chapter. Although differences were observed between the strength of the two types of markers using various analysis methods, the SNPs were equally able to detect the population differentiation between West and East coast *H. midae* populations related to the biogeographic barrier at Cape Agulhas. Varying levels of significance obtained with SNP and microsatellite data can be attributed to 1) SNPs being bi-allelic and microsatellites multi-allelic thereby exhibiting different degrees of polymorphism; 2) differences in mutation rate which could affect genetic diversity at equilibrium; and 3) SNPs often being associated with genes and thus more likely to

be under selection compared to microsatellites. The number of SNPs used proved to be sufficient in verifying population structure of *H. midae* but in general the number of SNP loci required depends mainly on the degree of polymorphism of the selected loci and the inherent genetic variation between the sampling populations. Including this study, there is enough evidence that when applying carefully ascertained and validated SNPs, the number of markers necessary for population structure analysis can be reduced substantially. The research reported in this chapter strongly supports the use of SNPs as an alternative or additional molecular marker to microsatellites for the inference of population genetic structure in *H. midae*.

References

Aguilar A and Garza JC (2008) Isolation of 15 single nucleotide polymorphisms from coastal steelhead (*Oncorhynchus mykiss*). *Molecular Ecology Resources* **8:** 659–662.

Anderson EC and Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* **172:** 2567–2582.

Artamonova VS (2007) Genetic markers in population studies of Atlantic Salmon *Salmo salar* L.: Analysis of DNA Sequences. *Russian Journal of Genetics* **43:** 341–353.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (2000) *GENETIX, logiciel sous Window TM pour la génétique des populations* Laboratoire Génome et Populations, Montpellier, France.

Ben-Ari G, Zenvirth D, Sherman A, Simchen G, Lavi U, Hillel J (2005) Application of SNPs for assessing biodiversity and phylogeny among yeast strains. *Heredity* **95:** 493–501.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18:** 249–255.

Cenadelli S, Maran V, Bongioni G, Fusetti L, Parma P, Aleandri R (2007) Identification of nuclear SNPs in gilthead seabream. *Journal of Fish Biology* **70:** 399–405.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes* **7:** 747–756.

Elfstrom CM, Gaffney PM, Smith CT, Seeb JE (2005) Characterization of 12 single nucleotide polymorphisms in weathervane scallop. *Molecular Ecology Notes* **5:** 406–409.

Elfstrom CM, Smith CT, Seeb JE (2006) Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. *Molecular Ecology Notes* **6:** 1255–1259.

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

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48:** 361–372.

Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*) *Genetics* **172:** 1845–1853.

Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity* **88:** 62–65.

Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* **156:** 439–447.

Lao O, van Duijn K, Kersbergen P, de Knijff P, Kayser M (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *American Journal of Human Genetics* **78:** 680–690.

Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK, Kruhøffer M, Grønkjaer P, George SG, Dyrskjøt L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology* **16:** 4674–4683.

Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of Single Nucleotide Polymorphisms and microsatellites in inference of population structure. *BMC Genetics* **6:** S26

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by “Monmonier’s algorithm”. *Human Biology* **76:** 173–190.

Morin PA and Hancock BL (2007) Development and application of single nucleotide polymorphisms (SNPs) for bowhead whale population structure analysis. Paper SC/59/BRG8 presented to the Scientific Committee of the International Whaling Commission, May 2007.

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S (2007) Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). *Molecular Ecology Notes* **7:** 626–630.

Morin PA, Luikart G, Wayne RK and SNP-Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19:** 208–216.

Nei M (1987) Molecular Evolutionary Genetics. New York, Columbia University Press, 356pp.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154:** 931–942.

Nielsen R and Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* **63:** 245–255.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Adi Fledel-Alon, Tanenbaum DM, Ciuvello D, White TJ, Sninsky JJ, Adams MD, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3:** e170

Ostrowski MF, David J, Santoni S, Mckann H, Reboud X, Le Corre V, Camilleri C, Brunel D, Bouchez D, Faure B, Bataillon T (2006) Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Molecular Ecology* **15:** 1507–1517.

Pariset L, Cappuccio I, Ajmone Marsan P, Dunner S, Luikart G, Obexer-Ruff G, Peter C, Marletta D, Pilla F, Valentini A (2006) Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds. *Journal of Chromatagraphy B* **833:** 117–120.

Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, Hahoney MW, Drineas P (2007) PCA-Correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics* **3:** e160

Patterson N, Price A, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics* **2:** e190

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Raymond M and Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86:** 248–249.

Rengmark AH, Sletten A, Skaala O, Lie O, Lingaas F (2006) Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture* **253:** 229–237.

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

Rosenberg N, Mahajan S, Ramachandran S, Zhao C, Pritchard J, Feldman M (2005) Clines, clusters, and the effect of study design on the influence of human population structure. *PLoS Genetics* **1**: e70

Rosenberg NA (2004) DISTRUCT*:* a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137–138.

Rosenblum EB and Novembre J (2007) Ascertainment bias in spatially structured populations: A case study in the Eastern Fence Lizard. *Journal of Heredity* **98:** 331–336.

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* **14:** 503–511.

Seeb J, Antonovich A, Templin W, Seeb L, Habicht C, Johnston G, Pronzati K, Lardizabal E, Elfstrom C, Smith C (2005) Single Nucleotide Polymorphisms (SNPs) provide high throughput and high resolution DNA data for sockeye salmon. Annual American Fisheries Society meeting, September 2005.

Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu CT, Verrelli BC, Eanes WF (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics* **168:** 923–931.

Vignal A, Milan D, San Cristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34:** 275–305.

Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide polymorphisms-and inferences about human demographic history. *American Journal of Human Genetics* **69:** 1332–1347.

Weir BS and Cockerham CC (1984). Estimation *F*-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Chapter V

Phylogenetic relationships of *Haliotis midae* with other abalone species based on mtDNA and nuclear DNA sequences

5.1 Introduction

Of the 56 extant abalone species found worldwide none are distributed globally whereas four areas of endemism (Australia, South Africa, New Zealand and North America) have been recognised and confirmed by molecular phylogenetics (Lee and Vacquier, 1995; Geiger, 2000; Estes *et al.,* 2005). The first studies on the phylogeny of Haliotidae were based on allozyme frequencies and sperm lysin DNA sequence data and resulted in a broad separation of the north Pacific species (including west American and large Japanese species) from the remaining taxa (Brown, 1993; Lee and Vacquier, 1995). These studies were however criticized for low taxon sampling and the exclusion of the type species of the genus, *Haliotis asinina*. By incorporating additional taxa and morphological characters, Geiger (1999) identified three discrete groups namely the North Pacific, the Australian and the Indo-Pacific based on the geographic correspondence of the composing species, while New Zealand was the only geographic location that showed species (*H. australis and H. iris)* belonging to two different of the above mentioned phylogenetic groups. Further studies elaborated on these findings using a variety of molecular characters such as internal transcribed spacer sequences (ITS), lysin, cytochrome *c* oxidase subunit 1, mitochondrial 16S rRNA, haemocyanin and cytochrome oxidase II (Coleman and Vacquier, 2002; Estes *et al.,* 2005; Streit *et al.,* 2006; Degnan *et al.,* 2006) and found

two major clades, one representing the southern hemisphere abalone and one comprising of northern hemisphere species. For instance, Estes *et al.* (2005) investigated the evolution of body size in abalone using a four-gene data set (16S rRNA, cytochrome *c* oxidase subunit 1, ITS and lysin) and confirmed 1) the subdivision of Haliotidae into two major lineages, the European-Australasian clade and the North-Pacific clade and 2) the lack of geographic- phylogenetic correspondence of the New Zealand species. Furthermore, the phylogenetic relationships among taxa of both clades indicated a broad Indo-Pacific origin of abalone which is in full agreement with the evolutionary relationship and systematics previously proposed for the Haliotidae (Lindberg, 1992).

Haliotis midae is the only species among five South African taxa for which the phylogenetic positioning within the genus *Haliotis* has been examined previously (Lee *et al.,* 1995; Geiger, 1999; Streit *et al.,* 2006; Degnan *et al.,* 2006). Two of the species *H. spadicea* and *H. parva* have a similar geographic distribution as *H. midae* while *H. speciosa* and *H. queketti* have a limited distribution range and are considered to be rare. *H. spadicea* and *H. parva* are interesting in that despite an overlap in biogeographic distributions with *H. midae* (Figure 5.1), they both exhibit unique morphological features also found in other southern hemisphere abalone (Geiger, 1999, 2000). *Haliotis parva* is also much smaller in size and appears at a slightly different depth range (subtidal) than the two intertidal species, *H. midae* and *H. spadicea*.

Previous investigations grouped *H. midae* in the Australasian or southern hemisphere clade whilst its closest relative varied depending on the number of taxa and phylogenetic methods used (Lee *et al.,* 1995; Geiger, 1999; Streit *et al.,* 2006; Degnan *et al.,* 2006). In a study on the evolutionary relationship among abalone

species from the southern hemisphere, *H. midae* is grouped closest to the Australian endemics e.g. *H. rubra, H. laevigata, H. roei and H. scalaris* based on mitochondrial cytochrome *c* oxidase subunit 2 DNA sequence data (Degnan *et al.,* 2006). None of the previous studies addressed the phylogeny of the South African species as a group.

The aim of this chapter was therefore to investigate the evolutionary status of *H. midae* within the southern hemisphere group. At the same time the first phylogenetic classification of two additional South African species, *H. spadicea* and *H. parva* using mitochondrial and nuclear sequence data was made.

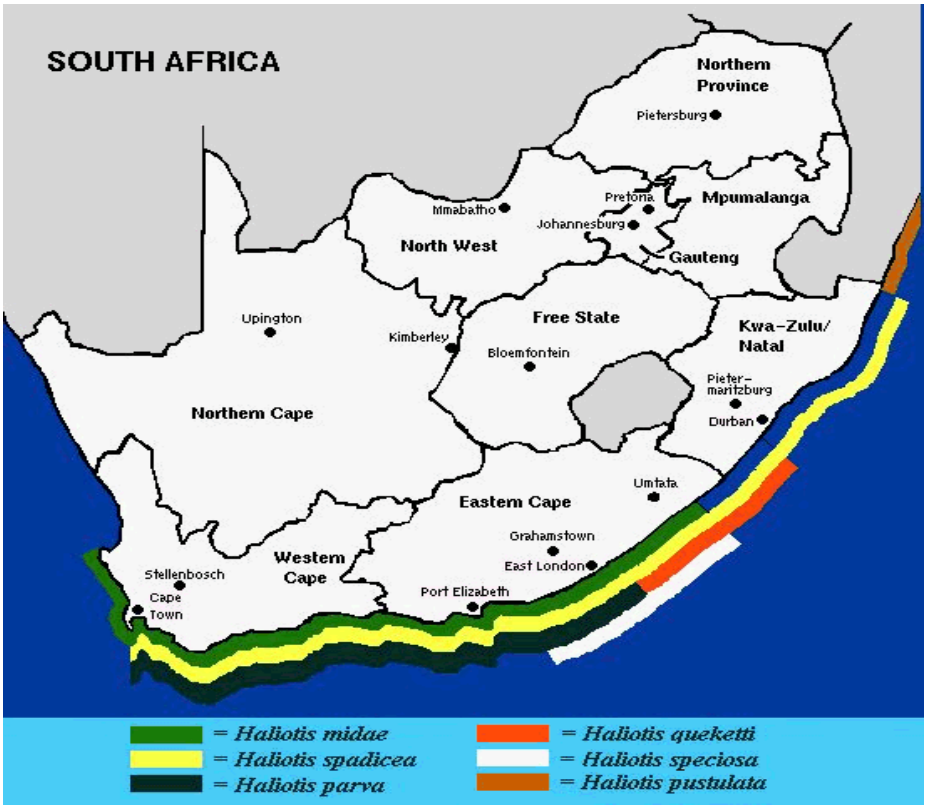

Figure 5.1 Map of South Africa indicating the distribution of the six different indigenous abalone species (ABNET[b])

[b] http://web.uct.ac.za/depts/zoology/abnet/

5.2 Materials and Methods

5.2.1 Taxon sampling

Twenty-five wild-caught abalone representing seven southern hemisphere and one northern hemisphere species were collected for this study (Table 5.1). Samples were received either as fresh tissues (mainly the South African species) or as a small piece of foot or mantle tissue preserved in absolute ethanol. Total genomic DNA was extracted using a Cell and Tissue Extraction kit (Puregene, Gentra). Approximately 0.01g of tissue was added to the recommended volume of cell lysis solution, followed by overnight incubation at 55ºC and the manufacturer's instructions.

Table 5.1 Origin and number of samples

Species	Origin	N
H. midae[a]	South Africa	2
H. spadicea[a]	South Africa	3
H. parva[a]	South Africa	4
H. rubra[b]	Australia	2
H. asinina[c]	Thailand	4
H. iris[d]	New Zealand	2
H. australis[d]	New Zealand	2
H. fulgens[e]	California	2

[a] Samples collected by A. Bester
[b] Samples collected by S. Appleyard
[c] Samples collected by P. Jarayabhand
[d] Samples collected by E. Watts
[e] Samples collected by K. Gruenthal

5.2.2 PCR amplification

A fragment of approximately 1100bp of the mitochondrial NADH-dehydrogenase subunit 1 (ND1) gene was amplified using primers developed from the *Haliotis* ND1 sequences available in the NCBI GenBank (*HmND1F*: TTTTACTTAAACTACCCCTCG; *HmND1R*: CTTTCGTTGGTTGGTCTGTTC). All samples were amplified in a total reaction volume of 50µl containing 100ng genomic template DNA, 200µM of each dNTP, 0.25µM of each primer, 2.5mM $MgCl_2$ and 1.0 units *Taq* polymerase (Promega). Cycling conditions were as follows: an initial denaturation of 10 min at 95℃ followed by 10 cycles of 1 min at 94℃, 1 min at 55℃, 2 min at 72℃, followed by 30 cycles of 1 min at 94℃, 1 min at 50℃, 72℃ for 2 min and ended by a final elongation step of 10 min at 72℃.

