View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Stellenbosch University SUNScholar Repository

Development of Gene-linked Molecular Markers in South African Abalone (*Haliotis midae*) Using an *In silico* Mining Approach

by

Clint Rhode

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

Pectora roborant cultus recti

Supervisor: Dr. R. Roodt-Wilding

Department of Genetics

March 2010

Declaration

By submitting this dissertation 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.

March 2010

Copyright © 2010 Stellenbosch University.

All Rights Reserved

“*But then with me the horrid doubt always arises whether the convictions of man’s mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy.*”

- Charles Darwin, 3 July 1881 (in a letter to William Graham) -

Abstract

The South African abalone, *Haliotis midae*, is the only endemic species of commercial value. Aquaculture remains the only avenue for expanding the industry, since the closure of the fishery. The current focus is on implementing a molecular breeding programme; thus the development of molecular markers for linkage mapping and QTL analysis is a priority. Various markers, mainly anonymous, have been developed for *H. midae*; however emphasis is being placed on the development of gene-linked type I molecular markers. The present study investigates and demonstrates the use of public sequence collections to develop type I markers for a species with limited genomic resources, via three strategies: Surveying anonymous *H. midae* microsatellite markers' flanking regions to find homology to gene sequences in public databases, cross-species marker transfer of anonymous markers from *H. rubra* and *H. discus hannai* demonstrating putative gene associations and lastly EST marker mining (SNP and microsatellites) from various Haliotids and testing transfer to the target species. Approximately 17% of *H. midae* anonymous markers showed significant similarity to genes. The current study also reports higher cross-species transferability from both *H. rubra* and *H. discus hannai* to *H. midae* (39% and 20.5%, respectively) than previously demonstrated and 15 EST-microsatellites and 16 EST-SNPs were successfully mined. Furthermore, the non-random distribution of microsatellites and high nucleotide diversity in the *H. midae* genome was confirmed. This is a low cost and time effective method for marker development and presents a continuous and dynamic resource that could be used for future marker development and characterisation as sequence information in public databases grow exponentially.

Opsomming

Die Suid-Afrikaanse perlemoen, *Haliotis midae*, is die enigste van vyf inheemse spesies van kommersiële waarde. Na die noodgedwonge sluiting van die vissery, is akwakultuur die mees praktiese oplossing om die perlemoen industrie uit te brei. Die huidige fokus is gerig op die implementering van ‘n molekulêre teel-program en dus is die ontwikkeling van molekulêre merkers vir genetiese kartering en kwantitatiewe kenmerk lokus analise, van uiterste belang. Tipe II merkers is voorheen vir die perlemoen ontwikkel, maar huidige tendense lê klem op die ontwikkeling van geen-gekoppelde tipe I merkers. Die huidige studie ondersoek die gebruik van publieke databasisse vir die ontwikkeling van tipe I molekulêre merkers vir ‘n spesie met beperkte genomiese bronne. Drie strategieë is geïmplementeer: Eerstens is ‘n opname gemaak van die homologie van perlemoen tipe II merker-vleuelende volgordes met geen volgordes in databasisse. Verder is die oordraagbaarheid van tipe II merkers vanaf *H. rubra* en *H. discus hannai* wat assosiasie met gene toon ondersoek. Laastens is ‘n Uitgedrukte Volgorde Merk (UVM) (Expressed Sequence Tag, EST) merker-ontginnings metode vanaf verskeie *Haliotis* spesies en toetsing van oordraagbaarheid na die teiken spesie uitgevoer. Ongeveer 17% van die tipe II *H. midae* merkers het geniese assosiasie getoon. ‘n Hoër tussen-spesie oordraagbaarheid vanaf beide *H. rubra* en *H. discus hannai* na *H. midae* (39% en 20.5%, onderskeidelik) word gerapporteer in vergelyking met vorige studies en 15 UVM-mikrosatelliete en 16 UVM-enkel nukleotied polimorfismes (single nucleotide polimorphism, SNP) is ontwikkel. Verder bevestig die studie die nie-lukrake verspreiding van mikrosatelliete en hoë nukleotied diversiteit in die perlemoen genoom. Die gebruik van publieke databasise vir die ontwikkeling en karakterisering van tipe I molekulêre merkers is tyd- en koste-besparend en bied ‘n volgehoue en dinamiese bron vir toekomstige gebruik.

Acknowledgements

Hereby I would like to thank the following persons (in alphabetical order) for technical assistance, advice, guidance and general support during the course of this project:

Van der Merwe, A
Du Plessis, N
Hepple, J
Jansen, S
Roodt-Wilding, R
Sandenbergh, L
Slabbert, R
Swart, B

The following institutions (in alphabetical order) are thanked for biological materials and financial support:

Abagold (Pty) Ltd
Aquafarm (Pty) Ltd
HIK Abalone Farm (Pty) Ltd
Irvin and Johnson Ltd
Innovation Fund
National Research Foundation (South Africa)
Roman Bay Sea Farm (Pty) Ltd
Stellenbosch University

Table of Contents

List of Figures

List of Tables

List of Abbreviations

%	Percentage
(Pty) Ltd	Property Limited
[ ]	Concentration
<	Less than
>	Greater than
®	Registered Trademark
µl	Microlitre
µM	Micromole
3’	Three prime
5’	Five prime
A	Adenine
A>G	Adenine to Guanine substitution
A>T	Adenine to Thymine substitution
AFLP	Amplified Fragment Length Polymorphism
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
C	Cytosine
cDNA	complimentary DNA
C>G	Cytosine to Guanine substitution
C>T	Cytosine to Thymine substitution
CATS	Comparative Anchor Tagged Sequences
d.$H_2$0	Distilled Water
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediamine Tetra-Acetic Acid ($C_{10}H_{16}N_2O_8$)
EST	Expressed Sequence Tag
F	Forward Primer
g	Grams
G	Guanine
G>T	Guanine to Thymine substitution
gDNA	genomic Deoxyribonucleic Acid

GenBank Acc.#	GenBank Accession Number at www.ncbi.nih.nlm.gov
HWE	Hardy-Weinberg Equilibrium
Inc.	Incorporated
L	Litre
LD	Linkage Disequilibrium
LE	Linkage Equilibrium
M	Molar (Moles per Litre)
MAS	Marker Assisted Selection
mg/ml	Milligram per Millilitre
$MgCl_2$	Magnesium Chloride
min	Minutes
ml	Millilitre
mM	Millimole
MYA	Million Years Ago
N/A	Not Applicable
ng	Nanograms
ng/ml	Nanogram per Millilitre
ng/µl	Nanogram per Microlitre
^{o}C	Degrees Celsius
p	Probability value (As a statistically significant limit)
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
pmol	Picomol
pp.	Pages
QTL	Quantitative Trait Locus
R	Reverse Primer
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
sec	Seconds
SNP	Single Nucleotide Polymorphism
SSCP	Single Strand Conformational Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
T	Thymine

T>A	Thymine to Adenine substitution
T>C	Thymine to Cytosine substitution
Taq	*Thermus aquaticus* DNA Polymerase
TBE	Tris-Borate-EDTA Buffer
T_m	Melting Temperature
™	Trademark
U	Units (enzyme)
UTR	Untranslated Region
v/v	Volume per Volume
w/v	Weight per Volume
ZAR	South African Rand (monetary currency)

Chapter I:
Introduction

1. An Overview

As a member of the gastropod class in the phylum *Mollusca*, the *Haliotidae* family is vast; with a variety of species distributed worldwide along primarily cool to temperate but also tropical waters off the coastlines of all the continents, with the exception of Antarctica (Geiger 2000). South Africa, in particular, has five endemic species, occurring across the country's sea border. Of these only *Haliotis midae* was found suitable for commercial harvest; mostly due to its relative abundance, compared to the other species, and its large growing nature (Roodt-Wilding and Slabbert 2006). Internationally the abalone industry contributed 629,842,000.00 US dollars to the world economy, corresponding to 14 400 metric tons of abalone, in 2006 (United Nations Food and Agriculture Organisation, www.fao.org/fishery). As such, *H. midae* has become increasingly more vulnerable to over exploitation by commercial fisheries and poaching by illegal syndicates due to the considerable profits to be gained with abalone related commerce. With dwindling natural stocks and the eminent threat of extinction, the establishment of abalone aquaculture enterprises has become a viable alternative to wild harvest so as to satisfy the global demand for abalone products.

The advantage of abalone aquaculture is two fold: Firstly in the light of conservation and sustainable utilisation of natural resources, these enterprises will decrease the dependency on wild stocks, allowing natural populations to recover from overexploitation. The artificial environments created can also serve as the epicentre for captive breeding programs to supplement diminishing wild populations thereby fast tracking population rehabilitation (Najmudeen and Victor 2004; Theodorou and Couvet 2004; Gutierrez-Gonzalez and Perez-Enriquez 2005; Dixon *et al.* 2006; Roodt-Wilding 2007). Secondly, the balance between the needs of man and that of nature is a fine one, often leading to conflict between local communities and the authorities (Balmford *et al.* 2001; Jha and Bawa 2006). *Haliotis midae* is valued as a source of revenue; therefore any effort in the preservation of this species or to regulate its harvest will inadvertently have socio-economic ramifications. Already small scale fishermen are finding themselves without means of income due to more stringent governmental policies. Abalone aquaculture enterprises can fill the need for employment as most of the daily tasks mandatory to operate an abalone farm require no specialised skills. Therefore poorly educated labourers, generally from poverty stricken communities can be solicited for employ (Troell *et al.* 2006). Furthermore, the reseeding and stock enhancement initiatives fuelled by abalone ranching as an off-shoot from culturing practices could facilitate the recovery of wild populations and may subsequently lead to the relaxation of harsh harvesting regulations, allowing small scale fishermen to once again provide for themselves (Najmudeen and Victor 2004; Gutierrez-Gonzalez and Perez-Enriquez 2005; Dixon *et al.* 2006; Roodt-Wilding 2007).

Abalone culture is now established in several countries including: Australia, New Zealand, Japan, Taiwan, China, Ireland, Iceland, the United States and Mexico. South Africa is relatively new to the industry with first initiatives starting in the late 1980's and 1990's. Nonetheless South Africa has made great strides in the industry to the point where it is now the world's largest supplier of abalone outside of the Orient (Troell *et al.* 2006).

It is accepted that genetics forms an essential part of innovative and holistic management strategies in animal production. This has become especially true in more recent times with the application of novel technologies aided by the development of molecular/DNA markers. Considerable headway has been made in this regard in traditional animals of economic value, e.g. cattle, sheep, swine and poultry (Beuzen *et al.* 2000; Vignal *et al.* 2002; Van Marle-Köster and Nel 2003). In aquaculture species progress has been slow due to the fact that many industries relied on natural fisheries rather than on culturing. However with the collapse of many fisheries, attention is shifting to the culturing of aquatic species. In the light of this new direction, the development and application of molecular markers are becoming more important in this sector. Currently the most common use of molecular markers in aquaculture species is for population management of both wild and cultured stock (Ferguson *et al.* 1995; Smith *et al*. 2005). However, greater emphasis is being placed on the construction of linkage maps for QTL analysis and selective breeding programs (Davis and Hetzel 2000; Elliot 2000; Hulata 2001; Liu and Cordes 2004). Such programs are vital for the continued supply of aquaculture products to the market, by selecting for production traits, subsequently increasing production output.

In essence a molecular marker is any stretch of DNA sequence that exhibits sufficient variation amongst individuals that adheres to Mendelian segregation and can be traced through a pedigree or phylogeny. An array of polymorphisms has been identified as markers, with older types including: restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), single strand conformational polymorphisms (SSCPs), minisatellites and random amplified polymorphic DNA (RAPDs). These have mainly fallen out of favour with many scientists due to either their taxing nature to produce results or to their low information content because of dominant inheritance or low polymorphism; nonetheless they remain in use where sequence information is limited (Vignal *et al.* 2002; Van Marle-Köster and Nel 2003).

Currently the most widely used molecular marker in animal genetics is microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs). Their popularity stems from the high levels of polymorphism displayed, their genome-wide distribution at relatively high frequency as well as their ease in genotyping and results interpretation, using PCR and computer programming (Beuzen *et al.* 2000; Vignal *et al.* 2002; Van Marle-Köster and Nel 2003). A new variety of polymorphism, the single nucleotide polymorphism (SNP), is rapidly gaining popularity due to its technical simplicity and promise of high throughput and total automation (Beuzen *et al*. 2000; Rengmark *et al*. 2006). Even though it is co-dominant, the bi-allelic nature of this marker lowers the information content in comparison to the multi-allelic microsatellite that confers higher information content. This disadvantage is easily overcome by the sheer number of SNPs distributed throughout a genome (Brumfield *et al.* 2003). The use and advantages of SNPs have been demonstrated in various applications including parentage assignment, individual identification (Werner *et al*. 2004), population and phylogenetic inferences (Brumfield *et al.* 2003; Morin *et al.* 2004; Seddon *et al*. 2005; Rengmark *et al*. 2006).

Molecular marker technology can play a vital role in the abalone aquaculture industry, especially considering that abalone domestication is in its infancy. Very little is known about the genetic contributions to abalone production traits. Therefore ensuring a broodstock population representative of the wild population will capture the total genetic diversity within the captive stock; allowing the captive population to adapt to the demands of a new artificial culture environment. This can be achieved by molecular marker-based estimation of diversity parameters and other population statistics (Mgaya *et al.* 1995; Evans *et al.* 2004). Molecular markers are also envisioned to play an important role in the general management and genetic improvement of cultured stock (Roodt-Wilding and Slabbert 2006). Furthermore, molecular markers are imperative to linkage analysis and the construction of genetic maps. Dense marker maps can serve as a framework for high resolution association studies for the identification of quantitative trait loci (QTLs) facilitating the implementation of marker assisted selection (MAS) (Borevitz and Chory 2004; Baranski *et al.* 2006; Liu *et al.* 2006; Hayes *et al.* 2007*a*; Khatkar *et al*. 2007). Molecular markers also revolutionised genomic annotation with dense maps illuminating anomalies regarding genomic structure, linkage disequilibrium and recombination hotspots (Pritchard and Przeworski 2001; Khatkar *et al*. 2007; Sekino and Hara 2007).

Traditional methods for the detection of molecular markers, in particular microsatellites and SNPs, are based on the construction of either genomic or cDNA libraries. In the case of microsatellites these are then screened using a synthetic oligonucleotide hybridisation probe representing a repetitive sequence. Constructs providing a positive signal are then sequenced for validation (Thiel *et al.* 2003). SNPs are detected based on locus-specific sequence variation amongst multiple individuals, using multiple alignments of sequences selected at random from the libraries. Depending on the sequence of origin, molecular markers can be classified into two categories: type I and type II markers. Type II, also known as random or anonymous markers, are by far the most common, derived from random genomic segments (genomic libraries) of which the context is not known; thus anonymous DNA (Gupta and Rustgi 2004; Serapion *et al.* 2004). This is currently the predominant marker type available for *H. midae.* On the contrary, type I markers are derived from known genic regions (e.g. Expressed Sequence Tags, ESTs) and are therefore gene-linked markers, which demonstrate their obvious advantage. Being gene-linked, these markers facilitate the identification of candidate genes, shed light on gene function, expression and regulation as well as filling the gaps in marker maps or producing transcriptional/functional maps. Type I markers will also simplify the identification of functional markers i.e. markers that demonstrate phenotypic causality (Gupta and Rustgi 2004; Serapion *et al.* 2004). This is essential to the abalone industry in South Africa for MAS and subsequent establishment of a domesticated, genetically enhanced strain of *H. midae* that will express superior production phenotypes.

Conventional methods of type I molecular marker development via molecular techniques are laborious, time-consuming and expensive. However the continual expansion of EST data in public databases has fashioned an alternative route for type I marker development (ESTs are cDNA-derived sequences and are thus representative of the transcribed region of the genome). This method is based on the *in silico* evaluation, using computer programming, of these ESTs. Previous investigations revealed the feasibility of this approach for SNPs (Hayes *et al.* 2007b,c; Quiling *et al.* 2007; Souche *et al.* 2007), microsatellites in plants (Thiel *et al.* 2003; Qureshi *et al.* 2004) and animal SSRs (Cnaani *et al.* 2002; Serapion *et al.* 2004; Antunes *et al.* 2006; Maneeruttanarungroj *et al.* 2006; Provan *et al.* 2007). Furthermore, coding sequences are more likely to be conserved across related taxa; this has been confirmed for EST-microsatellite flanking regions (Farber and Medrano 2003, 2004) and exonic sequences (Aitken 2004). This can therefore facilitate the development of cross-species molecular markers and synteny mapping, with marker transferability a linear function of phylogenetic distance. The advantage of this approach is that it allows for marker development in species with limited sequence data, using CATS (comparative anchor tagged sequences) primers, and more comprehensive conclusions regarding evolutionary relationships between species and genes (Decroocq *et al.* 2003;

Aitken 2004; Chagne *et al.* 2004; Fraser *et al.* 2004; Ju *et al.* 2005; Kumpatla and Mukhopadhyay 2005; Pérez *et al.* 2005).

The current study aims to develop type I/gene-linked molecular markers for the economically, yet vulnerable *Haliotis midae*, via a cost and time efficient *in silico* mining approach, utilising sequence data from native and other Haliotids. This will advance our understanding of structural, functional, and comparative genomics in the context of molluscan evolution and enhance selective breeding programs for the domestication of *Haliotis midae*.

2. General Biology, Ecology and Evolution of Abalone

2.1. Classification, Evolution, Phylogeny, and Distribution

The phylum *Mollusca* is an incredibly diverse animal taxon, second only to the *Arthropoda*, which includes the insects, arachnids and crustaceans. The molluscs are divided into several classes that represent independent lineages, thought to be derived from a common molluscan ancestor. The most commonly known molluscs are: *Polyplacophora* (chitons), *Bivalvia* (mussels, oysters, scallops etc.), *Cephalopoda* (octopi, squids and cuttlefish) and *Gastropoda* (snails and slugs) (Bieler 1992; Raven and Johnson 2002). Abalone, genus *Haliotis*, falls within the last mentioned and largest class of *Gastropoda* (Table 1.1). These molluscs are characterised by a single, spiralled shell and a distinctive muscular foot that forms the primary locomotive organ. Typically it is this foot that is the most obvious under the shell of the animal (note that in the case of slugs the loss of the shell is a secondary trait). Most distinctive of all gastropods is the ontogenic phenomenon of torsion: the anticlockwise twisting of the viscera and mantle in 180º along the anterior-posterior axis. It is this that provides the unique morphological and anatomical features of Gastropods (Bieler 1992).

Table 1.1: Scientific classification of the abalone (*Haliotis*)

Kingdom:	*Animalia*
Phylum:	*Mollusca*
Class:	*Gastropoda*
Order	*Orthrogastropoda*
Family:	*Haliotidae*
Genus:	*Haliotis*

Historically the Gastropod class was subdivided into three subclasses, based on morphological data, in particularly the locality of the gills in relation to the heart (*Opisthobranchia* and *Prosobranchia*) or the absence of gills as with the *Pulmonata* (Bieler 1992; Winnepenninckx *et al.* 1998). This classification has mostly fallen out of favour due to a re-evaluation of molluscan phylogeny (Hanszpruner 1988) and more resent molecular evidence that suggests the paraphyly of the *Prosobranchia* (Winnepenninckx *et al.* 1998), thus placing *Haliotis* in the monophyletic *Orthrogastropoda*, previously affiliated with the *Prosobranchia*. Furthermore, with an inner lining of mother-of-pearl in the shell, a pair of diotocardic bipectinate ctenidia with bursicles and a streptoneurous nervous system, *Haliotis* is grouped in the *Vetigastropoda* based on the anatomical features associated to this order (Hanszpruner 1993). As a member of the *Haliotidae* family, Haliotids shares these common traits: peripheral row of several tremata on the left side of the flattened shell, a well-defined, hypertrophied epipoduim and symmetrical radular teeth with marginal teeth demonstrating denticulate cusps and a well-developed rachidian tooth (Geiger 1999).

The position of *Gastropoda* and its relation to possible sister taxa within the molluscan phylogeny remains unresolved, with several hypotheses postulated (Winnepenninckx *et al.* 1996 and references therein). What remains clear, however, is the monophyletic origin of *Gastropoda* (Bieler 1992) and the basal position of *Vetigastropoda* within the *Gastropoda*; an indication of the early divergence of this group from the rest of the Gastropods (Winnepenninckx *et al.* 1998). The exact ancestor of the Haliotids is unknown, but fossils dating back to the Cretaceous from California suggest their existence since almost 70 MYA (Lindberg 1992 and references therein). Controversial fossils from the same period found in Europe were thought not to abalone-like in nature (Lindberg 1992 and references therein). However, recent phylogeographical studies hypothesise a European origin of at least modern Haliotids and that they radiated from Europe (Mediterranean) in an easterly direction to Australasia, Africa, Asia, and North America (Fig. 1.1) (Estes *et al.* 2005; Streit *et al.* 2006).

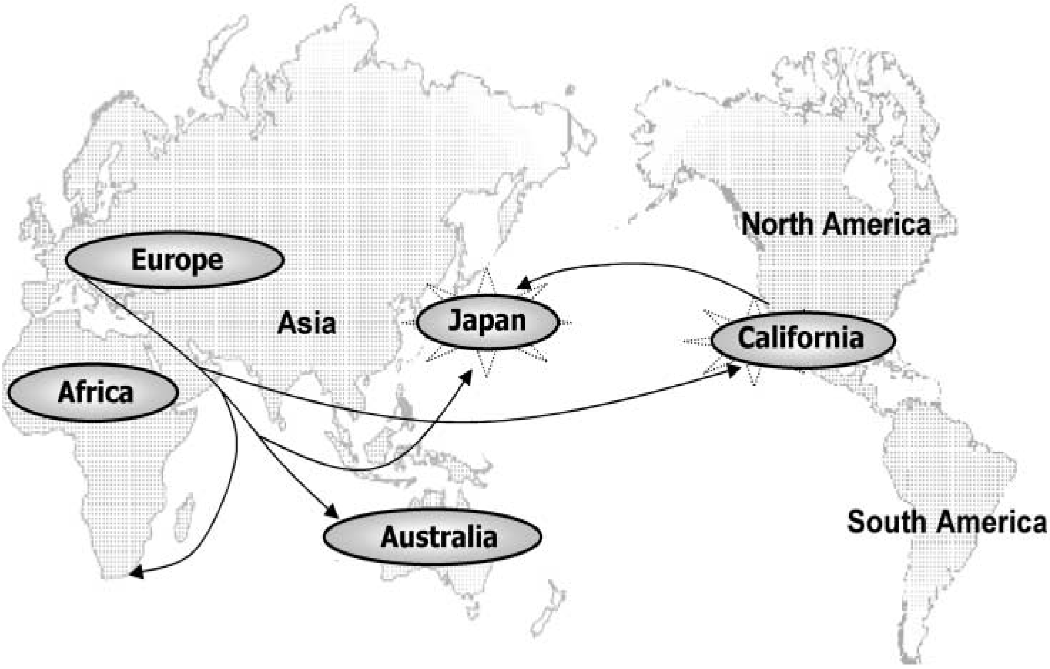


Figure 1.1: This image taken from Streit *et al.* (2006) depicts the possible spread of modern Haliotids from Europe.

By the beginning of the Miocene, Haliotids already demonstrated a world-wide distribution (Fig. 1.1) (Lindberg 1992). In the modern oceans abalone are endemic to the coastlines of Europe, Asia, the western seaboard of North America, southern Africa, Australia and New Zealand (Lindberg 1992; Streit *et al.* 2006). Globally 56 species of abalone are currently recognised (Geiger 2000; Degnan *et al.* 2006). Phylogenetic structure of the *Haliotidae* confers strictly to geographical range, with two clades predominating: a Northern Pacific clade (North American and Japanese species) and an European-Australasian clade (European, Australian, New Zealand and southern African species) (e.g. Estes *et al.* 2005; Degnan *et al.* 2006) (Fig.1.2). Furthermore, within the southern hemisphere species, two distinct groupings were recently documented, consisting of the southern African species and the remainder of southern hemisphere species (Bester-Van der Merwe 2009). This expanded radiation could be viewed as evidence against the grouping of all abalone into a single genus, as genetic distance reveals many species to be divergent to such an extent that can generally be recognised as separate genera (Brown and Murray 1992). This is probably due to the evolutionary age of *Haliotis* (Brown and Murray 1992; Lindberg 1992).

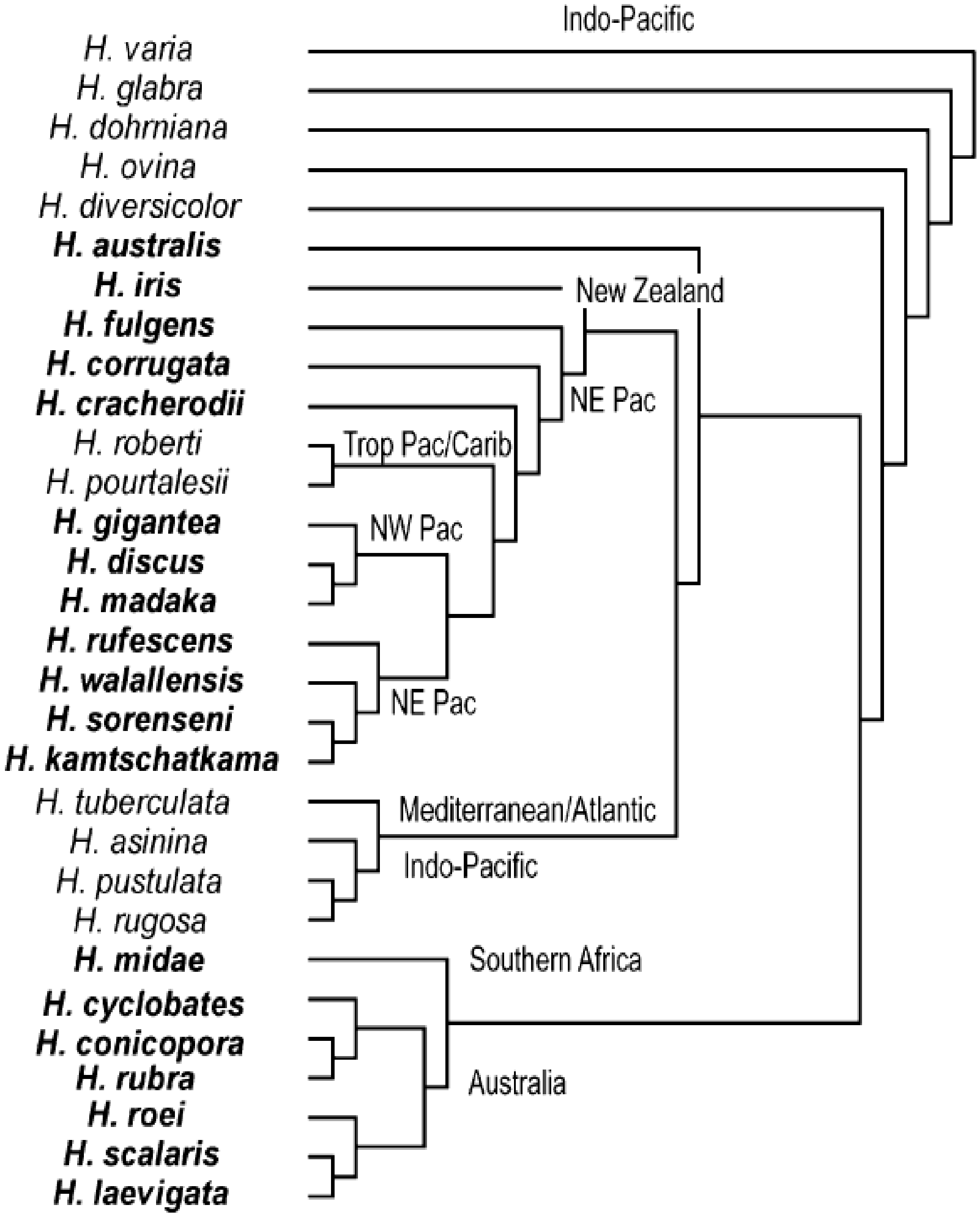


Figure 1.2: The *Haliotis* molecular phylogeny based on maximum parsimony analysis of the *16S*, *COI*, *ITS* and *lysine* genes (Figure taken from Estes *et al.* 2005).

The South African abalone, *Haliotis midae*, is the most well-known and characterised of the five endemic species. It has the second largest range after *H. spadicea* (Lindberg 1992), stretching from Cape Columbine in the west to Transkei in the east (Fig.1.3) (Troell *et al.* 2006). The other three species are *H. parva, H. queketti* and *H. speciosa,* with little known about these faunas. The origin and relation of *H. midae* to other abalone species remains unresolved. Molecular data strongly advocates an European ancestor, grouping *H. midae* in the European-Australasian clade (Streit *et al.* 2006). Recent investigation also suggests Australian species, in particular *H. rubra,* as sister taxon to *H. midae*, however relationships to the tropical Indo-Pacific species cannot be excluded (Estes *et al.* 2005; Degnan *et al*. 2006; Streit *et al.* 2006). The monophyly of the South African species is supported by the recent findings, grouping *H. midae* and *H. spadicea* as sister taxa, indicating recent divergence, whilst placing *H. parva* as a most-likely ancestral representative of the South African species. However complete phylogenetic analysis remains to be done using data of all five South African species (Bester-Van der Merwe 2009).

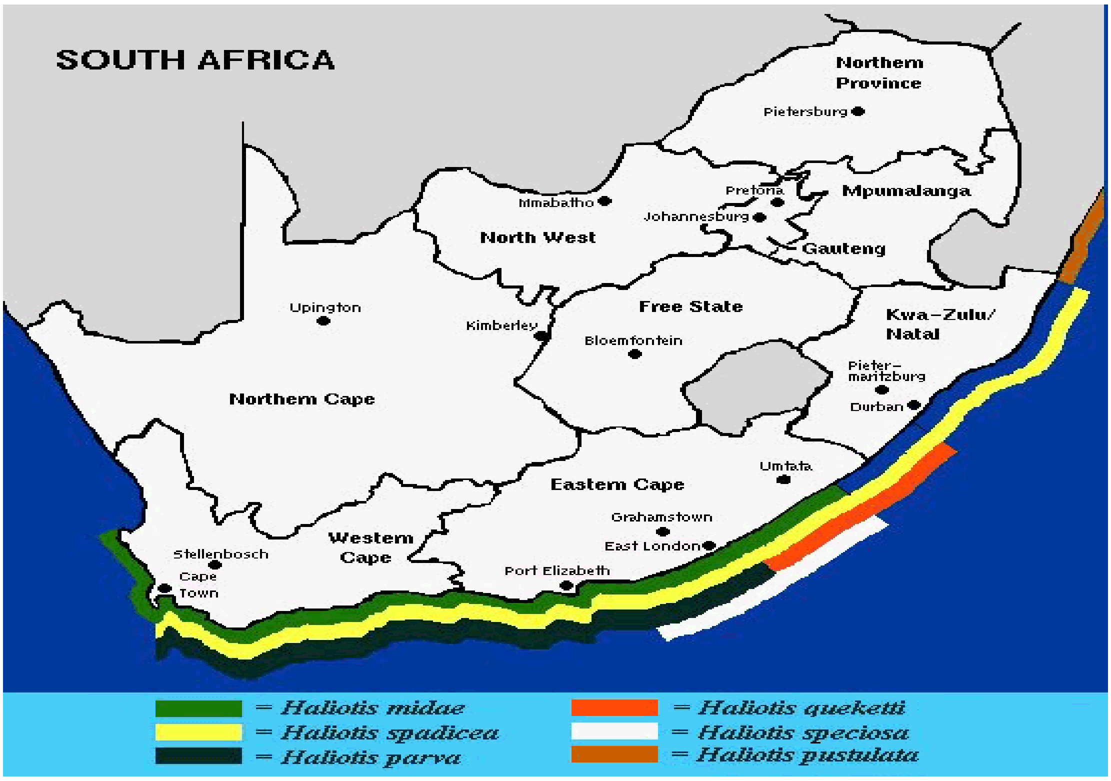


Figure 1.3: A map of the South African shore, showing the range of the five endemic species (www.abalone.cenrm.uwa.edu.au).

2.2. Life History and Ecology

Tropical abalone resides in shallow water, reef ecosystems and generally lodges onto stones or coral. On the other hand temperate species, such as *H. midae,* prefer rocky, kelp-bed habitats and utilise rocks and boulders as substrates. The inter-tidal zone, generally 10m deep, is where most animals are found, but individuals in waters up to 30m deep are not uncommon (Lindberg 1992). For most of its adult life abalone remains sedentary only moving occasionally to find improved foraging grounds. When a satisfactory locality has been found, an adult animal will often aggressively defend its territory from others (Tarr 1995). Larger adult animals are relatively inert, as locomotion seems to be a function of age, with younger individuals more prone to movement, possibly until such time as they find an optimal “home site” (Tarr 1995). Abalone are herbivorous, feeding on drifting seaweed trapped under its foot or grazing micro-algae from rocks (Barkai and Griffiths 1986; Tarr 1989; Wood and Buxton 1996). As such abalone play a vital role in the ecosystem as a primary consumer; freeing minerals and energy trapped by photosynthetic algae (Raven and Johnson 2002). *Haliotis midae* seems to be most active at night, feeding during the morning hours before sunrise. This nocturnal behaviour is suggested to be a strategy to avoid octopi and crabs, the most common predators of abalone (Wood and Buxton 1996). Even though abalone feed on a variety of algae, they prefer kelp and red algae (Barkai and Griffiths 1986; Wood and Buxton 1996). The choice of feed depends more on what is abundant in a particular environment, than on seasonality of selected algae types (Barkai and Griffiths 1986). There are also discrepancies in the diet of adult abalone compared to that of juvenile animals, which prefer thodophytes and calcified *Corallina* species. These deviations are indicative to the particular nutritional requirements of juveniles for growth and development (Wood and Buxton 1996).

Unlike the benthic adult abalone, larvae start off as part of the pelagic plankton (Fig. 1.4) (McShane 1992 and references therein). After the simultaneous release of ova and sperm into the water, constituting broadcast spawning (a common reproductive strategy of many marine molluscs) that allows for external fertilisation, the abalone zygote undergoes rapid cell cleavage to produce a morula, which in turn develops into the trochophore. When the trochophore hatches from the egg approximately 20 hours after fertilisation, it marks the end of the embryonic phase and the start of the larval stage as the veliger (Tarr 1989). The veliger remains planktonic for a further 5-7 days, during which time the larvae undergo the gastropod indicative process of torsion (Tarr 1989; McShane 1992).

Abalone larvae are lecithotrophic and as such they do not feed on external sources of nourishment, but rather rely on yolk supplies (McShane 1992). During the embryonic and larval stages the abalone young are extremely vulnerable to ocean currents and predation. The reproductive strategy of abalone to counter this loss is to release mass amounts of gametes in a synchronised event (broadcast spawning), because even though larvae do exercise some locomotion by means of beating cilia, the directionality of this movement is debatable (Tarr 1989; McShane 1992).

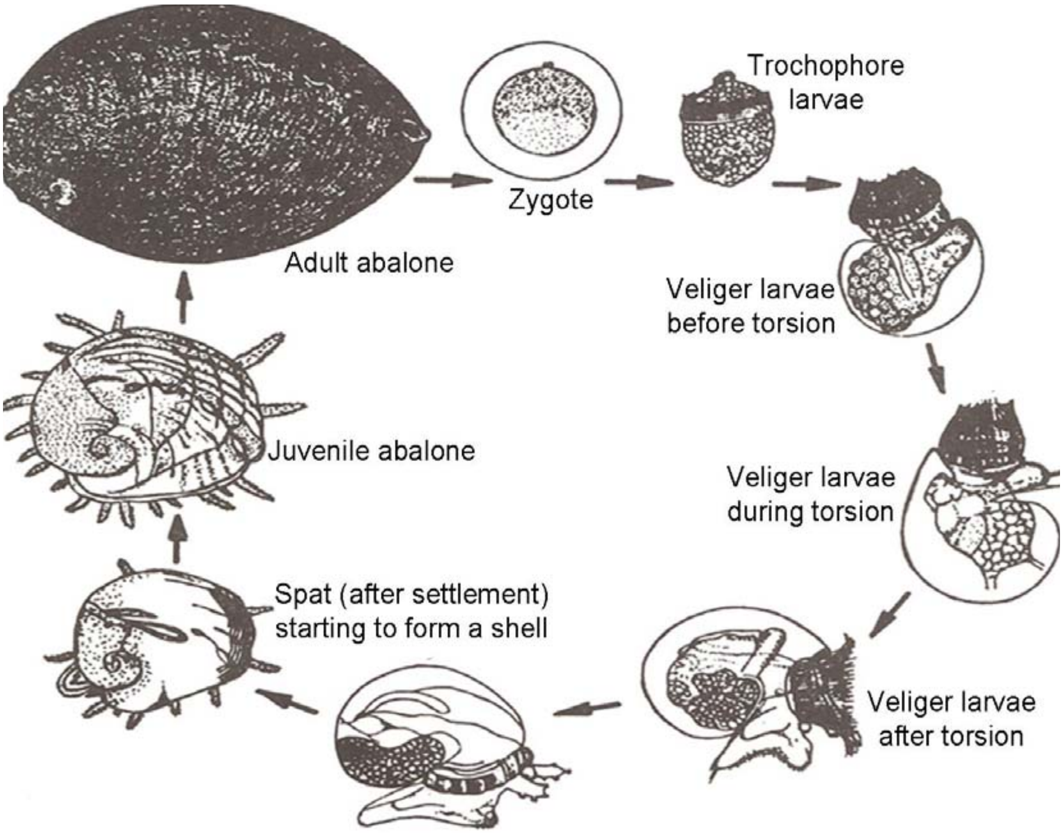

Trochophore larvae
Zygote
Adult abalone
Veliger larvae before torsion
Juvenile abalone
Veliger larvae during torsion
Spat (after settlement) starting to form a shell
Veliger larvae after torsion

Figure 1.4: Illustration of the abalone life cycle (adapted from www.itresourcing.com.au).

Various queues, such as substrate topology and water temperature have been postulated to facilitate the settlement of larvae (spat fall) on crustose coralline algae, however none is as important as the pheromonic action of gamma-aminobutyric acid (GABA), secreted by the diatom filaments. This chemical acts as a neurotransmitter that initiates metamorphosis of the larvae into juvenile abalone (Tarr 1989; McShane 1992; Day and Branch 2000). These abalone recruits (also known as spat) favour moderately thick, rough textured encrusting corallines that seem to offer greater shelter, protection from predators and improved grazing (Day and Branch 2000). As these recruits grow into juveniles, they are no longer sufficiently covered by the coralline, leading them to develop photophobia, seeking dark crevasses for protection (Tarr 1989). Juvenile abalone also develop a unique commensal symbiosis with urchins, hiding under their spines for fortification. A marked increase of predation on urchins by rock lobster has been found to lead to a decrease in juvenile abalone survival (Tarr *et al*. 1996). As the juveniles mature into adult abalone after 4 to 7 years, they progressively move from grazing micro-algae to trapping macro-algae under their foot. As such they will resume a more static life style after finding an optimum position to create a “home site” amongst boulders in the kelp-beds.

3. Fishery and Culture

3.1. History and Development of the Industry

Of the five local species of abalone in South Africa, *Haliotis midae,* more commonly known as perlemoen, is the only of economic value. Abalone fisheries are probably one of the oldest in the world as well as in South Africa. The Japanese have been harvesting abalone since 425 AD and in South Africa, stone-age-man collected these animals from as early as 125,000 years ago (Tarr 1989 and references therein). Abalone remains a highly prized commodity, especially in the Far East that primarily supports the export of abalone products from the Americas, Australasia and South Africa. Even though its one of the smallest fisheries based on tonnage caught annually, abalone fisheries are the most lucrative world-wide (Hauck and Sweijd 1999). The modern abalone fishing industry in South Africa commenced in 1949 and relied on the sub-tidal stocks off the coast of Saldanha through to Cape Agulhas, approximately 580km along the coast (Tarr 1989, 1992). To allow the animals sufficient breeding time before they are incorporated into the market chain, a size limit of 13.8cm (shell length) was imposed. Annual harvest increased incrementally to a record high of 2800 tons in 1965. However subsequent annual declines in abalone landings led to a growing concern regarding the sustainability of harvesting practises (Troell *et al.* 2006). This resulted in the implementation of a production quota system and a restricted fishing season (Troell *et al.* 2006). Further declines in stock numbers saw the successive decrease of the quota in following years and the revision to a whole mass system (the sum of the total allowable catch for each of the seven fishing zones to prevent regional overexploitation) (Fig. 1.5).

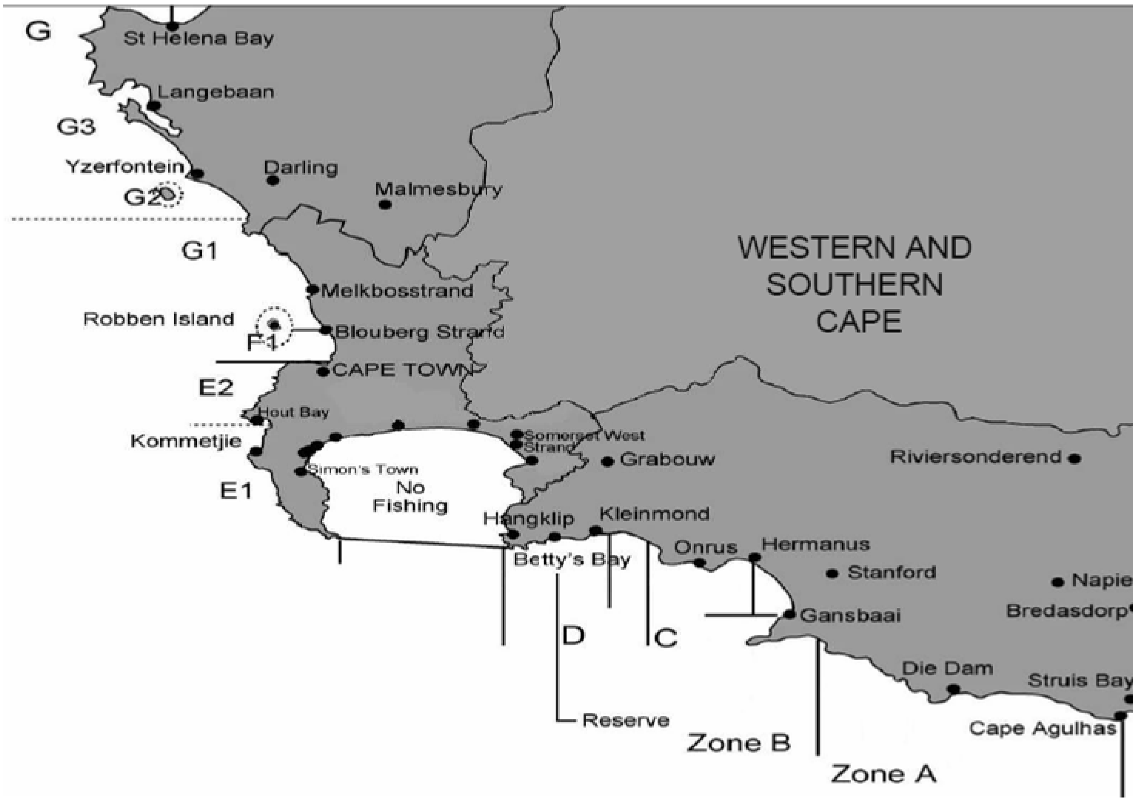


Figure 1.5: A regional map of the Western Cape coastline depicting the seven former abalone fishing zones (A-G) and sub-sections (e.g. E1, E2 etc.) (Figure taken from a presentation by A. du Plessis (2006) International Abalone Symposuim, Chile).

This quota was further reduced incrementally to 615 tons in 1995 (Tarr 1989, 1992). During the 2006/2007 season the total landings for abalone reached an all time low of 125 tons, forcing Government to decrease the number of fishing zones from seven to four (Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs and Tourism, Republic of South Africa – 25 October 2007). In a further effort to curb the growing recreational fishing sector, closed seasons were introduced from 1985. Recreational fishermen were also restricted to using standard snorkelling gear only (Dichmont *et al.* 2000).

With growing apprehension in regards to the depletion of the wild resource, maritime aquaculture became a viable alternative to fishery for the expansion of the industry. This is a global trend with abalone fisheries in all major producing countries collapsing at phenomenal rates (McShane *et al.* 1994; Altstatt *et al*. 1996; Hobday *et al*. 2001; Hobday and Tegner 2002). It seems that producing countries only turned to aquaculture when it became apparent that wild harvesting was untenable and that it could no longer supply the growing demand. Thus the start of major culturing endeavours coincides with the decline in natural stocks. In the USA this was during the mid-1960's when the commercial abalone fishery supported five species of economic value (Ebert 1992; Altstatt *et al.* 1996; Hobday *et al.* 2001); in Mexico during the 1970's (Garza and Bernal 1992) and in New Zealand in 1980 (Tong and Moss 1992). China probably has the world's oldest abalone culture facilities, with local research in this field dating back to the 1950's (Nie 1992; Fleming and Hone 1996 and references therein).

South Africa is new to the mariculture industry, with first initiatives only introduced during the late 1980's; nonetheless the country has made substantial headway. Other than declining stocks, the first incentives for cultivating the endemic abalone, *H. midae*, came with the ability to stimulate perlemoen to spawn in captivity and to be reared successfully afterwards (Genade *et al.* 1988; Troell *et al.* 2006). Efficient food conversion, improved growth rates in captivity and initial successes in the USA and New Zealand further encouraged the development of the South African abalone mariculture industry (Troell *et al*. 2006 and references therein). After this initial research, several abalone fishery enterprises invested in farming ventures. While the majority of facilities were erected on the southwest coast, a few occur as far as Port Nolloth on the west coast and Port Elizabeth in the east. During 1996 the first of the farms entered the commercial export phase, though not at full capacity (Cook 1998). During 1998, twenty two tons of cultured abalone were produced with a value of almost ZAR 6 million (Hoffman *et al.* 2000). The latest estimations for South Africa show that current abalone aquaculture is worth ZAR 268.20 million (Britz and Lee 2009). This value is expected to grow as the industry now soly relies on aquaculture due to the ban on fisheries (Fig. 1.6) (Cook 1998; Roodt-Wilding and Slabbert 2006; Troell *et al.* 2006).

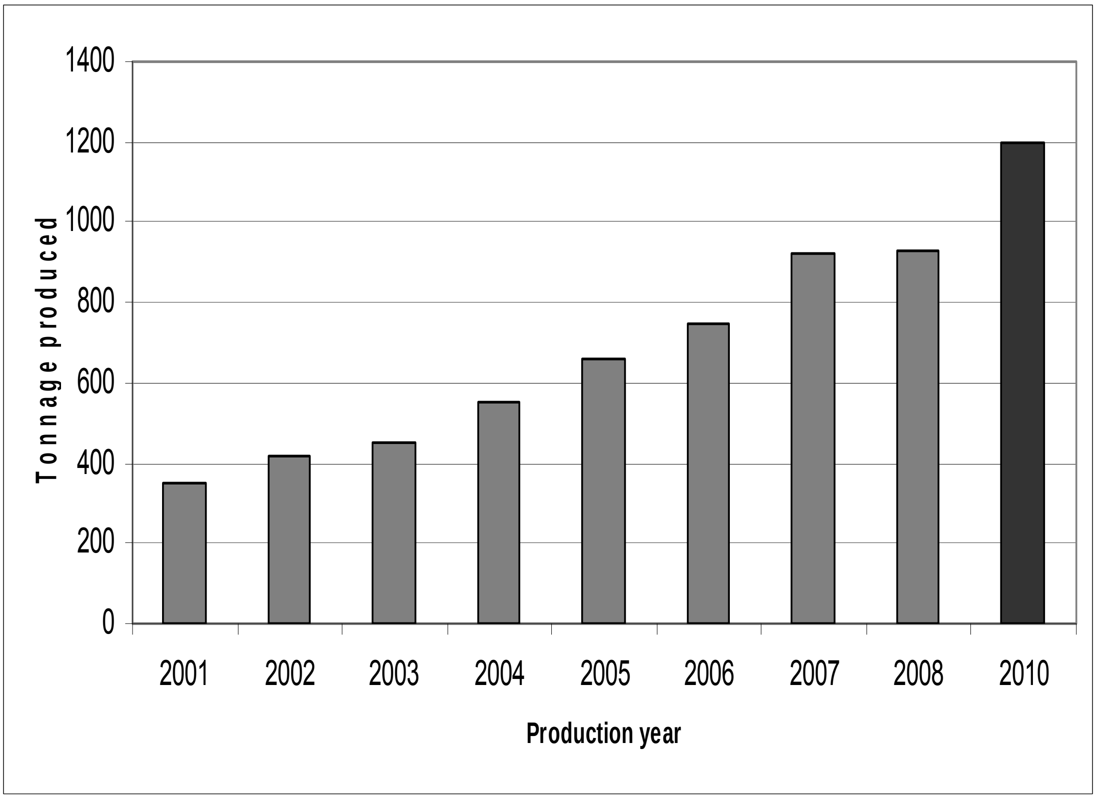


Figure 1.6: Abalone aquaculture production from 2001 to 2008, with projected figures for 2010 (This graph was constructed using data from a presentation by A. du Plessis (2006) International Abalone Symposuim, Chile and Britz and Lee 2009).

