UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

MICROSATELLITE MARKER DEVELOPMENT AND PARENTAGE ASSIGNMENT IN *HALIOTIS MIDAE*

BY

NICOL-CANDICE VAN DEN BERG

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science at the University of Stellenbosch

March 2008

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: _____

Date: _____

Abstract

The five leading abalone producers in South Africa have initiated a genetic enhancement program for *Haliotis midae* in a collaborative effort to improve economically valuable traits. Several independent objective-specific studies were initiated, including the establishment of a Performance Recording Scheme (PRS), utilised in this study, and necessary to monitor the ongoing performance of individuals as the move from mass-selection to marker assisted selection (MAS) is implemented.

The *primary objective* of this study was parentage assignment of F1 offspring mass-selected for size at approximately one year and allocated to either a "faster" or a "slower" growth group. Nine microsatellite markers were used to genotype juveniles and potential parents, with assignment completed using CERVUS 2.0.

Average growth results for Abagold and HIK were comparable for both growth groups. Slight environmental effects, although not statistically significant, were evident as growth advantages for juveniles within the faster growth group at two of the five locations and for juveniles within the slower growth group at one of the five rearing locations. Despite measures to standardise environmental influences, variables are difficult to control within the reality of a production environment; and potential genotype x environment interactions may require further investigation and factoring into future breeding programs.

The additional costs associated with MAS often make the technology prohibitive to most aquaculture operations, despite the significant genetic gains to be realised from its implementation. Cost-optimising routine processes such as DNA extractions may be one approach to reduce these additional costs. Chelex®100 appears to be a suitable alternative to the CTAB method – being quick and cost-effective to perform. Applying this method in combination with the high throughput of a robotic platform warrants further evaluation.

For the microsatellite development, 50% of positive recombinant clones contained inserts. Sequencing of these clones produced 16% perfect repeats and 47% imperfect repeats for which 52 primer sets were designed and tested. In total, 31 polymorphic microsatellite loci of different motifs and composition were developed. Sixty-one percent of sequenced clones were deemed redundant and pre-screening for both uniqueness and the presence of microsatellites would reduce unnecessary sequencing thus improving the efficiency of the FIASCO method and reducing costs.

Nine loci were selected for parentage assignments. Null alleles were present for all the selected markers; however, frequencies were below the critical level of 5%.

Parentage yielded 91% and 90% successful assignment for Abagold and HIK respectively; however, observations indicate that a measure of relatedness may exist between breeders. Recommendations with regards to future family breeding include, for both Abagold and HIK, retaining selected breeders based on their respective contributions to the F1 progeny while reassessing the potential of remaining breeding stock under more controlled breeding conditions.

No obvious trends were observed for growth with most individuals producing both faster and slower growing offspring. Juveniles will be reassessed at two years to determine whether the size advantage or disadvantages were maintained and to ascertain whether growth advantages/disadvantages may be gender specific.

Opsomming

Die vyf mees toonaangewende perlemoen produseerders in Suid Afrika het ‘n genetiese verbeteringsprogram vir *Haliotis midae* geinisieer in ‘n gesamentlike poging om ekonomiese belangrike eienskappe te verbeter. Verskeie onafhanklike fokus-spesifieke studies is geinisieer, insluitend die totstandkoming van ‘n groeiprestasie aantekenstelsel, soos gebruik in hierdie studie, en wat noodsaaklik is om die aaneenlopende prestasie van individue te moniteer soos daar beweeg word van massa seleksie tot merker bemiddelde seleksie.

Die *primêre fokus* van hierdie studie was die ouerskapsbepaling van F1 nageslag wat massa geselekteer is op ouderdom 1 jaar vir grootte en as of “vinniger” of “stadiger” groeiers geklassifiseer is. Nege mikrosatelliet merkers is gebruik om jong perlemoen individue en moontlike ouers te genotipeer, met die ouerskapstoekenning bereken deur CERVUS 2.0.

Groei resultate vir Abagold en HIK was vergelykbaar vir beide groei groepe op drie van die lokaliteite. Geringe omgewingseffekte, alhoewel nie statisties betekenisvol nie, was sigbaar as ‘n groei voordeel vir jong individue op twee van die vyf lokaliteite. Ongeag maatstawe om omgewingsinvloede te standardiseer, is varieerbares moeilik om te beheer in die produksie omgewing en genotipe x omgewings interaksies mag verdere navorsing vereis en behoort in ag geneem te word in toekomstige telingsprogramme.

Die onkoste wat met merker bemiddelde seleksie geassosieer word, maak die tegniek soms onaantreklik vir die meeste akwakultuur operasies; nie teen staande die genetiese voordele wat die gebruik daarvan veroorsaak. Die koste-optimiseering van roetine prosesse, soos byvoorbeeld, DNA ekstraksies, is dalk een aanslag om die addisionele koste te verminder. Chelex®100 blyk ‘n geskikte alternatief tot die CTAB metode te wees – die tegniek is vinnig en koste-effektief om uit te voer. Die gebruik van hierdie metode in kombinasie met die hoë deurvloei van ’n robotiese sisteem behoort verder ondersoek te word.

Vir die mikrosatelliet ontwikkeling het slegs 50% van die positiewe rekombinante klone invoegings bevat. Nukleotiedvolgorde bepaling van hierdie klone het 16% perfekte herhalings en 47% onderbroke herhalings bevat waaruit 52 inleierstelle ontwikkel en getoets is. In totaal is 31 polimorfiese mikrosatelliet loki van verskillende motiewe en samestelling ontwikkel. Een-en-sestig persent van die volgorde bepaalde klone is oortollig geag en vooraf sifting vir beide uniekheid en die teenwoordigheid van mikrosatelliete sal onnodige volgorde bepaling verhoed, die effektiwiteit van die FIASCO tegniek verhoog sowel as addisionele koste verminder.

Nege loki is geselekteer vir ouerskapsbepaling. Nul allele was teenwoordig vir al die geselekteerde merkers, maar die frekwensies was egter laer as die 5% kritieke waarde.

Ouerskap is 91% en 90% suksesvol bepaal vir Abagold en HIK onderskeidelik. Waarnemings dui egter daarop dat daar verwantskappe mag wees tussen van die broeidiere.

Voorstelle in terme van toekomstige familie teling sluit is, vir beide Abagold en HIK, om geselekteerde broei diere te behou gebaseer op hulle onderskeie bydraes tot die F1 nageslag asook die herevaluaring van die potensiaal van die oorblywende broei diere onder meer beheerde teling toestande.

Geen voor-die-handliggende tendense is waargeneem vir groei nie met die meeste individue wat beide vinniger en stadiger groeiende nageslag geproduseer het. Jong individue moet geherevalueer word op tweejarige ouderdom om te bepaal of die groei voordeel of nadele behou is en om te bepaal om groei voordele/nadele geslagspesifiek is.

Stellenbosch University

Supervisor: Dr Rouvay Roodt-Wilding
Department of Genetics
Division of Aquaculture

Signature: _____

Date: _____

Project Leader: Professor Danie Brink
Department of Genetics
Division of Aquaculture

Signature: _____

Date: _____

Acknowledgements

I would like to express my deepest gratitude to all those who were instrumental in the completion of this study with special thanks to the following people and companies:

Professor Danie Brink and Dr Rouvay Roodt-Wilding for my acceptance onto the Masters program and into their research group; for providing the educational and financial support needed to enable me to complete my studies and move towards a career in molecular genetics. For their advice, guidance and generous giving of their time over the last two years, I will always be extremely grateful. Thank you.

The Innovation Fund and the Industry Consortium, without whom, this opportunity to improve the genetic status of South African abalone would not have been possible. The support and hospitality we received at the farms made the tasks of regular sample collection, assessments and tagging, an efficient, rewarding and enjoyable process. Special thanks to: Louise Jansen, Adri Claasen, Stephen Ashlin, Lize Schoonbee, and Rudi Barnardo.

To Irvin and Johnson (Pty) Ltd for providing financial support in my first year of study and for the opportunity to spend a few occasions helping out at the farm and learning more regarding the practical aspects of abalone production. Special thanks to Nick Loubser and Lize Schoonbee for accommodating me in their busy schedules and for giving of their time and their experience.

Ruhan Slabbert, for sharing his knowledge and experience and for helping me to master the techniques required to complete this study with some measure of success; for his input into the running and interpretation of my results, his friendship and company on the cycle trail. I wish him all the best in his own studies going forward.

My friends and colleagues in the lab for being there throughout; for sharing their space, their suggestions, their sense of humour and the odd glass of wine on a Friday afternoon.

To Gloudi and Carel at the DNA Sequencing Facility; for your continuous support, knowledgeable suggestions, extreme patience and extra hours spent getting things through for me when time was tight. I appreciate both your assistance and your friendship. Thank you.

To my family, who have always encouraged me to follow my dreams; you have taught me to value, love and challenge myself continuously and to always reach further than I believed possible. I am eternally indebted to you for helping me become the person I am today. Thank you for your unconditional love; for encouraging me to be an independent thinker and to have the confidence to take on new pursuits which inspire me and give my life purpose. I love you always.

And last but not least, I wish to express my love, thanks and appreciation to my husband, Rob; without whose continuous support, understanding and belief in my abilities, I would never have had the courage to step back in order to move forward. For your encouragement, patience and unwavering love, I will always be grateful. Your sense of humour has seen me though the tough times and your love has made my life complete. Thank you for teaching me to see the beauty around me; to always expect more out of life but to give something back and to never ever give up. I love and respect you with all my heart.

"How inappropriate to call this planet Earth, when it is clearly Ocean"

Arthur C. Clarke

(Quoted in Beaumont and Hoare 2003)

Table of Contents

List of Figures

List of Tables

List of Appendices

Abbreviations, Acronyms, Formulae and Symbols

AA	Acrylamide
AbG	Abagold (Pty) Ltd
ABI	Applied Biosystems Incorporated
AFLP	Amplified Fragment Length Polymorphism
APS	Ammonium Persulphate
ATP	Adenine Tri-Phosphate
AMP	Ampicillin
BAA	Bisacrylamide
Kb	Kilo Base Pairs
BR	Black Rock Wild Abalone Population
BSA	Bovine Serum Albumin
CTAB	Cetyl Trimethyl Ammonium Bromide
dNTP	Dinucleotide Triphosphate
DNA	Deoxyribonucleic Acid
dsDNA	Double Stranded DNA
DTT	Di-Thio-Threitol
EDTA	Ethylene-Diamine-Tetra-Acetate
eSNP	Expressed Single Nucleotide Polymorphism
EST	Expressed Sequence Tag
EtBr	Ethidium Bromide
Excl1	Probability of excluding an unrelated individual when first parent is known
Excl2	Probability of excluding an unrelated individual when neither parent is known
F1	First Generation Offspring Individual
FCE	Feed Conversion Efficiency
FIASCO	Fast Isolation by AFLP of Sequences Containing Repeats
gDNA	Genomic DNA
GelDoc	Multi Genius Bio Imaging System
He	Expected Heterozygosity
Ho	Observed Heterozygosity
HWE	Hardy Weinberg Equilibrium
IDT	Integrated DNA Technologies
IF	Innovation Fund

IPTG	Isopropyl β-D-Thiogalactopyranoside
KCl	Potassium Chloride
LB	Luria Bertani Medium
LOD	Natural log of the overall likelihood ratio
MAS	Marker Assisted Selection
$MgCl_2$	Magnesium Chloride
*Mse*I	Restriction Enzyme
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
N	Number of individuals tested in a population
Na	Number of alleles for a given locus as calculated by CERVUS 2.0
NCBI	National Centre for Biotechnology Information
Ne	Effective Population Size
NGI	Next Generation Individual
Null Freq	Estimated Null allele frequency as calculated by CERVUS 2.0
PAA	Polyacrylamide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
PRS	Performance Recording Scheme
QTL	Quantitative Trait Loci
SDS	Sodium Dodecyl Sulphate
SSC	Sodium Citrate Solution
ssDNA	Single Stranded DNA
Taq	*Thermus aquaticus* DNA Polymerase Enzyme
TD-PCR	Touch Down Polymerase Chain Reaction
TE	Tris-Ethylene-Diamine-Tetra-Acetic Acid
TEMED	N' – Tetra-Methyl-Ethylene-Diamine
TBE	Tris-Borate-Ethylene-Diamine-Tetraacetic Acid
Tris-HCl	Tris-buffer with pH adjusted using hydrochloric acid
Ta	Optimum Annealing Temperature
Tm	Melting Temperature
UV	Ultra Violet
X-Gal	5-Bromo-4-Chloro-3-Indolyl-β-D-Thiogalactopyranoside

List of Unit Abbreviations

°C	Degrees Celsius
s	Seconds
U	Units
µg	Micrograms
µl	Microlitre
µg/µl	Micrograms per microlitre
µg/ml	Micrograms per millilitre
M	Molar
µM	Micromolar
mg	Milligrams
ml	Millilitres
mM	Millimolar
pmol	Picomoles
ng	Nanograms
ng/µl	Nanograms per microlitre
nM	Nanometres
v/v	Volume per volume
w/v	Weight per volume

Chapter One

Project Overview and Study Objectives

1.1 Introduction

Natural resources of South African abalone are under threat from a combination of environmental factors, intensive fishing pressures as well as illegal and highly organized poaching groups. An evaluation of these threats has lead to the speculation that commercial harvests of wild stocks will become unsustainable in the very near future (http://www.mcm-deat.gov.za/press/2007/10042007.html). This decline in natural resources and high international market prices has resulted in extensive interest in abalone culture in many countries, including South Africa, and currently over twenty species worldwide are either, at present, commercially cultivated (see Table 3 Chapter 2) or under investigation for market value and hence their potential for cultivation (Elliot 2000). With the world's supply from natural fishing being unable to meet the steadily increasing global demands, the abalone aquaculture industry both locally and internationally, has developed into a reliable year round source of high quality abalone products cultured to meet specific market needs (Oakes & Ponte 1996). Within a cultured environment, the lifecycle is a closed one and the main focus for farms has been on optimizing the surroundings of the animal i.e. culture techniques; nutrition and tank design (Elliot 2000).

Currently, two areas require the focus of the local abalone industry: Firstly, understanding and fully exploiting international market opportunities through establishing the traits or characteristics that the different markets (specific country to which the product is being exported) consider important or preferable; and secondly, the establishment of a genetic improvement program so as to exploit the biological potential of the commercial species *Haliotis midae* (Elliot 2000). For other aquaculture species within genetic improvement programmes, production gains of between 5% and 15% per generation have been recorded and computer modelling has shown that the implementation of marker assisted selection by the industry could lead to significantly higher rates of genetic gain than would be expected in response to phenotypic selection alone (Bentsen & Olesen 2002).

A significant amount of research on abalone aquaculture and genetic improvement programs is being conducted around the world culminating in the recent publications of linkage maps for Pacific abalone, *H. discus hannai* (Liu *et al*. 2006; Sekino & Hara 2006) and Blacklip abalone, *H. rubra* (Baranski *et al*. 2006).

For South African abalone farms to remain competitive and sustainable and in order to increase the quantity and quality of their supply to a highly competitive global market, the adoption of advanced management strategies and innovative biotechnologies, including genetic improvement programmes, has become necessary (Roodt-Wilding & Slabbert 2006). Table 1 illustrates the status of genetic improvement of the indigenous, commercially exploited abalone species, *Haliotis midae*, in comparison to aquaculture species as a whole and to other abalone species around the world.

Table 1 Summary of the status of genetic improvement of *Haliotis midae* in relation to aquaculture in general as well as to other abalone species around the world (Taken from: Project Proposal 2005: p7; Author: Prof. D. Brink)

	Fields of activity	Activity Levels		
		Aquaculture	**Abalone**	***H. midae***
a.	Domestication and strain evaluation	High	Low	Low
b.	Inbreeding and maintenance of genetic quality	High	Medium	Low
c.	Intraspecific crossbreeding	High	Medium	Low
d.	Interspecific hybridization	High	Medium	None
e.	Genetic selection	High	Medium	None
f.	Polyploidy	High	High	Low
g.	Sex manipulation	High	Low	None
h.	Gynogenesis, androgenesis and cloning	Medium	Low	None
i.	Gene transfer and genomics	Medium	Low	None
j.	Linkage mapping and quantitative trait loci	High	Medium	Low
k.	Marker-assisted selection	Medium	Low	None

Biological productivity needs to be addressed through the improvement of traits identified as economically important; such as feed conversion efficiencies, growth rates, disease resistance, meat to shell ratios and survival rates. Traits such as these show continuous variation as they are typically influenced by multiple genes as well as the environment and it is this natural variation, together with high fecundity and a closed life cycle, which makes abalone such an ideal candidate for genetic improvement (Bentsen & Olesen 2002).

1.2 Project Objectives

To achieve this output of a genetic enhancement program for South African abalone, a consortium was secured, consisting of the five leading companies within the South African abalone industry, together with scientists within the field of genetics. The primary products to be delivered out of this five year collaboration will be improved strains or genotypes of *Haliotis midae* to be utilized by the commercial hatcheries for increased delivery and profitability as a whole. Increased knowledge and understanding of the species and the different farm and wild populations; together with a growing pool of molecular markers to assist in the identification of superior genotypes and individuals, will move the South African abalone industry towards the implementation of marker assisted selection programmes and breeding schemes.

Within the next five years, the project as a whole aims to address the following core objectives, through delivery against concurrently running individual, objective-specific studies:

(1) To develop and implement innovative technologies directed towards the long term genetic improvement of the South African abalone species, *Haliotis midae* for competitive advantage within international markets.

(2) To produce novel/improved strains or genotypes with enhanced productivity thus improving the profitability of the industry as a whole and creating opportunities for economic development.

(3) To overcome specific scientific and technological barriers currently associated with the indigenous species: lack of knowledge pertaining to the genome of *Haliotis midae* and the absence of genetic characterisation; application of existing technologies while continuously evolving through the implementation of new and improved techniques; resolving issues related to biosafety and environmental risk assessments.

(4) To establish a collaborative network that will form the basis of a sustainable Aquatic Biotechnology Platform (ABP) for abalone aquaculture in South Africa.

(5) To develop human resources that will help secure the future development and application of key genetic and biotechnologies to advance aquaculture development in South Africa.

1.3 Study Objectives

This particular study aims to address the preliminary requirements of the project through the utilisation of a Performance Recording Scheme (PRS) to aid the identification of individuals with potentially superior genetic qualities. Current breeding stock and offspring for five abalone farms were evaluated using a set of nine molecular markers to assign parentage and assess breeding stock contributions or in other words, the spawning success of the breeders involved. Since the scope of this work was too large to be conducted singularly, it was divided into two separate yet linked studies of microsatellite marker development and parentage.

The ***primary objective*** of this study was to assign parentage to offspring produced by two of the participating farms: Abagold (Pty) Ltd and HIK Abalone (Pty) Ltd. The offspring were identified as either superior or inferior in terms of growth through an initial visual assessment followed by measurement of shell length and body weight as appraised against pre-determined size ranges. The assignment results that ensued were used to make recommendations regarding the inclusion or exclusion of the contributing breeding stock from future breeding programs.

The ***specific objectives*** designed to assist in the completion of the primary objective are indicated below and form part of the larger set of project objectives. The results and outcomes will aid in the delivery against the project's targets as a whole:

(1) To isolate and develop species-specific polymorphic microsatellite markers for the indigenous abalone species, *Haliotis midae*.

(2) To genotype all current breeding stock for Abagold and HIK as well as specific offspring selected from the PRS on the basis of growth performance; namely faster and slower growth groups.

(3) To assign parentage for offspring exhibiting either faster or slower growth rates for use in, or exclusion from, future MAS breeding programs.

A brief overview of the research design and methodology is outlined in the flow diagram (Figure 1). Each specific area of research is documented in a separate chapter, these following the introductory chapter and literature review. In turn, each chapter is structured to include the specific research methodology employed (presented as materials and methods), related results and discussion.

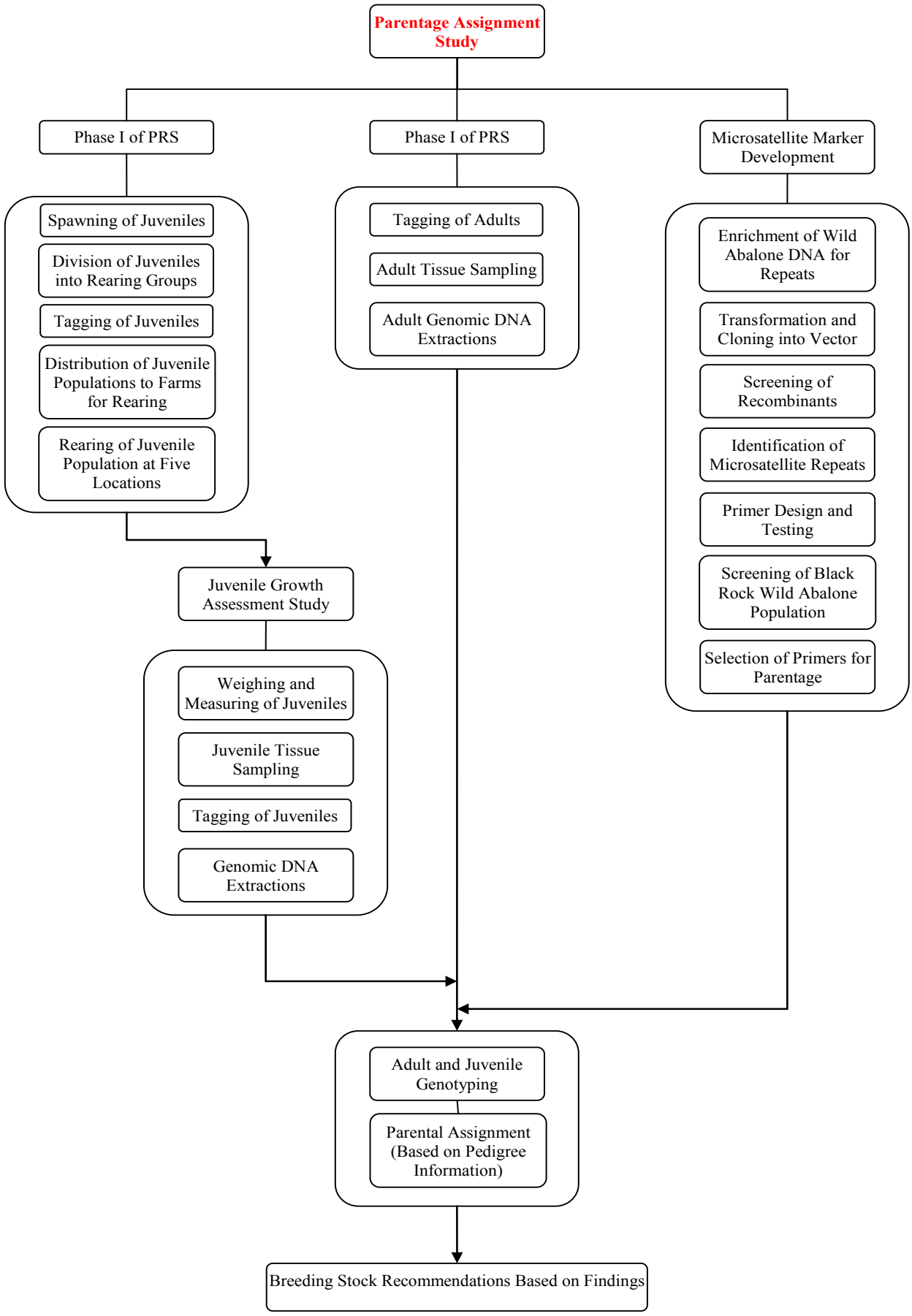


Figure 1 Abstract outline of the research design and methodology adopted in this study

Chapter Two

Literature Review

2.1 Background on Abalone

2.1.1 Classification

Abalone are classified as single-shelled (univalve) marine vetigastropod molluscs from the family Haliotidae with a worldwide distribution of fifty-six species in both temperate and tropical waters (Geiger 2000; Stevens 2003; Degnan *et al*. 2006). Table 2 provides a detailed classification of one of South Africa's endemic species, *Haliotis midae*, and the species under investigation within this study.

Table 2 Taxonomic classification of *Haliotis midae*
(Marine and Coastal Management/MCM [online], Hahn 1989)

Phylum	Mollusca
Class	Gastropoda
Subclass	Prosobranchia (winkles, whelks, limpets)
Order	Archaeogastropoda
Family	Haliotidae (ear shaped shells)
Genus	*Haliotis*
Species	*midae*
Common Name	Perlemoen, abalone

2.1.2 Biological Characteristics

Abalone are mass spawners and both male and female alike release their gametes into the surrounding water for external fertilization to occur. Few mechanisms (except temperature-related events) have been identified which may lead to synchronized spawning within the wild while spawning within a captive environment is artificially induced (Leighton 1989). Most temperate species have an annual reproductive cycle although the periodicity and duration of spawning was found to vary both intra- and inter-specifically (McShane 1992).

Fecundity for abalone may be high but at the same time so is mortality due to predation, larval viability and unfavourable end destinations on ocean currents (Leighton 1989); while further studies have revealed that large variations in longevity and growth rates exist between different species (Day & Fleming 1992).

2.1.3 Abalone in the Natural Environment

The majority of the world's temperate oceans are home to abalone, with the animals themselves being found along rocky substrates, reefs and crevices near to the shore, generally from sea level up to thirty metres. Depending on whether the species is temporally or sub-tropically distributed, water temperatures capable of sustaining abalone life, range anywhere from 2˚C to 30˚C. The species is characterized by a planktonic larval phase having highly specific requirements for settlement; while the adult phase is relatively sedentary with mature animals, in favourable circumstances, remaining in one place for some time (Leighton 1989; Degnan *et al*. 2006).

2.1.4 Abalone Feeding in the Wild and in Captivity

Abalone are nocturnal, slow-feeding herbivores within their natural environment. The adult diet consists predominantly of seaweeds while the juveniles feed on micro-algae and diatoms that are found attached to the surfaces onto which the abalone settles (Elliot 2000). In the past, the aquaculture of abalone was hindered somewhat by the limited choice of good quality, commercially manufactured diets but research is ongoing within this arena resulting in improved artificial feeds being continuously developed. Newly developed feeds aim to allow the farms to substitute up to 100% manufactured food while still improving growth rates, feed conversion efficiencies and ease of production (FishTech™Inc [online] 2007).

2.1.5 Global Distribution of Haliotids

Abalone are marine animals. Of the fifty-six identified species of Haliotids, eight of these are native to the north-eastern Pacific Ocean while the greatest variety within *Haliotis* is found in the south-western Pacific and Indian Oceans (Leighton 2000). The largest members of the genus *Haliotis* are located near the coasts of Southeast Asia, Australia, Japan, New Zealand, Africa and western North America; with wild and farmed abalone from Mexico, Australia, Taiwan, Japan, China and New Zealand accounting for the majority of supply to meet the world's volume demands (Stevens 2003).

Abalone can be distributed regionally from sea level to a depth of approximately 30 metres and there are no Haliotids that have a global distribution (Degnan *et al*. 2006). Geiger (2000) identified four distinct endemic areas: South Africa (five endemics), Australia (six endemics), New Zealand (three endemics) and North America (six endemics along the western coast) with this broad pattern being confirmed by molecular phylogenetics of selected species (Degnan *et al*. 2006). The twelve

species that are commercially cultured for the international market are presented in Table 3 with America and Australia having several species that are currently under commercial cultivation.

Table 3 Twelve commercially cultivated Haliotid species occur throughout the world and are listed with the species names (scientific and common) and origins (Stevens 2003)

Species Name	Common Name	Location
Haliotis rufescens	Red abalone	North America & Chile
Haliotis cracherodii	Black abalone	North America
Haliotis fulgens	Green abalone	North America
Haliotis corrugata	Pink abalone	North America
Haliotis kamtschatkana	Pinto abalone	North America
Haliotis midae	Perlemoen	South Africa
Haliotis laevigata	Greenlip abalone	South Australia
Haliotis rubra	Blacklip abalone	South Australia
Haliotis roei	Roe's abalone black-footed	Australia
Haliotis iris	Paua	New Zealand
Haliotis diversicolor supertexta	Small abalone	Taiwan
Haliotis discus hannai	Disk abalone	Japan & China

2.2 *Haliotis midae* (South African Abalone)

2.2.1 Natural Populations

Five recognised endemic abalone species are distributed along the South African coastline (Figure 2) only one of which is currently commercially utilised. The most widespread species are *Haliotis midae* (perlemoen), *H. parva* and *H. spadicea* (siffie) while *H. queketti* and *H. speciosa* are only found along the east coast of South Africa (Geiger 2000). *Haliotis midae* is the most abundant species in local waters and due to its large adult size it is the only species that is suitable for commercial cultivation (Roodt-Wilding & Slabbert 2006).

Between Cape Hangklip and Danger Point on the south west coast, a strong link has been observed between juvenile abalone of the species *Haliotis midae* and the sea urchin *Parechinus angulosus* (Tarr *et al*. 1996; Mayfield & Branch 2000). Tegner (1989) maintains that the juvenile abalone gain protection from predation by sheltering underneath the urchins and may benefit from access to extra food harvested by the sea urchins. The collapse of urchin numbers through lobster

predation and hence the destruction of this beneficial relationship and protective co-habitat has been implicated in the reduced number of juvenile abalone observed within this zone (Tarr *et al*. 1996).

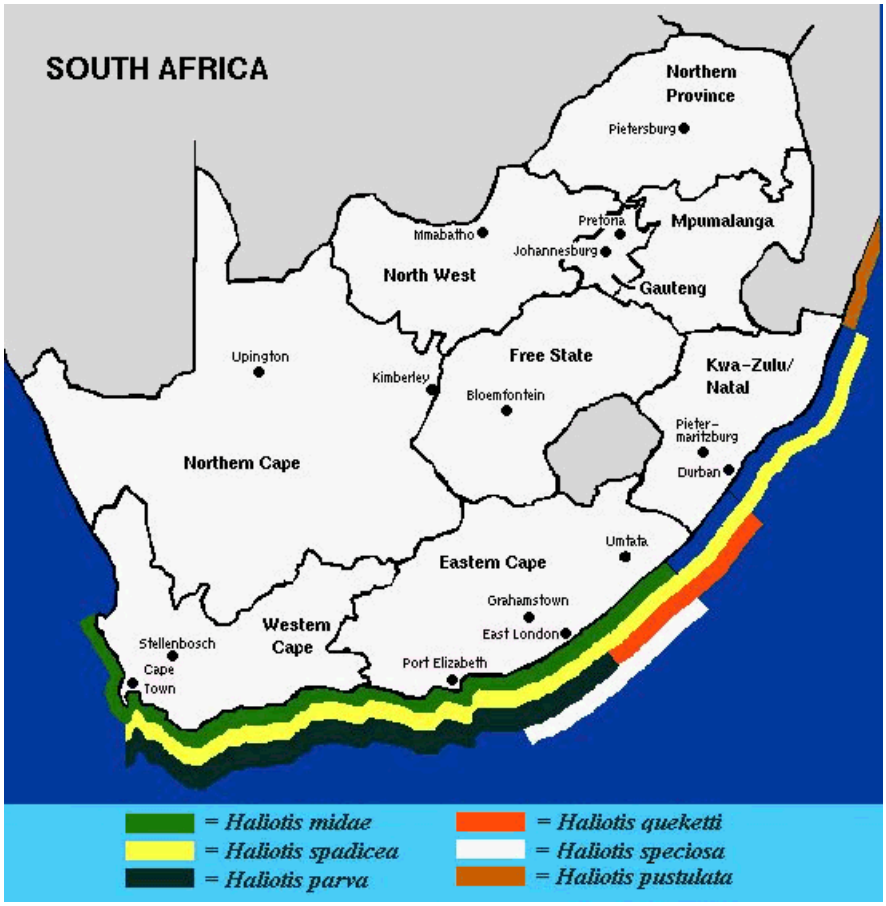


Figure 2 Distribution of six abalone species along the South African coast, only five of which are recognized as endemic to South Africa. *Haliotis pustulata* inhabits the Mozambique coastline (Taken from: http://web.uct.ac.za/depts/zoology/abnet/safrica.html)

2.2.2 Commercial Threats and Predation

The threats to wild abalone populations and their long term sustainability are various and often specific to coastlines. Among these are: commercial over-fishing, recreational catches and exploitation of legal bag limits, pollution of habitat, disease and inadequate wild stock management (Stevens 2003).

For *Haliotis midae* specifically, the impact of rock lobster predation and their destruction of the habitats (anemone or sea urchins) on which the larvae settle has been extensive as has the illegal harvesting of this species by large, well-organized poaching syndicates. The latter activity is driven by the high international market prices making illegal trade a lucrative business (Coastal and Marine Life online). Figure 3 illustrates the decline in natural abalone over 100mm in length to be

seen along the coast of South Africa where legal harvesting is still permitted. Transects of approximately 60 metres were measured within two to three metres from the shore.

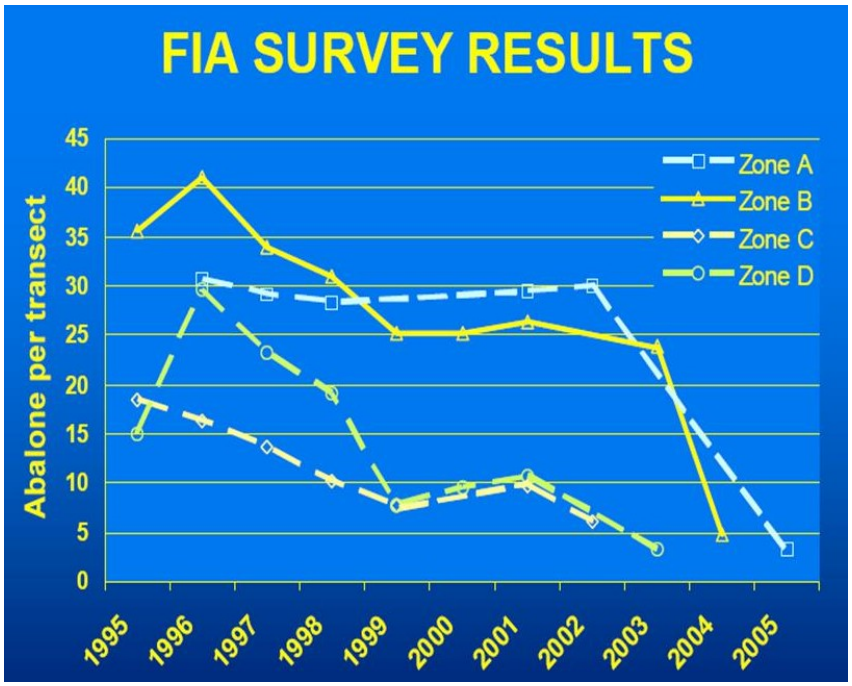


Figure 3 FIA Survey conducted in 2005 reveals the declines in natural abalone over 100mm in length that can be found across Zones A-D of the South African coast. (Taken from: http://www.abalone.cenrm.uwa.edu.au/__data/page/1963/A_brief_overview_of_the_abalone_industry_in_South_Africa5.pdf)

2.2.3 Legal Catch and Policy Control

In an effort to control the exploitation of natural abalone populations, legal catch was allocated to rights holders, while areas along the coast of South Africa were zoned into legal catch subareas thus further sub-dividing the total allowed catch. A small area of approximately 6.5km that lies between Cape Hangklip and Danger Point remains a protected marine reserve within which the harvesting of abalone and other bethnic species is prohibited (Mayfield & Branch 2000). The total allowable catch of abalone declined from approximately 400 tons in 2001/2002 to about 223 tons in 2004/2005, a clear indication of the over harvesting of this valuable resource (MCM online). These legal catch zones are illustrated in Figure 4. The harvesting of natural populations will no longer be permitted for either commercial or recreational purposes from 01 February 2008; with these bans being instituted in an effort to control the illegal harvesting of abalone by syndicated poaching groups and to protect the rapidly declining natural populations (Press Release: Environmental Affairs and Tourism 2007, http://www.engineeringnews.co.za/article.php?a_id=119867). A delay in instituting the ban has been permitted so as to allow all stakeholders adequate opportunity to

prepare for the implementation of the policy and the unavoidable impact thereof on individual and corporate businesses.

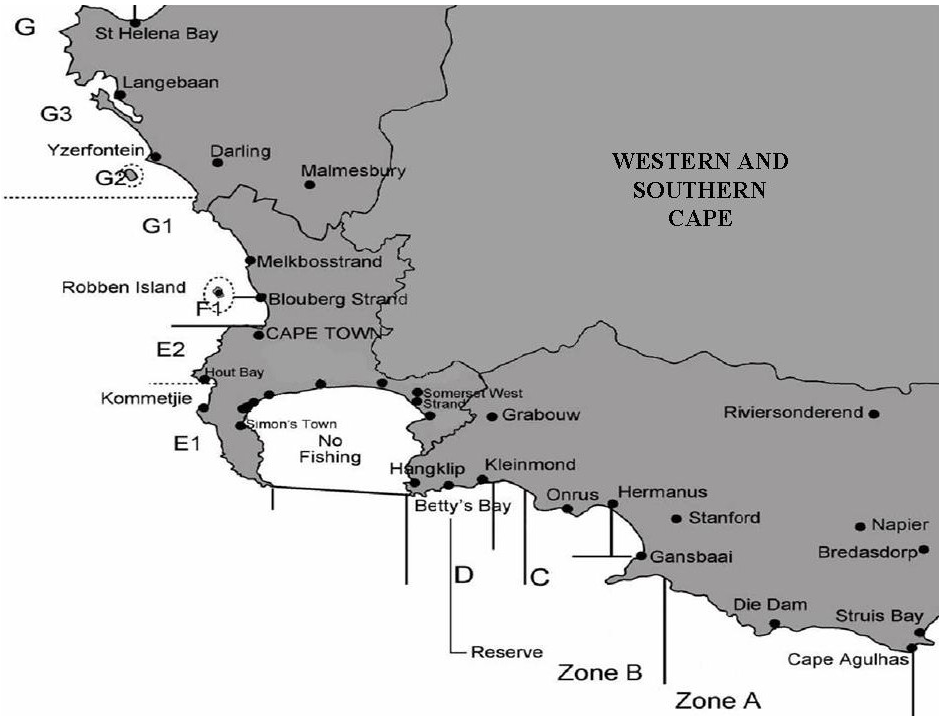


Figure 4 Distribution of legal catch sub-area allocations along the coast of South Africa (Taken from: http://www.abalone.cenrm.uwa.edu.au/__data/page/1963/A_brief_overview_of_the_abalone_industry_in_South_Africa5.pdf)

2.3 A Snapshot of the International and the Local Abalone Market

2.3.1 World Market Opportunities and Commercial Value

To date, twelve species have been identified to have traits of economic importance and/or market value and are currently commercially exploited (Table 3). From 1975 where a balance existed between supply and demand to predicted volumes for 2004, a shortfall in the supply to a niche market with a steady demand was anticipated (Figure 5). This imbalance in supply/demand persists and is one of the primary drivers of the increase in abalone aquaculture seen today. The strong world demand for this delicacy, however, does not automatically mean a premium price but remains very dependent on commercial producers to compete successfully within a highly dynamic international market (Gordon & Cook 2001). Understanding the world market and cultural traditions which determine product quality and desirable traits, is key to identifying the market standards that aquaculture producers need to satisfy (Oakes & Ponte 1996).

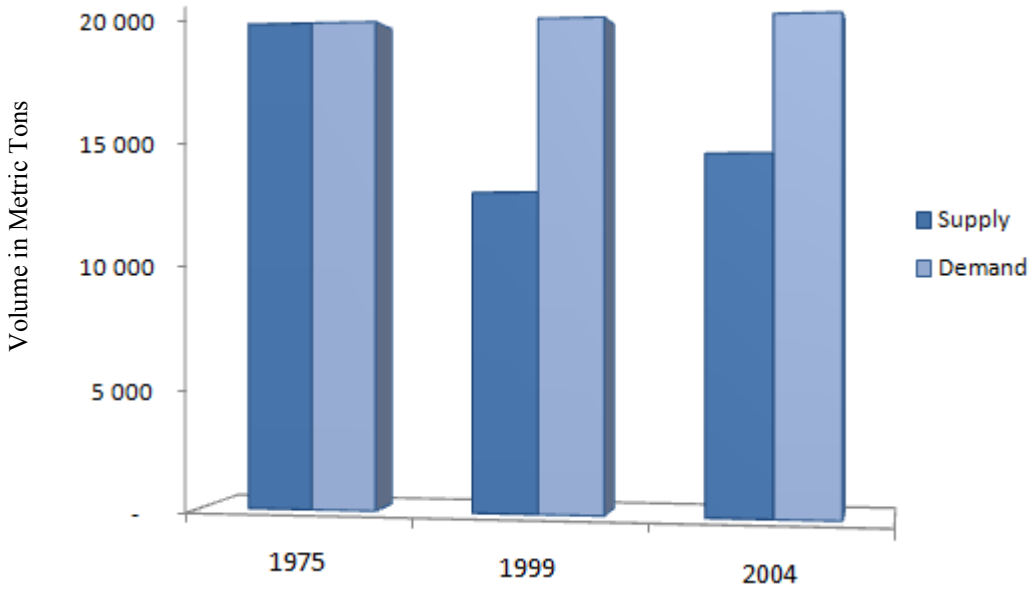


Figure 5 World market supply and demand of abalone from 1975 to predicted volumes in 2004. Cultured species *H. supertexta* is not included in "supply" (5000mt 1998, 6000mt 2004) (Taken from: Gordon & Cook 2001)

The major consumer countries of abalone are Japan and China (including South East Asia), while established markets also exist in the United States, Mexico, Europe and Korea. Live, fresh and frozen abalone are generally identified as those product forms having the highest commercial value on the world market. Appearance, colour, taste and texture are all important considerations in determining the value of the end product while within most premium markets, consistency of supply is weighted as importantly as the quality of product (i.e. taste and appearance) (Oakes & Ponte 1996).

2.3.2 Abalone Production and Marketing in Southern Africa

The 1980's saw the initiation of abalone farming in South Africa, with the first farms being constructed in 1992 and culminating in the thirteen land-based abalone farms that are currently located throughout the country (Loubser 2005). These farms operate almost exclusively using pump ashore systems, while two of the farms make use of additional re-circulation systems and a third one employs a partial re-circulation system in their operation (Cook 1998; MCM online).

Local abalone producers employ different feeding strategies for their cultured animals, with most of the farms using a mixture of harvested kelp (*Ecklonia maxima*); an artificial food source formulated as a dry pellet (Abfeed); and cultivated seaweed (*Gracilaria*). Two producers feed their animals solely on a diet of kelp; three farms feed a combination of kelp and cultivated seaweeds; while yet another producer substitutes 100% manufactured feed (MCM online).

The success of the South African abalone industry can be contributed in part to the availability of excellent sites with good water quality and partly to exclusive access to the internationally well accepted species, *Haliotis midae*. A joint international promotional effort is being considered by local abalone producers with the aim of making a firm stand against the strong Australian and Chilean competition (Loubser 2005). Figure 6 shows the distribution of the thirteen abalone aquaculture farms currently operating in South Africa.

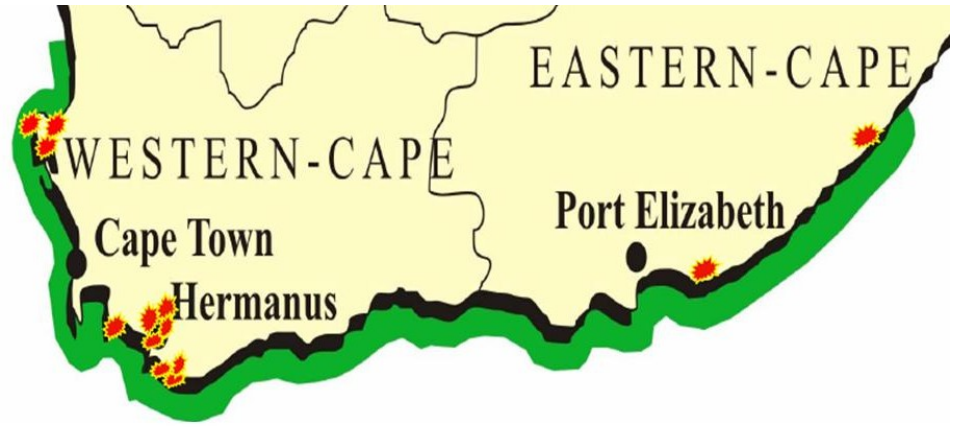


Figure 6 Thirteen abalone producers are distributed along the South African coast. The majority of these farms are located along the West Coast while two are situated in the Eastern Cape (Taken from http://www.abalone.cenrm.uwa.edu.au/__data/page/1963/A_brief_overview_of_the_abalone_industry_in_South_Africa5.pdf)

Industry production in South Africa has increased steadily over the last five years culminating in an annual volume of approximately 900 tons (whole animal in shell) with an export value of US$25 million in 2006 (van Schalkwyk 2007). A projected volume of close to 1200 tons is targeted for 2010. Some of the larger farms have already reached an annual output of 100 tons or greater while the contribution of the smaller farms is in the region of 50-60 tons per annum (Loubser 2005).