As the haemocyanin-specific primers of Streit *et al.* (2006) failed to amplify in *H. spadicea* and *H. parva*, novel primer sequences (*HaemF1*: ACAATAMTTTATGAACCAGG; *HaemF2*: TMAGGTTGTCAGTCTCAGC) were developed from related *Haliotis* haemocyanin sequences available in GenBank. Amplification products with variable product size (450-1100bp) were obtained using the same PCR conditions as described in Streit *et al.* (2006). PCR reactions contained 50 ng genomic DNA, 1 µM of each primer, 200 µM of each dNTP, 3 mM $MgCl_2$, and 2.5 units *Taq* polymerase in a final volume of 50µl. Samples were denatured for 2 min at 94ºC, followed by 45 amplification cycles (10 s at 94ºC, 30 s at 55ºC, 2 min at 72ºC) and final extension for 10 min at 72ºC.

All amplification products were separated by electrophoresis in a 1.5% agarose gel and purified with SigmaSpin post-reaction clean-up columns. Sequencing was performed in both directions using ABI PRISM BigDye Terminator version 3.1 Cycle

Sequencing kit following the specifications of the manufacturers and the 3100 Genetic DNA Analyser (Applied Biosystems).

5.2.3 Sequence and data analysis

Sequence data from both genes were manually edited in BioEdit version 7.0.0 (Hall, 1999) and confirmed by BLASTn at the NCBI website. The two sets of sequences were aligned using ClustalX as implemented in BioEdit and truncated to the longest possible common sequences to obtain a consensus data set across all eight species. The ND1 and haemocyanin sequences of the gastropods *Conus textile* (Sorbeoconcha: Conidae) and *Megathura crenulata* (Archaeogastropoda: Fissurrelidae) respectively were imported from GenBank to be used as outgroups (Accession numbers DQ862058 and AJ698330). Final alignments of 579bp and 371bp were saved and converted to the appropriate input file formats for phylogenetic analysis. As phylogenetic trees reconstructed with the mitochondrial and nuclear datasets were largely congruent, these datasets were combined and tested via three different methods of phylogenetic reconstruction. A haemocyanin only dataset which contained GenBank sequences of an additional five taxa (Accession numbers AJ749648; AJ749646; AJ749641; AJ749644; AJ749643) as well as the outgroup, *Megathura crenulata*, was compiled to allow direct comparison with the phylogenetic tree obtained by Streit *et al.* (2006).

Maximum parsimony analysis was conducted in MEGA version 4.0 (Tamura *et al.,* 2007), applying the close-neighbor-interchange algorithm and a random addition of 10 sequences to obtain initial trees. Clade support was assessed through a 1000 replicate bootstrap test. For maximum likelihood (ML), a tree was constructed with PHYML version 2.4.4 (Guindon and Gascuel, 2003) based on the most appropriate

model determined by the web-based ModelTest version 3.7 (Posada and Crandall, 1998) available at http://hcv.lanl.gov/cgi-bin/findmodel/findmodel.cgi. The best model was identified as the transversion model for nucleotide substitutions described by Tamura and Nei (1993) and the dataset was run with 1000 bootstrap replicates. MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003) was used for Bayesian inference of phylogeny with parameters nst=6 and rates=invgamma. For the combined data set of ND1 and haemocyanin, four MCMC chains were run under the GTR model for a total of 1 000 000 generations each (ngen= 1 000 000) and a burn-in of 50 000. Initially, 10 000 generations were added until the split frequency stabilized or dropped below 0.1. Default priors were used for the number of trees sampled (every 100 generations) and the percentage of trees discarded as the burn-in phase (the first 25%). Bayesian posterior probabilities were estimated from the last 7500 trees. These results were repeated at least once to ensure consistency. Consensus trees were constructed through the 'sumt' command and visualized in TREEVIEW version 1.6.6 (Page, 1996). For all three methods used, nodes were considered as unsupported when bootstrap values were less than 70% and Bayesian posterior probability were less than 0.95.

In addition, genealogical relationships between the species were examined by haplotype networks constructed by the reduced median-network approach (Bandelt *et al.,* 1995, 2000) in NETWORK version 4.2.0.1. According to the program's recommendations, the reduced median approach is particularly effective for sequence sets representing large geographical distances. A second network was also inferred from the sequences representing only the South African species. Finally the estimates for pair-wise genetic distances among the South African species were calculated in MEGA version 4.0.

5.3 Results

5.3.1 Combined dataset of ND1 and haemocyanin

The combined dataset of the ND1 and haemocyanin genes contained 21 individuals representing eight species of the genus *Haliotis*, as well as one outgroup sequence. After truncation of the original alignments and combination of the two datasets, the final alignment contained 960 characters, of which 581 (60%) were variable, 441 (46%) parsimony-informative and 140 were singletons. The transition: transversion ratio was 1,2:1 with a relatively strong bias towards transition substitutions with a rate of *R* = 9.677. Although base frequencies were unequal within sequences (A=38,4%, C=25,4%, G=14,1%, T=21,6%), base homogeneity was confirmed across all taxa excluding the outgroup. The web-based Modeltest selected the TrN+G (loglk= -4477.26642) as the substitution model that best fitted the sequence data .

The maximum parsimony (MP) analysis resulted in 63 equally parsimonious trees with length=1084, a consistency index (CI) of 0.733 and a retention index (RI) of 0.857. The high CI and RI values are indicative of a strong signal obtained with the data for all 63 trees. The consensus MP tree shows the bootstrap and Bayesian probability support of all three phylogenetic analysis methods used for this study (Figure 5.2). The ML tree was identical to the parsimony tree except for lower internodal bootstrap support within members of the same species. A Bayesian consensus tree with similar topology was produced by MrBayes with posterior probability values comparable to the ML and MP bootstrap values. The topologies of all three phylogenetic trees were similar and characterised by a strong separation of a minor clade (*H. fulgens* and *H. iris*) from the other species. Support values for the nodes between species were relatively high ranging from 62% to 100% bootstrap

values and 0.93 to 1.00 Bayesian posterior probability values. Strong support was obtained for the clustering of the South African and Australian clades together while the New Zealand species showed no correspondence between phylogeny and geographic distribution. Very high bootstrap (96% and 97%) and Bayesian posterior probability support (1.00) was given to a separate Indo-Pacific group comprising of *H. asinina* within the Southern-Pacific clade. Of most importance to this study, is the clustering of the South African species with the Australian endemic *H. rubra*, and the subclade separation of *H. parva* from *H. midae* and *H. spadicea*.

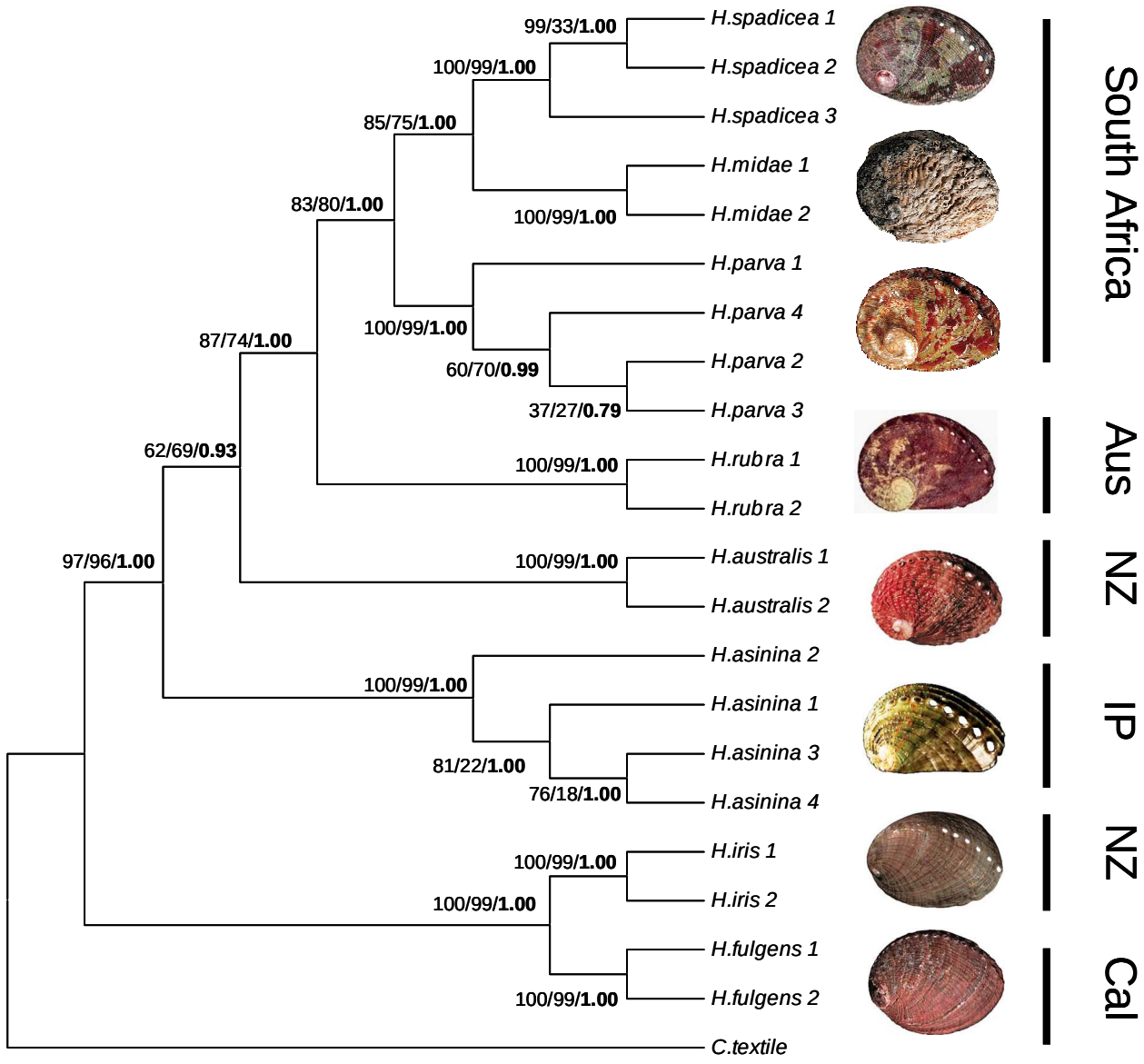

Figure 5.2 A maximum parsimony tree for *Haliotis* based on the combined ND1 and haemocyanin DNA sequence. Maximum likelihood and parsimony bootstrap values are shown in plain text and Bayesian posterior probabilities in bold. Clades are named according to geographical origin e.g. Australia (AUS), New Zealand (NZ), Indo-Pacific (IP) and California (CAL).

The reduced-median networks identified a total of 16 haplotypes and corresponded to the topology of the MP, ML and Bayesian trees. Eight lineages were distributed throughout the network but the two lineages of *H. iris* and *H. fulgens* were

clearly separated from the rest (Figure 5.3). The lineages of *H. asinina*, *H. midae* and *H. spadicea* were represented by a single haplotype each, whereas the lineage of *H. parva* contained four, the highest number of haplotypes for all the species included. The network based only on South Africa samples reflected the close relationship of the two larger species, *H. midae* and *H. spadicea*, with only two median vectors separating their haplotypes. As with the ML, MP and Bayesian trees, the *H. parva* haplotypes represented the evolutionary most distant species within the South African cluster. Similarly the pairwise genetic distances were slightly higher between *H. parva* and the rest of the South African species while the smallest genetic distance were obtained between *H. midae* and *H. spadicea* (Table 5.2).