3.2. Current Perspectives and Practises

The Ministry of Environmental Affairs and Tourism (currently the Ministry of Water and Environmental Affairs) declared a total ban on the wild harvest of abalone, causing the fisheries industry to formally suspend all operations, effective from February 2008. This followed after critically low abalone landings during the 2007/2008 season of only 75 tons (Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs and Tourism, Republic of South Africa – 04 December 2007). The declines in abalone populations were attributed to the increased influx of rock lobster (Tarr *et al*. 1996) and illegal overexploitation by poachers (Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs and Tourism, Republic of South Africa – 04 December 2007). Worldwide a collapse in natural stocks led to the subsequent closure of abalone fisheries, but was instituted too late, causing a delay in wild population recovery. South Africa has the particular advantage that the problem was identified promptly and the fishery will in most likelihood reopen in years ahead, if poaching can be brought under control (Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs and Tourism, Republic of South Africa – 04 December 2007).

Poaching is the major obstacle for the recovery of the natural abalone resource. By 2002 the amount of abalone confiscated from illicit dealers surpassed that of the legal catch (Steinberg 2005). The majority of animals caught are also undersized (Dichmont *et al.* 2000; Plagányi *et al.* 2001). The uncryptic sessile lifestyle, shallow intertidal habitats and high value of abalone makes it particularly vulnerable to poaching (Hauck and Sweijd 1999). Poaching in South Africa is complicated by the country's unique socio-economic circumstances and political history; therefore a co-operative management system between government and local communities are vital for the preservation of the wild resource (Hauck and Sweijd 1999). Even though impoverished fishing communities poach due to need, these communities are often exploited by the greed of syndicates, fuelling the black market and the ever growing organised crime sector (Hauck and Sweijd 1999). Other than the obvious overexploitation and looming extinction of the fished animal due to falling broodstock densities, there are other important ramifications of poaching such as for example a collapse of the ecosystem, because abalone are important grazers in kelp bed habitats. Various economic factors will also be impacted, because the market will be swamped with illegitimate products, subsequently the aquaculture industry will suffer losses. Lastly it could result in social unrest due to conflict between the legal operators and illicit poachers as well as the exploitation of the poor and the youth (Hauck and Sweijd 1999 and references therein).

With the closure of the commercial fishery, aquaculture has become the only, lawful, supply for the growing demand for abalone products. South Africa has benefited greatly from technology transfer and success in other countries and currently there are 18 local abalone aquaculture farms (Britz and Lee 2009). Most farms function on an on-growing, land-based husbandry system, where filtered seawater is pumped to the holding facilities. Because of the lack of sheltered bays and hostile sea currents, in-ocean cage systems for culturing in South Africa are impractical (Cook 1998). The production of abalone revolves around the life cycle of the animals. In South Africa, the on-farm hatchery broodstock populations are wild caught individuals, representative of the wild resource, kept in individual tanks. These are conditioned to spawn, using peroxide treatment or water temperature control to produce the seed for commercial production (Cook 1998; Sales and Britz 2001). Fertilisation is achieved by adding sperm to the eggs in a hatching bin, where embryonic development takes place. Thereafter larvae are flushed into a rearing tank (Spencer 2002). South African farmers prefer to induce spat fall by providing plastic plates prepared with coated, naturally grown algae/bacteria biofilms, rather than GABA treatment (Cook 1998). The newly settled recruits are then moved to an indoor nursery for 4 – 6 months before they are moved again to outdoor tanks (for the grow out phase) and weaned on macro-algae for 4 – 5 years untill they reach market size (Cook 1998; Spencer 2002).

3.3 Advanced Technology and Abalone Culture

The abalone cultivating industries are under considerable pressure to keep up with the world demand and therefore the development of a genetically superior abalone strain for production traits is paramount. It is therefore envisioned that, unlike traditional livestock whose domestication was achieved by centuries of selection and careful breeding, the genetic improvement and domestication of abalone will be reliant on both traditional animal breeding methodologies and novel technologies. These technologies, including ploidy manipulations, gene transfer and molecular marker technology will fast-track the domestication process to fill the void in the market.

Aquaculture species are unique in that polyploid individuals are viable, unlike birds and mammals. As a result, various fish and shellfish species have been targeted for ploidy induction (Dunham 2004). Ploidy manipulation in abalone is mainly directed at the production of triploid animals, in other words animals that possess three genomic chromosome sets. However because the first and second polar body is present right after fertilisation in molluscs, the production of tetraploid animals are theoretically possible (Dunstan *et al.* 2007). The allure of triploid animals is potential faster growth rate and larger animals (Elliott 2000). Higher growth rate stems from the redirection of metabolic resources from reproduction to somatic growth: because three homologous chromosomes cannot equivocally synapse, meiotic division is hindered, therefore such animals are sterile with underdeveloped gonads (Liu *et al.* 2004a). Furthermore, polyploidy gigantism produce larger animals, because their cells are larger: The cytoplasm to nucleus ratio is constant for most cells, thus when the nucleus is enlarged to accommodate the extra chromosomes, the cytoplasmic content increases accordingly, subsequently producing an animal with augmented dimensions (Dunham 2004; Dunstan *et al.* 2007). Triploidy induction in many abalone species has been performed using an array of chemical (e.g. cytochalasin B, 6-dimethylaminopurine, caffeine) and physical (e.g. temperature, pressure) stressors that prevent the expulsion of the first or second polar body after fertilisation. However there are various success rates of different methods in different species, ranging from 50% to 98% for induction success and 10% to 90% larval survival (Stepto and Cook 1998; Elliott 2000; Norris and Preston 2003; Liu *et al.* 2004a,b,c; Dunstan *et al.* 2007; Li Y *et al*. 2007; Okumura *et al.* 2007). In the South African endemic *H. midae*, triploid induction has been achieved using both chemical (cytochalasin B, Stepto and Cook 1998) and physical methods (hydrostatic pressure, De Beer 2004).

Gene transfer or recombinant DNA technology is the artificial incorporation of foreign/exogenous DNA into a host/target organism's genome. China was the first country to report the successful transfer of foreign DNA to an aquatic species (fish) in 1985 (Dunham 2004). Since then the majority of gene transfer research in aquaculture has been directed at increased growth rates and focus was placed on genes like growth hormone and insulin-like growth factors (Elliott 2000; Dunham 2004). The greatest advantage of gene transfer technologies is that gains are immediate, unlike conventional breeding that accomplishes significant gains only after a number of generations (Elliott 2000). Several methods are available for the incorporation of exogenous genes into the abalone genome, including: microinjection, electroporation or chemical mediation to ova and embryos (Powers *et al.* 1995; Elliott 2000). However sperm-mediated gene transfer has been suggested as an alternative to direct egg or embryo transfection, as sperm is more versatile and allows for the rapid, simultaneous treatment of many ova during fertilisation. It entails the transfection of sperm by means of aforementioned techniques after spawning (Sin *et al.* 1995; Tsai *et al.* 1997). Chen *et al.* (2006) took this further and argued that current sperm-mediated protocols remain laborious, time-inefficient and relies on expensive equipment, as both DNA constructs and gametes need to be prepared. They proposed and demonstrated the use of direct injection of foreign DNA into the testes of live males and subsequently the propagation of transgenic offspring as an alternative.

Ploidy manipulations and gene transfer are innovative technologies that hold the promise of major leaps for the aquaculture industry. However, in many instances the technologies, in their infancy, show variable results depending on methods used and species studied. Much refinement is needed before the technology could be implemented on a commercial scale. Some would also argue that it holds socio-political and ethical implications and that this may impact on the marketability of such produce. The current public perception of such technologies, especially in regards to animal manipulation, is generally negative. Thus due to technological constraints and current public thinking, ploidy manipulations and gene transfer may not be viable for the short term genetic improvement of abalone. Currently, molecular marker technologies (marker assisted selection), as an extension of conventional selective breeding is however an attractive option for the genetic improvement of abalone.

4. Molecular Markers

4.1. General Aspects and Older Molecular Markers

Molecular (DNA) markers have become the standard for measuring genetic variation and genomic annotation. However, the use of DNA markers as genetic markers is a relatively new development; a consequence of the molecular era. The first genetic markers employed were physical traits such as morphological characters. Of course such phenotypic characters governed by single genes are scarce and therefore limited in their applicability; resultantly the resource was soon depleted.

The next wave of genetic markers was biochemical markers, most notably blood group systems, such as the ABO blood groups in humans (Yamamoto 1990) and various other blood antigens (Weller 2001). Even allozymes, the first used molecular marker in main stream animal genetics, including fisheries science and aquaculture (e.g. May *et al.* 1980; Seeb and Seeb 1987), was a protein-based marker. This marker type is produced by differential migration of protein allelic variants during electrophoresis. Differences in motility are generally caused by the size, shape and charge variation between alternate alleles brought about by amino acid substitution, but protein truncations or insertions and deletions are not uncommon.

The shift from physical to biochemical markers represented a significant improvement, however biochemical procedures for protein isolation and analysis were often tedious and time consuming. The major drawback of protein-based markers however is that if the marker protein of interest was not expressed in a readily available tissue (e.g. blood or skin) the animal will have to be sacrificed or undergo a surgical procedure to collect a biopsy of the relevant tissue and often biopsy is not an option because protein isolation requires large amounts of tissue. In terms of animal breeding for conservation or commercial stud, where individual animals are highly valued, sacrifice and unnecessary stressors like surgery are unwarranted. Therefore, with the advent of the polymerase chain reaction (PCR) and DNA sequence technology it became much more feasible to use DNA-based markers. The most advantageous characteristic of DNA-based markers is that it provides a means to evaluate direct, genome-wide (coding and non-coding) genetic variation within and between individuals, populations and species.

A molecular marker can be defined as any sequence variation/polymorphism between individuals that is inherited in a Mendelian fashion, therefore traceable though a pedigree when examining successive generations and when these generations spread over evolutionary time, the molecular markers could be used to reconstruct phylogenies. Such sequence polymorphisms include: insertions and deletions, segment inversions and rearrangements, nucleotide base pair substitutions and variable number tandem repeats (Liu and Cordes 2004). Older types of molecular/DNA markers, such as AFLPs, RAPDs and RFLPs rely on the detection of such, aforementioned, variation without explicit identification of the causative mutation. On the contrary, more recent markers are based on the detection of the particular sequence variation, and include microsatellites and SNPs.

Molecular markers are broadly classified by means of isolation/detection, genomic context and association to functional variants (Dodgson *et al.* 1997; Dekkers 2004; Liu and Cordes 2004; Collard *et al.* 2005). Firstly in terms of means of isolation/detection, molecular markers can be classed as sequence-clone markers or fingerprint markers (Dodgson *et al.* 1997). Sequence-clone markers are isolated by the creation of DNA libraries and investigating individual cloned fragments for sequence polymorphism. Such markers include: SNPs, microsatellites and RFLPs. The most frequently used fingerprint markers in animal genetics are RAPDs and AFLPs. RAPDs are detected by employing short (8-10 mer) non-specific primers that randomly amplify genomic regions to create a DNA profile. The drawback of this marker is its reliability on stringent PCR conditions that impacts on the repeatability of the marker type (Dodgson *et al.* 1997). AFLPs, on the other hand, exploit a variety of genomic variation, including: restriction site polymorphism, indels and anonymous tandem repeats (Liu and Cordes 2004). Here genomic DNA is digested and universal adaptors are ligated to fragments and PCR amplified using adaptor annealing primers. This is then electrophoresed to generate a DNA fingerprint. The advantage of fingerprint markers over sequence-clone markers is that *a priori* sequence information is not required. As such fingerprint markers are based on genome-wide analysis of genetic variation through 'random amplifications' that create unique electrophoresis banding patterns for individuals. However, unlike the co-dominant sequence-clone markers, fingerprint markers are dominantly inherited, limiting their information content.

A further sub-classification based on detection, perhaps more on the visualisation or genotyping of markers, can also be made: hybridisation-based markers, PCR-based markers and sequence-based markers. The classification of markers in these three sub-classes also illustrates the evolution of molecular marker systems and technologies. The first molecular markers were detected via hybridisation techniques such as Southern blot analysis, e.g. RFLPs and minisatellites, where genomic DNA was enzymatically digested, electrophoresed and transferred to a membrane for hybridisation to a complementary visualisation probe (generally radioactive probes were used, but later fluorescently labelled probes became the norm). The advent of PCR largely saw the hybridisation markers fall out of favour, e.g. minisatellite being replaced by microsatellites or hybridisation markers being converted to a PCR system, e.g. RFLPs. With sequencing technology becoming evermore affordable there is a shift from PCR and gel electrophoresis (for unique banding patterns) to PCR and direct sequencing. A prime example is again RFLP markers where alternate alleles are created by a point mutation either creating or abolishing a restriction enzyme sites. With DNA sequencing, such point mutation can now be identified as particular nucleotide base pair substitutions and effectively be classified as a SNP. Even though RFLPs were the first DNA markers to be employed, today they mainly serve as a means for low-cost SNP genotyping.

In terms of genomic context, molecular markers are either classed as type I or type II markers (O'Brien 1999). Type I molecular markers are derived from known coding sequences, while type II markers are isolated from anonymous genomic segments. As such allozymes and EST-derived markers (e.g. EST-SNPs or EST-SSRs) are *de facto* type I markers, whereas RAPDs and AFLPs generated through random genomic amplifications, are type II markers. The overwhelming majority of molecular markers for aquaculture species, including abalone, are type II. The apparent usefulness of type I markers were overlooked by the early industry pioneers (Liu and Cordes 2004). A molecular marker associated to a coding region has a higher probability of conferring a phenotypic effect or being linked to (in phase with) a polymorphism in that gene that acts as a causative mutation of a phenotype. Furthermore, as gene sequences are better conserved between related species, type I markers will be easier to transfer from species to species, thus providing a platform for comparative genomics and synteny mapping. Perhaps more importantly for aquaculture, it allows for markers from a ‘marker-rich’ species to be transferred to a ‘marker-poor’ species; leading to the saturation of marker maps in less characterised fauna in an inexpensive and time saving manner. As such type I markers will find application in population studies for detecting signatures of selection and QTL analyses for the improvement of commercial stocks (Liu and Cordes 2004).

The third classification of molecular markers is derived from the association of markers to functional variants and three categories exist: 1) Direct markers, i.e. the molecular marker is the functional variant; 2) Linkage disequilibrium (LD) markers, the marker is in population-wide LD with the functional variant and 3) Linkage equilibrium (LE) markers, where the marker is in population-wide LE with the functional variant, however linkage occurs in family lines (Dekkers 2004). Unlike, the previous classifications that were based on the laboratory techniques for diagnosis and the material for isolation, this classification relies on the analysis of particular markers in pedigrees or populations and the identification of a genotype-phenotype correlation. In the case of LD markers this is achieved by candidate gene association tests, whereas LE markers are detected using linkage maps and pedigree analysis (Andersson 2001). It is however, more challenging to discern whether a marker, once a genotype-phenotype correlation has been established by association test or genome scan, is a direct marker (Dekkers 2004). As such, markers will generally remain LD/LE markers untill laboratory experiments (e.g. gene expression studies or protein activity assays) can prove functional causality. However with increasing bioinformatic capacity it has become easier to make sound predictions on the functionality of particular mutations that leads to putative direct markers. Table 1.2 summaries the attributes of the predominantly used molecular markers in aquaculture.

Table 1.2: A summary of key attributes of molecular markers routinely used in aquaculture [this table was constructed using elements from Liu and Cordes (2004); Schlötterer (2004); Collard *et al*. (2005)]

Marker	Detection	Type I/Type II	Mode of Inheritance (D/C)[3]	PIC[4]	Advantages	Disadvantages	Major/General Application
Allozyme	Protein isolation	I	C	Low	Relatively cheap, universal protocol	Tissue-specific, environmental factors may play a role, limited number of markers	Linkage mapping, population studies
RFLP	Sequence-clone, PCR	I or II[2]	C	Low	Robust, reliable	Bi-allelic, laborious and expensive to develop	Linkage mapping
RAPD	Fingerprint, PCR	II	D	Inter.	Easy, fast, inexpensive, analyse multiple loci	Low reproducibility	Population studies, strain ID
AFLP	Fingerprint, PCR	II	D	High	Simultaneous multiple loci analysis	Complicated methods for detection and analysis	Linkage mapping, population studies
SSR[1]	Sequence-clone, PCR	I or II[2]	C	High	Robust, reliable	Laborious and time consuming to develop, high mutation rates	Linkage mapping, population studies, parentage assignment
SNP	Sequence-clone, PCR	I or II[2]	C	Low	High genomic frequency, high-throughput, mutationally stable	Bi-allelic, expensive	Linkage mapping (fine mapping), population studies
mtDNA	Sequence haplotype	N/A	Maternally inherited	High	Multiple copies in cells	Only maternally inherited	Maternal lineage

1- SSRs (simple sequence repeats) are also known as microsatellites or short tandem repeats; 2- Depending on whether coding DNA, e.g. exon/EST or anonymous DNA was used for isolation; 3- Dominant or Co-dominant; 4- Polymorphism Information Content.

Over the years there have been definite trends in the development and use of molecular markers. At first, the choice of marker systems was limited by what was available; as such most early genetic studies relied on allozyme and RFLP data. Today, however, numerous molecular marker systems are in place. The choice of molecular marker should thus be based on what best suits the particular investigation at hand; taking in consideration budgetary constraints (Schlötterer 2004). Nonetheless, there is a clear decrease in the usage of RFLPs, RAPDs and allozymes and an increase in the usage of AFLPs, microsatellites, and SNPs (Liu and Cordes 2004). At first glance this may seem quite normal: older molecular markers making way for newer markers. But taking into consideration that older marker types are often less costly and remains relevant for particular applications, the shift may now seem unjustified. However, the failure of older markers lies in their inability to be "multi-applicable". As an example, SNPs and microsatellites may be employed for a variety of applications (including: linkage mapping, population studies, individual identification, parentage assignment, strain/species identification etc.), whereas RAPDs will suffice for strain/species identification but can not be readily employed for genetic mapping due to difficulties with reproducibility (Liu and Cordes 2004). Therefore it is sensible to invest in one or two marker systems that have wide applicability, rather than establishing multiple marker systems. This is of particular importance to the aquaculture industry where molecular markers are envisioned to play a significant role in diverse applications.

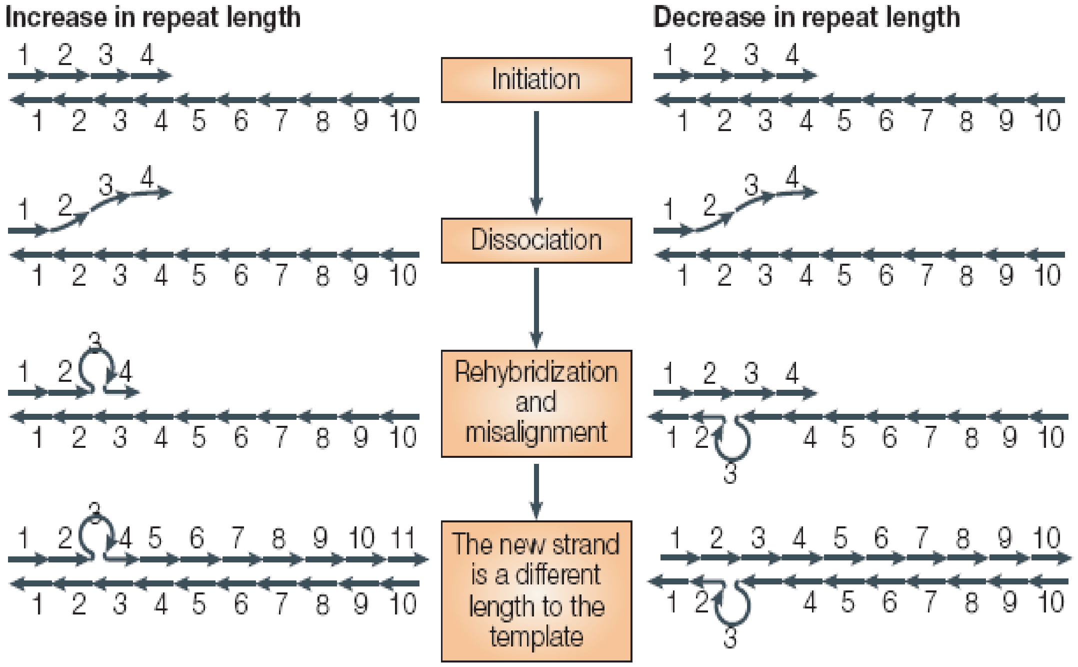
4.2. Microsatellites [Short Tandem Repeats (STRs), Simple Sequence Repeats (SSRs)]

Microsatellites are a group of repetitive DNA elements that are classed with minisatellites as variable number tandem repeats (VNTRs) (Nakamura *et al*. 1987). Minisatellite consists of larger, ten to fifteen, unit tandem repeats that may stretch up to 30 kb. On the contrary, microsatellites are two to six nucleotide base pair units that are repeated in tandem for a minimum array size of eight nucleotides, but much less than 1 kb (Chambers and MacAvoy 2000). The classification of microsatellites as di-, tri-, tetra-, penta-, and hexa-nucleotide repeats, based on the number of nucleotides per repeat unit have become standard practice. Chambers and MacAvoy (2000) further proposed a six class system of nomenclature based on the composition and structure of repetitive motifs at specific loci (Table 1.3) to more accurately reflect the pattern of repeat motif diversity.

Table 1.3: The six class nomenclature of Chambers and MacAvoy (2000).

Class	Description	Sequence example
Pure or perfect	The repetitive motif consist of a single repeated unit	$(AC)_n$
Interrupted pure	A perfect repeat motif that is interrupted by a non-repetitive sequence	$(AC)_n$ - AT - $(AC)_n$
Compound	The repeat motif consists of two or more perfect repeats that follow one another consecutively	$(CA)_n(CT)_n$
Interrupted compound	A compound repeat motif that is interrupted by a non-repetitive sequence	$(CA)_n$ - AGA - $(CT)_n$
Complex	Interrupted compound motif, where repeat units may differ periodically in base pair composition	$(AC)_n$ – ATT- $(CCTT)_n$ - $(CYKY)_n$
Interrupted complex	A complex motif where perfect repeats are interrupted	$(CA)_n$ – AT – $(CA)_n$ – ATT – $(CCTT)_n$ – $(CYKY)_n$

Microsatellite loci demonstrate a ‘life cycle’ of sorts; as such they have a genesis, a period of development and then ultimately demise. This is a process that may transcend generations through evolutionary time (Messier *et al*. 1996; Primmer and Ellegren 1998). It is postulated that microsatellites may have their origin in regions of ‘cryptic simplicity’, where rudimentary, imperfect repetitive elements are already in excess in the genome, subsequent chance point mutations may then lead to the creation of ‘proto-microsatellites’ that then expand into repetitive tracts (Tautz *et al*. 1986; Messier *et al.* 1996; Schlötterer 2000). It is generally accepted that the mutational mode of microsatellite, through their expansion phase, is slipped mispairing/replication slippage (Levison and Gutman 1987). The process is depicted in Fig. 1.7; during DNA replication the synthesised DNA strand temporarily dissociates from the template and re-anneals out of frame. This in turn generates allelic variation by the addition or loss of repeat units in the nascent strand (Ellegren 2000, 2004). From this model of microsatellite mutation it is plausible to envision that longer repeat tracts will more readily mutate as the probability of misalignment increases, because the region for possible misalignment is greater. Mutation rate may also depend on other factors; including specific repeat motifs (e.g. sequence or length of repeat units) and flanking sequence composition. Furthermore, it must be noted that even though the loss or gain of repeat units are theoretical equal per mutation event, empirical evidence suggest a bias toward gain-of-unit mutations. All of the aforementioned attributes of replication slippage, contributing to microsatellite evolution, is thought to be due to inherent limitations in the mismatch repair system that fail to restore larger segments of misalignment (Chambers and MacAvoy 2000; Ellegren 2000, 2004).



Increase in repeat length
Decrease in repeat length
Initiation
Dissociation
Rehybridization and misalignment
The new strand is a different length to the template
1 2 3 4
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11

Figure 1.7: An illustration of slipped mispairing, the mutational mechanism of microsatellites, leading to the addition or loss of repeat units (Figure taken from Ellegren 2004).

In theory, taking into account mutational bias for unit addition, microsatellites should be able to grow infinitely. In practice this is, however, not the case. Thus there exists an upper margin for microsatellite expansion (Nauta and Weissing 1996). This may in part be due to natural selection acting against superfluous expansions, leading to functional defects in the phenotype (Metzgar *et al.* 2000; Li *et al.* 2002). However, because most microsatellites are located in non-coding regions, the working premise is that they are selectively neutral. Thus alternative forces must be at play to create the observed expansion limits. One prevailing hypothesis, supported by numerous studies, seem to suggest that overly large microsatellite alleles are more prone to deletions, leading to shorter repeat tracts (Chambers and MacAvoy 2000; Ellegren 2000 and references therein). Ellegren (2000, 2004) also argues in support for the role of point mutations in the ‘death’ of microsatellites. Here chance point mutations accumulate in the repetitive tract, effectively generating interrupted pure microsatellites (Table 1.3). This slows the mutation rate by replication slippage, because of the shorter repeat arrays of interrupted microsatellites. As point mutations accumulate, the repetitive tract may revert back to a region of ‘cryptic simplicity’, bringing the microsatellite locus full circle. The incorporation of point mutations in the evolution of microsatellites puts forth a new model, where slippage and point mutations operate at opposites to create and decay microsatellites. It should be stressed that the mutational mechanisms discussed here are grossly over simplified and that the true evolutionary nature of microsatellites may also depend on several other molecular processes, most notably unequal crossing-over during recombination (Li *et al*. 2002), however the balance mutation model does provide a relatively well-understood and defined foundation for continued investigation and analysis.

Microsatellites are relatively uniformly distributed throughout the genomes of all known organisms at densities proportional to genome size (Tóth *et al*. 2000; Katti *et al*. 2001). However the randomness of particular localisation and type of repetitive motif represented in genomes remains a matter of debate. As such, translated regions of the genome are for the most part devoid of microsatellites, with the exception of trinuleotides enduring in some coding segments (Li *et al.* 2002). This is most probably due to negative selection pressure against repeat expansions that will cause frame-shift mutations (Metzgar *et al.* 2000). Furthermore there seems to be, in general, no differential for microsatellite abundance between intergenic regions and introns, in line with the neutrality theory and thus random distribution (Tóth *et al*. 2000). However genomic regions occur where the assumption of random base pair composition can not explain the overrepresentation of microsatellites (Bachtrog *et al*. 1999). Dinucleotide repeats are the predominant motifs and of these $(CA)_n/(TG)_n$ followed by $(AT)_n/(TA)_n$ repeat units are most abundant (Li *et al.* 2002; Ellegren 2004). This pattern of microsatellite dispersal seems to be present in molluscs as well (Cruz *et al*. 2005). Microsatellites have been found to be associated with transposable elements (Ramsy *et al.* 1999; Meglécz *et al.* 2007) that may partially explain dispersal patterns. But it has also been postulated that the biased distribution of microsatellites may allude to the possible genomic functionality of these sequences. Indeed, microsatellites have been shown to play a role in diverse functions including: chromatin organisation, gene expression and recombination (for thorough reviews: Li Y-C *et al*. 2002, 2004)

The unique evolutionary mechanism of microsatellites provides it with exceptionally high mutation rates several orders of magnitude greater than most other genomic regions at 10^{-3} to 10^{-4} per locus per generation (mutation rates may differ between loci, species) (Ellegren 2000). In turn this provides hyper-variable microsatellites with multiple alleles per locus, endowing microsatellites with the highest PIC-values (polymorphic information content) of all molecular markers (Liu and Cordes 2004). However the complex mutational mode of microsatellites makes it difficult to formulate appropriate population genetic models for microsatellite data analysis. Assuming selective neutrality, the most widely used models are the Infinite Allele Model (IAM) (Kimura and Crow 1964) and the Stepwise Mutational Model (SMM) (Ohta and Kimura 1973). None of these models however accurately reflect the observed mutational dynamics of microsatellites. Firstly, the IAM does not make provision for microsatellite size homoplasy. The mechanism of replication slippage may generate two alleles with identical repeat number, i.e. they would be identical by state. However as these alleles do not share a common ancestral lineage, they are not identical by decent; this constitutes homoplasy. Thus if homoplasy is taken into account 'identical by state' is misinterpreted as 'identical by decent' that subsequently diminishes the resolving power of microsatellite data (Chambers and MacAvoy 2000). On the contrary the SMM is thought to represent the addition and loss of repeat units by means of replication slippage more accurately, but still this model assumes infinite expansion is possible and also it does not make provision for bias toward gain-of-unit mutations. For this reason amendments have been made to the basic SMM, however these are mainly presented as independent models and thus a single consolidated model is still lacking (Chambers and MacAvoy 2000; Ellegren 2004). Microsatellites also suffer from various technical difficulties such as null alleles (failure of an allele to amplify due to primer binding site sequence variation) and stuttering (*in vitro* slippage of *Taq* polymerase causing multiple bandings of a single allele) leading to genotyping errors (Hoffman and Amos 2005; Girard and Angers 2008)

Nonetheless, high levels of polymorphism, co-dominant mode of inheritance, semi-automation and fluorescent dye capillary electrophoresis systems coupled to computer imaging programs for easy allele scoring has made microsatellites arguably the most popular molecular marker in use currently. Microsatellites have found application in diverse settings including: conservation, agriculture, evolutionary biology and forensic science. Here they are used to infer population history and dynamics, reconstruct phylogenetic relationships, identification of individuals or assignment of parentage, genome annotation and marker assisted selection.

Due to the power of microsatellites as genetic markers, a number of protocols have been developed to isolate microsatellites; it has proved not to be a trivial task. With very little genomic information readily available for non-model species, *de novo* microsatellite isolation is often required. Zane *et al*. (2002) provides a detailed review of major protocols for the isolation of microsatellites. In brief, most current protocols rely on three basic steps: construction of a partial genomic library, screening for positive clones and marker-specific primer design and optimisation. Traditionally, partial genomic libraries were constructed by selecting genomic fragments based on size (Fig. 1.8). Clones were then screened via colony hybridisation using repeat probes and positive clones are sequenced. From the sequences, repeat flanking primers are designed for PCR optimisation. This traditional procedure works well for microsatellite-rich genomes, but may be inefficient where densities are low.

Newer protocols endeavour to construct microsatellite-enriched libraries. One way of doing so is by employing RAPD amplicons (Fig. 1.8), often rich in repeat sequences. Southern blot analysis of RAPD profiles using repeat probes indicates bands containing microsatellites; these are then excised, cloned and sequenced. A more popular method, however, is selective hybridisation. These protocols are variations on the traditional method, where genomic fragments are selected to maximise the number of clones containing microsatellites. This is achieved by the introduction of a hybridisation (to a repeat probe) step to capture fragments containing repeat motifs. Captured fragments are then used to construct DNA libraries, generally followed by colony PCR and sequencing (Fig 1.9).

Once a marker system has been established for a focal species, time and effort can be saved for marker development in closely related species, as cross-species transfer of microsatellite can be done. It must be noted that the success of this strategy depends on the phylogenetic distance between the donor and recipient species. Furthermore, even though cross-amplification may be successful, high levels of polymorphism cannot be guaranteed (Chambers and MacAvoy 2002; Zane *et al*. 2002)

In recent years there has been a considerable effort to develop type I markers, especially for commercial molecular breeding projects (Liu and Cordes 2004). This led to adaptation of microsatellite isolation protocols from using genomic DNA to using cDNA (e.g. Cheng *et al*. 2007). However, as genomic and cDNA sequence information expands exponentially in public databases, *in silico* avenues are becoming ever more attractive. A large number of gene expression studies specifically are generating copious numbers of EST data for many species, and as these consist of coding sequence, they are more readily transferable between species. Furthermore, the growing number of computer software programs for identification of microsatellites in sequences makes it easy to identify repetitive motifs in large numbers of sequences simultaneously (Cnaani *et al.* 2002; Serapion *et al.* 2004; Antunes *et al.* 2006; Maneeruttanarungroj *et al.* 2006; Provan *et al.* 2007). The major advantage is thus that the extensive construction of DNA libraries and colony screening is omitted and sequences (as downloaded from the databases) are ready for analysis and primer design. Irrespective of the strategy followed for microsatellite isolation, a key aspect of marker development remains a comprehensive optimisation of PCR conditions. Here a balance must be struck between maintaining high primer fidelity for product specificity and the generation of a sufficient signal for detection. Once this has been achieved microsatellite loci are checked for polymorphism, generally via polyacrylamide gel electrophoresis and if this is confirmed the marker is implemented in a genotyping system.

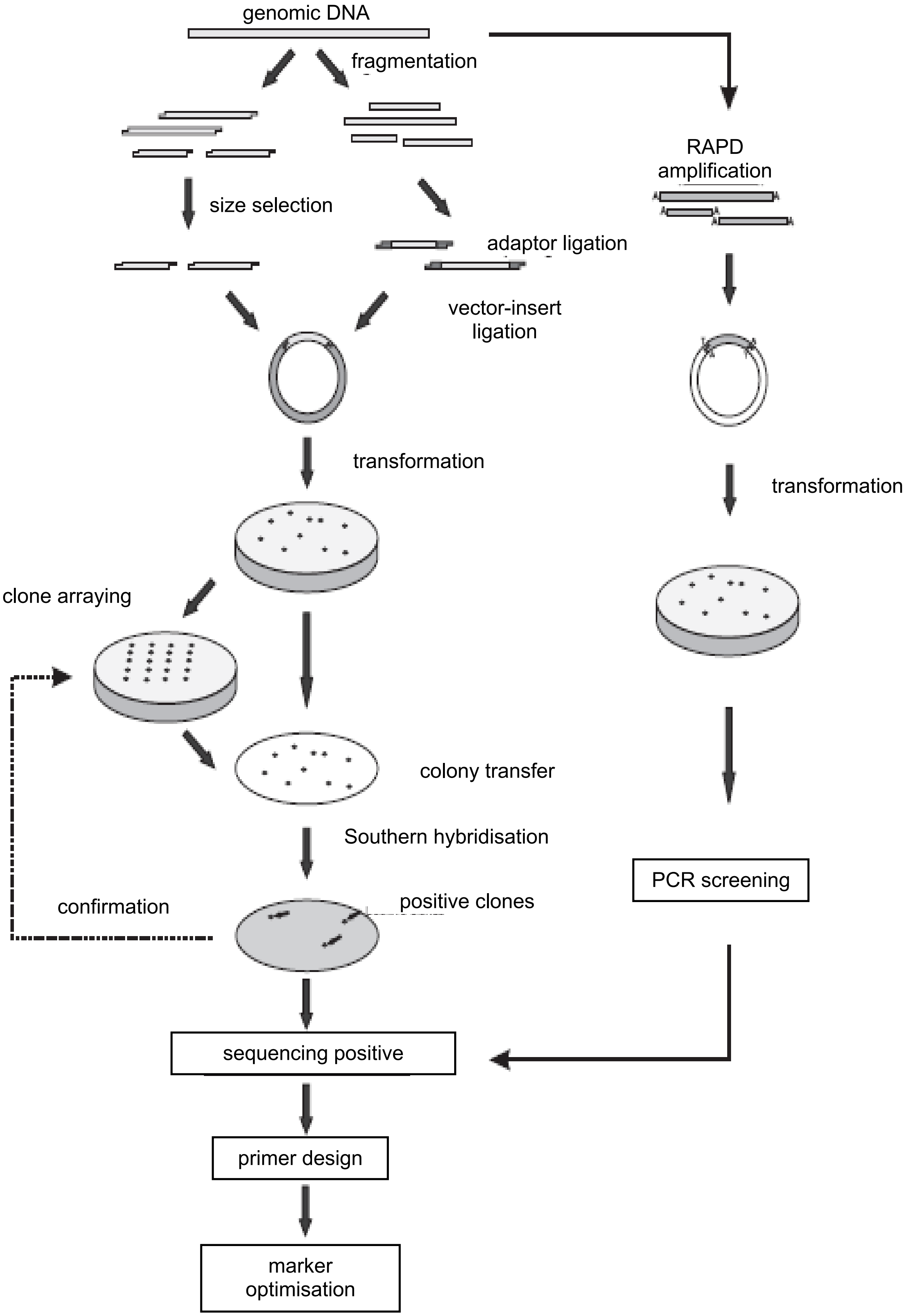


Figure 1.8: A diagram showing the traditional method, as well as the use of RAPD amplicons, for isolating microsatellites loci (Figure taken from Zane *et al*. 2002).

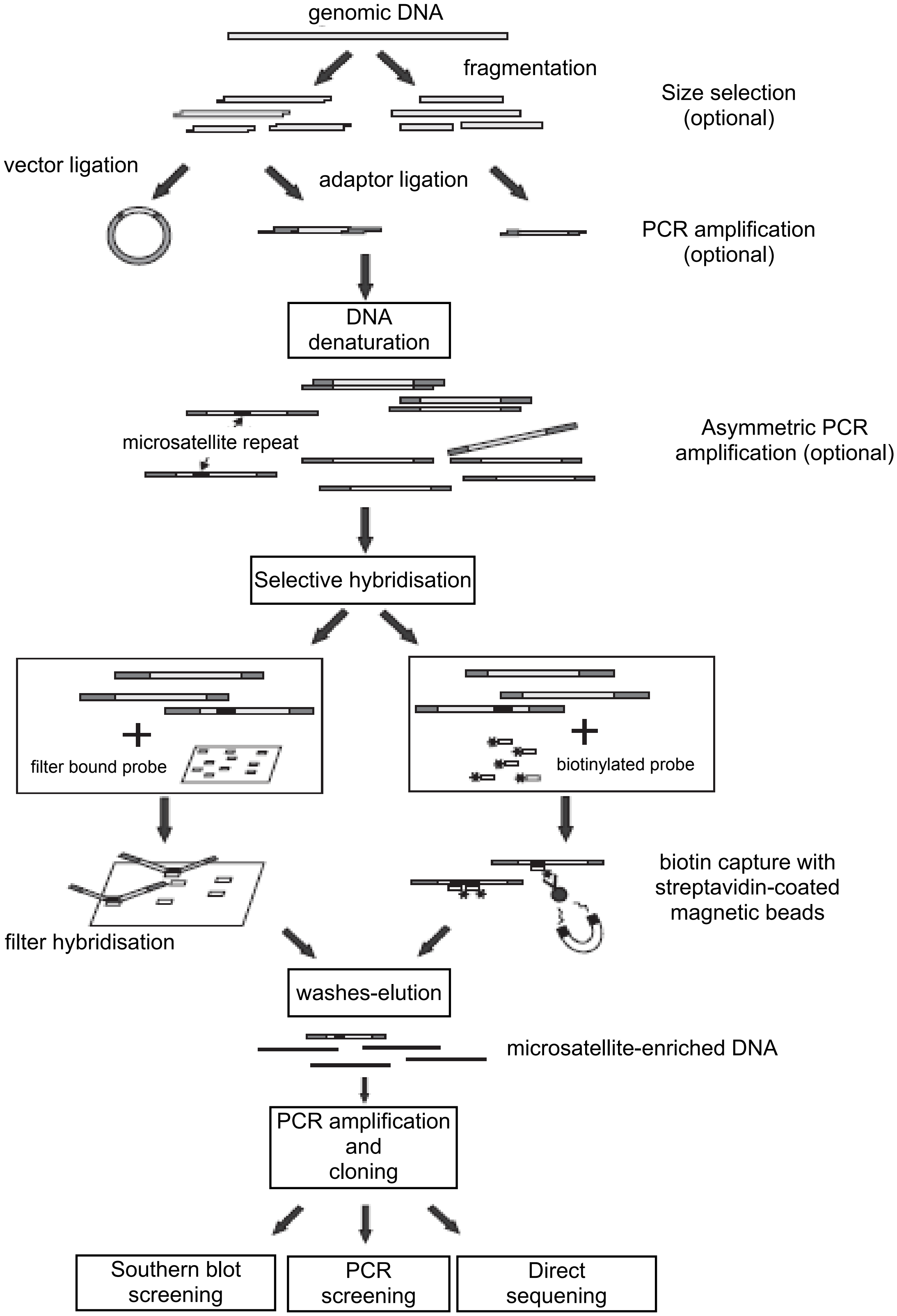


Figure 1.9: This diagram depicts the steps in generating a microsatellite enriched library for microsatellite isolation (Figure taken from Zane *et al*. 2002).

4.3. Single Nucleotide Polymorphisms (SNPs)

A new generation of molecular marker, the Single Nucleotide Polymorphism (SNP), is rapidly gaining popularity and is predicted to surpass microsatellites in utility in future. This is mostly due to the potential of adapting SNP marker systems for high throughput technologies like micro-arrays, DNA chips and mass spectrometry (Landegren *et al.* 1998; Beuzen *et al*. 2000; Rengmark *et al*. 2006). SNPs also demonstrate several other advantages that may deem them superior to the currently popular microsatellite: They are mutationally more stable and thus their inheritance conform more strictly to Mendelian expectations, also increasing their resolving power by being less prone to homoplasy and thus conforming more readily to population genetics theory (Beuzen 2000). Therefore, SNPs allows for more comprehensive evolutionary, population and pedigree deductions (Werner *et al*. 2004; Rengmark *et al*. 2006). Furthermore SNPs constitutes the majority of variation in the genome, coding and non-coding, constituting up to 90% of all polymorphisms (Collins *et al*. 1998). This characteristic allows researchers to negate SNP's greatest flaw: their bi-allelic nature, thus having low PIC values. But on average, screening 30-50 SNPs will provide equal information content to 10-15 microsatellites, depending on SNP heterozygosity (Beuzen 2000; Aitken *et al.* 2004). High frequencies (one SNP every 100-1000bp, depending on species and genomic region) also allow for denser marker maps to be compiled, aiding in association studies, gene detection, mapping of quantitative trait loci and phylogenetic comparisons (Morin *et al*. 2004; Tsang *et al*. 2005; Rengmark *et al*. 2006). Being frequent in coding regions or regulatory elements of genes, SNPs have increased probability of either conferring functional effect or being closely linked to a functional polymorphism or possibly being under selective pressures. This sheds light on gene identity, function, genotype-phenotype correlations and evolutionary forces acting on populations (Pariset *et al*. 2006; Rengmark *et al*. 2006). The robustness of SNPs makes it possible to repeat results across laboratories, thus facilitating the creation of standardised digital DNA signatures for individual animals or taxa identification (Vignal *et al*. 2002; Werner *et al*. 2004). Single nucleotide polymorphisms also minimize the effect of genotyping errors, which increase their exclusion power during parentage testing and individual identification (Weller *et al*. 2006). As a bi-allelic system, the most frequent error is the wrongful assignment of homo- and heterozygotes. In the case of the multi-allelic microsatellites the creation of a "new allele" due to genotyping artefacts is also a possibility (Vignal *et al*. 2002).

Single nucleotide polymorphisms are single nucleotide base pair positions that differ between the genomes of individuals. Generally a base pair variant is considered a SNP if the minor allele frequency is greater than 0.01 (Brookes 1999). This base pair variation is mostly introduced into genomes spontaneously, by means of replication error. The DNA polymerase enzyme incorrectly inserts a nucleotide approximately every 100 000 base pairs of which the majority are corrected by the proof reading mechanism to create an overall mutation rate of 10^{-7} – 10^{-9} per generation per locus (Crow 1993; Klug and Cummings 2003). As such SNPs may theoretically also be tri- or tetra-allelic; however this is rarely seen in nature, probably due to the low mutation rates. When a purine to purine/pyrimidine to pyrimidine substitution occurred the SNP is classified as a transition. On the other hand if a purine to pyrimidine/pyrimidine to purine substitution occurred the SNP is classified as a transversion (Brookes 1999). Transition vs. transversion ratio should be one to one; however transitions seem to more common, up to 1.7 times more in mammals and four times more in birds. It has been postulated that this bias is a consequence of high rates of spontaneous deamination of cytosine; leading to the overrepresentation of C to T/T to C transitions (Vignal 2002). Furthermore, SNPs that are located in coding regions may be categorised in one of four classes: synonymous (nucleotide substitution does not alter the amino acid called for by the codon), non-synonymous (codon is altered to call for a different amino acid), non-sense (codon is altered to create a premature stop codon) and read-through (stop codon is altered to call for an amino acid) (Brown 2002).

The principal method for SNP discovery is based on the locus-specific comparisons of sequence variation amongst diverse, unrelated individuals. This is generally termed amplicon resequencing. PCR primers are designed and genomic regions are amplified, sequenced and aligned. A variety of strategies exist depending on the template sequence origin. For animal species with well characterised, sequenced genomes or genes, a targeted gene approach for SNP discovery may be followed; specifically targeting exons or functional regions of the 5'/3' UTRs or introns (non-coding regions will readily demonstrate higher levels of polymorphism than coding regions) (Rafalski 2002a, b; Vignal *et al*. 2002; Ganal *et al*. 2009). Most species will however not have these genomic frameworks to aid SNP discovery. Thus as is the case with microsatellites, *de novo* sequence generation will be preceded by DNA library construction. Generally anonymous genomic fragments will then be selected for resequencing. This method could, however, be adapted to a form of gene targeting if cDNA libraries are constructed and the identity of genes determined by similarity searches via public databases (Vignal *et al*. 2002; Ganal *et al*. 2009). An alternative is to use sequence information on primer annealing sites of closely related species where a degree of sequence conservation is assumed. Thus homologous segments in non-focal species can be amplified (Slate *et al.* 2009).

Bioinformatic analysis of sequenced clones or EST data (self generated or downloaded from databases) has become a popular method, permitting that sufficient sequence redundancy is present and that libraries are constructed from multiple unrelated individuals (e.g. Hayes *et al.* 2007b,c; Quilang *et al.* 2007; Souche *et al.* 2007). The process of sequence analysis to detect SNPs constitutes a pipeline of contig assembly and sequence alignment, contig identification (if possible), and scoring sequence variation. Such identification of candidate polymorphisms is then validated via amplicon resequencing. Another method, a consequence of the EST boom, is SNP hybridisation arrays. Such arrays evaluate sample DNA hybridisation patterns to synthetic oligonucleotides generally derived from EST fragments. This method proves advantageous for several reasons: it allows for simultaneous analysis of many fragments, it facilitates gene expression studies and may be employed in comparative genomics (Ganal *et al*. 2009). In recent years the development of next generation high throughput sequencing technology has raised hopes for the fast and low cost discovery of large numbers of SNPs (Mardis 2008; Shendure and Ji 2008). Indeed, Novaes *et al.* (2008) demonstrated the success of such technologies for SNP discovery in the uncharacterised genome of *Eucalyptus grandis*.

A plethora of SNP genotyping techniques are in existence ranging from elementary gel banding pattern analysis to elaborate technologies for high throughput automated systems. These are thoroughly reviewed elsewhere (Landegren *et al*. 1998; Gut 2001; Syvänen 2001, 2005). Nonetheless, SNP genotyping systems rely on two components: Means of discriminating alleles and methods of analysis/visualisation. Allele discrimination requires the PCR amplification of template DNA, generally followed by either hybridisation, to allele-specific oligonucleotides, or enzymatic treatment. Gel-based analysis methods were one of the first means of visualisation, e.g. PCR-RFLPs or single strand conformational polymorphisms (SSCP) and are still in use where high throughput is not necessary. However with accumulating polymorphism data newer technologies include: hybridisation systems such as fluorescent reader analysis (e.g. realtime PCR high resolution melt analysis) and micro-arrays (DNA chip), and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Nonetheless amplicon sequencing remains the most direct form of SNP genotyping and the "gold standard" to which all other techniques are measured (Gut 2001). Ultimately, the choice of genotyping method relies on the research question and the capabilities of the laboratory.

In summary, SNPs promise to be the most versatile molecular marker. However, the wide use of SNPs is still prohibited by the lack of characterised SNPs in genomes, other than that of humans and model organisms (Seddon *et al*. 2005), the presence of rare alleles (frequencies less than 0.01) that lowers the information content (an allele frequency of 0.5 for each allows for optimal information content) (Vignal *et al*. 2002) and technical difficulties in the development and cost of high throughput technologies (Beuzen *et al*. 2000). Of these, the development of technology is probably the greatest hurdle. Even though there are a wide variety of methods, many of these remain to be optimised for individual research questions (Gut 2001). There is also a growing need for greater throughput, lower costs for genotyping, all while retaining high specificity (Landegren *et al.* 1998).

5. Molecular Markers and Abalone Culture

5.1. An Introduction

The exact history of animal domestication remains an enigma; it is not clear how man first made the decision to keep animals (for a review: Diamond 2002). Even so, the need for man to domesticate aquatic species to the extent of what is seen today in terrestrial animals such as dogs, fowl, cattle etc., have never been achieved, with the exception of goldfish and other carp species that have been domesticated by the ancient Chinese, Japanese and Romans (Dunham 2004; Komiyama *et al.* 2009). This could be in part due to the common misconception that the oceans and their resources are vast and infinite. However, in modern times the collapse of fisheries globally, has changed this perception and sparked aquaculture endeavours to supply the ever popular source of human nourishment.

The key focus of such aquaculture endeavours is to create domestic strains of economically important animals that will outperform their wild progenitors in terms of production traits (Hulata 2001). In doing so, there is no refuting the integral role that genetics plays in the holistic management of these animal populations. In fact the management and manipulation of genetic variation is the primary concern of animal domestication and breeding (Bourdon 2000), as genetic factors are the major biological causes responsible for observable phenotypic variation. In abalone culture this is no different. Traditional animal domestication relied primarily on phenotypic selection, with more sophisticated methods employed as knowledge of quantitative genetics and accompanying statistics grew during the reformation of animal breeding throughout the industrial revolution (Harris 1998). However, as noted earlier, molecular markers are the most direct way of measuring genetic variation and as such its role in domesticating abalone is paramount. Unlike traditional domestic animals of economic importance, aquaculture species' domestication will rely on a holistic integration of classical and modern techniques, of which molecular marker technology will be the most valuable for the immediate future.