Abalone products are marketed in a variety of forms to an up-scale international market. High international prices allow producers to recoup the costs of an expensive long term production investment and remain a profitable venture (McBride & Conte 1996). Currently, 62% of farmed animals are sold live, 33% in a canned format and 5% in other formats such as frozen meat; and in general, the demand for this valuable commodity continues to exceed supply (Gordon & Cook 2001). Major international markets for cultured abalone are: Japan, Hong Kong (and South East Asia), China, Taiwan, Mexico, Korea, Europe and the United States (Oakes & Ponte 1996).

Figure 7 illustrates the annual and projected growth (in tons) of abalone production throughout South Africa.

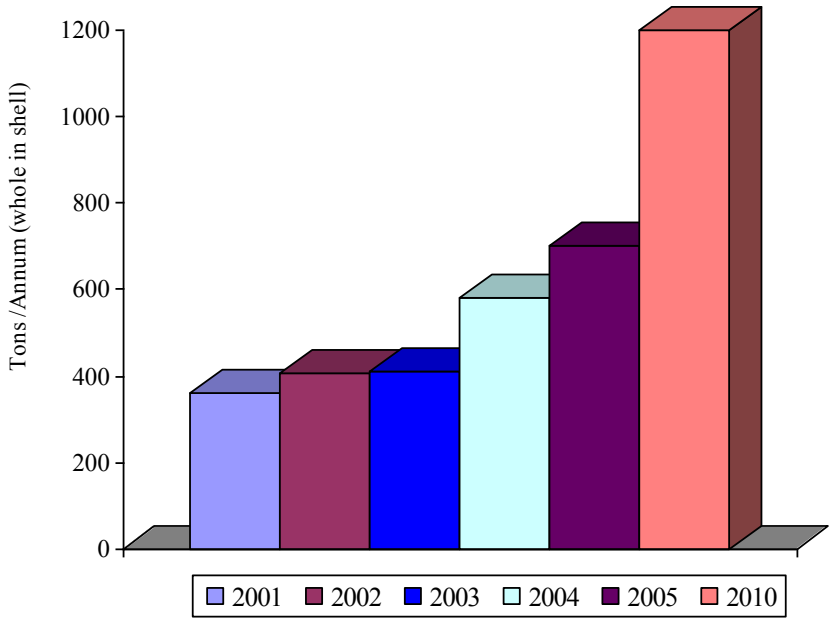


Figure 7 Annual volume growth in tons per annum of farm produced abalone in South Africa with a predicted volume of twelve hundred tons in 2010 (Taken from: Loubser 2005)

2.4 Gaining a Competitive Advantage in a Dynamic International Market

2.4.1 Genetic Improvement

Genetic improvement can be defined as "a gain in the cultured production of an abalone species through the exploitation or manipulation of the genetic variation present within the particular species" (Elliot 2000). The exploitation of inherent genetic variation can contribute in several ways to improve abalone: Firstly, through the improvement of commercially and economically important traits such as growth rates, faster growing individuals can be produced and ultimately production times are reduced thus affecting the end cost of market sized animals. This is a vital deliverable in an arena where slow growth rates are considered a hindrance to aquaculture potential.

Secondly, manipulation of genetic variation through controlled breeding can assist in the production of superior strains of abalone better suited to particular grow-out environments, more disease resistant, with higher survival rates or feed conversion efficiencies. These enhanced genotypes may have desirable market traits, better meat to shell ratios, higher fecundity, lower age at maturity and sterility or single sex production, among a whole host of other favourable characteristics (Elliot 2000).

2.4.2 Commercially Important Traits

The ultimate aim of a genetic improvement program is to improve the overall profitability of an industry or company. Commercially valuable traits, although definable, are not always measurable;

therefore, traits that are known to have some genetic relationship with the commercial trait of interest are used as selection criteria. These selection traits will be used to measure the outputs or gains of the genetic improvement program and therefore need to fulfil certain criteria: they need to be inexpensive to measure, they need to be heritable and most importantly, they must be genetically linked to the trait of interest (Anne 2006).

2.4.3 The Potential of Genetic Improvement Programs within Aquaculture

Biotechnology is defined as: "Any technique that uses living organisms to make or modify a product, to improve plants or animals or to develop microorganisms for specific uses" (US Congress Office of Technology Assessment as cited in Diaz & Neira 2005). Biotechnology has been one of the most significant contributors to the advancement seen in aquaculture over the last five decades; particularly in areas such as reproduction, pathology, nutrition and health and the genetic improvement of cultured species (Diaz & Neira 2005). Genetic improvement at the hand of biotechnology, however, should not be seen as a stand-alone answer to increasing the performance of a species but should be addressed together with other practical aspects of marketing, nutrition, health and husbandry practices (Holst 1999).

Genetic enhancement programs within aquaculture have the potential to deliver collectively and sustainably to the worldwide requirements of increased food production and food quality; the unearthing and the progression of new natural resources; as well as awareness and maintenance of both declining biodiversity and a severely challenged environment at the hands of modern day society (Davis & Hetzel 2000; Melamed *et al*. 2002). Beyond these deliverables, financial gain for the industries and aquaculture companies is a key requirement if delivery is to be sustainable and motivating (Davis & Hetzel 2000).

2.4.4 The Key to Unlock What Lies Beneath

Genomics is loosely defined as the "study of genes and their functions" and is an area of research that impacts on numerous avenues of everyday life ranging from reproduction, food safety and health care to law enforcement to name a few (Anne 2006). With the coming together of three key components, the power held within molecular data can today be accessed and otherwise unobtainable information revealed and studied: Firstly, the availability of efficient molecular techniques for examining specific areas of DNA; secondly, ongoing development of appropriate statistical methods and models for analyzing DNA data and thirdly, the design and availability of user-friendly computer packages (Tarr 2000).

2.5 Molecular Markers - Tools to Deliver Genetic Improvement Programs

2.5.1 Genetic Variation and DNA Molecular Markers in General

Genetic variation exists within and among individuals, species and higher taxonomic groups as a result of a combination of factors: mutations, interactions with the environment, fitness selection and genetic drift. Various types of mutations accumulate over time and the degree and number of mutations that exist, serve to define the variation or polymorphism within a given species. For genetic variation to be useful, it needs to be inheritable and discernable either phenotypically or genetically through molecular techniques (Jones *et al*. 2001). Genetic diversity can be estimated through the application of various molecular techniques and to date a series of different genetic markers have been explored and developed (Sekino & Hara 2006). Molecular genetic markers are comprised of two elements: the inheritable DNA region to be screened (locus) and the technique to be used to reveal the variation (alleles) inherent at each region (Sunnucks 2000).

Varying degrees of variation are required depending on the study being completed and this variation within a population can be measured through an assessment of gene diversity and allelic richness (Aho *et al*. 2006; Anne 2006). Intra-population studies such as parentage analyses require high to very high levels of variation at the DNA level. Choosing the right molecular marker for a study relies to a large extent on our understanding of the evolutionary forces that either create or remove variability; for example, repetitive sequences such as microsatellite loci, are highly mutable and provide highly variable markers (Anne 2006).

2.5.2 Choosing the Right Molecular Marker and Technique

The choice of genetic marker depends on several criteria: firstly, the level of variability of the DNA region being investigated; secondly, the nature of the information (e.g. dominance, codominance, ploidy etc) that needs to be determined in order to answer the biological question being posed and thirdly, practical aspects such as which molecular techniques are feasible given the laboratory equipment and facilities available as well as the level of experience of the scientists involved (Sunnucks 2000).

Techniques for molecular marker analysis describe the various methods employed in the detection of variation at DNA level and encompass processes such as: initial template production, PCR and electrophoresis; through to sequencing and genotyping if the study requires this level of detail. Several techniques are currently available for analysing different molecular markers; with no single molecular technique being universally ideal and each of those currently available possessing both strengths and weaknesses. Ultimately, the choice of technique relies on a compromise between the same criteria considered when choosing the type of molecular marker most applicable to an area

of study: the research question to be resolved; the genetic resolution or depth of information required; and finally, any potential financial constraints (Anne 2006).

2.5.3 Types and Usefulness of Molecular Markers

Not all molecular genetic markers are equally suited to provide accurate biological inferences concerning the biological question of interest; and an understanding of the level of variation required and the nature of the information generated by the molecular marker is crucial to selecting the most appropriate marker for particular applications (Anne 2006). In saying this, however, all genetic markers are able to reveal variation at DNA level, although there is often a trade-off between precision and convenience (Tarr 2000).

One of the most useful measures of a molecular genetic marker is its Polymorphic Information Content (PIC value) which is an indicator of the ability of a marker to detect polymorphism in a population. PIC values are dependent on the number of alleles detected as well as their frequency distributions. The more alleles detected by a single molecular marker or the higher the frequency of a given allele revealed by a molecular marker, the higher its PIC value. Researchers can compare PIC values for different genetic molecular markers as a means to determine the power of different types of markers to address their specific research question (Liu & Cordes 2004).

Within aquaculture genetics several different types of markers have been employed in the past. The basic properties of these are indicated in Table 4.

Table 4 Basic properties of molecular markers and their major applications (Liu & Cordes 2004)

Name	Alias	Inheritance	Type	No. Locus	Allele No.	Polymorphism	Major applications
Allozyme		Mendelian codominant	Type I	Single	2 – 6	Low	Linkage maps, population studies
Mitochondrial DNA	mtDNA	Maternal inheritance	-	Multiple Haplotypes			Maternal lineage
Restriction fragment length polymorphism	RFLP	Mendelian codominant	Type I or Type II	Single	2	Low	Linkage maps
Random amplified polymorphic DNA	RAPD, AP-PCR	Mendelian dominant	Type II	Multiple	2	Intermediate	Fingerprinting in population studies, hybrid identific.
Amplified fragment length polymorphism	AFLP	Mendelian dominant	Type II	Multiple	2	High	Linkage maps, population studies
Microsatellites	SSR	Mendelian codominant	Mostly Type II	Single	Multiple	High	Linkage maps, population studies, paternity analysis
Expressed sequence tags	EST	Mendelian codominant	Type I	Single	2	Low	Linkage maps, physical maps, comparative maps
Single nucleotide polymorphism	SNP	Mendelian codominant	Type I or Type II	Single	2, but up to 4	High	Linkage maps, population studies
Insertions/deletions	Indels	Mendelian codominant	Type I or Type II	Single	2	Low	Linkage maps

Molecular markers can be categorized into two types: Type I markers are associated with genes of known function, while Type II markers are associated with unknown genomic sequences (Liu & Cordes 2004). Type II genetic markers are able to provide information in one of two ways: through a molecular phenotype as evidenced by the presence/absence of given fragments (dominant) or by means of a genotype wherein both alleles of a diploid individual are revealed (codominant) (Anne 2006).

2.5.4 Pros and Cons of Molecular Markers

The first tools used to assess diversity and differentiation were morphological traits, followed by allozymes (protein products of genes) and at the DNA level, restriction fragment length polymorphism (RFLP). These methods have, however, since been surpassed by newer, PCR driven techniques (Gaudeul *et al*. 2004). Liu and Furnier (1993: as cited in Huang *et al*. 2000) established that although allozymes analysis is the most cost-effective method, it is very often less informative than DNA based methods.

Maternally inherited mitochondrial DNA (mtDNA) was and is still used extensively in the analysis of stock structure and population differentiation but the molecule is considered a single locus due to its non-Mendelian mode of inheritance. PIC values for this marker, although higher than for allozymes are lower than for RAPD, AFLP, microsatellite repeats (SSR) and Single Nucleotide Polymorphisms (SNPs). mtDNA is also subject to the same problems, such as back mutations, that affect other DNA-based markers (Liu & Cordes 2004; Gilbey *et al*. 2005).

RFLP markers on the other hand, are codominant markers that are relatively easy to score; however, the low level of polymorphism observed for RFLP and the difficulties in developing such markers from a poorly understood genome or one where molecular information is limited, makes these DNA markers a less than ideal choice for many applications (Liu & Cordes 2004).

RAPD analysis is a fairly simple test to perform using dominant Mendelian markers, has high ability to detect polymorphisms and PCR primers can be designed without prior knowledge of DNA sequences. On the down side for these loci, allele scoring can be time-consuming, tricky and difficult to reproduce; while PIC values fall below those obtained for AFLP, microsatellites and SNPs, indicating that these molecular markers may not be as informative as the latter three options (Huang *et al*. 2000; Liu & Cordes 2004).

Mueller and Wolfenbarger (1999) advocate that the difficulty in the precise identification of alleles using AFLP-PCR is more than offset by the fact that this is a fast, reliable, cost-effective PCR based method with which to generate a large number of high resolution markers. As a result, AFLP has been found useful in biodiversity surveys, population and conservation genetics,

fingerprinting and kinship analyses and QTL mapping. These authors suggest that a combination of dominant AFLP and codominant microsatellite marker information, together with sequencing information, are likely to become the primary tools in the study of genetic variation (Mueller & Wolfenbarger 1999). In a parentage study by Gerber *et al*. (2000), AFLP was also shown to provide encouraging results, with the conclusion that 100 to 200 dominant loci (AFLP) or less than 10 highly polymorphic codominant loci (microsatellites) would be adequate to assign parentage. Development costs would be a driving factor in the decision as to which marker to select for a study, as would the biological question under consideration. Microsatellite development requires cloning, sequencing, identification of repeats followed by primer design and testing. It is therefore time consuming, costly and shown to be very species-specific but it presents the advantages of multi-allelic and codominant markers which are numerous throughout the genome. AFLP, in contrast, is more transferable across species and has lower initial development costs; however, lower informativeness requires that more AFLP markers are required to achieve a comparable level of assignment when employed in parentage analyses (Gerber *et al*. 2000; Gaudeul *et al*. 2004). In an individual-based population assignment study involving whitefish, *Coregonus clupeaformis*, Campbell *et al*. (2003) show that within a weakly structured population, AFLP provides a suitable alternative to the logistical issue of screening the large numbers of microsatellite loci required to provide enough statistical power to be considered useful.

In contrast to applications in parentage analyses, Bernatchez and Duchesne (2000) reveal that within population assignments, increasing the number of loci (dominant AFLP), is considerably more important that increasing the number of alleles (codominant microsatellites). For research projects that aim at large-scale understanding and description of the population structure, the long term benefits of microsatellite development would need to be considered over the immediate cost benefits of AFLP development (Campbell *et al*. 2003).

Microsatellite markers, however, are still considered the marker of choice when these loci are available and are generally preferred when investigating genetic structure. AFLP are mostly employed when quick, reliable, large scale, cost-effective surveys are required. However, a combination of different kinds of molecular markers, when practical, will always prove to be the most enlightening; especially when each exhibit different characteristics in the form of mutation rates, segregation or response to evolutionary influences (Gaudeul *et al*. 2004).

SNPs, a more recent development, describe different alleles containing point mutations at a given base within a locus and are becoming a focus within molecular marker development for several reasons: (1) they are the most frequently occurring polymorphism within any organism; (2) SNPs have significantly lower mutation rates than microsatellites and are hence considered stable

molecular markers, an essential criteria in evolutionary studies, population biology and pedigree studies such as parentage analyses; (3) unlike microsatellites, which occur primarily in non-coding sequences within the genome, SNPs are often found in coding regions thereby having the potential to affect protein function and expression and consequently are subject to selection forces; (4) they are amenable to automation and can often reveal hidden polymorphisms undetectable through other methods (Li *et al*. 2004; Morin *et al*. 2004; *Rengmark et al*. 2005). Both SNPs and microsatellites have wide scale prospective applications for the long term management of farmed and wild aquaculture populations (Rengmark *et al*. 2005).

2.5.5 Application of Molecular Markers in Aquaculture

Rapid advances in the application of molecular genetics in aquaculture investigations has allowed for studies into genetic variation, ongoing monitoring and control of inbreeding, parentage assignment studies, strain and species differentiation, hybridisation and cross-breeding of species (Davis & Hetzel 2000; Jones *et al*. 2001; Anne 2006). The full spectrum of currently available molecular markers have been widely utilised and assessed through a variety of studies for the aquaculture industry. These technologies encompass the following: allozymes, RFLP, mDNA, RAPD, AFLP, microsatellite markers and more recently SNPs (Liu & Cordes 2004).

Well-designed studies using these markers singularly or in combination, have allowed for a greater understanding of various aquaculture species; culminating in the construction of high resolution genetic linkage maps for some (Coimbra *et al*. 2003; Baranski *et al*. 2006; Liu *et al*. 2006; Sekino & Hara 2006). The rapidly expanding wealth of information within the field of aquaculture will further hasten the identification of quantitative genes involved in controlling trait loci and the subsequent application of the QTL in marker assisted selection and genomics (Liu & Cordes 2004).

However, despite all the recent advances, the widespread application of genetic typing as routine within management practices is still restricted by cost, technical complexity and speed of existing methods. The use of more than one type of genetic marker in combination or the prudent selection of the appropriate marker (type and minimum requirement) to answer the biological question of interest may help to overcome some of these restrictions; producing fast, accurate and robust results (Gilbey *et al*. 2005).

2.6 Microsatellites - The Molecular Marker of Choice?

2.6.1 Microsatellite Markers and their Distribution in the Genome

Microsatellites are simple sequence repeat loci, consisting of two to six nucleotides repeated in tandem and they represent codominant genetic loci; meaning that heterozygous genotypes are distinguishable from homozygous genotypes (Chistiakov *et al*. 2005; Webster & Reichart 2005). Microsatellite markers are found with relative abundance throughout the entire genome of almost all living organisms (Sekino & Hara 2006). Lagercrantz *et al*. (1993) and Ross *et al*. (2003) found that the frequency, mutation rate and motif of these repeat sequences varies considerably across genomes of different species; with recent evidence indicating that their distribution may, indeed, not be random. Additionally, Kruglyak *et al*. (1998: as cited in Cruz *et al*. 2005) found that an inverse relationship exists between relative abundance and repeat motif; most likely explained by the higher rates of slippage experienced by shorter motifs (mono and dinucleotides) when compared to longer ones.

Microsatellites, consisting of the dinucleotide repeat (CA)n, are the most frequently found in most genomes with high levels of heterozygosity evident. Trinucleotide and tetranucleotide occur less frequently and with lower heterozygosity levels (Mueller & Wolfenbarger 1999; Chambers & MacAvoy al. 2000). According to Primmer *et al*. (1997), there appears to be a propensity for the density and length of microsatellites to increase as genome size increases whereas in contrast, Deka *et al*. (1999) suggest that the genomic frequency of microsatellites is most probably not directly related to genome size but rather a result of species-specific recombination rates and/or mutation or repair rates acting within the species under investigation.

It has been further suggested that the number of recombinant clones recovered from species-specific genomic libraries is an indirect indication of the microsatellite richness within the species as was observed in related species of bivalve molluscs (Cruz *et al*. 2005).

2.6.2 The Importance and Application of Microsatellite Markers in Genetics

The increasing interest in microsatellite markers as the marker of choice in applications across many different fields is motivated by their high degree of polymorphism and relative ease of scoring. One of the major drawbacks, however, is the need to isolate *de novo* from species in the initial stages of examination (Zane *et al*. 2002).

The size differences between alleles as a result of a varying number of repeat units at a given microsatellite locus, is the basis for this molecular markers polymorphism (Lui & Cordes 2004). The jury is still out as to the origin of these length differences but two likely models are presented. The first model by Schlötterer and Tautz (1992) suggests that microsatellites evolve through

frameshift mutations resulting from slippage events during DNA replication or repair; while Richards and Sutherland (1994) present a scenario of unequal recombination. The former model is likely to represent the principal mutational mechanism for this particular molecular marker.

Microsatellite markers can, in the majority, be classified as Type II markers unless, like Expressed Sequence Tags (EST) or expressed Single Nucleotide Polymorphisms (eSNP), they are associated with genes of known function (Jones *et al*. 2001). The establishment of tight linkage of a marker to a gene of interest and from which it is then possible to infer the presence of the gene from the presence of the marker, forms the basis of marker assisted selection (MAS). Breeding programs can then be used to track the inheritance of this gene and the availability of a clearly identifiable and quantifiable phenotypic trait/s associated with this same gene and marker is crucial in the successful implementation of MAS (Chistiakov *et al*. 2005).

To date, these loci have been adopted for a variety of biological applications: as inheritable genetic markers in breeding stock selection, characterisation of breeding stocks through parentage and genotyping analyses, breeding behaviour and population studies, DNA profiling in forensic investigations and the construction of high density linkage maps mapping economically valuable traits (QTL) for application in marker assisted breeding, to name but a few (Chambers & MacAvoy. 2000; Chistiakov *et al*. 2005). Ultimately, the effective use of microsatellite markers relies on good statistical theories and associated methods for the analysis of the data they produce (Chambers & MacAvoy 2000).

2.6.3 Microsatellite Multiplexing: Increased Productivity - Reduced Genotyping

The implementation of multiplex PCR, whereby the amplification of several loci within a single PCR reaction is carried out, is revolutionising DNA laboratories all over the world. Microsatellite multiplexing is a commanding technique that will allow productivity of genetic studies to be increased while simultaneously reducing costs (Neff *et al*. 2000a; Buckleton & Triggs 2006; Wesmajervi *et al*. 2006). An optimised multiplex protocol can provide high fidelity and band resolution and several protocols have been published for both fluorescently labelled primers using automated detection methods as well as for radioisotope labelling (Neff *et al*. 2000a; Wesmajervi 2006). Yet despite the clear advantages, many laboratories continue to use single PCR reactions. This is likely due to the lack of suitable automation required to reduce the complexity of analyses as well as concern regarding the consistency of genotyping output (Neff *et al*. 2000a).

As the number of available microsatellite markers rises and the number of loci pooled into one reaction expands, so does the flexibility to create various combinations of markers. However, as a result, the chance increases that some loci occur on the same chromosome, with these loci being

sufficiently close in some instances to complicate parentage and pedigree analyses (Buckleton & Triggs 2006). Linkage mapping will prove useful in distinguishing which loci may be linked and proposed multiplexes adjusted accordingly.

An extension to current multiplexing strategies has been reported where the DNA of all the individuals under investigation within a populations are pooled and simultaneously screened with primers flanking a microsatellite locus. Only one PCR is employed thereby significantly reducing the amount of genotyping required. This technique of "selective sweeping" has been proposed as an approach to identify genes involved in adaption and the result is a complex pattern of peaks faithfully reflecting the combined pattern of individual alleles (Thomas *et al*. 2007) Schlötterer (2003: as cited in Thomas *et al*. 2007) claims that this systematic scanning could allow population-specific and locus-specific loss of variation to be traced.

2.7 Genotyping Errors and Artefacts Associated with Microsatellite PCR

2.7.1 Common Sources of Error

Microsatellite genotyping errors will occur in all except the smallest of data sets; and their potential to weaken downstream analyses reinforces the need to make allowance for these errors within analysis techniques and software programs (Hoffman & Amos 2005). In reality, genotyping errors may occur frequently even under optimized PCR conditions and with high quality template DNA. This is especially true for data where repeat typing is limited or not feasible as a result of cost or time constraints or limited DNA (Wang 2004).

The most common sources of error are: PCR slippage mutations producing stutter bands which lead to erroneous allele calling; allele drop-out or null alleles; "misprinting" where amplification products are generated and incorrectly interpreted as alleles; as well as data entry errors (Hoffman & Amos 2005; Wang 2005). Hoffman and Amos (2005) found scoring errors to be the most prevailing source of genotyping errors and also found a positive correlation between locus polymorphism and/or product size and PCR failure rates and/or allele drop-out rates. They suggest that primers should either be designed to amplify smaller fragments or to avoid loci producing large products sizes as a mean to controlling the impact of this phenomenon.

2.7.2 PCR Slippage and Stutter Profiles

Talbot *et al*. (1995: as cited in Wang 2004) report the mutation rates observed for microsatellite loci to be as high as 1.4×10^{-2} per generation. Mutational slippage results from polymerase slippage during DNA replication and leads to a change in the number of repeat units (Miller & Yuan 1997). Lai and Sun (2004) propose that slippage mutations within repeat sequences dominate over point

mutations in non-repetitive sequences and that a linear relationship exists between the number of repeat units and microsatellite slippage mutation rate. Lai and Sun (2003) further found that when PCR slippage occurs, the probability for contraction of the repeat was higher than the probability for the repeat to expand and that a threshold repeat number exists under which the mutation rate becomes too small to be observed. The assumption from their observations in these two studies was that each repeat unit within a microsatellite mutates independently.

A complication occurring during the analysis of microsatellite data is the appearance of secondary or shadow bands resulting in extended banding patterns known as stutter patterns or stutter profiles (Miller & Yuan 1997; Mueller & Wolfenbarger 1999). These extra bands are the products of PCR slippage and are often a hindrance to assigning alleles correctly in microsatellite genotyping since they may overlap in heterozygous alleles. The combined intensity forms a skewed composite pattern leading to possible misinterpretation of the data as a homozygous pattern (Mueller & Wolfenbarger 1999).

The problem of stutter profiles is observed to be most acute where dinucleotide repeats are concerned while greatly reduced in tri and tetranucleotide repeats; with further evidence reporting that longer microsatellite repeats are more prone to producing stutter than shorter microsatellites (Mueller & Wolfenbarger 1999). The latter observation was made during studies in which mutation rates for microsatellite slippage extensions and contractions was modelled as a linear function of repeat unit number (Miller &Yuan 1997). Stutter bands can be reduced and sometimes eliminated through the optimization of individual PCR conditions and through the use of “hot start” PCR (Mueller & Wolfenbarger 1999).

2.7.3 Microsatellite Null Alleles

Null alleles continue to pose persistent challenges in populations genetic studies. A microsatellite null allele is, by definition, any allele at a microsatellite locus that repeatedly fails to amplify to detectable levels via PCR (Dakin & Avise 2004). Three processes are currently identified as potentially leading to null alleles at microsatellite loci:

One of the causes of null alleles recurring is point mutations at primer sites resulting in poor primer annealing. In addition, null alleles can be generated by preferential amplification of allele size variants due to the competitive nature of PCR and the more efficient amplification of shorter alleles. The larger allele is reported to “drop-out” and a null allele resulting from this preferential amplification is sometimes referred to as a “partial null”. Loading more of the sample or adjusting contrast may, in some cases, allow this partial null to be observed. A final cause of null alleles, and one which can be controlled to some extent, involves inconsistent PCR failure as a result of poor

DNA quality or low DNA quantity. The unpredictability of PCR amplification due to poor DNA quality makes it difficult to identify when this is indeed the cause of the null allele (Dakin & Avise 2004).

Since the presence of null alleles can severely bias estimations of allele and genotype frequencies, it is important to accurately identify them as the source of homozygous excess and estimate their occurrence (van Oosterhout *et al*. 2006). Chakraborty *et al*. (1992: as cited in Dakin & Avise 2004) reflect that it is important to recognise, however, that populations that are out of Hardy Weinberg equilibrium and the subsequent heterozygote deficiency associated with this occurrence, may not always be as a result of null alleles but could also be due to other biological factors such as: population substructure, inbreeding, selection occurring at or near the locus or sex linkage. Marshall *et al*. (1998) recommend that any loci with a null allele frequency greater than 0.05 (5%) should be excluded from future analyses.

Solutions presented to alleviate non-amplifying alleles encompass the lowering of PCR stringency conditions to allow for base pair mismatches or to redesign primers further away from the microsatellite where it is proposed that mutations more readily accumulate (Pemberton *et al*. 1995). In a study by Dakin and Avise (2004), approximately ninety percent of reported studies were found to include loci with null alleles at a frequency less than $p < 0.40$ and usually less than $p < 0.20$, but that only a very small portion of these studies integrated statistical corrections to accommodate the effects thereof. Many software programs designed to analyse microsatellite data for various end applications, take cognisance of and factor in the presence of null alleles into their calculations while providing an estimation of the frequency thereof in the data set under analysis. Examples include CERVUS 2.0 (Marshall *et al*. 1998), the program used in the parentage assignment analyses within this study.

2.8 Parentage Assignments

2.8.1 Assignment of Individuals to Families and Populations

Genetic breeding programs incorporate information for selection candidates as well as relatives in an effort to increase the accuracy of selection and hence the subsequent genetic gain (Villanueva 2002). One of the difficulties encountered with the implementation of effective breeding programs is the establishment and maintenance of pedigree information. The rearing of families in isolation until animals are large enough to be physically tagged is impractical and introduces confounding environmental effects which may mask the genetic effects under study (Norris *et al*. 2000).

The effective tracing of individuals to both families and populations has a powerful application within an aquaculture environment where the species are highly prolific and minimising inbreeding

as a means to maximising response to artificial selection is an important goal (Norris *et al*. 2000; Jerry *et al*. 2004). As a result, the use of genetic tags as opposed to physical tracking and the establishment of pedigrees and parentage using codominant molecular markers are finding increasing application within aquaculture genetic breeding programs (Norris *et al*. 2000; Selvamani *et al*. 2001). The use of genetic tags in aquaculture enhancement programs allows the reproductive potential of hatchery based breeding stock to be assessed. However, the successful advancement of genetic gain is affected by factors such as the number, quality and variability of the molecular markers used, the completeness of sampling of all candidate animals to be used in pedigree analyses and assignments, the number of potential parents and the heritability of the trait of interest (Selvamani *et al*. 2001; Dodds *et al*. 2005). The most significant benefit of DNA parentage assignments is that the technique is non-invasive and can be applied in retrospect, once the growth of the animals has been evaluated (Jerry *et al*. 2004).

Not all microsatellite markers that are developed are suitable for assigning parentage and a thorough assessment should be conducted prior to including them in analyses. Many loci have segregating null alleles, are difficult to score reproducibly or do not amplify reliably. The inheritance of all markers should be tested against known parent-offspring relationships and any problematic markers should be excluded from future assignments (Jerry *et al*. 2004).

2.8.2 Software for Parentage Assignment - An Overview

2.8.2.1 Deciding Which Method is Best

Analysing the genetic diversity inherent in every living species is crucial in our understanding of the evolutionary process at both population and genome level (Excoffier & Heckel 2006). To fully comprehend the genetic variation present for numerous economically valuable traits, family lines need to be established and evaluated through the use of pedigree information of mass-spawned offspring (Elliot 2000). A barrage of statistical approaches and related software packages for the analysis of parentage within natural and experimental populations has followed in the wake of the DNA molecular marker boom and the details and implementation of the various techniques differ in ways that can influence the final output (Jones & Ardren 2003; Taggart 2007). The available programs offer both standard and more complex analyses; however, they rely to a large extent on assumptions that are fundamental to the correct interpretation of results (Excoffier & Heckel 2006).

When evaluating a potential software program, the underlying algorithm and its ability to handle null alleles, genotyping errors and weak linkage between loci, are important criteria against which to assess the prospective program's suitability to address the research problem at hand. It has further been suggested that the analysis of a data set using of a number of different statistical

assignment programs, with different strengths and weaknesses, would boost an assignment project; as would a realistic assessment of potential genotyping errors and their consequences in parentage studies (Taggart 2007).

The majority of current programs are based either on a complete exclusion principle or on categorical/likelihood methods, both of which have associated advantages and disadvantages. Likelihood assignment techniques should be considered as a means to compensate for the shortcomings often associated with a data set. These could include, for example, insufficient genetic variation, incomplete sampling, mutations, genotyping scoring errors and null alleles which can all confound the processing of parentage analyses and occur with relatively high frequency when working with large data sets (Jones & Ardren 2003).

Jones and Ardren (2003), in their overview of parentage analysis methods, encompass some guidelines to selecting an appropriate statistical package for parentage reconstruction, taking cognisance of the completeness of the samples and constraints as a consequence thereof. Four methods are discussed in their review, two of which, fractional allocation and parental reconstruction, are not relevant to this study. Both exclusion probabilities and categorical likelihoods were considered as appropriate for assigning parentage within the context of this thesis and are elaborated on.

2.8.2.2 Exclusion-Based Methods of Assignment

Exclusion techniques work on the principle of incompatibilities between parents and offspring; and on this basis, specific parent-offspring combinations are rejected. However, since genotyping errors, null alleles and mutations are only weakly tolerated by this method, and very few studies achieve perfection, false exclusions are a reality when applying an exclusion method.

As the amount of data being analysed increases so the number of loci required increases; which in turn leads to an increased likelihood of errors and mutations. As a result, exclusion-based methods become less accurate. Put differently, exclusion-based techniques, the earliest and simplest techniques to be conceptualised, are most accurate when highly polymorphic markers are available, and when employed in studies where few candidate parents are under consideration (Jones & Ardren 2007).

2.8.2.3 Likelihood-Based Methods of Assignment

Categorical likelihood methods assign offspring to non-excluded candidate parents based on likelihood scores derived from their genotypes. The likelihood of an individual to be the correct mother or father is positively correlated to the individual's homozygosity, or in other words; a

compatible individual that is homozygous for a particular locus has a higher likelihood of being the parent than would a compatible individual that is heterozygous for the locus. The LOD score, on which assignment is based, is derived from the logarithm of the likelihood ratio; which in turn is calculated by determining the likelihood of an individual (or pair of individuals) being the parent/s divided by the likelihood of these individuals being unrelated. Likelihood methods are more tolerant of genotyping errors and mutations; and in some instances, programs incorporate the occurrence of errors and null alleles into the algorithmic calculations (Jones & Ardren 2003).

2.8.2.4 Pros and Cons of Available Software Programs

The extent to which an algorithm is capable of detecting and handling errors and null alleles is crucial given that these factors can lead to false exclusion in parentage assignments (Jones & Ardren 2003). CERVUS (Marshall *et al*. 1998) makes allowance for genotyping errors and null alleles by allowing this parameter, the error rate, to be set by the researcher and further calculating an estimation of null allele frequency. MICRO-CHECKER is a WINDOWS®-based program that aids in the identification and correcting of genotyping errors in microsatellite data from diploid populations (van Oosterhout *et al*. 2004). These genotyping errors encompass null alleles, allele dropout and errors in scoring of stutter peaks. In addition, the program provides null allele frequency estimations (van Oosterhout *et al*. 2004). PAPA (Duchesne *et al*. 2002) or PARENTE (Cercueil *et al*. 2002) do not contain an algorithm for detecting null alleles and both these programs treat these sources of error as any other mutation or genotyping error (Cercueil *et al*. 2002; Duchesne *et al*. 2002; Jones & Ardren 2003). CERVUS, PAPA and PARENTE all have distinct algorithms with which to deal with mutations or genotyping errors although each program is unique in its approach to handling mismatches that result from scoring errors and mutations (Jones & Ardren 2003).

Current programs are not well equipped to accommodate strongly linked loci and linkage disequilibrium and it is best to avoid loci that show significant departure from Hardy Weinberg equilibrium (Jones & Ardren 2003). In saying that, weak linkage and HW disequilibrium are, in some cases, tolerated, for example within CERVUS. The assumption of independent assortment of alleles is made by most researchers but studies by Ardren *et al*. (1999) and Jones *et al*. (2001) show that this assumption is not always valid (as cited in Jones & Ardren 2003).

One of the assumptions made by most parentage programs is that the candidate parents are unrelated to one another. When relatives occur among candidate individuals, this poses some serious challenges to the accuracy of assignment. A strict exclusion approach will still be effective under conditions of relatedness; however, the number of loci required for complete exclusion to be

valid may become limiting in itself. Some programs are designed to test hypotheses of relationship between individuals; for example, KINSHIP (Goodnight & Queller 1999) which uses likelihood methods to test non-inbred relationships between individuals and calculations can be performed for both diploid and haploid populations (Goodnight & Queller 1999). FAP (Family Analysis Program) (Taggart 2007), on the other hand, is a DOS based assignment software that calculates exclusion-based assignment probabilities within families assuming all candidate parent genotypes are known.

Finally, confidence that the program employed reliably and accurately addresses the assignments at hand is important to both the researcher and any invested parties for whom the output will form the basis of future decisions and actions. Room for improvement exists for all programs currently available with regards to predicting statistical confidence (Jones & Ardren 2003).

2.10 What of the Future?

2.10.1 Linkage Mapping in *Haliotis midae*

A linkage map is also sometimes called a "recombination map". It defines the order of loci along a chromosome as well as the genetic distance between these markers and is constructed on the basis of inheritance of these loci within families or mapping populations (Chistiakov *et al*. 2005). To be able to map QTL and identify the position of potentially valuable genes, linkage maps need to be constructed. The increasing interest in and application of comparative genomics has allowed for noteworthy progress to be made in the development of these genetic maps for aquatic species; with slower progress evident for the marine gastropod *Haliotidae* (Baranski *et al*. 2006; Sekino & Hara 2006).

Molecular markers suitable for mapping purposes need to be polymorphic, reliable, well distributed throughout the genome and amenable to automation for high throughput analyses. To date, several classes of molecular markers (RAPD, AFLP, microsatellite markers and SNPs) have been employed to establish linkage maps of varying density within fish and shellfish species (Martinez 2006; Sekino & Hara 2006) and to provide sufficient genome coverage. However, microsatellite marker-based linkage maps are considered by Sekino and Hara (2006) to be essential in addressing the challenges of evolutionary complexity, genetic divergence and incongruities between morphological/ecological differences present in some species of *Haliotis*. Baranski *et al*. (2006) concur with this statement although stipulate that the fairly frequent presence of null alleles and segregation distortions observed with microsatellite markers tend to pose some complications.

2.10.2 Quantitative Trait Loci (QTL) and QTL Mapping

Traits which reveal continuous variation are generally controlled by several genes exerting varying degrees of influence on the trait. Although mapping the genes that affect quantitative variation in natural populations has become more feasible with the recent development of molecular markers such as microsatellites and SNPs (which highlight polymorphisms at population level), together with PCR-based screening of these loci; QTL mapping remains difficult for many organisms as the technique requires many polymorphic loci dispersed throughout the genome. The heritability of the trait, its genetic nature and the number of genes influencing it, all contribute to the success of mapping a QTL.

Jonasson *et al*. (1999: as cited in Robinson *et al*. 2003) show that the heritability of growth rate increases with age in red abalone; meaning that early selection, based on size, would be an inaccurate and poor predictor of performance at a later age. However, should molecular markers linked to growth genes be available, these in turn could be utilised as a more effective and accurate assessment of future performance and hence employed as an early selection tool for this valuable trait (Robinson *et al*. 2003). The identification of QTL will further enable the improvement of economically important traits that are difficult to breed for, such as: disease resistance and overall health, meat quality and yield, sexual maturation and feed conversion efficiencies (FCE) through the implementation of marker assisted selection and breeding programs (Jones *et al*. 2001; Martinez 2006).

Since information about the genome of most organisms is limited, generating enough molecular markers through traditional techniques to enable a suitably dense map to be constructed is often restricted by time and costs (Sekino & Hara 2006). For mapping of single trait QTL, it is possible to reduce the number of individuals that need to be genotyped through selective genotyping and DNA pooling methods; thereby reducing the often restrictive costs of time and finances (Robinson *et al*. 2003). Looking to the future with the newly developed animal models and genomic tools that are available to smooth the progress of QTL mapping; SNPs maps, improved software programs and finally transgenic procedures are all helping to narrow the search for candidate genes (Abiola *et al*. 2003).

The initial stage of QTL mapping involves coarse mapping wherein loci are mapped to a chromosome segment to within a specified range. Before proceeding to finer QTL mapping, linkage should be confirmed through independent crosses, the construction of congenic strains or a short term selective breeding study (Abiola *et al*. 2003). Once QTLs are known and selective breeding has achieved lines or breeds containing constructive alleles; the design of new genotypes combining favourable alleles from multiple loci can begin, using genetic building strategies (Dekkers &

Hospital 2002). According to Charmet *et al*. (1999: as cited in Dekkers & Hospital 2002) and Hospital *et al*. (2000), producing individuals that are homozygous for a large number of loci may necessitate successive generations of mating and selection. Environmental influences are also known to contribute to the phenotypic expression of a QTL and it is important to keep this in mind when constructing new genotypes which may perform differently under diverse environmental conditions (Robinson *et al*. 2003).

Our understanding to date, of aquaculture genomics, has relied on the use of information from well studied species. The development of Quantitative Trait Loci (QTL) maps for *H. midae* will be a key outcome from the rapid molecular marker development currently taking place within aquaculture and for abalone in particular. QTL mapping is further driven by an interest in specific production traits and performance unique to a species and the subsequent identification of QTL related to these traits (Liu & Cordes 2004).

2.10.3 Marker Assisted Selection (MAS)

MAS offers a potential early selection strategy whereby future breeders/breeding stock are chosen based on a combination of overall phenotype and genetic conformation in terms of known markers linked to genes or traits of economic importance. Genotypic information is determined through the application of molecular genetic markers (Bentsen & Olesen 2002; Liu & Cordes 2004). Microsatellite loci are particularly suitable for gene mapping and MAS programs (Robinson *et al*. 2003; Martinez 2006). Using these DNA markers, it is then possible to theoretically observe and exploit genetic variation in the entire genome. Within specifically bred family lines, animals showing the best possible phenotypic traits, in combination with the best possible genetic makeup in terms of available marker information, could then be selected for ongoing breeding for improved traits (Robinson *et al*. 2003).

2.10.4 Pedigree Records versus Mass Selection

Although breeding programs based on mass selection procedures are simpler and more cost-effective to implement than pedigree based techniques requiring the physical or genetic tagging of the individuals involved; these designs can only be used to select for phenotypic traits that can be visually observed and recorded for the selection candidates and are best employed when selecting for a single trait (Bentsen & Olesen 2002). The most valuable trait within a breeding program is very often growth rate which can easily be selected for through a mass selection process. In order for mass selection to be effective, breeding designs need to be developed which reduce the rate of inbreeding, thereby maximising the response to selection while minimising the associated risks.

Alternatively, MAS could provide an effective means to select the best future breeders for a breeding program. Since selection for one important trait may be at the expense of another, it is essential for a well-designed MAS program to consider all the economically valuable traits before embarking (Liu & Cordes 2004). It is expected that the implementation of MAS would provide a greater selection response to that achieved with purely pedigree-based techniques since the within family variation would be exploited to predict breeding values (Martinez 2006).

A combination of biotechnology and more traditional selective breeding techniques such as marker assisted selection (MAS) is likely to represent the longer term future for genetic improvement (Elliot 2000). Family-based breeding programs have several distinct advantages over mass selection programs for aquaculture species: the improvement of a wider range of traits with better control over inbreeding, and the implementation of MAS within family-based structures provides an accurate means of estimating the effects of QTL segregating within families, which reciprocally determines the success of MAS (Hayes *et al*. 2007).

One of the limitations of implementing MAS is the cost associated with genotyping full-sib families of the selection candidates for the trait of interest; making it an inaccessible tool to most aquaculture programs. A selective genotyping strategy evaluated against the effect on selection response as well as a cost-benefit analysis would need to be investigated to determine the most profitable approach to implementing MAS (Hayes *et al*. 2007).

2.10.5 Walk-back and Optimum Contribution Selection Schemes

Walk-back selection is a protocol which achieves extreme family-based selection by exploiting the high fecundity of aquaculture animals, while minimising the level of inbreeding. It assumes that selection is for a trait that can easily be recorded such as size or body weight.

The individual with the highest phenotypic value is selected as a breeder and DNA fingerprinted; where after, the animal with the second largest phenotypic value is selected and similarly DNA fingerprinted. This second animal will only be added to the pool of potential breeders should it not be related to the first individual. The walk-back process (from the upper end of the size distribution) is repeated until a sufficient number of breeders are obtained (Doyle & Herbinger 1995; Sonesson 2005). Doyle and Herbinger (1995) estimated that walk-back selection exceeds combined selection using physical tagging methods in the order of one to three standard deviations, while being thirty to forty times more cost-effective in the long run.

Sonesson (2005) reports a novel breeding scheme for fish which combines walk-back selection with optimum contribution selection. Optimum contribution selection is a group selection procedure that maximizes genetic gain while placing a restriction on the inbreeding coefficient. It is a dynamic

scheme that uses both discrete and overlapping generations and is adaptable to the current selection candidates. Within this combined two-stage protocol, a reduced number of individuals are genotyped; thus, reducing costs while maintaining a high level of genetic gain at a constrained rate of inbreeding. Sonesson (2005) maintains that two-stage selection schemes are very effective particularly when there is a strong correlation between the first and second stages of selection.