Table 5.2 Pairwise genetic distances (below diagonal) and standard errors (above diagonal) among the South African species

	1	*2*	*3*
H. midae	–	0.023	0.026
H. spadicea	0.054	–	0.030
H. parva	0.061	0.071	–

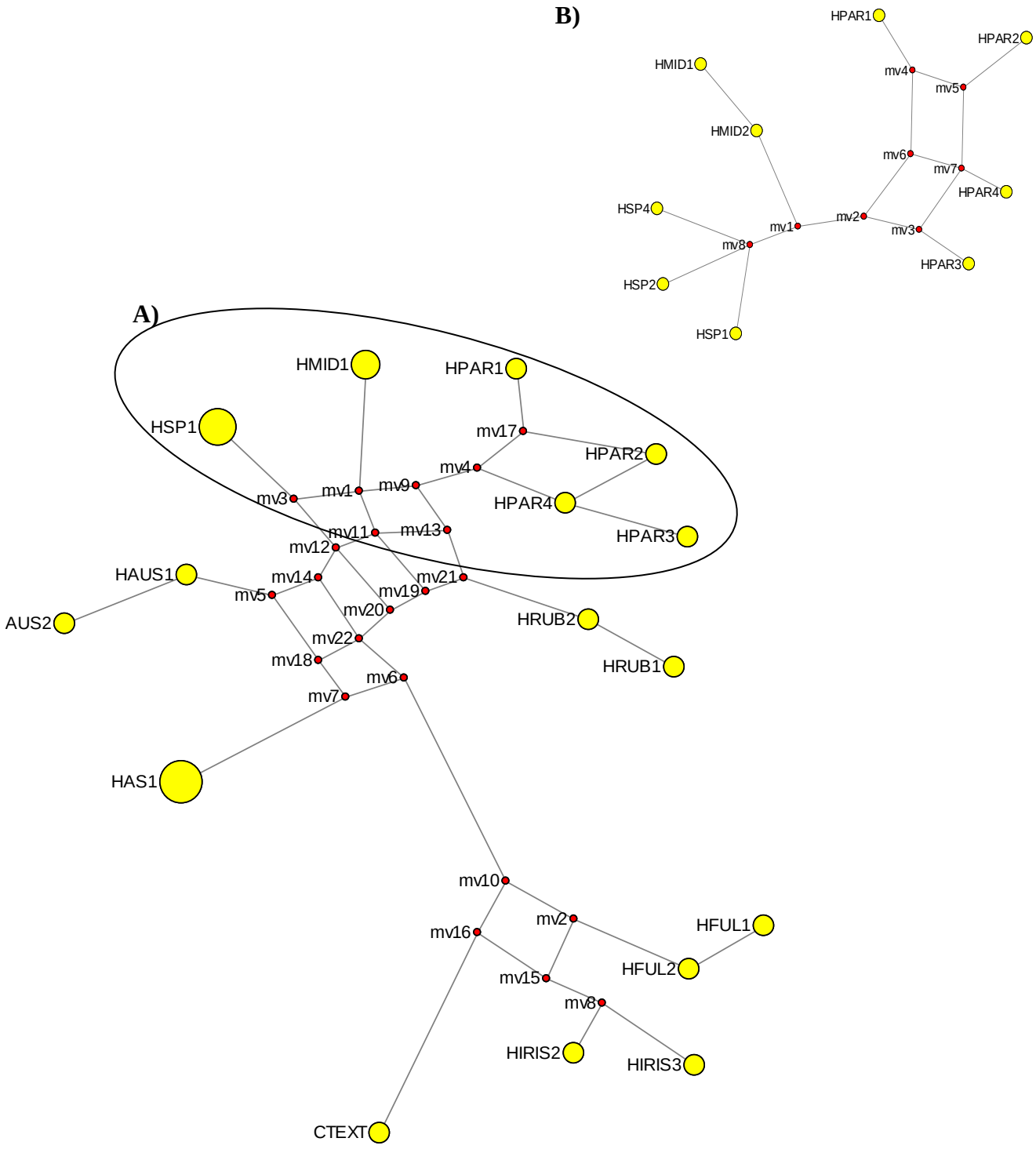

Figure 5.3 Reduced median-networks using combined ND1 and haemocyanin DNA sequence data showing genetic relationship among A) eight *Haliotis* lineages and B) the South-African lineage. Haplotype frequency is indicated by the size of the circle. HAUS= *H. australis*; HAS= *H. asinina*; HFUL= *H. fulgens*; HMID= *H. midae;* HPAR= *H. parva*; HRUB= *H. rubra*; HSP= *H. spadicea*; CTEXT= *C. textile* (outgroup).

5.3.2 Haemocyanin dataset

It was interesting to note the significant differences in product size (450-1100bp) of the haemocyanin sequences obtained for the various *Haliotis* species. This was mainly attributed to the highly variable length of the intron sequence incorporated in the fragment. The largest intron in this study was found in the *H. iris* gene fragment. With the GenBank sequences included, the final alignment of haemocyanin sequences of 26 taxa (representing 12 *Haliotis* species) contained 371 characters in total. The dataset was characterised by very high nucleotide diversity and revealed 315 (85%) variable sites. The same *Megathura crenulata* haemocyanin type two fragment used by Streit *et al.* (2006) was used as outgroup in this study. Sixty-six percent of the characters were parsimony informative, indicative of the fragments' suitability for species level phylogenetic reconstruction as observed by Streit *et al.* (2006). The MP, ML and Bayesian analyses resulted in similar topologies with slightly different nodal support values. Similarly to the results obtained with the combined dataset, species were grouped into two major clades, the North and the South Pacific group both with relatively strong support (bootstrap >90; posterior probability >95%). The two clades represented the same species as with the combined data, but in contrast the South Pacific clade was comprised of two subclades; one comprising of all three the SA species and one including the rest of the southern hemisphere species. Once again, the New Zealand species (*H. iris* and *H. australis*) resided in different clades, with *H. iris* in the North Pacific and *H. australis* in the South Pacific clade. Despite the poor internodal support within the South African subclade, the topology was in general agreement with the phylogenetic tree presented by Streit *et al.* (2006). To allow for easy comparison

with their data the Bayesian tree with bootstrap and posterior probability support values is shown in Figure 5.4.

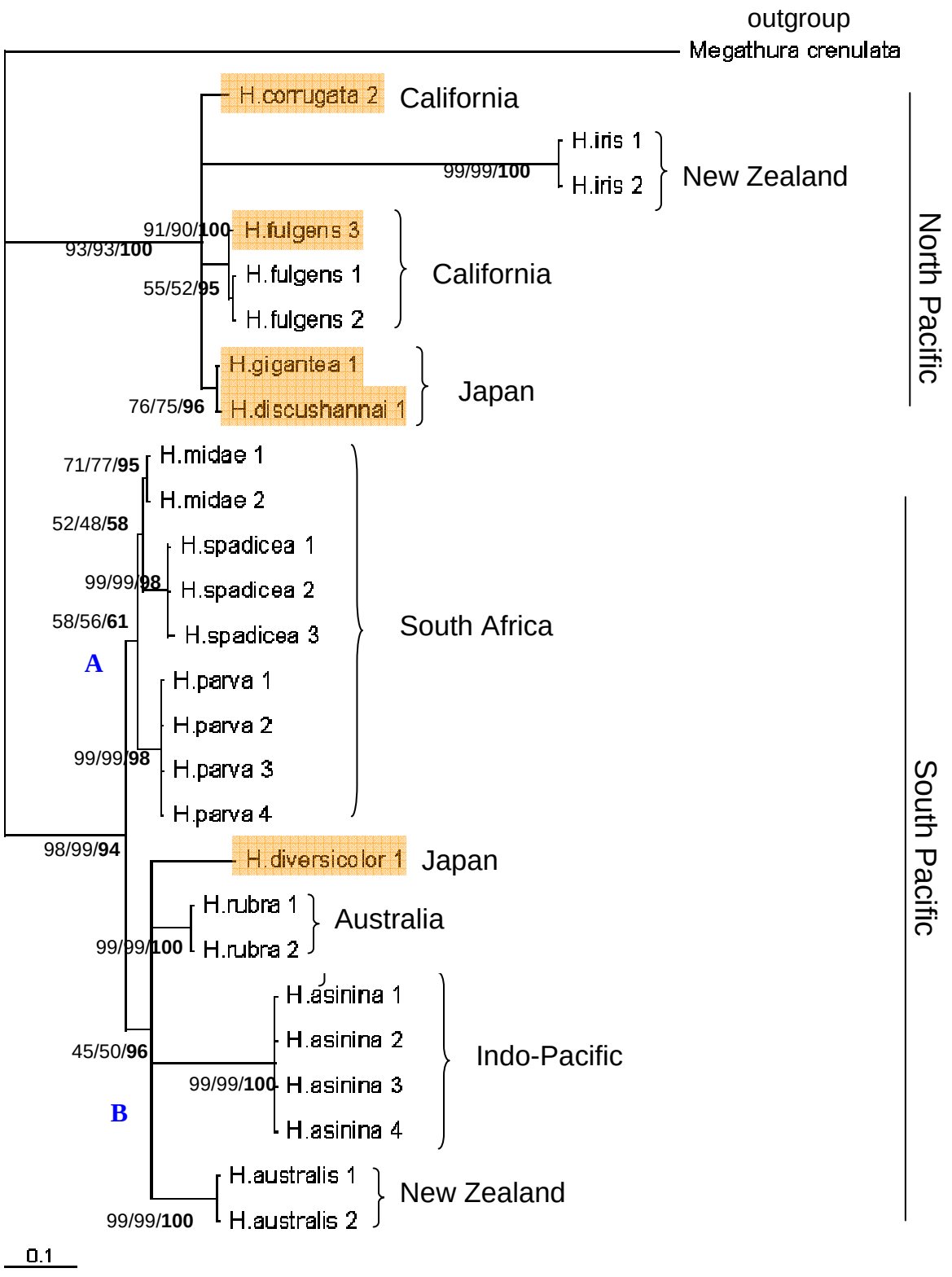

Figure 5.4 A Bayesian probability phylogram for 12 *Haliotis* species based on the haemocyanin isotope type 1 sequence. Maximum likelihood and parsimony bootstrap values are shown in plain text and Bayesian posterior probabilities in bold. Two clades (North Pacific and South Pacific) are identified of which the South Pacific clade comprises of two subclades (A and B). GenBank sequences are highlighted.

5.4 Discussion

5.4.1 Phylogeny of Haliotis

In the last decade, several studies addressed the world-wide phylogeny of abalone based on a range of molecular markers and different methodological approaches. Although taxon sampling has been a major caveat in most of these studies, consensus has been reached about either an ancient Tethyan (Geiger and Groves, 1999) or a broad Indo-Pacific origin (Lindberg, 1992; Briggs, 1999; Geiger, 2000) of Haliotidae and the existence of two major lineages with Northern and Southern-Pacific distribution. Except for the relationship among the three New Zealand endemics that vary somewhat between different data sets, most other well studied *Haliotis* species have been assigned to distinct evolutionary groups or lineages. The results presented here are in general agreement with the phylogenetic patterns observed previously. The major significance of this study lies in the inclusion of a complementary mitochondrial gene region, ND1, and two additional South African species, *H. spadicea* and *H. parva*.

The large congruence of this data with the more recent phylogenetic investigations by Geiger (1999), Estes *et al.* (2005), Streit *et al.* (2006) and Degnan *et al.* (2006) firstly confirms the species' identity of the taxa and secondly, the efficiency of ND1 as a novel phylogenetic marker for abalone. The presence of two clades, one representing the southern hemisphere abalone and the other consisting of one North American (*H. fulgens*) and one New Zealand species (*H. iris*), is a consistent feature throughout this study. The only obvious discrepancy is with the total evidence cladistic analysis of Geiger (1999) which proposes an isolated position for the type species, *H. asinina* based on combined analysis of allozyme frequencies, lysin and 16S mtDNA sequence and morphological data. In this study,

the position of *H. asinina* within the South-Pacific clade is well supported based on ND1 and haemocyanin data. Geiger further classifies abalone into four subgenera 1) *Haliotis, senso stricto* or the type species of the genus, *H. asinina* 2) *Nordotis*, the North Pacific species including the large Japanese (*H. discus*, *H. gigantea*) and the west American species (*H. fulgens, H. rufescens, H. sorenseni, H. cracherodii*, *H. corrugata*), 3) *Notohaliotis*, the endemic Australian species (*H. rubra*, *H. roei*, *H. scalaris*, *H. laevigata*) and 4) *Sanhaliotis,* the Indo-Pacific species (*H. ovina*, *H. varia*, *H. diversicolor*). Although in the current study taxon number is limited and not necessarily representative of all groups, at least one species of three of the proposed subgenera are included in the combined dataset. *H. diversicolor* represents also the fourth subgenera, *Sanhaliotis,* in the haemocyanin dataset.

From the analysis of the haemocyanin dataset presented here, an interesting observation was the identification of two groups within the southern hemisphere, represented by (1) the South African species and (2) mainly the Indo-Pacific and Australian species. Although Streit *et al.* (2006) distinguished a similar southern versus northern hemisphere phenomena using haemocyanin, the exclusion of *H. rubra*, the type species for *Notohaliotis*, as well as the New Zealand species did not allow for the identification of subclades within the southern group.

In general, the two major clades identified in this study correspond to the geographical distribution of abalone species south and north of the Indo-Pacific Ocean, while excluding the New Zealand endemics. Similar to the findings of Briggs (1999) and Geiger (2000) this study coincides with the possible origin of Haliotidae in the Indo-Pacific Ocean but due to incomplete sampling of taxa, other theories of origin such as the Tethys (Lee and Vacquier, 1995; Geiger and Groves, 1999) or the Pacific Rim (Talmadge, 1963) models can not be easily discarded. Geiger and

Groves (1999) documented the diploid chromosome number of 14 *Haliotis* species across its biogeographic range and postulated a model of progressive increase in the number of chromosomes from the Mediterranean (2n=26), through the Indo-Pacific (2n=32) to the North Pacific (2n=36). This suggested the earliest abalone were of Tethyan origin and radiated eastwards. Despite evidence for different places of origin, all three biogeographic models (the Indo-Pacific, the Tethys and the Pacific Rim) therefore coincide with an eastward radiation of haliotids.

Another feature consistent with previous findings is the somewhat controversial positioning of the two New Zealand endemics included in this study, especially *H. iris* grouping with the Californian species. The highly supported association of *H. iris* with *H. fulgens* and the basal positioning of *H. australis* within the southern hemisphere clade, suggest the presence of two independent lineages. Although this current study only resolves two of the three New Zealand species (*H. iris*, *H. australis* and *H. virginea*), it is comparable to the most recent evolutionary placement of these endemics by Degnan *et al.* (2006). Contrary to Geiger's (2000) total evidence cladistic analysis that suggested three individual colonizations of the island, the study by Degnan *et al.* (2006) clearly shows the isolated or separate divergence of *H. australis* from *H. iris* and *H. virginea.* Interestingly, Miocene fossil records are available for *H iris*, whereas more recent Pleistocene records are available for *H. australis* and *H. virginea* (reviewed in Geiger and Groves, 1999), suggesting a more recent speciation event for the latter. While the current placement of the three species can either be explained by vicariance or dispersal events, e.g. ancient lineages that became isolated during the drifting of the Gondwana landmass (Degnan *et al.,* 2006) or dispersal from the North Pacific to the South Pacific, the pattern of multiple colonization of New Zealand remains somewhat surprising.