Flint and Woolliams (2008) maintain that there are three goals for effective animal breeding: (1) broaden the scope and accuracy of the prediction of breeding outcomes, (2) secure animal welfare by preventing the introduction of deleterious genetic effects, and (3) the effective management of genetic resources and diversity. Requirements to reach such goals are: the establishment of (1) definite breeding objectives within the context of the environment, (2) robust models for the extent of genetic variation for important traits and, (3) a thorough understanding of the molecular genetic processes that underpin the genotype-phenotype correlation. In the light of this, the status and future of the abalone domestication and the use of molecular marker technology in South Africa may be assessed.

5.2. Establishing a Founder Population

In accord with the key equation for animal breeding, the rate of genetic change in a population is directly proportional to the accuracy and intensity of selection and genetic variation and inversely proportional to the generation time (Bourdon 2000). Thus, the first step in the creation of a genetically enhanced abalone strain is to establish a founder population representing the maximum genetic diversity of the wild population (in South Africa the current broodstock are mostly wild caught animals). This will, in turn, form the baseline population for downstream artificial selection. A thorough understanding of the genetic history and structure of natural stocks is thus essential. Molecular markers have long been used as a means to differentiate management units for fishery stocks, identifying population structure, gene flow, as well as taxon discrimination and phylogenetic relationships (Sweijd *et al*. 2000; Garant and Kruuk 2005; Chistiakov *et al*. 2006; Wenne *et al.* 2007; Hauser and Seeb 2008). Population genetic studies have been conducted on the wild stock of a number of abalone species (Sekino *et al.* 2005; Gruenthal and Burton 2008; Miller *et al*. 2009), including the South African endemic, *H. midae* (Evans *et al.* 2004a; Bester-Van der Merwe 2009). The majority of these studies were based on neutral type II markers, particularly microsatellites, thus elucidating demographic and historic processes. However in recent times more emphasis is being placed on understanding local adaptation and selective forces and thus an ever-growing need for type I markers to be employed in such studies (Vasemägi 2005; Wenne *et al.* 2007; Namroud 2008).

A loss of genetic diversity in the founder population leads to a loss of genetic variation that could have been capitalised on during subsequent selection programs for traits of contemporary and future importance (Bourdon 2000; Flint and Woolliams 2008). It is therefore advised that the founder population consists of a sufficient number of individuals that will ensure all allelic variation at random loci with allele frequencies greater than 0.05, is captured with 95% certainty (Marshall and Brown 1975). Hayes *et al.* (2006) demonstrated the use of molecular markers to maximise genetic diversity for aquaculture selective breeding schemes, by minimising kinship amongst broodstock members and selecting broodstock to maximise population heterozygosity.

5.3. Hatchery Management

The establishment of the broodstock also simulates a bottleneck effect. The crucial factor to note here is the pronounced effect of genetic drift in small populations, leading to chance losses of potentially advantageous alleles or the fixation of detrimental alleles (Roodt-Wilding 2007). This again illustrates the importance of maximising founder population diversity and ensuring the survival of broodstock animals. A further consequence of reduced population size is the increased probability of inbreeding. The number of individuals harbouring advantageous alleles [causing differential effects on the fertility and fecundity of broodstock members (Elliott 2000)] may be few. Thus when animals are retained for subsequent breeding, the probability of these being related is higher than that expected under natural conditions. This is further exacerbated by high fecundity and broadcast spawning of abalone (Davis and Hetzel 2000; Lemay and Boulding 2009). Both of these phenomena, genetic drift and inbreeding, reduce genetic diversity and genetic erosion has been reported for numerous hatchery reared abalone (Smith and Conroy 1992;

Evans *et al.* 2004b; Li Q *et al.* 2004, 2007; Hara and Sekino 2007), including South Africa's perlemoen (Evans *et al.* 2004b; Slabbert *et al*. 2009).

Thus management of hatcheries should be geared at ensuring long term retention of genetic diversity that will ensure sustainable animal breeding and long term genetic gains through artificial selection (Rauw *et al.* 1998; Gamborg and Sandøe 2005; Jensen and Andersson 2005; Cardellino and Boyazoglu 2009). This can generally be done by minimising the breeding of relatives and maximising effective population size through spawning as many animals as possible, equalling the sex ratio and ensuring even family sizes (Doyle *et al.* 2001; Taniguchi 2003). These parameters can accurately be estimated by evaluating pedigree data. However, housing individual families are impractical in mass spawning, mixed family commercial settings and tag losses make it difficult to track individual animals (Kube *et al.* 2007). Nonetheless, parentage assignment can be conducted using molecular markers as molecular signatures for individual/family identification and subsequently inferring pedigrees. Furthermore, from such pedigrees, family performances and trait heritabilities for genetic correlations and selective breeding programmes may also be deduced (Mousseau *et al*. 1998; Vandeputte *et al.* 2005; Lucas *et al.* 2006; Kube *et al.* 2007; Gheyas *et al*. 2009). As such, management decisions can be made in terms of the number of individuals contributing to production and which animals should be retained for subsequent mating (Jackson *et al.* 2003; Sekino *et al.* 2003; Jerry *et al.* 2006; Roodt-Wilding and Slabbert 2006; Wenne *et al*. 2007; Lemay and Boulding 2009).

5.4. Genetic Improvement

The main aim of the genetic improvement of animal populations is to increase the number of individuals with favourable genotypes. This is achieved by only selecting those animals with the best genotypes to contribute to the gene pool; thereby causing the phenotypic value to deviate from the mean of the previous generation in such a manner as to achieve the breeding objective. Alternatively stated, genetic improvement of a species for agriculture/aquaculture is the manipulation and exploitation of the inherent genetic variation in view of gaining production value (Elliott 2000). Classical animal breeding techniques estimate the breeding value of individual animals via inferences regarding its genotype by evaluating its own performance, that of its ancestors (pedigree data) and that of its offspring (progeny data); using statistical models for quantitative inheritance such as the selection index procedure and BLUP (best linear unbiased prediction) (Harris 1998; Flint and Woolliams 2008). In regards to abalone culture in South Africa where phenotypic selection (breeding value of an individual is estimated based only on its own performance) is mainly practised, the lack of more sophisticated breeding systems is due to a lack of sufficient data. In the current production system, pedigrees rarely stretch further than the F_1-generations and record keeping of progeny performance is only now being initiated.

The South African abalone aquaculture industry currently has one primary breeding objective namely an increase in the mean growth rate of cultured abalone, thereby increasing production output rates. However, limited knowledge is available on genetic contributors to this trait, consequently also a lack in sufficient genetic models and knowledge of molecular constituents. The value of molecular markers in this instance is undeniable: Molecular markers are routinely used to construct genetic linkage maps (Baranski *et al.* 2006; Sekino and Hara 2007). These maps, in turn, serve as the basis of mapping genetic information and identifying and positioning quantitative trait loci and genes associated to important traits (Davis and Hetzel 2000; Massault *et al*. 2008). This sheds light on molecular genetic elements governing phenotypic performance. Incorporating this into a breeding programme, referred to as marker assisted selection, increases the accuracy of breeding value prediction, as breeding value is directly estimated from the observable genotype and not, as per usual, via the phenotype (Dekkers and Hospital 2002). This holds particular merit where desirable phenotypes are only expressed at a later age (allowing early detection of genetically superior animals) or detectable in one sex only (thus the performance of offspring of the opposite gender to the parent may be deduced by parental scrutiny) (Dekkers 2004). Quantitative trait loci associated to growth-related traits have been mapped for *H. discus hannai* (Liu *et al.* 2007) and *H. rubra* (Baranski *et* *al.* 2008) and Hayes *et al.* (2007a) demonstrated the feasibility of incorporating marker assisted selection into a BLUP model for abalone. However for *H. midae* such data is not yet available, although trials are in progress.

5.5. Trade and Regulation

As several domesticated strains and/or species of abalone are introduced into the market a means of discriminating between different produce (species/strain identification), often after processing, will become more important (Hayes *et al.* 2005; Kitaoka *et al.* 2008). Firstly, this will ensure that illegitimate products (e.g. wild harvested animals) do not enter the market. Further, a method of quality control for product labelling to hold producers liable for the products that they market (tracing origin of produce) must be established, thereby ensuring that the consumer receives goods of such a high standard as what was paid for. Lastly it affords the producer a means of protecting their intellectual property. The development of specialised strains will be a considerable investment for aquaculture enterprises and as market competitiveness and rivalry increase, so too will the drive to protect investments. Thus to prevent the unlawful use of developed stock, a measure of identification and ownership is vital (Scholtz and Mamabolo 2006; Ogden and Weigel 2007; Tvedt *et al*. 2007). As each species or strain will possess unique DNA polymorphisms, these can be used to create molecular identikits for individual strains/species available on the market; thus simplifying the traceability of produce beyond processing. The aforementioned provides a means for forensic investigation, law enforcement and property protection.

5.6. Conclusion

Molecular markers are without a doubt a fundamental part of a commercial abalone aquaculture enterprise. At present molecular marker application is focused on matters of management and more specifically improvement of cultured stock. The sheer body of literature on the subject of molecular breeding advocates the merit of such application (e.g. Buitkamp and Epplen 1996; Meuwissen and Goddard 1996; Beuzen *et al*. 2000; Davis and Hetzel 2000; Elliot 2000; Dekkers and Hospital 2002; Andersson and Georges 2004; Dekkers 2004; Collard *et al*. 2005; Williams 2005; Kadarmideen *et al.* 2006; Allan and Smith 2008).

6. Aims and Objectives

The ever-expanding body of sequence information in public databases creates a powerful resource for the development of type I molecular markers, associated with genes, which is the primary aim of this investigation. From an animal production viewpoint, type I molecular markers have a higher likelihood of conferring phenotypic effects or being linked to a causative variant, because of the close marker-gene association. Therefore these molecular markers are of greater value for quantitative trait loci (QTL) identification and mapping, as well as application in marker assisted selection (MAS). Traditional development of type I markers relies on technically demanding laboratory techniques for mRNA extraction and cDNA library construction. On the contrary, a wealth of sequence information on related Haliotids and model organisms are readily available, thus facilitating the transfer of sequence information from well characterised genomes to the less characterised genome of *H. midae*.

Three strategies will be followed to reach the aim of this investigation: Firstly, a set of well characterised type II microsatellite markers have previously been developed for *H. midae* (Bester *et al*. 2004; Slabbert *et al*. 2008, in press, in prep.). The objective would be to search for possible similarity between the microsatellite flanking regions and genes in public databases to convert these markers from type II to type I markers bioinformatically. Secondly, *H. rubra* and *H. discus* are two abalone species that could be viewed as “marker-rich” species, in comparison to the “marker-poor” *H. midae*. These have published linkage maps (Baranski *et al.* 2006; Sekino and Hara 2007) that consist of numerous microsatellite markers (type II). As with the *H. midae* microsatellites, markers from *H. rubra* and *H. discus* will be bioinformatically converted to type I markers and their transferability to perlemoen tested. This strategy has the added advantage of being the first steps of evaluating regions of synteny between the three economically important abalone species. Finally, the numerous number of Haliotid EST data allows for the investigation of the prevalence of microsatellites in expressed sequences. Also with sufficient redundancy it allows for the *in silico* detection of single nucleotide polymorphisms. Both EST-microsatellites and EST-SNPs will be tested for transferability to the South African endemic. With this a greater understanding of polymorphism distribution and conservation of marker flanking sequences in abalone coding regions may be gained.

References

1. Aitken N, Smith S, Shwartz C, Morin PA (2001) Single nucleotide polymorphisms (SNP) discovery: a targeted-gene approach. Mol Ecol 13: 1423-1431.
2. Allan MF, Smith TPL (2008) Present and future applications of DNA technologies to improve beef production. Meat Sci 80: 79-85.
3. Altstsatt JM, Ambrose RF, Engle JM, Haaker PL, Lafferty KD, Raimondi PT (1996) Recent declines in black abalone *Haliotis cracherodii* on the mainland coastal of central California. Mar Ecol Prog Ser 142: 185-192.
4. Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet 2: 130-138.
5. Andersson L, Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet 5: 202-212.
6. Antunes A, Gharbi K, Alexandrino P, Guyomard R (2006) Characterisation of *transferrin*-linked microsatellites in brown trout (*Salmo trutta*) and Atlantic Salmon (*Salmo salar*). Mol Ecol Notes 6: 547-549.
7. Bachtrog D, Weiss S, Zangert B, Brem G, Schlotterer C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Mol Biol Evol 16: 602-610.

8. Balmford A, Moore JL, Brooks T, Burgess N, Hansen LA, Williams P, Rahbek C (2001) Conservation conflicts across Africa. Science 291: 2616-2619.

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

10. Baranski M, Rourke M, Loughnan S, Hayes B, Austin C, Robinson N (2008) Detection of QTL for growth rate in the blacklip abalone (*Haliotis rubra* Leach) using selective DNA pooling. Anim Genet 39: 606-614.

11. Barkai R, Griffiths CL (1986) Diet of the South African abalone *Haliotis midae.* S Afr J Mar Sci 4: 37-44.

12. Bester AE, Slabbert R, D'Amato ME (2004) Isolation and characterisation of microsatellite markers in South African abalone (*Haliotis midae*). Mol Ecol Notes 4: 618-619.

13. Bester-Van der Merwe AE (2009) Population genetic structure and demographical history of South African abalone, *Haliotis midae,* in a conservation context. Unpublished PhD thesis, Stellenbosch University, South Africa.

14. Beuzen ND, Stear MJ, Chang KC (2000) Molecular markers and their use in animal breeding. Vet J 160: 42-52.

15. Bieler R (1992) Gastropod phylogeny and systematics. Annu Rev Ecol Syst 23: 311-338.

16. Borevitz JO, Chory J (2004) Genomic tool for QTL analysis and gene discovery. Curr Opin Plant Biol 7: 132-136.

17. Bourdon RM (2000) Understanding Animal Breeding 2nd ed. Prentice-Hall. pp. 198-226.

18. Britz PJ, Lee B (2009) AISA Aquaculture Benchmarking Survey. Volume 1: Primary Production. A report for the Aquaculture Institute of South Africa and SwissContact, produced by Enviro-Fish Africa (Pty.) Ltd. 73p.

19. Brookes AJ (1999) The essence of SNPs. Gene 234: 177-186.

20. Brown LD, Murray ND (1992) Genetic relationships within the genus *Haliotis.* In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds)

21. Brown TA (2002) Genomes 2nd ed. Bios Scientific Publishers. pp. 428-430.

22. Brumfield RT, Beerli P, Nickers DA, Edwards SV (2003) The utility of single nucleotide polymorphism in inferences of population history. Trends Ecol Evol 18: 249-256.

23. Buitkamp J, Epplen JT (1996) Modern genome research and DNA diagnostics in domestic animals in the light of classical breeding techniques. Electrophoresis 17: 1-11.

24. Cardellino RA, Boyazoglu J (2009) Research opportunities in the field of animal genetic resources. Livest Sci 120: 166-173.

25. Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Phys B 126: 455-476.

26. Changè D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio J-M, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109: 1204-1214.

27. Chen H-L, Yang H-S, Huang R, Huai T (2006) Transfer of a foreign gene to Japanese abalone (*Haliotis diversicolor supertexta*) by direct testis-injection. Aquaculture 253: 249-258.

28. Cheng L, Liao X, Yu X, Tong J (2007) Development of EST-SSR by an efficient FIASCO-based strategy: A case study in rare minnow (*Gobiocyrpis rarus*). Animal Biotech 18: 143-152.

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

30. Cnaani A, Ron M, Hulata G, Seroussi E (2002) Fishing *in silico*: searching for tilapia genes using sequences of microsatellite DNA markers. Anim Genet 33: 474-476.

31. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196.

32. Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8: 1229-1231.

33. Cook P (1998) The current status of abalone farming in South Africa. J Shellfish Res 17: 601-602.

34. Crow FJ (1993) How much do we know about spontaneous human mutation rates. Environ Mol Mutagen 21: 122-129.

35. Cruz F, Pérez M, Presa P (2005) Distribution and abundance of microsatellites in the genome of bivalves. Gene 346: 241-247.

36. Davis GP, Hetzel DJS (2000) Integrating molecular genetic technology with traditional approaches for genetic improvement in aquaculture species. Aquac Res 31: 3-10.

37. Day E, Branch GM (2000) Relationships between recruits of abalone *Haliotis midae*, encrusting corallines and the sea urchin *Parechinus angulosus*. S Afr J Mar Sci 22: 137-144.

38. De Beer M (2004) Induction of triploidy in the South African abalone, *Haliotis midae*, by the use of hydrostatic pressure. Unpublished MSc Thesis, Stellenbosch University, South Africa.

39. Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106: 912-922.

40. Degnan SM, Imron, Geiger DL, Degnan BM (2006) Evolution in temperate and tropical seas: Disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae). Mol Phylogenet Evol 41: 249-256.

41. Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies and Lessons. J Anim Sci 82 (E Suppl): E313-E328.

42. Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3: 22-32.

43. Diamond J (2002) Evolution, consequence and future of plant and animal domestication. Nature 418: 700-707.

44. Dichmont CM, Butterworth DS, Cochrane KL (2000) Towards adaptive approaches to management of the South African abalone *Haliotis midae* fishery. S Afr J Mar 22: 33-42.

45. Dixon CD, Day RW, Huchette SMH, Shepherd SA (2006) Successful seeding of hatchery-produced juvenile Greenlip abalone to restore wild stocks. Fish Res 78: 179-185.

46. Dodgson JB, Cheng HH, Okimoto R (1997) DNA marker technology: a revolution in animal genetics. Poultry Sci 76: 1108-1114.

47. Doyle RW, Perez-Enriquez R, Takagi M, Taniguchi N (2001) Selective recovery of founder genetic diversity in aquacultural broodstock and captive, endangerd fish populations. Genetica 111: 291-304.

48. Dunham RA (2004) Aquaculture and Fisheries Biotechnology: Genetic Approaches. CABI Publishing. pp. 1-6, 22-53.

49. Dunstan GA, Elliot NG, Appleyard SA, Holmes HH, Conod N, Grubert MA, Cozens MA (2007) Culture of triploid abalone (*Haliotis laevigata* Donovan) to market size: Commercial implications. Aquaculture 271: 130-141.

50. Ebert EE (1992) Abalone aquaculture: a North American regional review. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 570-582.

51. Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16: 551-558.

52. Ellegren H (2004) Microsatellite: simple sequences with complex evolution. Nat Rev Genet 5: 435-445.

53. Elliott NG (2000) Genetic improvement programmes in abalone: what is the future? Aquac Res 31: 51-59.

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

55. Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliott NG (2004a) Population genetics structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. Mar Ecol Prog Ser 270: 163-172.

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

57. Farber CR, Medrano JF (2003) Putative *in silico* mapping of DNA sequences to livestock genome maps using SSLP flanking sequences. Anim Genet 34: 11-18.

58. Ferguson A, Taggart JB, Prodöhl PA, McMeel O, Thompson C, Stone C, McGinnity P, Hyens RA (1995) The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo.* J Fish Biol 47: 103-126.

59. Fleming AE, Hone PW (1996) Abalone Aquaculture. Aquaculture 140: 1-4.

60. Flint APF, Woolliams JA (2008) Precision animal breeding. Philos T Roy Soc B 363: 573-590.

61. Fraser LG, Harvey CF, Crowhurst NR, De Silva HN (2004) EST derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108: 1010-1016.

62. Gamborg C, Sandøe P (2005) Sustainability in farm animal breeding: a review. Livest Prod Sci 92: 221-231.

63. Ganal MW, Altmann T, Röder M (2009) SNP identification in crop plants. Curr Opin Plant Biol 12: 211-217.

64. Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters in the wild. Mol Ecol 14: 1843-1859.

65. Garza AES, Bernal RS (1992) Development and present status of abalone culture in Mexico. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 538-546.

66. Geiger DL (1999) A total evidence cladistic analysis of the Haliotidae (Gastrooda, Vetigastropoda). Unpublished PhD Thesis, University of Southern California, USA.

67. Geiger DL (2000) Distribution and biogeography of the *Haliotidae* (Gastropoda: Vetigastropoda) world-wide. Boll Malacol 35: 57-120.

68. Genade AB, Hirst AL, Smit CJ (1988) Observations on the spawning, development and rearing of the South African abalone *Haliotis midae* Linn. S Afr J Mar Sci. 6: 3-12.

69. Gheyas AA, Woolliams JA, Taggart JB, Sattar MA, Das TK, McAndrew BJ, Penman DJ (2009) Heritability estimation of silver carp (*Hypophthalmichthys molitrix*) harvest traits using microsatellite based parentage assignment. Aquaculture 294: 187-193.

70. Girard P, Angers B (2008) Assessment of power and accuracy of methods for detection and frequency-estimation of null alleles. Genetica 134: 187-197.

71. Gruenthal KM, Burton RS (2008) Genetic structure of the California Black abalone (*Haliotis cracherodii* Leach, 1814) a candidate for endangered species status. J Exp Mar Biol Ecol 355: 47-58.

72. Gupta PK, Rustgi S (2004) Molecular markers from transcribed/expressed regions of the genome in higher plants. Funct Integr Genomics 4: 139-162.

73. Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. Hum Mutat 17: 475-492.

74. Gutierrez-Gonzalez JL, Perez-Enriqez R (2005) A genetic evaluation of stock enhancement of blue abalone *Haliotis fulgens* in Baja California, Mexico. Aquaculture 247: 233-242.

75. Hara M, Sekino M (2007) Genetic differences between hatchery stocks and natural populations in Pacific abalone (*Haliotis discus*) estimated using microsatellite DNA markers. Mar Biotechnol 9: 74-81.

76. Harris DL (1998) Livestock improvement: Art, Science, or Industry. J Anim Sci 76: 2294-2301.

77. Haszpruner G (1988) On the origin and evolution of major gastropod groups, with special reference to *Streptoneura*. J Mollus Stud 54: 367-441.

78. Haszpruner G (1993) The Archaeogastropoda: A clade, grade or what else. Am Malacol Bull 10: 165-177.

79. Hauck M, Sweijd NA (1999) A case study of abalone poaching in South Africa and its impact on fisheries management. ICES J Mar Sci 56: 1024-1032.

80. Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. Fish Fish 9: 473-486.

81. Hayes B, Sonesson AK, Gjerde B (2005) Evaluation of three strategies using DNA markers for traceability in aquaculture species. Aquaculture 250: 70–81.

82. Hayes B, He J, Moen T, Bennewits J (2006) Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. Aquaculture 255: 573-578.

83. Hayes B, Baranski M, Goddard ME, Robinson N (2007a) Optimization of marker assisted selection for abalone in breeding programs. Aquaculture 265: 61-69.

84. Hayes BJ, Nilsen K, Berg PR, Grindflek E, Lien S (2007b) SNP detection exploiting multiple sources of redundancy in large EST collections improves validation rates. Bioinformatics 23: 1692-1693.

85. Hayes B, Laerdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Høyheim B (2007c) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. Aquaculture 265: 82-90.

86. Hobday AJ, Tegner MJ (2002) The warm and the cold: Influence of temperature and fishing on the local population dynamics of the red abalone. CalCOFI Rep 43: 74-96.

87. Hobday AJ, Tegner MJ, Haaker PL (2001) Over-exploitation of a broadcast spawning marine invertebrate: Decline of the white abalone. Rev Fish Biol Fisher 10: 493-514.

88. Hoffman LC, Swart JJ, Brink D (2000) The 1998 production status of aquaculture in South Africa. Water SA 26: 133-135.

89. Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches common sources and consequences for parental exclusion. Mol Ecol 14: 599-612.

90. Hulata G (2001) Genetic manipulation in aquaculture: a review of stock improvement by classical and modern technologies. Genetica 111: 155-173.

91. Jackson TR, Martin-Robichauda DJ, Reith ME (2003) Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. Aquaculture 220: 245-259.

92. Jensen P, Andersson L (2005) Genomics meets Ethology; A new route to understanding domestication, behaviour, and sustainability in animal breeding. Ambio 34: 320-324.

93. Jerry DR, Evans BS, Kenway M, Wilson K (2006) Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon.* Aquaculture 255: 542–547.

94. Jha S, Bawa KS (2006) Population growth, human development, and deforestation in biodiversity hotspots. Conserv Biol 20: 906-912.

95. Ju Z, Wells MC, Martinez A, Hazlehood L, Walte RB (2005) An *in silico* mining for simple sequence repeats from expressed sequenced tags of zebrafish, medaka, *Fundulus*, and *Xiphophorus. In Silico* Biology 5: 439-463.

96. Kadarmideen HN, Von Rohr P, Janss LLG (2006) From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. Mamm Genome 17: 548-564.

97. Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18: 1161-1167.

98. Khatkar MS, Zenger KR, Hobbs M, Hawken RJ, Cavanagh JAL, Barris W, McClintock AE, McClintock S, Thomson PC, Tier B, Nicholas FW, Raadsma H (2007) A primary assembly of a bovine haplotype block based on a 15,036-single nucleotide polymorphism panel genotyped in Holstein-Freisian cattle. Genetics 176: 763-772.

99. Kimura M, Crow JF (1964) The number of alleles that can be maintained in an infinite population. Genetics 49: 725-738.

100. Kitaoka M, Okamura N, Ichinose H, Goto M (2008) Detection of SNPs in fish DNA: Application of the fluorogenic ribonuclease protection (FRIP) assay for authentication of food contents. J Agric Food Chem 56: 6246-6251.

101. Klug WS, Cummings MR (2003) Concepts of Genetics 7th ed. Prentice Hall, New Jersey. pp 374-379.

102. Komiyama T, Kobayashi H, Tateno Y, Inoko H, Gojobori T, Ikeo K (2009) An evolutionary origin and selection process of goldfish. Gene 430: 5-11.

103. Kube PD, Appleyard S, Elliott NG (2007) Selective breeding greenlip abalone (*Haliotis leavigata*): preliminary results and issues. J Shellfish Res 26: 821-824.

104. Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48: 985-998.

105. Levinson G, Gutman GA (1987). Slipped-strand mis-pairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203-221.

106. Landergren U, Nilsson M, Kwok P-Y (1998) Reading bits of of genomic information: Methods for single-nucleotide polymorphisms analysis. Genome Res 8: 769-776.

107. Lemay MA, Boulding EG (2009) Microsatellite pedigree analysis reveals high variance in reproductive success and reduced genetic diversity in hatchery-spawned northern abalone, Aquaculture 295: 22-29.

108. 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.

109. 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. Aquacult Res 38: 981-990.

110. Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465.

111. Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, Function, and Evolution. Mol Biol Evol 21: 991-1007.

112. Li Y, Li X, Qin JG (2007) Triploidy induction in the Australian greenlip abalone *Haliotis laevigata* (Donovan) with cytochalasin B. Aquac Res 38: 487-497.

113. Lindberg DR (1992) Evolution, distribution and systematics of *Haliotidae*. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds)

Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 3-18.

114. Liu W, Heasman M, Simpson R (2004a) Induction and evaluation of tripliody in the Australian blacklip abalone, *Haliotis rubra*: a preliminary study. Aquaculture 233: 79-92.

115. Liu W, Heasman M, Simpson R (2004b) Evaluation of cytochalasin B (CB) treatments for triploidy induction in the blacklip abalone, *Haliotis rubra* (Leach, 1814). Aquac Res 35: 1062-1075.

116. Liu W, Heasman M, Simpson R (2004c) Optimization of triploidy induction in the blacklip abalone, *Haliotis rubra* (Leach, 1814), using 6-dimethylaminopurine. Aquac Res 35: 1076-1085.

117. Liu X, Liu X, Gou X, Gau Q, Zhao H, Zhang G (2006) A preliminary genetic linkage map of the Pacific abalone *Haliotis discus hannia* Ino. Mar Biothecnol 8: 386-397.

118. Liu X, Liu X, Zhang G (2007) Identification of quantitative trait loci for growth-related traits in the Pacific abalone *Haliotis discus hannai* Ino. Aquac Res 38: 789-797.

119. Liu ZJ, Cordes JF (2004) DNA marker technologies and their application in aquaculture genetics. Aquaculture 238: 1-37.

120. Lucas T, Macbeth M, Degnan SM, Knibb W, Degnan BM (2006) Heritability estimates for growth in the tropical abalone (*Haliotis asinina*) using microsatellites to assign parentage. Aquaculture 259: 146-152.

121. Maneeruttanarungroj C, Pongsonboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J, Li Y, Whan V, Chu K-H, Li CP, Tong J, Glenn K, Rothschild M, Jerry D, Tassanakajon A (2006) Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). Anim Genet 37: 363-368.

122. Mardis ER (2008) The impact of next generation sequencing technology on genetics. Trends Genet 24:133-141.

123. Marshall DR, Brown AHD (1975) Optimal sampling strategies in genetic conservation. In: Frankel OH, Hawkes JG (eds) Crop genetic resources

for today and tomorrow. Cambridge University Press, Cambridge, UK. pp. 53-80.

124. Massault C, Bovenhuis H, Haley C, De Koning D-J (2008) QTL mapping designs for aquaculture. Aquaculture 185: 23-29.

125. May B, Stoneking M, Wright JE (1980) Joint segregation of biochemical loci in *Salmonidae*: Linkage associations from a hybridized *Salvelinus* genome (*S. namaycush* x *S. frontinalis*). Genetics 95: 707-726.

126. McShane PE, Mercer SF, Naylor JR (1994) Spatial variation and commercial fishing of New Zealand abalone (*Haliotis iris* and *Haliotis australis*). New Zeal J Mar Fresh 28: 345-355.

127. Meglecz E, Anderson SJ, Bourguet D, Bucher R, Caldas A, Cassel-Lundhagen A, D'Acier AC, Dawson DA, Faure N, Fauvelot C, Frank P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju C, Patt A, Petenian F, Silvian J-F, Wilcock HR (2007) Microsatellite flanking region similarities among different loci within insect species. Insect Mol Biol 16: 175-185.

128. Messier W, Li S-H, Stewart C-B (1996) The birth of microsatellites. Nature 381: 483.

129. Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10: 72-80.

130. Meuwissen THE, Goddard ME (1996) The use of marker haplotypes in animal breeding schemes. Genet Sel Evol 28: 161-176.

131. Mgaya YD, Gosling EM, Mercer JP, Donlon J (1995) Genetic variation at three polymorphic loci in wild and hatchery stocks of abalone, *Haliotis tuberculata* Linnaeus. Aquaculture 136: 71-80.

132. Miller HJ, Maynard BT, Mundy CN (2009) Genetic diversity and gene flow in collapsed and healthy abalone fisheries. Mol Ecol 18: 200-211.

133. Morin PA, Luikart G, Wayne RK, The SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends Genet 19: 208-216.

134. Mousseau TA, Ritland K, Heath DD (1998) A novel method for estimating heritability using molecular markers. Heredity 80: 218-224.

135. Najmudeen TM, Victor ACC (2004) Seed production and juvenile rearing of tropical abalone *Haliotis varia* Linnaeus 1758. Aquaculture 234: 277-292.

136. Nakamura Y, Leppert M, O'Connel P, Wolff R, Holm T, Cluver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number tandem repeat (VNTR) markers for human gene mapping. Science 235: 1616-1622.

137. Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. Mol Ecol 17: 3599-3613.

138. Nauta MJ, Weising FJ (1996). Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics 143: 1021-1032.

139. Nei ZQ (1992) A review of abalone culture in China. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 592-602.

140. Norris BJ, Preston NP (2003) Triploid induction in the tropical abalone, *Haliotis asinina* (Linne), with 6-dimethylaminopurine. Aquacult Res 34: 261-264.

141. Novaes E, Drost DR, Farmerie WG, Pappas GJ jr., Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalypyus grandis,* an uncharacterised genome. BMC Genomics 9: 312.

142. O'Brien (1991) Mammalian genome mapping: Lessons and Prospects. Curr Opin Genet Dev 299: 211-217.

143. Okumura S-I, Arai K, Harigaya Y, Eguchi H, Sakai M, Senbokuya H, Furukawa S, Yamaori K (2007) Highly efficient induction of triploid Pacific abalone *Haliotis discus hannai* by caffeine treatment. Fisheries Sci 73: 237-243.

144. Ogden ER, Weigel K (2007) Can you shrinkwarp a cow? Protections for the intellectual property of animal breeding industry. Anim Genet 38: 647-654.

145. Ohta T, Kimura M (1973) The model of mutation appropriate to calculate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22: 201-204.

146. Pariset L, Cappuccio I, Joost S, D'Andrea M, Marletta D, Ajmone-Marsan P, Valentini A, The ECONOGENE Consortium (2006) Characterization of single nucleotide polymorphisms in sheep and their variation as evidence of selection. Anim Genet 37: 290-292.

147. Pérez F, Ortiz J, Zhinaula M, Gonzabay C, Calderón J, Volckaert FAM (2005) Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei, Litopenaeus stylirostris,* and *Trachypenaeus birdy.* Mar Biotechnol 7: 554-569.

148. Plagányi EE, Butterworth DS, Brandão A (2001) Toward assessing the South African abalone *Haliotis midae* stock using an age-structured production model. J Shellfish Res 20: 813-827.

149. Primmer CR, Ellegren H (1998) Pattern of molecular evolution in avian microsatellites. Mol Biol Evol 15: 997-1008.

150. Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: Models and data. Am J Hum Genet 69: 1-14.

151. Provan J, Beatty GE, Maggs CA, Savidge G (2007) Expressed sequence tag-derived microsatellites for the cool-water marine copepod *Calanus finmarchicus.* Mol Ecol Notes 7: 1369-1371.

152. Powers DA, Kirby VL, Cole T, Hereford L (1995) Electroporation as an effective means of introducing DNA into abalone (*Haliotis rufescens*) embryos. Mol Mar Biol Biotech 4: 369-375.

153. Quilang J, Wang S, Li P, Abernathy J, Peatman E, Wang Y, Wang L, Shi Y, Wallace R, Guo X, Liu Z (2007) Generation and analysis of ESTs from the eastern oyster *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. BMC Genomics 8: 157-167.

154. Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: A new class of genetic marker in cotton. The Journal of Cotton Science 8: 112-123.

155. Rafalski A (2000a) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5: 94-100.

156. Rafalski A (2000b) Novel genetic mapping tool in plant: SNPs and LD-based approaches. Plant Science 162: 329-333.

157. Ramsey L, Macaulay M, Cardle L, Morgante M, Ivanissevich SD, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in the barley. Plant J 17: 415-425.

158. Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ (1998) Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest Prod Sci 56: 15-33.

159. Raven PH, Johnson GB (2002) Biology 6th ed. McGraw-Hill, New York, pp. 571-590, 899-905.

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

161. Roodt-Wilding R (2007) Abalone ranching: a review on genetic considerations. Aquac Res 38: 1229-1241.

162. Roodt-Wilding R, Slabbert R (2006) Molecular markers to assist the South African abalone industry. S Afr J Sci 102: 99-102.

163. Sales J, Britz PJ (2001) Research on abalone (*Haliotis midae* L.) cultivation in South Africa. Aquac Res 32: 863-874.

164. Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

165. Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? Nat Rev Genet 5: 63-68.

166. Scholtz MM, Mamabolo J (2006) A developing country perspective on recent developments in animal breeders and intellectual property rights. S Afr J Anim Sci 36: 22-25

167. Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecology and conservation studies: a test in the Scandinavian wolf population. Mol Ecol 14: 503-511.

168. Seeb JE, Seeb LW (1986) Gene mapping of isozyme loci in chum salmon. J Heredity 77: 399-402.

169. Sekino M, Saitoh, K, Yamada T, Kumagai A, Hara M, Yamashita Y (2003) Microsatellite-based pedigree tracing in a Japanese flounder

Paralichthys olivaceus hatchery strain: implications for hatchery management related to stock enhancement program. Aquaculture 221: 255-263.

170. Sekino M, Saido T, Fujita T, Kobayashi T, Takami H (2005) Microsatellite DNA markers of Ezo abalone (*Haliotis discus hannai*): a preliminary assessment of natural populations sampled from heavily stocked areas. Aquaculture 243: 33-47.

171. Sekino M, Hara M (2007) Linkage maps for Pacific abalone (Genus *Haliotis*) based on microsatellite DNA markers. Genetics 175: 945-958.

172. Serapion J, Kucuktas H, Feng J, Liu Z (2004) Bioinformatic mining of type 1 microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Mar Biotechnol 6: 364-377.

173. Shendure J, Ji H (2008) Next generation DNA sequencing. Nat Biotechnol 26: 1135-1145.

174. Sin FYT, Mukherjee UK, McKenzie JC, Sin IL (1995) Electroporation of abalone sperm enhances sperm-DNA association. J Fish Biol 47 (Suppl. A): 20-28.

175. 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.* J World Aquacult Soc 39: 429-435.

176. Slabbert R, Bester AE, D’Amato MED (2009) Analysis of genetic diversity and parentage of the abalone *Haliotis midae* Linnaeus using microsatellite markers. J Shellfish Res 28: 1-7.

177. Slabbert R, Hepple J, Venter A, Nel S, Swart L, Van den Berg NC, Roodt-Wilding R (in press) Isolation and segregation of 44 microsatellite loci in the South African abalone, *Haliotis midae* L*.* Anim Genet. DOI: 10.1111/j.1365-2052.2009.02003.x.

178. Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guideline and future directions. Genetica 136: 97-107.

179. Smith CT, Elfstorm CM, Seeb LW, Seed JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol Ecol 14: 4193-4203.

180. Smith PJ, Conroy AM (1992) Loss of genetic diversity in hatchery produced abalone, *Haliotis iris*. New Zeal J Mar Fresh 26:81-85.

181. Souche EL, Hellemans B, Van Houdt JKJ, Canario A, Klages S, Reinardt R, Volckaert FAM (2007) Mining for single nucleotide polymorphisms in expressed sequence tags of European sea bass. Journal of Integrative Bioinformatics 4: 73-82.

182. Spencer BE (2002) Molluscan Shellfish Farming. Blackwell Publishing. pp. 185-202.

183. Steinberg J (2005) The illicit abalone trade in South Africa. Institute for Security Studies paper 105.

184. Streit K, Geiger DL, Lieb B (2006) Molecular phylogeny and the geographic origin of *Haliotidae* traced by haemocyanin sequences. J Mollus Stud 72: 105-110.

185. Stepto NK, Cook PA (1998) Induction of triploidy in the South African abalone using cytochalasin B. Aquacult Int 6: 161-169.

186. Sweijd NA, Bowie RCK, Evans BS, Lopata AL (2000) Molecular genetics and the management and conservation of marine organisms. Hydrobiologia 420: 153-164.

187. Syvänen A-C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev 2: 930-942.

188. Syvänen A-C (2005) Toward genome-wide SNP genotyping. Nat Genet 37: 55-60.

189. Tanigichi N (2003) Genetic factors in broodstock management for seed production. Rev Fish Biol Fisher 13: 177-185.

190. Tarr RJQ (1989) Abalone. In: Payne AIL, Crawford RJM (eds) Oceans of Life off Southern Africa. Vlaeberg Publishers, Cape Town, pp.62-68.

191. Tarr RJQ (1992) The abalone fishery of South Africa. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 438-447.

192. Tarr RJQ (1995) Growth and movement of the South African abalone *Haliotis midae*: a Reassessment. Mar Freshwater Res 46: 583-590.

193. Tarr RJQ, Williams PVG, MacKenzie AJ (1996) Abalone, sea urchin and rock lobster: a possible ecological shift that may affect traditional fisheries. S Afr J Mar Sci 17: 319-323.

194. Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is the major source of genetic variation. Nature 322: 652-565.

195. Theodorou K, Couvet D (2004) Introduction of captive breeders to the wild: Harmful or beneficial? Conserv Genet 5: 1-12.

196. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare L.*). Theor Appl Genet 106: 411-422.

197. Tong LJ, Moss GA (1992) The New Zealand culture system for abalone. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 583-591.

198. Troell M, Robertson-Anderson D, Anderson RJ, Bolton JJ, Maneveldt G, Halling C, Probyn T (2006) Abalone farming in South Africa: An overview with perspectives on kelp resources, abalone feed, potential for on-farm seaweed production and socio-economic importance. Aquaculture 257: 266-281.

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

200. Tsai H-I, Lai C-H, Yang H-S (1997) Sperm as carrier to introduce and exogenous DNA fragment into the oocyte of Japanese abalone (*Haliotis diversicolor supertexta*). Transgenic Res 6: 85-95.

201. Tsang S, Zhonghe S, Luke B, Stewart C, Lum N, Gregory M, Wu X, Subleski M, Jenkins NA, Copeland NG, Munroe DL (2005) A comprehensive SNP-based genetic analysis of inbred mouse strains. Mamm Genome 16: 476-480.

202. Tvedt MW, Heimstra SJ, Drucker AG, Louwaars N, Oldenbroek JK (2007) Regulatory options for exchange, use and conservation of animal genetic resources: a closer look at property rights issues. Animal genetic resources information (FAO report) 41:91-99.

203. Vandeputte M, Kocour M, Mauger S, Dupont-Nivet M, De Guerry D, Gela D, Vallod D, Linhart O, Chevassus B (2005) Heritability estimates for growth related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpo* L.). Aquaculture 247: 31-32.

204. Van Marle-Köster E, Nel LH (2003) Genetic markers and their application in livestock breeding in South Africa: A review. S Afr J Anim Sci 33: 1-10.

205. Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene linked polymorphisms for detecting signatures of selection in Atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22:1067-1076.

206. Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other molecular markers and their use in animal genetics. Genet Sel Evol 34: 275-305.

207. Weller JI (2001) Quatitative trait loci analysis in animals. CABI Publishing. pp. 3-14.

208. Weller JL, Seroussi E, Ron M (2006) Estimation of the number of genetic markers required for individual animal identification accounting for genotyping errors. Anim Genet 37: 387-389.

209. Wenne R, Boudry P, Hemmer-Hansen J, Lubieniecki KP, Was A, Kuase A (2007) What role for genomics in fisheries management and aquaculture. Aquat Living Resour 20: 241-255.

210. Werner FAO, Durstewitz G, Habermann FA, Thaller G, Kramer W, Kollers S, Buitkamp J, Georges M, Brem G, Mosner J, Fries R (2004) Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Anim Genet 35: 44-49.

211. Williams JL (2005) The use of marker-assisted selection in animal breeding and biotechnology. Rev scitech Off int Epiz 24: 379-391.

212. Winnepenninckx B, Backeljau T, De Wachter R (1996) Investigation of molecular phylogeny on the basis of 18S rRNA sequences. Mol Biol Evol 13: 1306-1317.

213. Winnepenninckx B, Seiner G, Backeljau T, De Wachter R (1998) Details of gastropod phylogeny inferred from 18S rRNA sequences. Mol Phylogenet Evol 9: 55-63.

214. Wood AD, Buxton CD (1996) Aspects of the feeding biology of the abalone *Haliotis midae* (Linne, 1758) on the east coast of South Africa. 1. Feeding Biology. S Afr J Mar Sci 17: 61-68.

215. Yamamoto F, Clausen H, White T, Marken J, Hakamori S (1990) Molecular genetic basis of the histo-blood group ABO system. Nature 345: 229-233.

216. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11: 1-16.

Chapter II:
A bioinformatic survey of *Haliotis midae* microsatellites

Abstract

To date, no survey has been done to investigate the distribution and frequency of microsatellites in Haliotids. It has been shown by various studies that these loci demonstrate divergent patterns of distribution between genomic regions within a particular species and between the genomes of different species. Because whole genome data is not available for the South African abalone, *Haliotis midae*, a bioinformatic method that utilises microsatellite flanking region homology is proposed here to investigate the prevalence of microsatellites in genes and transposable elements. An understanding of microsatellite genomic distribution will facilitate more efficient use and development of this popular marker-type. It was found that 17.5% and 21% of the microsatellites had gene and/or transposable element associations, respectively. The association of di- and tetranucleotides with transposable elements could explain their high genomic frequencies across the genome. CA microsatellite units were the most abundant repeat motif, but were notability underrepresented in genic regions where GAGT repeats predominate. This alludes to a possible functional role for these microsatellite motifs. The data presented here supports the previous findings of non-random distribution of microsatellite sequences.

1. Introduction

Microsatellites are ubiquitous in genomes of all eukaryotes, comprising approximately 3% of the human genome (International Human Genome Sequencing Consortium 2001) and 1.3% - 3.2% of various puffer fish species' genomes (Chistiakov *et al.* 2006); their abundance and distribution seemingly correlated to genomic evolutionary events. It was generally assumed that microsatellites demonstrated a random distribution throughout genomes that could be explained by neutral mutation, as is the case for *Drosophila malonogaster* CA motifs (Bachtrog *et al.* 1999). However, cumulating evidence would suggest a non-random pattern of microsatellite dispersal, both inter- and intragenomically (Li *et al.* 2002 and references therein). There is a clear bias for particular microsatellite motifs between species, with vertebrates and arthropods demonstrating CA excess while fungal genomes favour CG repeats and rice showing an overrepresentation of TA motifs. Such taxon specific microsatellite bias cannot be explained by regular mutational mechanisms of microsatellite evolution, and is most probably a result of dynamic protein-DNA interactions creating differential selective pressures between lineages (Tóth *et al*. 2000; Grover *et al.* 2007). Furthermore it is suggested that there is a differential in microsatellite density between different chromosomes in the same genome (Subramanian *et al.* 2003; Guo *et al.* 2009). Although dinucleotides are the predominant microsatellite motif in many genomes, they are however glaringly underrepresented in exons (Tóth *et al*. 2000). Thus, even though microsatellites are comparatively scarce in protein coding regions due to selective pressures on frame shift mutation, their high frequency in 5'/3' - untranslated and intronic sequences exceeds that which is expected based on random nucleotide composition (Zeiss *et al.* 1998; Li *et al.* 2002).

It is this systematic distribution that alludes to the possible genomic functionality of microsatellites (Chambers and MacAvoy 2000; Li *et al.* 2002, 2004; Cruz *et al*. 2005). Microsatellites have been implicated in playing a role in diverse functions including: regulation of gene expression, intron splicing, chromatin organisation and meiotic recombination (Meloni *et al.* 1998; Zeiss *et al.* 1998; Albanèse *et al.* 2001; Chiba-Falek and Nussbaum 2001; Li *et al.* 2002, 2004; Guo *et al.* 2009). Recent studies have also shown a close association between microsatellites and interspersed repetitive DNA/dispersed repetitive elements (Ramsay *et al.* 1999; Meglecz *et al.* 2004, 2007). Such repetitive elements, e.g. transposons, play an important role in genome architecture and evolution; facilitating duplications, deletions, insertions etc. (for reviews: Bennetzen 2000; Kidwell 2002; Kazazian 2004). Their association to microsatellites suggests a further function for these sequences.

The popularity and usefulness of microsatellite markers has ensured that countless numbers of these sequences have been isolated for numerous species. However the vast majority of these, including microsatellites for many aquaculture species, were developed from anonymous genomic DNA fragments (Serapion *et al*. 2004) generally via genomic enrichment protocols as explained by Zane *et al.* (2002). To date 200 polymorphic microsatellites have been developed for *Haliotis midae* (Bester *et al.* 2004; Slabbert *et al.* 2008, in press, in prep.) and is currently being employed in the abalone culture industry for parentage assignment in selective breeding programs and to assess population diversity and stratification for effective management of the wild and cultured populations (Roodt-Wilding and Slabbert 2006; Slabbert *et al.* 2008). In ongoing projects the objectives are to develop a sufficient number of molecular markers to construct a dense linkage map for downstream applications, such as QTL analysis to facilitate the genetic enhancement of cultured stock by means of marker assisted selection. However all of these markers for *H. midae* are type II molecular markers.

In recent years emphasis has been placed on the development of type I markers, associated to genes. Type I markers have higher probability of conferring phenotypic effect or being closely linked to a causal mutation (Gupta and Rustgi 2004; Liu and Cordes 2004). Therefore, from an animal production standpoint, type I markers are of greater value. Furthermore, they provide insights into genome structure and evolution and facilitate comparative genomics and synteny mapping (Liu and Cordes 2004; Serapion *et al.* 2004).

Microsatellites are flanked by unique DNA regions; these are exploited to design primer sequences for the PCR amplification of these loci. However the location of microsatellites in genic regions and other genomic features (Li *et al*. 2004) suggest that at least a comparable portion of anonymous (type II) microsatellites will locate in conserved functional sequences, such as regulatory motifs, exons and transposable elements (Ramsay *et al.* 1999; Faber and Medrano 2003). Public databases provides a powerful resource for finding sequence similarity between flanking regions of microsatellites and conserved genomic regions in other species. Sequence information in such databases has been growing exponentially since their induction.

The aim of this investigation is thus to survey the current microsatellites of *Haliotis midae* to better understand the dynamics of these genomic features in this species; in particular their association to genes and transposable elements. In the case of gene associations, type II markers may be converted to type I markers. Unlike previous studies (Herron *et al.* 1998; Farber and Medrano 2003, 2004), this study will expand the bioinformatic analyses to not only use these microsatellites as gene tags, but also to deduce possible microsatellite locality within these genes and thus gain an understanding of putative function of the associated microsatellites. Furthermore, investigating the association of microsatellite to dispersed repetitive elements may provide an indirect measure of gene associations, as many transposable element classes (e.g. SINEs) locate in high gene density regions of the genome, particularly 5'-UTR. Transposable elements have also been implicated in actively altering gene regulation and function (Medstrand *et al*. 2005). In some cases the transposable element has explicit genomic or biochemical/physiological functionality, e.g. an L1-related retrotransposon in *Drosophila melanogaster* plays a vital role in the maintenance of chromosomal telomeres (Levis *et al*. 1993) and endogenous retroviral elements are thought to contribute to normal placental development in humans (Mi *et al*. 2000).