Chapter Three

Sample Processing and PRS Data

Abstract

Assessment of Performance Recording Scheme individuals for Abagold and HIK yielded comparable results for both the faster and slower growing groups at three of the five rearing locations. Further it was observed that for the faster growth group, for both Abagold and HIK progeny, juveniles that were reared at HIK and Roman Bay were on average heavier and longer than those of the same age at Aquafarm, I&J and Abagold. Similarly for the slower growth group, Abagold and HIK offspring reared at Aquafarm and Roman Bay had marginally higher weights and shell lengths than offspring at the remaining three rearing locations. These differences, despite not being statistically significant, are worth mentioning since they may be an indication of environmental effects contributing to the slight growth advantages seen for the juveniles being reared at these locations.

Within a controlled laboratory environment it is much simpler to manage external influences than it is to control these factors within the reality of a production environment; however, there is indication that genotype x environment interactions may require further investigation and potentially be factored into future selective breeding programs. Potentially, rearing juveniles across different locations in an effort to minimize environmental influences may not be enough when utilizing current production and husbandry practices. This may be particularly relevant during the initial stages of the PRS when one is selecting candidate animals for monitoring; and standardising feeding and other aspects of the rearing process may be worth consideration.

Abagold lost two breeding individuals, one male and one female, prior to sampling and hence sampling was 98% complete while sampling was 100% complete for HIK. Incomplete sampling has obvious implications for the success of parentage assignment and reconstruction of genotypes of missing parents may be a means to evaluating their contribution.

Chelex®100 resin was shown to be a cost-effective and time reducing alternative method of DNA extraction to a standard CTAB or phenol-chloroform protocol, given any future cost or time constraints, and may be feasible in combination with the high through-put of a robotic platform.

3.1 Introduction

A sound experimental design and the ability to accurately measure traits of interest very much influences the ability to find gene markers (Sekino & Hara 2006). Since a long term objective of this industry project would entail the location of quantitative genes influencing growth, disease resistance, meat quality and other economically important traits; the initial preparations and groundwork, as well as the successful completion of preliminary objectives such as those set for this study, go a long way to realising these future goals.

A holistic approach to chapter content was adopted for the presentation of the research findings from this study; and to this end, the following sections are encompassed in chapter three: materials and methods presents the structure of a PRS (Phase I) that was used to monitor F1 offspring within different rearing localities and hence environments; implementation of mass-selection techniques to select for potentially faster and slower growing individuals on the basis of their size at one year, together with the process of measuring selected individuals (weight and shell length); the activities of tissue sampling and tagging of the PRS offspring and finally, genomic DNA extraction for both adults and juveniles samples. Results document the processing of adult and offspring data; the concentrations of genomic DNA obtained and the interpretation of absorbance values; mass-selection outputs from the PRS in terms of size and weight measurements; as well as the bee-tag tagging procedure used to physically mark those offspring measured at one year old for later re-measurement.

Since the reproducibility and ultimate accomplishment of the outputs of a project are extremely reliant on an accurate, structured and well planned experimental design, the shortcomings and sources of error that were encountered, and which can potentially impact on the eventual parentage analyses, are highlighted. The discussion covers only the results presented in this chapter; where it evaluates the findings and presents suggestions for improvement of specific aspects.

3.2 Materials and Method

3.2.1 Breeding Stock Information

3.2.1.1 Abagold (Pty) Ltd

Four individuals of the same sex are held per tank at the Abagold hatchery facility and breeding stock are divided into three spawning groups. A total of ninety-nine individuals selected from across all three groups were used to produce offspring for the Performance Recording Scheme (PRS) and since these individuals are currently used in production, they were all already conditioned for spawning. The same Sample ID that was assigned to each 1.5ml eppendorf sample tube upon tissue collection was maintained throughout the project for both genotyping and parentage assignment and later used to identify the corresponding Tag ID for physical identification of relevant breeders. Details in terms of: Sample ID, Tag ID, gender, tank number and group number are provided in Appendix 1(a). Table 5 shows the distribution of breeding stock used in the PRS in terms of males and females across the three spawning groups.

Table 5 Distribution of male and female adults across Spawning Groups: 1, 2 and 3

	No. Males	No. Tanks	No. Females	No. Tanks	Total Animals	Total Tanks
Group 1	16	4	28	7	44	11
Group 2	20	5	23	6	43	11
Group 3	4	1	8	2	12	3
Total	**40**	**10**	**59**	**15**	**99**	**25**

3.2.1.2 HIK Abalone (Pty) Ltd

A varying number of same sex adult individuals are held per tank at the HIK Abalone hatchery facility. Group B and Group C are both conditioned spawning groups and one hundred and one individuals from these two groups were selected to produce the offspring entered in the PRS. The sample ID assigned to the 1.5ml eppendorf tube prior to tissue collection was maintained throughout the project for both genotyping and parentage assignment and was later used to identify the corresponding Tag ID for physical identification of contributing breeders. Details of the breeding stock in terms of: Sample ID, Tag ID, gender, tank number and tank position (where applicable) are provided in Appendix 1(b). Table 6 indicates the distribution of PRS males and females across Groups B and C.

Table 6 Distribution of male and female adults across Spawning Groups: B and C

	No. Males	No. Tanks	No. Females	No. Tanks	Total Animals	Totals Tanks
Group B	22	7	20	8	42	15
Group C	31	10	28	8	59	18
Total	**53**	**17**	**48**	**16**	**101**	**33**

3.2.2 Performance Recording Scheme (PRS)

3.2.2.1 Background

The availability of information concerning the genetic quality of commercial stocks of *Haliotis midae* is very limited and is further hindered by the lack of information exchange between farms. Although production technologies, husbandry, feeding and management strategies are already well established on commercial farms, in order for overall productivity, performance and profitability to be addressed, the genetic quality of the commercial stock needs to be better understood and then improved through the selection of superior animals for breeding purposes.

Through the securing of the five year collaboration between five of South Africa's leading abalone producers, the channels of communication were finally opened; and the subsequent implementation of a Performance Recording Scheme (PRS) made further provision for the following:

(1) The assessment and comparison of the production performances of participating hatcheries as an indicator of genetic qualities

(2) The identification of genetically superior genotypes from current breeders and commercially produced progeny groups

(3) The establishment of an effective evaluation protocol for future breeding populations to be developed in Phase I of the Innovation Fund (IF) project

The five abalone farms participating in the PRS are: Abagold (Pty) Ltd, Aquafarm Dev. Comp (Pty) Ltd, HIK Abalone (Pty) Ltd, Irvin and Johnson (Pty) Ltd and Roman Bay.

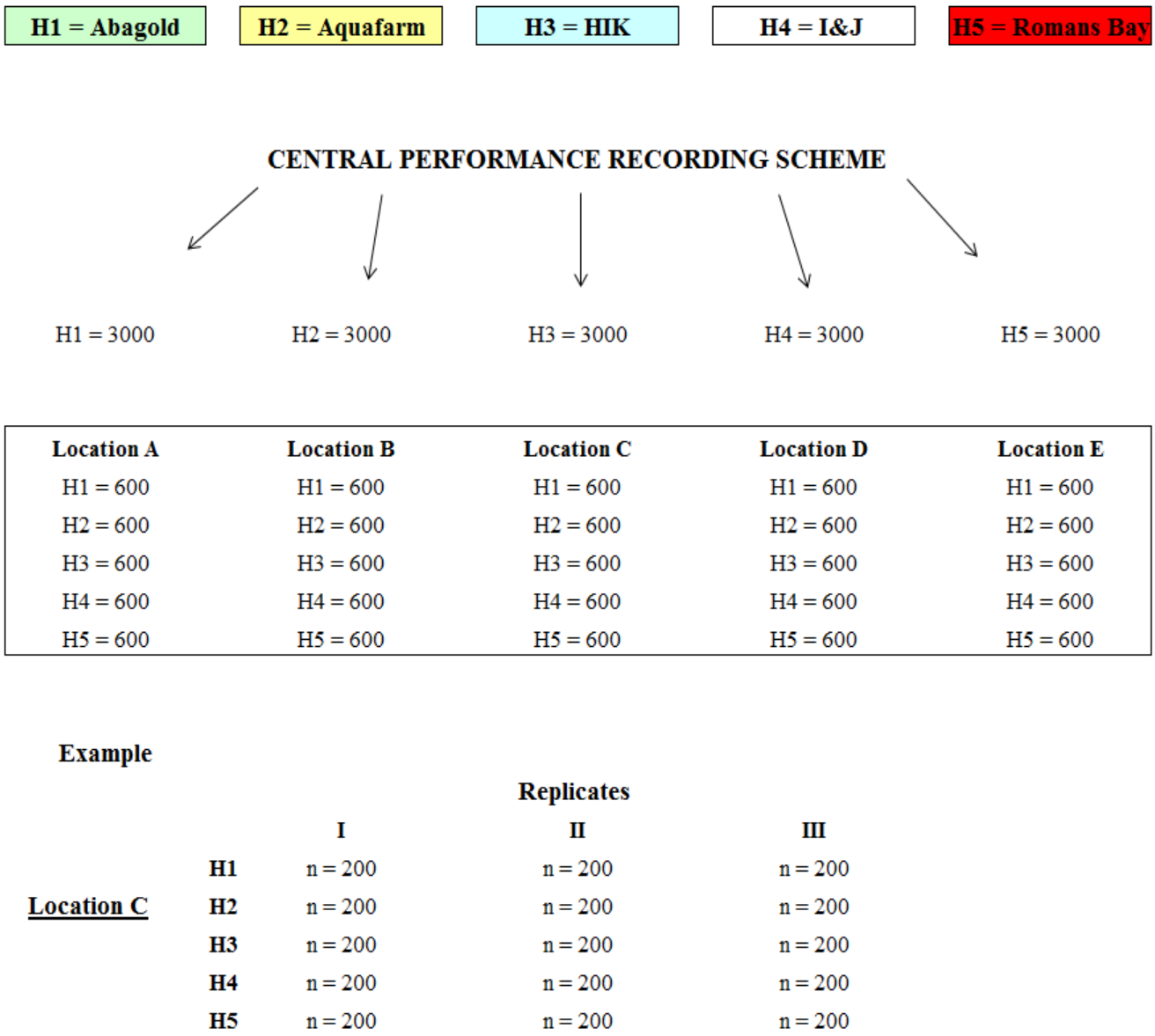
3.2.2.2 Structure of the Performance Recording Scheme

The PRS is divided into two phases, the first of which has already been implemented through the procedures indicated in the flow diagram (Figure 8). This phase is based on the assessment and comparison of animals produced by current breeders through mass spawning practices and randomly distributed across all farm locations for rearing. Due to the limited genetic information currently available for *Haliotis midae*, this phase is limited to the application of the most basic of selection procedures, mass-selection. The second phase will be based on the evaluation of a specialized breeding population with a defined family structure that will allow for the implementation of more advanced molecular selection techniques such as walk-back selection.

Three thousand animals were produced by each hatchery for subscription to the PRS. For the purpose of this study, the breeding animals detailed in Appendix 1(a) and (b) were used and spawning was conducted by each farm according to current breeding practices. All offspring submitted to the PRS were of uniform age and entered into the scheme on a prescribed date.

Each farm made provision to accommodate six hundred tagged offspring from each hatchery resulting in a total of three thousand test animals per farm. Once the abalone groups were randomly distributed to the different locations, the animals were allowed a six month adjustment period to adapt to their new environment prior to the commencement of evaluations. Measurement of the animals and assessment of specific environmental and hatchery effects were conducted at quarterly intervals. The protocol remains a dynamic document which is reviewed and adjusted on a regular basis based on the input from participating individuals. Ongoing feedback has contributed to the

identification and elimination of problems and in so doing improving accuracy and validity of the PRS going forward into Phase II.



Location A	Location B	Location C	Location D	Location E
H1 = 600	H1 = 600	H1 = 600	H1 = 600	H1 = 600
H2 = 600	H2 = 600	H2 = 600	H2 = 600	H2 = 600
H3 = 600	H3 = 600	H3 = 600	H3 = 600	H3 = 600
H4 = 600	H4 = 600	H4 = 600	H4 = 600	H4 = 600
H5 = 600	H5 = 600	H5 = 600	H5 = 600	H5 = 600

Example

		Replicates		
		I	II	III
Location C	H1	n = 200	n = 200	n = 200
	H2	n = 200	n = 200	n = 200
	H3	n = 200	n = 200	n = 200
	H4	n = 200	n = 200	n = 200
	H5	n = 200	n = 200	n = 200

Figure 8 Structure of the Performance Recording Scheme for Phase I detailing the random distribution of three thousand offspring to each of the five farms participating in the genetic improvement program of local abalone species, *Haliotis midae*

3.2.2.3 Production Traits to be Measured or Assessed

The following factors were measured or assessed every three months to provide a holistic data set of both economically important traits and production performance factors; the latter to be used as indicators of the genetic qualities of the animals they impact on: growth traits (weight and shell length), yields, mortalities, environmental influences, water temperature, water quality and densities.

Hatchery effects in terms of constant but unexplained performance differences between siblings across hatcheries were also monitored. For this study, only the growth traits were of interest as they provided an indication of those individuals who were potentially faster growing versus those that were potentially slower growing.

3.2.2.4 Collection of Offspring Data and Tagging

The offspring data required for this parentage study was collected in August 2006. Only individuals from replicate one were assessed and the weight (in grams) and shell length (in millimetres) for the upper and lower 10% of animals for Abagold and HIK across all locations were recorded. This translated into 20 larger individuals (fast growers) and 20 smaller individuals (slow growers) per location per farm. The size ranges against which these upper and lower 10% of individuals were selected were determined using previous size data collected in the PRS quarterly assessments. The ranges differed slightly for each farm's animals at each location as they were based on the prior performance of each farms test animals at the respective location. The ranges against which each group were measured are provided in Appendix 2.

Each animal, for which the size and the weight were recorded, was tagged with a bee-tag so that it could be located again at the age of two years and re-measured to determine whether the size and weight advantage or disadvantage had been maintained or not. The bee-tags were applied using water-resistant superglue and a sharp pointed tool to position and secure the tag to the shell. Size measurements were taken when the offspring had reached approximately one year in age.

3.2.3 Collection of Abalone Tissue Samples

3.2.3.1 Tissue Sampling of Abalone Breeding Stock

Tissue samples were collected in February 2006 for the majority of the breeding animals committed into the PRS by HIK. The remaining animals were tagged during the period February to August 2006 at which point a second tissue sampling took place together with the sampling, measurement and tagging of the offspring. For Abagold, tagging was only completed in August 2006 and sampling was done for both adults and offspring in this period.

For all adult animals, DNA samples were removed in a non-destructive manner (Slabbert & Roodt-Wilding 2006) without the animals being anaesthetized. Two to three epipodial tentacles were cut from each animal using sharp dissection scissors and tweezers and stored immediately in pre-labelled 1.5 ml eppendorf tubes containing 99.9% ethanol. After removal of the tissue sample from each animal, scissors and tweezers were wiped clean to prevent DNA cross-contamination. Tissue samples were stored at 4˚C until DNA extractions were conducted.

3.2.3.2 Tissue Sampling of Abalone Offspring

Given the very small size of some of the offspring, making handling difficult for the researcher and at the same time stressful to the animal; all offspring, in all replicates, were anaesthetized prior to sampling by using each farm's own practices and protocols. This was done in order to standardize external influences for all juveniles despite replicates two and three not being sampled.

Following the method in Slabbert & Roodt-Wilding (2006), snips of epipodia were taken from each offspring using sharp dissection scissors and tweezers and transferred immediately into pre-labelled 1.5 ml eppendorf tubes containing 99.9% ethanol. After removal of the tissue sample from each animal, scissors and tweezers were wiped clean to prevent DNA cross-contamination. Each offspring was then weighed, measured and tagged with a bee-tag. Tissue samples were stored at -20˚C until DNA extractions could be completed.

3.2.4 Genomic DNA Extractions

3.2.4.1 Genomic DNA Isolation from Abalone Breeding Stock Tissue

Isolation of genomic DNA from epipodial tentacles was performed using a standard CTAB extraction protocol as described by Saghai-Maroof *et al*. (1984). Tentacles were transferred to labelled 1.5ml eppendorf tubes containing 500µl of freshly prepared digestion buffer [2% (w/v) CTAB, 1.4M NaCl, 0.2% (v/v) β-mercapto-ethanol, 20mM EDTA pH 8.0, 100mM Tris-HCl pH 6.8] to which 2µl Proteinase K (10mg/ml, *Roche*) was added. Each sample was shaken gently by hand to mix the contents thoroughly and then incubated overnight in a waterbath at 60˚C. Given time constraints this incubation step may be reduced to between three to eight hours but it must be ensured that all the tissue has been digested before continuing. An equal volume of chloroform:isoamyl alcohol (24:1) was added to each tube, followed by gentle mixing for five minutes using a Vortex-Genie 2 (Scientific Industries). Samples were then centrifuged at high speed (13,200 rpm) for ten minutes using an Eppendorf Centrifuge 5810D set at 4˚C to separate the aqueous and organic phases. Each clear aqueous supernatant was transferred to a pre-labelled 1.5ml eppendorf tube and the chloroform-isoamyl step repeated until the interface was clean.

DNA was precipitated overnight at -20˚C by adding two-thirds (⅔) volume of cold isopropanol to the supernatant. Again, given any time constraints, this precipitation step can be reduced to a minimum of one hour at -20˚C or until a precipitate is evident. Samples were centrifuged at high speed (13,200 rpm) for twenty minutes at room temperature in an Eppendorf Centrifuge 5415D and the isopropanol pipetted or poured off, taking care not to dislodge or discard the DNA pellet. To wash the pellet of DNA, 200µl 70% ethanol was added to each sample tube and left for a few

minutes before centrifuging a second time at 13,200 rpm for ten minutes at room temperature. After careful removal of the ethanol, the DNA pellet was dried for 10-15 minutes at 55˚C in an oven. DNA was re-suspended in 50µl sterile distilled water and quantified using a NanoDrop® ND-1000 Spectrophotometer. DNA samples were stored at -20˚C until required for analysis.

3.2.4.2 Genomic DNA Isolation from Abalone Offspring Tissue

The amount of available tissue for each individual was significantly less than for the adult abalone; particularly for the lower 10% of animals, given their very small size and the need for non-destructive sampling (Slabbert & Roodt-Wilding 2006).

Automated DNA extraction was used for isolation of genomic DNA from offspring tissue samples using the Nucleospin® 96 Plant Extraction Kit (*Macherey-Nagel*). Automation was opted for to ensure maximized extraction of DNA, to reduce any risk of sample loss and to meet time constraints given that four hundred samples had to be processed within a short time frame. Several different DNA extraction kits were tested and cost versus result (in terms of DNA quantity and quality) was assessed before a final decision was made. In using the latter kit, economies of scale could be applied while still providing good quality DNA; and eluted DNA was ready-to-use for PCR without further purification required.

The following buffers were supplied with the kit and were prepared according to the provided instructions prior to starting extractions: C1 lysis solution, C4 buffer and C5 buffer. C4 buffer is stable for four months at room temperature and the recommendation by the supplier is that it be stored in the dark. CW buffer was used as provided without need for prior preparation of a working solution.

The Nucleospin® 96 Plant Protocol was modified slightly to include an overnight digestion step using C1 lysis solution and Proteinase K. Epipodia snips were transferred from their ethanol storage medium into pre-labelled eppendorf tubes to which 400µl C1 buffer and 2µl Proteinase K was added. Samples were shaken gently to mix contents and incubated overnight in a waterbath at 60˚C. Prior to loading sample tubes onto a Genesis RMP200 Automated Liquid Handler (*Tecan*) modified for vacuum processing, all samples were centrifuged for 20 minutes at high speed (13,200 rpm) at room temperature.

Tubes were loaded with the caps open into racks on the workstation and DNA extraction was then completed entirely by a robotic platform at the Central DNA Sequencing Facility located at the Genetics Department, Stellenbosch University.

Processing under a vacuum allows for complete automation without the need for centrifugation steps for membrane drying or elution. Optimized vacuum settings reduce the risk of cross-

contamination during the elution step. Specific scripts and general considerations for adapting an extraction kit to the workstation being used are available from *Macherey-Nagel* but the principle of extraction remains the same:

After digestion of the tissue and homogenisation if necessary, DNA is extracted with lysis buffers containing chaotropic salts, denaturing agents and CTAB lysis buffers. The CTAB removes contaminants while DNA remains in solution. Lysis mixtures should be cleared by centrifugation or filtration in order to remove contaminants and residual cellular debris. The clear supernatant is mixed with binding buffer and ethanol to create conditions for optimum binding of the DNA to a specialised silica membrane in the Nucleospin® Plant Binding Module. DNA is subsequently washed with two different buffers, CW and C5, and is eluted either in a low salt buffer or water as preferred. The eluted DNA can be used in subsequent reactions without any further processing. Elution methods and volume of elution buffer can be adapted depending on the end application of interest; since only the concentration, and not the purity of DNA, is affected by the volume of the elution buffer (Source: User Manual: Nucleospin® 96 Plant 2004). Offspring DNA was eluted using sterile distilled water and quantified using a NanoDrop® ND-1000 Spectrophotometer. DNA samples were stored at -20˚C until required for analysis.

3.2.4.3 Comparison between Extraction Methods and PCR Results

A variety of different methods are available for DNA extraction. Although a CTAB method (Saghai-Maroof *et al*. 1984) is very often used, extractions using phenol-chloroform instead of chloroform-isoamyl as well as Chelex®100 resin (Sigma) are reported; with the latter having been modified to extract DNA from various forensic samples such hair, sperm, semen and blood as well as from cells in culture (Walsh *et al*. 1991). More recently, it was the extraction method of choice for DNA extraction from caudal fin tissue of Senegal sole (Castro *et al*. 2006) as well as mantle tissue from Californian red abalone, *H. rufescens* (Kirby *et al*. 1998). The opportunity exists to use this in the future should it prove viable, especially when large scale extractions are required within a restricted time frame. The protocol used for the Chelex®100 DNA extractions is provided in Appendix 3.

Given the relative simplicity of the Chelex®100 method and the varied application, a comparison was made between the three extraction methods and PCR reactions were run on genomic DNA to ascertain whether there were any differences in amplification success. Each method was applied to three samples of abalone tissue and the concentrations quantified using a NanoDrop® ND-1000 spectrophotometer (Rockland, DE, USA). PCR was run using a Touch-

Down PCR program (Appendix 4) for microsatellite marker HmD11 (Bester *et al*. 2004) in the following reaction mix: 1µl undiluted DNA, 200µM of each primer for HmD11 (forward and reverse), 2mM $MgCl_2$, 1X *Taq* polymerase buffer, 200µM of each dNTP and 0.25U Go*Taq* Flexi DNA polymerase. Once amplification was confirmed, two samples were selected from each method and DNA concentrations standardised by dilution to 20ng/µl. A multiplex PCR was subsequently run on each sample using primers for loci HmD55 and HmD59 (Bester *et al*. 2004) in the following reaction: 20ng DNA, 100µM of each HmD55 and HmD59 primer (forward and reverse), 2mM $MgCl_2$, 1X *Taq* polymerase buffer, 200µM of each dNTP and 0.25U Go*Taq* Flexi DNA polymerase.

PCR products from both tests were resolved on a 2% 1X TBE agarose gel: 8µl PCR product combined with 2µl loading dye, 1.5µl Hyperladder IV (Bioline) and run for thirty minutes at 100V. Visualisation was with 0.03% (v/v) Ethidium Bromide stock solution (EtBr) (10mg/ml) and Ultra Violet light (UV).

3.2.4.4 Quantification of Isolated Genomic DNA

The quantity of DNA isolated from tissue samples was assessed using a NanoDrop® ND-1000 spectrophotometer for small volume UV analysis. In order to ensure complete sample homogeneity and accuracy in readings, samples were mixed by gently tapping the tube prior to loading the sample for analysis. A sample volume of 2µl was pipetted onto the fibre optic surface and a single reading recorded. The fibre optic surface was wiped thoroughly to remove all trace of the previous sample before loading the next one.

3.2.4.5 Preparation of Genomic DNA Dilutions

Quantification of genomic DNA samples allowed for working stocks to be prepared to the desired concentrations for the intended application and all dilutions were made using sterile distilled water. All working stocks for primer testing were made to approximately 20ng/µl when the initial concentration was sufficiently high to permit dilution; working stock solutions for genotyping were made to approximately 5ng/µl.

Where initial concentrations were very low, undiluted aliquots were used as the working solutions and volumes adjusted in the reactions to meet the required DNA concentrations. Working stocks were prepared as separate aliquots so as minimize the risk of cross-contamination of primary stock solutions. All primary and working stocks were stored at -20˚C.

3.3 Results

3.3.1 Performance Recording Scheme (PRS)

3.3.1.1 Background

In order to identify superior genotypes from current breeders and commercially produced Next Generation Individuals (NGI), it is imperative to have an understanding of spawning success and hence main male and female contributions to spawning need to be identified. To this outcome, the genotypes of potentially faster and slower growing individuals were compared to potential parents through parentage assignments and any trends existing with regards to growth were examined. This was the primary objective of the study and results thereof are presented in chapter five.

3.3.1.2 Production Traits Measured or Assessed

For the purpose of this study, only the growth traits of weight and shell length of the 10% faster and slower growing F1 progeny were required. Data was collected when the juveniles were approximately one year old and this was done for each farms test animals across all locations. The larger and smaller offspring were selected visually and then compared to pre-determined minimum and maximum length and weight criteria for allocation to the upper or lower growth groups.

3.3.1.3 Collection of Offspring Data and Tagging

Juveniles were tagged with bee-tags so that they could be identified later for re-measurement and re-assessment. Figure 9 illustrates a coloured and numbered bee-tag that was applied to one of the larger offspring. These animals had been tagged prior to them being distributed to the different locations for rearing so that the farm of origin could be identified whenever measurements were taken. Each farm was allocated a different tag colour and originally the offspring were tagged with a silicone tube through the respiratory pore.

Figure 9 Coloured and numbered bee-tags were applied to the shell of the juvenile using water-resistant superglue and a sharp pointed tool to position and press the tag onto the shell

Tag loss using the original method (silicone tags) was quite substantial for some of the animals and hence a different technique using bee-tags was tested. Tag loss will again be assessed at a later stage in the PRS but these results do not fall within the scope of this specific study.

3.3.1.4 Offspring Data for Abagold

The average weights and lengths of the upper and lower 10% of Abagold animals across all locations are summarized in Table 7. For the faster growing individuals the highest average weight recorded was 4.97g (HIK) and the lowest average weight was 3.54g (I&J). Although weights and lengths were not entirely correlated, there appears to be a link between weight and size/length of the animal. In other words, the heavier animals are also on average longer. For the slower growing individuals, the highest average weight was 1.21g (Aquafarm) and the lowest average weight was 0.88g (Abagold).

Table 7 Average weight (grams) and length (millimetres) for the upper 10% and lower 10% of Abagold PRS F1 progeny across all locations

	Fast Growers			Slow Growers	
	Avg. Weight	**Avg. Length**		**Avg. Weight**	**Avg. Length**
Location					
Abagold	3.56	27.16		0.88	16.23
HIK Abalone	4.97	30.54		1.03	17.08
Aquafarm	3.56	27.29		1.21	19.24
I & J	3.54	27.26		0.92	17.04
Roman Bay	4.36	30.15		1.04	17.83
Overall Average	**4.00**	**28.48**		**1.02**	**17.49**

For the faster growing individuals the difference between the maximum and minimum weight averages was 1.43g and difference between the maximum and minimum shell lengths averages was 3.38mm. Similarly, for the slower growing individuals the difference between weight averages (max. less min.) was 0.33g and shell length averages (max. less min.) was 3.01mm. Between faster and slower growing groups this difference was 4.09g (heaviest vs. lightest) and 14.31mm (longest vs. shortest) for weight and shell length respectively.

For the faster growth group, animals located at HIK and Roman Bay were slightly heavier and longer than those at the other three locations and similarly, for the slower growth group, those juveniles reared at Aquafarm were slightly larger than at the other farms. Animals located at the remaining three locations were almost identical in terms of weight and shell length for the faster

group while for the slower group, the measurements recorded for animals at the remaining four rearing locations were comparable.

3.3.1.5 Offspring Data for HIK Abalone

The average weights and lengths of the upper and lower 10% of HIK animals across all locations are summarized in Table 8. For the faster growing individuals the highest average weight was 5.36g (Roman Bay) and the lowest average weight was 3.86g (Aquafarm). Weights and lengths appear correlated, with the heaviest animals also being on average the longest. For the slower growing individuals, the highest average weight was 1.21g (Aquafarm) and the lowest average weight was 0.90g (I&J).

Table 8 Average weight (grams) and length (millimetres) for the upper 10% and lower 10% of HIK PRS F1 progeny across all locations

	Fast Growers			Slow Growers	
	Avg. Weight	**Avg. Length**		**Avg. Weight**	**Avg. Length**
Location					
Abagold	4.41	29.22		0.91	16.70
HIK Abalone	5.22	31.55		1.18	17.83
Aquafarm	3.86	28.56		1.21	19.11
I & J	4.01	28.88		0.90	17.62
Roman Bay	5.36	31.82		1.19	18.81
Overall Average	**4.57**	**30.01**		**1.08**	**18.01**

For the faster growing individuals the difference between the weight averages (max. less min.) was 1.36g and shell length averages (max. less min.) was 3.26mm. Similarly, for the slower growing individuals, the difference between highest and lowest weight averages was 0.31g and shell length averages was 2.41mm. Between faster and slower growing groups the difference was 4.32g (heaviest vs. lightest) and 15.12mm (longest vs. shortest) for weight and shell length respectively.

It was again noted that for the faster growth group, animals being reared at HIK and Roman Bay were on average longer and heavier than those at the other three locations and for the slower growth group, animals at Aquafarm showed a marginal but not statistically significant advantage.

3.3.3 Collection of Abalone Tissue Samples

3.3.3.1 Tissue Sampling of Abalone Breeding Stock

Tissue samples were available from all of HIK adult individuals in the PRS and parental genotypes were therefore complete. For Abagold, two individuals, one male and one female, died prior to tissue samples being collected. Incomplete sampling has obvious implications for parental pair assignment success although the possibility still exists to make either a maternal or a paternal assignment.

3.3.3.2 Tissue Sampling of Abalone Offspring

No tissue samples were lost therefore sampling was complete for all four hundred F1 progeny across Abagold and HIK. Genotypes could be generated for all individuals although scoring of alleles was not always possible across all loci typed for an individual.

3.3.4 Genomic DNA Extractions

3.3.4.1 Genomic DNA Isolation and Quantification of Abalone Breeding Stock

All DNA concentrations were quantified on a NanoDrop® ND-1000 spectrophotometer. DNA extractions using a standard CTAB extraction method produced variable results with lower purity than preferred. Despite the samples not being "pure", there were no apparent problems with PCR amplification as a result of these contaminants; although, mismatches in juvenile and parental genotypes, that may be attributable to partial nulls or allele dropout, could result from less than adequate DNA quality.

3.3.4.2 Genomic DNA Isolation and Quantification of Abalone Offspring

Aside from the initial manual preparation of an overnight digestion step, all DNA extractions for offspring were completed through a robotic platform. No DNA samples were lost during the extraction process and quantification was completed using a NanoDrop® ND-1000 spectrophotometer. Samples were of adequate concentrations and purity for subsequent PCR reactions.

3.3.4.3 Comparison between Extraction Methods and PCR Results

Three DNA extraction methods were compared: a standard CTAB method, extraction using phenol-chloroform and a Chelex®100 protocol. Concentrations are shown in Table 9 along with the absorbance ratio for 260/280 and 260/230 wavelengths. The ratio of sample absorbance at wavelengths 260nm and 280nm (i.e. 260/280) provides an indication of the purity of the sample,

with a ratio of approximately 1.80 for DNA indicating a pure sample with relatively few contaminants. A 260/280 ratio lower than 1.8 can mean one of two things: Firstly, that the sample contains contaminants such as proteins; or secondly, that the DNA was not solubilised completely when the reading was taken. A ratio higher than 2.0 is evidence of contaminants such as phenol or chloroform. The ratio 260/230 shows the sample absorbance at 260nm and 230nm respectively. This is a secondary measure of DNA purity and in the instance of "pure" DNA these values are often higher than the 260/280 ratios, commonly within the range of 1.8-2.2. Much lower values may indicate the presence of co-purified contaminants.

Table 9 Concentration and genomic DNA purity comparison of three extraction methods tested

Sample ID	Extraction Method	ng/µl	260/280 ratio	260/230 ratio
1	CTAB	11.89	2.01	1.25
2	CTAB	119.81	1.93	1.71
3	CTAB	128.11	1.93	1.72
4	phenol/chloroform	175.64	1.94	1.98
5	phenol/chloroform	6.96	1.81	0.41
6	Chelex®100	62.95	1.45	0.31
7	Chelex®100	80.22	1.33	0.26
8	Chelex®100	89.86	1.38	0.31

The gDNA concentrations from CTAB extraction were variable across the three samples. Both absorbance values indicate contaminants such as chloroform-isoamyl alcohol which may not have been effectively removed.

One sample from the phenol-chloroform extractions was lost during the initial stages of the protocol due to handling error. The remaining two also show very variable concentrations, with a minimum of 6.96ng/µl and maximum of 175.64ng/µl. Absorbance values also indicate likely contaminants such as phenol or chloroform.

Chelex®100 extractions provide the most constant concentrations across the three samples with a minimum of 62.95ng/µl and maximum of 89.96ng/µl. Low average absorbance values, however, indicated that these samples were the least pure, containing contaminants such as proteins.

Single primer PCR reactions run on the samples prior to standardization of DNA concentrations indicated that all three amplify successfully for microsatellite HmD11 $(TCTG)_8$. Fortunately the combination of microsatellites and PCR lends itself to amplification of degraded DNA and/or less

pure DNA despite these potential hindrances; one of the reasons for this being the molecular marker of choice in various applications (Webster & Reichart 2005). Figure 10 shows the comparative results of PCR amplification products produced from all gDNA samples tested.

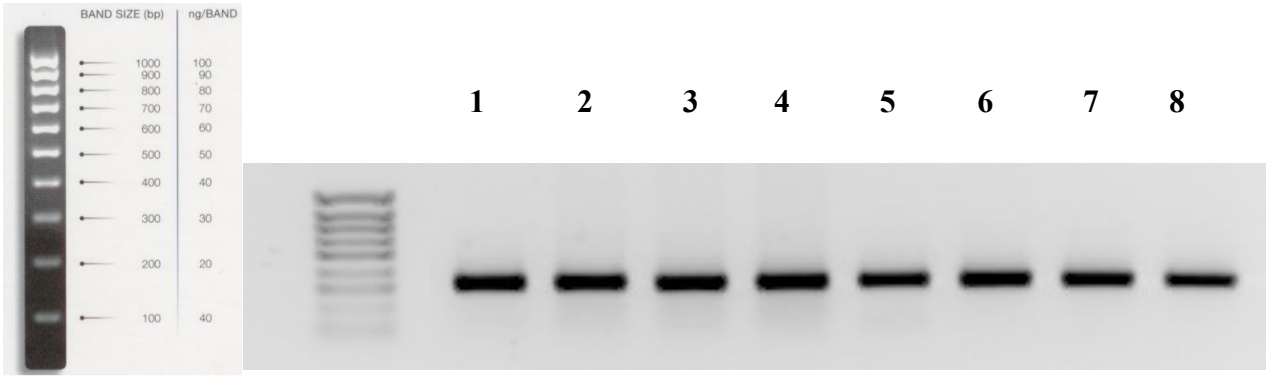


Figure 10 Single primer PCR results using samples from three extraction methods: 1-3 CTAB; 4-5 Phenol-Chloroform; 6-8 Chelex®100. All samples were amplified with a primer set representing microsatellite HmD11 (NCBI AY303341) with expected fragment range of 292-352bp

PCR on all DNA samples resulted in amplified product; producing fragments of approximately 300bp. To confirm results, two DNA samples from each method were subsequently diluted to an approximate concentration of 20ng/µl. A multiplex reaction incorporating primer sets for microsatellites HmD55 $(GTGA)_{12}$ and HmD59 $(CA)_{15}$ was then carried out on these selected samples using a Touch-Down PCR program. Results are shown in Figure 11. The expected size ranges were 183-211bp and 106-150bp respectively.

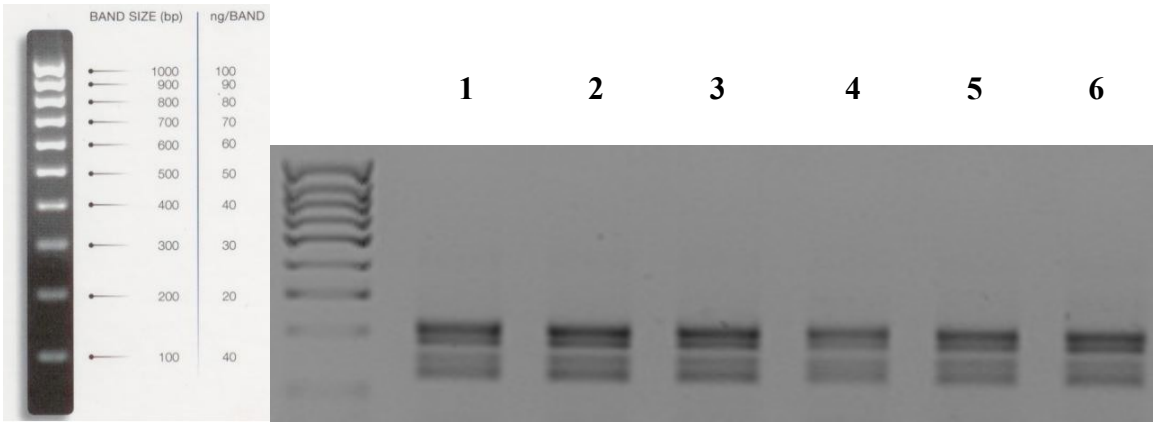


Figure 11 The multiplex PCR results are shown in the order: 1-2 CTAB; 3-4 Phenol-Chloroform; 5-6 Chelex®100. PCR was run using primers representing microsatellites HmD55 (NCBI AY303337) and HmD59 (NCBI AY303338). HmD55 produced fragments in the region of 200bp with those generated by HmD59 being slightly smaller at about 150bp

Amplification was consistent across all three extraction methods producing equally positive results and from this it would seem that the Chelex®100 resin may be an effective substitute for either of the other two methods employed regularly within our research group. The Chelex®100 protocol appears to offer the following benefits – time saving, effective and cost-effective – and its future utilization would obviously be at the discretion of the researcher and any constraints applicable within a particular study.

3.3.5 Shortcomings and Sources of Error

(1) Loss of breeding animals contributing to the Performance Recording Scheme, particularly Abagold where tissue samples were not available for these deceased animals. Incomplete sampling may lead to an increased percentage of unassigned offspring and inaccurate estimation of the breeding stock contributions to spawning. In addition, these animals are no longer available for inclusion in future breeding programs should they be found to be making significant contributions.

(2) The variability of concentrations achieved and the fairly low levels of purity realised for some samples. This may impact on genotyping should a combination of molecular markers be used for genotyping, especially with respect to QTL mapping and the development of linkage maps. These markers may require better quality DNA samples.

(3) The bee-tagging method was trialled for the first time so no prior experience as to their success or anticipated tag loss was known. Should the loss be significant, the re-measurement of these PRS animals at two years will be impossible thus hindering comparative results in Phase II.

(4) Only the upper 10% and lower 10% of offspring from one replicate were selected. This amounted to small sample sizes of 100 animals in each group for each farm of origin. The question exists as to whether these sample groups are completely representative of the larger population despite the random sampling employed.

3.4 Discussion

The current status of genetic improvement for *Haliotis midae* is lagging behind that of other countries commercially exploiting abalone so speed to market for the local industry is paramount. Laying the groundwork of a project is often the most laborious and costly component but is a crucial element in determining the success of the project as a whole. Comparative genomics is growing in application (Baranski *et al*. 2006; Sekino & Hara 2006) but is relevant only when Type I markers have been established and hence the location of genes is under investigation. A high level

of species specificity exists otherwise and despite research findings published for other species of abalone, such as heritability estimates (Lucas *et al*. 2006), loss of genetic variation at microsatellite loci (Evans *et al*. 2004; Li *et al*. 2004) as well as the first linkage maps for two abalone species, *H. rubra* and *H. discus hannai* (Baranski *et al*. 2006; Liu *et al*. 2006; Sekino & Hara 2006); a large amount of work and preparation needs to be completed in order to better understand our indigenous species, *H. midae*, to bridge the current gap and to gain a competitive advantage. To this effect, a Performance Recording Scheme was established with the collaboration of local industry providing an F1 generation of offspring for initial assessment and selection for growth. These animals are tracked by ongoing monitoring for growth benefits given that heritability of growth rate (which is defined as the genetic contribution to the total variance observed for a phenotypic trait) increases with age. This observation regarding heritability was made by Jonasson *et al*. (1999) for red abalone (as cited in Robinson *et al*. 2003). That size at an early age may be a poor predictor of later performance is further motivation for basing early selection on pedigree information and it highlights the need to monitor the animals for the duration of the PRS (from juvenile to maturity) so as to provide support, or contrary evidence, to the potential breeders selected for MAS on the basis of earlier assessments of juveniles.

One of the strengths of the PRS is the ability to rear animals from a specific origin/farm at five different locations, thereby minimizing the environmental influences to some extent. Management practices such as feeding strategies, design of rearing tanks, use of water re-circulation or on-shore pump systems and other husbandry methods were specific to the farms and allowed to continue as normal without any standardisation of these factors.

In a study performed by Jerry *et al*. (2006a) on shrimp families reared in commercial ponds, a pond effect on the growth of animals in one of the four rearing/grow-out ponds was observed. This resulted in the mean weight of individuals from this population being lower than for the other three ponds.

Within this study, the differences observed between animals in a growth group reared at separate locations were marginal (not statistically significant) and hence no conclusions can be drawn based on this preliminary performance assessment as to whether these same animals will maintain their faster or slower growth. The differences observed for animals at HIK, Aquafarm and Roman Bay when compared to the remaining farms may signify that there are environmental effects providing a slight advantage/disadvantage to the different growth groups at these locations; an indication that these effects may not be entirely standardised through the PRS. The fact that these differences are fairly small, however, suggests that these effects, although present, may not be contributing significantly. The effects observed for HIK juveniles were very similar to those obtained for the

Abagold groups and given the lack of statistical significance, no sound inferences could be made regarding the breeding population or the ongoing growth performance of the offspring. The environmental impacts may be inconsequential enough not to confound the assessment of potential genetic effects.

The animals will be reassessed at two years and then again closer to sexual maturity to determine whether they maintained their faster or slower growth and these results will form part of a separate study within the research group. Environmental effects may need to be monitored further and possibly factored in to future *H. midae* selective breeding programs. It is also evident from the findings that factors that can be controlled or standardized within the PRS should be managed accordingly so as to reduce as many variables as possible. This makes scientific sense when setting up an unbiased experimental design since multiple variables will otherwise confound the factor under investigation. Standardisation of variables should allow any growth advantages resulting from genetic factors to be better represented and hence distinguishable from environmental effects. It is clear that growth benefits are as a result of both genetic factors and favourable environmental conditions but the latter is not being evaluated in this study and needs to be controlled if it is realistic to do so.

The average weight and shell length recorded for selected juveniles at approximately one year old for both Abagold and HIK, compared more than favourably with average shell lengths (18.1mm) and weights (1.47g) recorded for *H. asinina* offspring of the same age (Lucas *et al*. 2006); a species of abalone reported to be one of the fastest growing, reaching sexual maturity at one year old. For Abagold these measurements were 28.48mm and 4.0g and for HIK, 30.01mm and 4.57g respectively. It was further noted that HIK juveniles, on average, particularly for the faster growth group, were slightly heavier and longer at one year of age than Abagold offspring. This was seen across all five rearing locations; however, although visually apparent, differences were not considered to be statistically significant. Genetic variation exists within and among individuals and despite breeding stock for Abagold and HIK being harvested from the same natural population, the genetic variation inherent in the HIK population may be slightly higher than for the Abagold population providing a slight fitness advantage within the F1 generation of HIK. A comparison was made of allelic richness within these two hatchery populations and observations presented as Addendum A.

Figure 8 shows the structure of Phase I of the PRS utilised in this study. At each location, the six hundred animals supplied by a particular farm were further subdivided into three replicates, each containing two hundred offspring. When selecting candidates for the faster and slower growth groups, the upper 10% and the lower 10% of one replicate only were evaluated, providing twenty

individuals within each group at each location. This amounted to a total of two hundred animals out of the original three thousand that were subsequently evaluated. These were relatively small sample sizes when compared to a parentage study conducted by Jerry *et al*. (2006a) in which the original sample size was one hundred and thirty-two thousand individuals. The upper 10% of each pond's weight distribution, up to three hundred animals per pond were genotyped to assess the representation of families within this upper weight class.