5.4.2 Phylogeny of South African haliotids

As mentioned before, the primary objective of this work was to place the South African endemics within the larger evolutionary framework of Haliotidae and to explore particular affinities within the group. With particular reference to *H. midae*, the topology obtained with the combined dataset concurs with the original sperm lysin data of Lee *et al.* (1995) up to the most recent observations by Degnan *et al.* (2006). The close relationship of *H. midae* with the Australian endemics is well documented and *H. midae* has often been referred to as the sister taxa of *H. rubra*, the type species of the *Notohaliotis* genus. Together with fossil data of both *H. midae* and *H. rubra* from the Pleistocene era (Geiger and Groves, 1999) and their geographical positioning in relation to each other, there is little reason to question the evolutionary status of the South African species within the temperate species clade. The haemocyanin only dataset is quite unique in that it represents an exclusively South-African (monophyletic) clade which is separated from the subclade containing the rest of the southern hemisphere species (*H. rubra*, *H. australis*, *H. asinina*). According to this topology, the South African species are placed in the same evolutionary context as the Australian and Japanese (Indo-Pacific) species and similar to Streit *et al*. (2006) suggests a common ancenstry of Tethyan origin. However, the basal positioning of *H. rubra* to the South African group as reflected with the combined dataset and the recent account of the chromosome number of *H. midae* of 2n=36 (Van der Merwe and Roodt-Wilding, 2008), is more suggestive of a recent radiation from the Indo-Pacific area via Australia. As no other study has included more than one South African species, the equivalent ND1 sequences of the species used in the study by Streit *et al.* (2006) as well as chromosomal data of

more southern radiated species, might help to resolve the evolutionary origin of the SA species completely.

Considering the geographical distribution range and endemic nature of the South African species, it is not surprising that all of them are contained within a single, highly supported clade. Of particular interest is the topology within the clade that signifies a more recent divergence between the two larger species, *H. midae* and *H. spadicea*, while the lesser abundant species, *H. parva*, is placed ancestrally to the rest of them. *Haliotis parva* has in the past been classified in synonymy with some of the Australian species such as *H. scalaris,* because of similar shape and its distinctive spiral ridge on the shell and supports the colonization of South Africa via Australia through the ancestral species *H. parva.* Also, the earlier diversification of *H. parva* is supported by every phylogenetic analysis in this study as well as the reduced median-networks which illustrate the *H. parva* lineage in very close relationship with *H. rubra.*

When the South African group is compared with the phylogeny of for example the Australian group (Degnan *et al.,* 2006), the short branch lengths that characterize this clade is suggestive of a relatively modern radiation of all the SA species. This for the most part rules out the direct radiation of the SA species from the Tethyan Sea as proposed by the haemocyanin data alone. Due to insufficient sampling and the absence of fossil data, in particular for *H. parva*, the early history of the South African species remains complex. In view of their derived position in most phylogenetic trees of this study, a more recent colonization of Indo-Pacific origin is proposed as the most likely scenario for the evolutionary history of South African abalone.

The particular phylogenetic relationship of the South African species to other southern-temperate species could either be explained by speciation events related to

founder dispersal or vicariance. Trans-oceanic larval dispersal has been suggested for several invertebrate and fish species as a possible mode of allopatric speciation (Craig *et al.,* 2004; Teske *et al.,* 2005; Lessions and Robertson, 2006). Wells and Kilburn (1986) for example recorded an apparent transport of three temperate gastropod species between the continents of South Africa and Australia. An in-depth study on *Patelloida profunda* (marine gastropod: *Lottidae*), however highlights the difficulty of determining evolutionary directionality from phylogenetic trees in a dispersal-driven system (Kirkendale and Meyer, 2004) and shows that even with a well-resolved topology, several scenarios might explain diversification within a species group. In view of the short-lived larval stages characteristic of *Haliotis* (Hamm and Burton, 2000; Imron *et al.,* 2007) and no evidence for other methods of contemporary introductions such as rafting of abalone larvae, dispersal can not necessarily be regarded as the sole reason for cladogenesis within the temperate *Haliotis* species included in this study. From the fossil records which dates back only to the Pleistocene and the karyotypes, the geographical origin of the South African species is more likely explained by founder dispersal than vicariance. Invasion of the African coastline as a result of dispersal from the Indian Ocean has been well-documented in the literature (Gibbons *et al*., 1995; Gibbons, 1997; Connolly *et al*., 2003; D'Amato *et al*., 2008).

Speciation within the South African group is again most likely a consequence of ecological specialization. The split between *H. parva* and the rest of the South African species in all probability occurred because of differential depth range (subtidal vs intertidal) whereas the more recent split between *H. midae* and *H. spadicea* could be a consequence of differential adaptation to habitat.

5.5 Conclusions

In summary, the present study makes successfully a first classification of South African abalone as a monophyletic group within the Haliotidae family. Throughout the study, phylogenetic inference positioned the SA group within the southern hemisphere clade while subclade analysis suggests that *H. parva* is the older lineage of the SA group. The placement of the SA species within phylogenetic trees and median-joining networks render evidence for their close affinity to other southern radiated species (Australia and New Zealand). Together with the karyotype, the overall topology based on the combined mitochondrial and nuclear sequences is highly suggestive of a relatively recent radiation of the SA species from the Indo-Pacific basin from Australia via the Indian Ocean. Despite incomplete sampling, this study corroborates all the major evolutionary studies of the Haliotidae family to date. If the two remaining SA species, *H. queketti* and *H. speciosa,* are included and they all cluster together, a common origin to South African abalone can be hypothesized while speciation within the group may be related to ecological specialization.

References

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141:** 743–753.

Bandelt HJ, Macaulay VA, Richards MB (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution* **16:** 8–28.

Briggs JC (1999) Coincident biogeographic patterns: Indo-West Pacific Ocean, *Evolution* **53:** 326–335.

Brown LD (1993) Biochemical genetics and species relationships within the genus *Haliotis* (Gastropoda: Haliotidae). *Journal of Molluscan Studies* **59:** 429–443.

Coleman AW and Vacquier VD (2002) Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*). *Journal of Molecular Evolution* **54:** 246–257.

Connolly SR, Bellwood DR, Hughes TP (2003) Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. *Ecology* **84:** 2178–2190.

Craig MT, Hastings PA, Pondella DJ (2004) Speciation in the Central American seaway: the importance of taxon sampling in the identification of trans-isthmian geminate pairs. *Journal of Biogeography* **31:** 1085–1091.

D'Amato ME, Harkins GW, de Oliveira T, Teske PR, Gibbons MJ (2008) Molecular dating and biogeography of the neritic krill *Nyctiphanes Marine Biology* (online article)

Degnan SM, Imron, Geiger DL, Degnan BM (2006) Evolution in temperate and tropical seas: disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae*). Molecular Phylogenetics and Evolution* **41:** 249–256.

Estes JA, Lindberg DR, Wray C (2005) Evolution of large body size in abalone (*Haliotis*): patterns and implications*. Paleobiology* **31:** 591–606.

Geiger DL (1999) A total evidence cladistic analysis of the Haliotidae (Gastropoda: Vetigastropoda). Ph.D. dissertation, University of Southern California, Los Angeles, 423pp.

Geiger DL (2000) Distribution and biogeography of the recent Haliotidae (Gastropoda: Vestigastropoda) world-wide. *Bollettino Malacologico* **35:** 57–120.

Geiger DL and Groves LT (1999) Review of fossil abalone (Gastropoda: Vetigastropoda: Haliotidae) with comparison to recent species. *Journal of Paleontology* **73:** 872–885.

Gibbons MJ (1997) Pelagic biogeography of the South Atlantic. *Marine Biology* **129:** 757–768.

Gibbons MJ, Barange M, Hutchings L (1995) Zoogeography and diversity of euphausiids around southern Africa. *Marine Biology* **123:** 257–268.

Guindon S and Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systems Biology* **52:** 696–704.

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

Hamm DE and Burton RS (2000) Population genetics of the black abalone, *Haliotis cracherodii*, along the central Californian coast. *Journal of Experimental Marine Biology and Ecology* **254:** 235–247.

Imron, Jeffrey B, Hale P, Degnan BM, Degnan SM (2007) Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. *Molecular Ecology* **16:** 289–304.

Kirkendale LA and Meyer CP (2004) Phylogeography of the *Patelloida profunda* group (Gastropoda: Lottidae): diversification in a dispersal-driven marine system. *Molecular Ecology* **13:** 2749–2762.

Lee YH and Vacquier VD (1995) Evolution and systematics in Haliotidae (Mollusca, Gastropoda): inference from DNA sequences of sperm lysin. *Marine Biology* **124:** 267–278.

Lee YH, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm. *Molecular Biology and Evolution* **12:** 231–238.

Lessions HA and Robertson DR (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society of London. Series B* **273:** 2201–2208.

Lindberg DR (1992) Evolution, distribution and systematics of Haliotidae. In: *Abalone of the world: biology, fisheries and culture.* (ed. Shepherd SA and Tegner M, SA) Blackwell Scientific, Oxford, pp3–18.

Page RDM (1996) Treeview: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12:** 357–358.

Posada D and Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Ronquist F and Huelsenbeck JP (2003) MrBayes: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Streit K, Geiger DL, Lieb B (2006) Molecular phylogeny and the geographic origin of Haliotidae traced by haemocyanin sequences. *Journal of Molluscan Studies* **72:** 111–116.

Talmadge RR (1963) Insular haliotids in the western Pacific. *The Veliger* **5:**129–139.

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) version 4.0. *Molecular Biology and Evolution* **24:** 1596–1599.

Teske PR, Hamilton H, Palsbøll PJ, Choo CK, Gabr H, Lourie SA, Santos M, Sreepada A, Cherry MI, Matthee CA (2005) Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. *Marine Ecology Progress Series* **286:** 249–260.

Van der Merwe M and Roodt-Wilding R (2008) Chromosome number of the South African abalone *Haliotis midae. African Journal of Marine Science* **30:**195–198.

Wells FE and Kilburn RN (1986) Three temperate water species of South African gastropods recorded for the first time in southwestern Australia. *Veliger* **28:** 453–456.

Chapter VI

Concluding comments on the population dynamics of *Haliotis midae*: implications on future management and conservation strategies

6.1 Population dynamics in Haliotis midae

Based on the microsatellite and SNP genotypes analysed in Chapters III and IV, strong evidence for subtle differentiation within the intertidal abalone, *Haliotis midae* was found. Results from both marker systems gave an appreciation for the various methodologies needed to depict low levels of structuring and highlighted the concern for drawing conclusions from simplistic analysis such as summary statistics alone. Varying levels of support were obtained with different analysis methods but based on the AMOVA, FCA and clustering results, it was reasonable to reject the hypothesis of panmixia. As no significant signs supporting a founder event, such as reduced allelic richness or smaller population sizes at the extremes of the species range were noticeable, the easterly range expansion from a founder event proposed by Evans *et al.* (2004) was also rejected. The most prominent physical feature responsible for limiting gene flow appeared to be the barriers originating from the ocean currents along the South African coast: the southward flowing Agulhas on the East coast and the Benguela upwelling system on the West coast. Although there is no substantial evidence for two evolutionary separate lineages, historical signatures in structure might be concealed by these currents and the causal events behind the observed structure can therefore only be speculated on.

Taking into consideration for instance the dispersal ability of the species (a 5-7 day larval period) and an average temperature difference of 7℃ recorded between the two coastal regions, it is most likely that contemporaneous current and temperature effects together with historical divergence shaped the population structure and dynamics of *H. midae* today.

In a marine realm where boundaries to gene flow are mostly diffuse rather than distinct, a number of mechanisms have been suggested to explain population structure. From a historical perspective, the long-term effects of global climatic changes on contemporary marine populations have been well recognized in particular for Northern Hemisphere taxa (Hellberg *et al.,* 2001; Hickerson and Ross, 2001; Gysels *et al.,* 2004; Hickerson and Cunningham, 2005; Adams *et al.,* 2006). Especially coastal glaciation during the ice ages is believed to have had a major impact on marine species, most often creating a founder effect when new habitats are colonized (Widmer and Lexer, 2001; Olsen *et al.,* 2004; Fraser and Bernatchez, 2005; Wilson, 2006). For South African abalone, the direct effect of glaciations on population structure is expected to be less pronounced than for Northern hemisphere abalone since glaciers were almost absent from southern Africa during the latest ice age (Denton and Hughes, 1981). However, associated consequences of glaciations such as latitudinal shifts in water temperature or sea level changes could have been the underlying shapers of population structure in contemporary populations of *H. midae*. It is possible that abalone along the southern most tip of Africa were not able to survive or adjust to the much lower water temperatures during inter-glacial periods, giving rise to isolated refugia between the West and East coast of South Africa. Sea-level fluctuations during glacial to inter-glacial periods in excess of 100 m are also expected to have had a large impact on the

circulation of currents in both southern and northern hemisphere coastal environments (Cutler *et al*., 2003; Miller *et al*., 2005). An analogous case in the southern hemisphere represents the marked disjunction between West and East coast Australian nerites (marine gastropod) associated with a paleogeographic barrier formed during the Pleistocene glacial maxima, and which is in striking contrast to the species dispersal ability (Waters *et al.,* 2005).