The association of microsatellites with dispersed repetitive elements have further implications on the practicality of microsatellite isolation, which remains a taxing task. An important aspect in the development of applicable microsatellites is primer design. Many microsatellite loci isolated for *H. midae* show significant violation of Hardy-Weinberg expectations; indicative of the possible presence of null alleles (Meglecz *et al.* 2004). We have previously demonstrated that microsatellite flanking regions in *H. midae* are hyper-variable and can serve as a rich source of single nucleotide polymorphisms (SNPs) (Rhode *et al.* 2008); however the sequence variation amongst individuals complicates primer design for the respective microsatellite loci. Furthermore due to the evolutionary age of molluscs it is suggested that their genomes might be rife with repetitive elements, both tandem, e.g. microsatellites, and dispersed (Kourtidis *et al.* 2006). The association of microsatellites to the latter holds a particular challenge to primer design: If primers were designed to anneal to such repetitive elements in the flanking region of a microsatellite, the PCR reaction can yield non-specific or superfluous products; thus rendering the particular microsatellite locus inadequate or warranting redesign of primers if possible (Temnykh *et al.* 2001; Meglecz *et al.* 2004).

A greater understanding of microsatellite genomic dynamics will facilitate the refinement of future microsatellite marker development. Furthermore the conversion of type II to type I markers will aid in the identification of candidate genes, when such markers are employed for QTL analysis.

2. Materials and Methods

Microsatellite sequences were either downloaded from NCBI (www.ncbi.nlm.nih.gov), using their accession numbers (Bester *et al.* 2004; Slabbert *et al.* 2008, in press, in prep.) or retrieved from the laboratory database (Molecular Aquatic Research Group). These were then saved as a text file in FASTA format. A survey of microsatellite repeat motifs was conducted by count. The sense and anti-sense strands as well as all overlapping sequence combinations were grouped for each motif, e.g. GAGT is equivalent to AGTG and their reverse compliments CTCA and TCAC and so on. All sequences were subsequently subjected to RepeatMasker (www.repeatmasker.org/cgi-bin/WEBRepeatMasker) to mask the tandem repeat motifs by replacing the nucleotide sequence with a series of “N’s”. This was done to ensure that significant hits are due to homologous flanking sequences and not to the repeat motif. To identify gene associations the masked sequences were then used to conduct BLASTX and BLASTN (Altschul *et al.* 1990; http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches to the nr-protein and refseq_rna NCBI databases respectively. Sequences that remained anonymous were then also screened against the nr-nucleotide database. Result files were retrieved in HTML format and downloaded as text files. For the BLASTX results, hits with a Score (S) < 33 and an E-value > 2.3 were discarded (Farber and Medrano 2004); similarly for BLASTN, hits demonstrating a Score (S) < 50 and an E-value > 1e-04 were discarded (Farber and Medrano 2003). As a further stringency parameter, BLASTX and BLASTN hits demonstrating a Positives-value < 50% or an Identities-value < 70% respectively, were eliminated (all other parameters were at default). Further manual inspection was done to reject significant hits due to repeat motifs not detected by RepeatMasker. Those genes that demonstrated the most statistically significant hit to the query microsatellites were assumed to be the most probable homolog.

The BLAST alignments were used to deduce possible microsatellite position within the gene. With regards to BLASTX, if the aligned segment did not include the flanking sequences on both sides of the microsatellite, the microsatellite was assumed to be intronic or in the 5'/3' untranslated region, pending on the alignment to the subject sequence. Open reading frame analysis was done in ORF Finder (NCBI, www.ncbi.nlm.nih.gov/gorf/gorf.html) to validate this positioning for each particular microsatellite query. Only the open reading frame positioned to the corresponding segment of the original BLAST alignment was considered.

The association of microsatellites to dispersed repetitive elements was investigated by using the masked sequences to screen the Repbase database via the CENSOR program (Jurka *et al*. 2005; www.girinst.org/repbase). To further access the functional role of microsatellites associated to genes, the microsatellites were also screened against the TRANSFAC® 7.0 Public 2005 database using the AliBaba2 program (Wingender *et al.* 2000; www.gene-regulation.com) to predict whether transcription factors would bind to particular tandem repeat motifs.

3. Results

Across all *H. midae* microsatellites investigated, CA repeats were the most abundant sequence motif, at ~30%, followed by GAGT and CGTG representing ~23 % and ~8% respectively. The most abundant length motif was tetranucleotide repeats at ~47%, followed by dinucleotide and trinucleotide repeats at ~37% and ~11% respectively (Fig. 2.1, 2.2).

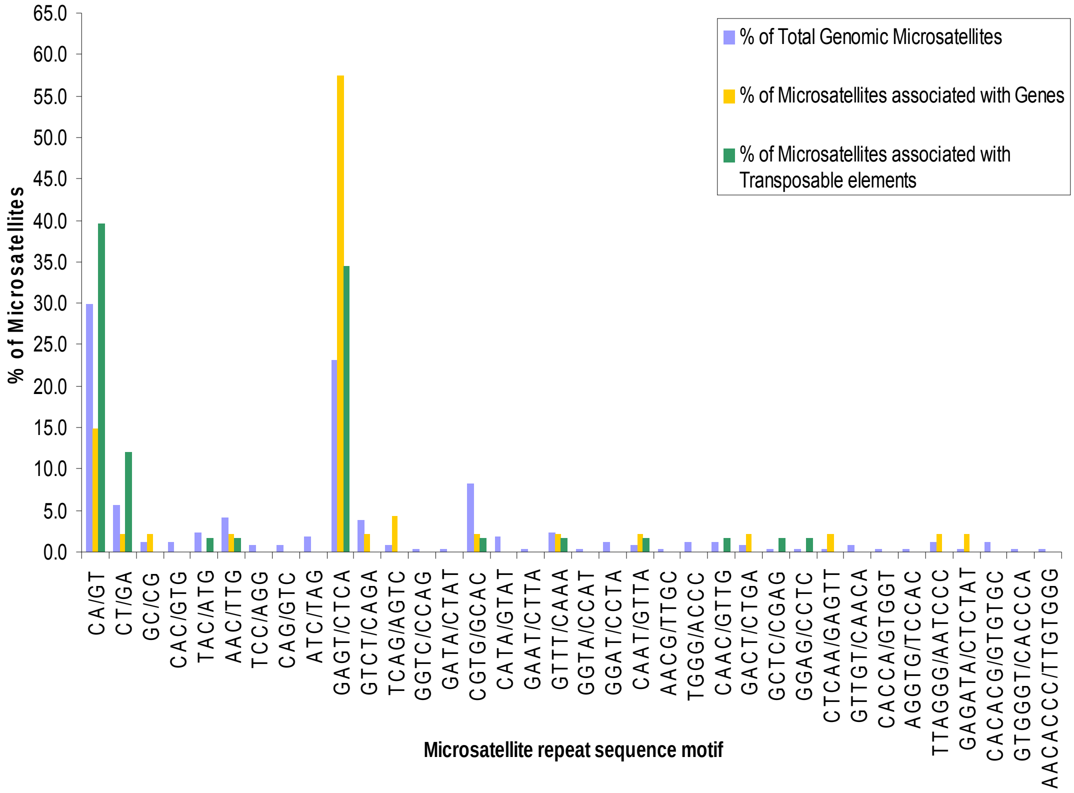


Figure 2.1: The distribution of microsatellite repeat sequence motifs across all genomic microsatellites, gene-associated microsatellites and transposon-associated microsatellites.

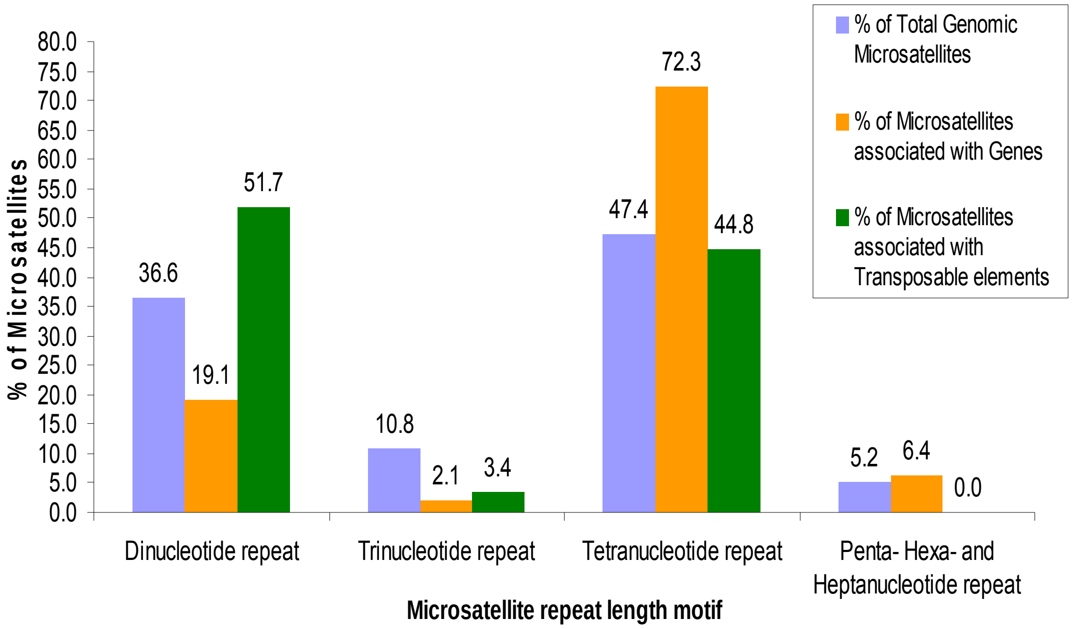


Figure 2.2: The distribution of microsatellite repeat length motifs across all genomic microsatellites, gene-associated microsatellites and transposon-associated microsatellites.

Thirty five of the 200 microsatellites surveyed (~17.5%) presented significant hits to protein and/or gene sequences in the NCBI databases. Of these, eight gave significant BLASTX hits, 30 significant BLASTN hits and three gave both significant BLASTX and BLASTN hits. Open Reading Frame and BLAST alignment analysis revealed that none of the microsatellites were placed within protein coding regions of the genes to which they aligned. Twelve microsatellites were found to be in the 5'-untranslated region (5'-UTR), 23 were intronic and three located in the 3'-UTR (Table 2.1, 2.2). Of these gene-associated microsatellites, GAGT was the most represented repeat sequence motif (~57%) and in general tetranucleotides was the most abundant length motif (~47%). This was followed by CA repeats and dinucleotides in general (Fig. 2.1, 2.2).

Table 2.1: **Microsatellite loci that produced significant BLASTX hits with S > 33 and E < 2.3.**

Microsatellite (Genbank Acc#[1])	Tandem Repeat Motif	BLASTX Hit (Genbank Acc#[1])	E-value	Score	Gene Name/ Function	Microsatellite Position
HmRS36T (DQ785753)	CTCA	Hypothetical protein *Strongylocentrotus purparatus* (XP_001183516.1)	4.0E-18	94.4	5-formyltetrahydrofolate cyclo-ligase	5'-UTR
HmRS117M (DQ785765)	(GAGT)n(GCGT)n	Vertebrate crystallin zeta *Danio rerio* (NP_001093446.1)	0.95	36.6	Quinone reductase (Energy metabolism)	Intronic
HmLCS55T (DQ993226)	GTGA	Unknown protein 6 *Haliotis diversicolor* (ABY87369.1)	0.95	36.6	Unknown (possible lysin)	Intronic
HmNR54H (EF063103)	TTAGGG	Hypothetical protein *Caenorhabditis briggsae* (XP_001670723.1)	0.65	36.6	UDP-Glucosyl transferase	Intronic
HmNR120T (EF121745)	TGAG	SH2 domain containing 3C *Danio rerio* (XP_687225.1)	3.0E-05	51.2	Signal transduction	Intronic
HmNS38T (EF367113)	TCAC	Unknown protein 6 *Haliotis diversicolor* (ABY87369.1)	0.52	37	Unknown (possible lysin)	Intronic
HmNST7T (EF455618)	CACT	Novel protein *Danio rerio* (XP_686458.2)	1.0E-06	55.8	Vertebrate inositol 1,4,5-triphosphate-3-kinase B	Intronic
Hmid4009D (GQ927128)	GT	Hypothetical protein *Branchiostoma flofidae* (XP_002586316.1)	3.0E-10	68.2	Unknown	Intronic

1 – GenBank accession number.

Table 2.2: Microsatellite loci that produced significant BLASTN hits with S > 50 and E < 1e-04.

Microsatellite (Genbank Acc#[1])	Tandem Repeat Motif	BLASTN Hit (Genbank Acc#[1])	E-value	Score	Gene Name/ Function	Microsatellite Position
HmD33 (AY303334)	(GAGT)n(AAGT)(GAGT)n	Cellulase gene *Haliotis discus hannai* (AB125892.1)	6.0E-16	89.7	Cellulose metabolism	5'-UTR
HmD30 (AY303342)	(AGTC)n(GGTC)(AGTC)n	Cellulase gene *Haliotis discus hannai* (AB125892.1)	2.0E-14	87.8	Cellulose metabolism	5'-UTR
HmG16D (DQ785744)	GTGA	Partial hemocyanin gene (H1) *Haliotis tuberculata* (AJ252741.1)	4.0E-18	100	Oxygen transporter	Intronic
HmRS83M (DQ785757)	(GTTT)n(GT)n(TTTG)n	Receptor protein tyrosine phosphatase delta *Haliotis discus* (FJ940467.1)	1.0E-17	98.7	Signal transduction	Intronic
HmRS54D (DQ785774)	CA	Lysin precursor, gene intron 4 *Haliotis rufescens* (AF076822.1)	4.0E-17	96.9	Fertilisation	Intronic

HmRS61H (DQ785776)	GAGATA	G-alpha signal transduction protein, gene intron 5'end *Haliotis rufescens* (AF070959.1)	7.0E-08	66.2	Signal transduction	Intronic
HmLCS5M (DQ825705)	(GCTA)n(ACTC)n	ATPase alpha subunit *Haliotis rubra* (AY043205)	3.0E-19	104	Energy metabolism	Intronic
HmLCS73T (DQ993219)	GAGT	Cellulase gene *Haliotis discus hannai* (AB125892.1)	5.0E-10	73.4	Cellulose metabolism	5'-UTR
HmLCS67M (DQ993222)	(GAGT)n(GT)n(GC)n	Cellulase gene *Haliotis discus hannai* (AB125892.2)	2.0E-20	107	Cellulose metabolism	5'-UTR
HmLCS55T (DQ993226)	GTGA	Lysin precursor, gene *Haliotis rufescens* (AF076824.1)	9.0E-35	172	Fertilisation	Intronic
HmNS31D (EF033333)	(GT)n(CT)(GT)n	Cellulase gene *Haliotis discus hannai* (AB125892.1)	2.0E-15	91.5	Cellulose metabolism	5'-UTR
HmDL131M (EF054867)	(AC)n(TC)n	Actin A2 gene *Haliotis iris* (AY921238.1)	4.0E-36	156	Cytoskeleton dynamics	Intronic
HmDL151T (EF054868)	CTGA	Cellulase gene *Haliotis discus hannai* (AB125892.1)	8.0E-39	169	Cellulose metabolism	5' UTR
HmDL214T (EF054871)	TGAG	Cellulase gene *Haliotis discus hannai* (AB125892.2)	9.0E-17	62	Cellulose metabolism	5' UTR
HmNR191T (EF121752)	GAGT	Actin A2 gene *Haliotis iris* (AY921238.1)	6.0E-41	176	Cytoskeleton dynamics	Intronic
HmNS38T (EF367113)	TCAC	Lysin precursor, gene *Haliotis rufescens* (AF076824.1)	2.0E-43	185	Fertilisation	Intronic
HmNS100T (EF367114)	GAGT	Cellulase gene *Haliotis discus hannai* (AB125892.1)	1.0E-18	102	Cellulose metabolism	5'-UTR
HmNS58D (EF367119)	GTT	Partial hemocyanin gene exons 1-15 (H2) *Haliotis tuberculata* (AJ297475.1)	3.0E-06	60.8	Oxygen transporter	Intronic
HmNST7T (EF455618)	CACT	Cellulase gene *Haliotis discus hannai* (AB125892.1)	6.0E-22	113	Cellulose metabolism	5' UTR
HmNR281P (EF512274)	CTCAA	Actin A2 gene *Haliotis iris* (AY921238.1)	4.0E-43	183	Cytoskeleton dynamics	Intronic
Hmid2015 (GQ927124)	GTCT	Sperm lysin *Haliotis corrugata* (FJ940473.1)	1.0E-58	235	Fertilisation	Intronic

Hmid2044T (GQ9272126)	GAGT	Hemocyanin *Haliotis diversicolor* (GQ352369.1)	8.0E-30	140	Oxygen transporter	Intronic
HLCS147T (GQ927134)	GAGT	Cellulase gene *Haliotis discus hannai* (AB125892.1)	4.0E-21	111	Cellulose metabolism	5'-UTR
Hm3A11F (GQ927135)	CA	Cytidine deaminase *Haliotis diverscolor* (EU101721.1)	1.0E-09	73.4	Catalyze the deamination of cytidine	3'-UTR
Hm2H6F (GQ927136)	CACT	Actin *Haliotis discus* (EF103363.1)	2.0E-11	78.8	Cytoskeleton dynamics	3'-UTR
HmS104 (GQ927137)	GAGT	Cellulase gene *Haliotis discus hannai* (AB125892.1)	1.0E-14	89.7	Cellulose metabolism	5'-UTR
HmR16 (GQ927138)	CAGT	Cytidine deaminase *Haliotis diversicolor* (EU101721.1)	2.0E-49	205	Catalyze the deamination of cytidine	3'-UTR
HmidPS1.374T (GU256684)	GAGT	Sperm lysin *Haliotis discus* (FJ940391.1)	7.0E-09	68	Fertilisation	Intronic
HmidPS1.588C (GU256700)	CACT	ATPase alpha subunit *Haliotis rubra* (AY_043205.1)	8.0E-08	64.4	Energy metabolism	Intronic
HmidPS1.1007C (GU256729)	$(ACTC)_nX(TCAA)_nX(CAAT)_n$	H1 hemocyanin *Haliotis tuberculata* (AJ252741.1)	4.0E-05	55.4	Oxygen transporter	Intronic

1 – GenBank accession number.

Forty two microsatellite's flanking regions (21%) showed significant hits to known dispersive repetitive elements in the Repbase database with similarities ranging from ~68% to 90% (Table 2.3). Here again di- and tetranucleotides were most prominent (~52% and 45% respectively); CA and GAGT were also the dominant sequence motifs (~39% and 34% respectively) (Fig. 2.1, 2.2).

Table 2.3: Microsatellite loci that produced significant hits to known dispersed repetitive elements in the Repbase database (Jurka *et al*. 2005).

Microsatellite (Genbank Acc#[1])	Tandem repeat motif	Class of dispersed repetitive element	Similarity (%)	Score
HmRS36T (DQ785753)	CTCA	Endogenous retrovirus	79.6	253
HmRS129D (DQ785766)	GT	DNA transposon	80	216

		(EnSpm)		
HmRS62D (DQ785777)	GT	DNA transposon (MuDR)	86	217
HmLCS73T (DQ993220)	GAGT	LTR retrotransposon (Gypsy)	87.5	249
HmLCS55T (DQ993226)	GTGA	Non-LTR retrotransposon I	76.5	284
HmLCS48M (DQ993227)	(CT)$_n$(CA)$_n$	DNA transposon (Polinton)	74.4	291
HmDL131M (EF054867)	(AC)$_n$(TC)$_n$	Non-LTR retrotransposon (CR1) & DNA retrotransposon	74.3 & 78.3 respectively	253 & 202
HmDL151T (EF054868)	GAGT	Non-LTR retrotransposon (SINE)	80	217
HmNS38T (EF367113)	CACT	Non-LTR retrotransposon I	68.7	282
HmD5 (AY303336)	CA	DNA transposon (MuDR)	86	217
HmNSp31 (EU126856)	(CAA)$_n$(CAG)n (CAA)$_n$	LTR retrotransposon (DIRS)	85.2	237
HmSP42 (EU126858)	(ATG)$_n$X(ATG)$_n$X (ATG)$_n$(TGT)$_n$	Non-LTR retrotransposon	76.4	719
Hmid0065M (GQ927111)	(CT)$_n$(AC)$_n$	DNA transposon (Polinton)	68.3	323
Hmid0136T (GQ92711)	ACTC	Interspersed repeat	86.8	253
Hmid0315 (GQ927120)	(GCGT)$_n$(GT)$_n$	DNA transposon	74	350
Hmid0558 (GQ97123)	GT	DNA transposon (Mariner)	77.4	360
Hmid2009BT (GQ927114)	GT	DNA transposon (hAT)	81.4	373
Hmid2015M (GQ927124)	(TG)$_n$(TGTC)$_n$	Endogenous retrovirus & DNA transposon (Sola)	73.1 & 75 respectively	627 & 285
Hmid2031C (GQ927125)	(GA)$_n$(AG)$_n$(GT)$_n$ (GAGG)$_n$	DNA transposon	76.9	271

Hmid2047AD (GQ927127)	GT	Non-LTR retrotransposon	72.85	312
Hmid4022C (GQ927130)	GT	DNA transposon (Sola)	74.8	489
HmidPS1.42C (GU256657)	ATCC	DNA transposon (MuDR)	72.5	369
HmidPS1.197T (GU256671)	GT	DNA transposon	88.4	456
HmidPS1.222T (GU256675)	GAGT	DNA transposon (Mariner)	76.3	329
HmidPS1.332D (GU256680)	AC	Endogenous retrovirus 2	77.6	290
HmidPS1.370C (GU256683)	$(CAACC)_nX(CACT)_n$	Interspersed repeat	77.4	275
HmidPS1.375C (GU256685)	GAGT	DNA transposon (hAT)	69.9	315
HmidPS1.382D (GU256687)	GT	Endogenous retrovirus 2	74.3	289
HmidPS1.405T (GU256689)	CAAC	Interspersed repeat	77.6	319
HmidPS1.484C (GU256693)	$(GAGT)_nX(GTGA)_nX$ $(GTGA)_n$	Non-LTR retrotransposon (L1)	78.7	244
HmidPS1.549C (GU256696)	GT	DNA transposon (hAT)	77.3	349
HmidPS1.559C (GU256698)	$(CA)_n(TGTA)_n$	Non-LTR retrotransposon (CR1)	80.2	327
HmidPS1.635D (GU256702)	CA	Non-LTR retrotransposon (CR1)	74.2	246
HmidPS1.638T (GU256703)	GAGT	Interspersed repeat	80	227
HmidPS1.728D (GU256706)	CA	Non-LTR retrotransposon (CR1)	76.8	304
HmidPS1.811C (GU256710)	$(TTGT)_n(GT)_n$	Endogenous retrovirus 1	81.3	268
HmidPS1.868T (GU256717)	GAGT	Interspersed repeat	90.6	236
HmidPS1.874C (GU256720)	$(CACG)_n(CA)_n$	DNA transposon (EnSpm)	89.8	395
HmidPS1.890M (GU256721)	$(CACT)_n(CT)_n$	DNA transposon (hAT)	78.3	289

HmidPS1.1038T (GU256734)	GTGA	DNA transposon (Mariner)	72.3	386
HmidPS1.1063C (GU256736)	$(TC)_n(CGTG)_n$	Endogenous retrovirus 1 & DNA transposon	80.5 & 85.7	210 & 391

1 – GenBank accession number.

Screening the microsatellite sequences against the TRANSFAC® 7.0 Public 2005 database showed that various tandem repeat motifs could confer transcription factor binding sites (Table 2.4).

Table 2.4: Repeat motifs that confer predicted transcription factor binding sites.

Tandem repeat motif	Transcription Factor
CA	Wilms' tumor suppressor *Sox2* *Rap1*
GA	*GATA1*
CGTG	*Sp1* *RAP1*
GAGT	*AP1* *GATA1*
CAGT	*JunD* *AP1*

4. Discussion

Seventeen percent of the microsatellite loci presented significant hits to genes in other species. Previous studies have found: 1.7% for the mouse (Herron *et al.* 1998); porcine, chicken and bovine with 2.8%, 3.3%, and 3.8% significant hits respectively (Farber and Medrano 2003) and 8.4% for the horse (Farber and Medrano 2004). This increase in hits is indicative of the exponential accumulation of sequence data in public databases in recent years (Farber and Medrano 2004). However it must be noted that less stringent parameters were set for the selection of significant hits during this study. Similar probability (E)- and score (S)-values as Farber and Medrano (2003, 2004) were used, however Herron *et al*. (1998) used an E-value equal to 1.0e-7 and a Score value of 200 for initial BLASTN selection. Furthermore, Farber and Medrano (2003) used as secondary selection criteria: a cross-species identity of 80%. During this investigation the value was lowered to 70% for BLASTN and for the BLASTX, a Positives-value of 50% was used. The use of the Positives-value, rather than the Identity-value, for the BLASTX was done to compensate for amino acid substitutions that may still confer similar biological function. The relaxed stringency may increase the false positive rate, but taking into consideration the relatively limited molluscan sequence information available, compared to mammalian data, homology with conserved biological function might be overlooked due to the evolutionary divergence of *H. midae* to organisms overrepresented in the NCBI databases.

The overrepresentation of hits to for example the cellulase gene, seemingly more than a third of the gene hits (Table 2.2), could be explained by the presence of multi-gene families (various proteins with similar structure and function; Van Holde *et al.* 2001; Lo *et al.* 2003; Sehring *et al.* 2007). Due to the fact that the actins, cellulases and hemocyanins (Table 2.2) all represent multiple gene families, caution must therefore be taken in the interpretation of these results. For this reason previous studies eliminated significant hits to members of multigene families (Herron *et al.* 1998; Farber and Medrano 2003, 2004). However, unlike those studies where the focus was on *in silico* synteny mapping, the aim of this investigation was the identification of microsatellites associated to genes and their possible function. Therefore the presence of microsatellites in multiple members of the same gene family is further evidence of the functional role that these polymorphisms may play in a particular gene family. Based on the differing Identity-values (Table S1, Appendix) and the lack of fully-conserved microsatellite nucleotide composition, it can therefore be assumed that significant hits to the *H. iris* Actin A2 gene is not necessarily the *H. midae* Actin A2 homolog *per se,* even though they align to the same region within the gene. A similar argument is presented for significant hits to the *H. discus hannai* cellulase gene. Based on this, the existence of at least 11 cellulase/cellulase-like genes and 3 actin/actin-like genes is predicted in *H. midae*. Several actin genes have been isolated in *Haliotis*, with at least six and three actin genes for *H. iris* and *H.*

virginea respectively (Sin *et al.* 2007). Thus far, to our knowledge, two genes of the cellulase gene family, a glycosyl hydrolase family 5 member (Ootsuka *et al.* 2006) and a glycosyl hydrolase family 9 member (Suzuki *et al.* 2003) have been characterised in *H. discus hannai*. The presence of multiple members, of this gene, has also been demonstrated in other invertebrates (Watanabe and Tokuda 2001). However, none of these studies were to the extent of what has been suggested in the current study. This could be due to the fact that these investigations were based on the isolation of proteins or cDNA from particular tissues of animals at a specific life stage. Thus the occurrence of multiple members of the cellulase gene families might have been overlooked. Nonetheless the persistence of up to 11 cellulase/cellulase-like genes in *H. midae* seems unlikely (but not impossible). On secondary investigation of the BLAST results, many microsatellite loci demonstrated multiple significant hits (but lower than what is reported in Table 2.1, 2.2). In an extreme case locus *HmLCS67M* had ten significant matches (Table 2.5). This is strong support for a conserved functional roll for this particular microsatellite locus across Haliotids for a variety of genes. A comparable argument could be given for other loci with similar tendencies. However the risk of incorrectly identifying specific genes associated to a particular locus is much greater.

Table 2.5: Significant matches for locus *HmLCS67M*, demonstrating a possible functional role for this repeat motif in a variety of genes thoughout various Haliotids.

BLAST Hit	Genbank Acc#[1]	E-value	Score	Microsatellite Position
Cellulase gene (*Haliotis discus hannai*)	AB125892.2	2.0E-20	107	5'-UTR
Cytidine deaminase (*Haliotis diversicolor*)	EU101721.1	3.0E-19	104	3'-UTR
Hemocyanin (*Haliotis diversicolor*)	GQ352369.1	9.0E-19	102	Intronic
Hemocyanin isoform H1 (*Haliotis tuberculata*)	AJ252741.1	1.0E-17	98.7	Intronic
Hemocyanin isoform H2 (*Haliotis tuberculata*)	AJ297475.1	1.0E-11	78.8	Intronic

Microphage migration inhibition factor (*Haliotis diversicolor*)	FJ1955326.1	4.0E-11	77.1	Intronic
ATPase alpha subunit (*Haliotis rubra*)	AY043205.1	2.0E-09	71.6	Intronic
Peroxirdoxin (*Haliotis discus discus*)	EF103356.1	7.0E-08	66.2	Intronic
Indoleamine 2,3 dioxygennase (*Sulculus diversicolor*)	AB183456.1	3.0E-06	60.8	Intronic
Glutathione-s-transferase (*Haliots discus discus*)	EF103347.1	1.0E-05	59	3'-UTR

1 – GenBank accession number.

Alignment and open reading frame analyses indicate that none of the microsatellites are located in protein coding sequences. This is expected, taking into account that the majority of the repetitive motifs are di- or tetranucleotide repeats. The expansion of such repeats will disrupt the open reading frame, subsequently placing the expansion under negative selective pressure (Metzgar 2000). Tetranucleotides are the most abundant microsatellite motif in genic regions (excluding exons) of *H. midae* (72%, Fig. 2.2); of these GAGT constitutes approximately 57%. This is unusual for animal genomes, which in general show a bias toward CA repeats in UTR's; only 14% of repeats in these regions of *H. midae* constitute CA motifs (Fig. 2.1, 2.2). The low frequency of CG and apparent absence of CCG and CGG repeats agrees with current evidence for selection against CpG-like motifs as in other invertebrates. CpG-like motif in vertebrate genes frequently locate in 5'-UTR where they serve as protein binding sites regulated by DNA-methylation, so-called CpG-islands. However in the absence of methylation, (invertebrates do not generally methylate DNA; Tóth *et al.* 2000), these CpG-like motifs proves to be highly mutable and may cause aberrant gene effects if persistent in genic sequences. It is for this reason that vertebrate introns are devoid of such motifs as it may incorrectly alter splicing signals (intron sequences are not methylated). These motifs are also prone to forming hairpin structures that may alter mRNA secondary structures, further prohibiting the formation of mature mRNA (Li *et al.* 2004). The intronic microsatellites seem to be located close to the exon-intron-boundary indicating a possible role in correct intron-splicing and/or alternative splicing (Sirand-Pugnet *et al.* 1995; Ejima *et al.* 2000; Pagani *et al.* 2000; Gabellini 2001). It has been demonstrated that intronic microsatellites also play a role in transcriptional regulation; both up-regulating and causing gene silencing (Meloni *et al.* 1998; Gebhardt *et al.* 1999, 2000; Albanèse *et al.* 2001; Li *et al.* 2004).

Microsatellites located in the 5'-UTR could confer binding sites for transcription factors (Table 2.4) and/or serve as a mechanism to stabilise the structure of the transcriptional complex; in this manner microsatellites could regulate and facilitate gene expression (Calkhoven *et al.* 1994; Sandaltzopoulos *et al.* 1995; Toutenhoofd *et al.* 1998; Timchenko *et al.* 1999; Chiba-Falek and Nussbaum 2001; Agnèse *et al.* 2009). Actual tandem repeat length has also been implicated in playing a role in the effective expression of genes. Gebhardt *et al.* (1999, 2000) found a decrease in transcriptional activity of the epidermal growth factor receptor gene with an increase of the CA tandem repeat number in the promoter of this gene. Three prime UTR microsatellites have been shown to influence transcription slippage that aids in the transcription of long mRNA or facilitates transcription termination (Zeiss *et al.* 1998; Li *et al.* 2004). The microsatellites with hits to the 5'-UTR of the *H. discus hannai* cellulase gene (Table 2.2) all align to the same region of the nucleotide sequence, approximately 399 base pairs upstream. Within this region the *H. discus hannai* cellulase gene exhibits a CACT repeat (GAGT reverse compliment). The majority of microsatellites share this repeat with the subject sequence, however in a few loci this motif is not strictly conserved; nonetheless all exhibit a C/G rich tetranucleotide repeat structure. The exception to this is the dinucleotide motif of *HmNS31D*, which in all probability is being read as a functional GTGT tetranucleotide repeat by the transcriptional complex. This phenomenon could in part be explained as a mechanism for differential gene expression of different members of a multigene family. The need for such differential expression of members of the cellulase gene family in abalone is probably due to the variety of cellulose sources that these animals utilise. Depending on the life cycle stage, juvenile abalone feed on a range of diatoms and other micro-algae to macro-algae, such as kelp, in adulthood (Barkai and Griffiths 1986; Wood and Buxton

1996). However it must be stressed, as noted earlier, that not all these microsatellite loci might be associated to a cellulose gene. The GAGT and CAGT motifs are in fact recognition site for AP-1 and GATA transcription factors (Table 2.4) that facilitate transcription of a wide range of genes. Thus the high number of hits to cellulase may be artefactual, as demonstrated in Table 2.5 for the *HmLCS67M* locus.

Both *HmLCS55T* and *HmNS38T* gave significant BLASTX hits to an unknown protein of *Haliotis diversicolor* and significant BLASTN hits to a lysin gene. It is therefore thought that this unknown protein is in most likelihood a lysin protein. Similarly *HmNST7T* had positive BLASTX and BLASTN hits to Vertebrate inositol 1,4,5-triphosphate-3-kinase B and Cellulase respectively. This can be clarified as possible overlapping open reading frames of two independent genes for an inositol kinase and a cellulase gene or the existence of a single protein that has both inositol kinase and cellulase activities. Such divergent multi-functionality of proteins are known, e.g.: phosphoglucose isomerase/neuroleukin that functions as an enzyme in glycolysis, a cytokine and a growth factor (Jeffery 1999) and maize sucrose synthase that functions both as a signal transducer and in sugar metabolism (Subbaiah *et al.* 2006). If the first scenario proves to be the case, this microsatellite might possess dual functionality as an intronic microsatellite for the inositol kinase gene and a 5'-UTR microsatellite for the cellulase gene.

Microsatellites in various species have been found to be associated to mobile elements, such as transposons (Ramsay *et al.* 1999; Kidwell 2002; Biémont and Viera 2005; Grover *et al.* 2007; Meglécz *et al.* 2004, 2007 and reference therein). In *H. midae*, 21% of the microsatellite flanking regions showed similarity to known transposable elements (Table 2.3) suggesting an association with microsatellites in this species as well. Twenty one percent is considerably more than what was found for *Drosophila melanogaster* which demonstrated a 6.4% microsatellite-transposable element association, but rather low in comparison with that of *Anopheles gambiae* (mosquito) (40.1%) (Meglécz *et al.* 2007).

Transposable elements are hypothesised to play an important role in the origin, distribution and evolution of microsatellites in the genome (Temnykh *et al.* 2001; Meglecz *et al.* 2007). The relationship between microsatellites and mobile elements is suggested to be mutualistic, and proponents argue in favour of a co-evolution, as the tandem repeats are postulated to facilitate transposition by serving as a navigational tool, directing homologous interaction at target sequences for integration (Nadir *et al.* 1996). The CA and GAGT motifs are comparatively equal in the percentage of loci associated to transposable elements in *H. midae* (Fig. 2.1) (with dinucleotides, in general, marginally higher, Fig. 2.2) and may in part explain their high genomic frequency. Similar scenarios were postulated for the abundance of AC(A/G)G repeats in *Drosophila* that is associated to the *SGM* element (Miller *et al.* 2000) and TA repeats in the rice genome linked to the *Micron* transposon (Grover *et al.* 2007).

The association of microsatellites to duplicated sequences hold implications for the use of microsatellites as a molecular marker in *H. midae* and perhaps in other abalone species as Baranski *et al*. (2006) also reported regions of similarity amongst microsatellite flanking regions for the Blacklip abalone, *H. rubra*. As previously noted, primer design is imperative to the creation of a workable, ‘well behaving’ microsatellite marker. Primers that anneal to such duplicated regions could lead to non-specific amplifications during PCR. Primmer *et al.* (1997) noted that the lack of microsatellite associations to SINE/LINE elements in avian genomes facilitated easier microsatellite isolation. A further implication of microsatellite-transposable element association is the increase of the false positive rate when identifying microsatellite-gene associations as described above. It can not be excluded that at least some of the hits identified by the BLAST analyses are due to interspersed repetitive elements. Evidence for this is the high number of significant hits to the cellulase gene. However it is well known that multi-gene families arise due to duplication events and that such duplication could be facilitated by transposable elements. Furthermore various mobile elements locate in gene-rich regions and actively contribute to gene evolution and regulation (Bennetzen 2000; Medstrand *et al.* 2005), arguing that superfluous hits may still be associated to genes. In fact it is not improbable that the GAGT repeat originated in a transposable element that located in 5’-UTRs or introns of a gene that subsequently altered gene function favourably, leading to the propagation of this motif.

Overall, CA repeats have the highest frequency of all sequence motifs in the *Haliotis midae* genome. This is an anticipated result as most higher eukaryotes, including humans (Subramanian *et al.* 2003), most other vertebrates, arthropods (Tóth *et al.* 2000) and bivalve molluscs (Cruz *et al.* 2005) have demonstrated an overrepresentation of this repeat motif. This uniform distribution, across taxa, may be indicative of its evolutionary neutrality as has been suggested in the case of *D. melanogaster* (Bachtrog *et al.* 1999). However *Caenorhabditis elegans* marginally favours AT motifs, while CG repeats are more common in fungi (Tóth *et al.* 2000). In some insects (Bees and Bumble bees) (Estoup *et al*. 1993) and the European flat oyster (*Ostrea edulis*) (Naciri *et al*. 1995), CT motifs predominate. This suggests negative selection against CA microsatellites only in these species and perhaps then a highly conserved function in the majority of animal lineages. Indeed highly conserved CA microsatellite loci have been identified in sharks (Martin *et al.* 2002) and the 3'-UTR CA microsatellite locus of the dystrophin gene in mammals has shown wide cross-species amplification in animals as diverse as primates, bovids and rodents. Zeiss *et al.* (1998) propose a conserved function for this microsatellite as a transcriptional stabiliser, because the dystophin gene takes approximately 16 hours to complete one round of transcription. Furthermore, dinucleotide (including CA motifs) motifs have the ability take on alternative forms of DNA; most notably the Z-conformation that may facilitate protein binding for recombination (Li *et al.* 2002). A recent study showed a positive correlation between microsatellite densities on chromosomes and recombination rate (Guo *et al*. 2009).

In *H. midae*, TA motifs seem to be absent (perhaps more realistically in very low frequency) in the genome. This is surprising because TA repeats are the second most plentiful dinucleotide after CA repeats in most organisms (Tóth *et al.* 2000). On average CA repeats are 2.3 times more prevalent than TA repeats (Christiakov *et al.* 2006), therefore the expected TA frequency for this abalone species will be ~13%. There is thus a strong bias against TA microsatellite loci, however whether this bias is a true biological bias or a mere artefact of the hybridisation protocol's inability to isolate palindromic motifs, remains to be investigated (Powell *et al.* 1996). The second most abundant microsatellite motif in *H. midae* is GAGT. Contrary to most other taxa, tetranucleotides and not dinucleotides are the most abundant length motifs throughout the genome of *H. midae* (Fig. 2.1). The percentage of dinucleotide repeats (~37%) are comparable to what was found for genome-wide dinucleotide repeats in Japanese puffer fish (34%), yet the tetranucleotide distribution is in stark contrast between *H. midae* (~47%) and this fish (21%) (Fig. 2.2) (Edwards *et al.* 1998). This coupled with a high percentage of tetranucleotide-gene association, points to a functional role of tetranucleotides, especially GAGT, in the South African abalone.

5. Conclusion

During this study, data was added to the growing body of evidence that supports a non-random distribution of microsatellites within and between genomes. It is hypothesised that this is due to a selection differential; however possible ascertainment biases due to the cloning procedure for microsatellite isolation can not be wholly excluded (Agnèse *et al.* 2009). Furthermore a thorough understanding of microsatellite genomic dynamics has practical implications for future marker development and use in *H. midae*.

The phenotypic effects of microsatellite variation are well documented in human disease phenotypes e.g. the CAG expansion in the *HD* gene that causes Huntington's disease and the role of deleted tandem repeat motifs in the development of cancer syndromes (Li *et al.* 2002, 2004). In production animals, phenotypic variation of quantitative traits of economic value, such as egg production in chickens, milk yield in cattle and growth rate in fish have been found to be associated to microsatellite loci, using linkage analyses and association studies (Sakamoto *et al.* 1999; Streelman and Kocher 2002; Cnaani *et al.* 2003; Hansen *et al.* 2005; Schnabel *et al.* 2005; Chatterjee *et al.* 2008). Obviously in such QTL studies the microsatellite may not necessarily be the causative agent *per se*, but could however be closely linked to such a variant. Nonetheless the direct effect of microsatellite variation on production phenotypes have been suggested in some aquaculture species (Agnèse *et al.* 2009). In this study it has been demonstrated that anonymous microsatellites could be converted to type I markers via similarity searches to public databases. It is envisioned that this method will gain power, in future, due to the ever-expanding number of sequences in these databases and the refinement of search algorithms. This will facilitate candidate gene, even causal variant, identification for phenotypes with value. It must be noted though that positive gene identification may be hindered by conserved functional motifs in UTRs persisting in many genes, as postulated to be the case in *H. midae.* Transposable elements too are powerful mutagenic agents and are known to locate in intronic and regulatory elements of genes, altering expression and function (Bennetzen 2000). Indeed, transposable elements have been shown to be under positive selection, illustrating a role in adaptation (Maside *et al.* 2002; Schlenke and Begun 2004). Association of transposable elements to microsatellites, as shown, allows for such elements of functional importance to be detected and employed in marker assisted breeding schemes; however this association also impedes the design of unique primer sets for microsatellite development.

Microsatellites are a popular and powerful molecular tool. However the nature of the biological dynamics of this genomic feature remains under debate. The lack of such consensus may impact the utility of microsatellites and the development of appropriate models for this marker system. It is clear however that cumulating evidence suggest that microsatellites are far from passive DNA, as judged by their non-random distribution.

References

1. Agnèse J-F, Adépo-Gourène B, Nyingi D (2009) Functional microsatellite and possible selective sweep in natural populations of the black-chinned tilapia *Sarotherodon melanotheron* (Teleostei, Cichlidae). Marine Genomics 1: 103-107.
2. Albanèse V, Biguet NF, Kiefer H, Bayard E, Mallet J, Meloni R (2001) Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite. Hum Mol Genet 10: 1785-1792.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

4. Bachtrog D, Weis S, Zangerl B, Brem G, Schlötter C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Mol Biol Evol 16: 602-610.

5. Baranski M, Rourke M, Loughnan S, Austin C, Robinson N (2006) Isolation and characterization of 125 microsatellite DNA markers in the blacklip abalone, *Haliotis rubra.* Mol Ecol Notes 6: 740-746.

6. Barkai R, Griffiths CL (1986) Diet of the South African Abalone *Haliotis midae.* S Afr J Mar Sci 4: 37-44.

7. Bennetzen JF (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42: 151-169.

8. Bester AE, Slabbert R, D’Amato ME (2004) Isolation and characterisation of microsatellite markers in South African abalone (*Haliotis midae*). Mol Ecol Notes 4: 618-619.

9. Biémont C, Vieir C (2005) What transposable elements tell us about genome organization and evolution: the case of *Drosophila.* Cytogenet Genome Res 110: 25-34.

10. Calkhoven CF, Bouwman PR, Snippe L, Ab G (1994) Translation start site multiplicity of the CCAAT/enhancer binding protein alpha mRNA is detected by a small 5’ open reading frame. Nucleic Acids Res 22: 5540-5547.

11. Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Phys B 126: 455-476.

12. Chatterjee NR, Sharma RP, Mishra A, Dange M, Bhattacharya TK (2008) Variability of microsatellites and their association with egg production traits in chicken. International Journal of Poultry Science 7: 77-80.

13. Chiba-Falek O, Nussbaum RL (2001) Effect of allelic variation at the NACP-Rep1 repeat upstream of the α-synuclein gene (SNCA) on transcription in cell culture luciferase reporter system. Hum Mol Genet 10: 3101-3109.

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

15. Cnaani A, Hallerman EM, Ron M, Weller JI, Indelman M, Kashi Y, Gall GAE, Hulata G (2003) Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F_2 tilapia hybrid. Aquaculture 223: 117-128.

16. Cruz F, Pérez M, Presa P (2005) Distribution and abundance of microsatellites in the genome of bivalves. Gene 346: 241-247.

17. Edwards YJ, Elgar G, Clark MS, Bishop MJ (1998) The identification and characterization of microsatellites in the compact genome of the Japanese puffer fish, *Fugu rubripes:* perspectives of functional and comparative genomic analysis. J Mol Biol 278: 843-854.

18. Ejima Y, Yang L, Sasaki MS (2000) Aberrant splicing of the ATM gene associated with shortening of the intronic mononucleotide tract in human colon tumor cell lines: a novel mutation target of microsatellite instability. Int J Cancer 86: 262-268.

19. Estoup AM, Solignac MH, Cornuet J-M (1993) Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res 21: 1427–1431.

20. Farber CR, Medrano JF (2003) Putative *in silico* mapping of DNA sequences to livestock genome maps using SSLP flanking sequences. Anim Genet 34: 11-18.

21. Farber CR, Medrano JF (2004) Identification of putative homology between horse microsatellite flanking sequences and cross-species ESTs, mRNAs and genomic sequences. Anim Genet 35: 28-33.

22. Gabellini N (2001) A polymorphic GT repeat from the human cardiac Na^+ Ca^{2+} exchanger intron 2 activates splicing. Eur J Biochem 268: 1076-1083.

23. Gebhardt F, Zanker KS, Brandt B (1999) Modulation of epidermal factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem 274: 13176-13180.

24. Gebhardt F, Burger H, Brandt B (2000) Modulation of EGFR gene transcription by a polymorphic repetitive sequence – a link between genetics and epigenetics. Int J Biol Marker 15: 105-110.

25. Grover A, Aishwarya V, Sharma PC (2007) Bias distribution of microsatellite motifs in the rice genome. Mol Genet Genomics 277: 469-480.

26. Guo W-J, Ling J, Li P (2009) Consensus features of microsatellite distribution: Microsatellite contents are universally correlated with recombination rates and are preferentially depressed by centromeres in multicellular eukaryotic genomes. Genomics 93: 323-331.

27. Gupta PK, Rustgi S (2004) Molecular markers from transcribe/express regions of the genome in higher plants. Funct Integr Genomics 4: 139-162.

28. Hansen C, Yi N, Zhang YM, Xu S, Gavora J, Cheng HH (2005) Identification of QTL for production traits in chickens. Anim Biotechnol 16: 67-79.

29. Herron BJ, Silva GH, Flaherty L (1998) Putative assignment of ESTs to the genetic map by use of the SSLP database. Mamm Genome 9: 1072-1074.

30. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

31. Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci 24: 8-11.

32. Jurka J, Kapitonov VV, Pavlick A, Klownowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462-467.

33. Kazazian HH (2004) Mobile elements: Drivers of genome evolution. Science 303: 1626-1632.

34. Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49-63.

35. Kourtidis A, Drosopoulou E, Pantzartzi CN, Chintiroglou CC, Scouras ZG (2006) Three new satellite sequences and a mobile element found inside the HSP70 introns of the Mediterranean mussel (*Mytilus galloprovincialis*). Genome 49: 1451-1458.

36. Levis RW, Genesan R, Houtchens K, Tolar LA, Sheen FM (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell 75: 1083-1093.

37. Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465.

38. Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 21: 991-1007.

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

40. Lo N, Watanabe H, Sugimura M (2003) Evidence for the presence of a cellulose gene in the last common ancestor of the bilaterian animals. Proc R Soc Lond B (Suppl.) 270: S69-S72.

41. Martin AP, Pardini AT, Noble LR, Jones CS (2002) Conservation of a dinucleotide simple sequence repeat locus in sharks. Mol Phylogenet Evol 23: 205-213.

42. Maside S, Lee AW, Charlesworth B (2002) S-element insertion are associated with the evolution of the *Hsp70* genes in *Drosophila melanogaster*. Curr Biol 12: 1686-1691.

43. Medstrand P, Van de Lagemaat LN, Dunn CA, Landry J-R, Svenback D, Mager DL (2005) Impact of transposable elements on the evolution of mammalian gene regulation. Cytogenet Genome Res 110: 342-352.

44. Meglecz E, Petenian F, Danchin E, D' Acier AC, Rasplus J-Y, Raure E (2004) High similarity between flanking regions of microsatellite detected within each of two species of Lepidoptera: *Paenassius apollo* and *Euphydryas aurinia*. Mol Ecol 13: 1693-1700.

45. Meglecz E, Anderson SJ, Bourguet D, Bucher R, Caldas A, Cassel-Lundhagen A, D' Acier AC, Dawson DA, Faure N, Fauvelot C, Frank P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju C, Patt A, Petenian F, Silvian J-F, Wilcock HR (2007) Microsatellite flanking region similarities among different loci within insect species. Insect Mol Biol16: 175-185.

46. Meloni R, Albanèse V, Ravassard P, Treilhou F, Mallet J (1998) A tetranucleotide microsatellite, located in the first intron of tyrosine hydroxylase gene, acts as a transcription regulatory element *in vitro*. Hum Mol Genet 7: 423-428.

47. Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10: 72-80.

48. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith LC, McCoy JM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403: 785-789.

49. Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. Mol Biol Evol 17: 1597-1609.

50. Naciri YY,. Vigouroux J, Dallas E, Desmarais CD, Bonhomme F (1995) Identification and inheritance of (GA/TC)n and (AC/GT)n repeats in the European flat oyster *Ostrea edulis* (L.). Mol Mar Biol Biotechnol 4: 83–89.

51. Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Nat Acad Sci USA 93: 6470-6475.

52. Ootsuka S, Saga N, Suzuki K, Inoue A, Ojima T (2006) Isolation and cloning of an endo-β-1,4-mannanase from Pacific abalone *Haliotis discus hannai*. J Biotechnol 125: 269-280.

53. Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE (2000) Splicing factors induce cystic transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem 275: 210141-210147.

54. Primmer CR, Raudsepp T, Chowdhary BP, Møller AP, Ellegren H (1997) Low frequency of microsatellites in the avian genome. Genome Res. 7: 471-482.

55. Powell W, Macrhay GC, Provan J (1996) Polymorphisms revealed by simple sequence repeats. Trends Plant Sci 1: 215-222.

56. Ramsey L, Macaulay M, Cardle L, Morgante M, Ivanissevich SD, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in the barley. The Plant Journal 17: 415-425.

57. Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite flanking regions: a SNP mine in South African abalone (*Haliotis midae*). Anim Genet 39: 329.

58. Roodt-Wilding R, Slabbert R (2006) Molecular markers to assist the South African abalone industry. S Afr J Sci 102: 99-102.

59. Sakamoto T, Danzmann RG, Okamoto N, Ferguson MM, Ihssen PE (1999) Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 173: 33-43.

60. Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, Becker PB (1995) Dual regulation of the *Drosophila hsp26* promoter *in vivo.* Nucleic Acids Res 23: 2479-2487.

61. Schlenke TA, Begun DJ (2004) Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proc Nat Acad Sci USA 101: 1626-1631.

62. Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet 36: 408-416.

63. Sehring IM, Mansfeld J, Reiner C, Wagner E, Plattner H, Kissmehl R (2007) The actin multigene family of *Paramecuim tetreaurelia*. BMC Genomics 8: 82-97.

64. Serapion J, Kucuktas H, Feng J, Liu Z (2004) Bioinformatic mining of Type 1 microsatellites from expressed sequence tags of Channel Catfish (*Ictalurus punctatus*). Mar Biotechnol 6: 364-377.

65. Sin FYT, Bryant MJ, Johnstone A (2007) Molecular evolution and phylogeny of actin genes in *Haliotis* species (Mollusca: Gastropoda). Zool Stud 46: 734-745.

66. Sirand-Pugnet P, Durosay P, Brody E, Marie J (1995) An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken β-tropomyosin pre-mRNA. Nucleic Acids Res 23: 3501-3507.

67. Slabbert R, Ruivo NR, Van den Berg NC, Lizamore DL, Roodt-Wilding R (2008) Isolation and characterization of 63 microsatellite loci for the abalone, *Haliotis midae*. J World Aquacult Soc 39: 429-435.

68. Slabbert R, Hepple J, Venter A, Nel S, Swart L, Van den Berg NC, Roodt-Wilding R (in press) Isolation and segregation of 44 microsatellite loci in the South African abalone, *Haliotis midae* L. Anim Genet. DOI: 10.1111/j.1365-2052.2009.02003.x.

69. Streelman JT, Kocher TD (2002) Microsatellite variation associated with prolactin expression and growth of salt-challenged Tilapia. Physiol. Genomics 9: 1-4.

70. Subbaiah CC, Palaniappan A, Duncan K, Rhoads DM, Huber SC, Sachs MM (2006) Mitochondrial localization and putative signalling function of sucrose synthase in Maize. J Biol Chem 281: 15625-15635.

71. Subramanian S, Mishra R, Singh L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol 4: R13

72. Suzuki K, Ojima T, Nishita K (2003) Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. Eur J Biochem 270: 771-778.

73. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441-1452.

74. Timchenko NA, Welm AL, Lu X, Timchenko LT (1999) CUG repeat binding protein (CUGBP1) interacts with the 5’ region of C/EBPbeta mRNA and regulates transcription of C/EBPbeta isoforms. Nucleic Acids Res 27: 4517-4525.

75. Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and Analysis. Genome Res 10: 967-981.

76. Toutenhoofd SL, Garcia F, Zacharias DA, Wilson RA, Strehler EE (1998). Minimum CAG repeat in the human calmodulin-1 gene 5' untranslated region is required for full expression. Struct Expr 1398: 315-320.

77. Van Holde KE, Miller KI, Decker H (2001) Hemocyanins and invertebrate evolution. J Biol Chem 276: 15563-15566.

78. Watanabe H, Tokuda G (2001) Animal cellulases. Cell Mol Life Sci 58: 1167-1178.

79. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruß M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 28: 316-319.

80. Wood AD, Buxton CD (1996) Aspects of the feeding biology of the abalone *Haliotis midae* (Linne, 1758) on the east coast of South Africa. 1. Feeding Biology. S Afr J Mar Sci 17: 61-68.

81. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11: 1-16.

82. Zeiss CJ, Trepanier LA, Aguirre GD, Ray K (1998) A highly conserved microsatellite in the dystrophin gene of diverse mammalian species. Anim Genet 29: 224-227.

Chapter III:
Microsatellite transfer from *Haliotis rubra* and *Haliotis discus hannai* to *Haliotis midae*

Abstract

Microsatellites are popular and widely used molecular markers; however their isolation remains technically challenging and expensive via classical *de novo* strategies. On the other hand, cross-species transfer of microsatellites is a quick and inexpensive method that exploits existing marker data from other species. Reports of microsatellite cross-species transfer in abalone have however demonstrated comparatively low success rates. This study attempted to increase transfer success by using public databases to identify microsatellites that may be associated with genes and therefore show greater sequence conservation between species. A further advantage is that this approach allows for the concurrent development of cross-species type I molecular markers. Ten loci were developed from two species, *H. rubra* and *H. d. hannai*, demonstrating moderate levels of polymorphism, as to be expected. An increase of microsatellite transfer success was achieved in comparison to previous studies, however statistically significant differences in transfer rate between type I versus type II markers were only found for the more distantly related species *H. d. hannai*. This illustrates that the cross-species transfer approach, employed here, may be more valuable when source and target species are phylogenetically less related. Additional factors including null allele frequencies and repeat motif length may also play a role in transfer success.

1. Introduction

Microsatellites, even though a popular and widely used molecular maker, remain technically cumbersome and costly to develop via traditional *de novo* methodologies. These strategies rely on the construction of partial genomic libraries, often enriched for repetitive sequence motifs (for a review, Zane *et al.* 2002). This is the primary strategy currently employed for microsatellite isolation in *Haliots midae* as well (Bester *et al.* 2004; Slabbert *et al.* 2008, in press). There is a growing need to develop “universal markers” that will successfully amplify products across a variety of species that can be used for evolutionary studies, investigating population divergence and speciation events (Palo *et al.* 2001; Noor and Feder 2006). Furthermore, cross-species transfer allows genomic information from a well characterised species to be superimposed on that of a less well known species. This is of particular importance in aquaculture where QTLs may be linked to the same marker locus in related species, facilitating comparative and functional mapping (Chistiakov *et al.* 2006).

Cumulating marker information available on a variety of model and well characterised species, facilitated the initiation of a large number of studies, especially examining microsatellite transferability between species (e.g. Moore *et al.* 1991; Schlötterer *et al.* 1991; FritzSimmons *et al.* 1995; Primmer and Merilä 2002; Freitas *et al.* 2007; Miles *et al.* 2009). In some organisms, microsatellite flanking regions demonstrate surprisingly high levels of conservation, e.g. in fish (Rico *et al.* 1996) and marine turtles (FritzSimmons *et al.* 1995), allowing for amplification even after 470 and 300 million years of divergent evolution, respectively. Nonetheless, the success of microsatellite transfer displays a high degree of variation amongst taxonomic groupings. With regards to percentage loci amplifying within genera, invertebrates (excluding arthropods), reptiles, birds and mammals achieved the greatest successes. However, in terms of polymorphic loci transferred, birds and reptiles outperform invertebrates (including arthropods). With regards to cross-genus transfer, invertebrates perform even more poorly (Fig. 3.1, Barbará *et al.* 2007).

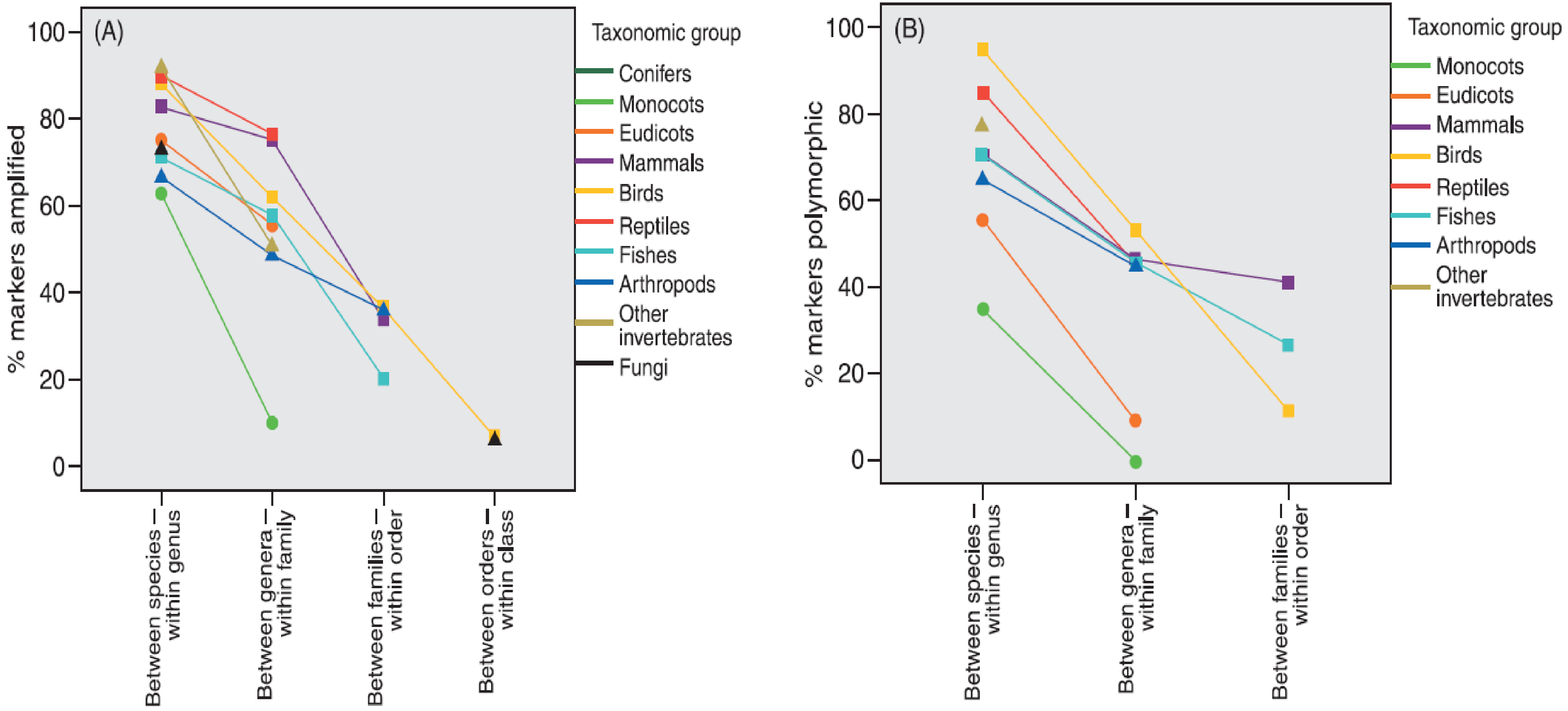


Figure 3.1: This figure taken from Barbará *et al.* (2007) shows the relative successes of cross-species microsatellite transferability across a wide spectrum of plant and animal taxa. These results were obtained by surveying literature published on 611 cross-species transferability studies from 1997 to mid 2006.

Cross-species microsatellite transfer in molluscs appear to be somewhat lower, in most cases with regards to amplification success rate, in comparison to the estimates for invertebrates (excluding arthropods) in general (Barbará *et al.* 2007) (Fig. 3.1). Nonetheless substantial variation is reported between taxa (Table 3.1).

Table3.1: A summary of microsatellite transfer rates (within genera and between genera) and percentage polymorphic markers in various molluscan species.

Species	Transfer Rate (%)	Polymorphic (% of transferred markers)	Reference
Cross-species transfer			
Limpet (*Patella* sp.)	45.5% and 63.6%	18% and 36%	Peréz *et al.* (2007)
Oysters (*Crassostrea* sp.)	20%-70%	50%-100%	Cruz *et al.* (2007)
Scallop (*Nodipecten* sp.)	74.3%	92.3%	Ibarra *et al.* (2006)
Oyster mussel (*Epioblasma* sp.)	90%	100%	Jone *et al.* (2004)
Mucket (*Lampsilis* sp.)	100%	66.7%	Eackles and King (2002)
Cross-genus transfer			
Aplexa to *Physa* (*Gastropoda*)	6.7%-40%	0% - 20%	DuBois *et al.* (2008)
Pectinidae family (scallop, *Bivalvia*)	0%-31.4%	undetermined	Ibarra *et al.* (2006)

It must however be noted that Barbará *et al.* (2007) imposed stringent criteria for including a study's findings in their survey. Consequently their estimate for non-arthropod invertebrates was based on only two publications of which one represented a mollusc (Eackles and King 2002). Therefore their result may be highly biased.

The first attempt to test microsatellite transferability in abalone was conducted by Huang and Hanna (1998). They evaluated only three microsatellite loci, originally from *H. rubra*, but surveyed 15 species over five geographic regions (including two South African endemics: *H. midae* and *H. parva*). They failed to produce significant amplification beyond Australian endemics. Evans *et al.* (2001) increased the number of markers in their study to 22, also originally isolated in *H. rubra*, and tested marker transfer across 12 species. They found the highest transfer rate between *H. rubra* and *H. conicopora* (a proposed sub-species of *H. rubra*; Geiger 2000). The lowest transfer results were obtained for the tropical abalone, *H. asinina* and two New Zealand species, *H. iris* and *H. australis,* with 22% each. An intermediate result was obtained for the South African *H. midae,* with 45% transfer success of which 60% proved to be polymorphic. A later study testing transfer from *H. rubra* to *H. laevigata* reported 71.2% amplification success with 75.3% polymorphic loci of those transferred (Baranski *et al.* 2006a). High transfer success was reported between the two Pacific abalone subspecies (*H. discus discus* and *H. d. hannai*) (100% transfer and polymorphism), however transfer to *H. midae* and *H. rubra* was low; 16.7% and 0% respectively (Sekino and Hara 2007a). American abalone show similar trends within the North American species group, but demonstrates low transferability to southern hemisphere species (Cruz *et al.* 2005; Díaz-Viloria *et al.* 2008).

Several factors are postulated to affect microsatellite transferability between species. The most prominent is that of phylogenetic distance between source and target organism (Primmer *et al.* 2005). Thus marker loci fail to amplify in the target species due to an increased accumulation of sequence mutations over time. Also, even if the locus produces a product in the target species there is no guarantee that the marker would have remained polymorphic, as the microsatellite repeat might be in a process of decay, halting mutation via replication slippage. This is particularly true taking into account the life cycle hypothesis of microsatellite evolution as discussed in Chapter I (Chambers and McAvoy 2000). This dependence of microsatellite transfer success on phylogenetic distance appears to be the norm in abalone as well (Fig. 3.2, Sekino and Hara *et al.* 2007a); however Panova *et al.* (2008) did note that phylogenetic distance was not always a good measure of microsatellite utility after amplification success.

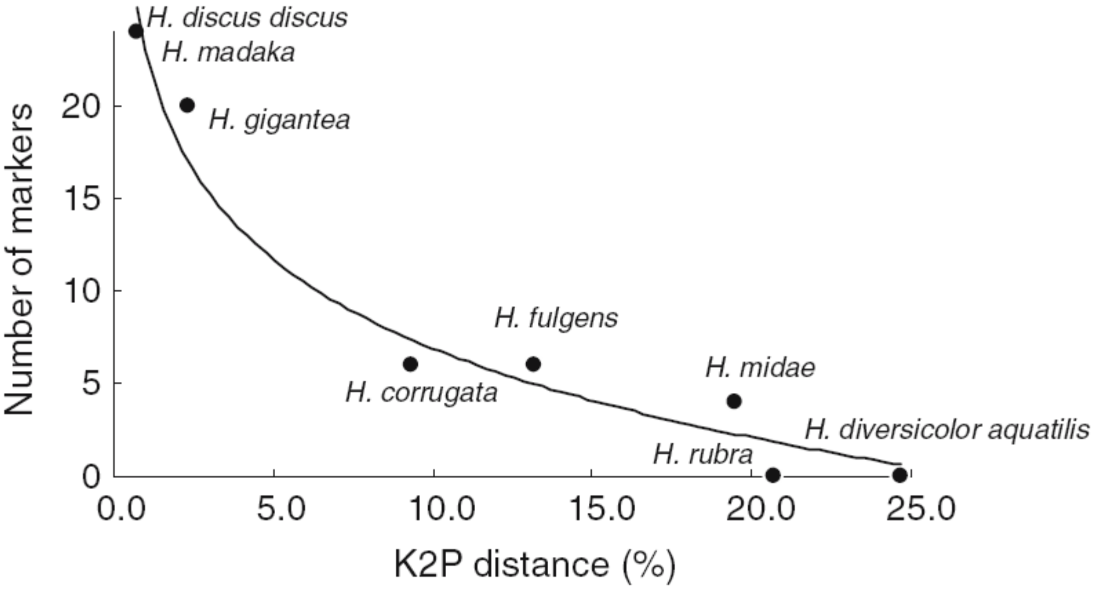


Figure 3.2: The graph depicts the number of microsatellite loci transferred from *H. d. hannai* to other abalone species as a logarithmic regression function of phylogenetic distance calculated here as the K2P distance (Kimura 1980) based on the mitochondrial *cytochrome c oxidase subunit I* nucleotide sequence. The graph shows a clear negative correlation between phylogenetic distance and microsatellite transferability (Graph taken from Sekino and Hara 2007a).

Several other factors have been suggested to impact microsatellite transferability. These include: a negative correlation between genome size and microsatellite transfer, as it is generally thought that the larger the genome of an organism the higher the accumulated sequence variation and thus the greater the probability of failed locus amplification. Organisms with short generation times and/or practising selfing as a mating systems (e.g some snails, DuBois *et al.* 2008) demonstrate excessive sequence variation, because of rapid metabolism, leading to increased mutation rates and/or small effective population sizes (Barbará *et al.* 2007). Furthermore, the number of perfect repeats in the source species is directly proportional to trans-species amplification success and polymorphism in the target species (Neff and Gross 2001). A number of species and laboratory "random effects" may also hamper the accurate estimation of microsatellite transferability. Species effects are most notably sequence variation at the primer binding site that becomes fixed in the target species creating a null allele effect. Thus if the primer binding site was slightly offset to this point of variation, a successful amplification could have been scored. Laboratory "random effects" include the quality and number of target DNA specimens, thermo-cycler ramping time and gel visualisation strategies (Primmer *et al.* 2005).

In terrestrial animals, microsatellite loci appear to be conserved across a moderate evolutionary time scale (~20-30 Million years) (Harr *et al.* 1995; Pépin *et al.* 1995; Ellegren *et al.* 1997; Noor *et al.* 2001). However as noted earlier, marine animals demonstrate sequence conservation over extremely long evolutionary periods, between 35 and 470 million years (Shlötterer *et al.* 1991; FritzSimmons *et al.* 1995; Rico *et al.* 1996; Martin 2002). Two hypotheses have been put forth to explain this phenomenon, the first postulates that mutagenic processes in aquatic environments are less prevalent. The second argues in favour of a discrepancy between mutation rates for endo- and ectothermic animals, where endotherms demonstrate high mutation rates because of high metabolism. Current microsatellite transfer data for abalone seem inconsistent with both these hypotheses, as abalone displays comparatively low levels of microsatellite transferability, with significant decay after less than 13 million years (Evans *et al.* 2001; Sekino and Hara *et al.* 2007a). A similar observation has been made for the Pacific oyster (*Crassostrea gigas*, Hedgecock *et al.* 2004). It is postulated that in the case of these molluscs, the high fecundity of females favours the accumulation of mutations due to elevated meiotic divisions. A similar argument is proposed for elevated mutation rate in males of most species that produce copious amounts of sperm (male driven evolution, Li *et al.* 2002). Thus in these molluscs where both males and females generate excessive amounts of gametes, the probability of mutation accumulation is increased.

In this study it is aimed to increase the microsatellite transfer rate to *H. midae* by selecting microsatellite markers that demonstrate association to genes. Increased sequence depositories in public databases have simplified the identification of such markers. It is postulated that these markers will demonstrate less sequence divergence due to possible selective pressures. As source species, two economically important abalone in Australia (*Haliotis rubra*) and Asia (*Haliotis discus hannai*) have been selected. Considerable work has been done on these species and linkage maps have been published for both (Baranski *et al.* 2006b; Sekino and Hara 2007b), thus allowing for comparative mapping between these three important abalone species.

2. Materials and Methods

Mapped microsatellite markers for both *H. rubra* and *H. d. hannai* were downloaded from NCBI using their accession numbers as reported in Baranski *et al.* (2006b) and Sekino and Hara (2007b) and references therein. These markers were then systematically analysed as described in Chapter II to identify putative gene association, i.e. bioinformatic conversion to type I markers. The exception to this protocol, as in Chapter II, was that initial BLAST searches were only done in the non-redundant (nr) protein and refseq_rna databases to ensure at least partial exonic association; thought to demonstrate greater sequence conservation. From *H. rubra* and *H. d. hannai* 11 and 12 markers respectively, that remained anonymous after initial BLAST, were selected for comparison. The selection criteria for the anonymous markers were based on traditional measures that could be used to select microsatellite markers for transferability, but depended on the particular information available in the initial or subsequent publications (Evans *et al.* 2000, 2001; Hara and Sekino 2005; Baranski *et al.* 2006a,b Sekino *et al.* 2005, 2006; Sekino and Hara 2007a,b). These criteria included: low null allele frequencies, previous cross-species transfer and long perfect tandem repeats. After the transfer experiments, type II markers that demonstrated cross-species transfer where then again subjected to BLAST searches. This time the non-redundant (nr) nucleotide database was also included to evaluate the possible association of these loci to conserved genic features other than coding regions.

All primer sequences were taken from the original literature and initial PCR conditions were used as reported to test transfer in a preliminary panel of four individuals (Evans *et al.* 2000, 2001; Hara and Sekino 2005; Baranski *et al.* 2006a,b; Sekino *et al.* 2005, 2006; Sekino and Hara 2007a,b). Further optimisation was done where required, altering annealing temperatures appropriately (increasing incrementally if non-specific products persisted or decreasing incrementally if product signal was low) and performing buffer and/or magnesium chloride titrations for optimal concentration determination or adopting a touchdown cycle program. Go*Taq*® Flexi DNA Polymerase (Promega) or KAPA2G^{TM} Fast HotStart DNA polymerase (KAPA Biosystems) was used. All PCR reactions were done in a final volume of 10µl with final reagent concentrations as follows: for Go*Taq*® - 20ng gDNA, 1-2X Buffer, 1.5-4mM $MgCl_2$, 0.3mM dNTPs, 0.3µM of each primer and 0.5U *Taq*; for KAPA2G^{TM} – 20ng gDNA, 1X Buffer, 1.5-1.7mM $MgCl_2$, 0.2mM dNTPs, 0.2µM of each primer and 0.25U *Taq*. All PCR cycling reactions were performed on the Applied Biosystems 2720 Thermo Cycler (for final conditions for each locus refer to Table S2 in the Appendix). Amplification success was evaluated by agarose gel (2% w/v, 1X TBE; Appendix) electrophoresis, loading 3µl of PCR product mixed with 1µl of loading dye (6X Bromophenol Blue, Appendix). The electrophoresis ran for 1 hour at 120 volts. The Promega 100bp ladder was run with all samples for preliminary fragment size scoring. Optimised PCR products (single definite band), were purified for sequencing using the SigmaSpin^{TM} Post-Reaction Cleanup Columns (Sigma) as per manufacturer's specifications. This was followed by bi-directional sequencing via standard Sanger sequencing chemistry (BigDye® terminator V3.1 cycle sequencing kit, Applied Biosystems) and sent to the Stellenbosch University Central Analytical Facility (DNA sequencing unit) for capillary electrophoresis. Subsequently, sequences (reverse compliment where appropriate) were aligned with the original EST (contig or singleton) sequence by means of BioEdit version 7.0.9.0 Sequence Alignment Editor computer software (Hall 1999), using the ClustalW (Thompson *et al*. 1994) function for multiple alignments, to evaluate the presence of false positives. False positives were discarded. Thus, successful transfer was defined as a single scorable band on agarose gel, with no non-specific products that confirmed the homolgous locus in the original sequence as determined by sequencing and multiple alignments.

Microsatellite polymorphism was tested in an initial panel of eight individuals (of wild origin). PCR products were subjected to electrophoresis for two hours at 150 volts on a 12% polyacrylamide (49:1 – acrylamide:bis-acrylamide, see Appendix) gel to detect size variants. A microsatellite was deemed polymorphic when two bands were distinguishable in a single individual, thus an identifiable heterozygote, and/or there were clear size discrepancies between bands of different individuals. Fluorescently labelled primers were designed for microsatellite loci that demonstrated polymorphism. Original optimised PCR conditions for all labelled loci were used with minor adjustments for annealing temperature where necessary. PCR products using labelled primers were sent to the Central Analytical Facility of Stellenbosch University for capillary electrophoresis (ABI Genetic Analyser). A panel of 32 animals (16 each from Witsand and Saldanha) was genotyped for marker characterisation, by scoring allele size using GeneMapper® version 4 software (Applied Biosystems) (Fig. 3.3).

To test whether gene-liked markers transferred with greater efficiency, 2X2 contingency tables were constructed (for markers from *H. rubra* and *H. d. hannai* respectively) and Fisher's exact test performed (Monte Carlo approximation based on 10000 simulations) (Excell Macros downloaded from: Murdoch University, School of Chemical and Mathematical Science: (www.cms.murdoch.edu.au/areas/maths/statsnotes/inference/excelprocs2.html). The analysis was first done using only data generated in this study, but because relatively few type II markers were used, additional analyses was done incorporating type II marker data from Sekino and Hara 2007a (*H. d. hannai*) and Evans *et al.* 2001 (*H. rubra*). Departure from Hardy-Weinberg equilibrium [exact probability test; enumeration method for loci with less than four alleles and Markov chain method (20 batches, with 1000 dememorizations per batch) for loci with more than four alleles] was employed; allele frequencies, observed and expected heterozygosity, pairwise linkage disequilibrium; F_{is}-statistics and null allele frequencies (Brookfield 1996) were computed for all marker-loci using Genepop version 4 software (Rousset 2008).

The polymorphic information content (PIC) was also calculated, for all markers, using the following formula:

$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_i^{\,2}$, where *k* is equal to the total number of alleles detected and p_i is the allele frequency of the *i*-th allele.

Furthermore, to test the assumption of neutrality, an Ewens-Watterson homozygosity test was performed using the algorithm by Manly (1985) (1000 simulations) in POPGENE version 1.32 (Yeh 1999, www.ualberta.ca/~fyeh/pr01.htm).

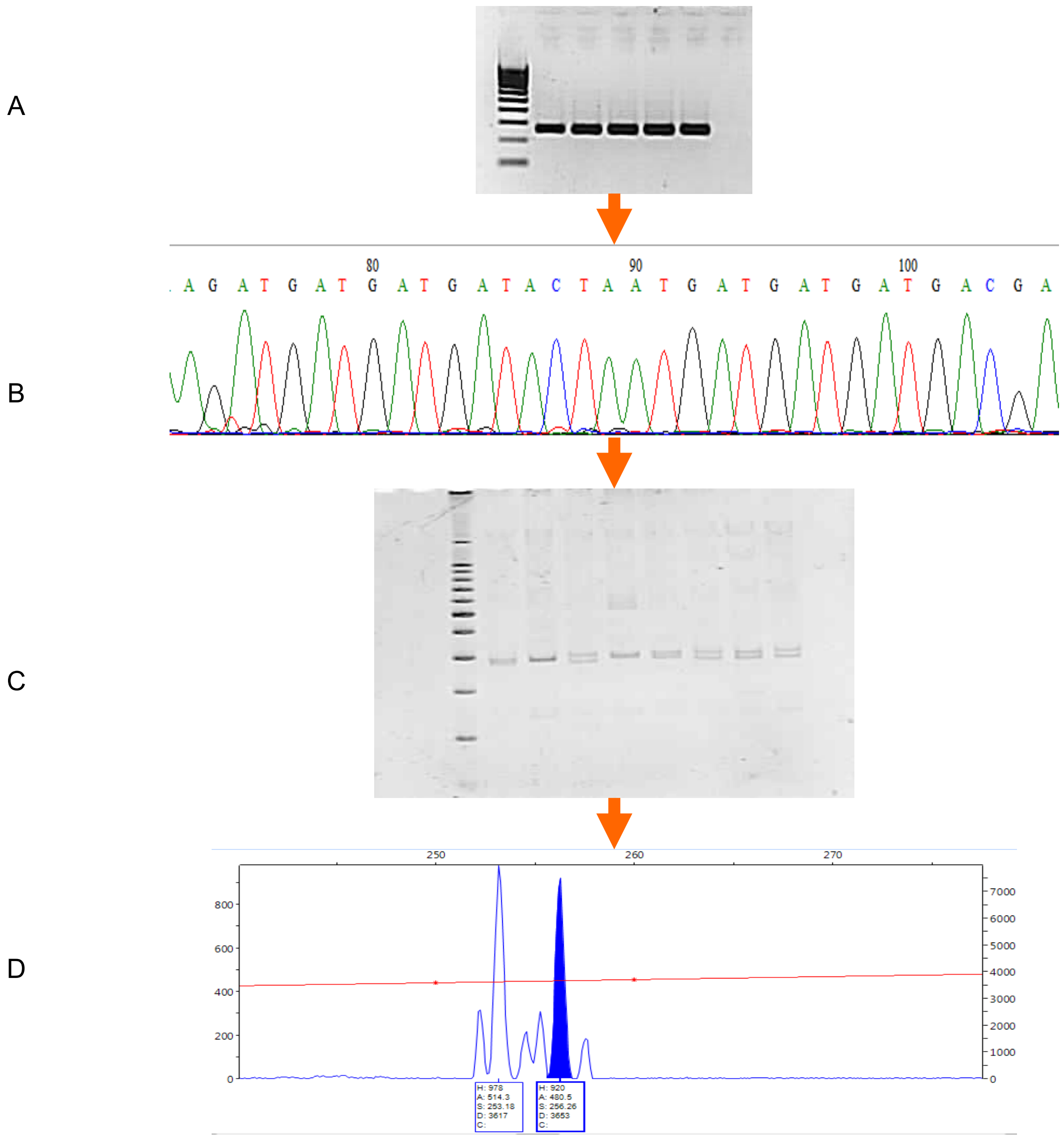


Figure 3.3: A schematic representation of the methodology followed for the identification of polymorphic cross-species microsatellites. A) Primer optimisation seen on agarose gel, single visible band with good signal. B) Sequence confirmation of the orthologous locus. C) Confirmation of polymorphism on PAGE with clear heterozygotes. D) Fragment size analysis, electropherogram for allele size scoring and genotyping.

3. Results

The initial BLAST searches resulted in 22 microsatellites (14.7%; 150 mapped, Baranski *et al.* 2006b) in *H. rubra* and 22 microsatellites (12.2%; 180 mapped, Sekino and Hara 2007b) in *H. d. hannai* being putatively gene-linked, i.e. type I markers (Table 3.2, 3.3). However, after subsequent BLAST analysis of type II markers that transferred, four additional markers were identified as type I, one from *H. rubra* (*Hrub6.C04*) and three from *H. d. hannai* (*Awb028*, *Awb041* and *Awb098*) (Table 3.4). The total number of type

I markers identified in this study was therefore 23 (15.33%) and 25 (13.8%) for *H. rubra* and *H. d. hannai* respectively, only considering the mapped markers in these species. The number of type II markers subsequently tested were ten for *H. rubra* and nine for *H. d. hannai*.

Table 3.2: Initial BLAST (nr-protein and refseq_rna) results for *H. rubra* microsatellites with gene associations (Type I markers).

Microsatellite (Genbank Acc#[1])	Tandem Repeat Motif	BLAST Hit			
		Gene (organism)	Genbank Acc#[1]	Score	E-value
Hrub1.D12 (DQ277993)	$(GT)_n(GTCT)_n$	Metalloexopeptidase (*Pichia stipitis*)	XP_001387424.1	35	2.1
Hrub10.E02 (DQ278001)	TG	Aldehyde dehydrogenase (*Branchiostoma floridae*)	XP_002217168.1	105	1.0E-21
Hrub10.H10 (DQ278006)	$(TG)_n(TC)_n$	Solute carrier family 16 (*Mus musculus*)	NM_ 028247.4	68	4.0E-09
Hrub11.A07 (DQ278009)	TG	Predicted protein (*Nemastostella vectensis*)	XP_001638172	38.5	0.19
Hrub11.A12 (DQ278011)	$(GT)_n(G)_n$	Transposase (*Strongylocentrotus purpuratus*)	XP_787284.1	43.1	0.008
Hrub12.A02 (DQ278017)	CA	Transposase (*Haemochus contortus*)	AF099908.1	41.6	0.022
Hrub12.D02 (DQ278021)	AC	Ribosomal ATPase (*Plasmoduim berylei*)	XP_680324.1	35.1	1.2
Hrub12.E10 (DQ278024)	GATG	Syndcan Domain (*Drosophila pseudoobscura*)	XP_001361704.2	76.6	6.0E-13
Hrub12.F06 (DQ278027)	$(GA)_n(GTTT)_n$	Collagen binding adhesin (*Streptococcos mutans*)	BAD13529.1	58.9	1.0E-07
Hrub15.A01 (DQ278045)	CAGA	Predicted protein (*Monosiga brevicollis*)	XP_001744239	36.2	0.03
Hrub16.D06 (DQ278047)	TG	Predicted protein (*Danio rerio*)	XP_001920139.1	35.4	1.6
Hrub16.F04 (DQ278048)	$(GT)_n(GC)_n$	Exportin (CRM1) (*Ixodes scupalaris*)	EEC067331.1	55.8	1.0E-06
Hrub16.F08 (DQ278050)	AT	NADH dehydrogenase (*Geobacter lovleyi*)	YP_001952792.1	35	2.1

Hrub16.G08 (DQ278052)	$(GT)_n(GCGT)_n$	Soduim/aminoacid transporter (*Aedes aegypti*)	XP_001654961.1	40.4	0.049
Hrub17.D11 (DQ278053)	TCCA	Immunoglobin (*Trichplax adhaeres*)	XP_002112188.1	35	2.1
Hrub4.B09 (DQ278072)	$(TA)_n(TG)_n$	Olfactory receptor (*Monodelphis domestica*)	XP_001365937.1	35	2.1
Hrub4.F07 (DQ278075)	ATGG	Unknown protein 6 (*Haliotis diversicolor*)	ABY87369.1	37.4	0.41
Hrub7.G10 (DQ278092)	AC	GAG-like protein (*Biomphalaria glabrata*)	ABN58713.1	52.8	9.0E-06
Hrub8.D02 (DQ278097)	ATGG	Immunoglobin (*Culex quinquefaciatus*)	XP_001844411.1	45.4	0.002
Hrub8.F05 (DQ278098)	CA	Discoiidin (*Branchiostoma floridae)*	XP_002209537.1	37	0.54
Hrub9.C11 (DQ278107)	GT	Unknown protein 6 (*Haliotis diversicolor*)	ABY87369.1	37	0.54
Hrub9.E04 (DQ278108)	AC	Solute symporter (*Nematostella vectensis*)	XP_001636363.1	35	2

1 – Genbank accession number.

Table 3.3: Initial BLAST (nr-protein and refseq_rna) results for *H. d. hannai* microsatellites with gene associations (Type I markers).

Microsatellite (Genbank Acc#[1])	Tandem Repeat Motif	BLAST Hit			
		Gene (organism)	Genbank Acc#[1]	Score	E-value
Afa037 (AB239614)	AC	Predicted protein (*Stronglocentrotus purpuratus*)	XP_001176024.1	38.1	0.59
Afa050 (AB239622)	AC	ATP binding cassette SubFam C (*Danio rerio*)	XP_00133398.2	100	7.0E-20
Afa068 (AB239627)	$(CA)_nA(AC)_nX(CA)_n$	Predicted protein (*Nematostella vectensis*)	MX_001628561.1	62.6	2.0E-07
Afa115 (AB239645)	TCAC	Transposase (*Salmo salar*)	ACI68988.1	74.7	5.0e-12
Afa129 (AB239651)	$(AC)_n$(TCAC)n$(AC)_n$	Predicted protein (*Nasomia vitripennis*)	XP_001602822	37	1.2
Afa162	$(CA)_nCG(CA)_n$	Mitogen activated	AAI44487.1	108	4.0E-

(AB239667)		kinase 8 (*Homo sapiens*)			22
Afa172 (AB239669)	(AGACAC)$_n$(AGACAT)$_n$ (AGACAC)$_n$(AC)$_n$X(AC)$_n$	Predicted protein (*Caenorhabditis elegans*)	T29881	36.2	2.1
Afa194 (AB239713)	TC	Predicted protein (*Paramecium tetraurelia*)	XP_001437243.1	36.2	2.2
Afa207 (AB239716)	CTCA	Dynein heavy chain (*Branchiostoma floridae*)	XP_002245841	71.6	5.0E-11
Awb022 (AB177914)	TG	Predicted protein (*Theileria parva*)	XP_765158.1	41.2	0.029
Awb044 (AB177926)	(AC)$_n$(GC)$_n$X(CT)nX(CA)$_n$	Periplasmic transporter (*Branchiostoma floridae*)	XP_002235269.1	40.4	0.049
Awb083 (AB177936)	ATC	Transposase tcb2 (*Salmo salar*)	ACI67316.1	41.6	0.086
Awb089 (AB177937)	CT	Transposase (*Salmo salar*)	ABI31711.1	80	2.0E-13
Awb101 (AB177940)	AG	Transposase (*Salmo salar*)	ACI68988.1	87.4	1.0E-15
Eab059 (AB272714)	CA	Predicted protein (*Apis mellifera*)	XP_001119920	42	0.13
Eab638 (AB272734)	CAAA	Transposase (*Portunus pelagicas*)	CAP20054.1	96.3	6.0E-18
Eab790 (AB272738)	CACT	Low density lipoprotein receptor (*Drosophila melanogaster*)	XP_001359537	41.6	0.073
Eab1125 (AB272746)	CTCA	Transbilayer Amphipath transporter (*Caenorhabditis elegans*)	NP_499363	40.4	0.16
Hd715 (AB17874)	CTCA	Unknown protein (*Haliotis diversicolor*)	AB178074	40	0.15
Ahdh553 (AB178066)	CTCA	Mitochondrial ATP synthase (*Haliotis discus*)	ABO26646.1	55.5	6.0E-06
Ahdh644	CA	Tubulin-tyrosine	AAI62916.1	55.1	5.0E-

(AB178072)		ligase (*Danio rerio)*			06
Ahdh1029 (AB178081)	TGAG	Cellulase synthase (*Escherichia fergusonii*)	YP_002384598.1	36.2	2.2

1 – Genbank accession number.

Table 3.4: Initial type II markers that transferred to *H. midae* and subsequently subjected to further BLAST analysis including nr-nucleotide database.

Microsatellite (Genbank Acc#[1])	Tandem Repeat Motif	BLAST Hit			
		Gene (organism)	Genbank Acc#[1]	Score	E-value
Hrub6.C04 (DQ278083)	CTGT	IDO-like myoglobin (*Sulculus diversicolor*)	D83984.1	60.8	5.0E-06
Awb028 (AB177917)	AC	G-alpha protein (*Haliotis rufescens*)	AF070959	284	5.0E-73
Awb041 (AB177924)	ATG	Hemocyanin (*Haliotis diversicolor*)	AB177924	131	7.0E-27
Awb098 (AB177939)	AC	Tyrosien phosphatase delta (*Haliotis discus*)	AB178066	131	2.0E-33

1 – Genbank accession number.

Overall, cross-species amplification was successful for 13 of the 33 (39%, of which 58% was polymorphic) and 7 of the 34 (21% of which 38% was polymorphic) markers for *H. rubra* and *H. d. hannai* respectively. In terms of number of type I markers for *H. rubra*, 8 of the 23 transferred (34.8%, of which 50% was polymorphic). For *H. d. hannai* 7 of 25 (28%, 43% polymorphic) type I markers transferred. In comparison 5 out of 10 (50%, 60% polymorphic) of *Haliotis rubra* type II markers amplified, while none of the *H. d. hannai* type two markers transferred (Table 3.5, 3.6). Fisher's exact test initially revealed, no statistically significant differences between the transfer efficiency of gene-associated and anonymous markers for both species to *H.* midae, using only the data generated in this study (*H. rubra*: p = 0.458 and *H. d. hannai*: p = 0.076; significant p < 0.05). However, with the incorporation of data from Sekino and Hara 2007a (*H. d. hannai*) and Evans *et al.* 2001 (*H. rubra*), the result changed slightly. Transfer rate differences remained insignificant for *H. rubra* markers (p = 1.0), but was significant for *H. d. hannai* markers (p = 0.018). Five of the *H. rubra* microsatellite markers demonstrated conserved or nearly conserved sequence repeat motif in *H. midae*, while only two markers from *H. d. hannai* demonstrated comparable conservation. Polymorphic markers across both species were more likely to maintain repeat motif conservation (Table 3.5, 3.6).

Table 3.5: *Haliotis rubra* markers tested for transfer to *H. midae*.

	Locus	Successful Transfer (Y/N)^1	Polymorphic (Y/N)^1	Microsatellite Conservation
Type I (Gene-associated markers)	*Hrub1.D12* (DQ277993)	N	-	-
	Hrub10.E02 (DQ278001)	N	-	-
	Hrub10.H10 (DQ278006)	N	-	-
	Hrub11.A07 (DQ278009)	Y	Y	Conserved GT motif
	Hrub11.A12 (DQ278011)	N	-	-
	Hrub12.A02 (DQ278017)	N	-	-
	Hrub12.D02 (DQ278021)	Y	N	(T)$_n$X(CA)$_n$X(TA)$_n$ complex (original motif: CA)
	Hrub12.E10 (DQ278024)	Y	Y	(GA)$_n$X(GT)$_n$X(CT)$_n$ complex (original motif: GAGT)
	Hrub12.F06 (DQ278027)	N	-	-
	Hrub15.A01 (DQ278045)	Y	Y	Conserved CAGA
	Hrub16.D06 (DQ278047)	N	-	-
	Hrub16.F04 (DQ278048)	N	-	-
	Hrub16.F08 (DQ278050)	N	-	-
	Hrub16.G08	N	-	-

	(DQ278052)			
	Hrub17.D11 (DQ278053)	Y	Y	Conserved TCCA
	Hrub4.B09 (DQ278072)	N	-	-
	Hrub4.F07 (DQ278075)	Y	N	Complex-interrupted T mononucleotide tract (original motif: ATGG)
	Hrub7.G10 (DQ278092)	N	-	-
	Hrub8.D02 (DQ278097)	N	-	-
	Hrub8.F05 (DQ278098)	N	-	-
	Hrub9.C11 (DQ278107)	N	-	-
	Hrub9.E04 (DQ278108)	Y	N	(G)$_n$X(A)$_n$ complex - interrupted (original motif: AC)
	Hrub6.C04 (DQ278083)	Y	N	(CTT)$_n$X(CT)$_n$ complex-interrupted (original motif: CTGT)
Type I Markers Amplified and Polymorphic (%)		8 (34.8%)	4 (50% of transferred markers)	-
Type II (anonymous markers)	*Hrub2.G01* (DQ278060)	N	-	-
	Hrub1.D03 (DQ277991)		False Positive	
	Hrub14.A04 (DQ278043)	N	-	-
	Hrub4.E05 (DQ278073)	Y	N	(C)$_n$X(T)$_n$ complex-interrupted (original motif: CA)
	Hrub1.H08 (DQ277997)	N	-	-
	Hrub2.B01 (DQ278057)	N	-	-
	Hrub12.B10	Y	Y	Conserved CAA

	(DQ278020)			
	Hrub13.F06 (DQ278037)	Y	Y	Interrupted poly-G-mononuleotide tract (original motif: GT)
	Hrub9.B05 (DQ278104)	Y	Y	AT (original motif: AC)
	Hrub16.G01 (DQ278051)	Y	N	Interrupted-T-mononuleotide tract (original motif: GT)
Type II Markers Amplified and Polymorphic (%)		5 (50%)	3 (60% of transferred markers)	-
Total Markers Amplified and Polymorphic (%)		13 (39%)	7 (53.8% of transferred markers)	-

1 – Yes/No. For primer information refer to Appendix.

Table 3.6: *Haliotis discus hannai* markers tested for transfer to *H. midae*.

	Locus	**Successful Transfer (Y/N)**[1]	**Polymorphic (Y/N)**[1]	**Microsatellite Conservation**
Type I (Gene-associated markers)	*Afa037* (AB239614)	N	-	-
	Afa050 (AB239622)	N	-	-
	Afa068 (AB239627)	Y	N	Conserved CA motif
	Afa115 (AB239645)	N	-	-
	Afa129 (AB239651)	N	-	-
	Afa162 (AB239667)	N	-	-
	Afa172 (AB239669)	N	-	-
	Afa194 (AB239713)	Y	N	Complex-interrupted $(GT)_nX(CT)_n$ (original motif: CT)
	Afa207	Y	N	Complex-

	(AB239716)			interrupted CACT (original motif: CACT)
	Awb022 (AB177914)	N	-	-
	Awb044 (AB177926)	N	-	-
	Awb083 (AB177936)	Y	Y	TC-complex-interrupted (original motif: ATC)
	Awb089 (AB177937)	N	-	-
	Awb101 (AB177940)	N	-	-
	Eab059 (AB272714)	N	-	-
	Eab638 (AB272734)	N	-	-
	Eab790 (AB272738)	N	-	-
	Eab1125 (AB272746)	N	-	-
	Hd715 (AB17874)	N	-	-
	Ahdh553 (AB178066)	N	-	-
	Ahdh644 (AB178072)	False positive		
	Ahdh1029 (AB178081)	N	-	-
	Awb028 (AB177917)	Y	N	Complex motif (original motif: AC)
	Awb098 (AB177939)	Y	Y	Conserved AC
	Awb041 (AB177924)	Y	Y	Interrupted TTAGGG (original motif: ATG)
Type I Markers Amplified and Polymorphic (%)		7 (28%)	3 (42.9% of transferred markers)	-

Type II (anonymous markers)	Afa005 (AB177904)	N	-	-
	Awb068 (AB177931	N	-	-
	Awb039 (AB177923)	N	-	-
	Awb052 (AB177927)	N	-	-
	Afa195 (AB239714)	N	-	-
	Awb033 (AB177918)	N	-	-
	Afa066 (AB239626)	N	-	-
	Afa107 (AB239642)	N	-	-
	Afa185A (AB239675)	N	-	-
Type II Markers Amplified and Polymorphic (%)		0	0	-
Total Markers Amplified and Polymorphic (%)		7 (20.5%)	3 (37.5% of transferred markers)	-

1 – Yes/No. For primer information refer to Appendix.

The number of alleles per locus ranged from 2 to 10 and average observed and expected heterozygosity was 0.57 (range: 0.28 - 0.84) and 0.58 (range; 0.26 - 0.83), respectively. The average PIC value was equal to 0.57 (range: 0.25 - 0.81). Over all loci there were no apparent heterozygous excess or deficit (average F_{is} = -0.003, range: -0.69 - 0.59). Only two loci deviated from Hardy-Weinberg equilibrium and no significant linkage disequilibrium ($p < 0.05$) was observed. None of the observed F-values for homozygosity (Ewens-Watterson test) fell outside the 95% confidence levels, indicating neutrality (Table 3.7).

Table 3.7: Population statistics for polymorphic transferred microsatellite markers.