The size of adult abalone in relation to the size of adult shrimp are obviously not comparable nor are the required rearing conditions for each species; hence the scope for stocking a much larger number of juveniles into 1ha ponds may have been more feasible with the latter study. However, a potential drawback of the small sample sizes evaluated for growth in this study, although allowing for contribution of breeders to be assessed, is that no significant conclusions can be drawn regarding the impact of the breeders' contribution to the next generation. In both studies, the upper 10% and lower 10% of individuals was essentially sampled, but the differences were the effective population size (Ne) from which each sample was drawn and that the current study evaluated one replicate only while the shrimp study took animals from each pond/replicate. The downside of this latter, more extensive method was that it ultimately provided more animals to be genotyped with the associated increase in project costs and time required to run the tests, but in saying that, possibly provided a more representative sample for which more accurate conclusions could be drawn.

Consideration should be given to sampling 10% of candidates for each growth group from each replicate across all locations. This strategy would then provide a total of six hundred individuals per farm of origin (three hundred per growth group). The implication would obviously be increased genotyping costs but these could potentially be offset by optimising the following factors:

Cost of DNA extractions through the use of robotic platform and Chelex®100 (Walsh *et al*. 1991; Hoff-Olsen *et al*. 1999; Zilberman *et al*. 2005; Castro *et al*. 2006) as opposed to using more costly commercial kits or time and resource demanding techniques such as phenol-chloroform extraction. High throughput DNA extraction would, in addition, reduce the time taken to complete extractions thus freeing up resource for molecular marker development and screening.

Reducing the number of microsatellite loci needed through the selection of markers with a very high degree of informativeness; a prerequisite to assign parentage with an acceptable level of success. This will become more feasible given the increasing pool of markers being developed within the research group from which to select. Studies have shown that five highly polymorphic loci may be sufficient to assign parentage (Selvamani *et al*. 2001; Jerry *et al*. 2004; Lucas *et al*. 2006) but this would obviously be dependent on the number of candidate parents and the number of offspring to be assigned. The isolation and development process for microsatellite markers requires

optimisation and suggestions for pre-determining unique clones thereby reducing the number of redundant clones currently being sequenced, are elaborated on in the discussion of chapter four.

Juveniles selected for growth were physically tagged using bee-tags but no prior trials had been conducted using this method of tagging. The establishment of genetic tags will eventually circumvent the need for physical tagging but in the initial stages of the PRS, these were still necessary. Tag loss still needs to be quantitatively assessed but if it is as substantial as the previous method using silicone tubing, this would mean that a large number of the original two hundred offspring measured for each farm may no longer be traceable for follow-up measurements. Physical tagging methods have been employed in numerous other studies on abalone yet details regarding the type of tag used and the application thereof are not documented in the publications; neither is the degree of tag loss experienced using these particular methods. It may be worthwhile to correspond with other researchers working on abalone and who are at various stages within selective breeding programs to ascertain how they may have overcome the limitations of physical tagging methods. Tag loss would have serious implications going forward into Phase II of the PRS and solutions should be sought sooner rather than later.

Concentrations of extracted DNA samples were highly variable as was the purity of the DNA samples. The most commonly observed source of contamination in breeding stock samples was the presence of chloroform or isoamyl given that 260/280 ratios were greater than 1.8. A comparison was subsequently made between three extraction methods: standard CTAB, phenol-chloroform and Chelex®100, the latter method being used in an increasing number of studies (Li *et al*. 2003; Hoffman & Amos 2005; Castro *et al*. 2006). It was determined that Chelex®100 resin may be a quicker, cheaper alternative to the standard CTAB method which is more time-consuming, costly and resource intensive. Contaminants present in the samples extracted using Chelex®100 appeared to be proteins.

Since comparisons were very limited in terms of the samples tested, it is recommended that a larger number of samples be analysed and that these samples are also genotyped using a range of microsatellite markers to assess the quality of the fingerprint and the ease of allele scoring. Poor DNA quality is known to be a cause of null alleles due to inconsistent PCR failure (Dakin & Avise 2004) and the potential for this phenomenon to increase using Chelex®100 extracted samples, which may contain protein contaminants, needs to be assessed prior to adopting this method.

Only two animals were lost prior to collection of tissue samples (from the Abagold population) and this loss may not be sufficient enough to warrant reconstruction of genotypes in order to improve assignment success or ascertain whether these two adults (one male and one female) made significant contributions to spawning. The additional work and cost involved in genotype

reconstruction would need to be weighed up against the potential benefits of having a complete sample set of genotypes for Abagold from which to select potential breeders; especially in light of the fact that these animals are no longer available to future breeding programs.

The results discussed in this section indicate that there is definitely room for improvement with regards to the structure of the PRS going forward as well as DNA extraction protocols. The obvious benefits of reducing the costs of routinely applied yet fairly large-scale procedures such as genomic DNA extractions, is that a greater proportion of a projects budget is available to conduct value-adding research. Results achieved, however, were adequate enough for progression to the next study objective of microsatellite isolation and development.

Chapter Four

Microsatellite Marker Development

Abstract

Out of a total of 909 potentially recombinant clones, 50% were found to contain inserts ranging from 150bp to 750bp. Sequencing of these 462 positive clones produced 16% perfect repeat sequences, 47% imperfect/interrupted repeats sequences, 31% having no repeats evident while 6% were undetermined due to poor sequence quality. Fifty-two primer sets were designed yielding thirty-one polymorphic microsatellite loci of different motifs and composition. To determine the suitability of a marker set for parentage analyses, a set of summary statistics for each locus is required providing the allele number, allele frequency, observed (Ho) and expected (He) heterozygosity, the polymorphic information content as well as the potential null allele frequency. Since a high frequency of null alleles segregating with a locus can seriously impede the success of parentage assignments, it is best to select those loci with a frequency below 5% (Marshall *et al*. 1998). A wild abalone population was screened with a set of eleven polymorphic markers to determine their informativeness and thus suitability in assigning parentage within two hatchery populations. Five loci met the requirements and were selected as part of a set of nine markers to be used in genotyping PRS juveniles and breeders from Abagold and HIK. These loci were: HmNS38, HmNST7, HmNS56, HmNS28 and HmNS58 all of which had: allele numbers above 10.0, PIC values greater than 0.82 (except for HmNS38 which was at 0.50), observed heterozygosity greater than 0.74 and null allele frequencies below the critical value of 5%.

4.1 Introduction

With the advent of microsatellite molecular marker analyses in the late 1980's, came the realisation that these may well be the most powerful neutral, codominant Mendelian markers as yet discovered. Since then, these loci have been widely applied in a range of studies; from their potential association to genetic diseases, suitability for linkage mapping and MAS breeding programs, effectiveness in kinship and parentage investigations as well as within population studies. Their usefulness within the field of population genetics, however, remains an ongoing debate given the null alleles that tend to segregate with some of these loci (Jarne & Lagoda 1996). Microsatellite molecular markers were developed to facilitate the delivery of the primary objective of this study, parentage assignment; and chapter four maintains the holistic approach initiated in chapter three, documenting the materials and methods used to isolate, develop and assess these microsatellite loci, together with the development of suitable multiplexed reactions to screen thirty-two individuals from a wild abalone population harvested at Black Rock, along the east coast of South Africa.

The wild population screening was employed as means to test the microsatellite markers under consideration and to select the best loci for subsequent parentage analysis of two hatchery

populations. All relevant results are presented together with identified shortcoming and sources of error. A brief discussion of the results and suggestions for addressing the shortcomings, where applicable, is also included.

4.2 Materials and Methods

4.2.1 Microsatellite Isolation

The Fast Isolation by AFLP of Sequences Containing Repeats (FIASCO) method (Zane *et al*. 2002) was used to isolate species-specific polymorphic microsatellite markers from genomic DNA of wild *Haliotis midae* abalone populations across South Africa. In total, two enrichments were performed for each of the following repeat types in the individuals indicated: dinucleotide repeat $(AC)_n$ in individuals 22 and 37 (collected from Gansbaai, 2003); trinucleotide repeat $(CAA)_n$ in individual 9 (collected from Gansbaai, 2003) and individual 21 (collected from Witsand, 2004); tetranucleotide repeat $(GTGC)_n$ in individual 13 (collected from Gansbaai, 2003) and individual 47 (collected from Rietpoint, 2003). The biotinylated oligonucleotide probes used to isolate the various repeat sequences were supplied by Integrated DNA Technologies Inc. (*IDT*) in stock concentrations of 200µM from which 10µM working stocks were prepared through dilution with sterile distilled water. Working stocks of probes were stored at -20˚C.

4.2.1.1 Genomic DNA from Wild Abalone Populations

The six individuals selected (see 4.2.1) were chosen at random based on the high concentration and relative purity of their genomic DNA (gDNA). DNA extractions for the wild population samples were processed by Mrs Aletta van der Merwe (Genetics Department, Stellenbosch University) using a phenol-chloroform extraction protocol.

Extracts were quantified through agarose gel electrophoresis. Undiluted gDNA samples were resolved on 2% (w/v) agarose gel against a Lambda (λ) standard (250ng/µl).

The gels were run at 100V allowing for slow resolution and hence clearer band separation. Visualisation of the DNA was achieved through the addition of 1.5µg of EtBr per millilitre gel mixture – i.e. 0.03% (v/v) EtBr stock solution (10mg/ml) and viewed under ultra violet light (UV). The intensity of the band of gDNA was measured relative to the intensity of the standard and an estimation of the concentration in ng/µl was made for each sample.

Samples with a high concentration and relatively few impurities, evident as a thick smear or band running ahead of the band of gDNA, were selected. The concentrations of undiluted aliquots of the selected individuals were reassessed on a NanoDrop® ND-1000 spectrophotometer and working stocks of approximately 250ng prepared. Working stocks were stored at -20˚C.

4.2.1.2 Buffer Stock Solutions for FIASCO Protocol (Zane *et al*. 2002)

The following stock solutions were prepared in advance and stored at room temperature: 1M acetic acid, 1M sodium hydroxide (NaOH), 1M Tris buffer (Tris-HCl pH 8.0), 5M sodium chloride (NaCl), 0.5M ethylene-diamine-tetra-acetic acid (EDTA pH 8.0), 10% (w/v) sodium dodecyl sulphate (SDS), 5X Tris-borate-EDTA buffer (TBE pH 8.0), 10X Tris-EDTA buffer (TE pH 8.0) and 20X SSC (3M NaCl, 0.3M Sodium-Citrate pH 7.0). Details regarding buffer preparation are provided in Appendix 5.

4.2.1.3 Digestion and Ligation of Genomic DNA followed by PCR Amplification

*Mse*I restriction enzyme digestion and subsequent ligation with *Mse*I-adaptors was performed in a single 25µl reaction: 250ng gDNA, 1X OnePhor All Buffer, 5mM DTT, 50µg/ml BSA, 1µM *Mse*I-adaptors, 200µM ATP, 2.5U *Mse*I restriction enzyme and 1U T4 DNA Ligase. All reagents were combined in a microcentrifuge tube and *Mse*I restriction enzyme and T4 DNA Ligase kept on ice and added to the reaction just prior to the incubation step. The mixture was incubated in a GeneAmp® PCR System 2700/2720 (*Applied Biosystems*) for three hours at 37˚C.

A 10X (1:9) dilution was made of the mixture using sterile distilled water prior to AFLP PCR amplification. PCR amplification was achieved using *Mse*I-adaptor specific primers also referred to as *Mse*I-N primers. This primer cocktail contained all possible "selective" bases at the 3′- end that could match the first nucleotide occurring after the original restriction site.

The AFLP-PCR reaction was performed in a 20µl volume: 5µl ligation mixture, 120ng *Mse*I-N primer cocktail, 1X *Taq* polymerase buffer, 1.5mM $MgCl_2$, 200µM of each dNTP (prepared as a cocktail) and 0.4U Go*Taq* Flexi DNA polymerase.

The PCR was run using the following program: 94˚C for 30s, 53˚C for 60s and 72˚C for 60s and amplification was repeated for a range of cycles (14, 17, 20, 23 and 26) to determine the optimum number of cycles required to produce a size range of amplified product. Go*Taq* Flexi DNA polymerase and all associated PCR reagents were supplied by *Promega*.

Products were resolved using 2%, 1X TBE agarose gel run at 100V for forty-five minutes and visualised under UV light, with a smear indicating a positive result. The PCR reaction was rerun at the optimized cycle number to provide sufficient yield of amplified fragments and amplification samples were stored at 4˚C.

4.2.1.4 Enrichment through Hybridization with Biotinylated Probes

Amplified DNA from the six wild abalone individuals was hybridized using biotinylated probes (refer to section 4.2.1). The microsatellite enrichment step was carried out in a 100µl volume: 250-

500ng amplified DNA (20µl AFLP PCR reaction), 50-80pmol biotinylated probe, 4.2X SSC (42µl), 0.7% (w/v) SDS (7µl) and sterile distilled water to volume. Denaturation was performed in a PCR Thermal Cycler at 95˚C for three minutes and the annealing step at room temperature for twenty-five minutes. To dilute the hybridization mixture, 300µl TEN100 (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 100mM NaCl pH 7.5) was added before storing at 4˚C.

4.2.1.5 Selective Capturing of DNA-Probe Hybrids and PCR Amplification

DNA molecules hybridized to biotinylated probes were selectively recaptured using streptavidin coated beads (Streptavidin Magnetic Particles, *Boehringer Mannheim*). A volume of 100µl (1mg) of beads was washed four times in an equal volume of TEN100 buffer. After each wash the beads were held separate from the solution through the application of a magnetic field and the buffer pipetted off. After the final wash the beads were re-suspended in 40µl fresh TEN100 buffer and 10µl of non-related PCR product (source: plant DNA or human DNA) was added to minimize non-specific binding in subsequent steps.

In the next step, using a 1.5ml eppendorf tube, the following were combined: 1mg prepared beads (40µl beads solution and 10µl non-related PCR product), 400µl DNA-probe hybrid solution (100µl AFLP reaction mix and 300µl TEN100) and 300µl TEN100. The mixture was incubated at room temperature for thirty minutes with the application of constant and gentle agitation using a Vortex Genie 2. The above beads-probe-DNA complex was separated from the buffer through the application of a magnetic field and the buffer was discarded. Non-specific unbound DNA was subsequently removed through three non-stringency washes followed by three stringency washes.

The non-stringency washes were performed through the addition of 400µl TEN1000 (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 1M NaCl pH 7.5) to the beads complex with gentle mixing for five minutes at room temperature. A magnetic field was applied to separate the beads from the buffer and the supernatant from the first two washes was pipetted off and discarded. The supernatant from the final wash was kept and stored at -20˚C for later use (NS).

The stringency washes encompassed the addition of 400µl 0.2X SSC and 0.1% (w/v) SDS to the complex with gentle mixing for five minutes at room temperature. A magnetic field was applied to separate the beads from the solution and the buffer from the first two washes discarded as previously. The supernatant from the final wash was again kept and stored at -20˚C for later use (S).

DNA was separated from the probe-DNA complex by two denaturation steps. The first denaturation was completed by adding 50µl 1X TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0) to the beads and incubating for five minutes at 95˚C using a hot-block (*Lasec*). The supernatant was removed and stored *immediately* at -20˚C for later use (D1). The second

denaturation was done by adding 15µl 0.15M NaOH to the beads. The supernatant was removed and neutralised through the addition of 1µl 0.1667M acetic acid and 34µl 1X TE buffer to a final volume of 50µl prior to storing at -20˚C (D2).The streptavidin beads were discarded after these steps.

DNA was precipitated from the four supernatants (NS, S, D1, and D2) through the addition of one volume cold isopropanol and 0.15M sodium acetate. The supernatants were left overnight at -20˚C to facilitate precipitation and then centrifuged at 13,200 rpm for thirty minutes. The supernatant was poured off carefully so as not to dislodge the pellet; the DNA resuspended in 50µl sterile distilled water and then stored at -20˚C.

A second AFLP-PCR was run for each of the four DNA suspensions in a 20µl reaction volume: 2µl DNA suspension, 120ng *Mse*I-N primer cocktail, 1X *Taq* polymerase buffer, 1.5mM $MgCl_2$, 200µM of each dNTP and 0.4U Go*Taq* Flexi DNA polymerase. The PCR was run on the following program: 94˚C for 30s, 53˚C for 60s and 72˚C for 60s. The program was repeated for 30 cycles with 15˚C ∞ added for overnight reactions.

Products were resolved on 2% agarose gel run at 100V for forty-five minutes against a HyperLadder IV (*Bioline*) and visualised under UV light. A smear of fragments greater than 200bp indicated the desired result. Ideally, the PCR of the final stringency wash suspension (S) should not yield any product.

D1 is expected to contain the majority of the enriched fragments and post-PCR cleanup was therefore performed on suspension 3 only (D1) using SigmaSpin™ Post-Reaction Purification Columns (*Sigma*). This was completed following the supplier protocol. The purification column removes small molecules such as primer-dimers and salts. Samples were run post clean up on 2% agarose gel as described previously to ensure sufficient PCR product was still available to proceed with ligation and transformation into competent cells. Samples were stored at 4˚C.

4.2.1.6 Cloning through Ligation and Transformation of PCR Products

Cloning of PCR product was carried out in the pDrive Cloning Vector (3.85kb) supplied by QIAGEN® (Appendix 6) through a three step process: (1) Ligation of PCR product into pDrive Vector; (2) Transformation of competent cells with pDrive Vector; (3) Plating of competent cells onto Luria Bertani medium (LB) plates for overnight colony culture and screening through blue-white selection of recombinant clones. Ten LB medium plates were prepared per transformation: 1% (w/v) Bacto-Tryptone, 0.5% (w/v) Bacto-Yeast, 1% (w/v) NaCl and 1.5% (w/v) Bacterial Agar, adjusted to pH 7.0 with NaOH. LB medium was autoclaved and cooled to approximately 55˚C in a waterbath before addition of: 100µg/ml ampicillin (*Roche Applied Science*), 50µM isopropyl β-D-

thiogalactopyranoside (IPTG) (*Sigma*) and 80µg/ml 5-bromo-4-chloro-3-indolyl β-D-thiogalactopyranoside (X-Gal) (*Southern Cross Biotechnology* (*Pty*) *Ltd*). Plates were prepared in a laminar flow cabinet either the same day or one day prior to cloning. If prepared the previous day, plates were individually sealed with parafilm, group wrapped with clingwrap and stored overnight at 4˚C. It is recommended that long term storage of plates (> one week) be avoided due to the possibility of decreased antibiotic activity (QIAGEN® PCR Cloning Handbook, 2001).

All reagents for the ligation reaction were mixed thoroughly to avoid localised salt concentrations and kept on ice during preparation of the 10µl reaction mixture: 1µl pDrive Cloning Vector (50ng/µl), 4µl PCR product and 5µl 2X Ligation Master Mix. All reagents were supplied by QIAGEN® and were added in this order as recommended by the supplier. The ligation mixture was mixed briefly and then incubated for two hours at 4˚C.

An extended incubation time was applied to increase the number of recombinants but this may be reduced to a minimum of fifteen minutes. The ligation mixture was either stored at -20˚C until the following day or used immediately in the transformation.

In preparation for competent cell transformation, SOC medium [2% (w/v) Bacto-Tryptone, 0.5% (w/v) Yeast Extract, 0.05% (w/v) NaCl, 10ml 250mM KCl, 18ml 20% (w/v) sterile Glucose, 5ml 2M sterile $MgCl_2$, pH 7.0] was thawed to room temperature. QIAGEN® EZ competent cells were thawed on ice and 2µl ligation mixture added, mixed gently and incubated on ice for five minutes. The cells were then heat-shocked for thirty seconds at 42˚C in a waterbath and incubated again on ice for two minutes.

SOC medium (250µl) was added to the competent cells mixture, mixed thoroughly and 50µl pipetted directly on to each LB plate. Individual plates were sealed with parafilm, inverted and incubated overnight at 37˚C. A second incubation of a few hours at 4˚C was later performed to enhance the blue colour development thereby facilitating the differentiation between white (positive) and blue (negative) recombinants.

4.2.2 Microsatellite Screening

Screening of positive clones was done through direct PCR using vector primers (M13 F/R or T7/SP6) with visualisation of PCR products through agarose gel electrophoresis to determine those vectors containing abalone DNA inserts. Sequencing was completed for those vectors positive for an insert; screened for microsatellite repeat sequences followed by the design of primers in the DNA regions flanking the repeat should positioning of the insert in the vector have allowed for primer design.

4.2.2.1 Direct Amplification of Positive Clones through Colony PCR

Single white colonies were removed from the plates using sterile toothpicks and transferred to microcentrifuge strip-tubes containing 10.7µl sterile distilled water. Vector primers were used in a 20µl reaction to screen the clones for fragment inserts: DNA (single colony), 05µM of each primer (M13F/ T7 and M13R/SP6), 1X *Taq* polymerase buffer, 1.5mM $MgCl_2$, 200µM of each dNTP and 0.4U Go*Taq* Flexi DNA polymerase. A colony PCR program was utilised: 94˚C for 10 minutes; twenty-five cycles of 94˚C for 60s, 55˚C for 60s and 72˚C for 60s with a final extension step of 72˚C for 10 minutes with PCR products resolved on 2% agarose gel against HyperLadder IV and visualised using UV light. Only samples showing single bands greater than 300bp were considered to contain abalone DNA fragments and these PCR products were cleaned using SigmaSpin™ Post-Reaction Purification Columns, quantified on the NanoDrop® ND-1000 spectrophotometer and stored at -20˚C until sequencing.

4.2.2.2 Automated Direct Sequencing of Selected PCR Fragments

Direct sequencing of PCR fragments greater than 300bp was performed in a 10µl reaction: 10-20ng/µl DNA, 1.6µM primer (either forward or reverse primer) and 2µl BigDye® Terminator v3.1 (*Applied Biosystems*) using the following PCR program: 95˚C for 2 minutes; twenty-five cycles of 95˚C for 10s, 55˚C for 10s and 60˚C for 4 minutes and 15˚C ∞ (for overnight reactions). Samples were submitted for analysis to the Central DNA Sequencing Facility at the Genetics Department, Stellenbosch University.

Samples were sequenced on an ABI 3130 X L Automatic Sequencer using POP 7 and 50cm capillaries and against a LIZ600® size standard (*Applied Biosystems*). Screening and editing was performed using CHROMAS 1.45 (http://www.technelysium.com.au/chromas14x.html) to remove vector DNA sequence. In some instances, clones were sequenced in both directions in order to cover as much of the fragment insert as possible. This was especially necessary when long microsatellite repeats or very large fragment inserts were present. The edited forward and reverse sequences were aligned in BIOEDIT Sequence Alignment Editor Version 7.0.5.3 and consensus sequences created (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Both CHROMAS 1.45 and BIOEDIT v 7.0.5.3 are available to researchers as free software.

All sequences containing repeats were checked online against the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nih.gov/BLAST/) and by means of a local blast against all microsatellite loci in the process of development within the research group so as to avoid duplication. Text files were created for all sequences containing repeats and used further for the development of primers.

4.2.3 Primer Development

4.2.3.1 Design of PCR Primers

Microsatellite primer sets were designed using one of two software options in order of preference: OLIGO 4.1 (Rychlik & Rhoads 1989) Software or Primer 3 Input 0.4.0 (https://sourceforge.net/project/showfiles.php). The latter program is an online based program which automatically generates a range of primers sets listed in order of suitability based on a set of parameters selected by the user. Primer sets can then be checked individually for specific criteria and the most appropriate selected. General guidelines for primer design, as per in the QIAGEN® PCR Cloning Handbook, 2001 are provided in Appendix 7.

Unlabelled forward and reverse primers were ordered from *Whitehead Scientific* (*Pty*) *Ltd*. Primer sets were received in powder form, centrifuged for 20 minutes at 13,200 rpm and re-suspended using sterile distilled water to 100µM stock solutions. Primary stocks were stored overnight at room temperature or 4˚C to allow for complete dissolution before working stocks of 10µM were prepared by dilution (10X). All stocks were stored at -20˚C. A list of all primer sets developed in this study including status (i.e. either polymorphic or monomorphic), testing conditions and NCBI accession number are provided in Appendix 8.

4.2.3.2 Primer Testing for Detection of DNA Polymorphisms and Optimization

To assess the ability of designed primer sets to detect microsatellite polymorphisms, DNA from individuals 1-8 from the Black Rock wild abalone population was used. DNA working stocks of approximately 20ng/µl were prepared and stored at -20˚C. Primer sets were tested initially on a PX2 Gradient Thermal Cycler set to cover a range of temperatures (50.1˚C to 65.5˚C) to determine at which temperature optimum amplification occurred. Only one individual was used for temperature testing in a 10µl reaction volume: 20ng DNA (20ng/µl), 0.2µM of each primer (forward and reverse), 1X green *Taq* polymerase buffer, 2mM $MgCl_2$, 200µM of each dNTP and 0.25U Go*Taq* Flexi DNA polymerase on PCR program: 94˚C for 5 minutes; thirty cycles of 94˚C for 30s, *(50˚C-65.5˚C) for 30s (* varied across *Ta* gradient), 72˚C for 60s, and a final extension step of 72˚C for 10 minutes.

All primer sets showing successful amplification of the correct fragment size were rerun at their optimum annealing temperature (as determined above) on Black Rock individuals 1-8 using a standard PCR reaction of 15µl: 20ng DNA (20ng/µl), 0.2µM of each primer (forward and reverse), 1X green *Taq* polymerase buffer, 2mM $MgCl_2$, 200µM of each dNTP and 0.25U Go*Taq* Flexi DNA polymerase.

The PCR program used was: 94˚C for 5 minutes; thirty cycles of 94˚C for 30s, *optimal *Ta* for 60s (*primer-specific *Ta* temperature), 72˚C for 60s, and a final extension step of 72˚C for 10 minutes. Product from these standard reactions was run on 2% agarose gel to confirm amplification and then on 12% polyacrylamide (PAA) gels for resolution of polymorphisms if present.

Any primer sets that failed to produce amplified product on a gradient cycle were either re-designed (if for a perfect repeat sequence) or discarded (if an imperfect repeat sequence). If low amplification was achieved on a gradient cycle for some temperatures, primer sets were tested using a Touch-Down (TD) PCR strategy (Rahman *et al*. 2000) with the same reaction mix concentrations as for the gradient or standard PCR.

The TD-PCR program is detailed in Appendix 4. If no result was obtained on a TD-PCR, primers were either re-designed (perfect repeats) or discarded (imperfect repeats). All unique polymorphic primer sets were submitted to *GenBank* for assignment of accession numbers (http://www.ncbi.nlm.nih.gov/Genbank/submit.html).

4.2.3.3 Polyacrylamide Gel Electrophoresis

Resolution of PCR products using Polyacrylamide Gel Electrophoresis (PAGE) was carried out for all primer sets successfully amplifying fragments of the desired size. Resolution of polyacrylamide gels is significantly higher than for agarose and allows for double stranded DNA fragments differing by only a few base pairs to be visualised as separate bands. Basic guidelines for selecting the appropriate gel percentage for a required range of resolution are provided in Table 10 for both agarose and polyacrylamide gels.

Vertical, non-denaturing 12% (w/v) polyacrylamide (PAA) gels were run at 150V in a mini-vertical unit (*Cleaver Scientific Limited*) and made from a mixture of: 3ml 40% (v/v) PAA [5% cross-linkage acrylamide (AA): bisacrylamide (BAA) ratio of 19:1], 20% (v/v) 5X TBE buffer and 46.7% (v/v) distilled water. A 10% (w/v) ammonium persulphate (APS, $H_8N_2O_8S_2$) solution (300µl) and 30µl of N, N, N′, N′-tetramethylethylenediamine (TEMED) was subsequently added to produce an end volume of 10ml.

Electrophoresis was performed in 1X TBE buffer for approximately ninety minutes. EtBr was not added to the gel mix; in order to facilitate the visualisation of DNA fragments under UV light,gel were soaked for approximately ten minutes in a separate container containing 200ml 1X TBE buffer and 0.03% (v/v) EtBr.

Table 10 General guidelines for both agarose and polyacrylamide gels used for selecting the gel percentage most appropriate for resolving a specific size range of fragments (Taken from: Sambrook *et al*. 1998)

% Agarose	Range of Resolution Linear dsDNA (kbp)	% Acrylamide	Range of Resolution dsDNA (bp)	ssDNA (nt)
0.5	30-1.0	3.5	100-1000	750-2000
0.7	12-0.8	5.0	75-500	200-1000
1.0	10-0.5	8.0	50-400	50-400
1.2	7-0.4	12.0	35-250	
1.5	3-0.2	15.0	20-150	
		20.0	5-100	

The EtBr was added to the buffer about thirty minutes prior to staining and mixed thoroughly so as to prevent concentrated spots of EtBr appearing on the gel. The stain mixture was stored in a dark cupboard when not in use to minimize degradation of the EtBr due to light exposure. Several gels could be stained consecutively in the same solution before it became necessary to refresh the EtBr. Fragments were visualised under UV light using a Multi Genius Bio Imaging System (GelDoc) (*Vacutec*).

4.2.3.4 Fluorescent Labelling of Selected Primers

Polymorphic primers (either forward or reverse) were labelled with one of the following fluorescent dyes: VIC (green), NED (yellow), FAM (blue) or PET (red) (*Applied Biosystems*). For labelling, some basic criteria were adhered to:

(1) The longer of the two primers is preferred for labelling.

(2) The primer with the highest G/C content at the 3′ should be selected for labelling.

(3) The primer closest to the repeat sequences is considered most ideal for labelling.

(4) If the primer meeting the above labelling requirements has a G/C base at the 5′ end, select the opposite primer in the set.

4.2.3.5 Optimization Process of Labelled Polymorphic Primers for Genotyping

Optimization of labelled primers was initiated by running each primer set at its optimum annealing temperature (*Ta*) as determined prior to labelling. Any primer sets that failed to amplify were rerun on a gradient PCR to determine whether the *Ta* had changed as a result of labelling and

thirdly, if required as a final resort, non-amplifying primer sets were tested on a TD-PCR. Since a primer set may fail to amplify after labelling, it is best to confirm this before continuing with genotyping analyses.

4.2.4 Black Rock Abalone Population and Selection of Primers

In order to ascertain their level of informativeness, primer sets was tested on thirty-two individuals from a wild abalone population, Black Rock, which inhabits the east coast of South Africa in the region of Port Elizabeth. Samples for these individuals were collected and extracted in 2003 for inclusion in a separate population study. Genotypes were generated for these thirty-two individuals and analysed for statistics pertinent to determining the primer sets suitability for parentage analyses: allele number (Na), expected heterozygosity (He) versus observed heterozygosity (Ho) and exclusion1 and exclusion2 probabilities.

The screening process was additionally used as an opportunity to develop and test PCR multiplexes for later use in parentage analyses.

4.2.4.1 Preparation of Black Rock Genomic DNA samples

The concentrations for individuals 1-32 were quantified using a NanoDrop® ND-1000 spectrophotometer. Working stocks of 20ng/µl were prepared through dilution using sterile distilled water and these were re-quantified to confirm diluted concentrations. Stocks were stored at -20˚C.

4.2.4.2 Genotyping of Individuals using Polymorphic Primers

Eleven polymorphic primers for perfect microsatellite repeat sequences were used to genotype the thirty-two Black Rock individuals. The overlap of the primers was determined by firstly, calculating an estimated allele size range for each primer set by adding and subtracting the repeat type [e.g. tetra = (4bp)] multiplied by number of repeat units (e.g. 7) from the expected fragment size determined through initial primer design. The process of range calculation is shown in Table 11. The estimated allele size range for a microsatellite locus was later, more accurately defined by the smallest and largest alleles determined through genotyping. This range may vary from population to population

A multiplex PCR reaction cannot contain primer sets with the same colour label if they produce overlapping fragment sizes since their alleles will not be distinguishable in the genotyping analysis. Primers labelled with different fluorescent labels may be combined despite overlap as their alleles will be distinguishable based on peak colour.

Table 11 Example of estimated primer range calculation for ascertaining range overlap and the potential for combination of selected primers into multiplexed PCR reactions

Example	Fragment Size	Repeat Type	Repeat No.	Range (bp)
(CACT)7	500bp	4	7	500 ± 28 = 472 - 528
(CAT)5	426bp	3	5	426 ± 15 = 411 - 441
(CT)11	594bp	2	11	564 ± 22 = 542 - 586

Once estimated allele size ranges had been established using the method demonstrated in Table 11, a physical layout was compiled showing the overlapping fragments and the respective fluorescent labels and colours (Figure 12).

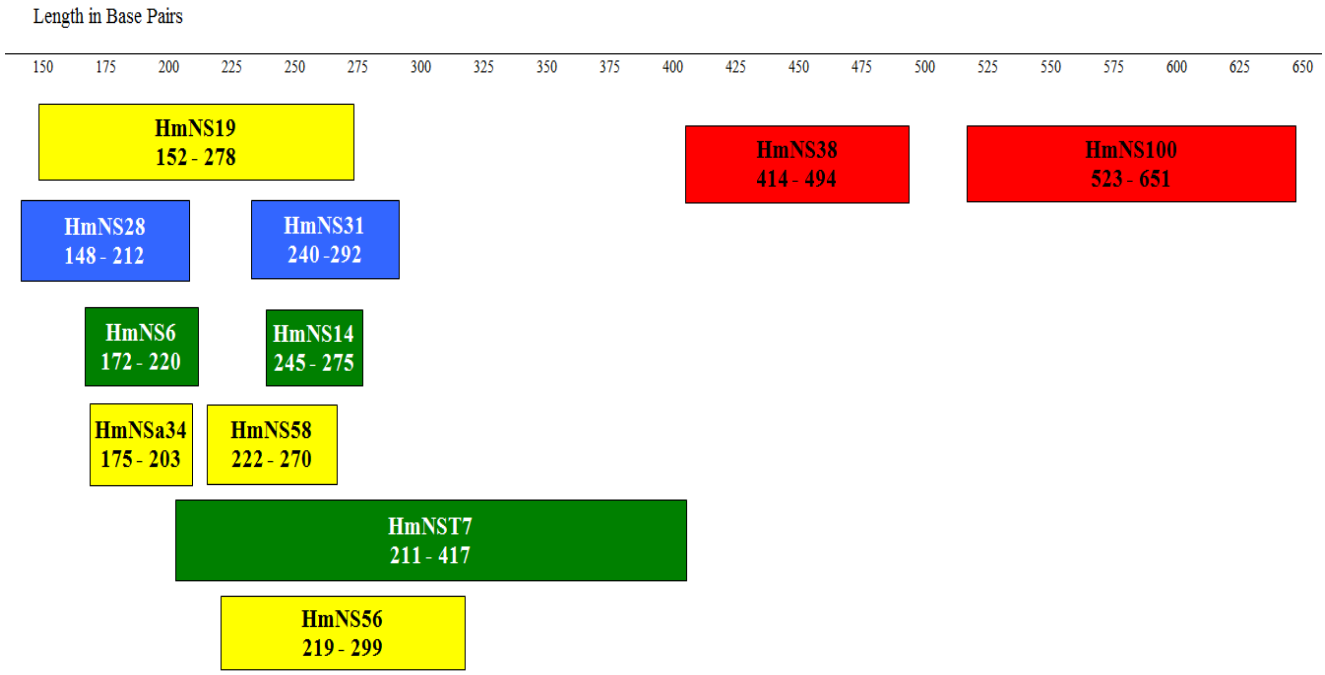


Figure 12 Estimated size range overlap for eleven microsatellites used in genotyping of Black Rock wild abalone population. Black Rock 1-32 was used as a test population to screen microsatellite markers for their informativeness and therefore usefulness in future parentage assignments

Individual standard PCR reactions of 10µl volumes were run for each primer set at their specific annealing temperatures: 5ng DNA, 0.2µM of each primer (forward and reverse), 1X *Taq* polymerase buffer, 2.0mM $MgCl_2$, 0.2µM dNTP mix, and 0.25U Go*Taq* Flexi DNA polymerase. A standard PCR program was run for each primer set: 94˚C for 5 minutes, thirty cycles of 94˚C for 30s, *55˚C-65.5˚C for 60s, 72˚C for 60s with a final extension step 72˚C for ten minutes (* indicates primer-specific annealing temperature (*Ta*) for each primer set).

Primers were pooled in a 1:1 ratio post-PCR into the following multiplexes and submitted for genotyping: Multiplex One combined four primers sets: HmNS19L, HmNS31D, HmNS38T and HmNS6T; Multiplex Two combined three primers sets: HmNS100T, HmNS58R and HmNST7T; while Multiplex Three combined four primers sets: HmNS14R, HmNS56D, HmNS28D and HmNSa34D. Samples were submitted to the Central DNA Sequencing Facility at the Genetics Department, Stellenbosch University for genotyping.

4.2.4.3 Optimization of Multiplexed Reactions for Genotyping

Three multiplexed reactions were used to genotype Black Rock individuals. Genotyping reactions were optimized through the systematic manipulation of different reagents:

(1) DNA concentrations of 1ng, 5ng, 10ng and 20ng were tested to optimize allele peaks for ease of scoring.

(2) Different *Taq* DNA polymerases were substituted in PCR reactions in an attempt to reduce non-specific amplification: Go*Taq* Flexi DNA polymerase (*Promega*) and Di-Kapa*Taq* (*KapaBiosystems*).

(3) Pooling strategies were varied: the first strategy comprised a multiplex PCR program with adjusted/increased concentrations of: $MgCl_2$ (2.5mM), Go*Taq* Flexi DNA polymerase (0.5U) and dNTPs (0.5mM) with all primer sets present in one reaction mix; compared to a second strategy of post-PCR pooling of individual standard PCR reactions run at each primer sets specific *Ta*.

Pooling strategy one versus strategy two provided comparable results although the intensities of some of the peaks in the multiplexed reaction were lower than for the second strategy. At the risk of allele drop-out it was decided to use the second strategy of post-PCR pooling despite the increased number of PCR reactions required.

4.2.4.4 Analysis and Primer Selection

Genotypes were generated for all thirty-two individuals and analysed using GENETIX 4.03 (Belkhir *et al*. 2000) for the following statistics: number of alleles (Na), expected heterozygosity (He) and observed heterozygosity (Ho). GENETIX 4.03 is a population genetics software package available free to researchers via the internet and can be used to perform exact tests and compute other classical population parameters for use in population genetics studies. CERVUS 2.0 was subsequently employed to calculate exclusion probabilities (Excl1 and Excl2), and null allele frequencies (Null Freq). Based on these results, the five most informative markers were chosen for use in parentage assignments.

4.3 Results

4.3.1 Microsatellite Isolation

4.3.1.1 Genomic DNA from Wild Abalone Populations

DNA extractions were processed by Mrs Aletta van der Merwe (Genetics Department, Stellenbosch University) using a phenol-chloroform method. Purity of the DNA and concentrations were originally determined by agarose gel electrophoresis against a Lamda (λ) standard (250ng/µl), but for the purpose of this study, the concentrations were reassessed using the NanoDrop® ND-1000 spectrophotometer. Table 12 indicates these results and where available the absorbance ratios. Six individuals were used in a series of enrichments and dilutions of the DNA were made to approximately 250ng for use in the ligation reaction.

Table 12 Initial genomic DNA concentrations for six individuals selected at random from the wild abalone populations shown and used in microsatellite enrichments

Individual	Location	Year	ng/ul	260/280	260/230
22	Gansbaai	2003	488.2	0.00	0.00
9	Gansbaai	2003	1048.56	1.99	2.32
13	Gansbaai	2003	1389.49	1.98	2.27
37	Gansbaai	2003	882.29	1.99	2.3
21	Witsand	2004	1 473.49	1.98	2.03

The concentrations for these samples were very high indicating extremely successful extraction from the mantle tissue used. Where 260/280 ratios were available, the slightly higher values of, on average 1.98, indicated some contaminants. These may be residues from the phenol-chloroform steps given that the value is higher than 1.8. The second indicator of purity, the 260/230 ratio is higher than the former ratio, as preferred, and very close to the 1.8-2.2 range expected for fairly clean DNA samples.

4.3.1.2 Digestion and Ligation of Genomic DNA followed by PCR Amplification

Several enrichments were performed during this study but since the expected outcome was the same for each, only one example will be included in these results to show what was considered a successful digestion and ligation.

Figure 13 shows the result of PCR performed on wild gDNA digested into smaller fragments using *Mse*I restriction enzyme and subsequently ligated with *Mse*I-adaptors. A range of fragment sizes is desired for cloning into pDrive vector which is evident by the smear of fragments greater than 200bp. The products were visualised on 2% agarose gel (run at 100-120V for 30 minutes).

DNA was subsequently enriched for a dinucleotide (AC)-n. Hyperladder IV provides an indication of the size range of the fragments produced by PCR and in this example fragments, are all greater than 250bp.

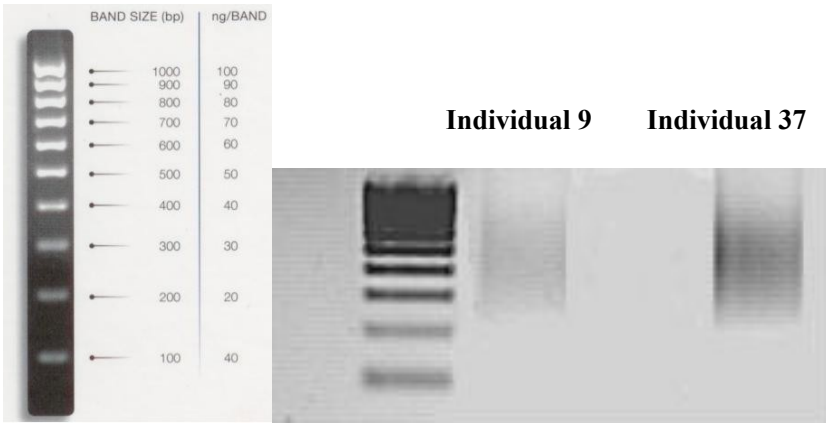


Figure 13 Genomic DNA from Individuals 9 and 37 collected from Gansbaai (2003) digested with restriction enzyme *Mse*I and ligated with *Mse*I-adaptors. PCR was completed using a *Mse*I-N primer cocktail

4.3.1.3 Selective Capturing of DNA-Probe Hybrids and PCR Amplification

Prior to cloning, PCR using *Mse*-N primers was run on each of the supernatants (NS, S, D1 and D2) removed during the selective recapture of enriched DNA. A range of fragment sizes (visible as a smear) indicates the desired result and this is demonstrated clearly in Figure 14 by supernatant D1. PCR of D1 was expected to produce the best result in terms of size range of fragments and the quantity of fragments amplified, given the higher concentration of recaptured DNA that is present in this supernatant. Very little or no product should be amplified in NS and S supernatants and ideally, there should be no DNA present in the final supernatant (D2), although this is rarely the case as recapture of hybridized DNA is not absolute.

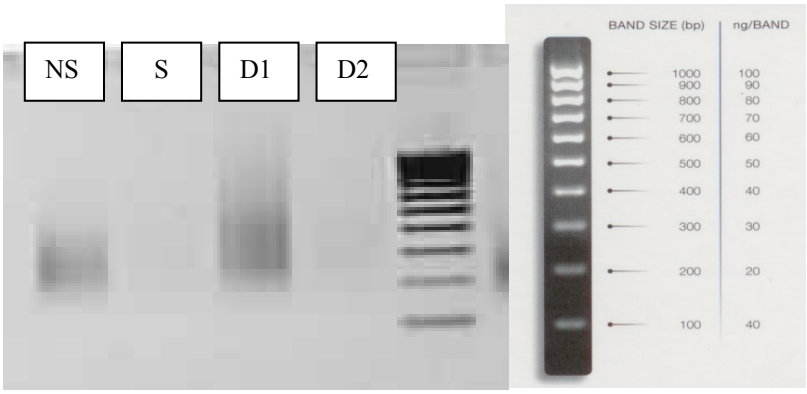


Figure 14 Supernatants NS, S, D1 and D2 amplified by PCR using *Mse*I-N primers. Genomic DNA of individual 21 collected at Witsand (2004) was enriched for a trinucleotide (CAA)-n. Hyperladder IV indicates that the fragments are greater than 200bp

4.3.1.4 Cloning through Ligation and Transformation of PCR Products

The success rate of transformation was very variable and several factors were changed in an attempt to increase the number of recombinant clones produced. These included the following, varied for different cloning procedures and never simultaneously: **1**. An extra 10 minute extension step was added to the PCR program used to generate fragments from D1 in an effort to increase the "A" overhangs on the fragments and hence improve their ligation into the vector; **2**. An extended incubation (2 hours) at 4˚C of the ligation reaction mix was performed as recommended in the ligation protocol; **3**. Two different temperatures (4˚C and 16˚C) were tested for the incubation step of the ligation process, these being the minimum and maximum temperatures of the range suggested; **4**. Purified PCR product was used for the ligation.

The application of a higher incubation temperature produced a low level of recombinants although whether this was indeed due to the temperature was not confirmed by repeat experiment. The QIAGEN® troubleshooting guide warns that higher temperatures e.g. greater than 25˚C give rise to increased background and thereby a reduced number of recombinants, hence an incubation temperature of 4˚C was used for all future ligations.