Besides historical processes, population structure in marine species has also been ascribed to a range of contemporaneous physical and ecological forces mainly as a consequence of geographical distance, oceanography, environmental transitions and life-history changes (Hemmer-Hansen *et al.,* 2007). In the case of *H. midae*, where all sampling was done within one generation time, life-history shifts such as different time and place of spawning between sites were expected to have little impact on structure. The main contributing factors predicted are 1) geographical distance; 2) ocean hydrodynamics together with temperature gradients and 3) habitat adaptation.

1) In the absence of isolation-by-distance and any significant pattern of pairwise differences, geographical distance *per se* did not seem to have a substantial effect on the structuring of *H. midae* populations. Although this is not surprising in marine populations with their high potential for long-distance dispersal, a number of studies have shown the contrary. Different levels of correlation between geographical and genetic distance has, for example, been reported for European eel (Wirth and Bernatchez, 2001); cod (Hutchinson *et al.,* 2001); European flat oyster (Launey *et al*., 2002); Atlantic herring (Mariani *et al.,* 2005); sea scallop (Kenchington *et al*., 2006) and common cuttlefish (Pérez-Losada *et al.,* 2007). Even more

important is the direct association of geographical distance with restriction to larval dispersal. In Atlantic salmon for instance assignment tests detected restricted dispersal that was significantly correlated with geographical distance where the commonly used Mantel test failed to reveal an increase of genetic differences with distance (Castric and Bernatchez, 2004). In *H. midae*, one would also expect larvae to follow a stepping-stone model of dispersal (Slatkin, 1993) but with a relatively short larval period characteristic of marine gastropods, survival of larvae within a given habitat rather than geographical distance *per se* are expected to determine the pattern of dispersal.

2) Ocean currents and thermal fronts may also influence larvae dispersal, by acting as an oceanographic barrier and thereby keeping larvae (and eggs) from dispersing naturally. As mentioned before, the primary oceanographic currents along the South African coast are the south-flowing Agulhas current on the East coast and the northward flowing Benguela current on the West coast. For *H. midae* in particular the inshore thermal front created by the Agulhas current can have a substantial effect on larvae retention in that area. It seems likely that the barrier created by the retroflection of the Agulhas current (Dijkstra and de Ruijter, 2001), may prevent larvae from dispersing further or act as a barrier to the dispersal of larvae from west to east. Recently, Zardi *et al.* (2007) found unexpected structure between indigenous mussel populations from the West and East coast of South Africa and suggested that together with other historical and contemporary factors, the Agulhas current could be the main feature responsible for the genetic disjunction.

Similarly, the thermal front at Port Elizabeth could affect dispersal patterns along the east coast of South Africa although not reflected in the population structure predicted here for *H. midae*.

3) The increasing evidence for association of intraspecific lineages with the three major marine biogeographic provinces of South Africa (Ridgeway *et al.,* 1998; Teske *et al.,* 2006; Edkins *et al*., 2007; Teske *et al.,* 2007), supports the idea that region-specific oceanographic processes plays a significant role in the structuring of species along the coast. Another example is the phylogeographic break found in the crustacean *Upogebia africana* that coincides with the boundary demarcated between the cool-temperate and warm-temperate provinces of South Africa (Teske *et al*., 2006). Species distributed along more than one of these coastal biogeographic provinces, are often subjected to diverse physical and biological environments, leading to local adaptation and genetic divergence. With *H. midae*, the broad morphological similarity between West and East coast samples suggests that abalone is well adapted to a wide range of environmental factors, including temperature, However, the possible contribution of differential selective forces to the apparent genetic discontinuity can only be excluded based on experiments specifically designed to detect adaptive divergence.

Considering all contemporary factors it is more likely that the biogeographic break at Cape Agulhas is the main restriction to gene flow and that what is observed today is a classic example of formerly allopatric populations that came into secondary contact following range expansion in both directions (also supported by clinal variation observed with the microsatellite data). As with most other studies

of genetic structure in the marine environment (Roques *et al.,* 2002; Nielsen *et al.,* 2004; Perez-Losada, 2007) several evolutionary mechanisms may be acting simultaneously to structure populations of *H. midae*.

6.2 Conservation and Management implications

Regardless whether the genetic discontinuity between West and East coast abalone is a consequence of historical or present-day factors, this study provides evidence that populations of abalone on either side of Cape Agulhas are to some degree differentiated. The recommendation based on the results presented herein is that these populations / regions should be considered separately in terms of management. Lack of isolation by distance or further structure within either of the groups supports two panmictic populations adapted to West and East coast environmental conditions respectively. To date, *H. midae* has been treated by government authorities and commercial fisheries as a single reproductive stock. Management of the South African abalone in the last few decades relied on measures such as minimum legal size (MLS), annual quotas, total allowable catch (TAC), closed seasons and marine reserves (Tarr, 1992; 2000). Increasing threat to the resource made a complete ban of abalone fishing in February 2008 inevitable, but still leaves the remaining stocks vulnerable to genetic mismanagement. Repercussions thereof together with the huge threat of illegal trade could jeopardise any efforts to rebuild depleted stocks in the long run. As recovery of abalone fisheries around the world has proven to be a major challenge (Campbell *et al*., 2000; Seki and Taniguchi, 2000; Tegner, 2000) an integration of genetic management with knowledge and removal of ecological threats are paramount to ensure conservation of this species.

First and foremost, the subtle but significant genetic structure found between West and East coast *H. midae* populations should be integrated into any management program intended for *H. midae*. Throughout literature, populations are typically recognised as separate management units when significant genetic differentiation is found between them (Moritz, 1994; Carvalho and Hauser, 1994). For *H. midae* this means one management unit spanning approximately 480km from Cape Agulhas up to St Helena Bay on the West coast and another spanning almost 900km from Cape Agulhas to the northern coast of Transkei. The main objective should be to maintain *H. midae* populations in their respective West and East coast natural settings to which local adaptation may have occurred. As pointed out by Crandall *et al*. (2000) and Merilä and Crnokrak (2001) significant adaptive divergence may have occurred even if molecular markers show little or no genetic difference between geographically separated groups. Crandall *et al*. (2000) also introduced the concept of exchangeability where populations are defined as conservation units only when the hypothesis of genetic as well as ecological exchangeability on recent and historical time scales can be rejected. Therefore, in *H. midae* where ecologically data is for the most part absent, management of the West and East coast groups as distinct evolutionary units based on genetic data alone, could in fact compromise the adaptive diversity and therefore evolutionary potential of the species (Crandall *et al*., 2000; Fraser and Bernatchez, 2001). A more realistic approach would thus be to treat the species as populations connected by various degrees of gene flow while aiming to preserve the natural connectivity between them. However, management actions should also depend on the species' need for conservation (Laikre *et al*., 2005) and in the case of *H. midae*

which is under a great deal of pressure, a precautionary attitude in this respect seems justifiable.

Since the current ban on commercial fishing of *H. midae* to some extent nullifies the potential risk for loss of genetic diversity or compromising stocks due to selective fishing, management actions should also be targeting commercial abalone fisheries and intended recovery programs. While regulations have already been placed on the location of commercial farms to reduce negative impact on the immediate surroundings of wild stocks, collection of brood stock are still performed without genetic control monitoring. Farms should be encouraged to develop protocols where the impact of wild brood stock removal on the natural populations is reduced and distinct genetic diversity are preserved by using brood stock that is highly representative of the surrounding wild abalone. The transport of larvae or juveniles between farms situated in different coastal regions should also be regulated whilst minimizing accidental release of commercial stock into the wild. It is furthermore recommended that genetic diversity within commercial stocks is evaluated on a regular basis to avoid unintentional contamination of the wild stocks with animals from a potential different adaptive background. As recovery and enhancement programs entail intentional artificial seeding of cultured abalone, it ought to be carried out under the same or even more stringent regulatory conditions. Most importantly, the hatchery-reared animals should match the genetic diversity of the wild gene-pool where they are going to be released. According to Goodsell *et al*. (2006) and Joaquim *et al*. (2008) it is also advisable and more productive to release fewer but larger abalone which could survive better in the wild whilst being less disruptive to the immediate surroundings. Ideally this means that in South Africa, potential recovery of abalone depleted populations from currently

available commercial stocks should only be allowed if performed within the West vs. East coast structure framework.

Despite cessation of commercial fishing, the role of marine protected areas (MPAs) in the future conservation of South African abalone remains invaluable (Attwood *et al*., 1997). Already, MPAs are believed to have prevented complete exhaustion of abalone stocks in areas such as Robben Island on the West coast, De Hoop and Storms River mouth on the South coast and Bird Island on the East coast (Attwood *et al*., 2000). Evaluation of MPAs for Californian abalone for example revealed that protected sites contained significantly more abalone than nearby fished sites and that the mean size and potential egg production was also larger (Rogers-Bennett *et al*., 2002). Along the South African coast, the 19 existing MPAs intended to conserve ecosystems should be evaluated for their use as conservation sites for abalone. It is expected that MPAs that protect species such as kelp and sea urchins will also indirectly benefit abalone. In the laboratory, juvenile abalone survived better in predation experiments where sea urchins were available for shelter and demonstrated how MPAs designed for one fished species could potentially help to protect another (Rogers-Bennett and Pearse, 2001). Of relevance to *H. midae* is a study on the diet of the South African rock lobster (*Jasus lalandii*) that showed that their consumption of sea urchins and juvenile abalone decreased dramatically in areas where black mussels were available (Mayfield *et al*., 2001). Particularly along the West coast where rock lobsters are considered to be a major threat to the survival of juvenile abalone, MPAs designed to enhance mussel production can therefore indirectly assist in the protection of suitable habitat for abalone.

In addition, the level of structure identified in *H. midae* could have relevant implications for the scientific procedures used to enhance enforcement on illegal fishing and exploitation of abalone. To date DNA forensics has only been implemented to identify and distinguish *H. midae* from other South African species (Sweijd *et al*., 1998; Roodt-Wilding and Bester, 2006) but assignment analysis using pure West and East coast abalone as reference samples could in the future help to define from which genetic stock or coastal region confiscated abalone most likely originated from. This in turn could assist in the identification of areas most vulnerable to illegal activities and encourage the development of multidisciplinary compliance efforts most beneficial to such regions.

In conclusion, the conservation measures discussed above underlines the importance of integrating genetic and environmental information for use in management decisions regarding TACs, fishery openings or closures, MPAs and enhancement sites for the recovery and conservation of South African abalone. In the future, experiments of ecological exchangeability in which adaptive divergence in relevant traits is tested, could assist in a more integrative designation of conservation units for *H. midae.* Difference in water temperature tolerance between West and East coast abalone for example could give further evidence towards non-exchangeability between groups and increase the support for separate management units. Also, the fact that the population sizes decline remained undetected in this study should by no means detract from the value that the generated historical demographic data has when compared to similar data measured in two to three generations' time. Seeing that the same generation subjected to poaching was sampled, the decline is probably too recent to be detected by the utilized methods. As for now, it is still common practice to define

management or conservation units on genetic differentiation alone and the information on the genetic population structure provided in this study should be regarded as crucial for sustainable management of *H. midae*. Integrating this data with the objectives of all potential role players (government, industry and public) could without a doubt benefit the protection and sustainability of *H. midae* in the future. As an example, findings of this study have recently been communicated to the Abalone Association of South Africa (AASA) and support was given for adding this information in a response to the 'Guidelines for Marine Ranching and Stock Enhancement' issued on the 13th June 2008 by the Department of Environmental Affairs and Tourism, SA.

References

Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology* **15:** 1109–1123.

Attwood CG, Mann BQ, Beaumont J, Harris JM (1997) Review of the state of marine protected areas in South Africa. *South African Journal of Marine Science* **18:** 341-368.

Attwood CL, Moloney CL, Stenton-Dozey J, Jackson LF, Heydorn AEF, Probyn TA (2000) Conservation of marine biodiversity in South Africa. In: *Marine Biodiversity Status Report for South Africa, at the end of the 20th Century* (ed. Durham BD and Pauw JC), pp68-83.

Campbell A (2000) Review of northern abalone, *Haliotis kamtschatkana*, stock status in British Columbia. In: Workshop on Rebuilding Abalone Stocks in British Columbia (ed. Campbell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 41–50.

Carvalho GR and Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* **4:** 326–350.

Castric V and Bernatchez L (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology* **13:** 1299–1312.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **17:** 390–395.

Cutler KB, Edwards RL, Taylor FW, Cheng H, Adkins J, Gallup CD, Cutler PM, Burr GS, Bloom AL (2003) Rapid sea-level fall and deep ocean temperature change since the last interglacial period. *Earth and Planetary Science Letters* **206:** 253–271.

Denton GH and Hughes TJ (1981) *The Last Great Ice Sheets*. Wiley, New York.

Dijkstra HA and de Ruijter WPM (2001) On the physics of the Agulhas Current: Steady retroflection regimes. *Journal of Physical Oceanography* **31:** 2971–2985.

Edkins MT, Teske PR, Papadopoulos I, Griffiths CL (2007) Morphological and genetic analyses suggest that southern African crown crabs, *Hymenosoma orbiculare*, represent five distinct species. *Crustaceana* **80:** 667–683.

Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliot NG (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology Progress Series* **270**: 163–172.

Fraser DJ and Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology* **10:** 2741–2752.

Fraser DJ and Bernatchez L (2005) Allopatric origins of sympatric brook charr populations: colonization history and admixture. *Molecular Ecology* **14:** 1497–1509.

Goodsell PJ, Underwood AJ, Chapman MJ, Heasman MP (2006) Seeding small numbers of cultured black-lip abalone (*Haliotis rubra* Leach) to match natural densities of wild populations. *Marine and Freshwater Research* **57:** 747–756.