Locus Name (Acc#)[1]	k[2]	PIC[3]	Fr(A)[4] Range	Allele Size Range (bp)[5]	H_O[6]	H_E[7]	p-value for HWE[8]	F_{is}	Fr(A_{nul})[9]
Hrub11.A07 (DQ278009)	10	0.67	0.02-0.53	88-118	0.69	0.68	0.07	-0.013	0.0
Hrub13.F06 (DQ278037)	7	0.66	0.02-0.48	203-219	0.28	0.69	0.0*	0.59	0.26
Hrub12.B10 (DQ278020)	8	0.72	0.02-0.29	237-250	0.66	0.74	0.16	0.11	0.085
Hrub9.B05 (DQ278104)	8	0.81	0.02-0.29	180-190	0.84	0.81	0.87	-0.03	0.159
Hrub17.D11 (DQ278053)	3	0.25	0.02-0.85	216-250	0.29	0.26	1.0	-0.13	0.0
Hrub15.A01 (DQ278045)	9	0.81	0.02-0.22	278-302	0.73	0.83	0.28	0.12	0.05
Hrub12.E10 (DQ278024)	4	0.44	0.03-0.72	825-837	0.56	0.46	1.0	-0.24	0.0
Awb083 (AB177936)	4	0.4	0.03-0.75	176-210	0.5	0.41	0.84	-0.23	0.0
Awb041 (AB177924)	2	0.5	0.45-0.55	80-90	0.84	0.5	0.0002*	-0.69	0.35
Awb098 (AB177939)	5	0.57	0.02-0.75	88-157	0.41	0.42	0.74	0.04	0.006

1 – Genbank Accession number; 2 – Number of Alleles; 3 – Polymorphic information content; 4 – Allele frequency; 5 – base pairs; 6/7 – Observed and Expected Heterozygosity respectively; 8 – Hardy-Weinberg Equilibrium ("" – significant deviation, p < 0.05); 9 – Null allele frequency (for primer information, refer to the Appendix).*

4. Discussion

Microsatellite cross-species transfer is deemed a quick and inexpensive alternative for *de novo* marker development. This method has been employed for numerous species of conservational and/or economic importance, especially those with limited genomic resources (e.g. Fritsimmons *et al.* 1995; Morin *et al.* 1998; Cairney *et al.* 2000; Primmer and Merliä 2002; Paterson *et al.* 2004; Piñera *et al.* 2006*;* Freitas *et al.* 2007; Augustinos *et al.* 2008; Galarza *et al.* 2009; Miles *et al.* 2009; Stratikopoulos *et al.* 2009; Tian *et al.*

2009; Zhao *et al.* 2009). Nevertheless, the success of such marker transfer endeavours demonstrate great variation amongst taxa (Fig. 3.1, Barbará *et al.* 2007) and molluscs in general appear to display moderate to low levels of cross-species microsatellite transfer (Winnepenninckx and Backeljau 1998; Eackel *et al.* 2002; Jones *et al.* 2004; Zhan *et al.* 2005; Ibarra *et al.* 2006; Cruz *et al.* 2007; Pérez *et al.* 2007; DuBios *et al.* 2008; Panova *et al.* 2008). Estimates for abalone concur with the general molluscan observation of relatively lower transfer rates over short evolutionary periods (Huang and Hanna 1998; Evans *et al.* 2001; Cruz *et al.* 2005; Baranski *et al.* 2006a; Sekino and Hara 2007a; Díaz-Vilotia *et al.* 2008). The aim here, was thus to improve microsatellite transfer rates to the South African abalone *Haliotis midae* from two other economically important species: *Haliotis rubra* (Australian species) and *Haliotis discus hannai* (Asian species). To achieve this, public databases were used to find genic homology to anonymous microsatellites that theoretically will transfer with greater ease. In the process, cross-species type I markers for future comparative and functional mapping were also established.

Success of microsatellite transfer agreed with the expected inverse relationship between phylogenetic distance and transfer rate (Sekino and Hara 2007a): amplification success was higher for *H. rubra* (a confirmed sister taxon, Bester-Van der Merwe 2009) to *H. midae* (39%) than for *H. d. hannai* to *H. midae* (23.5%). Previous studies of cross-species amplification of microsatellite markers from *H. rubra* to *H. midae* reported a positive amplification result of 45% (10 of 22 markers). However taking into account that two markers demonstrated non-specific amplification and that the authors postulate that a further two markers were likely to be false positives, based on the definition of successful cross-species transfer used in the current study, their success rate was only 27.3% (6 of 22 markers) (Evans *et al.* 2001). The higher value reported here, 39%, for marker transfer from *H. rubra* to *H. midae* could be due to the use of type I (gene-linked) markers. However, differential cross-species transfer of type I versus type II markers failed to reach statistical significance for *H. rubra* microsatellites and thus warrant further investigation in future, perhaps including a larger marker cohort of type I and II markers. On the contrary after the addition of anonymous marker data from Sekino and Hara (2007a), a statistically significant result was obtained for cross-species transfer between type I and type II microsatellites from *H. d. hannai.* This explains the higher transfer success reported in this study, where only type I markers produced a scorable product (23.5% versus 5%, Sekino and Hara 2007a). It should be noted that Sekino and Hara (2007a) tested 24 *H. d. hannai* microsatellites for transfer to *H. midae* and reported five initial amplifications, however one locus presented non-specific products and three loci only produced a PCR product in one individual. Furthermore, their paper makes no reference to any method for false positive elimination, thus the locus that demonstrated non-specific by-products was viewed as an unsuccessful transfer attempt and the three loci demonstrating limited individual amplifications were disregarded in subsequent analyses.

It would therefore seem that genic association of microsatellites markers becomes more important as the phylogenetic distance between the source and target species increases. This is most probably due to closer related species still sharing sufficient conservation of microsatellite flanking regions in both genic and intergenic sequences; while in distantly related species there is a marked difference in genic (which is functionally constrained) versus intergenic (subject to neutral evolution) sequences. However, this does not account for the relatively low transfer rate in closely related species. “Random species effects”, that alter primer binding sites may explain this phenomenon, thus primer redesign should rectify the problem, at least in some instances (Primmer *et al.* 2005). Support for this in abalone is the relatively high SNP frequency (Bester *et al.* 2008), as well as the hyper-variable sequence flanking regions of microsatellites (Rhode *et al.* 2008). Both these features of abalone genomes increase the probability of primer binding site sequence variation, the extent of which remains to be evaluated. A further factor that could influence transfer rate is the criteria for selecting type II markers. In this study the type II markers from *Haliotis rubra* were selected from Baranski *et al.* (2006a). The authors gave sufficient accounts of null alleles, repeat motifs and cross-species amplification in other Australian abalone. Coupled with the fact the South African species and Australian species share a recent common ancestor, a combination of all these characteristics could over shadow the single factor of being gene-linked. Therefore traditional selection criteria remain important when selecting markers for transfer. The information available on *H. d. hannai* markers was not as complete, subsequently only one or two factors were taken into consideration upon type II marker selection.

The pattern of microsatellite sequence conservation amongst these species demonstrated results concurring with the general expectations. Microsatellite structure tended to remain relatively conserved among type I markers, with four (of eight) of the *H. rubra* type I markers retaining the core repeat motif in *H. midae* (Table 3.5) and three (of seven) of the *H. d. hannai* type I marker (Table 3.6). It is also noted that the conserved perfect microsatellite repeats were more prone to exhibiting multiple alleles (Chambers and McAvoy 2000). The conservation of type I microsatellite repeats is indicative of a possible conserved functional role of these loci (Zeiss *et al.* 1998; Li *et al* 2004). Monomorphic loci presented, in most cases, imperfect (complex, interrupted, compound) microsatellites, explaining their lack of polymorphism. These imperfect repeats are produced by point mutation accumulation in the repetitive tract, in agreement with the life cycle hypothesis of microsatellites. Ultimately this will lead to the decay of the microsatellites and formation of a region of cryptic simplicity that then theoretically may jumpstart the regeneration of a new microsatellite (Ellegren 2004). Such mutations in abalone microsatellites seem to be prevalent within species. A quick survey of microsatellites revealed ~26.4% of the microsatellites in *H. rubra* to be imperfect (Baranski *et al.* 2006a), ~36.5% for *H. midae* (Bester *et al.* 2004; Slabbert *et al.* 2008, in press, in prep.) and as much as 70.7% for *H. d. hannai* (Sekino *et al.* 2006). It is thus expected that such tandem repeat interruptions will transcend across species boundaries (Liu and Ely 2009) and may in fact be the lead cause for the failure of microsatellite transfer (Chambers and McAvoy 2000; Primmer *et al.* 2005). Furthermore, two loci (*Hrub12.D02*, *H. rubra* and *Afa068*, *H. d. hannai*) demonstrated a conserved repeat motif (*Hrub12.D02* has a complex repetitive structure, however the CA motif remains prominent), however did not show multiple alleles. Both these loci are gene linked (Table 3.2, 3.3) thus it can be hypothesised that the monomorphism of these loci could possibly be attributed to the fixation of particular alleles at the point of ancestral divergence which was subsequently maintained by purifying selection. Restricted tandem repeat length has been implicated in gene functionality (Gebhardt *et al.* 1999, 2000).

Interestingly three loci (*Hrub9.B05*, *Awb083* and *Awb041*; Table 3.5, 3.6) demonstrated a new microsatellite motif in *H. midae*, differing from that originally reported for the source species. At present, based on the available data, at least two postulates may be put forth to explain this observation. Firstly, these loci may probably not represent orthologs, but rather the amplification of a paralog. Considering that microsatellites often have their origin in transposable elements (Nadir *et al.* 1996; Miller *et al.* 2000; Grover *et al.* 2007) this is highly probable. Thus these loci as paralogs may share a common transposable element that located in two separate genomic regions allowing their associated microsatellites to evolve independently, leading to alternate repetitive motifs. This may hold particular relevance for locus *Awb083*, noting its significant BLAST hit to a *transposase* gene (Table 3.3). On the other hand assuming that these loci are true orthologs, point mutations may pose an opportunity for microsatellite transitions, where such a point mutation creates a new repeat unit that is subsequently propagated via replication slippage (Ellegren 2004). The prevalence of compound microsatellites, i.e microsatellites demonstrating two unique, yet continuous repeat motifs is testament to this. With closer examination of locus *Hrub9.B05,* remnants of the original CA motif can still be seen, as the first two repeat units are still CA. However a point mutation leading to a C > T transition changed the microsatellite to the current predominant TA motif that is responsible for the observed polymorphism. The other two loci under discussion (*Awb083* and *Awb041*), do not at all demonstrate a compound repeat structure. This may be explained by differential repeat motif selection during the speciation event, leading to the decay of the non-functional repeat motif. This may be the case particularly for microsatellites that are associated to genes, as is *Awb083* and *Awb041.*

Several of the microsatellite markers from *H. rubra* investigated here has shown significant association to growth-related traits in a recent study by Baranski *et al.* (2008) (type I markers: *Hrub12.D02*, *Hrub16.F06*, *Hrub7.G10*, *Hrub8.F05*, *Hrub9.C11*, *Hrub9.E04* and type II: *Hrub2.G01*, *Hrub1. D03*, *Hrub12.B10*, *Hrub12.B10*) and may confer QTLs in *H. midae* as well.

Surprisingly only two of the type I markers (*Hrub12.D02* and *Hrub9.E04*) successfully transferred to *H. midae* and these failed to be polymorphic. The lack of polymorphism is to be expected if these loci represent a possible speciation event, also noting their complex repeat structure (Table 3.5). However if a microsatellite holds possible functionality within a gene it is likely that it may be conserved and thus the lack of transfer may be due to the fixation of null alleles in *H. midae*, therefore primer redesign is warranted for future investigation. A similar argument may be presented for type II markers even though they are not currently known to be associated to genes, the possibility cannot be excluded. One type II microsatellite in particular is noteworthy: *Hrub12.B10*. This marker demonstrates transfer, polymorphism and a conserved CAA repeat motif in *H. midae.* Furthermore, it confers a highly statistically significant QTL ($p < 0.001$) in *H. rubra* (Baranski *et al.* 2008). The combination of these factors suggests conserved functionally of this locus and makes it a candidate QTL for growth in *H. midae.* None of the *H. d. hannai* markers are currently known to be associated to a quantitative trait (Liu *et al.* 2007).

It is to be expected that at the individual locus level more type II than type I markers would demonstrate polymorphism, as type I markers may be functionally constrained (Metzger *et al.* 2000). Furthermore, the higher percentage of polymorphic *H. rubra* cross-species microsatellites, in comparison to *H. d. hannai* cross-species microsatellites is consistent with the hypothesis that microsatellite polymorphism decreases with phylogenetic distance between source and target species, taking into account the life cycle hypothesis for microsatellite evolution (Rubinsztein *et al.* 1995; Morin *et al.* 1998). The foregoing polymorphism of transferred markers is supported by the diversity estimates for these loci. The heterozygosity, PIC-values and number of alleles (Table 3.6) exhibits a particular disposition associated with moderately polymorphic microsatellites, in contrast to their original estimates in the source species showing higher values (Sekino *et al.* 2005; Baranski *et al.* 2006a). Similar studies reported comparable results (Jones *et al.* 2004; Piñera *et al.* 2006). Only two loci (*Hrub13.F06* and *Awb041*) deviated from Hardy-Weinberg expectation, most likely due to the persistence of null alleles, as both these loci also demonstrated the highest null allele frequencies (Table 3.6).

During the course of this study seven polymorphic, type I cross-species microsatellite markers have been developed for *H. midae* (three from *H. d. hannai* and four from *H. rubra*) and three polymorphic, type II cross-species microsatellites from *H. rubra*. Eleven monomorphic markers (five and six, from *H. d. hannai* and *H. rubra* respectively) were also identified. Where it is envisioned that polymorphic markers could be employed in general application for abalone aquaculture, including linkage mapping and diversity estimates for population management, the utility of monomorphic markers will be invaluable in forensic applications for individual species identification. Also these markers provide future utility for comparative genomic mapping and evolutionary investigation of abalone speciation.

5. Conclusions

Microsatellite cross-species transfer provides important insight into the genomic organisation of related species; and is thus routinely used in comparative and evolutionary studies. From a molecular breeding standpoint it allows genomic information from a well characterised and mapped genome (as those from *H. rubra* and *H. d. hannai*) to be superimposed on a less understood species such as *H. midae*. Several microsatellite markers investigated here have been shown to be growth-related QTLs in *H. rubra* and once transfer success has been achieved, these loci may act as candidate QTLs in *H. midae* in future investigations. Previous microsatellite cross-species transfer attempts in abalone reported low success rates. To counter this, a bioinformatic conversion of type II to type I markers was proposed as a selection criterion for markers to be transferred. The results demonstrated marginally increased transfer success, however failed to achieve statistical significance for microsatellite markers from *H. rubra*. Thus this approach may yield greater success for distantly related species where a marked difference in sequence conservation between genic and intergenic genomic regions is apparent. Furthermore, traditional selection criteria, null allele frequency, repeat motif character and previous cross-species transfer success remains important determinants; in fact combing genic and traditional criteria will most likely achieve the greatest success rate. Microsatellite motif conservation varied, however type I markers did show slightly improved marker conservation, indicating a possible conserved functional role of these loci.

References

1. Augustinos AA, Stratikopoulos EE, Drosopoulou E, Kakani EG, Mavragani-Tsipidou P, Zacharopoulou A, Mathiopoulos KD (2008) Isolation and characterization of microsatellite markers from the olive fly, *Bactrocera oleae*, and their cross-species amplification in the *Tephritidae* family. BMC Genomics 9: 618.
2. Baranski M, Rourke M, Loughnan S, Austin CM, Robinson N (2006a) Isolation and characterization of 125 microsatellite DNA markers on blacklip abalone, *Haliotis rubra.* Mol Ecol Notes 6: 740-746.
3. Baranski M, Loughnan S, Austin CM, Robinson N (2006b) A microsatellite linkage map of the blacklip abalone, *Haliotis rubra*. Anim Genet 37: 563-570.
4. Baranski M, Rourke M, Loughnan S, Hayes B, Austin C, Robinson N (2008) Detection of QTL for growth rate in blacklip abalone (*Haliotis rubra* Leach) using selective DNA pooling. Anim Genet 39: 606-614.
5. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol Ecol 16: 3759-3767.
6. Bester AE, Slabbert R, D'Amato ME (2004) Isolation and characterisation of microsatellite markers in South African abalone (*Haliotis midae*). Mol Ecol Notes 4: 618-619.
7. Bester AE, Roodt-Wilding R, Whitaker HA (2008) Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae*: a targeted EST approach. Anim Genet 39: 321-324.
8. Bester-Van der Merwe AE (2009) Population genetic structure and demographical history of South African abalone, *Haliotis midae,* in a

9. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453-455.

10. Cairney M, Taggart JB, Høyheim B (2000) Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Mol Ecol 9: 2175-2179.

11. Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Phys B 126: 455-476.

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

13. Cruz P, Ibarra AM, Fiore-Amaral G, Galindo-Sánchez CE, Mendoza-Carrión 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. Mol Ecol Notes 5: 857-859.

14. Cruz P, Yáñez-Jacome B, Ibarra AM, Rangel-Becrril J (2007) Isolation and characterization of microsatellite loci in the Pacific pleasure oyster, *Crassostrea corteziensis*, and their cross-species amplification in four other oyster species. Mol Ecol Notes 7: 448-450.

15. Díaz-Viloria N, Pérez-Enríquez R, Fior-Amaral G, Burton RS, Cruz P (2008) Isolation and cross-amplification of microsatellites in pink abalone (*Haliotis corrugata*). Mol Ecol Res 8: 701-703.

16. DuBios M-P, Nicot A, Jarne P, David P (2008) Characterization of 15 microsatellite markers in the freshwater snail *Aplexa marmorata* (*Mollusca, Gastropoda*). Mol Ecol Res 8: 1062-1064.

17. Eackles MS, King TL (2002) Isolation and characterization of microsatellite loci in *Lampsilis abrupta* (*Bivalvia: Unionidae*) and cross-species amplification with in the genus. Mol Ecol Notes 2: 559-562.

18. Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution – a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol Biol Evol 14: 854-860.

19. Ellegren H (2004) Microsatellite: simple sequences with complex evolution. Nat Rev Genet 5: 435-445.

20. Evans B, White RWG, Elliot NG (2000) Characterization of microsatellite loci in the Australian Blacklip abalone (*Haliotis rubra* Leach). Mol Ecol 9: 1183-1182.

21. Evans B, Conod N, Elliot NG (2001) Evaluation of microsatellite primer conservation in abalone. J Shellfish Res 20: 1065-1070.

22. Freitas P, Jesus CL, Galetti PM (2007) Isolation and characterization of new microsatellite loci in the Pacific white shrimp *Litopenaeus vannamei* and cross-species amplification in other penaeid species. Mol Ecol Notes 7: 324-326.

23. FritzSimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol Biol Evol 12: 432-440.

24. Galarza JA, Boulay R, Cerdá X, Doums C, Federici P, Magalon H, Monnin T, Rico C (2009) Development of single nucleotide sequence repeat markers for the ant *Aphaenogaster senilis* and cross-species amplification in *A. iberica*, *A. gibbosa*, *A. subterranean*, and *Messor maroccanus*. Conserv Genet 10: 519-521.

25. Gebhardt F, Zanker KS, Brandt B (1999) Modulation of epidermal factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem 274:13176-13180.

26. Gebhardt F, Burger H, Brandt B (2000) Modulation of EGFR gene transcription by a polymorphic repetitive sequence – a link between genetics and epigenetics. Int J Biol Marker 15: 105-110.

27. Geiger DL (2000) Distribution and biogeography of the *Haliotidae* (Gastropoda: Vetigastropoda) world-wide. Boll Malacol 35: 57-120.

28. Grover A, Aishwarya V, Sharma PC (2007) Bias distribution of microsatellite motifs in the rice genome. Mol Genet Genomics 277: 469-480.

29. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95-98.

30. Hara M, Sekino M (2005) Genetic difference between Ezo-awabi *Haliotis discus hannai* and Huro-awabi *H. discus discus* populations: microsatellite-based population analysis in Japanese abalone. Fisheries Sci 71: 754-766.

31. Harr B, Zangerl B, Brem G, Schlötterer C (1995) Conservation of microsatellite locus variability across two *Drosophila* sibling species. *D. melanogaster* and *D. simulans.* Mol Biol Evol 15: 176-184.

32. 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.* J Shellfish Res 23: 379-385.

33. Huang B, Hanna PJ (1998) Identification of three polymorphic microsatellite loci in blacklip abalone, *Haliotis rubra* (Leach), and detection in other abalone species. J Shellfish Res 17: 795-799.

34. Ibarra AM, Petersen JA, Famula TR, May B (2006) Characterization of 35 microsatellite loci in the Pacific lionpaw scallop (*Nodipecten subnodosus*) and their cross-species amplification in four other scallops of the *Pectinidae* family. Mol Ecol Notes 6: 153-156.

35. Jones JW, Cluver M, David V, Struthers J, Johnson NA, Neves RJ, O'Brien SJ, Hallerman EM (2004) Development and characterization of microsatellite loci in the endangered oyster mussel *Epioblasma capsaeformis* (Bivalvia: Unionidae). Mol Ecol Notes 4: 649-652.

36. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111-120.

37. Li W-H, Yi S, Makova K (2002) Male-driven evolution. Curr Opin Genet Dev 12: 650-656.

38. Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 21: 991-1007.

39. Martin AP, Pardini AT, Noble LR Jones CS (2002) Conservation of a dinucleotide simple sequence repeat locus in sharks. Mol Phylogenet Evol 23: 205-213.

40. Liu J-X, Ely B (2009) Complex evolution of a highly conserved microsatellite locus in several fish species. J Fish Biol 75: 442-447.

41. Liu X, Liu X, Zhang G (2007) Identification of quantitative trait loci for growth-related traits in Pacific abalone *Haliotis discus hannai* Ino. Aquac Res 38: 789-797.

42. Manly BFJ (1985) Linkage disequilibrium and selection at two or more loci. In: Usher MB, Rosenzwig ML (eds) The statistics of natural selection on animal populations. Chapman and Hall, London, pp. 309-341.

43. Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10: 72-80.

44. Miles LG, Lance SL, Isberg SR, Moran C, Glenn TC (2009) Cross-species amplification of microsatellites in crocodilians: assessment and applications for the future. Conserv Genet 10:935-954.

45. Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. Mol Biol Evol 17: 1597-1609.

46. Moore SS, Sargeant LL, King TJ, Mattick JS, George M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10: 654-660.

47. Morin PA, Mahboubi P, Wedel S, Rogers J (1998) Rapid screening and comparison of human microsatellite markers in baboons: Allele size is conserved, but allele number is not. Genomics 53: 12-20.

48. Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Nat Acad Sci USA 93: 6470-6475.

49. Neff BD, Gross MR (2001) Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. Evolution 55: 1717-1733.

50. Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. Nat Rev Genet 7:851-861.

51. Noor MAF, Kliman RM, Machado CA (2001) Evolutionary history of microsatellites in the Obscura group of *Drosophila*. Mol Biol Evol 18: 551-556.

52. Palo JU, Mäkinen HS, Helle E, Stenman O, Väinölä R (2001) Microsatellite variation in ringed seals (*Phoca hispida*): genetic structure and history of the Baltic Sea population. Heredity 86: 609-617.

53. Panova M, Mäkinen T, Fokin M, André C, Johannesson K (2008) Microsatellite cross-species amplification in the genus *Lottorina* and detection of null alleles in *Littorina saxatilis*. J Mollus Stud 74: 111-117.

54. Paterson S, Piertney SB, Knox D, Gilbey J, Verspoor E (2004) Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Mol Ecol Notes 4: 160-162.

55. Pépin L, Amigues Y, Lépingle A, Berthier J-L, Bensaid A, Vaiman D (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Example of use in parentage testing and phylogeny analysis. Heredity 74: 53-61.

56. Pérez M, Branco M, Llavona A, Riberio PA, Santos AM, Hawkins SJ, Dávila PA, Presa P, Alexandrino P (2007) Development of microsatellite loci for the black-footed limpet, *Patella depressa*, and cross-amplification in two other *Patella* species. Conserv Genet 8: 739-742.

57. Piñera JA, Bernardo D, Blanco G, Vázquez E, Sánchez JA (2006) Isolation and characterization of polymorphic microsatellite markers in *Pagellus bogaraveo*, and cross-species amplification in *Sparus aurata* and *Dicentrarchus labrax*. Mol Ecol Notes 6: 33-35.

58. Primmer C, Meriliä J (2002) A low rate of cross-species microsatellite amplification success in Ranid frogs. Conserv Genet 3: 445-449.

59. Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J (2005) Factors affecting avian cross-species microsatellite amplification. J Avian Biol 36: 348-360.

60. Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite flanking regions: a SNP mine in South African abalone (*Haliotis midae*). Anim Genet 39: 329.

61. Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. P R Soc Lond B Bio 263: 549-557.

62. Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8: 103-106.

63. Rubsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li S-H, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution – Evidence for directionality and variation in rate between species. Nat Genet 10: 337-343.

64. Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. Nature 354: 63-65.

65. Sekino M, Hara M (2007a) Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus: *Haliotis*). Conserv Genet 8: 823-841.

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

67. Sekino M, Saido T, Fujita T, Kobayashi T, Takami H (2005) Microsatellite DNA markers of the Ezo abalone (*Haliotis discus hannai*): a preliminary assessment of natural populations sampled from heavily stocked areas. Aquaculture 243: 33-47.

68. Sekino M, Kobayashi T, Hara M (2006) Segregation and lineage analysis of 75 novel microsatellite DNA markers in pair crosses of Japanese Abalone (*Haliotis discus hannai*) using the 5'-tailed primer method. Mar Biotechnol 8: 453-466.

69. Slabbert R, Ruivo NR, Van den Berg NC, Lizamore DL, Roodt-Wilding R (2008) Isolation and characterization of 63 microsatellite loci for the abalone, *Haliotis midae*. J World Aquacult Soc 39: 429-435.

70. Slabbert R, Hepple J, Venter A, Nel S, Swart L, Van den Berg NC, Roodt-Wilding R (in press) Isolation and segregation of 44 microsatellite loci in the South African abalone, *Haliotis midae* L. Anim Genet. DOI: 10.1111/j.1365-2052.2009.02003.x.

71. Stratikopoulos EE, Augustinos AA, Pavlopoulos ID, Economou KP, Mintzas A, Mathiopoulos KD, Zacharopoulou A (2009) Isolation and

characterization of microsatellite markers from the Mediterranean fruit fly, *Ceratitis capitata*: cross-species amplification in other *Tephritidae* species reveals a varying degree of transferability. Mol Genet Genomics 282: 283-306.

72. Thomson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequences weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

73. Tian Y-S, Miao G-D, Shoa C-W, Liao X-L, Chen S-L (2009) Isolation and characterization of polymorphic microsatellite loci from repeat-enriched genomic libraries of stone flounder (*Kareius bicoloratus*) and cross-species amplification. Conserv Genet 10: 1041-1043.

74. Winnepenninckx B, Backeljau T (1998) Isolation and characterization of microsatellites markers in the periwinkle *Littorina striata* King and Broderip, 1832 (*Mollusca, Gastropoda, Prosobranchia*). Mol Ecol 7: 1253-1254.

75. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11: 1-16.

76. Zeiss CJ, Trepanier LA, Aguirre GD, Ray K (1998) A highly conserved microsatellite in the dystrophin gene of diverse mammalian species. Anim Genet 29: 224-227.

77. Zhan A-B, Bao ZM, Wang XL, Hu JJ (2005) Microsatellite markers derived from bay scallop *Argopecten irradians* expressed sequence tags. Fish Sci 71: 1341-1346.

78. Zhao L, Shao C, Liao X, Ma H, Zhu X, Chen S (2009) Twelve novel polymorphic loci for the Yellow grouper (*Epinphelus awoara*) and cross-species amplifications. Conserv Genet 10: 743-745.

Chapter IV:
Development of EST-microsatellites and EST-SNPs from various *Haliotidae* for transfer to the endemic *Haliotis midae*

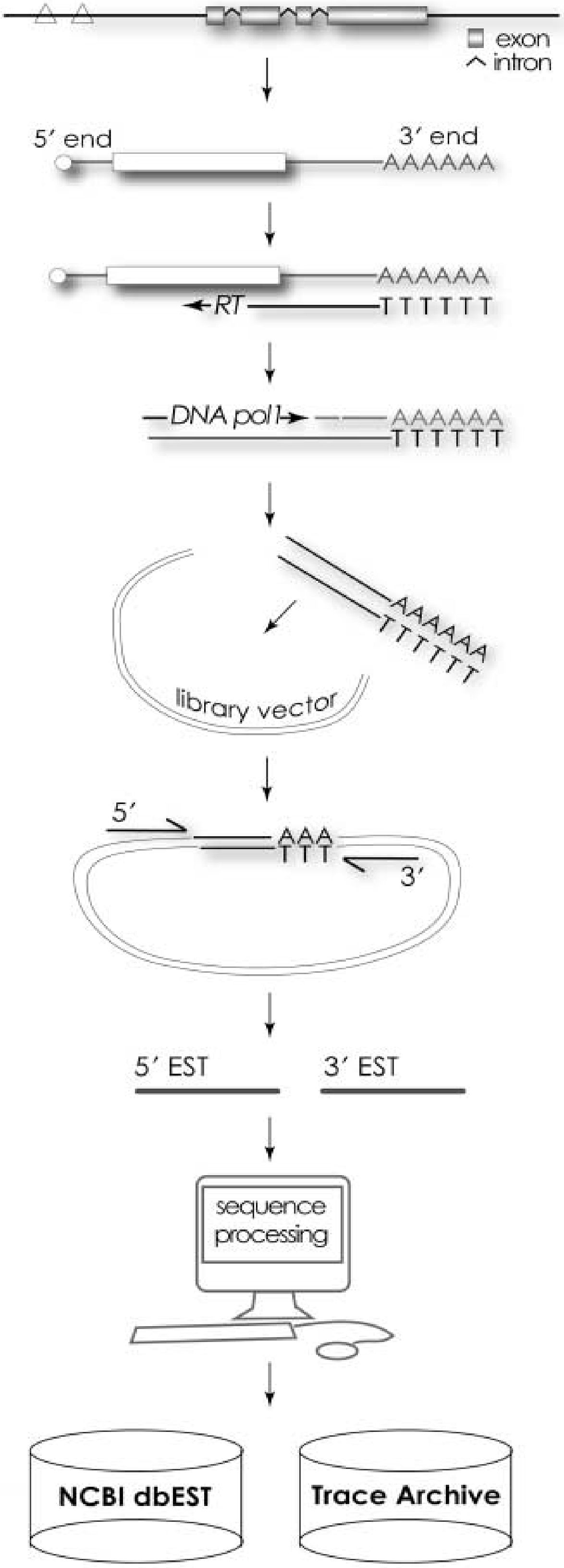
Abstract

Recent transcriptome sequencing projects, as an alternative to whole genome sequencing, and the development of micro-array technologies for gene expression studies have led to the rapid accumulation of EST data for many species. The data provides a ready source for type I molecular marker development, even in the absence of sequence data for the focal species, as these genic sequences demonstrate higher conservation and thus will transfer more readily between related species. In the present study, this resource is utilised to develop microsatellite and SNP markers associated to expressed sequences for the abalone, *Haliotis midae*. Because *H. midae* is currently underrepresented in these databases, a cross-species transfer strategy was adopted using EST sequences from other Haliotids. Sixteen SNPs and 15 microsatellites were characterised and the majority of these markers showed moderate polymorphism. This is probably due to functional constraints, in agreement with current evidence presented in similar studies. Putative function could be assigned to 11 sequences, based on similarity to known or putative genes. This strategy proved to be successful, but not without limitations. Thus it would not replace main-steam laboratory techniques, such as genomic library construction and amplicon resequencing, but remains a cost- and time-efficient method for saturating marker maps and developing molecular markers for diversity studies.

1. Introduction

Expressed sequence tags (generally abbreviated as ESTs) are probably the most abundant DNA sequence data currently available in the public domain (Bouck and Vision 2007). Typically, ESTs constitute single-read, automatically processed sequences derived from clones of cDNA libraries (Fig 4.1). Such cDNA libraries are routinely constructed to provide snapshots of gene expression within a tissue/organ/organism at particular developmental phases and/or under certain environmental conditions (Rudd 2003). These cDNA libraries were first used as a means of gene discovery and later as a tool for genomic annotation of the functional regions of the genome (Putney *et al.* 1983; Brenner 1990; Adams *et al.* 1991). Furthermore, with the advent of cDNA array-based technologies the utility of ESTs as an integral component of the experimental analysis become paramount (Gress *et al.* 1992).

In recent years transcriptome sequencing has become an alternative to whole genome sequencing; as the latter still remains costly in many instances, transcriptome sequencing has proven to be comparatively more amenable to modest resources (Bouck and Vision 2007). Furthermore, in some cases it is currently computationally impossible to construct whole genome assemblies, as is the case in many higher plants. In the latter instance genome evolution is characterised by substantial duplication and transposition events which create a currently unbridgeable encumbrance for computational positioning of overlapping genomic contigs during sequence assembly (Kent and Haussler 2001; Hoskins *et al.* 2002). Using cDNA (subsequently EST) data highlights genomic regions responsible for the vast majority of phenotypic variation, while circumventing genome structural features such as dispersed repetitive elements (Rudd 2003). The annotation (identifying function) of EST sets broadens the avenue toward comparative genomics and extrapolating data from related and model species to the focal organism. Such annotations are generally conducted via similarity searches to non-redundant gene or protein databases (Ronning *et al.* 2003; Nagaraj *et al.* 2006), thus providing a means of evaluating the number of orthologous genes shared amongst species, as well as measuring the degree of gene sequence conservation or similarity (Rudd 2003; Bouck and Vision 2007).



Genomic sequence of the gene, contains exons, introns and regulatory elements (triangles).

Processed mRNA after transcription in cells, which will be extracted.

Reverse Transcription to create the cDNA strand from the mRNA.

Double stranded cDNA is produced by first digesting the mRNA using RNase H and then synthesising a new DNA molecule in its place, catalysed by DNA polymerase 1.

Double stranded cDNA is inserted into cloning vector for cDNA library construction.

Sequencing of cDNA, from either or both sides, using universal primers generates ESTs. This is then processed and stored in databases, eg. NCBI's dbEST.

Figure 4.1: Diagrammatic representation of the processes followed to generate EST sequences (Figure adapted from Bouck and Vision 2007).

Expressed sequence tags are not without limitations: The most obvious is that of gene representation. The genic diversity of any cDNA library depends on the specific mRNA pool present at a particular point in time, taking into account the physiological status of the organism. Therefore poorly expressed genes will have limited representation, while genes not expressed at all will be omitted from the collection; as such a typical library may only represent approximately 60% of the total gene content (Rudd 2003). A further concern is that of sequence quality. Expressed sequence tags are partial single read fragments of the original cDNA and generally little to no editing is done, subjecting the sequence to base-calling errors; as many as 3% of bases may be called incorrectly (Hillier *et al.* 1996; Nagaraj *et al.* 2006). Also when clustering ESTs to reconstruct the original mRNA/cDNA, the contamination of ESTs with vector or polylinker sequences, as well as inability to discriminate between alternative alleles and splisoforms from paralogues, makes such assemblies prone to errors (Wang *et al.* 2004). Lastly, ESTs give no information on gene order or position and limited information on introns and untranslated regions (UTRs); these must be inferred via predictive software or additional laboratory investigation (Bouck and Vision 2007). Such limitations must be kept in mind; nonetheless it does not severely impact the utility of ESTs as shown by the continuously expanding public database depositories for EST sequences (Rudd 2003; Bouck and Vision 2007).

The key question now, at least in terms of the South African abalone industry is: How best to incorporate such EST data into a practical breeding and management plan for abalone? The answer lies, at least in part, in ESTs as a source of molecular markers. Molecular breeding relies on the construction of dense genetic maps for QTL analysis and subsequent marker assisted selection; molecular markers are needed to construct these maps. As discussed in Chapter I, EST-derived markers are *de facto* type I markers, and with this comes obvious advantages: EST-derived markers have a higher probability of associating with quantitative traits, either as a causative variant or closely linked to a causative variant, because of its genomic proximity to a gene. Furthermore, EST-marker QTLs facilitate the identification of candidate genes associated with complex traits. Also, because EST sequences should be more conserved than anonymous DNA, it enables the use of related species' ESTs for marker development for a species that may be underrepresented in EST databases as is the case with *H. midae* (Decroocq *et al.* 2003; Ju *et al.* 2005; Pérez *et al.* 2005; Chen *et al.* 2007; Feng *et al.* 2009). Linkage mapping of EST markers furthermore resolves the question of gene order and position (e.g. Rohrer *et al.* 2002; Chagné *et al.* 2003; Hong 2008). The expanding number of ESTs in public databases creates a valuable resource for marker development and negates the necessity for cDNA library construction, thus reducing the cost and time of marker development via traditional means (Bouck and Vision 2007).

There are several ways in which ESTs can be employed for the development of various marker types (Gupta and Rustgi 2004). Here the focus will be on microsatellites and SNPs as these markers are the most widely applied currently. The detection of microsatellites are based on the direct analysis of sequence repeat motifs in EST singleton sequences or contigs (after clustering of ESTs) generally via a variety of computer programs that have been developed for this purpose (Theil *et al.* 2003; Fraser *et al.* 2004; Qureshi *et al.* 2004; Serapion *et al.* 2004; Kumpatla and Mukhopadhyay 2005; Lindqvist *et al.* 2006; Maneeruttanarungroj *et al.* 2006; Ramesh *et al.* 2007; Westgaard *et al.* 2007). Single nucleotide polymorphisms discovery from EST sets relies on the inherent redundancy of these collections, where sequence variants can be directly detected from EST alignment during cluster analysis (Picoult-Newberg *et al.* 1999; Barker *et al.* 2003; Batley *et al.* 2003; Kota *et al.* 2003).

Even in the absence of sequence redundancy or obvious repeat motifs, ESTs might still serve as a source for marker development via less conventional routes. Expressed sequence tags may be the only sequence information available for a particular species. If a homolog in a related species was shown to be associated to a trait of interest, this could allow for a targeted gene approach for marker detection via amplicon resequencing (Bouck and Vision 2007). Exonic sequences (thus also ESTs) may not demonstrate particularly high nucleotide diversity; however aligning EST contigs to genomic sequences (of a related species if necessary) predicts the locality of introns, often conserved between related species (Ku *et al.* 2000). Introns are more likely to accumulate sequence variation and therefore present a source of molecular markers. Thus, from the EST sequence, primers may be designed that will anneal to the exon, but span across the intron. This approach has been termed exon priming, intron crossing (EPIC) (Bouck and Vision 2007).

In this study the aim is to mine for both microsatellites and SNPs via direct sequence investigation and exploiting redundancy of EST collection in the NCBI public database. However, currently the South African endemic abalone, *Haliotis midae*, is relatively poorly represented in the EST database. Therefore a cross-species EST transferability approach will be adopted, using ESTs from a variety of Haliotids. A set of 'conserved orthologous markers' (Fulton *et al.* 2002) that could serve as 'comparative anchor tagged sequences' (CATS) (Lyons *et al.* 1997) for later comparative genomic studies will thus be generated. These molecular markers will therefore have an increased range of applicability; in more traditional uses of molecular markers for diversity estimates, parentage assignments and molecular breeding, but also as the initial steps toward comparative analysis of Haliotid genomes and synteny mapping.

2. Materials and Methods

2.1. EST Download, Processing and Candidate Marker Identification

All available ESTs for *Haliotis* species up until 1 December 2008 were downloaded in FASTA format from the NCBI (http://www.ncbi.nlm.nih.gov/) EST database, using an appropriate search string (e.g. *Haliotis* AND EST, *Vetigastropoda* AND EST, *Gastropoda* AND EST) and saved as text. Expressed sequence tag sequences were subdivided according to species (*H. discus*, *H. asinina*, *H. diversicolor supertexta*, *H. midae*) and saved in individual text files. These sequences were then subjected to the web-based programs TRIMEST (http://emboss.bioinformatics.nl/cgi-bin/emboss/trimest) to remove any poly-A/T tails (minimum length for recognition of poly-A/T tail was five nucleotide A/T repeat at 5'/3' end). In the case of a poly-T tail at the 3' end, the reverse compliment of the sequence was used for further analysis (option for reverse compliment selected). To remove possible contaminating vector sequences, the TRIMEST output sequences were screened using VecScreen (http://www.ncbi.nlm.nih.gov/) (preset/default parameters). Next a cluster analysis was performed to detect redundancy and construct consensus sequences (contigs) of the cDNA/mRNA via contig assembly, using CAP3 (Huang and Madan 1999). Overlap length was set to a minimum of 30 nucleotides, sharing 85% identity, with a similarity score of equal to or greater than 500. CAP3 output files were saved.

Expressed sequence tag sequence redundancy was used to investigate the presence of putative single nucleotide polymorphisms, by *in silico* identification of sequence variation using the web-based program SNPServer (Savage *et al.* 2005; hornbill.cspp.latrobe.edu.au/cgi-binpub/autosnip/index_autosnip.pl). The CAP3 ACE extension output file was imported into SNPServer and submitted. Result files were obtained and saved. As a secondary control, the CAP3 capout sequence alignment files were also manually inspected to detect sequence variation in clusters with a minimum of four aligned sequences of which two sequences demonstrated an alternative allele. Putative microsatellite/SSR markers were identified using FastPCR's (Kalendar 2003; http://www.biocenter.helsinki.fi/bi.bare-1_html/dawnload.htm) SSR loci search function by importing FASTA format EST sequence files of the contigs and singletons created by CAP3 into the program. This program identifies most repetitive DNA sequences, however for this study only perfect, compound and interrupted di- to pentanucleotide microsatellites with a clear tandem repetitive motif, were selected. Furthermore, to qualify as a putative microsatellite marker, dinucleotide repeats had to demonstrate at least six tandem motifs, tri- at least five, tetra- four and pentanucleotides at least three tandem motifs. Result files were saved as text in FASTA format.

2.2. Validation of Marker Transferability, Marker Characterisation and Putative Gene Function

Putative molecular markers identified *in silico* based on DNA sequences of related species, were tested for transferability to the *Haliotis midae* genome employing PCR using *Haliotis midae* genomic DNA [previously extracted using the standard CTAB extraction protocol (Saghai-Maroof *et al*. 1984), DNA was retrieved from the Laboratory stock, Molecular Aquatic Research Group]. Primers were designed (using Primer3, Rozen and Skaletsky 1998, www.genome.wi.mit.edu/genome_software /other/primer3.html) to anneal to the microsatellite or SNP flanking region. Optimum product size was set to 100-600 base pairs, GC content ranging between 40% and 60%, with an optimum of 50% and melting temperature was set to an optimum of 55ºC allowing a maximum of 2ºC difference between forward and reverse primers. After each successive round of PCR for each primer pair, amplification success was evaluated by agarose gel (2% w/v, 1X TBE; Appendix) electrophoresis, loading 3µl of PCR product mixed with 1µl of loading dye (6X Bromophenol Blue, Appendix) and running for 1 hour at 120 volts. The Promega 100bp ladder was run with all samples for preliminary fragment size scoring. Several rounds of optimisation was done using Go*Taq*® Flexi DNA Polymerase (Promega) and KAPA2G^{TM} Fast HotStart DNA polymerase (KAPA Biosystems); altering annealing temperatures appropriately: increasing incrementally if non-specific products persisted or decreasing incrementally if no product could be observed. All PCR reactions were done in a final volume of 10µl with final reagent concentrations as follows: for Go*Taq*® - 20ng gDNA, 1-2X Buffer, 1.5-4mM $MgCl_2$, 0.3mM dNTPs, 0.3µM of each primer and 0.5U *Taq*; for KAPA2G^{TM} - 20ng gDNA, 1X Buffer, 1.5-1.7mM $MgCl_2$, 02mM dNTPs, 0.2µM of each primer and 0.25U *Taq*. Buffer and magnesium chloride titrations were also done to increase primer annealing specificity were necessary. All PCR cycling reactions were performed on the Applied Biosystems 2720 Thermo Cycler (Table S3, S4; Appendix).

Where successful amplification occurred, PCR products were purified for sequencing using the SigmaSpin^{TM} Post-Reaction Cleanup Columns (Sigma) as per manufacturer’s specifications. This was followed by bi-directional sequencing via standard Sanger sequencing chemistry (BigDye® terminator V3.1 cycle sequencing kit, Applied Biosystems) and sent to the Stellenbosch University Central Analytical Facility (DNA sequencing unit) for capillary electrophoresis. Subsequently, sequences (reverse compliment where appropriate) were aligned with the original EST (contig or singleton) sequence by means of BioEdit 7.0.9.0 Sequence Alignment Editor computer software (Hall 1999), using the ClustalW (Thompson *et al*. 1994) function for multiple alignments, to evaluate the presence of false positives. False positives were discarded.

To validate the putative SNP markers, as identified *in silico*, an initial panel of 24 animals (of wild origin: eight each from Riet Point, Saldanha, and Witsand) were sequenced for each EST locus demonstrating *in silico* nucleotide variation. Visual identification of sequence variation in multiple alignments [BioEdit (Hall 1999), ClustalW (Thompson *et al*. 1994)] was done and confirmed by investigating individual chromatograms. A SNP was confirmed if clear double peaks, (beyond possible noise) could be identified in heterozygous individuals. As an additional control all trace files were also imported into *Novo*SNP and inspected at various quality cutoffs (10 to 25) (Weckx *et al.* 2005) for validation. For final confirmation of SNPs in alignments that showed polymorphism, an additional 24 individual animals (of wild origin: eight each from Riet Point, Saldanha, and Witsand) were sequenced and scored; a minor allele frequency of greater than 0.01 was prerequisite for final confirmation of a SNP marker (Fig. 4.2).

To test microsatellite polymorphism, an initial panel of eight individuals (from wild origin) were used to PCR amplify each putative microsatellite locus. These PCR products were then subjected to 12% polyacrylamide (49:1 – acrylamide:bis-acrylamide, see Appendix) gel electrophoresis, running for 2 hours at 150 volts to identify multiple alleles. A microsatellite was deemed polymorphic when two bands were distinguishable in a single individual, thus an identifiable heterozygotes, and/or there were clear size discrepancies between bands of different individuals. Fluorescently labelled primers were designed for microsatellite loci that demonstrated polymorphism. Original PCR conditions for all labelled loci were used with minor adjustments for annealing temperature where necessary. PCR products using labelled primers were sent to the Central Analytical Facility of Stellenbosch University for capillary electrophoresis (ABI Genetic Analyser). A panel of 32 animals (16 each from Witsand and Saldanha) was genotyped for marker characterisation, by scoring allele size using GeneMapper® version 4 software (Applied Biosystems).

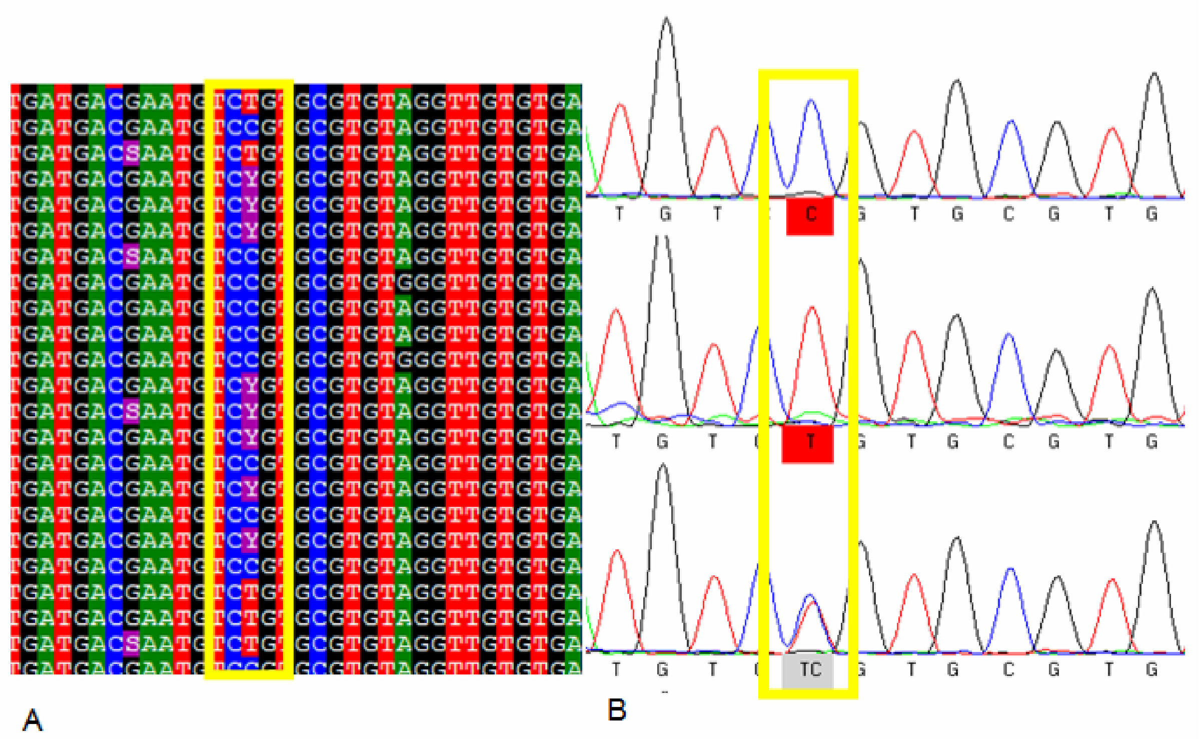


Figure 4.2: A) A multiple alignment depicting and C>T SNP (Yellow frame). B) The electropherograms of two homozygous individuals (CC and TT respectively) and a heterozygous individual, demonstrating a clear double peak (Yellow frame).

All markers were tested for Hardy-Weinberg equilibrium via an exact test (probability test); for loci with less than four alleles the complete enumeration method was used and for loci with more than four alleles the Markov chain method (20 batches, with 1000 dememorizations per batch) was employed. Furthermore, allele frequencies, observed heterozygosity, expected heterozygosity, linkage disequilibrium and F_{is}-statistics (as a measure of heterozygous deficiency or excess) were computed for all marker loci. For microsatellites, null allele frequencies were also estimated (Brookfield 1996). Genepop version 4 software (Rousset 2008) was used to perform the aforementioned analyses. The polymorphic information content (PIC) was calculated, for all markers, using the following formula:

$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_i^{\ 2}$, where *k* is equal to the total number of alleles detected and p_i is the allele frequency of the *i*-th allele.

An Ewens-Watterson test for neutrality, using the algorithm by Manly (1985) (1000 simulations) in POPGENE version 1.32 (Yeh 1999, www.ualberta.ca/~fyeh/pr01.htm) was also done. In order to confer putative gene function to these molecular markers, the bioinformatic protocol elaborated on in Chapter II was followed.

3. Results

In total 4762 ESTs were downloaded (1 December 2008) representing five Haliotid species/sub-species*: Haliotis asinina* (1760 ESTs), *H. discus discus* (1627 ESTs)*, H. d. hannai* (1293 ESTs), *H. midae* (70 ESTs) and *H. diversicolor supertexta* (12 ESTs). Nine hundred and ninety five species-specific contigs were assembled, with an overall average of 2.8 ESTs per contig and 1858 singleton ESTs failing to cluster. A total of 57 putative EST-SNPs were identified in 27 contigs. The majority of these were detected in the *H. d. hannai* clusters (52 SNPs in 23 contigs), probably because this species had the highest EST redundancy (3.7 ESTs per contig) (Table 4.1). Thirty four of these putative SNPs represented transitions and 23 represented transversions (transition to transversion ratio ~ 1.5 :1).