The density of white clones in the transformation shown in Figure 15 was not very high. The photo reveals an LB plate showing blue-white selection for recombinant clones. All suggestions for improving the success of recombination were taken from the trouble-shooting guide, pages 16-20, in the QIAGEN® cloning handbook (2001).

White clones indicate uptake of DNA into pDrive vector

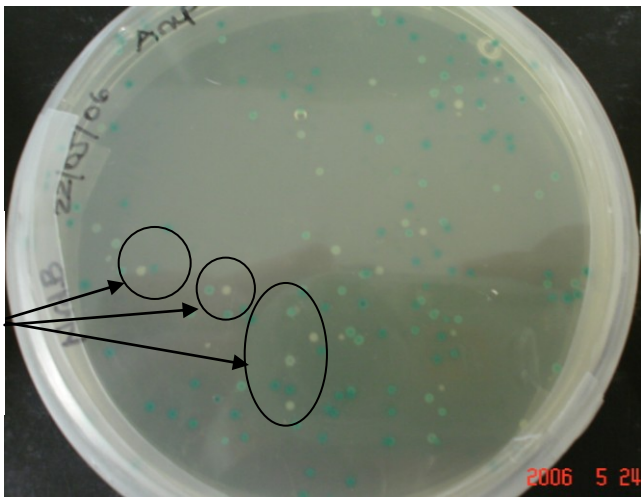


Figure 15 Blue-white selection for recombinant clones. An extra incubation at 4˚C was applied to increase the intensity of the blue colour thus aiding the identification of positive (white) clones given the low density of recombinants

4.3.2 Microsatellite Screening

Screening of positive clones was done through direct PCR of potentially recombinant clones using vector primers (see Appendix 6 for primer binding sites) with resolution of PCR products by agarose gel electrophoresis to identify those vectors containing abalone DNA inserts, indicated by the size of the fragment. Sequencing of the positive clones allowed for the design of primers within the regions flanking any identified microsatellite repeat sequences.

4.3.2.1 Direct Amplification of Positive Clones through Colony PCR

Only samples showing single bands greater than 300bp were presumed to contain abalone DNA. The distance between M13F and M13R primer binding sites is approximately 258bp (see Appendix 6 for the DNA region surrounding the cloning site) and any foreign insert will therefore be positioned within this region, hence increasing fragment size up from 258bp.

Figure 16 is an example of fragments produced by a colony PCR of positive clones. All nine clones appeared to be recombinants containing inserts of abalone DNA of approximately 340bp and are hence well above than the minimum requirement of 300bp. All these samples were sequenced and screened for microsatellites. Comparison to Hyperladder IV shows that the fragments are all above 600bp indicative of inserts of abalone DNA greater than 340bp.

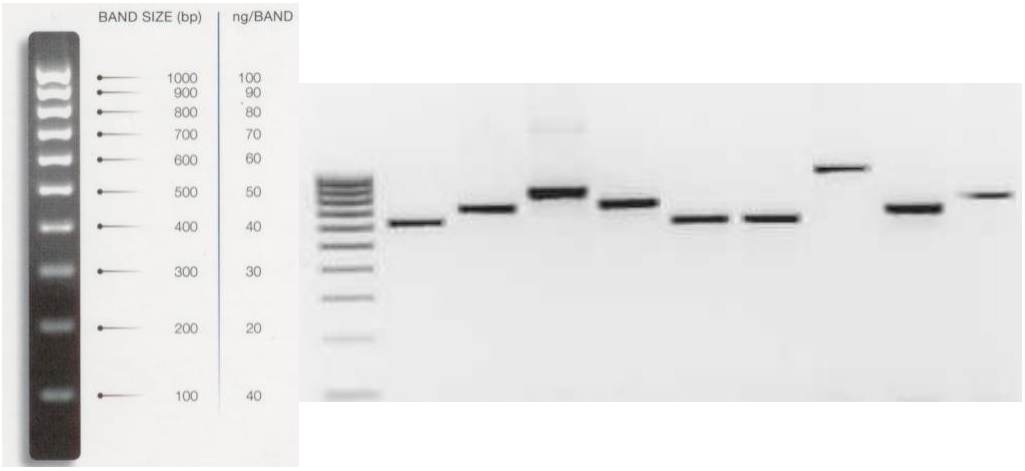


Figure 16 Recombinant clones containing inserts of wild abalone gDNA. Fragments were sequenced directly post purification using vector primers (T7, SP6 or M13 F/R). The minimum size of a recombinant clone is 300bp

A total of **909** clones were screened, of which approximately **50%** (four hundred and sixty-two) contained an insert of abalone DNA ranging from 150bp to 750bp. These clones were sequenced and screened for repeat sequences.

4.3.3.2 Automated Direct Sequencing of Selected PCR Fragments

Table 13 shows a summary of perfect repeats, imperfect repeats and gDNA containing no repeat sequences evident in the 462 sequenced clones. In some instances the quality of the sequence was such that it could not be screened adequately and this was then classified as "unknown". All sequences containing perfect repeats were subsequently compared to sequences already on NCBI as well as against all clones being developed with the research group to check for duplication.

Only **39%** of sequences containing perfect repeats were found to be unique, resulting in **61**% redundant clones. Results do not include the imperfect repeats, of which only fourteen were developed at a much later stage in the study. A blast of these fourteen clones revealed all of the sequences containing these imperfect repeats to be unique. The imperfect repeats classification in this study encompasses both imperfect/interrupted and compound repeats as defined by Chambers and MacAvoy (2000).

Table 13 Summary of perfect and imperfect repeats evident in cloned gDNA fragments. Findings are presented as a value and as a percentage of the total number of clones screened

	Number	Percentage (%)
Perfect Repeats	74	16%
Imperfect Repeats	219	47%
No Repeats	143	31%
Unknown	26	6%
TOTAL	**462**	**100%**

Figure 17 shows an electropherogram of a perfect repeat (ACGC)-6 edited using CHROMAS 145-95.

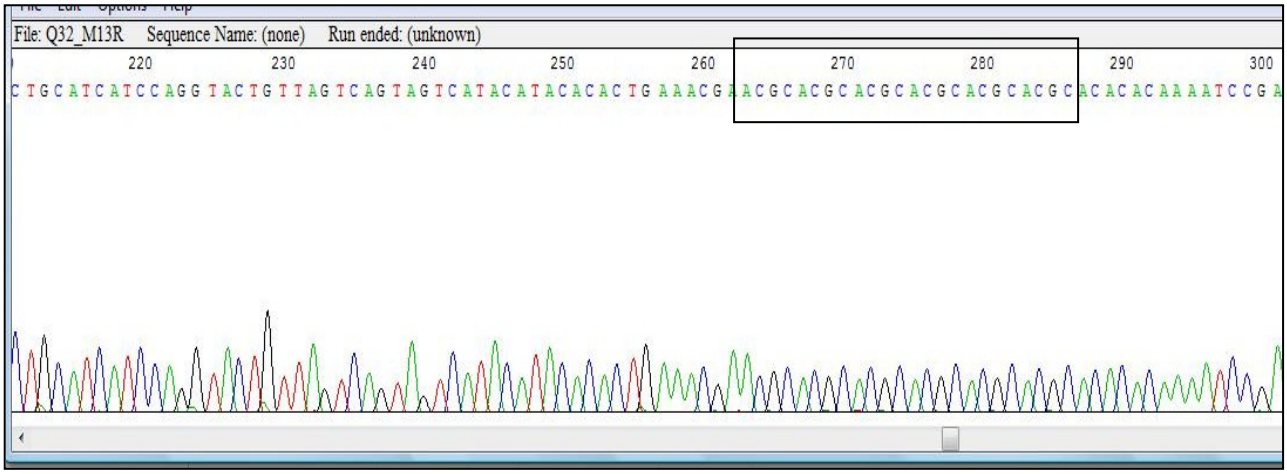


Figure 17 Electropherogram demonstrating a unique perfect repeat (ACGC)-6 identified in clone Q32 - the box indicates the position of the repeat sequence within the fragment of abalone DNA

The division of perfect repeats into dinucleotides, trinucleotides, tetranucleotides, hexa-nucleotides and heptanucleotides (minisatellites) was 39, 7, 18, 1 and 9 respectively. Dinucleotides, therefore, comprised the majority of the perfect repeat sequences identified with the tetra-repeats being the second highest in frequency.

4.3.3 Primer Development

4.3.3.1 Design of PCR Primers

Primers were developed for unique microsatellites only with preference given to perfect repeat sequences as these are reportedly easier to score accurately; an important consideration when microsatellites are to be used in parentage analyses. A total of **52** primer sets were developed and tested of which 14 failed to amplify.

Appendix 8 reports only those primer sets for which results were achieved. Two sets of primers were developed for microsatellite HmNS100T; the second primer set of which produces a fragment smaller than 600bp. Redevelopment of this primer was necessary since the ladder LIZ600® only allows accurate allele scoring up to 600bp. However, only the redesigned primer sequences used further in this study are included in the results and appendices for microsatellite HmNS100T.

4.3.3.2 Primer Testing for Detection of DNA Polymorphisms

All products from optimized primer sets were resolved on 12% PAA gels to detect polymorphisms if present. If non-specific banding occurred in the region of the expected fragment size, these primer sets were re-optimized in an effort to reduce this non-specificity. Table 14 provides a summary of the fifty-two primers tested and their status in terms of polymorphism.

Table 14 Summary of the total number of primer sets tested and their status: polymorphic, monomorphic or non-amplifying. These results are further divided into perfect and imperfect repeat sequences

	Perfect Repeats	Imperfect Repeats	Total
Polymorphic	17	14	**31**
Monomorphic	5	1	**6**
Non-Amplifying	6	9	**15**
Total	**28**	**24**	**52**

Examples of a monomorphic marker and a polymorphic marker as visualised on a polyacrylamide gel are shown in Figure 18 (a) and 18 (b) respectively. PCR products were resolved on 12% PAA gels run at 150V for 90 minutes. Visualisation was facilitated with EtBr staining and

UV light and fragments were approximately 200bp.

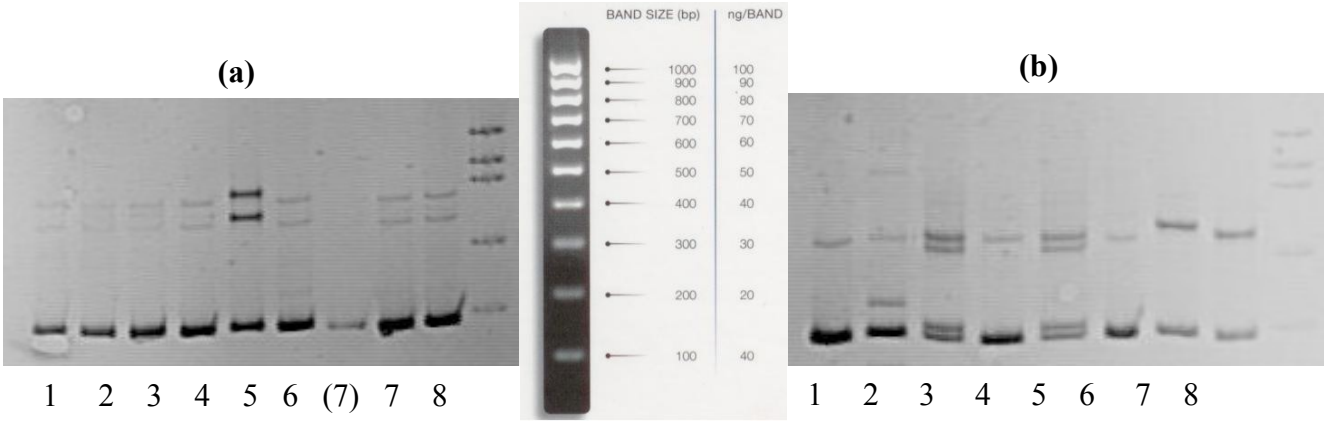


Figure 18 PCR products generated for Black Rock individuals 1-8 by **(a)** monomorphic marker HmNS26R $(GTA)_{-6}$ and **(b)** polymorphic marker HmNS6T $(ACGC)_{-6}$

The monomorphic marker HmNS26R is not suitable for parentage analysis and simply remains part of the pool of markers developed for *Haliotis midae*. The polymorphic marker HmNS6T was entered onto NCBI and can be identified with accession number EF367117.

4.3.3.3 Fluorescent Labelling of Selected Primers

Although, in total, 17 primer sets representing perfect microsatellite repeats were developed, only 14 were completed at the time of screening for suitable markers for parentage assignment. These fourteen sets plus two representing imperfect repeats were fluorescently labelled in preparation for screening of the wild abalone population collected at Black Rock (2003). The Black Rock population served as a test population to determine the level of informativeness of the primer sets developed and to select five of the best to use in parentage analyses. Two imperfect repeats were included to compare their ease of allele scoring to that of the perfect repeats.

Specific information such as the fluorescent label used and label colour, whether the forward or reverse primer was selected for labelling, and the microsatellite type to be amplified is provided in Appendix 9. The labels were assigned to minimize the overlap of the same colours (i.e. NED, PET, VIC or FAM) so that multiplexes could be tested. Three multiplexes were developed containing eleven of the sixteen labelled primers. HmNS7T and HmNS59D no longer amplified post labelling while HmNSS1H, HmNS17bT and HmNS18M were difficult to score producing multiple alleles in some individuals.

4.3.3.4 Optimization Process of Labelled Polymorphic Primer Sets for Genotyping

All optimization was carried out using Black Rock individuals (1-8) and all labelled primer sets were run at their optimum annealing temperatures as determined prior to labelling to see whether amplification still occurred. PCR products were visualised on 2% agarose gel using EtBr staining. The majority of the labelled primer sets amplified well at their specific *Ta* temperatures and no further optimisation was required.

Where amplification failed (HmNS7T and HmNS59D), these primer sets were run on a gradient PCR to determine whether the *Ta* had simply changed or whether amplification no longer occurred at all. For these two primer sets, no product was produced at any temperature and these will need to be redesigned if they are to be included in the pool of primers being developed for this abalone species.

It was further ascertained from these trials that primer sets HmNSS1H, HmNS17bT and HmNS18M produced multiple alleles in some of the individuals, despite combining these primer sets in different multiplexes. These were hence deemed not suitable for parentage analyses.

Several attempts were made to improve the specificity of the primers for microsatellites HmNSS1H, HmNS17bT and HmNS18M through increasing the annealing temperature where feasible and testing an alternative *Taq* polymerase, DiKapa *Taq* from *Kapa Biosystems*. Despite these reaction changes, no further success was had.

Primers for microsatellite HmNSS1 (multi-allelic) and HmNS59 (not amplifying) were subsequently redesigned and retested. PCR amplification was achieved for HmNSS1 but primers were not relabelled hence suitability for genotyping was not determined. The new primers for HmNS59 did not amplify specifically but produced a smear despite testing different annealing temperatures. Since the initial primer set for this locus amplifies well prior to labelling, the option exists to re-label the opposite primer as this may improve the result. Primers for HmNS7T, HmNS17bT and HmNS18M will need to be redesigned and retested.

Figure 19 illustrates the multiple alleles produced with primers for microsatellite HmNSS1H in multiplex one. Primers for this locus were run in a multiplexed reaction with primer sets for HmNS19L, HmNS31D, HmNS38T and HmNS6T. Products were pooled post-PCR in a ratio of 1:1.

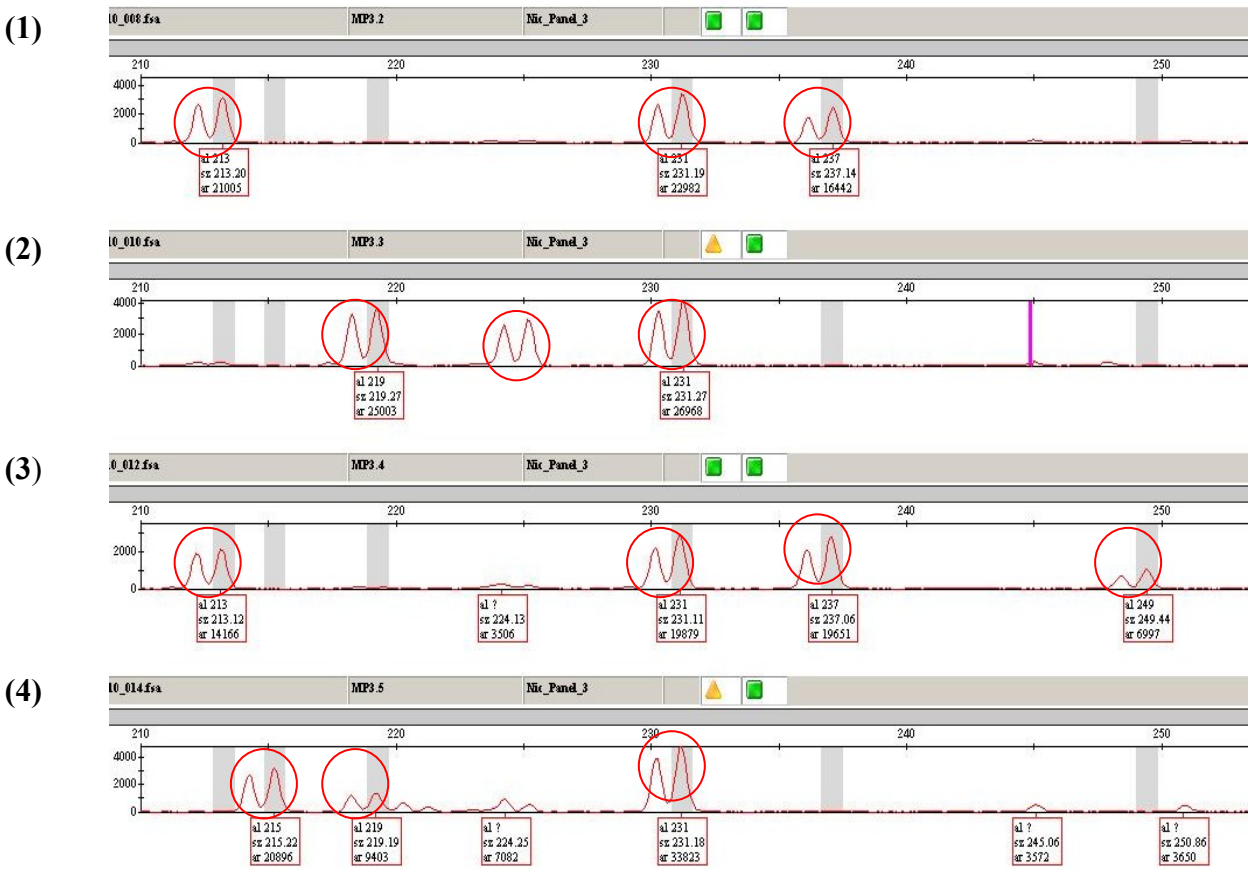


Figure 19 Multiple alleles for microsatellite HmNSS1H. Panels indicate Black Rock individuals 1-4 with the multiple alleles indicated by means of red circles

4.3.4 Black Rock Abalone Population and Selection of Primers

4.3.4.1 Preparation of Genomic DNA samples

Genomic DNA extractions had already been processed by Mrs Aletta van der Merwe (Genetics Department, Stellenbosch University) for wild abalone samples collected at Black Rock in 2003. Samples for individuals one to thirty-two were quantified using the NanoDrop® ND-1000 spectrophotometer and concentrations were sufficiently high to allow for dilutions to be prepared. Dilutions of 5ng/µl were used in microsatellite screening.

The absorbance ratio 260/280 indicated some phenol/chloroform contamination while the lower 260/230 ratio supported the evidence of co-purified contaminants. However, the purity of the samples was high enough not to affect subsequent PCR reactions.

4.3.4.2 Genotyping of Individuals using Labelled Polymorphic Primer Sets

The multiplex primer set combinations indicated in Table 15 were used to genotype the Black Rock population. Reactions were run at the *Ta* temperature shown using a standard PCR program. Pooling was post-PCR in a 1:1 ratio.

Table 15 Multiplex primer set combinations used for genotyping the Black Rock population together with expected fragment size and fluorescent label

Multiplex	Primer	Repeat	Size (bp)	Label	*Ta* (optimum)	Assession Number
MP1	HmNS19L	(AACACCC)9	215bp	NED	55°C	EF033330
	HmNS31D	(GT)4(CT)(GT)8	266bp	FAM	55°C	EF033333
	HmNS38T	(TCAC)10	454bp	PET	55°C	EF367113
	HmNS6T	(ACGC)6	196bp	VIC	55°C	EF367117
MP2	HmNS100T	(GAGT)16	406 bp	PET	60°C	EF367114
	HmNS58R	(GTT)8	246bp	NED	60°C	EF367119
	HmNST7	(CACT)26	313bp	VIC	60°C	EF455618
MP3	HmNS14R	(TTG)5	260bp	VIC	60°C	EF367115
	HmNS56D	(CA)20	259bp	NED	55°C	EF455619
	HmNS28D	(CA)16	180bp	FAM	55°C	EF033332
	HmNSa34D	(AC)7	189bp	NED	50°C	EF367118

4.3.4.3. Optimization of Multiplexed Reactions for Genotyping

Various combinations of the eleven remaining primer sets were tested ensuring no overlap peaks of the same colour. Samples were submitted to the Central DNA Sequencing Facility at the Genetics Department, Stellenbosch University for genotyping. Very little optimization was required although different DNA concentrations were tested to reduce the peak intensities of some of the primer sets while ensuring that allele drop out did not occur for other microsatellite loci being amplified within the same reaction. Allele drop out can result due to insufficient template DNA being available and therefore preferential amplification of the larger allele taking place.

Concentrations of approximately 1ng, 5ng, 10ng and 20ng per reaction were tested and no obvious differences were observed between 5ng, 10ng and 20ng levels. At lower levels (1ng), where previously two alleles had been observed, only one was present in some instances. To reduce this potential risk of allele drop out, a concentration of approximately 5ng was used.

4.3.4.4 Statistical Results and Selected Polymorphic Primers for Parentage Assignment

Genotypes were generated for the Black Rock population and analysed using GENETIX 4.03 and CERVUS 2.0 for statistics pertinent to determining their informativeness. These statistics encompassed: allele number (Na), observed heterozygosity (Ho) expected heterozygosity (He), PIC values, exclusion1 (Excl1) and exclusion2 (Excl2) probabilities and estimated null allele frequency (Null Freq). Table 16 shows the results of these analyses. The primers highlighted in red typeface were selected for inclusion in subsequent parentage assignments. The allele frequency results for all loci are provided in Appendix 10.

Table 16 Statistical results from GENETIX 4.03 and CERVUS 2.0 based on genotype input files for Black Rock individuals 1-32

Locus	Na	Ho	He	PIC	Excl1	Excl2	Null freq
HmNS19	12	0.5630	0.9030	0.878	0.632	0.776	0.2185
HmNS31	14	0.3440	0.7980	0.764	0.434	0.613	0.4001
HmNS38	11	0.7420	0.8330	0.495	0.495	0.667	0.0457
HmNS6	9	0.5940	0.7800	0.735	0.387	0.565	0.1275
HmNS58	10	0.7810	0.8560	0.824	0.520	0.688	0.0324
HmNS100	14	0.5310	0.9170	0.895	0.670	0.803	0.2572
HmNST7	21	0.9350	0.9450	0.926	0.754	0.859	-0.0033
HmNS14	4	0.0320	0.2110	0.199	0.022	0.109	0.7016
HmNS28	19	0.9060	0.9430	0.924	0.747	0.855	0.0115
HmNSa34	3	0.0970	0.3560	0.318	0.061	0.178	0.5608
HmNS56	16	0.8390	0.8540	0.826	0.536	0.700	0.0002

Key Allele number (Na), observed heterozygosity (Ho), expected heterozygosity (He), polymorphic information content (PIC), exclusion probability 1 (excl1), exclusion probability 2 (excl2) and estimated null allele frequency (Null freq)

A further five primer sets that were developed by N. Ruivo (Aquaculture Division, Genetics Department, Stellenbosch University), were selected for parentage assignment based on the same selection criteria used above. This set of ten primers was used to assign parentage to all offspring selected for fast or slow growth and a common marker suite was used so as to allow for results from the two independent studies to be compared and published. HmNR224T was subsequently excluded from further analyses for this study due to the multiple alleles produced during the screening of Black Rock 1-32 and inconsistencies with the alleles amplified when re-genotyping the same individual.

4.3.5 Shortcomings and Sources of Error

(1) Variability in transformation success and the number of recombinant clones produced. This needs to be addressed so as to reduce the number of transformations necessary to produce sufficient positive clones for microsatellite screening.

(2) The high level of redundant clones (61%) that were sequenced unnecessarily.

(3) Random positioning of the abalone gDNA insert within the vector so that it was often too close to either the 3´ or 5´ end to allow for primer design.

(4) The quality of the sequences was not always good enough to use requiring either re-sequencing or the abandonment of clones and potential microsatellites.

(5) Typing errors that may have occurred when compiling data and input files as well as during processing of output files for relevant statistics.

(6) Non-amplification of primer sets; fourteen out of the fifty-two sets developed failed to amplify under any conditions. This either meant redevelopment of the primer set or discarding of the microsatellite in question.

(7) Allele scoring errors. Genotyping was not repeated for individuals unless the allele scored was in doubt. When individuals were re-genotyped and allele scores differed from the first set, these results were re-looked and if the difference was a homozygous genotype versus a heterozygous genotype, the heterozygous genotype was used in the analyses. If the alleles differed in entirety, alleles were not scored for the individual at that specific locus.

(8) The null allele frequency of a specific locus, an estimate of which is provided by CERVUS 2.0. A high null allele frequency can reduce the level of assignment and primer sets for loci with null allele frequencies greater than 5% should be redesigned.

(9) Allele dropout/partial null alleles due to preferential amplification of the shorter allele or non-amplification due to inconsistent DNA quality or low template quantity within the multiplex. This may result in an apparent heterozygote deficiency which in turn could lead to a potential mismatch in genotypes between parents and offspring thereby reducing assignment success.

4.4 Discussion

Several parentage studies conducted for a variety of fish and mollusc species, including abalone, have employed microsatellites as the marker of choice (Herbinger *et al*. 1995; Garcia de Leon *et al*. 1998; Castro *et al*. 2004; Li *et al*. 2004; Dong *et al*. 2006). Due to their high levels of polymorphism, reliability and ubiquitous distribution throughout the genome of most eukaryotes, these markers are also extremely powerful for linkage mapping (Sekino & Hara 2006). Since a linkage map for *H. midae* is one of the broader project objectives, the development of a large number of molecular markers, amongst these microsatellites, by the research group is a prerequisite. Effectively, the employment of microsatellite loci in this parentage assignment study allowed two objectives, one study specific and the other project specific, to be addressed. An additional draw card of microsatellite analysis is that it requires only a very small amount of DNA thus facilitating non-lethal sampling. Since all breeding stock and juveniles within the PRS needed to be preserved for possible inclusion into future MAS programs, the employment of non-lethal sampling is essential.

A number of techniques for the establishment of microsatellite enriched genomic libraries have been published (Edwards *et al*. 1996; Hamilton *et al*. 1999; Paetkau 1999). More recently the FIASCO method by Zane *et al*. (2002) was reported, a technique which allows for the isolation of microsatellite repeat sequences directly from genomic DNA through similar hybridization techniques to those presented by Refseth *et al*. (1997). In so doing, the laborious process of constructing a genomic library prior to enrichment is avoided. FIASCO was the method choice for this study since our group had already had prior experience and success using this technique. Additionally FIASCO was recently used to isolate microsatellite markers in blue crabs (Steven *et al*. 2005) with a comparable measure of success to that already obtained by our laboratory.

Enrichment techniques circumventing the construction of a genomic library provide the following advantages: speed (three days if DNA extractions are already concluded), efficiency, the requirement of only the most basic of skills using limited laboratory equipment and reasonable reagent costs (Zane *et al*. 2002).

Low fragment yields in the initial digestion-ligation steps can seriously impede the success of downstream procedures such as transformation and cloning. QIAGEN® provide several criteria to take note of when performing the ligation step as well as suggestions in the troubleshooting guide should a low number of recombinants be generated (QIAGEN® Cloning Handbook, 2001, p12 & p17). All of the following criteria were applied in this study: confirming that a non-proofreading DNA polymerase is used in order to generate the necessary "A" overhangs required for successful ligation and adding an additional ten minute extension step to increase these overhangs; using purified DNA in the ligation and ensuring an adequate DNA: pDrive vector ratio; and finally increasing the ligation time to allow sufficient opportunity for ligation to take place. The numerous digestion-ligation procedures performed in this study produced variable results in terms of the quantity of fragments amplified although always with the desired spread of fragment sizes. Subsequent enrichment for different repeat types, $(AC)_n$, $(CAA)_n$ and $(GTGC)_n$, isolation of DNA-probe hybrids and PCR of supernatant D1, in turn produced a range of fragment sizes with the density of the band indicating that sufficient amplified product was produced. These results were consistently achieved for all enrichment steps performed.

Transformation and cloning results were more variable despite adopting all recommendations in the QIAGEN® Cloning Handbook for increasing the number of recombinant clones. Cruz *et al*. (2005) suggests that the number of recombinants obtained is directly related to the microsatellite richness within the species; however, it is known, from a study completed by Ross *et al*. (2003), that the genomes of molluscs' exhibit sufficient genome coverage by a variety of microsatellite repeat motifs. This fact is supported by the recent publication of medium density linkage maps for

two abalone species based solely on microsatellite markers (Baranski *et al*. 2006; Sekino & Hara 2006). The variable results obtained for cloning suggests that digestion-ligation using a single restriction enzyme, *Mse*I, may not have lead to a representative sample of the genome being produced due to the nature of enzyme/linker combinations and their variable performance within different species. Recognition sites for *Mse*I may have low presence within the genome of *Haliotis midae*, a complication that might be overcome by employing a combination of restriction enzymes and a universal linker as described in Hamilton *et al*. (1999).

A total of 909 positive (white) clones were directly screened using a colony PCR, 50% of which contained inserts ranging from 150bp to 750bp. These results are similar to those obtained for microsatellite isolation performed on Atlantic salmon using an enrichment technique described by Refseth *et al*. (1999) in which 70% and 50% positive clones (i.e. clones containing inserts of 300-400bp) were generated for a dinucleotide and tetranucleotide enrichment respectively. In contrast, a study by Li and Kijima (2007) on Pacific abalone (*H. discus hannai*) showed that only 24% of 350 white colonies screened with PCR appeared to contain inserts of abalone DNA. An enrichment technique employing magnetic bead-hybridisation as the selection method was also utilised and hence the output can be used as a means to evaluate the success of our results.

Sequencing of these positive clones yielded the following split: 16% containing perfect repeats (various motifs), 47% containing imperfect/interrupted repeats, 31% containing no repeat sequences and 6% undetermined due to poor sequence quality. Put differently, a total of 63% of sequenced clones contained repeat sequences of differing compositions.

In a study conducted on Atlantic salmon (Refseth *et al*. 1999), 46% of clones were found to contain repeat sequences, while the split was 69% perfect repeats sequences versus 23% imperfect/interrupted repeats. The higher level of perfect repeats in the genome of Atlantic salmon when compared to abalone could be explained by the fact that motifs, mutation rates and profusion of repeat types vary between species (Ross *et al*. 2003). In a study on Pacific abalone, Li and Kijima (2007), found only 13.1% of the positive clones to contain repeat sequences with 73.3% of these being perfect repeats and the remaining 26.7% being an equal split between imperfect and compound repeats as defined by Weber (1990: as cited by Li & Kijima 2007).

Both these latter studies show a higher proportion of perfect repeat sequences when compared to the results obtained in this study. Alternatively, Edwards *et al*. (1996), employing a genomic library in their enrichment protocol, found an average of 60% of clones (out of a total of 140 sequenced clones) contained at least one microsatellite locus with insert sizes between 250bp to 900bp. These results are in line with ours, although no split between repeat motifs is provided by Edwards *et al*. (1996) so no further comparisons can be drawn. Their study also provides no

indication of the species that was under investigation since it was the technique itself that was being evaluated.

Further breakdown of the 16% of clones containing perfect repeats revealed the following split in motifs in order of decreasing frequency: 53% dinucleotides, 24% tetranucleotides, 12% heptanucleotides (minisatellites), 10% trinucleotides and 1% hexanucleotides. Chistiakov *et al*. (2005), in a review of microsatellites and their genomic distribution, evolution, function and application, noted that dinucleotides represent the majority of microsatellites found in the genomes of vertebrates, more specifically finfish such as the puffer fish and the three-spined stickleback. They estimated this figure to be approximately 30% to 67% which is quite comparable to the 53% observed in this study. In the genome of Japanese pufferfish, a study by Edwards *et al*. (1998: as cited by Chistiakov *et al*. 2003) discovered that dinucleotides occurred with the highest frequency (34%) followed by tetranucleotide repeats to the magnitude of 21%, 19% trinucleotides, 16.5% mononucleotides, 6% hexanucleotides and 3% pentanucleotides. The order of the first three listed motifs is very similar to this study's findings for *H. midae* although the percentage of dinucleotides found was considerably higher and heptanucleotides had the third highest frequency.

A general finding with enrichment protocols is the sequencing of duplicate clones thus resulting in a large portion of redundant sequences (Paetkau 1999). This was experienced in this study when, after running a blast against the NCBI database as well as a local database of all microsatellites developed or in the process of development by our research group, it was found that only 39% of the clones containing perfect repeats were unique. This resulted in 61% redundancy of sequenced clones. Stevens *et al*. (2005), in a study on blue crabs, reported only 55.5% unique clones from a tetranucleotide enrichment, with one sequence having being cloned three times; while Piggott *et al*. (2006) observed a redundancy level of between 8% up to 63% for a range of invertebrate and vertebrate species investigated.

A criticism of the FIASCO protocol (Zane *et al*. 2002) is that it does not incorporate a means to predetermine, prior to sequencing, whether the positive clones in fact contain a repeat sequence or whether the insert within the vector is unique. With the ever increasing database of microsatellite markers being developed for *Haliotis midae*, the probability of duplication is escalating and consideration should be given to pre-screening all clones prior to sequencing. A method described by Piggott *et al*. (2006) provides a means to improve the efficiency of microsatellite identification through the use of SSCP. The study found that SSCP screening significantly reduced the amount of sequencing necessary to obtain a desired number of microsatellite markers and report that of 344 positive clones, pre-screening indicated that only 97 needed to be sequenced thus reducing the sequencing effort by 72%. A very conservative view was taken by Piggott *et al*. (2006) when

assessing the approach to mitigate the risk of missing potentially unique microsatellite sequences yet the improvement was still substantial.

With regards to identifying whether inserts contain a microsatellite locus without having to first sequence the clone, the possibility of hybridization with a fluorescently labelled oligo probe could be investigated. Refseth *et al*. (1997) report a hybridization technique for the identification of repeat containing positive clones and a cost-benefit assessment would need to be performed to determine whether this approach would be practical and useful to adopt.

Fragments sizes generated by the primer sets developed in this study ranged in length from 114bp to 618bp. Since shorter fragments are reportedly both easier to amplify consistently and to score when genotyping; designing primer sets to produce smaller fragment sizes is recommended for future. For those primer sets from this study that need to be redesigned, the preference for shorter fragments should be borne in mind and primer sets designed accordingly. Pemberton *et al*. (1995) propose that point mutations accumulate most readily in the flanking sequences immediately adjacent to the repeat sequence and both PCR primers should ideally be designed outside of these regions.

Accuracy of allele calling is crucial in ensuring high confidence in the results of a parentage study. Genotyping of the Black Rock population highlighted the need to further optimise multiplexed reactions. Signal intensities of the different primer sets within the multiplex were variable with some peaks exhibiting off-scale peak measurements while others were barely discernible at first glance. Dinucleotide repeat loci were particularly troublesome to score consistently due to the presence of stutter patterns. To circumvent the difficulties encountered when scoring dinucleotide repeat sequences as a result of the stutter profiles, the use of trinucleotide and tetranucleotide loci would be preferable. However, dinucleotide repeats remain the most prevalent microsatellites in the molluscan genome (Chambers & MacAvoy 2000) and it is not always feasible to work only with preferred motifs.

A few options exist for the improvement of allele calling: equalising of the signal intensities through a more accurate pre-determination of marker volumes required when in specific combination with other markers; removal of DNA *Taq* polymerase overhangs on PCR fragments through the application of T4 DNA polymerase to PCR products which should make allele calling simpler and more reliable (Gyapay *et al*. 1996); the application of an algorithm for binning the output of automated genotyping systems such as Genemapper®, the program used in this study. Amos *et al*. (2007) present their algorithm as an effective means to minimise binning errors arising from variable mobility of fragments, a phenomenon particularly evident for dinucleotide repeats. Assistance from this algorithm, however, is no substitute for careful scrutiny of data; repeat scoring

when in doubt and manual adjustments where necessary. In their review, Webster and Reichart (2005) provide some further guidelines for scoring allele data generated on an ABI system.

Screening of Black Rock individuals revealed that only five out of the eleven polymorphic loci complied with the minimum requirements for parentage analyses; a null allele frequency less than five percent and a high number and frequency of alleles per locus. The higher the frequency of alleles, the higher the PIC value of the marker while the overall PIC value of a potential marker suite provides an indication of the ability of this set of loci to address the research problem at hand; in this case, the accurate assignment of F1 offspring to candidate parents (Liu & Cordes 2004). Although null allele frequencies for HmNS19, HmNS31, HmNS6 and HmNS100 were above the acceptable level, the allele frequencies and PIC values were adequate and redesigning of the primer sets may allow for these loci to be used in future studies and to be included in the pool of molecular markers necessary for the construction of a linkage map for *H. midae*. Microsatellite loci HmNS38, HmNST7, HmNS56, HmNS28 and HmNS58 were selected for the next step: assignment of parentage in two hatchery populations of abalone.

Chapter Five

A Parentage Assignment Study

Abstract

The level of parentage assignment success for both Abagold and HIK was high, with a total of 91% and 90% achieved respectively. Offspring were assigned in the majority to a single parent pair while others were only assigned a specific mother or father. The similarity of genotypes between potential parents may have contributed to this occurrence as well as to the incidence of multiple potential parent pairs being assigned for some juveniles. These findings may be an indication that a measure of relatedness exists between breeders but this was not studied further since it fell outside the scope of this study. Based on contributions to spawning, it was recommended that eight breeding males and eight breeding females be retained within the Abagold hatchery. Similarly for HIK, it was recommended that ten males and nine females be retained on the basis of their larger contributions to the offspring studied. Within both hatcheries, those individuals not making significant contributions should either be removed or monitored carefully during more controlled breeding to reassess their potential. The small sample sizes investigated does not allow for significant conclusions to be drawn regarding the breeding population and caution should be exercised before discarding any potential breeders. Regarding growth potential, no obvious trends were observed, with most individuals producing both faster and slower growing individuals. Juveniles will be reassessed at two years to determine whether the size advantage or disadvantages observed at this early stage are maintained as the animals move towards sexual maturity; and in addition, to ascertain whether size advantages/disadvantages may be gender specific.

5.1 Introduction

In recent years, in studies of ecological and evolutionary processes, assignment methods and the genetic information they employ, have been used to determine whether individuals or groups of individuals belong to a specific population. The term “assignment method” is used broadly and encompasses genetic mixture analyses as well as parentage analyses since the same principles form their basis and both are active research fields (Manel *et al*. 2005).

Chapter five was structured in the same format as previous chapters and encompasses the following: Materials and Methods, which describes the development of three multiplexed reactions incorporating nine microsatellite markers. Multiplexing was adopted in an effort to reduce both genotyping costs and time allocation. All selected breeders and juveniles were genotyped using these selected loci and CERVUS 2.0 (Marshall *et al*. 1998) was utilised to match candidate parent/s to selected offspring in the PRS. Results of assignments are documented in terms of contribution of breeders to next generation individuals, in other words spawning success was assessed together with any trends related to the growth potential of juveniles derived from specific breeders.

Shortcomings and sources of error are elaborated on followed by a discussion of the results contained specifically within this chapter.

5.2 Materials and Methods

5.2.1 Genotyping of Abalone Adults and Juveniles from the PRS

Breeding individuals and juveniles were genotyped using the same set of microsatellite markers to generate a unique fingerprint for each. The data for offspring were then compared to all candidate adults through parentage assignment analysis using CERVUS 2.0 (Marshall *et al*. 1998); from which the most likely parent or parental combination was determined.

5.2.1.1 Genotyping of Performance Recording Scheme Individuals

A total of 598 individuals were genotyped for parentage analysis – 198 breeding individuals and 400 F1 progeny. Of the nine primer sets used to genotype these animals, four were developed by N. Ruivo and five by the author. In order to reduce the number of genotyping submissions required, the following multiplexed reactions were developed: For N. Ruivo's primer sets, two PCR reactions were performed and products from these reactions pooled in a 1:1 ratio into a single genotyping multiplex. The author's primer sets were run in three PCR reactions and pooled in a 1:1 ratio into two genotyping multiplexes.

5.2.1.2 Multiplexing of Polymorphic Primers for Parentage Assignment

The multiplex combinations for the genotyping of PRS individuals were based on the potential fragment overlaps illustrated in Figures 20 and Figure 21 respectively:

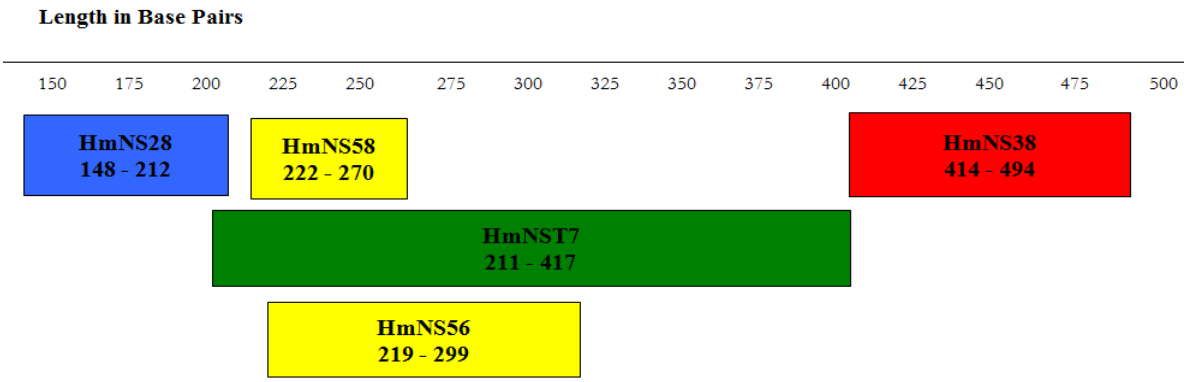


Figure 20 Overlap of five primer sets (author) selected for genotyping breeding stock and offspring for parentage assignments. Two genotyping multiplexes were developed composed of three PCR reactions with post-PCR pooling of products

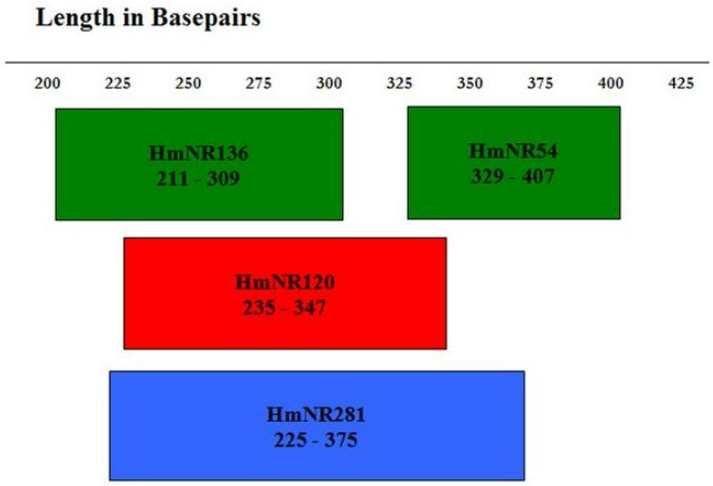


Figure 21 Overlap of four primer sets (N. Ruivo) used in genotyping breeding stock and offspring for parentage assignments. One genotyping multiplex was developed and prepared through two PCR reactions with post-PCR pooling of products

Multiplex One (Author): Three primer sets representing microsatellite loci HmNS38T, HmNST7T and HmNS56D were combined; products for which were prepared through two separate PCR reactions and then pooled in a 1:1 ratio.

PCR Reaction 1: 5ng DNA, 0.2µM of each primer for HmNS38T (forward and reverse), 0.1µM of each primer for HmNST7T (forward and reverse), 1X *Taq* polymerase buffer, 2.5mM $MgCl_2$, 0.5mM dNTP mix and 0.5U Go*Taq* Flexi DNA polymerase run on multiplex PCR program: 94˚C for four minutes, thirty-five cycles of 94˚C for 60s, 58˚C for 10s, 72˚C for 20s and 72˚C for ten minutes.