Gysels ES, Hellemans B, Pampoulie C, Volckaert FAM (2004) Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the

colonization of the Mediterranean and the North Sea. *Molecular Ecology* **13:** 403–417.

Hellberg ME, Balch DP, Roy K (2001) Climate-driven range expansion and morphological evolution in a marine gastropod. *Science* **292:** 1707–1710.

Hemmer-Hansen J, Nielsen EEG, Grønkjaer P, Loeschcke V (2007) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.) *Molecular Ecology* **16:** 3104–3118.

Hickerson MJ and Cunningham CW (2005) Contrasting quaternary histories in an ecologically divergent sister-pair of low-dispersing intertidal fish *(Xiphister)* revealed by multi-locus DNA analysis *Evolution* **59:** 344–360.

Hickerson MJ and Ross JRP (2001) Post-glacial population history and genetic structure of the northern clingfish (*Gobbiesox maeandricus*) revealed from mtDNA analysis. *Marine Biology* **138:** 407–419.

Hutchinson WF, Carvalho GR, Rogers SI (2001) Marked genetic structuring in localized spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series* **223:** 251–260.

Joaquim S, Gaspar MB, Matias D, Ben-Hamadou R, Arnold WS (2008) Rebuilding viable spawner patches of the overfished *Spisula solida* (Mollusca: Bivalvia): a preliminary contribution to fishery sustainability. *ICES Journal of Marine Science* **65:** 60–64.

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology* **15:** 1781–1796.

Laikre L, Palm S, Ryman N (2005) Genetic population structure of fishes: implications for coastal zone management. *Ambio* **34:** 111–119.

Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y (2002) Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism. *Journal of Heredity* **93:** 331–351.

Mariani S, Hutchinson WF, Hatfield EMC (2005) North Sea herring population structure revealed by microsatellite analysis. *Marine Ecology Progress Series* **303:** 245–257.

Mayfield S, de Beer E, Branch GM (2001) Prey preference and the consumption of sea urchins and juvenile abalone by captive rock lobsters (*Jasus lalandii*). *Marine and Freshwater Research* **51:** 773–780.

Merilä J and Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative genetic traits. *Journal of Evolutionary Biology* **14:** 892–903.

Miller KG, Kominz MA, Browning JV, Wright JD, Mountain GS, Katz ME, Sugarman PJ, Cramer BS, Christie-Blick N, Pekar SF (2005) The Phanerozoic Record of Global Sea-Level Change. *Science* **310:** 1293–1298.

Moritz C (1994) Defining “Evolutionarily Significant Units” for conservation. *Trends in Ecology and Evolution* **9:** 373–375.

Nielsen EE, Nielsen PH, Meldrup D, Hansen MM (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple

hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. *Molecular Ecology* **13:** 585–595.

Olsen JL, Stam WT, Coyer JA, Reusch TB, Billingham M, Bostroem C, Calvert E, Christie H, Granger S, Lumiere RL, Milchakova N, Oudot-le Secq M, Procaccini G, Sanjabi B, Serrao E, Veldsink J, Widdicombe S, Wyllie-Echeverria S (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology* **13:** 1923–1941.

Pérez-Losada M, Nolte MJ, Crandall KA, Shaw PW (2007) Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis. Molecular Ecology* **16:** 2667–2679.

Ridgeway TM, Stewart BA, Branch GM, Hodgson AM (1998) Morphological and genetic differentiation of *Patella granularis* (Gastropoda: Patellidae) along the coast of South Africa. *Journal of Zoology, London* **245:** 317–333.

Rogers-Bennett L and Pearse JS (2001) Indirect benefits of Marine Protected Areas for juvenile abalone. *Conservation Biology* **15:** 642–647.

Rogers-Bennett L, Haaker PL, Karpov KA, Kushner DJ (2002) Using spatially explicit data to evaluate Marine Protected Areas for abalone in California. *Conservation Biology* **16:** 1308–1317.

Roodt-Wilding R and Bester AE (2006) The forensic fight against Perlemoen poaching. *Fisheries News* **31:** 214–215.

Roques S, Sévigny J-M, Bernatchez L (2002) Genetic structure of deep-water redfish, *Sebastes mentella*, populations across the North Atlantic. *Marine Biology* **140:** 297–307.

Seki T and Taniguchi K (2000). Rehabilitation of northern Japanese abalone, *Haliotis discus hannai*, populations by transplanting juveniles. In: Workshop on Rebuilding Abalone Stocks in British Columbia (ed. Campbell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 72–83.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47:** 264–279.

Sweijd NA, Bowie RCK, Lopata AL, Marinaki AM, Harley E, Cook, PA (1998) A PCR technique for forensic, species-level identification of abalone tissue. *Journal of Shellfish Research* **17:** 889–895.

Tarr RJQ (1992) The abalone fishery of South Africa. In: *Abalone of the world* (ed. Shepherd SA, Tegner MJ and Guzmán del Próo SA), Fishing News Books, Cambridge, pp438–447.

Tarr RJQ (2000) The South African abalone (*Haliotis midae*) fishery: A decade of challenges and change. In: Workshop on Rebuilding Abalone Stocks in British Colombia (ed. Campbell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 32–40.

Tegner MJ (2000). Abalone (*Haliotis* spp.) enhancement in California: what we've learned and where we go from here. In: Workshop on Rebuilding Abalone Stocks in British Columbia (ed. Campbell A), *Canadian Special Publication of Fisheries and Aquatic Sciences* **130:** 61–71.

Teske PR, Froneman CD, Barker NP, McQuaid CD (2007) Phylogeographic structure of the caridean shrimp *Palaemon peringueyi* in South Africa: further evidence for intraspecific genetic units associated with marine biogeographic provinces. *African Journal of Marine Science* **29:** 253–258.

Teske PR, McQuaid CD, Froneman CD, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology Progress Series* **314:** 283–293.

Waters JM, King TM, O'Louglin PM, Spencer HG (2005) Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology* **14:** 2789–2802.

Widmer A and Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology and Evolution* **16:** 267–269.

Wilson AB (2006) Genetic signature of recent glaciation on populations of a near-shore marine fish species (*Syngnathus leptorhynchus*). *Molecular Ecology* **15:** 1857–1871.

Wirth T and Bernatchez L (2001) Genetic evidence against panmixia in the European eel. *Nature* **409:** 1037–1040.

Zardi GI, McQuaid CD, Teske PR, Barker NP (2007) Unexpected genetic structure of indigenous (*Perna perna*) and invasive (*Mytilus galloprovincialis*) mussel populations in South Africa. *Marine Ecology Progress Series* **337:** 135–144.

Appendices

Appendix 1

Molecular Ecology Notes (2004) **4**, 618–619 doi: 10.1111/j.1471-8286.2004.00755.x

PRIMER NOTE

Isolation and characterization of microsatellite markers in the South African abalone (*Haliotis midae*)

ALETTA E. BESTER, RUHAN SLABBERT and MARÍA EUGENIA D'AMATO
Department of Genetics, University of Stellenbosch, Private BagXI, 7602 Matieland, Stellenbosch, South Africa

Abstract

We report the isolation and characterization of 11 polymorphic microsatellite loci in the South African abalone *Haliotis midae*. These loci showed a range of five to 21 alleles per locus and observed heterozygosities ranging from 0.14 to 0.93 in a wild population of 32 individuals. All loci except four conformed to Hardy–Weinberg expectations and did not show linkage disequilibrium. The polymorphism exhibited at these loci indicate that they would be useful in determining levels of genetic variability in natural and commercial *Haliotis midae* populations as well as in parentage and Quantitative Trait Loci (QTL) analysis in hatchery reared abalone.

Keywords: abalone, *Haliotis*, microsatellites

Received 18 May 2004; revision received 23 June 2004; accepted 23 June 2004

Haliotidae (Gastropoda) includes about 90 species of abalone distributed in coastal waters of all continents, of which 15 are subjected to commercial exploitation in Australia, Japan, South Africa, North America and Southeast Asia. In South Africa, *Haliotis midae* is the only one of the six South African species with commercial potential.

A combined effect of over-exploitation and illegal harvesting has caused the decline of this species, to the extent that local extinctions have been forecasted to take place in the next 5 years (http://eces.org/articles/000263.php). Research has been conducted to evaluate the feasibility of stock enhancement or ranching of *Haliotis midae* in South Africa (Sweijd *et al*. 1998; De Waal *et al*. 2003). Both restocking programmes and genetic management in aquaculture practices require the application of molecular markers. Here we report a set of 11 polymorphic microsatellites used to investigate the genetic variability of wild and commercial *Haliotis midae*.

Genomic DNA was isolated from mantle tissue following a standard CTAB extraction method (Saghai Maroof *et al*. 1984). Tissue was homogenized in 700 μL of CTAB lysis buffer containing 0.5 mg/mL Proteinase K and incubated at 60 °C. Following phenol-chloroform: isoamyl alcohol (25 : 24 : 1) extractions, the supernatant was precipitated with two volumes of 100% cold ethanol. DNA was redissolved in 100 μL of distilled water and stored at –20 °C.

In this study, microsatellite repeat sequences were isolated using an enrichment technique (FIASCO) described by Zane *et al*. 2002. An enriched partial genomic library was constructed using DNA from a single individual. For this, 250 ng DNA was simultaneously digested with *Mse*I and ligated to *Mse*I AFLP adaptors. DNA was selectively amplified using a mixture of four adaptor specific primers (*Mse*I-N) and hybridized independently with a biotinylated $(AC)_{12}$ and a $(GATC)_6$ probe. Repeat-containing fragments were recovered by streptavidin magnetic particles and cloned into a TOPO-TA cloning vector (Invitrogen) in order to produce a highly enriched microsatellite library.

Approximately 1200 recombinant clones were obtained, of which 250 were sequenced on an ABI 3100 Automated Sequencer to verify the presence of repeat sequences. Forty-five percent of the clones contained repeat sequences but only a small number had sufficient flanking regions for primer design. Oligonucleotide primers were designed for 20 loci using the program OLIGO™. version 4.0 (National Biosciences Inc.). Using the template DNA from which these loci were isolated, amplification products of expected size were obtained for 16 of the 20 microsatellite loci isolated.

A total of 32 individuals from Black Rock, on the east coast of South Africa, were genotyped to test the polymorphism of the markers. For each primer pair, one of the primers was labelled with FAM, NED, VIC or PET dyes. All polymerase chain reactions (PCR) were conducted in a Geneamp 2700 thermo cycler (Applied Biosystems) in 10 μL reactions containing 20 ng DNA, 0.3 μM of each primer,

Correspondence: A. E. Bester. Fax: ; E-mail: aeb@sun.ac.za

Table 1 Primer sequences and characteristics of 11 *Haliotis midae* microsatellite loci

Locus	Repeat Sequence	Primer Sequence (5′–3′)	*n*	Size Range	T_a °C	H_O	H_E	Allele no.	Acc Nr
HmD14	$(CA)_{10}$	F TAAGGCAAGTGAATGTCTAG	27	142–180	60	0.67	0.76	16	AY303333
		R ATTGCAAGAATCACAACTGC							
HmD33	$(GAGT)_{12}AAGT(GAGT)_6$	F TTGAAAGTGAACCAAAATCTG	22	129–205	59	0.32	0.87	11	AY303334
		R CATGGGTACAATGTGTAAAGC							
HmD36	$(GTGA)_{14}$	F AGATCGAATGACATCAGCTTC	23	220–304	60	0.43	0.89	15	AY303335
		R CATATAGCAAGCCTGAAACC							
HmD55	$(GTGA)_{12}$	F ATCAAGATAAAACGAGGCG	32	183–211	60	0.68	0.8	9	AY303337
		R ACCACTGTGAAAACGTCCA							
HmD59	$(CA)_{15}$	F TATACTGCCATTTCCGTCTG	32	106–150	60	0.78	0.84	15	AY303338
		R TCTGTATTCTGGTCCTGTCG							
HmD60	$(CA)_{16}$	F AAGTTGTTCTCCATAAAGTCGTA	14	155–171	60	0.14	0.86	8	AY303339
		R GAAGATCCGGGTTAGAACTG							
HmD61	$(CA)_{24}$	F GATATCCAACCCCTGATCAC	28	234–298	60	0.61	0.82	11	AY303340
		R GAACATCAACATCTCCATGG							
HmD11	$(TCTG)_8$	F AGCTCAGAAAAGTGGTGTACG	30	292–352	61	0.32	0.66	5	AY303341
		R TTACCTAGCTAAAGTTGACAACG							
HmD30	$(AGTC)_2GGTC(AGTC)_{11}$	F TGATGTTGCTGGAATATTGC	27	124–150	60	0.7	0.8	11	AY303342
		R CAATTTCATTTTCAACAGTTCA							
HmSP1	$(CA)_{10}(CGCA)_2(CA)_4$	F ATAGTGGTCATACAGTCATCACCT	21	192–276	61	0.48	0.93	21	AY303346
		R TAGGCATGTTTGAGTTCGTGT							
HmSP5	$(AC)_{13}$	F TTCGGCAAGTGAATGTCTAG	31	185–219	60	0.63	0.74	14	AY303344
		R ATGCGACACTTACTACACCG							

T_m, optimum annealing temperature; H_O, observed heterozygosity; H_E, expected heterozygosity.

200 μM dNTP's, 0.1 unit of *Taq* polymerase (Promega), 1 × PCR Buffer A (Promega) and 2 mM $MgCl_2$. PCR consisted of an initial denaturing step at 94 °C for 5 min followed by 35 cycles of 30 s at 94 °C, 30 s at 59–61 °C, and 1 min at 72 °C, and a final extension for 10 min at 70 °C. PCR products were separated on a ABI 3100 Automated Sequencer and analysed using the GENESCAN software program (Applied Biosystems).