Table 4.1: A summary of the contig assembly statistics and number of putative microsatellite and SNP markers detected *in silico* for the five species of represented abalone in the NCBI EST database, as well as the total statistics over all species.

Species	Number of ESTs	Number of contigs	Number of singletons	Average number of ESTs per contig	Number of putative microsatellite loci detected	Number of putative SNP loci detected
H. asinina	1760	606	339	2.3	23	2 in 2 contigs
H. discus discus	1627	199	749	3.7	27	52 in 23 contigs
H. discus hannai	1293	185	699	3.25	26	None
H. diversicolor supertexta	12	0	12	N/A	None	None
H. midae	70	5	59	2.2	6	3 in 2

						contigs
Total	4762	995	1858	2.8	81	57 in 27 contigs

The simple sequence repeat loci search identified 82 putative microsatellite markers. Fifteen different repeat sequence motifs were identified, of which CA repeats were the most prominent overall (26.8%; range: 7.4% – 50% between individual species) (Fig. 4.3). Four length motifs were observed with dinucleotides most abundant (50% overall; range: 34.8% – 66.7%), followed by trinucleotides (31.7% overall, range: 16.7% – 47.8%), tetranucleotides (17.1% overall; range: 11.5% – 22.2%) and pentanucleotides (1.2% overall; range: 0% – 3.7%) (Fig. 4.4). Assuming that each contig and singleton represents a unique gene, on average over the represented Haliotid genomes (excluding *H. d. supertexta*), 2.8% of gene transcripts will contain a microsatellite, thus either a microsatellite within the coding regions or the 5'/3' UTR [*H. asinina* (2.4%), *H. d. discus* (2.8%), *H. d. hannai* (2.9%), *H. midae* (9.4%)].

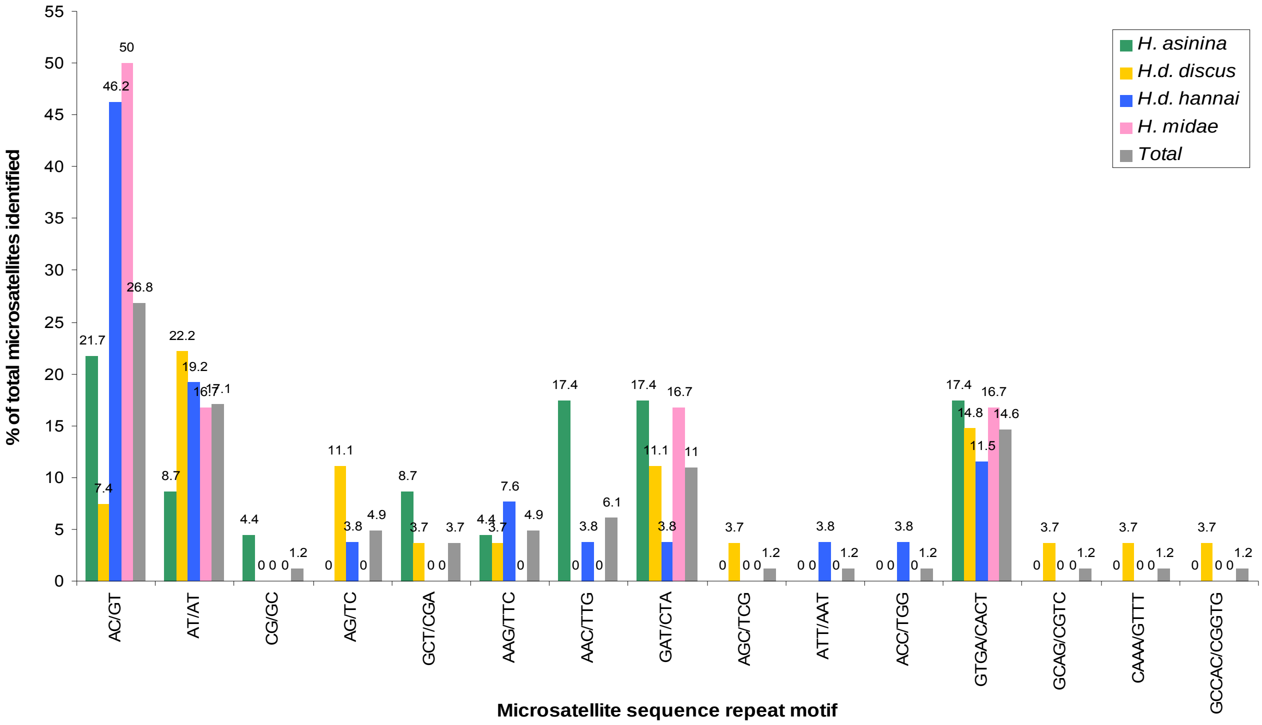


Figure 4.3: This graph depicts the microsatellite sequence repeat motifs as identified in the EST collections of represented Haliotid species, as well as an overall count across all ESTs downloaded.

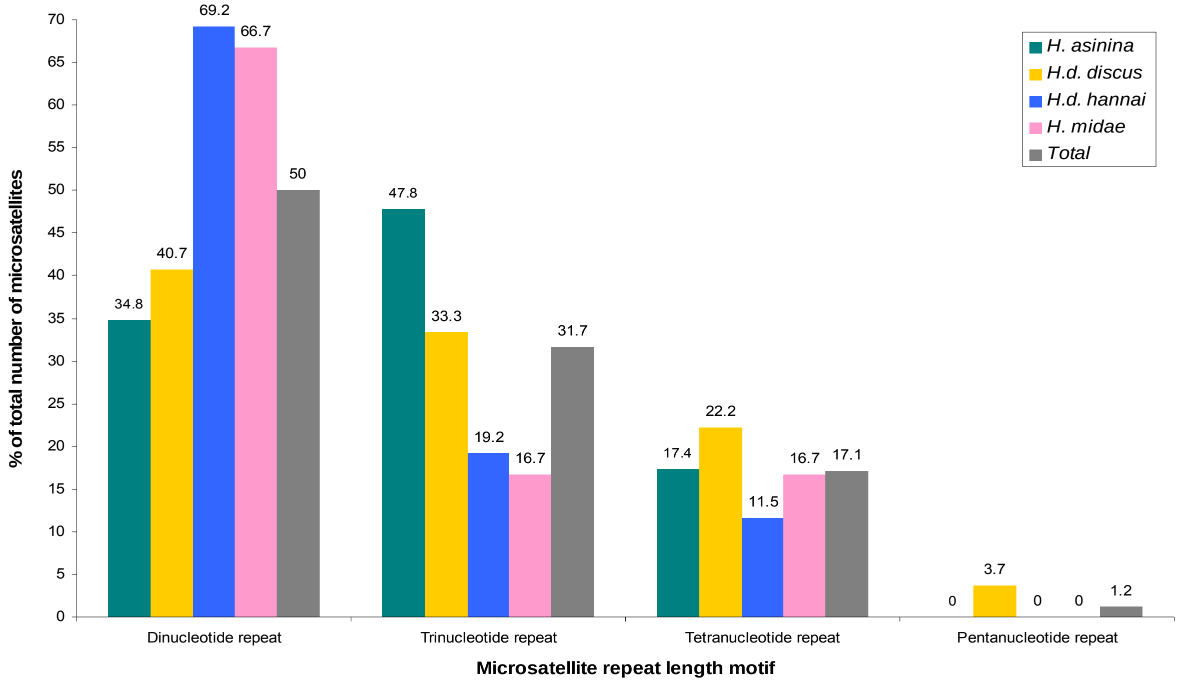


Figure 4.4: A graphical representation of the repeat length motifs as identified in the EST collection of all represented Haliotid species, as well as an overall count across all ESTs downloaded.

For putative molecular markers, 33 and 82 primer pairs that complied with the specifications could be designed for SNPs and microsatellites respectively. Twelve of the 33 (~36.4%) SNP-primer pairs optimised; nine pairs transferred from *H. d. discus* to *H. midae* (~27% transfer rate), one pair transferred from *H. asinina* to *H. midae* (50% transfer rate) and two of the three pairs from *H. midae* optimised. After sequencing and alignment with the original contigs, two *H. d. discus* primer pairs were discarded based on false positive amplification. Also, an additional two *H. d. discus*-originating fragments showed product sizes greater than what was expected, and where discarded because of low sequence quality. Of the remaining eight PCR fragments, amplicon resequencing in 24 individuals showed that only six fragments demonstrated sequence variation. Amplicon resequencing of an additional 24 individuals confirmed 16 new SNPs; of these 11 confirmed the SNPs originally identified during the *in silico* analysis while the remaining five constitutes *de novo* discoveries. Twelve SNPs were synonymous substitutions, three could not be assigned to a gene and one was non-synonymous (*Hdd.c148-885C>T* causes an amino acid substitution at position 272 in the peptide chain, Alanine to Valine) (Table 4.2). The transition to transversion ratio in this case was 1:1 and taking into account the total sequence length of all successful amplicons, the average SNP density was one SNP every 150bp. The average observed heterozygosity over all SNP loci was 0.78, the expected heterozygosity, 0.47 and the average PIC-value, 0.46. Thirteen SNPs demonstrated departure from Hardy-Weinberg expectations (p < 0.05). These SNPs also demonstrated severe heterozygous excess with an overall F_{is} equal to -0.6549.

Table 4.2: Characteristics of the sixteen EST-SNPs identified.

Locus Name	Minor Allele Frequency	PIC[1]	H_O[2]	H_E[3]	p-value for HWE[4]	F_{is}	Putative Function		
							Gene	E-value	NS/S[5]
Hdd.c2-427C>G	0.3	0.4	0.6	0.43	0.0041*	-0.378	Actin	0.0	S
Hdd.c2-487A>T	0.49	0.5	0.98	0.5	0.0*	-0.936	Actin	0.0	S
Hdd.c2-529C>T	0.48	0.5	0.88	0.5	0.0*	-0.750	Actin	0.0	S
Hdd.c2-553T>G	0.47	0.5	0.94	0.5	0.0*	-0.938	Actin	0.0	S
Hdd.c148-718C>G	0.2	0.32	0.44	0.37	0.2515	-0.046	Beta-Actin	0.0	S
Hdd.c148-742A>G	0.47	0.5	0.9	0.5	0.0*	-0.809	Beta-Actin	0.0	S
Hdd.c148-820T>C	0.49	05	0.94	0.5	0.0*	-0.938	Beta-Actin	0.0	S
Hdd.c148-885C>T	0.5	0.5	0.94	0.5	0.0*	-1.00	Beta-Actin	0.0	NS[5]
Hdd.c148-886A>T	0.5	0.5	1	0.5	0.0*	-1.00	Beta-Actin	0.0	S
Hdd.c106-556C>G	0.36	0.46	0.72	0.47	0.0002*	-0.538	Alpha-Tubulin	0.0	S
Hdd.c106-580C>A	0.44	0.48	0.88	0.5	0.0*	-0.676	Alpha-Tubulin	0.0	S
Hdd.c106-625C>T	0.33	0.43	0.65	0.44	0.0015*	-0.429	Alpha-Tubulin	0.0	S
Hdd.c106-688C>T	0.48	0.5	0.96	0.5	0.0*	-0.938	Alpha-Tubulin	0.0	S
Hm.c04-815C>T	0.46	0.5	0.75	0.5	0.0011*	-0.615	N/A	N/A	N/A
Hm.c05-	0.2	0.33	0.4	0.33	0.1704	-0.250	N/A	N/A	N/A

240T>G									
Ha.c500-207C>T	0.4	0.47	0.37	0.49	0.1287	0.255	N/A	N/A	N/A

1 – Polymorphic information content; 2 – Observed Heterozygosity; 3 – Expected Heterozygosity; 4 – Hardy-Weinberg Equilibrium, ("" – significant deviation, p < 0.05); 5 – Non-synonymous/Synonymous substitution; 6 – This variant substitutes an Alanine (C-allele) for an Valine (T-allele) in the peptide chain (for primer information, refer to Table S3 in the Appendix).*

For the microsatellite primer pairs, an overall success rate of 36% was achieved, with 29 of the 82 primer pairs optimised. Nine primer pairs transferred from *H. asinina* to *H. midae* (between species transfer rate: 39.1%), 10 primer pairs transferred from *H. d. discus* to *H. midae* (between species transfer rate: 37%), 4 primer pairs transferred from *H. d. hannai* to *H. midae* (between species transfer rate: 15.4%) and all six *H. midae* pairs produced a scorable band on agarose gel. Sequencing and alignment identified only one false positive, originating from a *H. d. discus* EST. Three primer pairs yielded longer fragments than anticipated. Of the 29 primer pairs optimised, only 21 loci demonstrated polymorphism on PAGE gel (six, seven, four and four from *H. asinina*, *H. d. discus*, *H. d. hannai* and *H. midae* respectively) and were subsequently labelled for further characterisation. After labelling, a further six loci (four from *H. d. discus* and two from *H. d. hannai*) were discarded, either because labelled primers failed to optimise or due to low peak quality, hindering accurate allele size scoring. Thus a final 15 microsatellites were characterised in 32 individuals. Both average observed and expected heterozygosity over all microsatellite loci was 0.62 and the average PIC-value was 0.61. Seven loci did not conform to Hardy-Weinberg expectations (p < 0.05) and overall F_{is} was -0.002. Putative function could be assigned to eight loci (Table 4.3).

The average observed and expected heterozygosity over all loci (SNPs and microsatellites) was 0.70 and 0.54, respectively and PIC over all loci was 0.53. Overall F_{is} suggests a heterozygous excess at -0.296 (Table 4.2 and 4.3). The exact test for linkage disequilibrium between pairs of loci showed that 18 pairs were significantly linked (p < 0.05) (Table 4.4). The test for neutrality demonstrated that eight SNP loci might be under selective pressure, while all microsatellite loci conferred neutrality (Table 4.5).

Table 4.3: Summary of the characterisation statistics for the 15 EST-microsatellite loci.

Locus Name (Acc#)[1]	Repeat Motif	k[2]	PIC[3]	Fr(A)[4] Range	Allele Size Range (bp)[5]	H_O[6]	H_E[7]	P-value for HWE[8]	F_{is}	Fr(A_{nul})[9].	Putative Function		
											Gene Name Acc#[1]	E-value	Location
HdSSRex495 (ex534495)	CA	17	0.84	0.02-0.35	180-216	0.7	0.86	0.005*	0.188	0.09	No Hit	N/A	N/A
HaSSRgd842 (gd241842)	$(CAG)_n$ (CAA)	6	0.73	0.05-0.38	118-133	0.8	0.74	0.66	0.081	0.01	B-Cell translocation gene *(Crassostrea gigas)* ACH92125	3.0E-26	Exonic
HaSSRdw239 (dw986239)	CAA	4	0.53	0.08-0.65	345-363	0.53	0.54	0.17	o.016	0.09	No Hit	N/A	N/A
HmSSRex489b (ex534489)	ACTC	4	0.6	0.07-0.57	95-107	0.53	0.61	0.27	0,125	0.16	Fertilisation protein (*Haliotis rufescens*) AF076827	8.0E-08	5’-UTR
HaSSRdw503 (dw986503)	ATG	2	0.19	0.11-0.89	253-256	0.53	0.61	0.31	0.213	0.9	No Hit	N/A	N/A
HmSSRex489a (ex534489)	CACT	5	0.66	0.06-0.52	230-246	0.71	0.67	0.16	-0.055	0.05	Fertilisation protein (*Haliotis rufescens*) AF076827	8.0E-08	5’-UTR
HdhSSRfe537 (FE041537)	GA	4	0.51	0.02-0.66	255-353	0.39	0.52	0.11	0.248	0.69	Putative mRNA (*Anoplopoma fimbria*) BT082209	9.0E-153	3’-UTR
HdSSRcx732 (cx726732)	AGC	2	0.49	0.44-0.56	225-234	0.88	0.5	0.00*	-0.771	0.00	No Hit	N/A	N/A

HmSSRex446a (ex534446)	GTGA	3	0.66	0.29-0.37	212-236	0.87	0.67	0.008*	-0.298	0.25	Cellulase gene (*Haliotis d. hannai*) AB125892	3.0E-39	5'-UTR
HaSSRc.571a (GU263799)	ATG	6	0.6	0.03-0.59	101-128	0.75	0.61	0.54	-0.237	0.00	No Hit	N/A	N/A
HmSSRex446b (ex534446)	ACTC	12	0.87	0.02-0.22	82-138	0.91	0.88	0.006*	-0.029	0.06	No Hit	N/A	N/A
HdhSSRc.60b (GU263800)	ACTC	16	0.9	0.02-0.19	80-120	0.66	0.91	0.00*	0.287	0.14	G-alpha signal transducing protein (*Haliotis rufescens*) AF070959	6.0E-86	5'-UTR[10]
HaSSRgd475c (gd272475)	GAA	2	0.3	0.19-0.81	97-100	0.38	0.31	0.57	-0.216	0.79	Supt5h protein (*Monodelphis domestica*) XM_001363183	1.0E-05	Exonic
HaSSRdy903 (dy402903)	TG	6	0.55	0.02-0.65	846-872	0.67	0.72	0.003*	0.404	0.70	Chaperonin (cpn10) (*Monodelphis domestica*) XP_001379358	9.0E-11	5'-UTR
HdSSRcx009 (CX726009)	GAA	5	0.7	0.05-0.39	108-144	0.68	0.72	0.009*	0.054	0.1	No Hit	N/A	N/A

1 – Genbank Accession number; 2 – Number of Alleles; 3 – Polymorphic information content; 4 – Allele frequency; 5 - base pairs; 6/7 – Observed and Expected Heterozygosity respectively; 8 – Hardy-Weinberg Equilibrium ("" – significant deviation, p < 0.05); 9 – Null allele frequency; 10 – This locus seems to locate in the intron of the original sequence in* Haliotis rufescens, *however because EST sequences do not contain introns it is postulated that this microsatellite is probably positioned in the 5'-UTR in* Haliotis asinina *(where the EST had its origin) (for primer information, refer to Table S4 in the Appendix)*.

Table 4.4: This table shows the calculated p-values [exact test in Genepop version 4 (Rousset 2008)] for significant linkage disequilibrium between pairs of loci.

	Hdd.c2-529C>T	***Hdd.c2-553T>G***	***Hdd.c148-742A>G***	***Hdd.c148-820T>C***	***Hm.c05-240T>G***	***HaSSRdw503***	***HmSSRex489b***	***HdhSSRfe537***	***HmSSRex446a***	***HaSSRc.571a***	***HdSSRcx009***
Hdd.c2-487A>T	0.031	0.032	-	-	-	-	-	-	-	-	-
Hdd.c2-529C>T	-	0.031	-	-	-	-	-	-	0.013	-	-
Hdd.c148-718C>G	-	-	0.016	0.031	0.005	-	-	-	-	-	-
Hdd.c148-742A>G	-	-	-	0.031	-	-	-	-	-	-	-
Hdd.c106-556C>G	-	-	-	-	-	0.018	-	-	-	-	-
HmSSRex489a	-	-	-	-	-	-	0.00	-	0.036	-	-
Ha.c500-207C>T	-	-	-	-	-	-	-	0.018	-	-	-
HdSSRcx732	-	-	-	-	-	-	-	-	0.004	-	-
HaSSRgd842	-	-	-	-	-	-	-	-		0.048	-
HaSSRgd475c	-	-	-	-	0.001	-	-	-	0.003	0.013	0.020

Table 4.5: Ewens-Watterson test for neutrality, showing the observed F (as a measure of homozygosity) and the 95% confidence levels for expected F. Loci marked with "*" demonstrate an observed F falling outside the confidence levels and thus deviation from neutrality.

Locus	Observed F	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Hdd.c2-427C>G	0.5957	0.5020	0.9692
Hdd.c2-487A>T	0.5005*	0.5020	0.9692
Hdd.c2-529C>T	0.5020	0.5020	0.9692
Hdd.c2-553T>G	0.5005*	0.5020	0.9692
Hdd.c148-718C>G	0.6763	0.5020	0.9692
Hdd.c148-742A>G	0.5005*	0.5020	0.9692
Hdd.c148-820T>C	0.5005*	0.5020	0.9692
Hdd.c148-885C>T	0.5000*	0.5020	0.9692
Hdd.c148-886A>T	0.5000*	0.5020	0.9692
Hdd.c106-556C>G	0.5421	0.5005	0.9683
Hdd.c106-580C>A	0.5176	0.5020	0.9692
Hdd.c106-625C>T	0.5749	0.5005	0.9683
Hdd.c106-688C>T	0.5005*	0.5020	0.9692
Hm.c04-815C>T	0.5000*	0.5044	0.9692
Hm.c05-240T>G	0.6686	0.5021	0.9683
Ha.c500-207C>T	0.5333	0.5021	0.9683
HdSSRex495	0.1556	0.0844	0.2056
HaSSRgd842	0.2711	0.2172	0.7022
HaSSRdw239	0.4672	0.3006	0.8728
HmSSRex489b	0.4017	0.3100	0.8728
HaSSRdw503	0.8052	0.5020	0.9692
HmSSRex489a	0.3377	0.2607	0.7903
HdhSSRfe537	0.4892	0.2902	0.8642
HdSSRcx732	0.5078	05020	0.9692
HmSSRex446a	0.3366	0.3699	0.9370
HaSSRc.571a	0.4009	0.2227	0.7188
HmSSRex446b	0.1328	0.1240	0.3457
HdhSSRc.60b	0.0977	0.0908	0.2446
HaSSRgd475c	0.6953	0.5020	0.9692
HaSSRdy903	0.4539	0.2189	0.7022
HdSSRcx009	0.2960	0.2575	0.7919

4. Discussion

In this chapter the aim was to investigate the feasibility of using public EST collections for type I molecular marker development in the commercially important South African abalone *Haliotis midae*. Haliotid species in general are scantily represented in public EST databases in comparison to traditional livestock (e.g. cattle: 1577950 ESTs, fowl: 600075 ESTs) and even in comparison to other aquaculture species, e.g. salmon (*Salmo salar*): 143714 ESTs, oyster (*Crassostrea virginica*): 9997 ESTs and tiger shrimp (*Penaeus monodon*): 7359 ESTs. However, it must be noted that since the commencement of this study (January 2008) the total EST collection for Haliotids more than quadrupled (21079 ESTs) with substantial additions to the *H. discus* species and *H. asinina* (NCBI's ESTdb, 05 November 2009); demonstrating a trend toward transcriptomics*.* Furthermore, because *Haliotis midae* is even more severely underrepresented than other Haliotids, a cross-species transferability approach was followed. The successful amplification of primers developed from related species to target species shows varying results depending on the taxa under investigation. Transfer rate is a function of phylogenetic distance with a negative correlation between the two variables (Estoup *et al.* 1995; Theil *et al.* 2003; Wang *et al.* 2007). Locus-specific dynamics is also suggested to have an influence, with loci in high mutagenic regions in a genome being less prone to cross-species transfer (Yue *et al.* 2003). It is however, commonly predicted that primers developed from exonic regions will transfer with greater ease, because of increased sequence conservation. Aitken *et al.* (2004) and Lyons *et al.* (1997) reported similar success rates for exon annealing primers, for SNP detection, amongst diverse mammalian taxa (marsupials to primates) ranging form 24% - 65% and 35% - 52% respectively, but Primmer *et al.* (2002) found only 20% amplification success between two related avian species (Flycatchers, *Ficedula* sp.). In terms of EST-microsatellite transfer between species, animals demonstrate significantly lower success rates in comparison to plants. Yue *et al.* (2004) and Wang *et al.* (2007) reported 35.7% and 62% transfer between carp species, respectively and Pérez *et al.* (2005) achieved 21% and 69% success between three shrimp species. Ju *et al.* (2005) demonstrated that as little as

0.6% of EST-microsatellites share significant flanking region conservation between four fish species to permit cross-genus transfer. Wang *et al.* (2008) reported the highest transfer success rate for animals in the oyster, ranging from 25% to almost 100%. On the contrary, plants demonstrate transfer rates of 40% - 100% within genera (Decroocq *et al.* 2003; Chagné *et al.* 2004; Cibrián-Jaramillo *et al.* 2008) and 33% - 77% between genera (Aggarwal *et al.* 2007; Feng 2009). It must also be noted that when primers are designed from EST sequences, the probability remains that these primers are situated over intron-exon-boundaries or that the product will include an intronic sequence too large to amplify under regular PCR conditions. This could lead to non-amplification or larger product sizes than expected (Pérez *et al.* 2005).

For this investigation, the underrepresentation of Haliotid ESTs is the most probable cause for low ESTs redundancy (2.8 ESTs per contig, Table 4.1). This was a major impediment for SNP discovery, which relied on sufficient sequence redundancy to identify putative SNPs. This explains why most of the putative SNPs were identified in *H. d. discus* that demonstrated the highest average ESTs per contig (Table 4.1). Irrespective of the low number of putative SNPs identified (57) in comparison to other similar studies employing *in silico* mining [1210 putative SNPs in the chicken (Kim *et al.* 2003), 101 putative SNPs in silk worm (Cheng *et al.* 2004), 232 putative SNPs in sea bass (Souche *et al.* 2007)], the number of SNP-markers validated per contig remains comparable; illustrating that should the number of ESTs included in future analyses increase, so too would the number of markers identified. This is supported by a recent publication by Qi *et al.* (2009) that identified 302 putative SNPs and validated 28 markers for *H. d. hannai*.

During the current study, 16 SNPs were validated in six contigs, providing a SNP to contig/gene ratio of 2.6:1. This is more than what was found for the chicken with 1.7:1 (Kim *et al.* 2003) and the sea bass with 1.6:1 (Souche *et al.* 2007), yet is marginally less than what was reported for *H. d. hannai* with 3.1:1 (Qi *et al.* 2009). The high SNP to contig ratio in abalone (Qi *et al*. 2009; present study) suggests that these molluscs have a high nucleotide diversity. This is further supported by the SNP frequency reported here of one SNP every ~150bp. Previous studies on *H. midae* by Bester *et al.* (2008) and Rhode *et al.* (2008) reported SNP frequencies of one SNP every 185 and

113bp respectively. Qi *et al.* (2009) reported one SNP every 100bp for *H. d. hannai*. It is well known that SNP frequencies fluctuate greatly between species and genomic regions within species, with frequencies equal to one SNP every 30 to 500bp in a variety of animals from molluscs to mammals (Brouillette *et al*. 2000; Vignal *et al*. 2002; Morin *et al*. 2004; Pariset *et al*. 2006; Quilang *et al*. 2007). However, high SNP frequencies are generally associated with neutral genomic regions and figures presented in this study seem to correlate with avian and canine intronic SNP densities of one SNP every 80 (Brumfield *et al*. 2003 and references therein) and 268bp (Brouillette *et al*. 2000) respectively. Taking into consideration that the SNPs developed in this study are derived from expressed sequences, the high frequency therefore seem unrealistic. This could be partially explained by the fact that the majority of SNPs reported here are non-synonymous substitutions and, therefore selectively neutral. Furthermore, molluscs first appeared ~530 million years ago (MYA) (during the Cambrian, Lydeard *et al.* 2000); in comparison the first mammalian ancestor only emerged during the Carboniferous (~310 MYA, Kumar and Hedges 1998). Given the evolutionary age of molluscs, they had an extended period to accumulate sequence variation, compared to mammals that demonstrate average, genome-wide SNP frequencies of approximately one SNP every 1000bp pairs (Brouillette *et al*. 2000; Brumfield *et al*. 2003; Morin *et al*. 2004; Seddon *et al*. 2005; Rengmark *et al*. 2006). High mutation accumulation may also be attributed to the reproductive strategy of abalone. As these animals are broadcast spawners their effective population size is theoretically much greater; resultantly the probability of losing new alleles through random drift, is less (Primmer *et al.* 2002). Also high fecundity leads to an increased mutation rate, because of excessive meiotic divisions (Hedgecock *et al.* 2004; Li *et al.* 2002). This is supported by a similar EST-SNP frequency in another broadcast spawning mollusc, the eastern oyster (*Crassostrea virginica*) with one SNP every 200bp (Quilang *et al*. 2007).

Direct EST amplicon resequencing studies yielded similar results to the *in silico* data obtained in this current study in terms of number of SNPs validated per contig in *Chenopoduim quinoa*, with 2.55 SNPs per contig. However, it must be noted that this food crop has a lower SNP frequency of one SNP every 462bp (Coles *et al.* 2005). Such a direct EST amplicon resequencing strategy in *H. midae* yielded a slightly higher average SNP to contig ratio of 3.3:1 (Bester *et al.* 2008). It is therefore suggested that because *H. midae* has such a high SNP frequency rate, amplicon resequencing will allow for greater detection of SNP-markers. A similar argument was posed by Qi *et al.* (2009) for *H. d. hannai* and is further supported by the discovery of five additional SNPs not initially found by the *in silico* search. The success of any *in silico* mining depends largely on the origin of the ESTs. Irrespective of the number of contigs constructed, redundant ESTs within a cluster must contain sufficient sequence variation to ensure detection of candidate SNPs. The only means to secure such redundant sequence variation is to construct cDNA libraries using multiple unrelated individuals. This approach was followed in the study by Souche *et al.* (2007) that derived EST data from five cDNA libraries constructed using 350 individuals. Subsequently they identified 112 candidate SNPs of which 69 were confirmed for the European sea bass. When obtaining EST data from public databases, as is the case here, it is however often difficult to ascertain the origin of particular sequences.

The transition to transversion ratio of the putative SNPs over all Haliotid genomes investigated here (~1.5:1) conforms to the general expectation of a higher transition rate in animals. Mammals demonstrate ratios between 1.4:1 and 1.7:1 (Collins *et al*. 1994; Picoult-Newberg *et al*. 1999), while birds show higher ratios from 2.3:1 to 4.0:1 (Smith *et al*. 2001; Vignal *et al*. 2002). Invertebrate estimates based on the silkworm found a transition to transversion ratio of 1.66:1 (Cheng *et al.* 2004). The persistence of this trend in molluscs is supported by the findings in oysters (*Crassostrea virginica*) that demonstrated a transition to transversion ratio of 1.3:1 (Quilang *et al*. 2007), in *H. d. hannai* this ratio is reported to be 2.2:1 (Qi *et al.* 2009) and in the scallop (*Patinopecten caurinus*), 2.4:1 (Elfstrom *et al.* 2005). However, looking at the confirmed transition to transversion ratio of *H. midae* in this study (1:1); this is seemingly less than the expected. This lower ratio is the same as previously reported for *H. midae* (Rhode *et al.* 2008) whereas Bester *et al.* (2008) even reported a transversion excess (1:1.5). The transition mutational mechanism is proposed to be correlated to the high mutability of CpG-like repeat units, where the spontaneous deamination rate of 5-methyl cytosine to thymidine is high (Brookes 1999; Vignal *et al.* 2002). Considering that *H. midae* demonstrates a low CpG-like repeat motif frequency (Chapter II, Fig. 2.1; Fig. 4.3) it is comprehendible that this could explain the lower transition rate. Nonetheless, definitive conclusions cannot be made until a thorough genome-wide assessment has been done.

Thirteen of the 16 SNPs deviated from Hardy-Weinberg expectations (Table 4.2). This is explained by the extreme heterozygous excess demonstrated by these SNPs as confirmed by the negative F_{is} and high observed heterozygosity values (Table 4.2). The SNPs derived from the contigs *Hddc.2* and *Hddc.148,* respectively, are most likely to be in linkage disequilibrium (Table 4.4). Furthermore, the aforementioned contigs gave significant hits to actin family members and *Hddc.106* to a tubulin family member. Both these protein families play important roles in cytoskeletal dynamics and show a molecular signature of purifying selection, however, also demonstrates an excess in amino acid polymorphism conserved between related species. This pattern of genetic diversity is indicative of balancing selection or heterozygous advantage (Wu *et al.* 2003; Bustamante *et al.* 2005) and is supported by both the negative F_{is}-values (Table 4.2) and the fact that the Ewens-Watterson test for neutrality shows SNPs in all three contigs that fall out of the range of the lower confidence boundary (Table 4.5). The majority of the SNPs reported for these contigs are synonymous substitutions and the one non-synonymous substitution (*Hdd.c148-885C>T*) substitutes an Alanine for Valine in the peptide chain and will probably not alter protein function drastically (as both are non-polar amino acids). Irrespectively, these SNPs are most likely in linkage disequilibrium with undetected causal variants that are responsible for the observed variation. It cannot be excluded that at least part of the observed variation is due to simultaneous amplification of paralogs and not true sequence variation within a single gene, as both actins and tubulins are members of multi-gene families (Hayes *et al.* 2007).

The three SNP loci (Table 4.2) that adhered to Hardy-Weinberg equilibrium frequencies demonstrated, for the greater part, heterozygosities in concordance with previous finding for *H. midae* (Bester *et al.* 2008; Rhode *et al.* 2008) and *H. d. hannai* (Qi *et al.* 2009). These heterozygosities also correlated well with those in other animals: ovines (Cappuccio *et al*. 2006;

Pariset *et al*. 2006), canines (Brouillette and Venta 2002; Seddon *et al*. 2005) and scallop (*Patinopectum caurinus*; Elfstrom *et al*. 2005). Coupled with moderately high PIC-values for SNPs (maximum PIC-values for bi-allelic markers are 0.5), it is feasible to include these markers in diversity studies for population management, breeding and conservational applications (Morin *et al*. 2004).

The transfer rate to *H. midae* from other Haliotids (15.4% - 50%) conforms to general expectations, as has been found for other animals. It does, however, tend toward the lower end of success rates (Yue *et al.* 2004; Pérez *et al.* 2005; Wang *et al.* 2007). This can be attributed to high divergent nucleotide diversity between species, exacerbated by the evolutionary age and reproductive strategy (as discussed earlier) of Haliotids that favour the accumulation of mutations and subsequently adversely affects primer binding sites. The transfer rates also show a general trend to correlate with phylogenetic distance, with *H. asinina* demonstrating higher transfer success than the *H. discus* sp. (Estes *et al.* 2005; Streit *et al.* 2005). Also there seem to be a species-specific factor in regards to SNP versus microsatellites loci. *Haliotis d. discus* EST-microsatellites show a 37% transfer rate, while EST-SNP demonstrates only 27%; *H. asinina* demonstrates the inverse relation, 39% and 50% for microsatellites and SNPs respectively. However, it is stressed, because the EST-SNP rate is based on only two primer pairs for *H. asinina*, an accurate conclusion can not be drawn and thus warrants further investigation. However, microsatellite flanking sequences have been shown to be hyper-variable [Blankenship *et al*. 2002; Dettman and Taylor 2004; Ablett *et al*. 2006 (and references therein); Rengmark *et al*. 2006; Rhode *et al.* 2008] and therefore a lower EST-microsatellite transfer rate is expected. The lower SNP-primer transfer rate for *H. d. discus* might therefore be artefactual.

There are currently no studies available that evaluated the direct marker transferability of anonymous/type II markers between *H. midae* and *H. asinina.* However, Evans *et al.* (2001) tested the transfer of type II microsatellite markers from *H. rubra* to *H. asinina.* They reported a transfer rate of 22% to *H. asinina*. As *H. rubra* and *H. midae* are sister taxa (Bester-Van der Merwe 2009), it can be deduced that *H. midae* will show a similar anonymous marker transfer rate from *H. asinina* as *H. rubra* and this is less than the 39% transfer rate reported here for EST-microsatellites. Sekino and Hara (2007) tested type II marker transfer from *H. discus* to *H. midae* and reported ~16.7% transfer success, again this is lower than the combined transfer rate reported for *H. discus* species in this study. The higher EST-microsatellite transfer rates are expected as these sequences are thought to demonstrate higher sequences conservation. A higher EST-marker transfer rate in comparison to genomic markers was also reported by various other studies (e.g. Pashley *et al.* 2006; Wang *et al.* 2007; Feng *et al.* 2009).

Within and between taxonomic groupings the number of ESTs containing microsatellites show considerable diversity. Various studies reported 1.5%-11.5% for fish (Edwards *et al.* 1998; Serapion *et al.* 2004; Ju *et al.* 2005; Siemon *et al.* 2005; Wang *et al.* 2007; Bouza *et al.* 2008; Kim *et al.* 2009). Shrimp seems to display similar proportions, with 2.2% (Wang *et al.* 2005), 7.2% (Pérez *et al.* 2005) and 13.7% (Maneeruttanarungroj *et al.* 2006) reported in various studies. Molluscs, on the other hand, tend toward the lower spectrum of microsatellite abundance within ESTs, with estimates of 3.9% for scallop (Zhan *et al.* 2005), 0.6% for the eastern oyster (Quilang *et al.* 2007) and 4.5% for the Pacific oyster (Wang *et al.* 2008). The percentage reported in this study of a 2.8% EST microsatellite abundance across Haliotid genomes is consistent with the general trend in molluscs. A relatively high value of 9.4% is reported for *H. midae*, however, because this abalone is severely underrepresented in comparison to the other species (only 70 ESTs), this value is in most likelihood skewed. Zhan *et al.* (2008) reported that 3.4% of *H. discus* ESTs contained microsatellites. In the current study it was reported to be 2.8% and 2.9% for *H. d. discus* and *H. d. hannai* respectively, but it must be noted that substantial manual curation was done that created more stringent screening criteria so as to optimise cross-species transfer.

Across all Haliotid ESTs, dinucleotides (50%), followed by tri- (31.7%) and tetranucleotides (17.1%) were most abundant, with the exception of *H. asinina* where trinucleotides were most abundant (Fig. 4.2). Amongst the dinucleotides, CA repeats (26.8%), followed by AT (17.1%) and CT (4.9%) repeats were the highest represented; the exception was *H. d. discus* that favoured AT above CA. This result concurs with the majority of findings for other animal EST-microsatellites, where dinucleotides and CA repeats, in particular, are the dominant marker (Serapion *et al.* 2004; Ju *et al.* 2005; Pérez *et al.* 2005; Quilang *et al.* 2007; Wang *et al.* 2007). However, Ju *et al.* (2005) did note that a teleost fish (*Fundulus* sp.) also, atypically of fish in general, favoured AT repeats. Furthermore Serapion *et al.* (2004) found that the channel catfish had as its second most frequent dinucleotide, GA repeats. Various plant species also display an affinity for GA repeats (Kantety *et al.* 2002; Kumpatla and Mukhopadhyay 2005; Aggarwal *et al*. 2007; Feng *et al.* 2009). In contrast some shrimp species appear to favour trinucleotides in expressed sequences (Pérez *et al.* 2005; Maneeruttanarungroj *et al.* 2006) consistent with a variety of plant species (Kantety *et al.* 2002). Contrary to the result of this study, Zhan *et al.* (2008) also reported a trinucleotide dominance for *H. discus.* In agreement with previous findings for invertebrates (Toth *et al.* 2000), fish (Serapion *et al.* 2004; Ju *et al.* 2005) and some plants (Aggarwal *et al.* 2007), CpG-like repeat motifs (e.g. CG, CCG) are in low frequency in genic regions of Haliotids, probably because of the high mutability of CpG sequences affecting functionality (Brookes 1999; Vignal *et al.* 2002). Furthermore, GAGT motifs enjoy a relatively high frequency across all Haliotis ESTs investigated here, suggesting a conserved function for these repeats in Haliotids (also refer to Chapter II).

Admittedly, it is often difficult to directly compare EST-microsatellite distribution among studies, because microsatellite detection relies heavily on the initial search parameters and number and constitution of ESTs in the collection (Zhan *et al.* 2005; Aggarwal *et al.* 2007). In fact, due to the very nature of trinucleotides not disrupting open reading frames, they are the most likely microsatellites to locate in exons (Toth *et al.* 2000). Thus taking into account that ESTs are short reads often from the 5' or 3' end of cDNA, it is comprehendible that 5'/3'-UTR are overrepresented in EST collections and this will provide a bias toward the detection of di- and tetranucleotides.

The number of EST-microsatellites demonstrating polymorphism, as well as the inherent level of polymorphism of individual loci, are expected to be lower than that of genomic microsatellites due to genic constraints. It was established for plants that approximately 46% - 47% of EST-microsatellites were polymorphic (Thiel *et al.* 2003; Feng *et al.* 2009). In fish, it was demonstrated that 42% of common carp focal species EST-microsatellites were polymorphic (Wang *et al.* 2007), 38.6% in the Atlantic salmon (Siemon *et al.* 2005) and as much as 81% of flounder EST-microsatellites (Kim *et al.* 2009). Molluscs appear to display lower proportions with estimates for the bay scallop at 13.8% (Zhan *et al*. 2005) and for *H. discus*, 20% (Zhan *et al.* 2008). The estimate for the number of polymorphic focal EST-microsatellites for *H. midae* is ~67%; however it is reiterated that only 6 EST-microsatellites were identified for this abalone and because of this limited sequence representation, accurate conclusions can not be drawn. The number of polymorphic EST-microsatellites transferred from related species also demonstrates a tendency to decrease with an increase in phylogenetic distance (Wang *et al.* 2007; Feng *et al.* 2009). The results presented here are in accord with this trend: 26% of *H. asinina* transferred EST-microsatellites showed polymorphism in *H. midae* and on average between the two *H. discus* sp., 21% of the markers were polymorphic.

Seven microsatellite loci departed from Hardy-Weinberg expectations (Table 4.3), however all conformed to neutrality (Table 4.4). In the case of four of these loci (*HdSSRex495, HdhSSRc.60b, HaSSRdy903, HdSSRcx009*), a heterozygous deficit was observed (positive F_{is} values) which is most likely because of the persistence of null alleles (Table 4.3). In regards to the remaining three loci (*HdSScx732, HmSSRex446a, HmSSRex446b*), they are likely in linkage disequilibrium with the SNP locus *Hdd.c2-529C>T*; this would explain the heterozygous excess (negative F_{is}) (Table 4.2, 4.3, 4.4). Average observed heterozygosity (0.62, range: 0.38-0.91), PIC-value (0.61, range: 0.19-0.9) and number of alleles per locus (6.3, range: 2-17) agrees with moderate EST-microsatellite polymorphism across various plant and animal taxa (Theil *et al.* 2003; Zhan *et al.* 2005; Aggarwal *et al.* 2007; Varsheny *et al.* 2007; Wang *et al.* 2007; Bouza *et al.* 2008; Cibrián-Jaramillo *et al.* 2008; Wang *et al.* 2008; Feng *et al.* 2009; Kim *et al.* 2009; Qi *et al.* 2009; Wang *et al.* 2009). Interestingly, Yue *et al.* (2004) reported that EST-microsatellites in carp showed higher polymorphism than that of genomic microsatellites and Coulibaly *et al.* (2005) reported no significant difference between the two marker types for trout. In contrast to this and in accordance to the norm (Chabane *et al.* 2005), *H. midae* transfer EST-microsatellites demonstrates lower polymorphism than that of genomic microsatellites isolated for the species. Bester *et al.* (2004) and Slabbert *et al.* (2008) reported average number of alleles per genomic microsatellite to be 12.4 and 13.1 respectively; double the number of alleles reported for EST-markers here. The lower polymorphism of these loci may indicate that these microsatellites might possess functional roles within the respective genes and therefore their allelic richness remains limited. Furthermore, the fact that all the microsatellite loci demonstrated neutrality, would suggest purifying selection; as such the observed variation remains neutral for all practical purposes (Gebhardt *et al.* 1999, 2000; Li Y-C *et al.* 2004).

Comparing the *in silico* method for microsatellite isolation as described to that of conventional cloning and hybridisation strategies generally followed, it would seem as if the *in silico* method demonstrates marginal superiority. Eighty one EST containing microsatellites were identified and 15 proved to polymorphic in *H. midae*, thus demonstrating a success rate of 18.5%. On the contrary Bester *et al.* (2004) identified 113 repeat containing clones, but could only isolate 11 polymorphic microsatellites (success rate: 9.8%) and Slabbert *et al.* (2008) isolated 63 polymorphic microsatellites from 462 repeat containing clones (success rate: 13.6). It is expected that as EST collections grow in public databases, especially for *H. midae* and more related species, such as *H. rubra*, these databases will become an even more powerful resource for type I marker development.

5. Conclusions

Molecular markers are vital for genetic management of any population. Their utility have been shown in various applications and in recent years there has been focus on the development of type I markers that show association to genic sequences. Concurrently numerous transcriptome projects have been initiated to provide insights into gene expression; leading to the accumulation of ESTs in public database. These in turn provide a substantial resource for type I molecular marker development. During this study it was demonstrated how microsatellites and SNPs could be developed from public EST collections, with little effort, for a species with minimal sequence representation, such as *Haliotis midae*. Pashley *et al.* (2006) argues that if only 2%-5% of ESTs contain microsatellites, a sufficient number of markers could be developed for diversity studies. Furthermore, SNPs are expected to be the marker of choice for future QTL analysis because they allow for the construction of dense marker maps and are probably responsible for the vast majority of phenotypic variation.

Sixteen EST-SNPs and 15 EST-microsatellites have been isolated from public EST collections of various Haliotids and transferred to the South African endemic during the course of this study. This approach apparently demonstrates a higher success rate than traditional cloning and hybridisation techniques for microsatellite isolation. However the number of SNPs discovered may be improved by means of amplicon resequencing due to the high nucleotide diversity demonstrated by *H. midae.* The data presented here conforms to the general trend for EST-markers in most other animal species. Furthermore a possible functional role for microsatellites across *Haliotis* is reaffirmed, especially in regards to the tetranucleotide repeat, GAGT.

Even though type I molecular markers demonstrate lower levels of polymorphism, they directly sample functional regions of the gene and as demonstrated here serve as a functional link between the genomes of related species. It is unlikely that an entire linkage map could be constructed only of type I markers, as marker density would be too low. Nonetheless, this strategy for marker development remains a dynamic and continual source of supplementary markers to saturate linkage maps (Rohrer *et al.* 2002; Maneeruttanarungroj *et al.* 2006). Furthermore, these markers may aid in the management and conservation of the genetic resources (Varshney *et al.* 2007). Their genic association makes it easier to detect possible divergent selection and population stratification (Vasemägi *et al.* 2005), and their lower allelic content may provide a more sensitive indicator of bottlenecks, than estimations derived from genomic microsatellites that seems to be impervious to the loss of rare alleles (Mgaya *et al*. 1995; Evans *et al*. 2004; Li Q *et al*. 2004)

References

1. Ablett G, Hill H, Henry RJ (2006) Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing. Mol Breed 17: 281-289.
2. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651-1656.
3. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analysis of coffee and related species. Theor Appl Genet 114:359-372.
4. Aitken N, Smith S, Schwartz C, Morin PA (2001) Single nucleotide polymorphisms (SNP) discovery: a targeted-gene approach. Mol Ecol 13: 1423-1431.
5. Barker G, Batley J, O'Sullivan H, Edwards KJ, Edwards D (2003) Redundancy based detection of sequence polymorphism in expressed sequence tag data using autoSNP. Bioinformatics 19: 421-422.
6. Batley J, Baker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertion/deletions in maize expressed sequence tag data. Plant Physiol 132: 84-91.
7. Benson G (1999) Tandem repeat finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573-580.
8. Bester AE, Slabbert R, D'Amato ME (2004) Isolation and characterisation of microsatellite markers in South African abalone (*Haliotis midae*). Mol Ecol Notes 4: 618-619.
9. Bester AE, Roodt-Wilding R, Whitaker HA (2008) Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae*: a targeted EST approach. Anim Genet 39: 321-324.
10. Bester-Van der Merwe AE (2009) Population genetic structure and demographical history of South African abalone, *Haliotis midae,* in a conservation context. Unpublished PhD thesis, Stellenbosch University, South Africa.

11. Blankenship SM, May B, Hedgecock D (2002) Evolution of a perfect sequence repeat locus in the context of its flanking regions. Mol Biol Evol 19: 1943-1951.

12. Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. Mol Ecol 16: 907-924.

13. Brenner S (1990) The human genome: the nature of the enterprise. CIBA Found Symp 149: 6-17.

14. Brookes AJ (1999) The essence of SNPs. Gene 234: 177-186.

15. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453-455.

16. Brouillette JA, Venta PJ (2002) Within-breed heterozygosity of canine single nucleotide polymorphisms identified by across-breed comparison. Anim Genet 33: 464-467.

17. Brouillette JA, Andrew JR, Venta PJ (2000) Estimate of nucleotide diversity in dogs with a pool-and-sequence method. Mamm Genome 11: 1079-1086.

18. Bouza C, Hermida M, Millán A, Vilas R, Vera M, Fernandez C, Calaza M, Pardo BG, Martínez P (2008) Characterization of EST-derived microsatellite for gene mapping and evolutionary genomics in turbot. Anim Genet 39: 666-670.

19. Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphism in inferences of population history. Trends Ecol Evol 18: 249-256.

20. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG (2005) Natural selection on protein-coding genes in the human genome. Nature 437: 1153-1157.

21. Cappuccio L, Pariset P, Ajmone-Marsan P, Dunner S, Cortes O, Erhardt G, Lühken G, Gutscher K, Joost S, Nijman J, Lenstra JA, England PR, Zundel S, Obexer-Ruff G, Beja-Pereira A, Valentini A, The ECONOGENE Consortium (2006) Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. Mol Ecol Notes 6: 992-997.

22. Chabane K, Ablett GA, Cordeiro GM, Valkoum J, Henry RJ (2005) EST versus genomic microsatellites markers for genotyping wild and cultivated barley. Genet Resour Crop Ev 52: 903-909.

23. Changè D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio J-M, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109: 1204-1214.

24. Chen SL, Ma HY, Jiang Y, Liao L, Meng L (2007) Isolation and characterization of polymorphic microsatellite loci from an EST library of turbot (*Scophthalmus maximus*) and cross-species amplification. Mol Ecol Notes 7: 848-850.