PCR Reaction 2: 2.5ng DNA, 0.2µM of each primer for HmN56D (forward and reverse), 1X *Taq* polymerase buffer, 2.0mM $MgCl_2$, 0.2mM dNTP mix and 0.25U Go*Taq* Flexi DNA polymerase run on PCR program: 94˚C for five minutes, thirty cycles of 94˚C for 30s, 60˚C for 60s, 72˚C for 60s and 72˚C for ten minutes.

Multiplex Two (Author): Two primer sets representing microsatellites HmNS28D and HmNS58R were combined in a single PCR reaction.

PCR Reaction 3: 5ng DNA, 0.1µM of each primer for HmNS28D (forward and reverse), 0.1µM of each primer for HmNS58R (forward and reverse), 1X *Taq* polymerase buffer, 2.5mM $MgCl_2$, 0.5mM dNTP mix and 0.5U Go*Taq* Flexi DNA polymerase run on multiplex PCR program: 94˚C for four minutes, thirty-five cycles of 94˚C for 60s, 52˚C for 10s, 72˚C for 20s and 72˚C for ten minutes.

Multiplex One (N. Ruivo): Four primer sets representing microsatellite loci HmNR54H, HmNR120T, HmNR136D and HmNR281P were combined; products for which were prepared through two separate PCR reactions and then pooled in a 1:1 ratio.

PCR Reaction 1: 5ng DNA, 0.1µM of each primer for HmNR54H (forward and reverse), 0.2µM of each primer for HmNR120T (forward and reverse), 1X *Taq* polymerase buffer, 2.5mM $MgCl_2$, 0.5mM dNTP mix and 0.5U Go*Taq* Flexi DNA polymerase run on multiplex PCR program: 94˚C for four minutes, thirty-five cycles of 94˚C for 60s, 52˚C for 10s, 72˚C for 20s and 72˚C for ten minutes.

PCR Reaction 2: 5ng DNA, 0.1µM of each primer for HmNR136D (forward and reverse), 0.2µM of each primer for HmNR281P (forward and reverse), 1X *Taq* polymerase buffer, 2.5mM $MgCl_2$, 0.5mM dNTP mix and 0.5U Go*Taq* Flexi DNA polymerase run on multiplex PCR program: 94˚C for four minutes, thirty-five cycles of 94˚C for 60s, 52˚C for 10s, 72˚C for 20s and 72˚C for ten minutes.

5.2.1.3 Optimization of Multiplexes for Parentage Assignment

As presented in more detail in chapter four: section 4.2.4.3, different factors were varied to ascertain the best combination of primer sets and the optimum test conditions for the multiplexed reactions used in genotyping. DNA concentrations between 2.5ng and 5ng worked optimally and were specific to primer sets; primer concentrations ranged from 0.1µM to 0.2µM per primer per reaction and T*a* was either 52˚C or 58˚C for the multiplexed reactions and 60˚C for locus HmNS56 which was run on a standard PCR program.

5.2.2 Parentage Assignment

5.2.2.1 Software Overview

The parentage assignment program CERVUS 2.0 (Marshall *et al*. 1998) was used to complete parentage analyses. This software makes use of likelihood ratios to infer paternity when polymorphic codominant markers are utilized since the use of exclusion may be insufficient to unambiguously resolve some issues of parentage. The program takes into account scoring errors at an estimated level of one percent and uses allele frequency data generated from the population under investigation (Marshall *et al*. 1998). Assignment is awarded to that parent or combination of parents that assumes the highest log-likelihood ratio (LOD score) and a LOD score of three is considered the critical value above which assignment can be accepted with ninety-five percent confidence (Slate *et al*. 2000).

5.2.2.2 Statistical Analysis

CERVUS 2.0 computes and presents its statistics in separate output files. For this study a genotyping error rate of 1% was employed. No simulation was used as this is primarily necessary

should one wish to ascertain how many loci would be needed to assign parentage unambiguously (Marshall *et al*. 1998). As pointed out by Jerry *et al*. (2004), however, simulation approaches are rather limited when deciding the number of loci required for assignment as they are not a very accurate representation of how the loci will perform under real conditions.

The allele frequencies generated through an allele frequency analysis of the population genotypic data (breeders and juveniles) was used to run the assignments. Confidence levels were specified at 80% (relaxed) and 95% (strict) as was required by the program. The parameters (number of candidate mothers and fathers in the population; proportion of breeders sampled and proportion of loci typed) were set for each population specifically.

The outputs of relevance to this study were the allele frequency analysis along with the following summary statistics: number of individuals typed, number of loci, mean number of alleles per locus (Na), mean proportion of individuals typed, mean observed homozygosity (Ho), mean expected heterozygosity (He), mean PIC value, combined exclusion probability (first parent), combined exclusion probability (second parent). Polymorphic Information Content (PIC), a measure of the informativeness of the locus, is related to the expected heterozygosity and is calculated from the allele frequency data.

Hardy Weinberg equilibrium, chi-square statistics and null allele frequency estimation (all three statistics being optional extras) were additional calculations selected for completeness of the statistical analysis. The allele frequency analysis data was presented for each locus individually as well as in a set of summary statistics for data across all loci. Within the parentage analysis summary output, only the following information was relevant to this study: candidate parent assigned against the known parent for a given juvenile (both males and females run separately) and respective LOD scores.

Outputs from the parentage analyses: candidate female run against known parent (male) and candidate male run against known parent (female) were subsequently combined to determine the most likely parental pair or mother or father based on the generated LOD scores. When LOD scores for a specific juvenile assignment were almost identical for two given sets of breeders and one parent was common to both sets, the genotypes of all three individuals were compared for potential mismatches and the reasons for these mismatches. The parental pair with the least mismatches was assigned. Should the number of mismatches for both potential parent pairs be equal then only the common parent was assigned with ninety-five percent confidence.

Percentage contributions to spawning, both female and male, as well as the ratio of fast growing individuals and slow growing individuals assigned to a particular parent were calculated from this output file and are represented graphically for interpretation.

Additionally, a comparison was made of the allele frequency data generated from genotypes of a wild population (Black Rock), hatchery stock and the F1 progeny from each farm. Changes in number of alleles per locus and expected heterozygosity levels can be an indication of a loss of genetic diversity. Loss of variation at microsatellite loci has been observed in several abalone populations as a result of inbreeding occurring within the small closed population commonly maintained in a hatchery environment (Boudry *et al*. 2002; Evans *et al*. 2004; Brown *et al*. 2005). With this in mind and out of interest, allelic richness was evaluated and compared for each population, one of these being a first generation population of individuals. The findings hereof are presented as Addendum A given that this exercise fell outside of the scope of this study.

5.3 Results

5.3.1 Genotyping of Performance Recording Scheme Individuals

Nine primer sets were used to genotype PRS animals: both breeders and next generation individuals (NGI). Within this study all progeny were still first generation individuals and were assumed to have been produced from non-related breeding stock. Table 17 provides information on these selected primer sets while more detail, including the forward and reverse primer sequences and testing conditions for these microsatellite markers, are provided in Appendix 11. Multiplex reaction details are provided in Material and Methods: section 5.2.1.2.

Table 17 The nine primer sets used in parentage analyses including repeat type, fragment size, NCBI accession number and estimated allele range for each locus

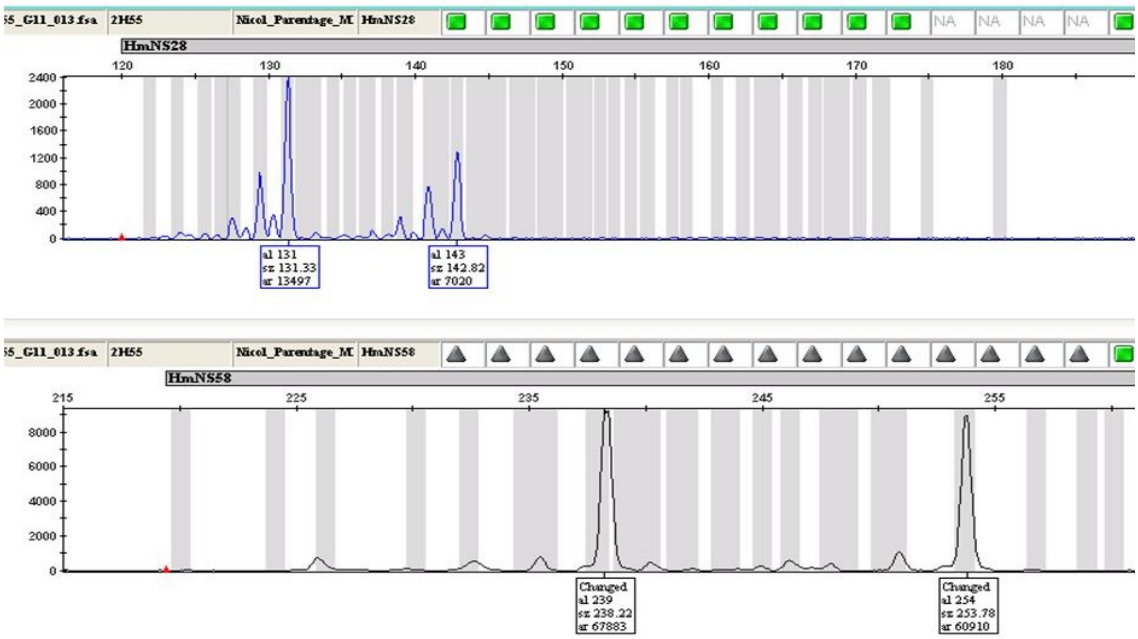
Name	Repeat	Size	Accession No.	Size Range
HmNS28D	$(CA)_{16}$	180bp	EF 033332	123 - 185
HmNS38T	$(TCAC)_{10}$	454bp	EF 367113	402 - 474
HmNS58R	$(GTT)_{8}$	246bp	EF 367119	233 - 272
HmNST7T	$(CACT)_{26}$	313bp	EF 455618	228 - 328
HmNS56D	$(CA)_{20}$	259bp	EF 455619	211 - 253
HmNR54H	$(TTAGGG)_{4}$	359bp	EF 063103	329-407
HmNR120T	$(TGAG)_{23}$	304bp	EF 121745	235-347
HmNR136D	$(CA)_{11}$	254bp	DQ 825710	211-309
HmNR281P	$(CTCAA)_{24}$	367bp	EF 512274	225-375

5.3.2 Optimization of Polymorphic Primers for Parentage Assignment

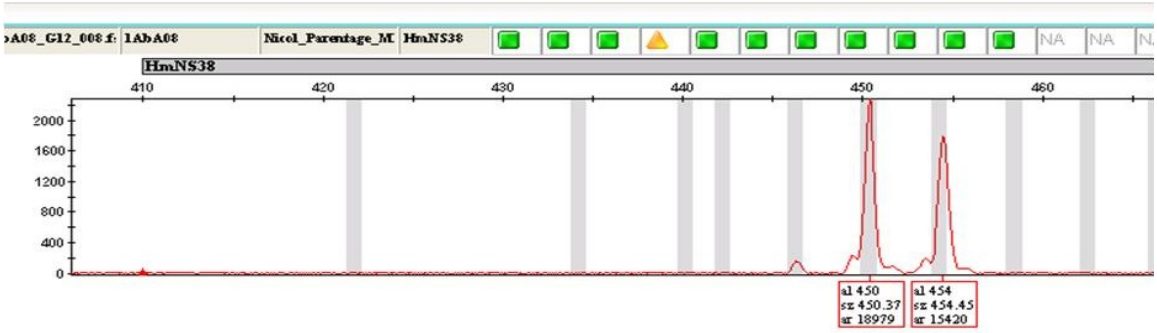
Two of the principal expenditures in parentage tests aside from the microsatellite marker development remain the genotyping costs and high time allocation, hence the use of multiplex reactions to reduce these two factors is considered advantageous. Once the final primer sets for parentage assignment were selected, these were tested in new multiplex combinations using a multiplex PCR in order to reduce the number of submissions for analysis. Variables tested included annealing temperature and DNA concentrations across the reactions.

The final combinations and the reaction conditions used are detailed in Section 5.2.1.2. Figures 22 (a) to (g) are examples of genotypes generated for different individuals, either adult or offspring, across the populations from Abagold and HIK for several of the nine microsatellite loci. The electropherograms were generated using Genemapper once scoring and manual adjustments had been completed.

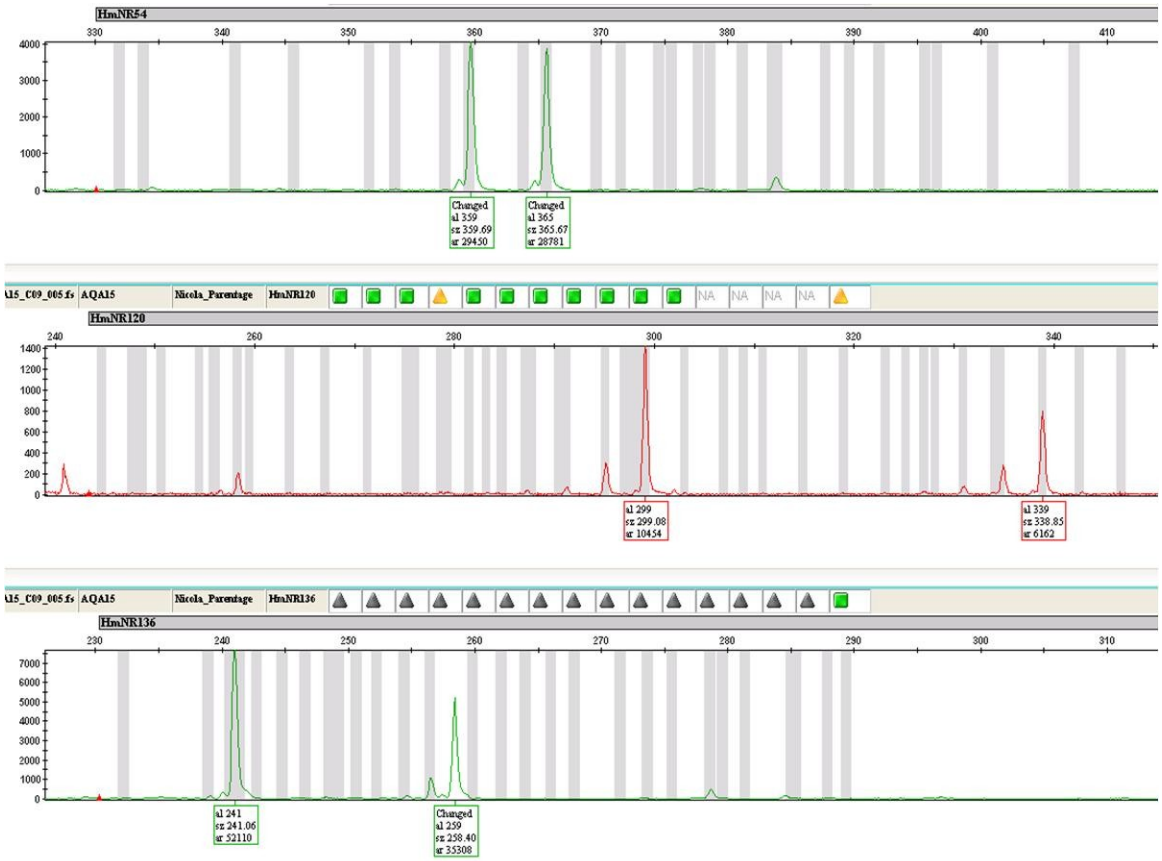
(a)



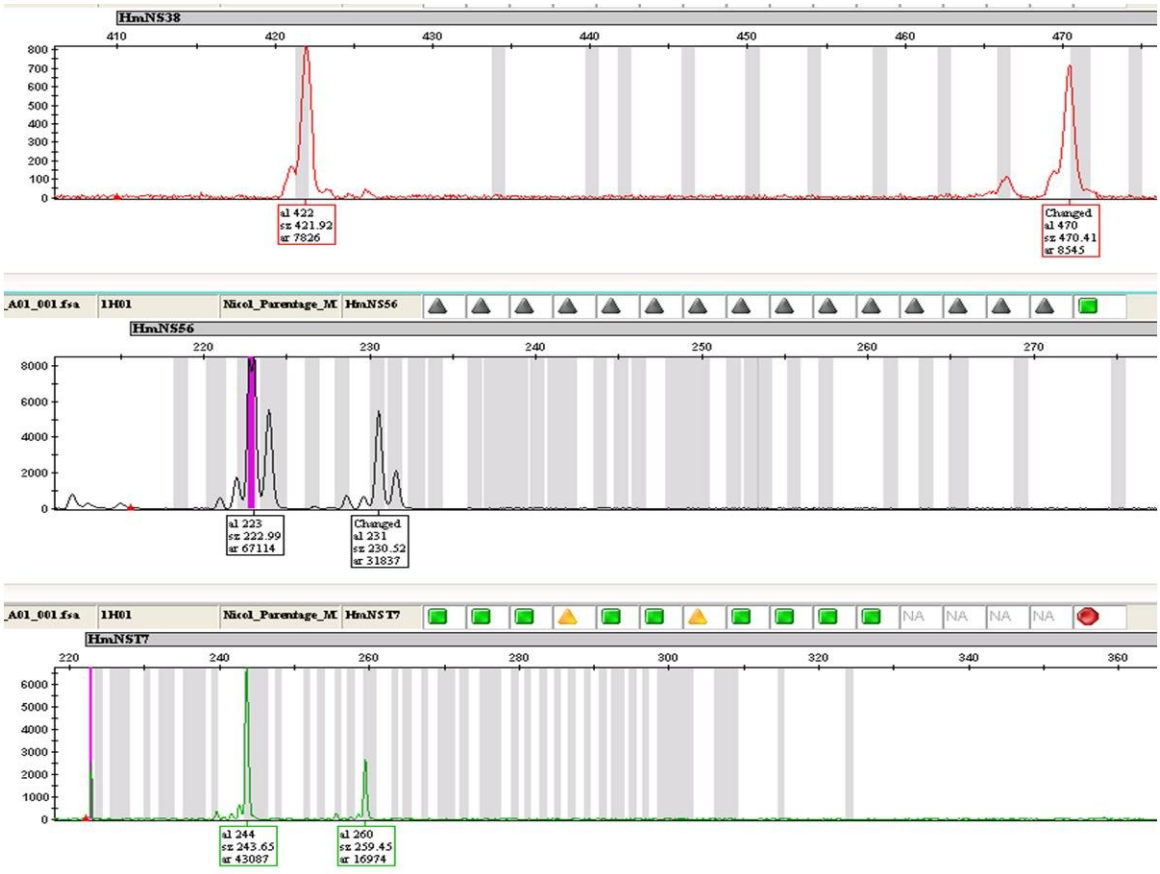
(b)



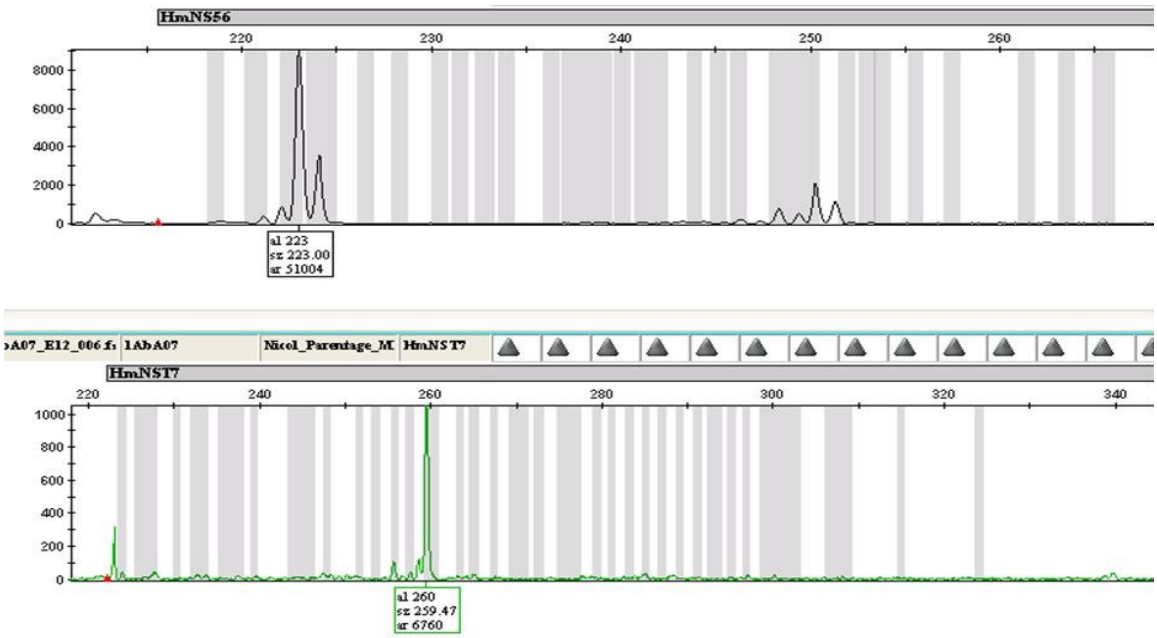
(c)



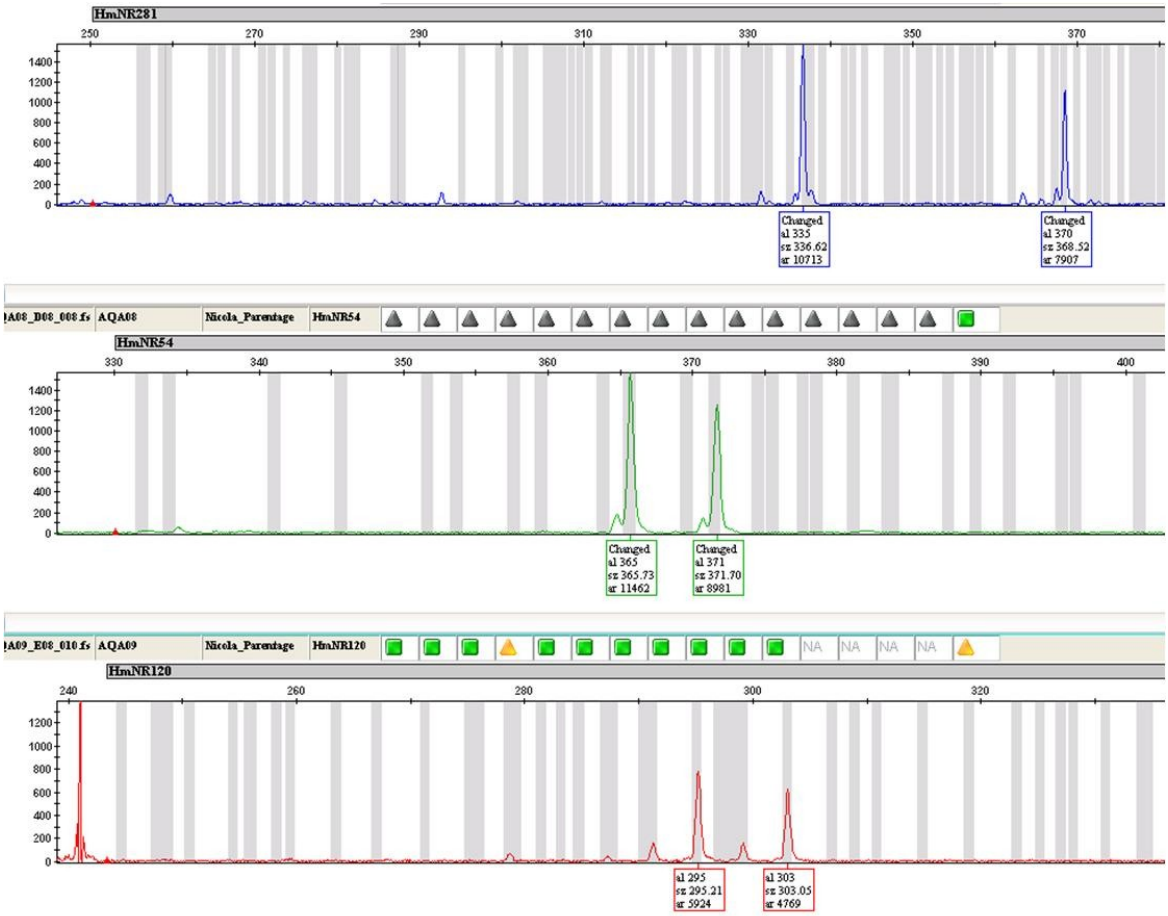
(d)



(e)



(f)



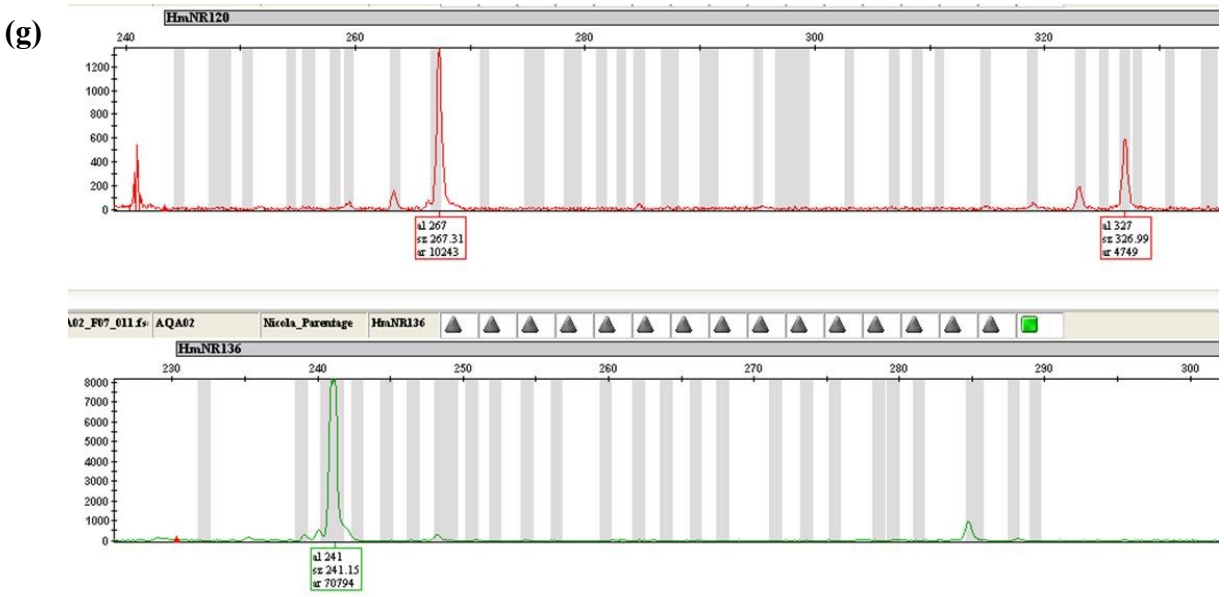


Figure 22 Electropherograms depicting genotypes for various F1 progeny from Abagold individuals reared at different locations as well as adult individuals from HIK based on different microsatellite markers: **(a)** adult male HIK55 (HmNS28, HmNS58); **(b)** juvenile located at Abagold (HmNS38); **(c)** juvenile located at Aquafarm (HmNR54, HmNR120, HmNR136); **(d)** adult male HIK01 (HmNS38, HmNS56, HmNST7) **(e)** juvenile located at Abagold (HmNS56, HmNST7); **(f)** juvenile located at Aquafarm (HmNR281, HmNR120, HmNR54); **(g)** juvenile located at Aquafarm (HmNR120, HmNR136)

5.3.2 Parentage Assignment

5.3.2.1 Statistical Results of Abagold Parentage Analyses

Allele frequency analysis of the Abagold population indicated the mean number of alleles per locus to be approximately 22.667 with a Ho of 0.682 versus a He of 0.804. The PIC of the marker suite used was high at 0.791. Excl1 represents the probability of excluding an unrelated candidate parent given that the genotype of the other parent is unknown and Excl2 represents the probability of excluding an unrelated candidate parent given that the genotype of a known parent of the opposite sex is available. The mean values of Excl1 and Excl2 were 0.549 and 0.684 respectively. When genotypes are missing, the exclusion probabilities may be lower than expected for a given marker suite. The mean null allele frequency was 8.5% which is higher than the preferred maximum of 5% for polymorphic loci.

Three loci, HmNS38, HmNS56 and HmNS58, deviated significantly from Hardy Weinberg equilibrium at the 1% level. Deviations at single loci may be as a result of natural selection acting on a nearby gene, while it is more common that the deviation indicates a potential problem in the genotyping at the specific locus such as segregating null alleles, allele drop out or other genotyping

errors. When Hardy Weinberg equilibrium was not calculated (NA), this does not, however, automatically mean that any underlying problems exist within the dataset, only that there were too few individuals to allow the program to assess the information. Deviations at several loci may indicate population substructure such as closely related or inbred individuals. Allele frequency data for each locus as well as mean values are presented in Table 18. Data was generated using CERVUS 2.0 and genotypes of both breeder and juvenile animals. The assumption made by the software is that the molecular markers are unlinked and in Hardy Weinberg equilibrium.

Table 18 Allele frequency data of the Abagold population generated by CERVUS 2.0 (n=297)

Locus	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	28	0.841	0.945	0.941	0.796	0.886	NA	0.058
HmNR136	21	0.444	0.491	0.479	0.146	0.325	NS	0.061
HmNR281	29	0.724	0.948	0.943	0.804	0.891	NA	0.132
HmNR54	11	0.732	0.743	0.702	0.344	0.520	NS	0.004
HmNS28	31	0.725	0.930	0.924	0.747	0.855	NA	0.124
HmNS38	13	0.639	0.836	0.815	0.510	0.678	**	0.132
HmNS56	22	0.407	0.504	0.490	0.153	0.332	**	0.133
HmNS58	21	0.759	0.899	0.890	0.661	0.796	**	0.083
HmNST7	28	0.866	0.940	0.935	0.780	0.876	NA	0.040
Mean values:	**22.7**	**0.682**	**0.804**	**0.791**	**0.549**	**0.684**		**0.085**

NS No significant deviation from Hardy Weinberg equilibrium

NA Hardy Weinberg equilibrium not calculated

** Indicates significant deviation from Hardy Weinberg equilibrium at the 1% level

Figure 23 illustrates that parentage could be assigned to a single parent pair for one hundred and seventy-eight individuals out of the two hundred selected offspring. One juvenile was assigned to a mother only and three juveniles to a father only resulting in a total of ninety-one percent assignment. The remaining eighteen individuals could not be assigned.

Two breeding individuals, AbG51 and AbG79 used in the spawning of PRS offspring died prior to tagging being completed and hence sampling of these animals had not yet been done. Incomplete sampling may have resulted in an increase in the percentage of unassigned offspring should these two parents have contributed to the PRS. However, this is not discernible from the results since other confounding factors, such as null alleles and genotyping errors are also potential contributors to reduced assignment.

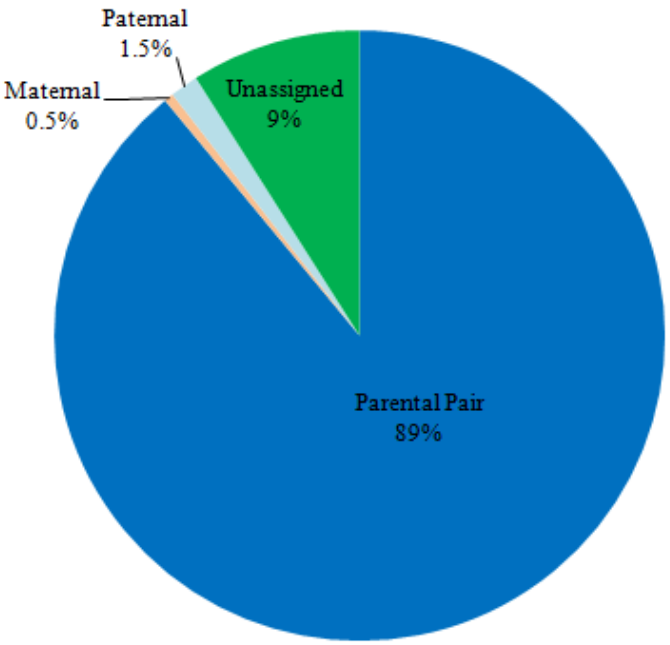


Figure 23 Percentage of F1 progeny (Abagold) assigned either to a single parent pair, a mother only or a father only while the remaining individuals remained unassigned. Assignment was completed at a genotyping error rate of 1%

A critical LOD score of a minimum of positive three was used to assign parentage with a statistical confidence of 95% (Slate *et al*. 2000). The minimum, maximum and average LOD scores obtained for each level of assignment are provided in Table 19 and show unambiguous assignment.

Table 19 The minimum, maximum and average LOD scores determined for each level of parentage assignment for Abagold

	Min LOD	Avg LOD	Max LOD
Parental Pair	3.01	4.98	9.49
Maternal	4.57	4.57	4.57
Paternal	3.48	3.93	4.34
Unassigned	1.83	2.57	2.99

Parentage models have various end applications of which reproductive success is but one of these applications (Neff *et al*. 2000b). Very often within a breeding population, only a few individuals are actually contributing to the production of the next generation. Within the Abagold population, approximately thirty percent of the F1 progeny that could be assigned were produced by twenty-one parental pairs while approximately seventy percent of the one hundred and seventy-six individuals assigned were produced by one hundred and twenty-six breeding pairs through single

offspring contributions. Figure 24 demonstrates this split in contributions. The parental pairs and the offspring to which they were assigned are provided in Appendix 12 (a) along with respective LOD scores. However, given the small sample size evaluated, it must be borne in mind that some contributions amounted to a single juvenile per adult and the question remains as to whether the contributions by these individuals can be considered value adding? Assignment would need to be performed for all three thousand offspring submitted to the PRS by each farm for a more realistic indication of the breeders' contributions to be obtained. This was not feasible for the purpose of this study nor was it an objective.

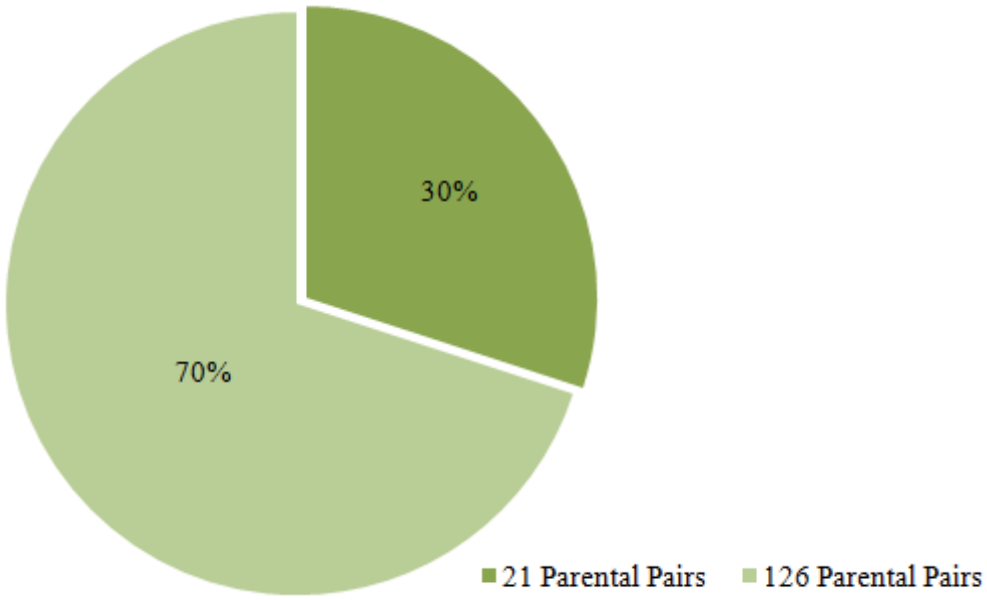


Figure 24 Contributions of individual parental pairs to assigned F1 progeny tested for faster or slower growth. Twenty-one breeding pairs produced thirty percent of the offspring tested for Abagold

The contributions can be further proportioned into maternal and paternal contributions. Due to the mass spawning practices of abalone, it is important to have an understanding of which males and females are contributing to the gene pool as a means to promote effective breeding programs and at the same time, to balance this against the potential negative effects of inbreeding depression as a result of breeding between relatives.

Within the Abagold population the contributions of female and male breeding stock to the one hundred and seventy-eight individuals assigned are depicted in Figure 25 and Figure 26 respectively. Not all breeders appeared to contribute (seven males out of forty and thirteen females out of fifty-nine were not assigned as parents to any offspring).

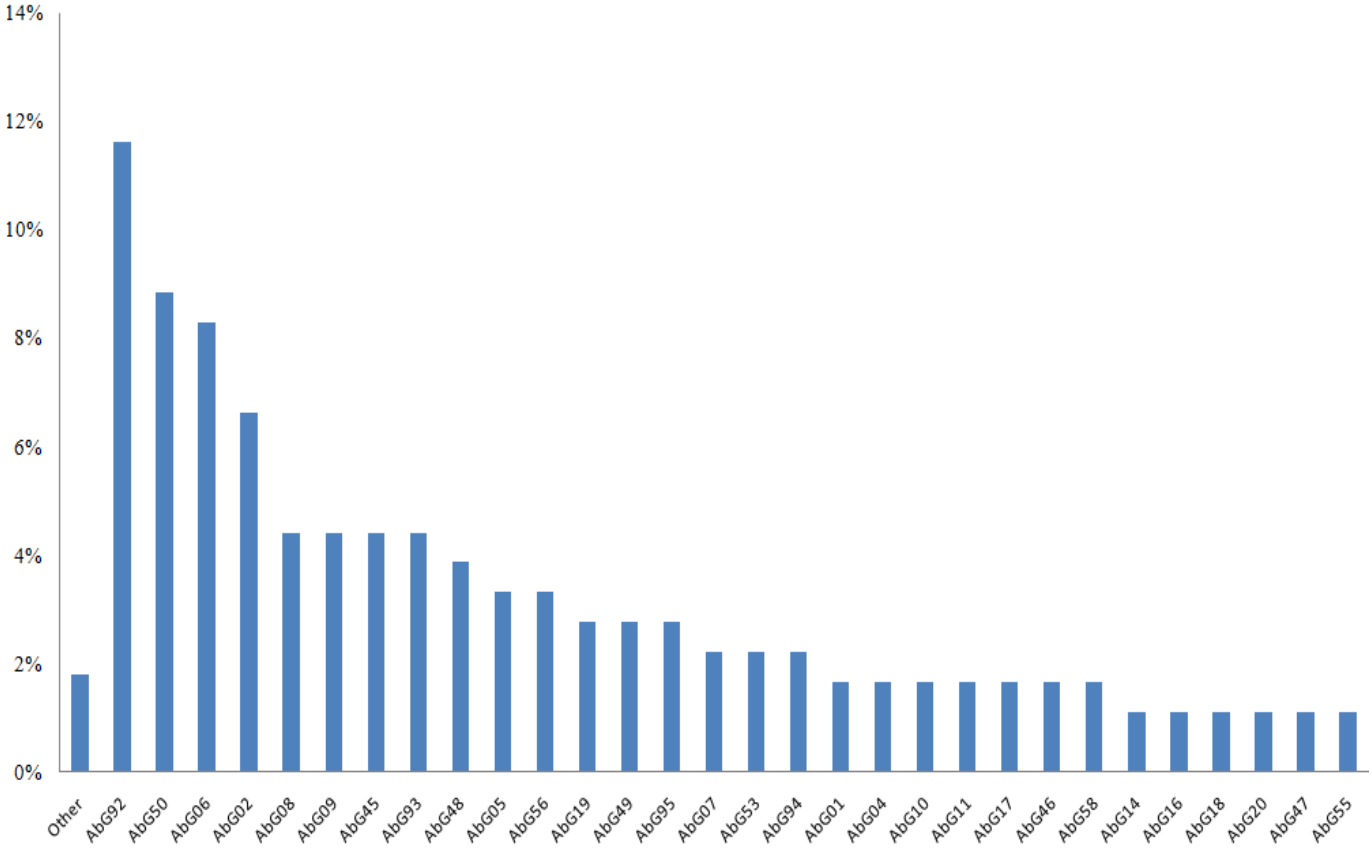


Figure 25 Percentage contributions of male breeding stock to F1 progeny (Abagold) assigned to a specific father. Of the forty individuals spawned, thirty-three males contributed to produce the PRS offspring

Thirty-three out of forty breeding males (82.5%) were found to have contributed to the one hundred and eighty-three PRS offspring assigned a specific father, although three of these breeding males only contributed a single individual. These single contributions have been grouped together under the term "other" and account for 1.8 percent of the contributions. The balance of assigned offspring was produced by the remaining thirty males. The highest contributions were made by individuals AbG92, AbG50, AbG06 and AbG02 of 11.6%, 8.8%, 8.3% and 6.6% respectively with individuals AbG08, AbG09, AbG45 and AbG93 each contributing 4.4%. The second lowest contributions were 1.1% each of the total assigned and these were made by individuals AbG14, AbG16, AbG18, AbG20, AbG47 and AbG55. The seven males making no obvious additions to the PRS should ideally be removed from the breeding program and individually monitored during controlled family breeding to determine their potential. Further recommendations are put forward in the discussion.

Forty-six out of fifty-nine breeding females (77.9%) were found to have contributed to the one hundred and seventy-nine PRS offspring assigned to a specific mother. Fourteen of these females only contributed a single offspring. These single contributions have been grouped together under the term "other" and account for approximately 8.4 percent of the contributions.

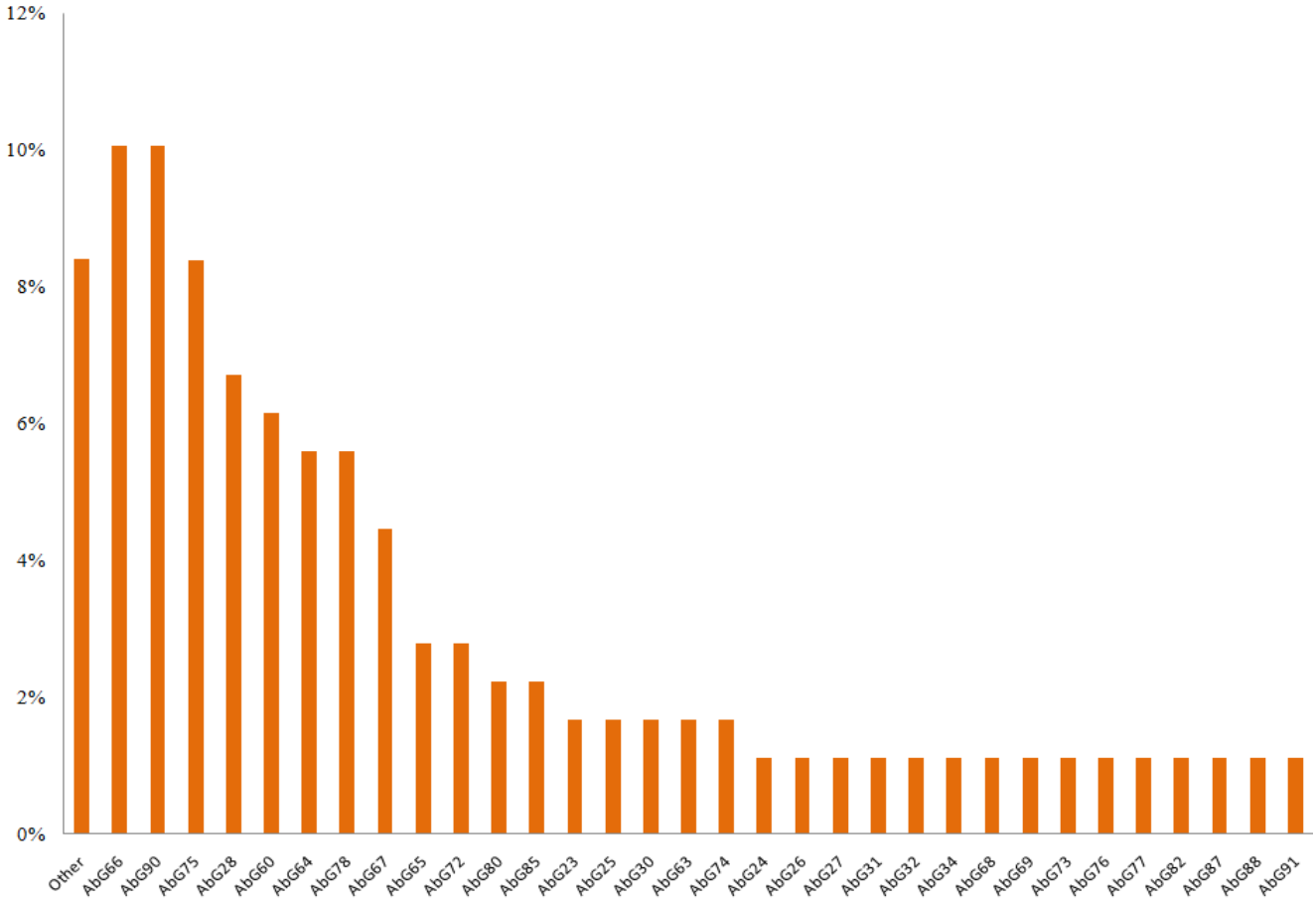


Figure 26 Percentage contributions of the female breeders to F1 progeny (Abagold) that were assigned to a specific mother. Of the fifty-nine individuals spawned, forty-six females contributed to the PRS

The balance of one hundred and sixty-five F1 progeny, were produced by the remaining thirty-two females. The highest contributions were made by individuals AbG66 and AbG90 of 10.1% each and the second highest contributions by individuals AbG75, AbG28 and AbG60 of approximately 8.4%, 6.7% and 6.1% respectively. It is recommended that those thirteen females that made no apparent contribution be removed from the breeding program and monitored during family breeding to provide a clearer assessment of their spawning success. The discussion encompasses further recommendations with regards the other females.