Eleven loci showed polymorphism. PCR primer sequences, optimal annealing temperature, repeat motif and allele size ranges are shown in Table 1. Observed and expected heterozygosities and probability of Hardy–Weinberg equilibrium (PHWE) were calculated using GENETIX version 4.02 (Belkhir *et al*. 2000). Allelic number ranged from five to 21. However, more alleles are likely to be present for loci HmD14 and HmSP5, due to an apparent allele dropout effect. An extensive PCR optimization (variation in DNA, primer and $MgCl_2$ concentration), following Goossens *et al*. (1998) could not overcome the problem. Presence of null alleles is suspected in loci HmD33, HmD36, HmD60 and HmSP1 because of their departure from HWE ($P < 0.01$). All other loci conformed to HWE expectations and no linkage disequilibrium was detected. In summary, these loci exhibited high levels of polymorphism and heterozygosity and provide an invaluable tool for analysing genetic diversity of wild and commercial *Haliotis midae* populations. In addition, these markers could be used for parentage analysis and possible QTL identification in hatchery reared abalone.

Acknowledgements

We are indebted to Brian Godfrey for collecting samples of the test population in Black Rock. This work was partially funded by a Claude Harris Leon Foundation research grant.

References

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (2000) *GENETIX, logiciel sous Window TM pour la génétique des populations*. Laboratoire Génome et Populations, Montpellier, France.

De Waal SWP, Branch GM, Navarro R (2003) Interpreting evidence of dispersal by *Haliotis midae* juveniles seeded in the wild. *Aquaculture*, **221**, 299–310.

Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology*, **7**, 1237–1241.

Saghai Maroof MA, Solima KM, Jorgenson RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences (USA)*, **81**, 8014–8018.

Sweijd N, Snethlage Q, Harvey D, Cook P (1998) Experimental abalone (*H. midae*) seeding in South Africa. *Journal of Shellfish Research*, **17**, 897–904.

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

Appendix 2

doi:10.1111/j.1365-2052.2008.01728.x

Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae*: a targeted EST approach

A. E. Bester, R. Roodt-Wilding and H. A. Whitaker

Animal Molecular Genetics Laboratory, Department of Genetics, Stellenbosch University, Stellenbosch, South Africa

Summary

In this study, we describe the first set of SNP markers for the South African abalone, *Haliotis midae*. A cDNA library was constructed from which ESTs were selected for the screening of SNPs. The observed frequency of SNPs in this species was estimated at one every 185 bp. When characterized in wild-caught abalone, the minor allele frequencies and F_{ST} estimates for every SNP indicated that these markers may potentially be useful for population analysis, parentage assignment and linkage mapping in *Haliotis midae*. No linkage disequilibrium was observed between SNPs originating from different EST sequences. These SNPs, together with additional SNPs currently being developed, will provide a useful complementary set of markers to the currently available genetic markers in abalone.

Keywords expressed sequence tag, *Haliotis midae*, single nucleotide polymorphism.

A commercially valuable species of abalone, *Haliotis midae*, inhabits the rocky shores of South Africa and is one of over 50 *Haliotis* species found worldwide. The increasing demand and overexploitation of this marine mollusc necessitated the development and implementation of genetic management practices in natural as well as commercial stocks. The population genetic status of *H. midae* has so far been investigated using allozymes, mtDNA sequence data and non-species-specific microsatellite markers (Evans *et al.* 2004). Although the isolation of species-specific microsatellites (Bester *et al.* 2004; Slabbert *et al.* 2008) promises a finer-scale inference of *H. midae* stock structure, SNPs are also considered for population genetic analysis in view of the potential limitations of microsatellite loci. Morin *et al.* (2004) reviews a number of advantages of SNPs over other marker types for the inference of population structure. Unlike microsatellite loci, SNPs occur more frequently in the genome, are extremely stable due to low mutation rates (Sachidanandam *et al.* 2001) and most importantly, SNP alleles are almost exclusively identical-by-decent (IBD), preventing scoring errors associated with homoplasy. From a technical perspective, the identification of SNPs can be made simple and inexpensive using high-throughput genotyping technology.

Address for correspondence

A. E. Bester, Animal Molecular Genetics Laboratory, Department of Genetics, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa.
E-mail: aeb@sun.ac.za

Accepted for publication 14 February 2008

Given the relative paucity of functional sequence data from *H. midae*, the need arose to create and sequence a cDNA library for this species. SNPs identified within these sequences can potentially be used as type I markers in population genetic analysis and linkage mapping of *H. midae*. In this study, we describe the detection of 20 SNP loci and the further characterization of 12 of them using a targeted gene approach. A collaborative effort to improve the genetic status of *H. midae*, while conserving natural stocks, is expected to benefit greatly from this marker information.

For the construction of a cDNA library, material was obtained from farmed abalone and was snap-frozen in liquid nitrogen. Total RNA was extracted from 200 mg gill tissue by homogenization with one-step RNA Reagent (Bio Basic Inc.). Approximately 150 μg of RNA was purified using a polyA Spin mRNA isolation kit (New England Biolabs) and 2 μg mRNA was used in cDNA library construction (Promega Universal Riboclone® cDNA Synthesis System). cDNA inserts were cloned into a pGEM® 4Z vector (Promega), and this was followed by plasmid miniprep DNA isolation (QIAprep Spin Miniprep kit; QIAGEN) and sequencing in at least one direction using vector-specific primers (M13). Vector and adaptor sequences were removed using CLC BIO GENE WORKBENCH SOFTWARE, and redundant ESTs and sequence reads were combined into contiguous sequences. Contiguous and unique single sequences were searched for similarity to known genes and proteins using the BLASTX and nucleotide MEGABLAST functions at NCBI.

Sequences from 18 clones were assembled into a single contiguous sequence that showed strong similarity to *Haliotis discus* ribosomal RNA. Thirty-nine clones formed 19 further contiguous sequences, whilst 53 clones gave unique

sequences or were 'singletons'. Following BLAST analysis, 26% of sequences showed no significant similarity to any known DNA or protein sequence, whilst 29% showed homology to known functional genes. A surprisingly high number of sequences (21%) were similar to microsatellite sequence submissions from other *Haliotis* species, whilst the remainder showed similarity to uncharacterized mRNA or genomic DNA sequence. All sequences are available in GenBank (EX534436–EX534505; EU135914–EU135918).

Six *H. midae* EST clones that showed significant similarity to known proteins (*E*-value ≤ 1e−5) were selected for the design of PCR primers using PRIMER3 software (Rozen & Skaletsky 2000). cDNA fragments of approximately 300 bp were targeted for amplification and care was taken to avoid possible exon/intron border sequences (AG/GT; He *et al.* 2003) when positioning primers. Genomic DNA was amplified from a test panel of eight *H. midae* individuals, selected from disparate sampling sites. Each PCR reaction contained 20 ng template genomic DNA, 200 μM dNTPs, 10 pmol primers, 1.5 mM $MgCl_2$ and 2 U GoTaq® Flexi DNA polymerase. Amplification was performed in a GeneAmp System 2700 thermal cycler (Applied Biosystems) as follows: an initial denaturation of 10 min at 95 °C; 30 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min; and a final elongation step of 10 min at 72 °C. This was followed by agarose gel electrophoresis and SigmaSpin post-reaction purification. Sequencing was performed in both directions using the ABI PRISM BigDye Terminator version 3.1 Cycle Sequencing kit and the 3100 Genetic DNA Analyser (Applied Biosystems).

Alignment of sequences was carried out in BIOEDIT, and sites that appeared to exhibit sequence variation, in either heterozygous or homozygous forms, were evaluated by manual inspection of the chromatograms. Heterozygous individuals were detected as multiple peaks at the same sequence position and were considered as putative SNPs when both alleles were observed at a frequency of greater than 10%. In a second validation step, internal primer sets were developed for the larger fragments to confirm the presence of SNPs in eight more individuals. The five internal primer sets were optimized using the same PCR amplification conditions as before, varying only the annealing temperature. For further characterization, 12 of the SNP loci were genotyped in 160 wild *H. midae* collected from five geographically distant locations. PCR conditions, automated sequencing and visual identification of SNPs were the same as described above. GENEPOP version 3.2 (Raymond & Rousset 1997) was used to test for departure from Hardy–Weinberg equilibrium while allelic frequencies, F_{ST} (Weir & Cockerham 1984) and observed and expected heterozygosities were calculated with GENETIX version 4.02 (Belkhir *et al.* 2000). To identify possible association among SNPs from different EST clones, genotypic disequilibrium was assessed according to Fisher's method (GENEPOP) based on 1000 permutations.

Genomic DNA fragments ranging from 240 to 1.2 kb were successfully amplified from the selected cDNA sequences (Table 1). Although alignments with BIOEDIT showed several potential sequence variations, fewer were considered as potential SNPs after closer inspection of the chromatograms

Table 1 The origin and positioning of 20 SNPs detected in *Haliotis midae*.

EST clone	Homology	GenBank accession	Primer sequences (5′–3′)	Product size (bp)	SNP identity	Base position
1A1	*Perlucin*	EU135915	TTTGTTAGCCTCGGTCCATC	1250	A>C	122
			CGATCACAGGGGACATCATT		G>A	148
					G>T	155
C12	*Cellulase*	EU135914	ATTTTTGTCGGTCACCTGGA	300	T>C	225
			GTAGGGCTTCCCAGAAGGAC		T>A	260
3B4	*Ribosomal protein L8e*	EU135916	GAAAGCCATTCCCTCAACAA	1230	T>A	57
			TGCTTTGTACCACGAAGACG		C>T	148
					A>T	272
					C>T	291
					C>G	320
					A>G	325
					T>A	492
					T>C	561
					T>A	563
3D10	*Haemocyanin*	EU135917	TGAGTGCGGCGTAAAAATAA	240	G>A	122
			AGTGGAACGCCAAATGTTGT			
2C3	*Fibroblast growth factor receptor*	EU135918	GGTGGAAACCCTTACGACAA	375	A>G	114
			AACCACGTTCGCTGTGGTAT		T>A	145
					A>C	280
2H9	*Ribosomal protein L22*	EX534490	TCCGACCTTCTTGATCTTGG	300	A>T	116
			AGGAAATGCTGGTGGACAAC		A>T	177

Table 2 Characteristics of 12 SNP loci derived from *Haliotis midae* EST clones.

EST clone	SNP	Internal primer sequences (5′–3′)	T_A (°C)	Minor allele frequency	*N*	H_O	H_E[1]	F_{ST}[2]
1A1	A>C	TTTCATGTTTTGCATCAAAC	57	C: 0.16	156	0.27	0.27	0.009
	G>A	AAGAAGGAACTGTATGGCTG		A: 0.47	156	0.95	0.50***	0.011
	G>T			T: 0.18	156	0.31	0.29	0.007
C12	T>C	ATTTTTGTCGGTCACCTGGA	55	C: 0.07	138	0.07	0.14**	0.008
	T>A	GTAGGGCTTCCCAGAAGGAC		A: 0.13	138	0.14	0.22**	0.009
3B4-1	C>T	AAACATCTGCAACATTTAGG	57	T: 0.47	153	0.52	0.48*	0.007
	A>T	GACAGCAAAACAAACATCAG		T: 0.43	155	0.40	0.47	0.008
	C>T			C: 0.49	155	0.40	0.49*	0.011
	A>G			G: 0.47	155	0.46	0.48	0.009
3B4-2	T>A	TAAGAATCCACAAGTTGGTG	57	A: 0.25	157	0.37	0.37	0.010
	T>C	ATGTATCATCACGGACAGG		C: 0.34	157	0.47	0.45	0.011
	T>A			A: 0.20	157	0.29	0.31	0.010

[1]Departure from HWE, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.
[2]Wright's *F*-statistics.

and allelic states. It is a common occurrence that some polymorphisms turn out to be artefacts of sequencing and therefore similar assessment criteria to those described in Primmer *et al.* (2002) were followed to minimize such data. The 20 candidate SNPs observed within the total of 3.7 kb DNA represented a rate of approximately one SNP every 185 bp. Depending on the species and genomic region targeted, SNPs generally occur every 200–500 bp. After verification with the five internal primer sets, the rate observed in *H. midae* was comparable to that found in other EST sequences (Fahrenkrug *et al.* 2002; Feau *et al.* 2007). Polymorphisms were present as a 1:1.5 ratio of transitions:transversions, which is slightly lower than what was expected. Wang *et al.* (1998) found that in the human genome, transitions occur more frequently than transversions due to the highly mutable site within CpG dinucleotides.

When characterized further, SNPs showed moderate levels of heterozygosity with frequencies of the minor allele ranging from 7% to 49% (Table 2). Five of the SNPs deviated significantly from Hardy–Weinberg proportions ($P < 0.05$). No significant linkage disequilibrium was apparent among SNPs derived from different EST sequences. In this study, the total amount of polymorphic information obtained from the SNPs was restricted to the number of EST clones screened. To increase chances of detecting unique EST-derived SNPs, one suggestion is to screen rapidly evolving gene classes (Nielsen *et al.* 2005) rather than conserved sequences. Alternatively, a higher number of EST sequences should be screened while aiming to isolate only one SNP per exonic region. Because of the lower information content expected of SNP markers in general, more SNPs must be accumulated in order to provide a powerful complementary set of markers to the available microsatellites in *H. midae.* Nonetheless, the SNPs developed here and in the future will be well suited for stock structure analysis, parentage assignment, linkage mapping and QTL mapping of both wild and cultured abalone. To our knowledge this is the first series of SNPs reported for the genus *Haliotis.*

Acknowledgements

We thank Belinda Swart and Lise Sandenbergh for excellent assistance with the SNP genotyping. This research was funded by the National Research Foundation (NRF) of South Africa and the Innovation Fund, Department of Science and Technology, South Africa.