25. Cheng T-C, Xia Q-Y, Qian J-F, Liu QC, Lin Y, Zha XF, Xiang Z-H (2004) Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain *Dazao*. Insect Biochem Molec 34: 523-530.

26. Cibrián-Jaramillo A, Marler TE, DeSalle R, Brenner ED (2008) Development of EST-microsatellites from the cycad *Cycas rumphii*, and their use in the recently endangered *Cycas micronesica*. Conserv Genet 9: 1051-1054.

27. Coles ND, Coleman CE, Christensen SA, Jellen EN, Stevens MR, Bonifacio A, Rojas-Beltran JA, Fairbanks DJ, Maughan PJ (2005) Development and use of expressed sequenced tag libaries in quinoa (*Chenopoduim quinoa* Willd.) for discovery of single nucleotide polymorphisms. Plant Sci 168: 439-447.

28. Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8: 1229-1231.

29. Coulibaly I, Gharbi K, Danzmann RG, Yao J, Rexroad CE (2005) Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. Anim Genet 36: 309-315.

30. Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106: 912-922.

31. Dettman JR, Taylor JW (2004) Mutation and evolution of microsatellite loci in *Neurospora*. Genetics 198: 1231-1248.

32. Edwards YJ, Elgar G, Clark MS, Bishop MJ (1998) The identification and characterization of microsatellites in the compact genome of the Japanese puffer fish, *Fugu rubripes:* perspectives of functional and comparative genomic analysis. J Mol Biol 278: 843-854.

33. Elfstrom CM, Gaffney PM, Smith CT, Seeb JE (2005) Characterization of 12 nucleotide polymorphisms in weathervane scallop. Mol Ecol Notes 5: 406-409.

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

35. Estoup A, Tailliez C, Cornuet JM, Solicnac M (1995) Size homplasy and mutational processes of interrupted microsatellites in 2 bee species *Apis mellifera* and *Bombus terrestris* (*Apidae*). Mol Biol Evol 12: 1074-1084.

36. Evans B, Conod N, Elliot NG (2001) Evaluation of microsatellite primer conservation in abalone. J Shellfish Res 20: 1065-1070.

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

38. Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009) Development, characterisation and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breeding 23:85-97.

39. Fraser LG, Harvey CF, Crowhurst NR, De Silva HN (2004) EST derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108: 1010-1016.

40. Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457-1467.

41. Gebhardt F, Zanker KS, Brandt B (1999) Modulation of epidermal factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem 274: 13176-13180.

42. Gebhardt F, Burger H, Brandt B (2000) Modulation of EGFR gene transcription by a polymorphic repetitive sequence – a link between genetics and epigenetics. Int J Biol Marker 15: 105-110.

43. Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. Mamm Genome 3609-619.

44. Gupta PK, Rustgi S (2004) Molecular markers from transcribed/expressed regions of the genome in higher plants. Funct Integr Genomics 4: 139-162.

45. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95-98.

46. Hayes B, Laerdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Høyheim B (2007). An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. Aquaculture 265: 82-90.

47. 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.* J Shellfish Res 23: 379-385.

48. Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N, DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucaba T, Lacy M, Le M, Le N, Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T, Schellenberg K, Marra M (1996) Generation and analysis of 280,000 human expressed sequence tags. Genome Res 6: 807-828.

49. Hong Y-B; Liang X-Q, Liu H-Y, Zhou G-Y, Li S-X, Wen S-J (2008) Construction of genetic linkage map based on SSR markers in peanut (*Arachis hypogaea* L). Agricultural Sciences in China 7: 915-921.

50. Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, Yasuhara JC, Wakimoto BT, Myers EW, Celniker SE, Rubin GM, Karpen GH (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biol: 3: research0085.1–0085.16.

51. Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Res 9: 868-877.

52. Ju Z, Wells MC, Martinez A, Hazlehood L, Walte RB (2005) An *in silico* mining for simple sequence repeats from expressed sequenced tags of zebrafish, medaka, *Fundulus*, and *Xiphophorus. In Silico* Biology 5: 439-463.

53. Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48: 501-510.

54. Kent WJ, Haussler D (2001) Assembly of the working draft of the human genome with GigAssembler. Genome Res 11: 1541-1548.

55. Kim H, Schmidt CJ, Decker KS, Emara MG (2003) A double-screening method to identify reliable candidate non-synonymous SNPs from chicken EST data. Anim Genet 34: 249-254.

56. Kim W-J, Kim Y-O, Nam B-H, Kong HJ, Park E-M, Kang J-H, Lee J-H, Kim K-K (2009) Development of 81 new polymorphic EST-derived microsatellite markers for the olive flounder, *Paralichthys olivaceus*. Conserv Genet 10: 1105-1111.

57. Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphism form large EST collections in barley (*Hordeum vulgare* L.). Mol Genet Genomics 270: 24-33.

58. Ku H-M, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proc Nat Acad Sci USA 97: 9121-9126.

59. Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392: 917-920.

60. Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48: 985-998.

61. 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.

62. Li W-H, Yi S, Makova K (2002) Male-driven evolution. Curr Opin Genet Dev 12: 650-656.

63. Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 21: 991-1007.

64. Lindqvist C, Scheen A-C, Yoo M-J, Grey P, Oppenheimer DG, Leebens-Mack JH, Soltis DE, Soltis PS, Albert VA (2006) An expressed sequence tag (EST) library from developing fruits of an Hawaiian endemic mint (*Stenogyne rugosa*, Lamiaceae): characterization and microsatellite markers. BMC Plant Biol 6: 16.

65. Lydeard C, Holznagel WE, Schnare MN, Gutell RR (2000) Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. Mol Phylogenet Evol 15: 83-102.

66. Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15: 47-56

67. Maneeruttanarungroj C, Pongsonboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J, Li Y, Whan V, Chu K-H, Li CP, Tong J, Glenn K, Rothschild M, Jerry D, Tassanakajon A (2006) Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). Anim Genet 37: 363-368.

68. Manly BFJ (1985) Linkage disequilibrium and selection at two or more loci. In: Usher MB, Rosenzwig ML (eds) The statistics of natural selection on animal populations. Chapman and Hall, London, pp. 309-341.

69. Mgaya YD, Gosling EM, Mercer JP, Donlon J (1995) Genetic variation at three polymorphic loci in wild and hatchery stocks of abalone, *Haliotis tuberculata* Linnaeus. Aquaculture 136: 71-80.

70. Morin PA, Luikart G, Wayne RK, The SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends Genet 19: 208-216.

71. Nagaraj SH, Gasser RB, Ranganathan S (2006) A hitchhicker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8: 6-21.

72. Pariset L, Cappuccio I, Joost S, D'Andrea M, Marletta D, Ajmone-Marsan P, Valentini A, The ECONOGENE Consortium (2006) Characterization of single nucleotide polymorphisms in sheep and their variation as evidence of selection. Anim Genet 37: 290-292.

73. Pashely CH, Ellis JR, McCauley DE, Bruke JM (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. J Hered 97: 381-388.

74. Pérez F, Ortiz J, Zhinaula M, Gonzabay C, Calderón J, Volckaert FAM (2005) Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei, Litopenaeus stylirostris,* and *Trachypenaeus birdy.* Mar Biotechnol 7: 554-569.

75. Picvoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs From EST databases. Genome Res 9: 167-174.

76. Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol Ecol 11: 603-612.

77. Putney SD, Herlihy WC, Schimmel P (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. Nature 302: 718-721.

78. Qi H, Liu X, Zhang G, Wu F (2009) Mining expressed sequences for single nucleotide polymorphisms in the Pacific abalone *Haliotis discus hannai.* Aquac Res 40: 1661-1667.

79. Quilang J, Wang S, Li P, Abernathy J, Peatman E, Wang Y, Wang L, Shi Y, Wallace R, Guo X, Liu Z (2007) Generation and analysis of ESTs from the eastern oyster *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. BMC Genomics 8: 157-167.

80. Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: A new class of genetic marker in cotton. The Journal of Cotton Science 8: 112-123.

81. Ramesh KA, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterisation and utilization of EST-

derived genic microsatellite markers for genome analysis of the coffee and related species. Theor Appl Genet 114: 359-372.

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

83. Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite flanking regions: a SNP mine in South African abalone (*Haliotis midae*)*.* Anim Genet 39: 329.

84. Rohrer GA, Fahrenkrug SC, Nonneman D, Tao N, Warren WC (2002) Mapping microsatellite markers identified in porcine EST sequences. Anim Genet 33: 372-376.

85. Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8: 103-106.

86. Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker J, Buell CR (2003) Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 131: 419-429.

87. Rudd S (2003) Expressed sequence tags: alternative or compliment to whole genome sequences. Trends Plant Sci 8: 321-329.

88. Saghai-Maroof MA, Solima KM, Jorgenson RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81: 8014-8018.

89. Savage D, Batley J, Erwin T, Logan E, Love CG, Lim GAC, Mongin E, Baker G, Spangenberg GC, Edwards D (2005) SNPServer: a real-time SNP discovery tool. Nucleic Acids Res 33: W493-W495.

90. Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecology and conservation studies: a test in the Scandinavian wolf population. Mol Ecol 14: 503-511.

91. Serapion J, Kucuktas H, Feng J, Liu Z (2004) Bioinformatic mining of type 1 microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Mar Biotechnol 6: 364-377.

92. Sekino M, Hara M (2007) Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus: *Haliotis*). Conserv Genet 8: 823-841.

93. Siemon HS, Chang A, Brown GD, Koop BF (2005) Type I microsatellite markers form the Atlantic salmon (*Salmo salar*) expressed sequence tags. Mol Ecol Notes 5: 762-766.

94. Slabbert R, Ruivo NR, Van den Berg NC, Lizamore, DL, Roodt-Wilding R (2008a) Isolation and characterization of 63 microsatellite loci for the abalone, *Haliotis midae.* J World Aquacult Soc 39: 429-435.

95. Smith EJ, Shi L, Drummond P, Rodriguez L, Hamilton R, Ramlal S, Smith G, Pierce K, Foster J (2001) Expressed sequence tags for the chicken genome from a 10-day-old White Leghorn whole embryo cDNA library: DNA sequence characterization and linkage analysis. J Hered 92: 1-8.

96. Souche EL, Hellemans B, Van Houdt JKJ, Canario A, Klages S, Reinhardt R, Volckaert FAM (2007) Mining for single nucleotide polymorphisms in expressed sequence tags of the European sea bass. Journal of Integrative Bioinformatics 4: 73-82.

97. Streit K, Geiger DL, Lieb B (2006) Molecular phylogeny and the geographic origin of *Haliotidae* traced by haemocyanin sequences. J Mollus Stud 72: 105-110.

98. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare L.*). Theor Appl Genet 106: 411-422.

99. Thomson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequences weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

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

101. Vareshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173: 638-649.

102. Vasemägi A, Nilsson J, Primmer GR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22: 1067-1076.

103. Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other molecular markers and their use in animal genetics. Genet Sel Evol 34: 275-305.

104. Wang D, Liao X, Cheng L, Yu X, Tong J (2007) Development of novel EST-SSR markers in common carp by data mining from public EST sequences. Aquaculture 271: 558-574.

105. Wang HX, Li FH, Xiang JH (2005) Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis.* Aquaculture 249: 107-114.

106. Wang JPZ, Lindsay BG, Leebens-Mark J, Cui L, Wall K, Miller WC, DePamphilis CW (2004) EST clustering error evaluation and correction. Bioinformatics 20: 2973-2984

107. Wang X, Song B, Qiu X, Meng X (2009) Development of EST-SSRs in scallop (*Patinopecten yessoensis*) from sequence database. Conserv Genet 10: 1129-1131.

108. Wang Y, Ren R, Yu Z (2008) Bioinformatic mining of EST-SSR loci in the Pacific oyster, *Crassotrea gigas*. Anim Genet 39: 287-289.

109. Weckx S, Del-Favero J, Radermakers R, Cleas L, Cruts M, De Jonghe P, Van Broekhoven C, De Rijk P (2005) NovoSNP, a novel computational tool for sequence variation discovery. Genome Res 15: 436-442.

110. Westgaard J-I, Tafese T, Wesmajervi MS, Nilsen F, Fjalestad KF, Damsgard B, Delghandi M (2007) Development of ten new EST-derived microsatellites in Atlantic cod (*Gadus morhua* L.). Conserv Genet 8: 1503-1506.

111. Wu W-L, Schaal BA, Hwang C-Y, Hwang M-D, Chiang Y-C, Chiang T-Y (2003) Characterization and adaptive evolution α-Tubulin genes in the *Miscanthus sinensis* complex (*Poaceae*). Am J Bot 90: 1513-1521.

112. Yue GH, Kovac B, Orban L (2003) Microsatellites from *Clarias batrachus* and their polymorphism in other in seven additional catfish species. Mol Ecol Notes 3: 465-468.

113. Yue GH, Ho MY, Orban L, Komen J (2004) Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. Aquaculture 234: 85-98.

114. Zhan A-B, Bao ZM, Wang XL, Hu JJ (2005) Microsatellite markers derived from bay scallop *Argopecten irradians* expressed sequence tags. Fish Sci 71: 1341-1346.

115. Zhan A, Bao Z, Wang M, Chang D, Yaun J, Wang X, Hu X, Liang C, Hu J (2008) Development and characterization of microsatellite markers for the Pacific abalone (*Haliotis discus*) via EST database mining. J Ocean Univ Chin 7: 219-222.

Chapter V:
Final Remarks, Synopsis and Conclusions

1. Abalone

Abalone are marine gastropod molluscs of which 56 species are currently recognised. Individual species are geographically localised, however abalone as a genus (*Haliotis*) are globally distributed along the coastlines of most continents (Geiger 2000; Degnan *et al.* 2006). South Africa in particular has five endemic species, with a sixth southern African species occurring in the waters of Mozambique further north. Of these, the best characterised is the South African endemic *Haliotis midae*, locally known as perlemoen. Abalone demonstrates a complex life cycle, starting as pelagic planktonic larvae and progressing to the characteristic adult form. As adults these animals prefer the inter-tidal zone where they lodge onto rocks and feed on algae. They remain passive for the greater part of their life demonstrating minimal locomotion (Tarr 1995).

Internationally, where abalone are found they have become a valuable fishing commodity with the global abalone trade estimated at 629,842,000.00 US Dollars (2006 estimate; United Nations Food and Agriculture Organisation, www.fao.org/fishery). This placed considerable pressure on natural abalone resources due to overexploitation and poaching and subsequently led to the closure of the fishery in many countries (McShane *et al.* 1994; Altstatt *et al*. 1996; Hobday *et al*. 2001; Hobday and Tegner 2002). In South Africa, abalone stocks have been in decline since the all-time high landings of 1965. This led to a variety of policy implementations to aid the conservation and sustainable use of the resource, including the introduction of restricted fishing seasons and marine preserves (Tarr 1989, 1992; Troell *et al.* 2006). However in February 2008 the fishery was formally suspended (Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs and Tourism, Republic of South Africa – 04 December 2007). Natural abalone stocks are still under threat, world-wide, as abalone-related commerce remains a lucrative enterprise.

2. Aquaculture and Molecular Markers

With the closure of abalone fisheries a void was created in the market and thus it comes as no surprise that the first abalone ventures coincided with the decline in fishery landings; this was a global trend (Garza and Bernal 1992; Tong and Moss 1992; Fleming and Hone 1996). South Africa entered the abalone aquaculture market during the late 1980’s when perlemoen was first kept and reared successfully under artificial conditions (Genade *et al.* 1988; Troell *et al*. 2006 and references therein). Initial research was focused on husbandry practices for optimal production and not much attention was given to matters of breeding and genetics (Sales and Britz 2001). However as the industry is maturing and husbandry protocols become established, focus will now be turning towards the development of truly domesticated abalone strains that will demonstrate superior production traits. This is vital if South African produces are to remain internationally competitive.

Traditional animal breeding has led to considerable improvements in production traits of traditional livestock (Harris 1998) and will undoubtedly play a vital role in abalone breeding. Nonetheless, advances in various genetic technologies will see an integration of modern biotechnology and classical breeding techniques in aid of abalone genetic improvement (Elliott 2000; Hulata 2001). Currently the most feasible and technically attainable of these modern innovations is molecular marker technology. Molecular markers have in many instances become the norm for measuring genetic diversity; extrapolating observations of biological variation to the most fundamental biological unit, the deoxyribonucleotide acid sequence. The utility of molecular markers for genetic improvement lies in the construction of genetic linkage maps and using these to identify genomic regions, quantitative trait loci, that are associated to particular phenotypic characters. This information is then incorporated into the breeding program via marker assisted selection. Marker assisted selection increases the accuracy of breeding value prediction as it surveys the genetic material directly. It also decreases the generation time because genetically superior animals can now be identified even before phenotypic characters have matured. Thus marker assisted selection increases the rate of genetic change in the population; fast-tracking the domestication process (Dekkers and Hospital 2002; Andersson and Georges 2004; Dekkers 2004; Collard *et al.* 2005). Molecular markers are versatile and the application stretches beyond marker assisted selection; Roodt-Wilding and Slabbert (2006) elaborate on various additional uses for molecular markers for the abalone culture industry including, diversity estimates for population management, parentage assignment and molecular tagging.

Over the course of the development of molecular marker technologies several marker systems have come into existence (Dodgson *et al.* 1997; Beuzen *et al.* 2000; Vignal *et al.* 2002; Liu and Cordes 2004). The use of these have been paralleled to that of fashionable trends (Schlötterer 2004), nonetheless the merit of a particular marker system should be judged by the research question at hand. Presently microsatellite markers are the most used molecular marker system in animal and aquaculture genetics. This popularity is due to the marker's exceptional allelic variability and co-dominant inheritance (Beuzen *et al.* 2000; Vignal *et al.* 2002). However, single nucleotide polymorphisms (SNPs) are rapidly gaining favour, because of its mutational stability, making it less prone to genotyping anomalies. Furthermore, SNPs are responsible for the majority of phenotypic variation due to genetic variation and the high genomic frequency allows for the construction of dense marker maps, which in turn facilitates the accurate positioning of QTLs (Beuzen *et al*. 2000; Brumfield *et al.* 2003; Morin *et al.* 2004; Rengmark *et al*. 2006).

3. Aim, Objectives and Findings of this study

In recent years there has been interest in the development of type I molecular markers, i.e. molecular markers that are associated with genes. In contrast to type II markers (markers developed from anonymous DNA fragments), type I markers directly surveys the genetic diversity at regions in the genome responsible for phenotypic variation. Thus, from a functional perspective, with reference to animal breeding, type I molecular markers have a higher probability of conferring a QTL, either as the causative variant or due to its close linkage to a causative variant elsewhere in the gene. The development of type I markers do however present additional challenges to already laborious and costly *de novo* marker development protocols, because of the inclusion of an mRNA isolation step (e.g. Liu *et al.* 1999; Cheng *et al.* 2007). The accumulating sequence information in public database does, however, offer a ready source for type I marker development via several routes. It was thus the primary aim of this study to investigate the viability of using these public sequence collections to develop type I molecular markers for the South African abalone, *Haliotis midae*; an animal with limited genomic resources.

To achieve this aim, three main objectives were set. The first was to investigate the prevalence of sequence similarity between the flanking regions of characterised anonymous microsatellite markers (previously developed for *H. midae*) and genic sequences in gene sequence depositories (Chapter II). This strategy was then extended to the microsatellite markers of two additional species (*H. rubra* and *H. discus hannai*) and further tested for cross-species transfer to *H. midae* (Chapter III). Lastly an EST-microsatellite and -SNP *in silico* mining approach, followed by cross-species transfer were applicable, was adopted (Chapter IV). The three strategies were all successful to varying degrees, however none can be deemed more superior to another, as each rely on unique premises and are in fact independent strategies. Here it is merely shown that irrespective of the genomic resources available for a particular organism, public databases may still provide a valuable commodity for type I (albeit putative) marker development.

As most aquaculture species will already possess sets of anonymous markers and as new anonymous markers are generated, the bioinformatic conversion via similarity identification may be the first avenue explored in aim of identifying marker-gene associations. This method is expected to grow in utility as gene identification and sequencing projects on various species increase in future. Also, as search algorithms are refined, such homologies may be identified with greater statistical confidence. Already in this study approximately 17% of microsatellites demonstrated possible genic association, in comparison to one of the first investigations of this sort that identified only 1.7% (Herron *et al.* 1998). The ten-fold increase over the past 11 years is an indication of the sequence accumulation in public databases, but species-specific prevalence of microsatellite-gene associations must also be taken into account.

The simplicity and ease of bioinformatic conversion lends itself to be exploited in more than one way: A number of microsatellite markers have been developed and mapped for two other economically important abalone, *H. rubra* and *H. d. hannai.* Subjecting these markers to similar bioinformatic analysis makes them prime candidates for cross-species microsatellite transfer experiments. The advantage of these markers are, not only that they are type I and thus expected to transfer with greater success (Aitken 2004), but they also provide a cross-species bridge for extrapolating functional data from the source species (Chistiakov *et al.* 2006). This method holds particular merit for a species with extremely limited genomic resources; perhaps a new species under investigation. Even though this approach only marginally increased the transfer success rate from *H. rubra* to *H. midae*, potentially valuable QTL data could still be retrieved. On the contrary, transfer success was significantly increased in the case of *H. d. hannai* markers, which suggests that this method will facilitate cross-species transfer between more distantly related species, where there may be a marked difference between sequence conservation of genic and intergenic regions. It must however be stressed that these bioinformatic conversions remain putative and the persistence of false positive hits cannot be excluded. However, EST-markers and converted markers demonstrate similar transfer rates and polymorphism. This provides good evidence that most of the microsatellite-gene associations were not artefactual and lends support to the use of this bioinformatic method.

Expressed sequence tag (EST) molecular markers are *de facto* type I markers. Thus the EST route is in all likelihood the surest way to obtain large numbers of type I markers, especially considering the current trend in transcriptome sequencing as alternative to whole genome sequencing (e.g. Rudd 2003; Chanderbali *et al.* 2008; Buell 2009). Irrespective of the fact that *H. midae* is presently underrepresented in EST databases, there was a sufficient collection of Haliotid ESTs in general to facilitate marker development via a cross-species approach. Interestingly microsatellite isolation from these ESTs demonstrated a slightly increased success rate than traditional methods (Bester *et al.* 2004; Slabbert *et al.* 2008). Single nucleotide polymorphism discovery was, however, hindered by low sequence redundancy. Nonetheless, the results presented here suggests that as EST sequence collections expand, especially ESTs from *H. midae,* the number of molecular markers identified will become more substantial.

Other than the explicit development of type I molecular markers for perlemoen, considerable insights were gained with regards to the particular biological nature of microsatellites and to a lesser extent SNPs in the *H. midae* genome. Firstly, the non-random distribution of microsatellites in the genome of *H. midae* is suggested, supporting the cumulating evidence of similar trends in other organisms (e.g. Tóth *et al.* 2000; Morgante *et al.* 2002). In particular, this study highlights the abundance of CA repeat motifs in intergenic regions and GAGT repeat motifs in genic regions; in contrast CpG-like repeat motifs are rare. This non-random distribution of microsatellites, coupled with moderate to low levels of genic microsatellite polymorphism and conservation of individual loci across species boundaries, suggests that these loci are likely to be functionally active and thus subject to selective pressures (Li *et al.* 2002, 2004). Furthermore, this study is the first to present tangible evidence for the association of microsatellites to transposable elements in abalone; a common occurrence in many species (Ramsey *et al.* 1999; Meglecz *et al.* 2004, 2007) and first suggested to be the case in abalone by Baranski *et al.* (2006).

It must be noted that unexpected discrepancies do however persist between the results of Chapter II and Chapter IV in terms of microsatellite genic distribution: Bioinformatic analysis seem to suggest a comparatively low repeat-gene association for dinucleotides, in contrast to what is suggested by the EST analysis. This discrepancy can be explained as a marker selection bias created by microsatellite isolation protocols, during which longer fragments are selected to maximise polymorphism. Therefore gene-linked dinucleotides may be overlooked due to their lack of or low levels of polymorphism. A further inconsistency between the two data sets was observed with regards to TA motif prevalence. TA motifs are well-represented in ESTs, but seemingly absent in the genomic set of microsatellites. Again this may be attributed to a selection bias of microsatellite isolation protocols against palindromic sequences (Powell *et al.* 1996). However a genuine low genomic frequency and a high genic frequency (possibly with functional importance) of TA repeats cannot be excluded. It is for this reason that caution must be taken when interpreting such results.

With regard to SNPs, this study reaffirms the high SNP frequencies and askew transition to transvertion ratio as previously reported for *H. midae* (Bester *et al.* 2008; Rhode *et al.* 2008). This high SNP frequency is probably due to the reproductive strategy of abalone, highly fecund broadcast spawning, which favours the accumulation of mutations (Li *et al.* 2002; Hedgecock *et al.* 2004). This high SNP frequency is ultimately responsible for the low success rate of cross-species marker transferability in abalone leading to primer binding site sequence variations and rapid decay of microsatellites. The aberrant transition to transverion ratio remains unresolved and further genomic investigation is necessary; however there might be a correlation between the lack of transitions and the lack of highly mutable CpG-like motifs in the *H. midae* genome (Vignal *et al.* 2002).

4. Final Conclusion

This study presents the results of using *in silico* resources for the development of type I, gene-associated molecular markers; microsatellites and single nucleotide polymorphism. These markers represents the next generation of genetic markers that have progressed from simple monogenic morphological traits to the present day's molecular marker. Their genic association makes these markers prime candidates for use in QTL and functional investigations. However their applicability may stretch much further. It has been shown that this *in silico* method may yield a sufficient number of molecular markers for initial diversity studies; even for an organism with modest genomic resources. Furthermore, many cross-species microsatellites demonstrated a pattern of cross-species variation indicative of speciation events, therefore these may provide insights into abalone evolution in future studies and may be used currently as species diagnostic markers. This study further presents the first attempts at cross-species characterisation of functional genomic regions of economically important abalone species. Although currently limited, it provides the first steps toward comparative and trans-species functional genomic mapping. Future investigations may build upon these results to extrapolate functional data from various abalone species in aid of gaining greater understanding of the molecular constituents responsible for phenotypic variation. In doing so, breeding objectives, whether for conservation or domestication, can be better defined in order to achieve the breeding goals. However, this strategy alone will not be enough to produce genetic maps of sufficient density to allow for the implementation of a molecular breeding program. Nonetheless it remains a dynamic and valuable resource for saturating and annotation of genomic maps, providing insights into possible candidate genes and other biological phenomena.

References

1. Aitken N, Smith S, Shwartz C, Morin PA (2001) Single nucleotide polymorphisms (SNP) discovery: a targeted-gene approach. Mol Ecol 13: 1423-1431.
2. Altstsatt JM, Ambrose RF, Engle JM, Haaker PL, Lafferty KD, Raimondi PT (1996) Recent declines in black abalone *Haliotis cracherodii* on the mainland coastal of central California. Mar Ecol Prog Ser 142: 185-192.
3. Andersson L, Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet 5: 202-212.
4. Baranski M, Rourke M, Loughnan S, Austin CM, Robinson N (2006) Isolation and characterization of 125 microsatellite DNA markers on blacklip abalone, *Haliotis rubra.* Mol Ecol Notes 6: 740-746.
5. Bester AE, Slabbert R, D’Amato ME (2004) Isolation and characterisation of microsatellite markers in South African abalone (*Haliotis midae*). Mol Ecol Notes 4: 618-619.

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

7. Beuzen ND, Stear MJ, Chang KC (2000) Molecular markers and their use in animal breeding. Vet J 160: 42-52.

8. Brumfield RT, Beerli P, Nickers DA, Edwards SV (2003) The utility of single nucleotide polymorphism in inferences of population history. Trends Ecol Evol 18: 249-256.

9. Buell CR (2009) Poaceae genomes: Going from unattainable to becoming a model clade for comparative plant genomics. Plant Physiol 149: 111-116.

10. Chanderbali AS, Albert VA, Ashworth VETM, Clegg MT, Litz RE, Soltis DE, Soltis PS (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. BioEssays 30: 386-396.

11. Cheng L, Liao X, Yu X, Tong J (2007) Development of EST-SSR by an efficient FIASCO-based strategy: A case study in rare minnow (*Gobiocyrpis rarus*). Animal Biotech 18: 143-152.

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

13. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196.

14. Degnan SM, Imron, Geiger DL, Degnan BM (2006) Evolution in temperate and tropical seas: Disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae). Mol Phylogenet Evol 41: 249-256.

15. Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J Anim Sci 82 (E Suppl): E313-E328.

16. Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3: 22-32.

17. Dodgson JB, Cheng HH, Okimoto R (1997) DNA marker technology: a revolution in animal genetics. Poultry Sci 76: 1108-1114.

18. Elliott NG (2000) Genetic improvement programmes in abalone: what is the future. Aquac Res 31: 51-59.

19. Fleming AE, Hone PW (1996) Abalone aquaculture. Aquaculture 140: 1-4.

20. Garza AES, Bernal RS (1992) Development and present status of abalone culture in Mexico. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 538-546.

21. Geiger DL (2000) Distribution and biogeography of the *Haliotidae* (Gastropoda: Vetigastropoda) world-wide. Boll Malacol 35: 57-120.

22. Genade AB, Hirst AL, Smit CJ (1988) Observations on the spawning, development and rearing of the South African abalone *Haliotis midae* Linn. S Afr J Mar Sci. 6: 3-12.

23. Harris DL (1998) Livestock improvement: Art, science, or industry. J Anim Sci 76: 2294-2301.

24. 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.* J Shellfish Res 23: 379-385.

25. Herron BJ, Silva GH, Flaherty L (1998) Putative assignment of ESTs to the genetic map by use of the SSLP database. Mamm Genome 9: 1072-1074.

26. Hobday AJ, Tegner MJ (2002) The warm and the cold: Influence of temperature and fishing on the local population dynamics of the red abalone. CalCOFI Rep 43: 74-96.

27. Hobday AJ, Tegner MJ, Haaker PL (2001) Over-exploitation of a broadcast spawning marine invertebrate: Decline of the white abalone. Rev Fish Biol Fisher 10: 493-514.

28. Hulata G (2001) Genetic manipulation in aquaculture: a review of stock improvement by classical and modern technologies. Genetica 111: 155-173.

29. Li W-H, Yi S, Makova K (2002) Male-driven evolution. Curr Opin Genet Dev 12: 650-656.

30. Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465.

31. Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 21: 991-1007.

32. Liu Z, Tan G, Li P, Dunham RA (1999) Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. Biochem Biophys Res Commun 259: 190-194.

33. Liu ZJ, Cordes JF (2004) DNA marker technologies and their application in aquaculture genetics. Aquaculture 238: 1-37.

34. McShane PE, Mercer SF, Naylor JR (1994) Spatial variation and commercial fishing of New Zealand abalone (*Haliotis iris* and *Haliotis australis*). New Zeal J Mar Fresh 28: 345-355.

35. Meglecz E, Petenian F, Danchin E, D' Acier AC, Rasplus J-Y, Raure E (2004) High similarity between flanking regions of microsatellite detected within each of two species of Lepidoptera: *Paenassius apollo* and *Euphydryas aurinia*. Mol Ecol 13: 1693-1700.

36. Meglecz E, Anderson SJ, Bourguet D, Bucher R, Caldas A, Cassel-Lundhagen A, D' Acier AC, Dawson DA, Faure N, Fauvelot C, Frank P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju C, Patt A, Petenian F, Silvian J-F, Wilcock HR (2007) Microsatellite flanking region similarities among different loci within insect species. Insect Mol Biol16: 175-185.

37. Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat Genet 30: 194-200.

38. Morin PA, Luikart G, Wayne RK, The SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends Genet 19: 208-216.

39. Powell W, Macrhay GC, Provan J (1996) Polymorphisms revealed by simple sequence repeats. Trends Plant Sci 1: 215-222.

40. Ramsey L, Macaulay M, Cardle L, Morgante M, Ivanissevich SD, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in the barley. The Plant Journal 17: 415-425

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

42. Rhode C, Slabbert R, Roodt-Wilding R (2008) Microsatellite flanking regions: a SNP mine in South African abalone (*Haliotis midae*). Anim Genet 39: 329

43. Roodt-Wilding R, Slabbert R (2006) Molecular markers to assist the South African abalone industry. S Afr J Sci 102: 99-102.

44. Rudd S (2003) Expressed sequence tags: alternative or compliment to whole genome sequences. Trends Plant Sci 8: 321-329.

45. Sales J, Britz PJ (2001) Research on abalone (*Haliotis midae* L.) cultivation in South Africa. Aquac Res 32: 863-874.

46. Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? Nat Rev Genet 5: 63-68.

47. Slabbert R, Ruivo NR, Van den Berg NC, Lizamore DL, Roodt-Wilding R (2008) Isolation and characterization of 63 microsatellite loci for the abalone, *Haliotis midae*. J World Aquacult Soc 39: 429-435.

48. Tarr RJQ (1989) Abalone. In: Payne AIL, Crawford RJM (eds) Oceans of life off southern Africa. Vlaeberg Publishers, Cape Town, pp.62-68.

49. Tarr RJQ (1992) The abalone fishery of South Africa. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 438-447.

50. Tarr RJQ (1995) Growth and movement of the South African abalone *Haliotis midae*: a Reassessment. Mar Freshwater Res 46: 583-590.

51. Tong LJ, Moss GA (1992) The New Zealand culture system for abalone. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds) Abalone of the World: Biology, Fishery and Culture. Fishing News Books. pp. 583-591.

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

53. Troell M, Robertson-Anderson D, Anderson RJ, Bolton JJ, Maneveldt G, Halling C, Probyn T (2006) Abalone farming in South Africa: An overview with perspectives on kelp resources, abalone feed, potential for on-farm seaweed production and socio-economic importance. Aquaculture 257: 266-281.

54. Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other molecular markers and their use in animal genetics. Genet Sel Evol 34: 275-305.

Appendices: Supplementary Information

Chemical Reagents and Electrophoresis Gels

5X Tris-Borate-EDTA-Buffer:

- 54g Tris-base ($C_4H_{11}NO_3$, M_r = 121.14)
- 27.5g Boric acid (H_3BO_3, M_r = 61.83)
- 3.725g EDTA disoduim salt ($C_{10}H_{14}N_2Na_2O_8.2H_2O$, M_r = 372.24)
- Bring to volume with autoclaved dH_2O (Final volume = 1L)

Bromophenol Blue loading dye:

- 22.7ml - Formamide (CH_3NO, M_r = 45.04)
- 0.168g - EDTA disoduim salt ($C_{10}H_{14}N_2Na_2O_8.2H_2O$, M_r = 372.24)
- 0.0125 - Bromophenol blue ($C_{19}H_{10}Br_4O_5S$, M_r = 669.96)
- Bring to final volume of 25ml, with dH_2O

2% Agarose Gel:

- 2g of agarose for every 100ml 1X TBE
- 0.05ng/µl of Ethidium Bromide

12% Polyacrylaminde Gel:

- 3ml - Acrylamide/Bis-acrylamide (49:1) (SIGMA®, prepared to manufacturer's specification)
- 2ml - dH_2O
- 300µl - Ammonium Persulfate ($NH_4.2S_2O_8$ M_r = 228.2) (10% m/v in dH_2O)
- 30µl - N,N,N',N' – Tetramethyl-ethylenediamine (Temed) ($C_6H_{16}N_2$, M_r = 116.2)
- Staining solution: 0.05ng/µl of Ethidium Bromide in 1X TBE-buffer

Identities, PCR conditions and cycling program

Identities

Table S1: Different microsatellite loci that demonstrated significant BLAST hits to the same gene, however not representing the same locus due to different Identity-values. This phenomenon could be explained by the presence of a conserved regulatory element in a gene family or a cluster of genes involved in similar biological processes.

Microsatellite	Gene Hit	Identities (%)
HmD33 (AY303334)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	82
HmD30 (AY303342)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	90
HmLCS73T (DQ993219)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	71
HmLCS67M (DQ993222)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	84
HmNS31D (EF033333)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	82
HmDL151T (EF054868)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	89
HmDL214T (EF054871)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	79
HmNS100T (EF367114)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	87
HmNST7T (EF455618)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	89
HLCS147T (GQ927134)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	81
HmS104 (GQ927137)	Cellulase gene *Haliotis discus hannai* (AB125892.1)	79
HmNS31D	Actin A2 gene	79

(EF033333)	*Haliotis iris* (AY921238.1)	
HmNR191T (EF121752)	Actin A2 gene *Haliotis iris* (AY921238.1)	80
HmNR281P (EF512274)	Actin A2 gene *Haliotis iris* (AY921238.1)	82
HmG16D (DQ785744)	H1 hemocyanin *Haliotis tuberculata* (AJ252741.1)	82
HmidPS1.1007C (GU256729)	H1 hemocyanin *Haliotis tuberculata* (AJ252741.1)	90

Primer information:

Table S2: PCR conditions and primer information for the cross-species microsatellite markers from *Haliotis rubra* and *Haliotis discus hannai*.

Locus Name	T_a (ºC)[1]	$[MgCl_2]$[2] (mM)/ Buffer[3] (X)	CP[4]	K/P[5]	Primer sequence
Hrub11.A07 (DQ278009)	55	2.5/1	4	P	**F**: *NED*-AAAGAACTTCTCGCCGAACA **R**: CAGCATGACCAAAACACCTG
Hrub12.D02 (DQ278021)	48	2.5/1	4	P	**F**: TTTGTCACAGCTTCGCATAA **R**: TGGGTGTTCTGCAATTAACA
Hrub12.E10 (DQ278024)	55	2.5/1	4	P	**F**: *VIC*-TGCAGCATAACACTTGCTCA **R**: CGTAGCTGCCTTCATCCTTC
Hrub15.A01 (DQ278045)	51	2.5/1	4	P	**F**:*FAM*-ACCGTTGGGATGACTGAAAG **R**: CCTGCGTAGGCGACATTTA
Hrub17.D11 (DQ278053)	56	2.5/1	4	P	**F**: *NED*-GTGGCTGAAAGGTTCAAACG **R**: GCCGAACGTTGAGGAGTATG
Hrub4.F07 (DQ278075)	54	2.5/1	4	P	**F:** AGTCGAACGAGGAGAACGAA **R**: GCTCCATTCATGGACAATCC
Hrub9.E04 (DQ278108)	55 (TD:65-55)	2.5/1	1	P	**F**: CACTGGATTGTGTGACCCTG **R**: CAGCGCAAAAGTGTTTTTCA
Hrub6.C04 (DQ278083)	52	2.5/1	4	P	**F**: CGTTGGTGGGTTCTCTTGA **R**: GGATGCTAGGGCATTATCCA
Hrub4.E05 (DQ278073)	55 (TD:65-55)	2.5/1	1	P	**F**: GTTTTGAAACCCGTTGCTGT **R**: CAATGCTCATTCCCACTCAC

Hrub12.B10 (DQ278020)	50	2.5/1	4	P	**F**: *VIC*-GGCGAGGTATTGCTTCTTTG **R**: GCGTAAGATAAAACCGTTTGAGA
Hrub13.F06 (DQ278037)	55	2.5/1	4	P	**F**: *PET*-GACAGGTGCTCCCCTATTCA **R**: CCAGGTGTCAACATGACCTG
Hrub9.B05 (DQ278104)	50	2.5/1	4	P	**F**: AATCCGGAATACTGCACTGG **R**:*NED*-AGGTCATATTGTCCACCGGA
Hrub16.G01 (DQ278051)	58	2.5/1	4	P	**F**: GCACTTGTTCCCATGGAATG **R**: CCCATGTTTCTTTTGCACCT
Afa068 (AB239627)	56	1.5/1	5	P	**F**: TTAAGGGGCTAAAACAAGGACTGG **R**: AATTTCGGTTTCGGGTTTAATGTC
Afa194 (AB239713)	50	1.5/1	5	P	**F**: ATGAGGTCGTCGATAACCCCATG **R**: CGCTAAAGGTTTTCGCTAGTCCAA
Afa207 (AB239716)	57	1.5/1	5	P	**F**: GAATATTTGCTAAAAGCGGCGTAA **R**: GCTAATGACACGGAAACCTCGAC
Awb083 (AB177936)	50	2.0/1.5	2	K	**F**: *NED*-GCTTAGAAGGGACATAACTCGCAATA **R**: AATAGACATTCTACAAGCGAGGAAA
Awb028 (AB177917)	50	1.5/1	6	P	**F**: CAGATCGTATAATCATCTGTAACAC **R**: GAGATGCTATCAGAACATATTAAGTA
Awb098 (AB177939)	60	1.5/1.5	3	K	**F**:*VIC*-ACATGGAACTGCGAGTCCTAGAAGC **R**: TGATTATTTTCAGATCGCCGTCATA
Awb041 (AB177924)	45	2.0/1.5	2	K	**F**:FAM-CAGCAATATTCTAGCATGACGGTGG **R**: ACATACGCATCATGTTGGAAAGCAC

1 – Annealing temperature; 2 – Magnesium chloride concentration in millimolar; 3 – Buffer concentration; 4 – PCR cycle program number, see below; 5 – KAPA^{TM} or Promega PCR Kit. Primers taken from: Bananski et al. (*2006a) (*Hrub.- *primers); Sekino* et al. *(2006)* (Afa-*primers); Sekino* et al. *(2004)* (Awb-*primers) – See Chapter 2.*

Table S3: PCR conditions and primer information for the EST-SNP contigs.

Contig Name	T_a[1] (ºC)	$[MgCl_2]$[2] (mM)/ Buffer[3] (X)	CP[4]	K/P[5]	Primer sequence
Hdd.c2 (GU263793)	50 (TD: 60-50)	1.5/1.2	1	P	**F**: AGGGTGTGATGGTCGGTATG **R**: AGGGCGTAACCCTCGTAGAT
Hdd.c148 (GU263794)	50 (TD: 60-50)	1.5/1.0	1	P	**F**: CACGCTATCATGAGGTTGGA **R**: CTTCATTGTGCTTGGTGCAA
Hdd.c106 (GU263795)	52 (TD:62-52)	1.5/1.0	1	P	**F**: GCTTCCTGATCTTCCACAGC **R**: AATCAGACGGTTGAGGTTGG
Hm.c04 (GU263796)	55 (TD:65-55)	1.5/1.2	1	P	**F**: TATCTGTCGGGCCATGGTAT **R**: TGACGTTACGTTAAGGGTTGA

Hm.c05 (GU263797)	52 (TD:62-52)	1.5/1.0	1	P	**F**: ATGTGCTGACATGGGCTGTA **R**: TCCAGTGATCAACAGCAAGG
Ha.c500 (GU263798)	50 (TD: 60-50)	1.5/1.0	1	P	**F**: TAAGAGCATGGGGGTGACTC **R**: TTCCCTGATGCAACCATACA

1 – Annealing temperature; 2 – Magnesium chloride concentration in millimolar; 3 – Buffer concentration; 4 – PCR cycle program number, see below; 5 – KAPA™ or Promega PCR Kit.

Table S4: PCR conditions and primer information for the EST-microsatellite markers.

Locus Name	**T$_a$ (°C)[1]**	**[MgCl$_2$][2] (mM)/ Buffer[3] (X)**	**CP[4]**	**K/P[5]**	**Primer sequence**
HdSSRex495 (ex534495)	55	1.5/1.5	2	K	**F**: CGCGGCATTAAGGAAATAAA **R**: *NED*-CAGTGTTAAACGTCGCATTGA
HaSSRgd842 (gd241842)	55	1.5/1.5	2	K	**F**: GAAAAGCAGTCTCAGCGTCAG **R**: *VIC*-CGAGATCAGCTTGTAGACTTGG
HaSSRdw239 (dw986239)	52	1.5/1.5	2	K	**F**: *VIC*-CACCAGGCAACAATCATCTG **R**: CCTGTTGTGTTTGACCGTTG
HmSSRex489b (ex534489)	55	1.5/1.5	2	K	**F**: *PET*-ATGAATCATGAAGCCAAGACG **R**: ACAATATGGCACCATGAGGAT
HaSSRdw503 (dw986503)	55	2.0/1.5	2	K	**F**: *FAM*-TAAGAGCATGGGGGTGACTC **R**:TTCCCTGATGCAACCATACA
HmSSRex489a (ex534489)	55	1.5/1.5	2	K	**F:** GGTTGCTCTGGTTCAAATCC **R**: *PET*-TCAAGGTATGCTCAGTGTGGA
HdhSSRfe537 (FE041537)	50	2.0/1.5	2	K	**F**: CGCCCGGGCAGGTACAAA **R**: *NED*-TCTATGTTAGGGTCCCTCCT
HdSSRcx732 (cx726732)	55	1.5/1.5	2	K	**F**: *NED*-GTTCTCTGACACGCCTCTCC **R**: CTGCGGTAGGCGATGTTCT
HmSSRex446a (ex534446)	56/ (TD: 66-56)	2.0/1.5	3	K	**F**: *PET*-GGTAGGGTGGGTTGGTTGAT **R**: AGACAAATCTCCTGAATCTCCA
HaSSRc.571a (GU263799)	55	1.5/1.5	2	K	**F**: CCAGTTGCCAAGGAGACACT **R**: *VIC*-ACAACCAACACGCACTGACAT
HmSSRex446b (ex534446)	55	1.5/1.5	2	K	**F**: *VIC*-AATGTGTGAAGCCCATTTCTG **R**: TCGAAGCACGTGTAAAATCCT
HdhSSRc.60b (GU263800)	55 (TD: 65-55)	2.0/1.5	3	K	**F**: *FAM*-CTAGGTTTGTCAGCACCATAC **R**: ACCCTGTGGAATAAGGTTTT
HaSSRgd475c (gd272475)	55	1.5/1.5	2	K	**F**: TCCTGGACAAAATGGTGTAGC **R**: *VIC*-TGTGTATCTGCTCCCTCCTTG
HaSSRdy903 (dy402903)	55	1.5/1.5	2	K	**F**: *FAM*-AGTGTGGAGGAACCAAGGTG **R**: GCAGTTTCCTTCCATTCTCC

HdSSRcx009 (CX726009)	56/ (TD: 66- 56)	2.0/1.5	3	K	**F**: AGACAAATGAGCTGCCTGAAG **R**: *NED*-GATACAAAGTCAGCAGGACACG:

1 – Annealing temperature; 2 – Magnesium chloride concentration in millimolar; 3 – Buffer concentration; 4 – PCR cycle program number, see below; 5 – KAPA^TM^ or Promega PCR Kit.

PCR Cycling Programs

Cycle-program 1: Touch Down (TD)

Note: Annealing temperatures (T_a) decrease with increments of 1ºC, all cycles are repeated twice with exception of the final T_a that cycles for 30 rounds.

Initial denaturing	05 min.	94ºC	

Denaturing	30 sec.	94ºC	
Annealing	30 sec.	$T_{a(1)}$ºC	X 02
Extension	30 sec.	72ºC	

Denaturing	30 sec.	94ºC	
Annealing	30 sec.	$T_{a(2)}$ºC	X 02
Extension	30 sec.	72ºC	

Denaturing	30 sec.	94ºC	
Annealing	30 sec.	$T_{a(3)}$ºC	X 02
Extension	30 sec.	72ºC	

Denaturing	30 sec.	94ºC	
Annealing	30 sec.	$T_{a(4)}$ºC	X 02
Extension	30 sec.	72ºC	

Denaturing	30 sec.	94ºC	
Annealing	30 sec.	$T_{a(5)}$ºC	X 02

Extension	30 sec.	72ºC	
Denaturing	30 sec.	94ºC	X 02
Annealing	30 sec.	$T_{a(6)}$ºC	
Extension	30 sec.	72ºC	
Denaturing	30 sec.	94ºC	X 02
Annealing	30 sec.	$T_{a(7)}$ºC	
Extension	30 sec.	72 ºC	
Denaturing	30 sec.	94ºC	X 02
Annealing	30 sec.	$T_{a(8)}$ºC	
Extension	30 sec.	72ºC	
Denaturing	30 sec.	94ºC	X 02
Annealing	30 sec.	$T_{a(9)}$ºC	
Extension	30 sec.	72ºC	
Denaturing	30 sec.	94ºC	X 02
Annealing	30 sec.	$T_{a(10)}$ºC	
Extension	30 sec.	72 ºC	
Denaturing	01 min.	94ºC	X 30
Annealing	01 min.	$T_{a(11)}$ºC	
Extension	01 min.	72ºC	
Final extension	07 min.	72ºC	

Cycle-program 2 (KAPA2G^{TM} Fast HotStart Program)

Initial denaturing	05 min.	95ºC

Denaturing	15 sec.	94ºC	X 30
Annealing	30 sec.	T_aºC	
Extension	30 sec.	72ºC	
Final extension	07 min.	72ºC	

Cycle-program 3 (KAPA2G^{TM} Fast HotStart Touch down Program)

Initial denaturing	05 min.	95ºC	
Denaturing	15 sec.	94ºC	X 10
Annealing	30 sec.	$T_{a(1)}$ºC	
Denaturing	15 sec.	94ºC	X 30
Annealing	30 sec.	$T_{a(2)}$ºC	

Cycle-program 4

Initial denaturing	10 min.	95ºC	
Denaturing	45 sec.	94ºC	X 30
Annealing	45 sec.	T_aºC	
Extension	45 sec.	72ºC	
Final extension	10 min.	72ºC	

Cycle-program 5 (Sekino *et al.* 2006)

Initial denaturing	12 min.	94ºC	
Denaturing	30 sec.	94ºC	X 35
Annealing	30 sec.	T_aºC	
Extension	01 min.	72ºC	

Final extension	20 min.	72ºC

Cycle-program 6 (Sekino and Hara 2001)

Initial denaturing	12 min.	94ºC	
Denaturing	30 sec.	94ºC	X 35
Annealing	30 sec.	T_aºC	
Extension	01 min.	72ºC	
Final extension	05 min.	72ºC	