Another important output of the parentage assignments in this study was the demonstration of any trends (if present at this early stage) regarding faster and slower growth and which male or female breeders may be contributing individuals to either of these groups. The results for males and females respectively are shown in Figure 27 and Figure 28 with no trend being evident. A single breeding individual contributed in most cases to both faster and slower growing F1 progeny in varying ratios. This result was expected given that as yet, no selection for growth is being made. The breeding program is still in its infancy and the outputs of this study will assist in laying the foundation for future MAS.

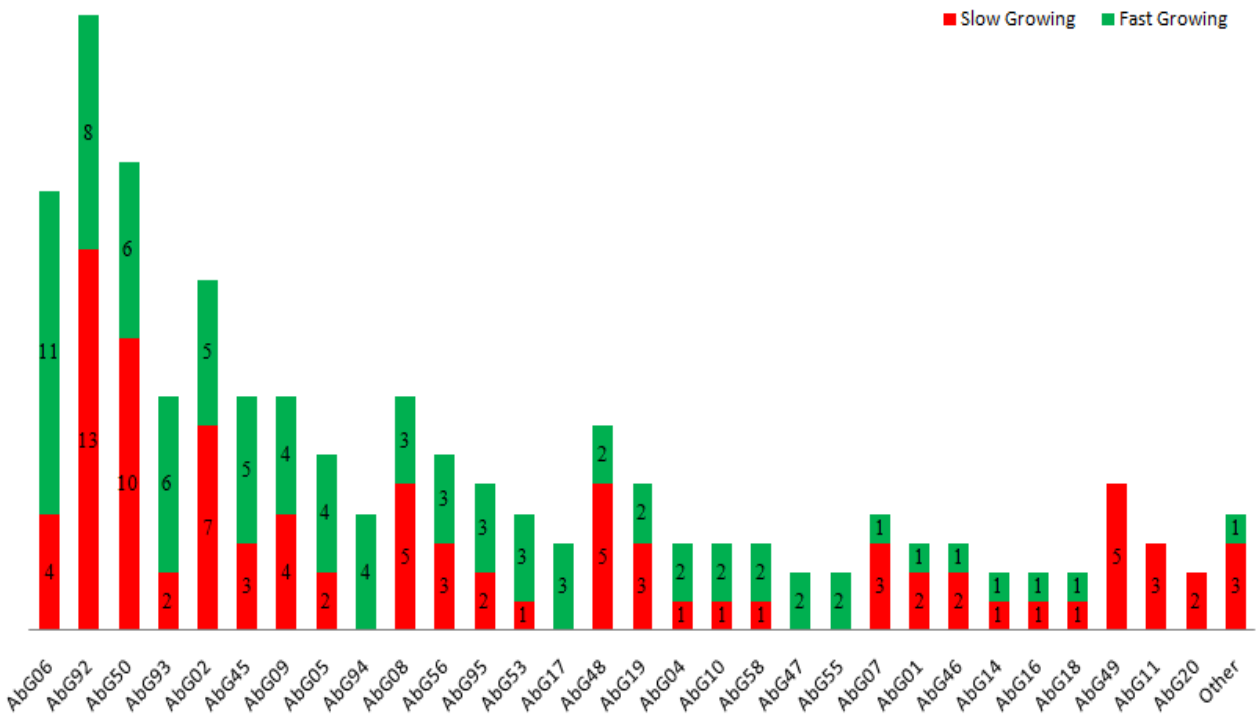


Figure 27 The contributions of breeding males (Abagold) to faster or slower growing F1 progeny indicated in terms of number of offspring. A single adult, in most instances, produced both faster and slower growing individuals

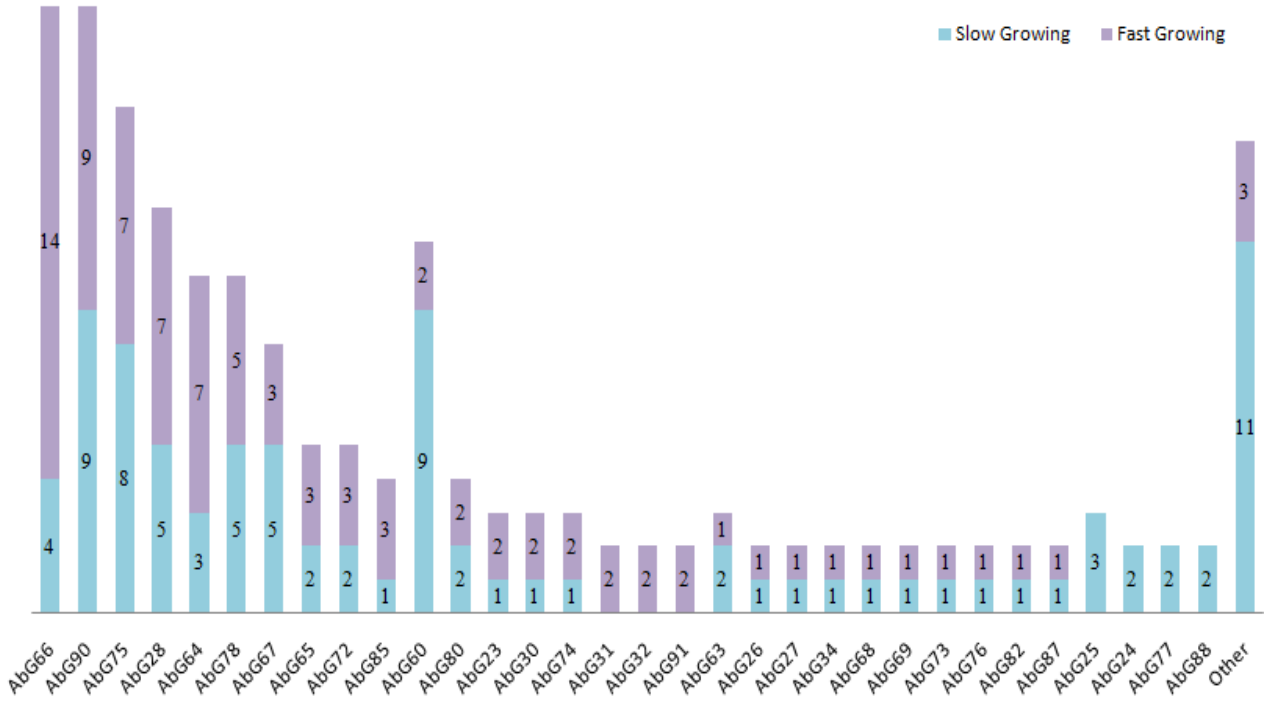


Figure 28 The contributions of breeding females (Abagold) to faster or slower growing F1 progeny shown in terms of number of offspring. In most instances, a single parent produced both faster and slower growing individuals

In general, the split between faster and slower growing individuals produced by a single adult is quite even. Individuals AbG11, AbG20 and AbG49 produced only slow growing individuals while AbG17, AbG47, AbG55 and AbG94 produced only fast growing individuals, however, the relatively small contributions made by each adult to this study sample do not allow for any conclusive deductions to be made regarding growth potential.

Similarly, given that these are the female partners to the breeding males, individuals AbG31, AbG32 and AbG91 contributed only faster growing individuals while AbG24, AbG25, AbG77 and AbG88 produced only slower growing individuals. Again, no significant conclusions can be drawn from such small sample numbers and it is important to note that the terms "faster" and "slower" are used only as indicators of potentially better or worse growth performance and that these positions may not be maintained at a later stage of the juveniles development especially once sexual maturity is reached.

5.3.2.3 Statistical Results of HIK Parentage Analyses

The allele frequency output of HIK's population calculates the mean number of alleles per locus to be approximately 25.0 with a mean Ho of 0.700 versus a mean He of 0.826. The PIC of the marker suite used was higher than for Abagold at a mean value of 0.814. Regarding exclusion probabilities, the mean values realised were 0.567 (first parent) and 0.707 (second parent). The mean null allele frequency was the same as for Abagold at a level of 0.085 (8.5%). Allele frequency data for all loci as well as mean values are provide in Table 20.

Table 20 Allele frequency analyses of HIK abalone population using CERVUS 2.0 (n=301)

Locus	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	28	0.796	0.938	0.933	0.774	0.872	**	0.0817
HmNR136	28	0.572	0.628	0.615	0.260	0.455	NS	0.0447
HmNR281	36	0.704	0.942	0.937	0.787	0.880	NA	0.1432
HmNR54	14	0.696	0.770	0.738	0.391	0.570	NS	0.0478
HmNS28	28	0.754	0.940	0.935	0.777	0.874	NA	0.1086
HmNS38	16	0.640	0.789	0.769	0.442	0.621	**	0.1023
HmNS56	21	0.446	0.585	0.571	0.218	0.409	**	0.1511
HmNS58	24	0.805	0.904	0.895	0.673	0.805	NS	0.0568
HmNST7	30	0.889	0.939	0.934	0.777	0.874	NA	0.0263
Mean values:	**25.0**	**0.700**	**0.826**	**0.814**	**0.567**	**0.707**		**0.085**

NS No significant deviation from Hardy Weinberg equilibrium

NA Hardy Weinberg equilibrium not calculated

** Indicates significant deviation from Hardy Weinberg equilibrium at the 1% level

Three loci, HmNS120, HmNS38 and HmNS56, deviated significantly from Hardy Weinberg equilibrium at the 1% level and could perhaps be attributable to a fairly high level of segregating null alleles, inadvertent scoring errors or linked microsatellite loci. Hardy Weinberg equilibrium was not calculated (NA) for three loci and for the remaining three loci, deviation from Hardy was not significant. The fact that deviations from HW occurred at more than two loci suggests that there may be some related individuals among the breeders and hence some degree of inbreeding may already be present.

Figure 29 shows that parentage assignment was achieved for one hundred and eighty out of the two hundred selected offspring which amounts to an overall assignment of ninety percent. One hundred and seventy-six offspring could be assigned to a parent pair (88%), two offspring were assigned to a mother only (1%) and two offspring to a father only (1%). The remaining twenty individuals could not be assigned. As previously, a minimum LOD score of three was used to assign parentage with a statistical confidence of ninety-five percent.

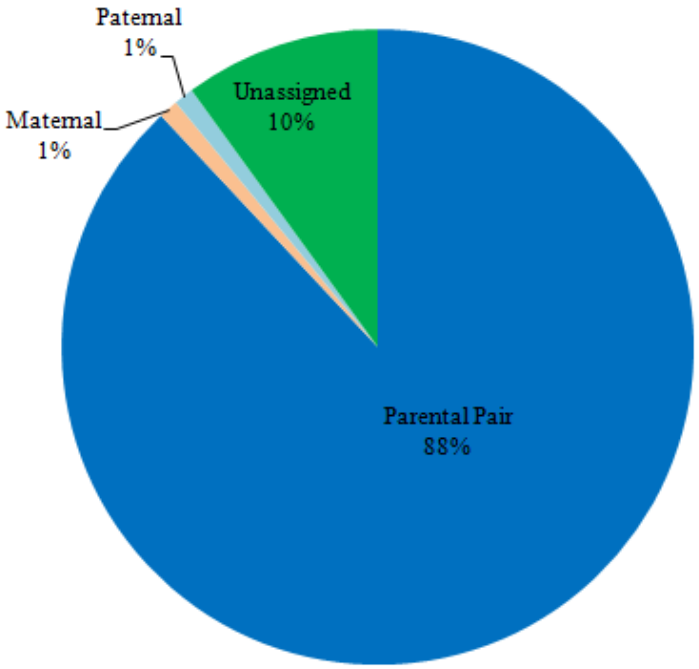


Figure 29 Percentage of F1 progeny (HIK) assigned either to a single parent pair, a mother only or a father only while the remaining individuals were unassigned. The assumption of 1% genotyping error rate was made

Seven of the breeders died subsequent to the completion of spawning for the PRS however DNA samples were available for inclusion in parentage analyses. The Sample ID and gender for each of these animals are provided in Table 21. The obvious implication is that these breeders are no longer available for future breeding programs.

Table 21 Sample ID, tag number and gender information pertaining to the deceased HIK breeders

Sample ID	Tag Number	Gender
HIK 13	D2	Male
HIK 33	Untagged	Female
HIK65	N1	Male
HIK 66	B5	Male
HIK 67	B4	Male
HIK92	Untagged	Female
HIK 93	Untagged	Female

The percentage contributions of the above individuals are provided in Figures 31 and Figure 32 for males and females respectively. The minimum, maximum and average LOD scores for each level of assignment for HIK are provided in Table 22.

Table 22 The minimum, maximum and average LOD scores determined for each level of parentage assignment at HIK.

	Min LOD	Avg LOD	Max LOD
Parental Pair	3.02	5.63	9.83
Maternal	6.05	6.56	7.08
Paternal	6.53	7.52	8.5
Unassigned	1.76	2.57	2.94

Within the HIK breeding population, approximately twenty-eight percent of the F1 progeny that were assigned were produced by twenty-one parental pairs while approximately seventy-two percent were produced by one hundred and twenty-six breeding pairs through single offspring contributions. These single contributions do not add value and the recommendation is to exclude these individuals from future breeding schemes or assess their performance when spawning the families for phase II of the PRS. Figure 30 demonstrates this split in contributions. The parental assignments with respective LOD scores for each juvenile are detailed in Appendix 12(b) and the same caution is recommended when drawing conclusions regarding breeding stock contributions as a whole given that such a small portion of PRS individuals were sampled for this study.

This contribution can be further divided into maternal and paternal contributions (Figure 30).

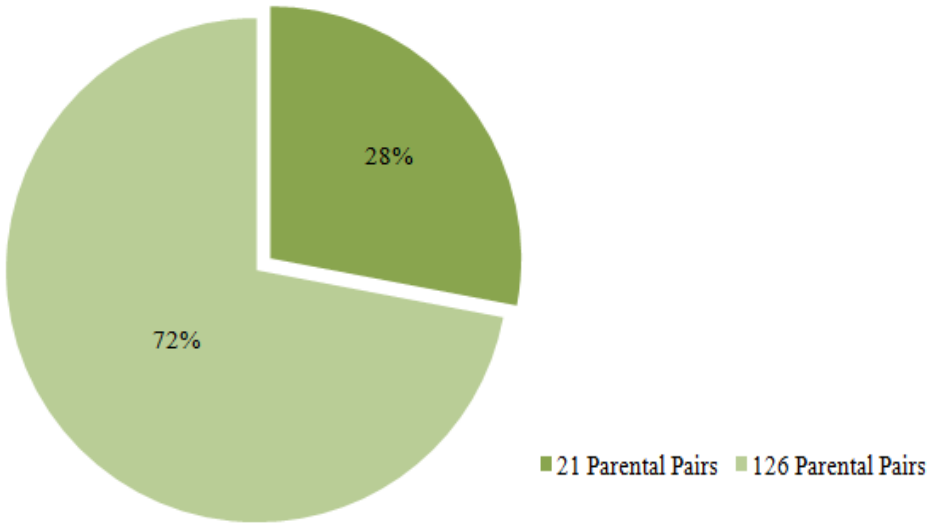


Figure 30 Contributions of individual parental pairs to assigned F1 progeny (HIK) selected for faster or slower growth. Twenty-one breeding pairs produced just over twenty-eight percent of the offspring tested for HIK

Within the HIK population the contributions of male and female breeding stock to the one hundred and eighty individuals assigned are depicted in Figure 31 and Figure 32 respectively.

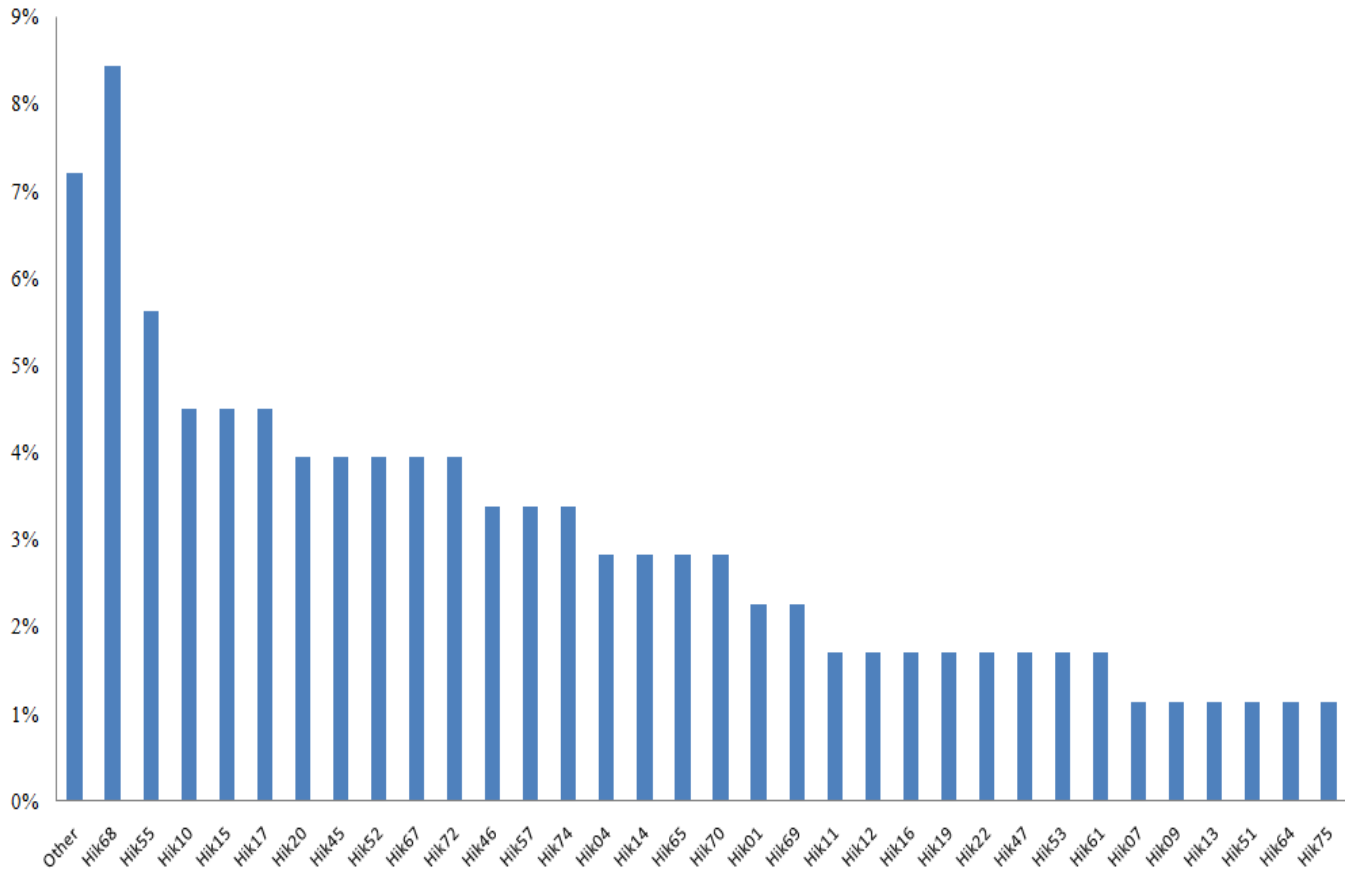


Figure 31 Percentage contributions of male breeding stock (HIK) to the F1 progeny assigned to a specific father. Of the fifty-three individuals spawned, forty-five males contributed to produce the PRS offspring

Forty-five out of fifty-three breeding males (85%) were found to have contributed to the one hundred and eighty PRS offspring assigned a specific father, although twelve of these breeding males only contributed by means of a single individual. These single contributions have been grouped together under the term “other” and account for approximately seven percent of the total contributions.

The remaining thirty-three males made larger contributions with the highest contributions made by individuals HIK68 and HIK1455 of 8.4% and 5.6% respectively while the second highest contributions were made by individuals HIK10, HIK15 and HIK17 of approximately 4.5% each. The second lowest contributions of 1.1% each of the total assigned were made by individuals HIK07, HIK09, HIK13, HIK51, HIK64 and HIK75.

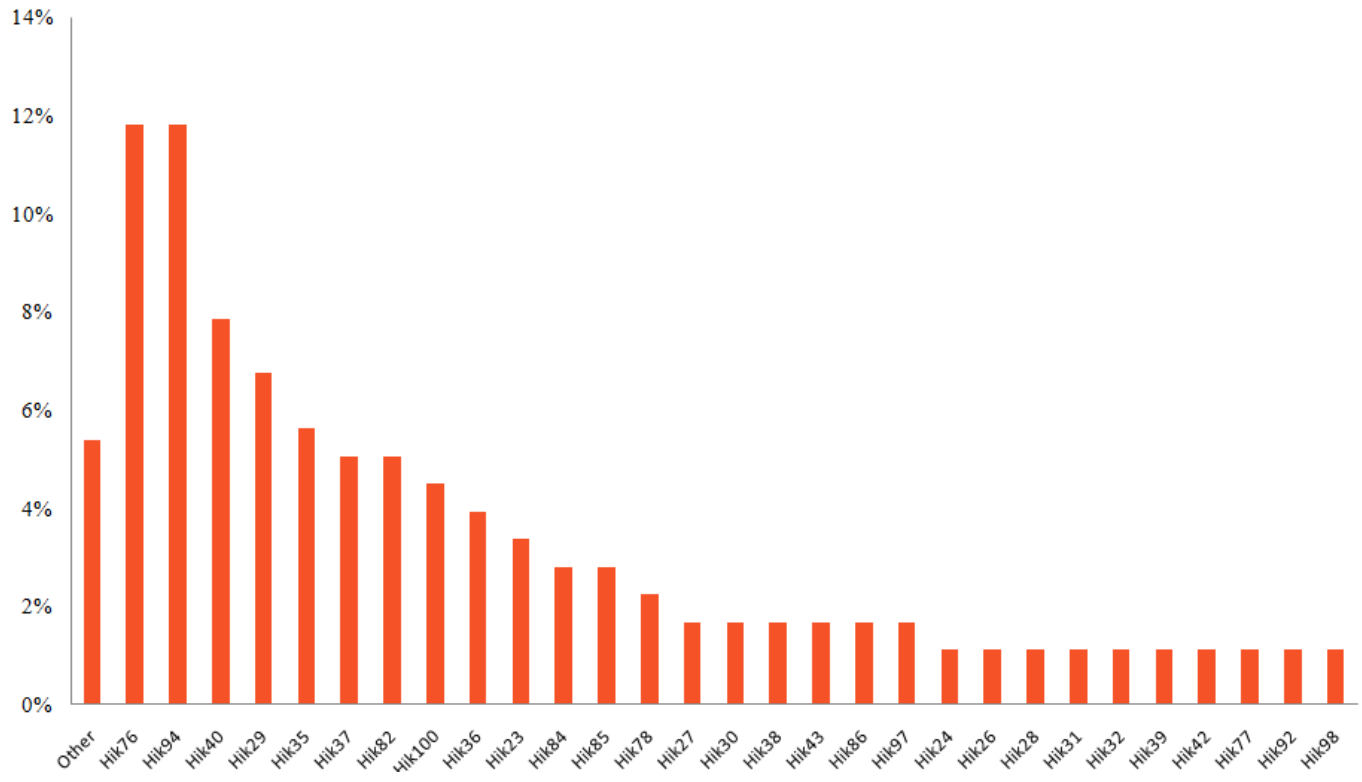


Figure 32 Percentage contributions of female breeding stock (HIK) to assigned F1 progeny. Of the forty-eight individuals spawned, thirty-eight females contributed to the PRS

Thirty-eight out of forty-eight breeding females (79%) made contributions to the PRS offspring assigned to a specific mother. Nine of these breeding females could only be assigned to a single offspring and therefore have been grouped together under the term “other”, accounting for approximately 5.4 percent of the contributions. The balance of assigned individuals was produced by the remaining twenty-nine females. The highest contributions were made by individuals HIK76 and HIK94 of 11.8% each with second highest contributions being made by individuals HIK40,

HIK29 and HIK35 of 7.9%, 6.7% and 5.6% respectively. Ten female breeders contributed only 1.1% each to the total assigned. The ten adult females which made no apparent contribution to the juveniles tested may have contributed to other offspring in the total PRS but it is unlikely, on the basis of these findings, that their additions would be significant and it is recommended that they be removed from the breeding program and observed independently under more controlled spawning conditions to reassess their potential.

With regards to potential trends in faster and slower growth and which male or female breeding individuals may be contributing to either of these groups, no obvious trends were observed within this study for this population. The results for males and females respectively are shown in Figure 33 and Figure 34. A single breeding individual contributed in most cases to both faster and slower growing F1 progeny in varying ratios. Again this was an expected result given that as yet, no selection for growth is being made.

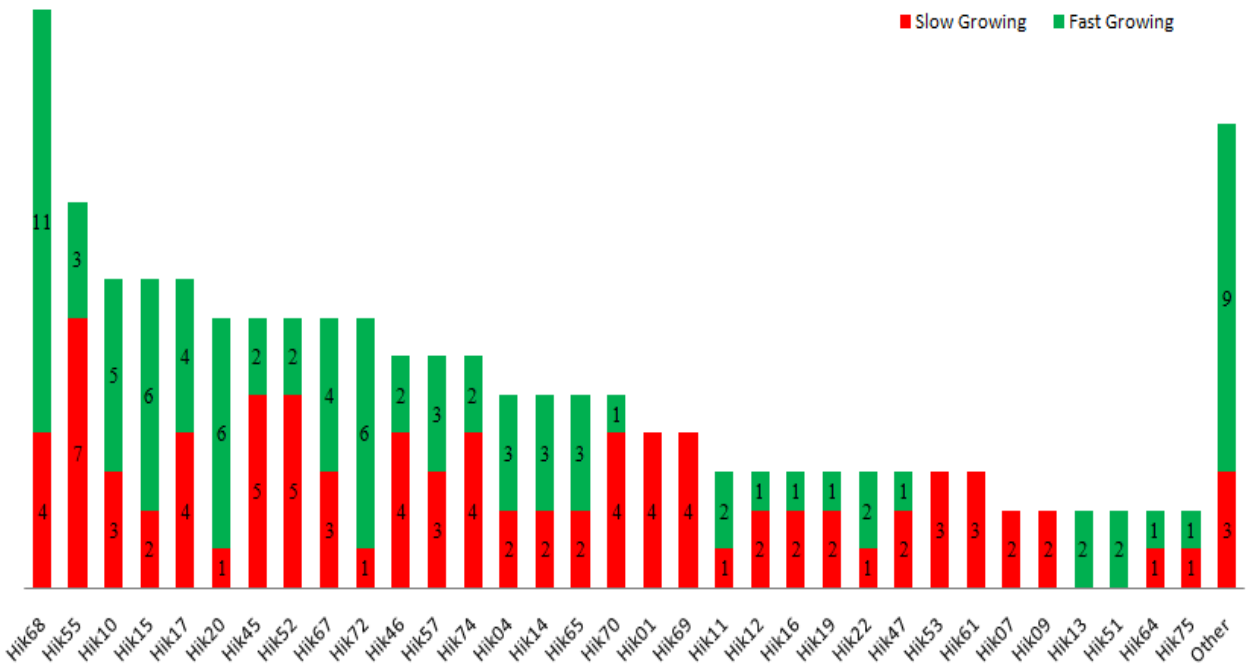


Figure 33 Contributions of breeding males (HIK) to faster or slower growing offspring indicated in terms of number of offspring. In the majority, both faster and slower growing individuals were produced by a single adult

Individuals HIK01, HIK07, HIK09, HIK53, HIK61 and HIK69 produced only slower growing individuals while HIK13 and HIK51 produced only faster growing individuals, however, the small contributions made to the F1 generation does not allow for any conclusive deductions to be made regarding growth potential in the longer term.

Similarly, for the breeding females, individuals HIK26 and HIK78 contributed only faster growing individuals while HIK32, HIK43 and HIK97 produced only slower growing individuals. Again, no significant conclusions can be drawn from such small numbers or at such an early stage of the PRS and growth assessments.

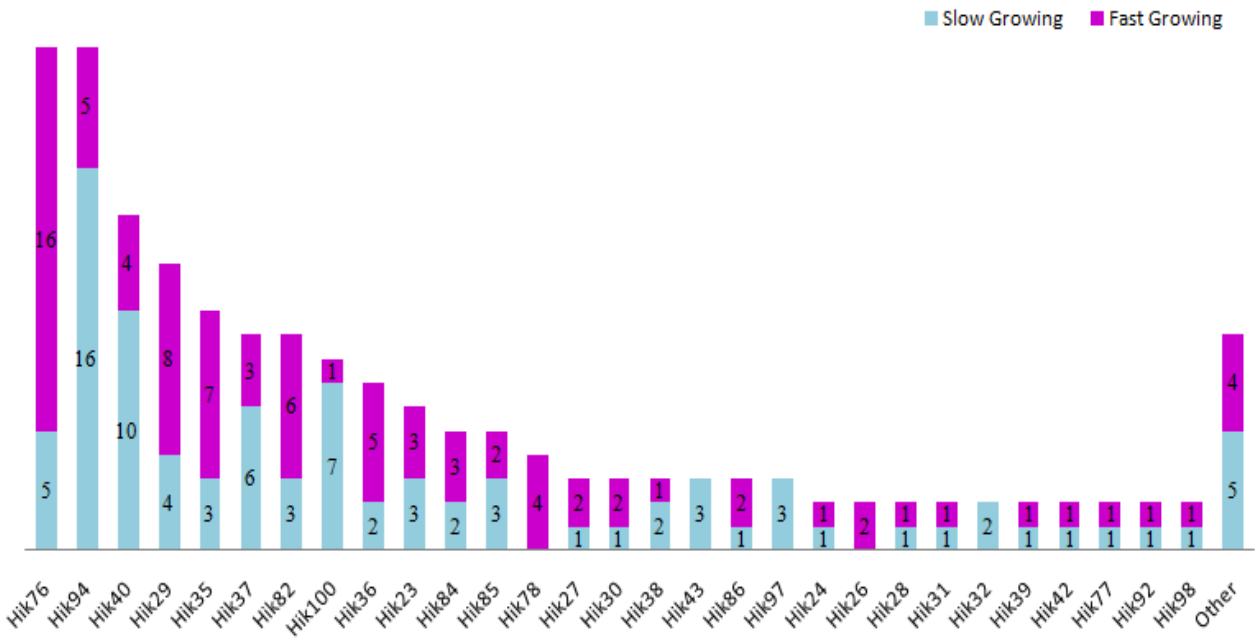


Figure 34 Contributions of breeding females (HIK) to faster or slower growing offspring indicated in terms of number of offspring. In the majority, a single adult individual produced both faster and slower growing offspring

5.3.3 Shortcomings and Sources of Error

(1) Loss of breeding stock animals contributing to the Performance Recording Scheme, particularly Abagold where tissue samples were not available for these deceased animals. Incomplete sampling may lead to an increased percentage of unassigned offspring and inaccurate estimation of breeding stock contributions to spawning.

(2) Typing errors that may occur when compiling data and input files as well as during processing of output files for relevant statistics.

(3) Allele scoring errors. Genotyping was not repeated for individuals unless scoring was uncertain. When individuals were re-genotyped and allele scores differed from the first set, these genotyped were reassessed and if the difference was a homozygous genotype versus a heterozygous genotype, the heterozygous genotype was used in the analyses. If the alleles differed in entirety, alleles were not scored for this individual at this specific locus.

(4) The null allele frequency of the locus, an estimate of which is provided by CERVUS 2.0. A high null allele frequency can reduce the level of assignment.

(5) Allele dropout due to changes in the DNA sequence at primer binding sites. This may result in the non-amplification of the allele in some individuals and hence a mismatch in genotypes between parents and offspring.

5.4 Discussion

Viana (2002: as cited in Lucas *et al*. 2006) emphasises that a vital priority for abalone producers is the increase of growth rate through selective breeding schemes. Primmer *et al*. (1999: as cited in Li *et al*. 2004), on the other hand, maintain that an essential criterion for successful hatchery management is the monitoring and maintaining of genetic variability within hatchery stocks. Both these statements are closely related as the latter is vital to achieving the former. Within hatchery environments, the success of family-based selective breeding programs is linked strongly to the effective population sizes utilized for establishing new generations and therefore promoting these high levels of genetic variation is of utmost importance (Aho *et al*. 2006).

Small founding populations ultimately result in lower genetic variation in subsequent generations and diversity may erode even further with the added effect of genetic drift. Maintaining an effective population size (Ne) is therefore a key consideration in ensuring a high level of diversity within an isolated breeding population (Aho *et al*. 2006); this is especially true for aquaculture species such as abalone, where contributions to spawning are invariably made by only a few individuals within the breeding group (Boudry *et al*. 2002; Evans *et al*. 2004, Brown *et al*. 2005). As a direct consequence of variable reproductive success, or in other words varying parental contributions to mass-spawning, a large reduction in Ne/N ratios (where N is the actual number of individuals) is observed in both the wild and hatchery populations. This extreme reproductive variability can be ascribed to three primary factors: variable sperm-egg interactions, gamete quality and differential viability of genotypes (Boudry *et al*. 2002) and has been observed in a parentage assessment study conducted by Herbinger *et al*. (1995) on rainbow trout.

Mass-spawning practices require no prior knowledge of existing pedigrees, however, the obvious implication of variable spawning success, is the inevitable inbreeding of related individuals (Norris *et al*. 2006) when multiple generations are involved in mass-spawnings. Employing more structured and controlled family breeding protocols, based on pedigree information for all breeding individuals, minimises the risk of unintentional breeding between relatives.

Since a selective breeding program is one of the broader project objectives and the realisation of genetic gains from the implementation of a family-based walk-back breeding scheme depends highly on a sound knowledge of existing pedigrees, a parentage assignment study was conducted for Abagold and HIK hatchery populations. The primary goal of these assignments was to evaluate

the reproductive success of the current breeding stock, to assess any potential trends in contribution to growth by specific individuals in succeeding generations (either faster or slower growth) and to make recommendations based on this output, regarding the inclusion or exclusion of specific animals in phase II of the PRS.

The number of loci required to assign parentage varies and is dependent on factors such as the overall informativeness of the marker suite (reflected in the mean PIC value), the combined exclusion probabilities, the number of potential parents and the number of offspring to be assigned (Norris *et al*. 2000). The higher the PIC value and combined exclusion probabilities and the lower the numbers of candidate parents and offspring to be assigned, the fewer loci will be required to accurately assign parentage. Nine loci were used in this study to assign parentage in two abalone populations consisting of approximately one hundred potential parents (for which gender was known) and two hundred offspring each. Within other parentage studies, anywhere from four to eleven polymorphic loci were employed, all with high levels of assignment achieved (Herbinger *et al*. 1995; Norris *et al*. 2000; Villanueva *et al*. 2002; Jerry *et al*. 2006b; Isberg *et al*. 2004; Dong *et al*. 2006).

The percentage assignment of appropriate candidate parents to juveniles was high for both Abagold and HIK populations. A total of 91% of individuals could be assigned within the Abagold population while 90% assignment was achieved for HIK. The breakdown of assignments into single parental pairs, maternal or paternal only, is provided in section 5.3.2.1 Figure 23 and section 5.3.2.3 Figure 29 respectively. These results are highly comparable to those achieved in a study of Chinese shrimp (Dong *et al*. 2006) wherein five microsatellite loci were used to successfully assign 90.7% of 215 progeny to a candidate parental pair within mixed family groups. Villanueva *et al*. (2002) in a study on fish populations with a finite number of parents and offspring, suggest that nine five-allele loci or four ten-allele loci are sufficient for assigning 99% parentage resulting from 100-400 crosses. These recommendations, however, are based purely on simulated data and may not accurately reflect the true scenario given the myriad of factors influencing assignment success that may not be taken into account when conducting a simulation. Caution should therefore be exercised when basing decisions on loci number using simulated data that is not supported by real data. Herbinger *et al*. (1995) demonstrated 91% parental pair assignment success out of one hundred candidate parents using four to five microsatellite markers; while in strong contrast, Jerry *et al*. (2004) assigned only 47% of Kuruma shrimp progeny spawned from 30 dams and 150 potential sires when using six loci. The assignment success achieved in this latter study was found to be much lower than what was expected from the simulation results using the same six markers. Norris *et al*. (2000) experienced variable accuracy in parentage assignments in Atlantic salmon, finding

that the precision of the markers was highly dependent on the number of putative parents involved and the number of loci and their inherent polymorphism. More families in a breeding program lead to a higher number of loci being required to assign parentage with 95% confidence. Since the results in this current study were high or, at the least comparable to other parentage studies, the number of loci used was deemed to be adequate. Should null allele frequencies be reduced through the redesign of specific primer sets, the overall informative of the marker suite may increase and therefore, the number of loci required for assignment may decrease even further.

Parentage assignment was completed using CERVUS 2.0 (Marshall *et al*. 1998) which is a likelihood-based assignment method. The program uses allele frequency data generated from actual population data to run its analyses and takes cognisance of genotyping errors and null alleles within the algorithm, allowing for better and more accurate assignments. Even though a likelihood method was used in its entirety for this parentage study, Taggart (2007) suggests that the employment of a number of different software programs with varying strengths and weaknesses and based on different principles of assignment (e.g. likelihood versus exclusion-based methods), will likely improve the success of assignments. For example, exclusion methods are very accurate when data is essentially error free or parentage is being used to assess heritability of a microsatellite marker within known pedigrees. However, this method is less robust when working with data where null allele frequencies, scoring errors or linkage between markers may be present. A combination of methods may allow for additional assignments to be made when one technique is unable to assign parentage to an individual based on the particular algorithm employed and the underlying assumptions of the equations. This is especially true when data may be in violation of these assumptions; for example HWE, unlinked or very weak linkage between markers etc.

Allele frequency data for Abagold revealed the following mean values: Na = 22.67, Ho = 0.682, He = 0.804, PIC = 0.791, Null Freq = 8.5%. Similarly the allele frequency analysis for HIK was: Na = 25.0, Ho = 0.70, He = 0.826, PIC = 0.814, Null freq = 8.5%. In both populations, three molecular loci deviated significantly at the 1% level from Hardy Weinberg equilibrium. These results are very comparable, although the Abagold population appears to be slightly less genetically diverse, evident by the lower average number of alleles per locus when compared to the HIK population. This however, is unlikely to be significant. Observed heterozygosity was lower for Abagold and since null allele frequency for both populations was equivalent, this may be due to a higher percentage of genotyping errors within this data set. The data set for Abagold was the first to be processed and scoring errors are likely to decrease as experience and familiarity with a technique increases. PIC value of the marker suite was also less for the Abagold population which may be as a result of incomplete sampling of this population. The genotypes for two individuals could not be

assessed due to DNA samples not being available for these animals.

Castro *et al*. (2004) in an assessment of potential sources of error within a parentage study on turbot found that many single mismatches in parent-offspring pairs were as a result of scoring errors that were not detectable prior to family analyses being conducted. Glaubitz *et al*. (2001: as cited in Castro *et al*. 2004) also suggests that parentage assignment can be severely impeded by allele-calling errors at several loci being greater than 5%.

Genotyping errors were not quantified in this study although several visual comparisons were made between genotypes of potential parents and offspring. The primary reason for mismatches appeared to be homozygous allele pairs, indicating a potential null allele; or alleles that differed by a repeat unit in each individual, indicating a potential allele calling error. This was most obvious with dinucleotide repeats, possibly as a result of the stutter patterns that were present with all the dinucleotide motifs employed in this study. Although a default error rate of 1% was used in CERVUS 2.0, this may in fact have been higher, as evident from the following studies.

An error rate of 2% to 3% per allele scored was reported by O'Reilly *et al*. (1998: as cited in Norris *et al*. 2000; Dong *et al*. 2006) in a parentage study on communally reared Atlantic salmon offspring. Further to this, a number of studies have been conducted on the effect of inaccurate pedigree data on longer term genetic gains and the impact can become catastrophic for a selection program if not managed appropriately; with decreases of 4% - 15% reported for dairy cattle as a result of 10% to 15% errors in genotype recording (Dodds *et al*. 2005).

Repeating genotyping of all individuals a minimum of three times at all loci would be an ideal situation. Simultaneously analysing all results for an individual for a particular locus would allow comparisons to be made and any scoring errors to be corrected prior to using the genotypes in a parentage analysis. Loci that are consistently difficult to score or are not reproducible across consecutive runs could then be excluded from future analyses. Obviously, financial and time constraints do not often allow for genotyping to be repeated, for accuracy and consistency of allele-calling to be confirmed or for the percentage of genotyping errors to be calculated.

For this study, several individuals were rerun and rescored. This repetition was completed when scoring was questionable or allele positions unclear. The two sets of results were then compared for confirmation purposes. If differences in calling were a homozygous genotype versus a heterozygous genotype, the latter call was scored for the locus. Should no overlaps in allele calling be evident, the individual was not scored for this locus. Despite repeat analysis being completed for some individuals, this was not done for all individuals typed. Genotyping errors may still have occurred in the fingerprints for those individuals for no repeat scoring was carried out.

The mean null allele frequency for both the test populations was 8.5%. This is higher than was observed for other parentage studies, where null allele frequency ranged from 0% to 0.41% (Dong *et al*. 2006), 0% to 3% (Castro *et al*. 2004) and 2% to 8% (Jerry *et al*. 2004). It is also higher than the maximum recommended level of 5% (Marshall *et al*. 1998). Given that for this same suite of markers tested on the Black Rock wild abalone population, the null allele frequency was only 2%, it is very likely that the increased frequency observed here is either as a result of increased genotyping errors within the larger data sets; allele dropout due to multiplexed reactions in which the template DNA concentration may have been too low to accommodate multiple primer sets or inadequate DNA quality (i.e. sample contained contaminants that affected PCR). A combination of all three factors may, in fact, have been at play. Weber and Wong (1993) report high mutations rates of 10^{-4} to 10^{-2} per locus per gamete per generation for microsatellite loci which may have resulted in point mutations at primer binding sites. Changes in the ability of a primer to bind to the DNA may be reflected by an increase in null alleles in subsequent generations. Primer sets exhibiting a high null allele frequency should be redesigned if representing a perfect repeat locus.

Three microsatellite loci were not in Hardy Weinberg equilibrium for both Abagold and HIK populations. Deviations at single loci may be as a result of natural selection acting on a nearby gene, while it is more common that a single deviation indicates a potential problem in the genotyping at the specific locus such as segregating null alleles, allele drop out and other genotyping errors. Deviations, however, at several loci may indicate population substructure such as closely related or inbred individuals (Dong *et al*. 2006). Determination of the relatedness amongst breeders should be further investigated so as to gain a better understanding of the degree of inbreeding that may be occurring as a result of the mass spawning practices currently applied.

The existence of varying degrees of relatedness between current breeding stock may provide suitable explanation as to why some individuals were assigned by CERVUS 2.0 to more than one parental pair, both having relatively similar LOD scores. In these instances, individual genotypes of all candidate parents were compared against the juvenile's genotype for mismatches and to ascertain reasons for mismatches. The parental pair with the least number of mismatches, or where mismatches could potentially be ascribed to a segregating null allele or scoring error, was then assigned.

Spawning success and hence contributions to the next generation was indeed found to be variable, and in some instances, negligible for some breeders. For both Abagold and HIK, twenty-one breeding pairs, contributed approximately thirty percent to the total individuals assigned. A further one hundred and twenty-six pairs contributed the balance through single spawnings. These single contributions do not add any value and it is recommended that all breeders making no

contribution or very small contributions in terms of offspring should be removed from the hatchery. For Abagold, thirty-three males out of forty contributed to the assigned offspring but only the following individuals made noticeable contributions: AbG92 (11.6%), AbG50 (8.8%), AbG06 (8.3%), AbG02 (6.6%), AbG93 (4.4%), AbG45 (4.4%), AbG09 (4.4%), AbG08 (4.4%). In total, these eight male breeders contributed 52.6% of the offspring assigned a specific father. For the females, forty-six out of fifty-nine animals contributed, however only the following eight individuals were deemed to have made a recognisable contribution and should be retained for future breeding: AbG66 (10.1%), AbG90 (10.1%), AbG75 (8.4%), AbG28 (6.7%), AbG60 (6.1%), AbG64 (5.6%), AbG78 (5.6%), AbG67 (4.5%). Together these females contributed 57.1% of the offspring assigned a specific mother.