References

Belkhir K., Borsa P., Goudet J., Chikhi L. & Bonhomme F. (2000) *GENETIX, logiciel sous Window TM pour la génétique des populations*. Laboratoire Génome et Populations, Montpellier, France.

Bester A.E., Slabbert R. & D'Amato M.E. (2004) Isolation and characterization of microsatellite markers in the South African abalone (*Haliotis midae*). *Molecular Ecology Notes* **4,** 618–9.

Evans B.S., Sweijd N.A., Bowie R.C.K., Cook P.A. & Elliott N.G. (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology–Progress Series* **270,** 163–72.

Fahrenkrug S.C., Freking B.A., Smith T.P., Rohrer G.A. & Keele J.W. (2002) Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Animal Genetics* **33,** 186–95.

Feau N., Bergeron M.-J., Joly D.L., Roussel F. & Hamelin R.C. (2007) Detection and validation of EST-derived SNPs for poplar leaf rust *Melampsora medusae* f. sp. *deltoidae*. *Molecular Ecology Notes* **7,** 1222–8.

He C., Chen L., Simmons M., Li P., Kim S. & Liu Z.J. (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics* **34,** 445–8.

Morin P.A., Luikart G., Wayne R.K. & The SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19,** 208–16.

Nielsen R., Bustamante C., Clark A.G. *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3,** e170.

Primmer C.R., Borge T., Lindell J. & Saetre G.P. (2002) Single nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology* **11,** 603–12.

Raymond M. & Rousset F. (1997) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

Rozen S. & Skaletsky H.J. (2000) PRIMER3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Ed. by S. Krawetz & S. Misener), pp. 365–86. Humana Press, Totowa, NJ.

Sachidanandam R., Weissman D., Schmidt S.C. *et al.* (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409,** 928–33.

Slabbert R., Ruivo N.R., Van den Berg N.C., Lizamore D.L. & Roodt-Wilding R. (2008) Isolation and characterisation of 63 microsatellite loci for the abalone, *Haliotis midae*. *Journal of the World Aquaculture Society* **39,** 429–35.

Wang D.G., Fan J.B., Siao C.J. *et al.* (1998) Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280,** 1077–82.

Weir B.S. & Cockerham C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38,** 1358–70.

Appendix 3

Summary of genetic variation at eight microsatellite loci at nine *Haliotis midae* sampling sites

		Loci								
Sample		*HmD14*	*HmD36*	*HmD55*	*HmD59*	*HmD11*	*HmSP5*	*HmAD102*	*CmrHr2.15*	*Mean*
Saldanha	*Ho*	0.542	0.391	0.745	0.917	0.383	0.563	0.378	0.383	0.538
n=48	*He*	0.554	0.901	0.869	0.92	0.616	0.556	0.942	0.802	0.770
	Na	14	24	20	21	8	14	26	13	
	f	0.033	0.573*	0.154*	0.014	0.387*	-0.002	0.606*	0.53*	0.312*
Robben Island	*Ho*	0.608	0.298	0.791	0.844	0.326	0.609	0.575	0.425	0.560
n=48	*He*	0.586	0.897	0.899	0.887	0.68	0.585	0.948	0.862	0.793
	Na	15	21	15	19	8	13	27	18	
	f	-0.028	0.674*	0.133*	0.06	0.53*	-0.029	0.403*	0.514*	0.305*
Kleinmond	*Ho*	0.553	0.349	0.636	0.872	0.467	0.553	0.511	0.267	0.526
n=48	*He*	0.518	0.924	0.884	0.905	0.504	0.494	0.798	0.729	0.720
	Na	13	23	17	19	7	12	23	12	
	f	-0.057	0.63*	0.291*	0.047	0.085	-0.11	0.369*	0.641*	0.279*
Gansbaai	*Ho*	0.563	0.222	0.766	0.813	0.447	0.583	0.565	0.539	0.562
n=48	*He*	0.642	0.925	0.861	0.91	0.562	0.657	0.945	0.844	0.793
	Na	13	24	14	19	8	15	30	15	
	f	0.134	0.764*	0.121	0.117*	0.215*	0.123	0.411*	0.373*	0.302*
Witsand	*Ho*	0.546	0.455	0.727	0.909	0.333	0.546	0.25	0.667	0.554
n=33	*He*	0.585	0.941	0.873	0.905	0.594	0.586	0.945	0.705	0.767
	Na	18	23	15	15	6	19	22	12	
	f	0.084	0.528*	0.182*	0.011	0.451*	0.085	0.743*	0.069	0.292*
Mossel Bay	*Ho*	0.472	0.373	0.68	0.765	0.389	0.434	0.408	0.571	0.512
n=54	*He*	0.498	0.923	0.847	0.859	0.674	0.504	0.948	0.726	0.747
	Na	18	19	15	17	10	16	27	11	
	f	0.053	0.603*	0.207*	0.119	0.431*	0.148	0.576*	0.223*	0.324*
Cape Recife	*Ho*	0.383	0.458	0.848	0.87	0.255	0.5	0.375	0.542	0.529
n=48	*He*	0.457	0.892	0.857	0.857	0.635	0.572	0.897	0.643	0.726
	Na	14	23	19	17	7	18	19	13	
	f	0.164	0.494*	0.022	-0.003	0.605*	0.136	0.589*	0.168	0.281*
Riet Point	*Ho*	0.692	0.36	0.62	0.787	0.255	0.75	0.442	0.596	0.563
n=53	*He*	0.685	0.911	0.827	0.881	0.62	0.688	0.935	0.671	0.777
	Na	19	25	13	21	7	19	25	10	
	f	-0.01	0.611*	0.26*	0.117	0.595*	-0.08	0.534*	0.121	0.284*
Black Rock	*Ho*	0.542	0.409	0.711	0.851	0.296	0.625	0.565	0.489	0.561
n=48	*He*	0.686	0.935	0.84	0.867	0.6511	0.688	0.953	0.722	0.793
	Na	18	25	11	19	6	19	29	14	
	f	0.212*	0.57*	0.165	0.029	0.554*	0.103	0.416*	0.332*	0.302*
Overall samples										
	Mean Ho	0.55	0.368	0.725	0.848	0.35	0.574	0.452	0.498	0.546
	Mean He	0.583	0.933	0.874	0.898	0.625	0.599	0.939	0.756	0.776
	Total Na	29	43	30	27	12	29	43	24	
	Mean *f*	0.046	0.605	0.171	0.057	0.428	0.042	0.519	0.330	
	PHWE	0.045	<0.001	0.001	0.038	<0.001	0.248	<0.001	<0.001	<0.001
	ETGD	<0.001	<0.001	0.004	0.001	<0.001	0.002	<0.001	<0.001	<0.001

Ho = observed heterozygosity; *He* = expected heterozygosity; *Na* = number of alleles; *f* = Fis of Weir and Cockerham, 1984; **f* = Fis significant at the 5% level; PHWE = probability of Hardy-Weinberg equilibrium; ETGD = Exact Test for Genic Differentiation.

Appendix 4

Summary of genetic variation at 12 SNP loci at nine *Haliotis midae* sampling sites

		Samples									
Loci		*SD*	*RI*	*KL*	*GB*	*WS*	*MB*	*CR*	*RP*	*BR*	*Mean Overall Samples*
SNP-1	*Ho*	0.310	0.094	0.313	0.281	0.563	0.355	0.387	0.375	0.281	0.329
	He	0.262	0.144	0.305	0.285	0.430	0.331	0.312	0.404	0.285	0.306
	Minor allele freq	0.155	0.078	0.188	0.172	0.313	0.210	0.194	0.281	0.172	0.196
	f	-0.167	0.363	-0.010	0.028	-0.295	-0.054	-0.224	0.088	0.028	-0.027
*SNP-2**	*Ho*	1.000	0.938	0.969	0.750	0.875	0.903	0.581	0.875	0.938	0.870
	He	0.500	0.498	0.500	0.500	0.492	0.495	0.412	0.492	0.498	0.487
	Minor allele freq	0.500	0.469	0.484	0.500	0.438	0.452	0.290	0.438	0.469	0.449
	f	-1.000	-0.879	-0.938	-0.488	-0.771	-0.818	-0.395	-0.771	-0.879	-0.771
SNP-3	*Ho*	0.448	0.188	0.094	0.188	0.500	0.452	0.419	0.375	0.344	0.334
	He	0.348	0.219	0.144	0.219	0.430	0.412	0.367	0.404	0.324	0.318
	Minor allele freq	0.224	0.125	0.078	0.125	0.313	0.290	0.242	0.281	0.203	0.209
	f	-0.273	0.158	0.363	0.158	-0.148	-0.080	-0.127	0.088	-0.046	0.010
*SNP-4**	*Ho*	0.185	0.000	0.000	0.125	0.031	0.107	0.000	0.000	0.077	0.058
	He	0.226	0.074	0.062	0.305	0.418	0.270	0.170	0.430	0.074	0.225
	Minor allele freq	0.130	0.039	0.032	0.188	0.297	0.161	0.094	0.313	0.039	0.143
	f	0.198	1.000	1.000	0.600	0.927	0.614	1.000	1.000	-0.020	0.702
*SNP-5**	*Ho*	0.259	0.000	0.194	0.219	0.000	0.107	0.000	0.194	0.115	0.121
	He	0.324	0.074	0.271	0.476	0.375	0.270	0.350	0.500	0.174	0.313
	Minor allele freq	0.204	0.039	0.161	0.391	0.250	0.161	0.226	0.516	0.096	0.227
	f	0.219	1.000	0.300	0.552	1.000	0.614	1.000	0.623	0.353	0.629
SNP-6	*Ho*	0.469	0.419	0.867	0.613	0.500	0.393	0.469	0.438	0.438	0.512
	He	0.488	0.492	0.498	0.481	0.469	0.484	0.500	0.498	0.451	0.484
	Minor allele freq	0.578	0.565	0.467	0.597	0.375	0.411	0.484	0.469	0.344	0.477
	f	0.055	0.163	-0.733	-0.258	-0.051	0.206	0.077	0.137	0.046	-0.040
SNP-7	*Ho*	0.406	0.419	0.500	0.563	0.500	0.393	0.500	0.438	0.281	0.444
	He	0.488	0.492	0.469	0.469	0.469	0.484	0.500	0.498	0.441	0.479
	Minor allele freq	0.422	0.436	0.375	0.375	0.625	0.589	0.500	0.531	0.328	0.465
	f	0.183	0.163	-0.051	-0.185	-0.051	0.206	0.016	0.137	0.376	0.088
SNP-8	*Ho*	0.375	0.419	0.563	0.594	0.500	0.393	0.500	0.438	0.250	0.448
	He	0.500	0.492	0.482	0.476	0.469	0.494	0.500	0.498	0.492	0.489
	Minor allele freq	0.500	0.436	0.406	0.391	0.625	0.554	0.500	0.531	0.563	0.501
	f	0.265	0.163	-0.151	-0.232	-0.051	0.223	0.016	0.137	0.504	0.097
SNP-9	*Ho*	0.469	0.419	0.594	0.594	0.500	0.393	0.469	0.438	0.406	0.476
	He	0.496	0.492	0.488	0.496	0.469	0.469	0.500	0.498	0.476	0.487
	Minor allele freq	0.453	0.436	0.422	0.453	0.625	0.625	0.516	0.531	0.391	0.495
	f	0.070	0.163	-0.202	-0.183	-0.051	0.180	0.077	0.137	0.162	0.039
SNP-10	*Ho*	0.375	0.323	0.563	0.344	0.438	0.333	0.429	0.323	0.281	0.379
	He	0.404	0.383	0.430	0.285	0.375	0.391	0.337	0.312	0.242	0.351
	Minor allele freq	0.281	0.258	0.313	0.172	0.250	0.267	0.214	0.194	0.141	0.232
	f	0.088	0.174	-0.295	-0.192	-0.151	0.164	-0.256	-0.017	-0.148	-0.070
SNP-11	*Ho*	0.438	0.419	0.594	0.344	0.500	0.533	0.429	0.387	0.375	0.447
	He	0.451	0.425	0.441	0.285	0.492	0.480	0.436	0.412	0.430	0.428
	Minor allele freq	0.344	0.307	0.328	0.172	0.438	0.400	0.321	0.290	0.313	0.324
	f	0.046	0.030	-0.333	-0.192	0.000	-0.094	0.036	0.077	0.143	-0.032
SNP-12	*Ho*	0.281	0.323	0.344	0.344	0.375	0.267	0.429	0.258	0.250	0.319
	He	0.324	0.383	0.324	0.285	0.342	0.320	0.337	0.271	0.219	0.311
	Minor allele freq	0.203	0.258	0.203	0.172	0.219	0.200	0.214	0.161	0.125	0.195
	f	0.147	0.174	-0.046	-0.192	-0.081	0.183	-0.256	0.063	-0.127	-0.015

Overall loci										
	Mean Ho	0.418	0.330	0.466	0.413	0.440	0.386	0.384	0.378	0.336
	Mean He	0.401	0.347	0.368	0.380	0.436	0.408	0.393	0.435	0.342
	Mean minor allele freq	0.333	0.287	0.288	0.309	0.397	0.360	0.316	0.378	0.265
	f	-0.014	0.223	-0.091	-0.049	0.023	0.112	0.080	0.142	0.033

Departure from HWE, **P*< 0.05***;*** *Ho* = observed heterozygosity; *He* = expected heterozygosity; *f* = Fis of Weir and Cockerham (1984)