For HIK, although forty-five out of fifty-three males were observed to have contributed to the assigned offspring, it is recommended that only ten of these be retained for future breeding: HIK68 (8.4%), HIK55 (5.6%), HIK10 (4.5%), HIK15 (4.5%), HIK17 (4.5%), HIK72 (3.9%), HIK67 (3.9%), HIK52 (3.9%), HIK45 (3.9%) and HIK20 (3.9%). In total, a contribution of 47% of the offspring assigned a specific father was made by these few individuals. Similarly for the females, it is recommended that only the following nine females be retained as breeders based on the size of their contributions, despite contributions having been made by thirty-eight of the forty-eight females: HIK76 (11.8%), HIK94 (11.8%), HIK40 (7.9%), HIK29 (6.7%), HIK35 (5.6%), HIK37 (5.1%), HIK82 (5.1%), HIK100 (4.5%), HIK36 (3.9%). A further recommendation is that those individuals from both farms that currently contributed less than 5%, be monitored during the regulated family-based spawning to gain a more accurate assessment of their breeding potential.

Regarding potential contribution to faster or slow growth, no trends were observed at this stage of mass selection, nor were they expected given that no selection for growth is as yet being made. Most contributing individuals produced both fast and slow growing juveniles with some contributing either to the fast or the slow growth groups. As mentioned previously, the small sample sizes tested in this study do not allow for firm conclusions to be drawn regarding the breeding population and the significance of their respective contributions, and recommendations are based primarily on a very preliminary view of breeding performance. The offspring were selected randomly on the basis of size, a prerequisite for unbiased sampling, but this is not to say that the other breeders did not contribute significantly to the remaining two thousand eight hundred individuals in the PRS that were not sampled for this study. However, the chances are quite slim that their performance should differ noticeably from that already observed.

An added assessment was made with respect to the genetic variation between the three populations (wild, hatchery and first generation) investigated in this study but since this does not fall under the scope of this study, the results of these are documented as Addendum A.

Chapter Six

Concluding Remarks and Suggestions for Future Research

Schilthuizen (2002: as cited in Pérez *et al*. 2005) predicted the phylum Mollusca to be one of the most evolutionary fertile groups; one which will become a central model in evolutionary radiation studies as a direct result of its role in morphological and molecular attempts to assess phylogeny of other major animal groups. Yet despite the fact that the molecular tools exist to understand the basic genomics of this group, molluscs remain behind the advancements seen in other eukaryotes.

Extensive research on commercially important abalone species is being undertaken around the world in an effort to rectify the lack of advancement within this phylum. South Africa is taking similar steps to remain competitive in a lucrative yet dynamic international market where genetic enhancement of the species is seen as the way forward to meeting the demand for this delicacy. The potential to realise significant genetic gains (10%-15% per generation) for abalone through the employment of well-structured genetic breeding programs, has been demonstrated for both mass selection and MAS (Elliot 2000; Hayes *et al* 2007). Selective breeding is therefore an attractive and attainable goal for the local industry given the right tools to it move forward towards implementation of family-based breeding as well as marker-assisted walk-back selection.

In the medium to long term, delivery on the broader project objectives will address the current lack of understanding and characterisation regarding the genome of the indigenous species *Haliotis midae*. In addition, a sustainable Aquatic Biotechnology Platform (ABP) for the genetic enhancement of the local commercially cultivated species *H. midae* will be established while building the expertise necessary to facilitate the effective use of biotechnologies such as stem cells and gene transfer. Delivery against the more immediate primary as well as specific objectives of this study provided the ground work that was necessary to move to Phase II of the PRS and the establishment of family-based selection.

All specific study objectives were successfully achieved:

(1) To isolate and develop species-specific polymorphic microsatellite markers for the indigenous abalone species, *Haliotis midae*.

Thirty-one polymorphic microsatellite markers of various repeat motifs were developed and optimum test conditions for individual reactions established. In addition, various combinations of loci were tested and several multiplexed reactions developed to reduce genotyping costs as well as number of PCR reactions required.

(2) To genotype all current breeding stock for Abagold and HIK as well as specific offspring selected from the PRS on the basis of growth performance; namely faster and slower growth groups.

Four hundred F1 juveniles were assessed for growth and subsequently genotyped along with one hundred and ninety-eight candidate parents using nine microsatellite markers.

(3) To assign parentage for offspring exhibiting either faster or slower growth rates for use in, or exclusion from, future MAS breeding programs.

Parentage was unambiguously assigned for 91% of F1 progeny from Abagold and 90% F1 progeny from HIK using likelihood methods of assignment. No growth trends were observed at this stage, involving mass-selection techniques only, however recommendations could be made regarding current hatchery stock at Abagold and HIK based on their contributions to F1 progeny.

Several conclusions were drawn from the results of this study:

Despite the FIASCO method being a quick, relatively effective technique for the enrichment and isolation of microsatellite repeat sequences, there are some inherent weaknesses which need to be investigated and addressed where feasible, in order to make the process of microsatellite marker development more robust. Given the high abundance of repeat sequences throughout the molluscan genome, the indication, from the low number of perfect repeats identified, is that the digestion-ligation step did not produce a representative sample of the genome for cloning into pDrive vector. One suggestion is to use a combination of restriction enzymes during the digestion-ligation step since enzyme/linker performance can be variable across species. The outcome of this suggestion should be investigated further.

A high level of redundant (not unique) clones (61%) were sequenced and the pre-screening of positive (i.e. insert-containing) clones using SSCP and fluorescent probe hybridisation should be considered since the risk of duplication occurring increases as the database of microsatellite markers for *H. midae* expands.

Repeat genotyping of each individual (at least twice) thus allowing for comparison of allele scores should be factored into the budget and resource/time requirements in order to reduce the number of allele-calling errors in a large data set.

A higher frequency of null alleles for each locus was observed for the hatchery stock and F1 progeny when compared to the wild population. It needs to be determined whether these are, in fact, true null alleles resulting from point mutations or whether they are a combination of PCR artefact and insufficiently optimised multiplexes together with increased genotyping errors.

Physical tagging methods are still necessary in the early stages of a PRS however, despite the numerous enhancement programs currently underway, very little is actually reported of the experience of other researchers, the methods employed and the success thereof. Networking between other researchers in this regard would be beneficial to addressing any limitations/tag losses

that may confound future assessments and selections. One of the requirements of the PRS was the conformation of the participating hatcheries to specific management and feeding practices in order to standardize environmental influences in the rearing of the juveniles. It is not clear whether this was adhered to. Standardisation of these practises should be implemented to remove as many variables as possible, thus allowing genetic effects to be more effectively evaluated.

The small sample sizes drawn did not allow for any statistically significant conclusions regarding the breeders and individual spawning success. There are three replicates in this project and consideration should be given to increasing the sampling to include all replicates and not just one, as was done for this study. One hundred animals were sampled per replicate for each growth group. Expanding this across three replicates with provide three hundred animals per growth group – a total of six hundred offspring to be genotyped.

Future Research

Further to this, the results from this study indicated some opportunities for future research or potential study areas encompassing:

(1) Further optimisation of multiplexes in terms of DNA concentrations and primer concentrations to ensure allele drop-out/null alleles is not as a result of these factors and can hence be alleviated or at least reduced.

(2) The determination of the molecular basis of the null alleles through sequencing if ascertained that PCR conditions are not the source.

(3) Establishing relatedness between hatchery individuals (either full-sib or half-sib) prior to family-based breeding commencing.

(4) The evaluation of changes in genetic diversity across several successive generations, once family-based breeding is established. This will allow an assessment of how effectively population size and inbreeding is being managed through the breeding structure and the use of pedigree records.

(5) Determination of cost-effectiveness of routine protocols such as DNA extractions. Chelex®100 was shown to be a possible alternative to the more time consuming and costly methods of CTAB extraction currently applied. This technique may also lend itself to high-throughput automation using the robotic platform in the Central DNA Sequencing Facility within the department of Genetics, University of Stellenbosch.

(6) Environmental effects should be investigated further and if necessary be factored in to the future breeding program.

(7) The institution of a complete ban on abalone harvesting in February 2008 may provide the opportunity to work with MCM to re-establish natural populations through re-seeding and ranching. Establishing a hatchery solely for the purpose of producing juveniles for re-stocking would allow for genetic diversity to be monitored and maintained through controlled breeding practices.

Appendices

Appendix 1 (a) Abagold Breeding Stock Data

Sample ID	Group No.	Tank No.	Tag ID	Gender
AbG01	2	13	AB 025	Male
AbG02	2	13	AB 027	Male
AbG03	2	13	AB 024	Male
AbG04	2	13	AB 026	Male
AbG05	2	16	AB 028	Male
AbG06	2	16	AB 029	Male
AbG07	2	16	AB 030	Male
AbG08	2	16	AB 031	Male
AbG09	2	22	AB 032	Male
AbG10	2	22	AB 034	Male
AbG11	2	22	AB 033	Male
AbG12	2	22	AB 035	Male
AbG13	2	33	AB 036	Male
AbG14	2	33	AB 037	Male
AbG15	2	33	AB 038	Male
AbG16	2	33	AB 039	Male
AbG17	2	40	AB 041	Male
AbG18	2	40	AB 040	Male
AbG19	2	40	AB 043	Male
AbG20	2	40	AB 042	Male
AbG21	2	14	AB 053	Female
AbG22	2	14	AB 056	Female
AbG23	2	14	AB 054	Female
AbG24	2	14	AB 055	Female
AbG25	2	15	AB 057	Female
AbG26	2	15	AB 058	Female
AbG27	2	15	AB 059	Female
AbG28	2	15	AB 060	Female
AbG29	2	23	AB 061	Female
AbG30	2	23	AB 062	Female
AbG31	2	23	AB 064	Female
AbG32	2	23	AB 063	Female
AbG33	2	25	AB 066	Female
AbG34	2	25	AB 065	Female
AbG35	2	25	AB 067	Female
AbG36	2	25	AB 068	Female
AbG37	2	27	AB 070	Female
AbG38	2	27	AB 069	Female
AbG39	2	27	AB 071	Female
AbG40	2	34	AB 072	Female
AbG41	2	34	AB 073	Female
AbG42	2	34	AB 074	Female
AbG43	2	34	AB 075	Female
AbG44	1	21	AB 005	Male
AbG45	1	21	AB 006	Male
AbG46	1	21	AB 007	Male
AbG47	1	21	AB 008	Male
AbG48	1	30	AB 008	Male
AbG49	1	30	AB 010	Male
AbG50	1	30	AB 011	Male

Sample ID	Group No.	Tank No.	Tag ID	Gender
AbG51	1	30	**Died**	Male
AbG52	1	36	AB 015	Male
AbG53	1	36	AB 014	Male
AbG54	1	36	AB 013	Male
AbG55	1	36	AB 012	Male
AbG56	1	39	AB 017	Male
AbG57	1	39	AB 018	Male
AbG58	1	39	AB 016	Male
AbG59	1	39	AB 019	Male
AbG60	1	17	AB 11	Female
AbG61	1	17	AB 12	Female
AbG62	1	17	AB 009	Female
AbG63	1	17	AB 010	Female
AbG64	1	20	AB 13	Female
AbG65	1	20	AB 14	Female
AbG66	1	20	AB 15	Female
AbG67	1	20	AB 16	Female
AbG68	1	24	AB 17	Female
AbG69	1	24	AB 20	Female
AbG70	1	24	AB 19	Female
AbG71	1	24	AB 18	Female
AbG72	1	29	AB 21	Female
AbG73	1	29	AB 22	Female
AbG74	1	29	AB 23	Female
AbG75	1	29	AB 24	Female
AbG76	1	31	AB 25	Female
AbG77	1	31	AB 26	Female
AbG78	1	31	AB 28	Female
AbG79	1	31	**Died**	Female
AbG80	1	32	AB 31	Female
AbG81	1	32	AB 30	Female
AbG82	1	32	AB 29	Female
AbG83	1	32	AB 32	Female
AbG84	1	35	AB 33	Female
AbG85	1	35	AB 34	Female
AbG86	1	35	AB 36	Female
AbG87	1	35	AB 35	Female
AbG88	3	18	AB 84	Female
AbG89	3	18	AB 87	Female
AbG90	3	18	AB 86	Female
AbG91	3	18	AB 85	Female
AbG92	3	19	AB 048	Male
AbG93	3	19	AB 049	Male
AbG94	3	19	AB 050	Male
AbG95	3	19	AB 051	Male
AbG96	3	28	AB 90	Female
AbG97	3	28	AB 89	Female
AbG98	3	28	AB 88	Female
AbG99	3	28	AB 91	Female

Appendix 1 (b) HIK Abalone Breeding Stock Data

Sample ID	Tank Position	Tank No.	Tag ID	Gender
HIK01		B34	93	Male
HIK02		B34	162	Male
HIK03		B34	O8	Male
HIK04		B34	O6	Male
HIK05		B35	O5	Male
HIK06		B35	O4	Male
HIK07		B35	T4	Male
HIK08		B35	T1	Male
HIK09		B35	T7	Male
HIK10		B35	T6	Male
HIK11		B35	T5	Male
HIK12	Back	B28	D9	Male
HIK13		**Died**	D2	Male
HIK14	Front	B32	D1	Male
HIK15	Front	B28	C7	Male
HIK16	Front	B24	Q6	Male
HIK17	Back	B23	H1	Male
HIK18	Back	B24	E3	Male
HIK19	Back	B32	H2	Male
HIK20	Front	B23	C9	Male
HIK21	Back	B25	C8	Male
HIK22		B13	N1	Male
HIK23		B13	P3	Female
HIK24		B13	P2	Female
HIK25		B13	P1	Female
HIK26		B13	O9	Female
HIK27		B14	L6	Female
HIK28		B14	P5	Female
HIK29		B14	P9	Female
HIK30		B14	S1	Female
HIK31		B17	180	Female
HIK32		B17	178	Female
HIK33		**Died**	untagged	Female
HIK34		B17	N6	Female
HIK35		B36	I9	Female
HIK36		B36	I2	Female
HIK37		B42	H4	Female
HIK38		B42	D4	Female
HIK39		B42	D5	Female
HIK40		B38	C6	Female
HIK41		C42	C5	Female
HIK42		C36	G8	Female
HIK43	Front	B11	I8	Female
HIK44	Back	B10	J2	Female
HIK45	Front	C25	G9	Male
HIK46	Back	C27	C3	Male
HIK47	Back	C25	C1	Male
HIK48		C34	U4	Male
HIK49		C34	M3	Male
HIK50		C34	U6	Male

Sample ID	Tank Position	Tank No.	Tag ID	Gender
HIK51		C34	R5	Male
HIK52		C34	S2	Male
HIK53		C35	Q7	Male
HIK54		C35	O2	Male
HIK55		C35	T3	Male
HIK56	Front	C23	J5	Male
HIK57	Back	C23	J3	Male
HIK58		C40	F3	Male
HIK59	Back	C24	B8	Male
HIK60	Front	C24	S6	Male
HIK61	Front	C24	J1	Male
HIK62	Front	C24	G3	Male
HIK63	Front	C24	B7	Male
HIK64		C41	B6	Male
HIK65		**Died**	N1	Male
HIK66		**Died**	B5	Male
HIK67		**Died**	B4	Male
HIK68		C40	B1	Male
HIK69		C40	F6	Male
HIK70		C35	179	Male
HIK71	Front	C26	A9	Male
HIK72	Front	C29	F5	Male
HIK73	Back	C30	F4	Male
HIK74	Back	C29	A1	Male
HIK75	Front	C30	A3	Male
HIK76		C12	O3	Female
HIK77		C12	N7	Female
HIK78		C12	N2	Female
HIK79	Front	C11	L7	Female
HIK80		C13	M6	Female
HIK81		C13	M9	Female
HIK82		C13	M8	Female
HIK83		C13	M4	Female
HIK84		C13	N3	Female
HIK85		C14	N8	Female
HIK86		C14	O1	Female
HIK87		C14	N9	Female
HIK88		C14	Q1	Female
HIK89		C19	182	Female
HIK90		C19	181	Female
HIK91		C19	183	Female
HIK92		**Died**	untagged	Female
HIK93		**Died**	untagged	Female
HIK94		C36	J8	Female
HIK95		C36	J7	Female
HIK96		C36	K2	Female
HIK97		C41	B9	Female
HIK98		C41	G2	Female
HIK99		C41	E5	Female
HIK100		C42	A8	Female
HIK101		C42	B2	Female

Appendix 2 Size Ranges for Assessment of PRS Juveniles for Fast or Slow Growth

Size ranges, against which the upper 10% (faster growers) and lower 10% (slower growers) of animals originating from Abagold and HIK hatcheries for rearing at different locations, as per the PRS were measured. Length is recorded in millimetres

Location Abagold

Origin	Length Range	
	Upper 10%	Lower 10%
Abagold	≥ 27	≤ 21
HIK	≥ 27	≤ 19

Location Aquafarm

Origin	Length Range	
	Upper 10%	Lower 10%
Abagold	≥ 31	≤ 20
HIK	≥ 34	≤ 20

Location HIK

Origin	Length Range	
	Upper 10%	Lower 10%
Abagold	≥ 29	≤ 20
HIK	≥ 31	≤ 22

Location I & J

Origin	Length Range	
	Upper 10%	Lower 10%
Abagold	≥ 26	≤ 13
HIK	≥ 29	≤ 18

Location Roman Bay

Origin	Length Range	
	Upper 10%	Lower 10%
Abagold	≥ 30	≤ 20
HIK	≥ 28	≤ 20

Appendix 3 Chelex®100 DNA Extraction Method

Tissue Used

1 x tentacle from adult abalone

Materials

Chelex®100 Resin (*Sigma*)

2 X 1.5ml eppendorf tubes per extraction

Double-autoclaved distilled water

Method

- Prepare a 10% (w/v) Chelex®100 resin solution in a sterile container using double distilled water and use fresh or store at 4°C until required
- Place the tissue sample in a 1.5ml eppendorf tube
- Add 200µl of 10% Chelex®100 solution. Pipette the solution in to the sample tube while at room temperature and under gentle agitation to ensure that the resin beads are distributed evenly in solution
- Add 2µl Proteinase K (10mg/ml) to sample tube
- Vortex samples for 10s using a Vortex Genie
- Place on an orbital incubator for 30 minutes at 56˚C
- Vortex samples for 10s using a Vortex Genie
- Pierce lids of tubes and incubate for 10 minutes at 95˚C in a hot-block
- Vortex samples for 10s using a Vortex Genie
- Centrifuge for 3 minutes at 16000rpm
- Transfer supernatant carefully to a clean 1.5ml eppendorf leaving Chelex®100 resin behind
- Measure DNA concentration on a NanoDrop® ND-1000 spectrophotometer
- Store DNA at -20˚C until required

Appendix 4 Touch-Down Polymerase Chain Reaction Program

94˚C	5 min	
94˚C	30 sec	
65˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
64˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
63˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
62˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
61˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
60˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
59˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
58˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
57˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
56˚C	30 sec	X2
72˚C	30 sec	
94˚C	30 sec	
55˚C	30 sec	X3
72˚C	30 sec	
72˚C	7 min	

Appendix 5 Stock Solutions Required for Microsatellite Enrichment and Isolation

Preparation of solutions required for the FIASCO protocol (Zane *et al*. 2002) with recommendations for storage where applicable (Sambrook *et al*. 1998). < ! > indicates where caution in handling is required. All solutions are made up to 1L but should be adjusted to personal requirements.

1M Tris-HCl (pH 8)

Dissolve 121g of Tris base in 800ml of distilled water while heating over a hotplate. Allow the solution to cool to room temperature and adjust the pH to 8.0 using concentrated HCl < ! >. Bring volume to 1L and sterilize by autoclaving.

20X SSC (pH 7)

Dissolve 175.3g of sodium chloride (NaCl) and 88.2g of sodium citrate in 800ml of distilled water. Adjust the pH to 7.0 using a few drops of concentrated HCl < ! > Bring the volume to 1L and sterilize by autoclaving. The final concentrations of the ingredients are 3M NaCl and 0.3M sodium citrate.

0.5M EDTA (pH 8)

Add 186.1g of disodium EDTA·2H2O to 800ml of distilled water. Stir vigorously with a magnetic stirrer. Adjust the pH to 8.0 with sodium hydroxide (NaOH < ! >) and bring volume to 1L. Sterilize by autoclaving. Note that the disodium salt of EDTA will not go into solution until the pH is adjusted.

1M NaOH

Slowly add 400g of sodium hydroxide pellets (NaOH) to 800ml of distilled water, stirring continuously. This reaction is extremely exothermic and should be performed with caution in a plastic beaker and preferably on ice. When pellets are completely dissolved, bring the volume to 1L. The solution should be stored at room temperature within a plastic container.

5M NaCl

Dissolve 292g of sodium chloride (NaCl) in 800ml of distilled water and then adjust to a final volume of 1L. Aliquot into smaller amounts and sterilize by autoclaving. Store NaCl solution at room temperature.

10% (w/v) SDS

Dissolve 100g sodium lauryl sulphate in 900ml of distilled water < ! >. Heat the solution to approximately 68°C and stir with a magnetic stirrer to assist dissolution. Adjust the pH to 7.2 by adding a few drops of concentrated HCl < ! >. Bring the volume to 1L and store at room temperature. Do not autoclave as sterilization is not necessary

1M (v/v) Acetic Acid

Add 159.5ml of acetic acid (CH_3COOH) to 1L of distilled water < ! >. Store the solution at room temperature.

10X Tris-EDTA (TE)

Add 100ml of Tris-HCl solution (100mM pH 8.0) and 20ml EDTA solution (10mM pH 8.0) to 800ml of distilled water. Confirm pH is at 8.0 and adjust if necessary. Bring to 1L volume. As the constituent solutions are sterilised prior to use, the final TE buffer solution does not need to be autoclaved again. Store the buffer at room temperature.

5X Tris-borate-EDTA (TBE)

Dissolve 54g Tris base and 27.5g boric acid into 800ml of distilled water. Heat the solution and stir with a magnetic stirrer to assist dissolution. Add 20ml EDTA (0.5M pH 8.0) and bring to 1L volume. TBE buffer is stored as a 5X solution as this is reportedly more stable than a 10X solution and solutes do not precipitate as readily during storage. Filtering the solution may further delay precipitation of solutes. Dilutions of the stock solution (1X) are made just prior to use their in the preparation of gel solutions and as electrophoresis buffer. Stock solutions are stored at room temperature.

Appendix 6 pDrive Cloning Vector indicating Primer Binding and Restriction Enzyme Sites

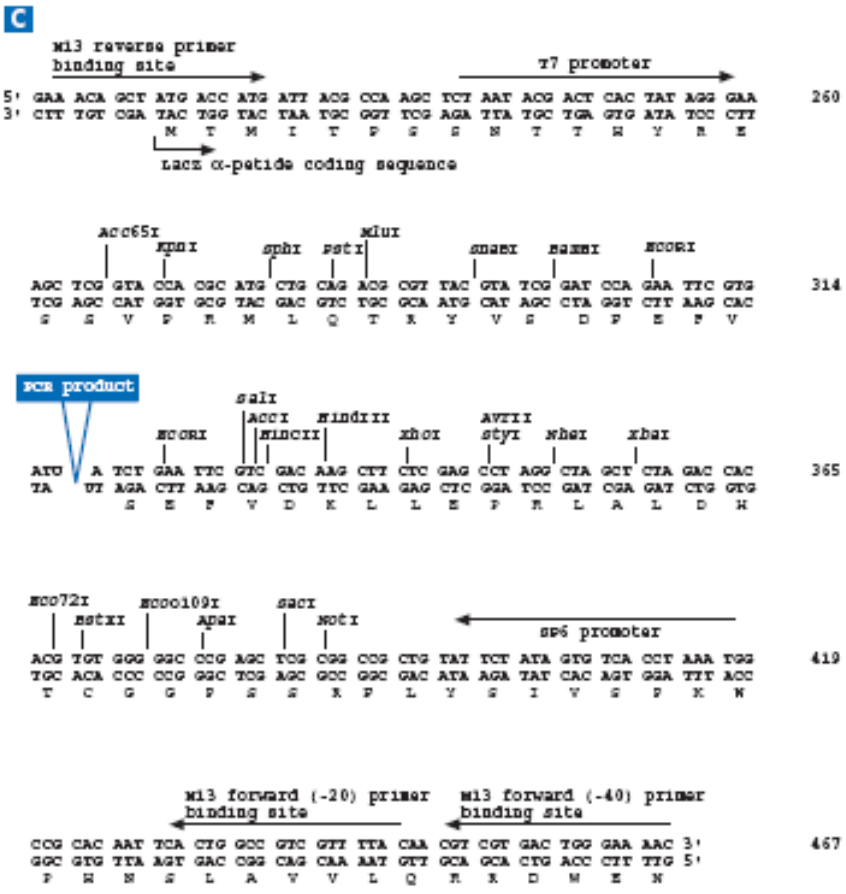
pDrive Cloning Vector information as supplied in the QIAGEN® PCR Cloning Handbook (2001) available as a PDF online at: http://www1.qiagen.com/literature/

Product Description

The pDrive Cloning Vector is supplied in a linear form, ready-to-use for direct ligation of PCR products. This vector allows ampicillin and kanamycin selection, as well as blue/white colony screening. The vector contains several unique restriction endonuclease recognition sites around the cloning site, allowing easy restriction analysis of recombinant plasmids. The vector also contains a T7 and SP6 promoter on either side of the cloning site, allowing in vitro transcription of cloned PCR products as well as sequence analysis using standard sequencing primers. In addition, the pDrive Cloning Vector has a phage f1 origin to allow preparation of single-stranded DNA. A map of the pDrive Cloning Vector and the sequence of the region surrounding the cloning site are provided by Figures C and A.

Figure C

DNA sequence of the region surrounding the cloning site - The amino acid sequence of the LacZ α-peptide is also given. The positions of the T7 and SP6 promoter sites and the M13 forward and reverse sequencing primer binding sites are shown, as are the cutting positions of the unique restriction endonucleases.



Primer sequences

M13 forward (–20): 5' GTAAAACGACGGCCAGT 3'

M13 forward (–40): 5' GTTTTCCCAGTCACGAC 3'

M13 reverse: 5' AACAGCTATGACCATG 3'

SP6 promoter: 5' CATTTAGGTGACACTATAG 3'

T7 promoter: 5' GTAATACGACTCACTATAG 3'

Figure A pDrive Cloning Vector Map A

Representation of the linearized pDrive Cloning Vector with U overhangs.

The unique restriction endonuclease recognition sites on either side of the cloning site are listed.

A

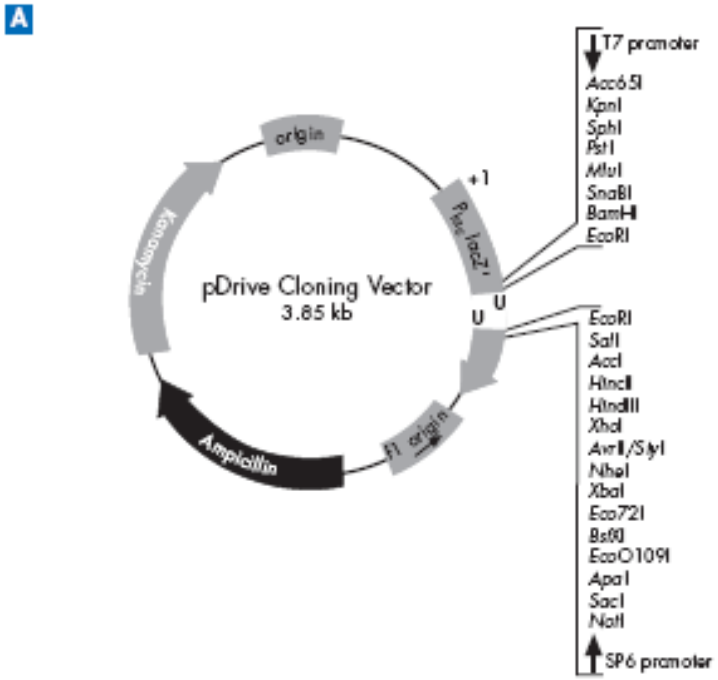


Table B Positions of various elements of the pDrive Cloning Vector

B

Element	Position (bp)
Multiple cloning site	266–393
LacZ α-peptide	216–593
T7 RNA polymerase promoter	239–258
T7 transcription start	256
SP6 RNA polymerase promoter	398–417
SP6 transcription start	400
Ampicillin resistance gene	1175–2032
Kanamycin resistance gene	2181–2993
pUC origin	3668
Phage f1 origin	588–1043
Primer binding sites:*	
M13 forward (–20)	431–447
M13 forward (–40)	451–467
M13 reverse	209–224
T7 promoter primer	239–258
SP6 promoter primer	400–418

Appendix 7 PCR Primer Design Guidelines

General guidelines for design of standard PCR primers (page 22) of the QIAGEN® PCR Cloning Handbook (2001)

Length: 18-30 nucleotides

G/C content: 40-60%

Tm: Simplified formula for estimating melting temperature (*Tm*):

Tm = 2°C x (A+T) + 4°C x (G+C)

Whenever possible, design primer pairs with similar *Ta* values. Optimal annealing temperatures may be above or below the estimated *Tm*. As a starting point, use an annealing temperature 5°C below *Tm*.

Sequence:

• Avoid complementarities of two or three bases at the 3' ends of primer pairs to reduce primer-dimer formation.

• Avoid mismatches between the 3' end of the primer and the target-template sequence.

• Avoid runs of 3 or more G or C at the 3' end.

• Avoid a 3'-end T. Primers with a T at the 3' end have a greater tolerance of mismatch.

• Avoid complementary sequences within a primer sequence and between the primer pair.

• Commercially available computer software (e.g., Primer Designer 1.0, Scientific Software, 1990; Oligo, Rychlik and Rhoads, 1989) can be used for primer design.

Concentration:

• Spectrophotometric conversion for primers:

1 A260 unit = 20–30µg/ml

• Molar conversions:

Primer Length	pmol/µg	20pmol
18mer	168	119ng
20mer	152	132ng
25mer	121	165ng
30mer	101	198ng

• Use 0.1-0.5 M of each primer in PCR. For most applications, a primer concentration of 0.2µM will be sufficient.

Appendix 8 Primer Information and Nomenclature

Primer information including status (i.e. polymorphic or monomorphic), testing conditions and NCBI accession numbers for all primer sets developed in this study. The following nomenclature was used for the primer names. Although a hepta-repeat is considered to be a minisatellite, one of these repeat types was developed.

Nomenclature

Hm	*Haliotis midae*	
NS	Researcher ID	
#	Number of the clone screened	
S	mono-nucleotide	microsatellite
D	dinucleotide	microsatellite
R	trinucleotide	microsatellite
T	tetranucleotide	microsatellite
P	penta-nucleotide	microsatellite
H	hexa-nucleotide	microsatellite
L	heptanucleotide	minisatellite
M	imperfect/compound	

Name	Repeat	Size	Primer	G/C Content	Annealing (*Ta*)	Result	NCBI Accession No.	PCR Conditions	Label	Colour
Hm NS7T F Hm NS7T R	(GAGT)6	517bp	ATT CCC CCA ACT CTA AGT CA CAA ATG AGC CAA GAT GTG TT	40%	60°C	Polymorphic	EF033329	TD-PCR	NED	Yellow
Hm NS19L F Hm NS19L R	(AACACCC)9	215bp	ACA ACA ACA AAG GTG GTC AA CAA TGA ATA GCT ATG GGT CG	40%	55°C	Polymorphic	EF033330	Standard PCR	NED	Yellow
Hm NS34T F Hm NS34T R	(GAGT)4	114bp	AGA TGT GAG CGA ATG AGT GA TGG AGT TCA CTA AAG CGG TA	45%	60.2°C	Monomorphic		Standard PCR		
HmNS28D F HmNS28D R	(CA)16	180bp	CAG TCA ATT TTC ATC GCA TT AGG TCG TTT TTC TCC TTC AG	35%	55°C	Polymorphic	EF033332	Standard PCR	FAM	Blue
HmNS31M F HmNS31M R	(GT)4(CT)(GT)8	266bp	CTC GGG TTC AGT TAC CTA CA CTT GCT GAC TTC GAT CAC AC	50%	55°C	Polymorphic	EF033333	Standard PCR	FAM	Blue
HmNS38T F HmNS38T F	(TCAC)10	454bp	CTG AGA CCC AAA GTT TTC TTT A ATC TAT GTT CAG GGT GTC AGT G	45%	55°C	Polymorphic	EF 367113	Standard PCR	PET	Red
HmNS17bT F HmNS17bT R	(CACT)31	618bp	TGT AGT AAA GTG GCC TGC AA CAG TCA GTG GGA GAG GAA GA	55%	65°C	Polymorphic	EF 367116	Standard PCR	FAM	Blue
HmNS14R F HmNS14R R	(TTG)5	260bp	GCT CTG GTG TAT GTT GTG TCA TTG ATC AAG TTG CAC ATG AAT	48%	60°C	Polymorphic	EF 367115	Standard PCR	VIC	Green
HmNS26R F HmNS26R R	(GTA)6	199bp	AGA CAA TCC AGT GAT CAA TAT CA AAC AAA CAA CAA CAA CTA CTG CT	35%		Monomorphic		TD-PCR		
HmNS6T F HmNS6T R	(ACGC)6	196bp	TGA GAG ACA TTT GAA GCA TTT A AAC ACT CAC GTA CGC ATA CAC	32%	55°C	Polymorphic	EF 367117	Standard PCR	VIC	Green
HmNSa34D F HmNSa34D R	(AC)7	189bp	CAT TCC ACG CTG AAG AAA TC TGA GAT GAG CGT GAA AAT GT	40%	50°C	Polymorphic	EF 367118	Standard PCR	NED	Yellow
HmNS58R F HmNS58R R	(GTT)8	246bp	TGC CAC TCA AAT GTT CCT TA CTA TTT CAG GTG TCC CCA GT	50%	60°C	Polymorphic	EF 367119	Standard PCR	NED	Yellow
HmNS69R F HmNS69R R	(CAA)7	246bp	GAT ATT CCG TGA TAG TCC CC ACC TTA TAC CAA TCT TGC TGA A	36%		Monomorphic		TD-PCR		
HmNS18M F HmNS18M R	(ACC)4(AGG)7 - (AAC)3(AG)(AAC)2 - (AG)(AAC)	209bp	AAT TGT CTC CTT TGT TCT TCT TT TAT TTG TGA CTT TAG GTG AGG AC	30%	60°C	Polymorphic	EF 367120	Standard PCR	PET	Red

Name	Repeat	Size	Primer	G/C Content	Annealing (*Ta*)	Result	NCBI Accession No.	PCR Conditions	Label	Colour
HmNST7T F HmNST7T R	(CACT)26	313bp	CAC ATG GGT ACA ATG TGT GAA G GGT AGC ACT GTT TCT CAC GA	50.0%	60°C	Polymorphic	EF 455618	Standard PCR	VIC	Green
HmNST3M F HmNST3M R	(GCAC)6(AG)45	237bp	AAC CGA GTC ATC GAG TCT GA AAC GCT TCT CCA GTG TGT GT	50.0%	55°C	Polymorphic	EF 455622	Standard PCR		
HmNS21M F HmNS21M R	(CT)5(C)2(CT)5(T)3(CT)21	237bp	ACG CTC ACA CAC AAA CTC AC AAT CGT ACA CAG GTT GAA GAA T	50.0%	60°C	Polymorphic	EF 455621	Standard PCR		
HmNS9M F HmNS9M R	(TCG)(TCC)2(C)2(TCC)4(GTC)(TCC)3	223bp	ACA ACA ACA GCA ATA ACA CCA CGT GAT CAA CAC TGC AAA GT	45.0%	55°C	Polymorphic	EF 455623	Standard PCR		
HmNS56D F HmNS56D R	(CA)20	259bp	TTC GGC AAG TGA ATG TCT AG CCG AGT TTG GAA TGT CTG AT	45.0%	55°C	Polymorphic	EF 455619	Standard PCR	NED	Yellow
HmNS59D F HmNS59D R	(GT)36	216bp	AAG CAT GAT AAA TAT GCG GA TAT GTA TGC AGC AGA GAG GG	50.0%	63°C	Polymorphic	EF 455620	Standard PCR		
HmNS32M F HmNS32M R	(CAA)5(GTC)(GTT)5(GTC)(GTT)2	156 bp	AAT CGT CGT TAC GAA GTT ATG TCC CAG GCT ATT TAC AGG CAA TCG	47%	50°C	Polymorphic	EF492144	Standard PCR		
HmNSb15M F HmNSb15M R	(GT)17(GCGT)3	186 bp	AGT GGA ATC GGT TTC TGA GC GGC CAT GAA GTA TGA TGT GG	50%	50°C	Monomorphic	Not submitted	Standard PCR		
HmNS17bT F HmNS17bT R	(CACT)31	245 bp	AGT AAA GTG GCC TGC AAT CTG GTG GCC AGT GGA ATT CTT TC	50%	58°C	Polymorphic	EF 367116	Standard PCR	PET	Red
HmNS100T F HmNS100T R	(GAGT)16	406 bp	CAG TTT TTG TTA GGG ATT TCA T GAA AAA GAC TGT TGA TGG GG	32%	60°C	Polymorphic	EF 367114	Standard PCR	PET	Red
HmNSp6M F HmNSp6M R	(CA)17(TA)(CA)13	203bp	GCA GTT GAC AAA CAC ACT GAC A GTG TCC AAA AAC TGG CAC AA	45%	60°C	Polymorphic	EU126855	Standard PCR		
HmNSp31M F HmNSp31M R	(CAA)4(CAG)3(CAA)3	286bp	ACT GGA GGA GTC GCT ACG AT CAA GTC AGG GTG GTC TTT CC	55%	60°C	Polymorphic	EU126856	Standard PCR		

Name	Repeat	Size	Primer	G/C Content	Annealing (*Ta*)	Result	NCBI Accession No.	PCR Conditions	Label	Colour
HmNSp34M F HmNSp34M R	(GAT)5(GTT)7	299bp	GTA CTT TCT CTG TTT GTG GGC ACA CAA AGT CAA CCC AAC AA	40%	60°C	Polymorphic	EU126857	Standard PCR		
HmNSp42M F HmNSp42M R	(AGT)3 - (AGT)13 - (AGT)6 - (TGT)6	315bp	CCT GAA TTT ATA GTA GTA GAC CTT G AAT CCA GTA AAG GTC AAT GTA AA	30%	60°C	Polymorphic	EU126858	Standard PCR		
HmNSp44M F HmNSp44M R	(CAA)8(CTA)(CAACCACACAACA)	181bp	GTC GAC AAG CAC GAC GAT AG CGG TTG TGG CTG ATA CTG AG	55%	60°C	Polymorphic	EU126859	Standard PCR		
HmNSq2T F HmNSq2T R	(GTGC)6	229bp	ACT TCC TCG TGG TGG ATG AC AGT TTT ACG CGG ACG TCA AT	55%	53°C	Monomorphic	Not submitted	Standard PCR		
HmNSq9M F HmNSq9M R	(TG)7(AG)11	180bp	GCA GAT CAC AGC AAC CCT TT TGC ACA CAC AAG CTC TCT CC	55%	60°C	Polymorphic	EU126860	Standard PCR		
HmNSq17T F HmNSq17T R	(TCAC)34	280bp	ACA AGC CAG AGA CGC ACA T TGG GTG TTG GTG TCA AGT TC	55%	54°C	Polymorphic	EU126861	Standard PCR		
HmNSq30M F HmNSq30M R	(CACT)2(CACG)8 - (CACT)4	153bp	GCG GCT AAA AAC CAA ACG ATT CCT CTC AAC GGA AAC TCC	50%	56°C	Polymorphic	EU126862	Standard PCR		
HmNSq43M F HmNSq43M R	(TGTA)4(TG)6(TATG)(TGCG)6(TG)12	259bp	GAA ATA TAT GAG TGG ACA GTT GC ATG ACG CCT TTT GAA CAG TAT	30%	54°C	Polymorphic	EU126863	Standard PCR		
HmNSq44T F HmNSq44T R	(GAGT)32	218bp	AGC TGT ATG CAA GCA AAC TT GAG TAA AGT GTT TGT CAC CC	45%	56°C	Polymorphic	Not submitted	Standard PCR		
HmNSq47T F HmNSq47T R	(GTGC)7	250bp	TGA GAG AGT AGG TGG TGG GTT T AAG AGC AGG TCA GCT TGG AG	55%	54°C	Monomorphic	Not submitted	Standard PCR		
HmNSr1R F HmNSr1R R	(ATC)9	211bp	GTT CCC ATG TGG AGG ACA AC TTG CCG TAT TGT TGG TCA GA	55%	56°C	Polymorphic	EU126864	Standard PCR		
HmNSr16D F HmNSr16D R	(CA)7	240bp	CAA GTA CTG TTT CAG GCA TGG T CAA AGT TCG CCA CCA GTG T	55%	56°C	Polymorphic	EU126865	Standard PCR		
HmNSr41D F HmNSr41D R	(CA)12	234bp	TCT TCT AAC GCA GGT CTT GGA AGC TCG ACA CAC AAC ACA GC	55%	56°C	Polymorphic	EU126866	Standard PCR		
HmNSS1H F HmNSS1H R	(GGGTTA)7	204bp	TCC AGC AGC TGA AAA ATG TG AAA CCA TCA ACC TCC CCC TA	50%	56°C	Polymorphic	EF033331	Standard PCR		

Appendix 9 Fluorescently Labelled Primers

Sixteen primers were fluorescently labelled and details are provided with regards to the label selected, the primer labelled and the microsatellite repeat to be amplified in genotyping analyses

Name	F/R	Primer Sequence (5'-3')	Primer Length (bp)	Repeat Type	Label	Colour
HmNS7T	Reverse	CAA ATG AGC CAA GAT GTG TT	20	(GAGT)6	NED	yellow
HmNS19L	Reverse	CAA TGA ATA GCT ATG GGT CG	20	(AACACCC)9	NED	yellow
HmNSS1H	Reverse	AAC AAC GAC CCT AAA CCA TC	20	(GGGTTA)7	PET	red
HmNS28D	Reverse	AGG TCG TTT TTC TCC TTC AG	20	(CA)16	FAM	blue
HmNS31D	Reverse	CTT GCT GAC TTC GAT CAC AC	20	GT(4)(CT)(GT)8	FAM	blue
HmNS38T	Reverse	ATC TAT GTT CAG GGT GTC AGT G	22	(TCAC)10	PET	red
HmNS100T	Reverse	GAA AAA GAC TGT TGA TGG GG	20	(GAGT)16	PET	red
HmNS14R	Reverse	TTG ATC AAG TTG CAC ATG AAT	21	(TTG)5	VIC	green
HmNS6T	Forward	TGA GAG ACA TTT GAA GCA TTT A	22	(ACGC)6	VIC	green
HmNSa34D	Forward	CAT TTC ACG CTG AAG AAA TC	20	(AC)7	NED	yellow
HmNS58R	Reverse	CTA TTT CAG GTG TCC CCA GT	20	(GTT)8	NED	yellow
HmNS18M	Reverse	TAT TTG TGA CTT TAG GTG AGG AC	23	(ACC)4(AGG)7 - (AAC)3(AG)(AAC)2 (AG)(AAC)	PET	red
HmNS56D	Forward	TTC GGC AAG TGA ATG TCT AG	20	(CA)20	NED	yellow
HmNS59D	Forward	AAG CAT GAT AAA TAT GCG GA	20	(GT)33	FAM	blue
HmNST7T	Forward	CAC ATG GGT ACA ATG TGT GAA G	22	(CACT)26	VIC	green
HmNS17bT	Forward	AGT AAA GTG GCC TGC AAT CTG	21	(CACT)31	PET	red

Appendix 10 Allele Frequency Analysis of Black Rock Wild Abalone Population

Allele frequency analysis for Black Rock individuals 1-32 generated by GENETIX 4.03 - selection of primer sets for parentage analysis was based on allele frequencies as well as other selection criteria (exclusion probabilities, null allele frequency)

HmNS19

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
175	1	1	0	0.0156	0.015
182	8	0	4	0.125	0.0614
189	2	2	0	0.0313	0.0302
196	7	1	3	0.1094	0.0614
203	4	0	2	0.0625	0.0302
210	4	2	1	0.0625	0.0457
217	6	4	1	0.0938	0.0774
224	4	2	1	0.0625	0.0457
231	12	10	1	0.1875	0.1798
238	9	7	1	0.1406	0.1271
245	5	5	0	0.0781	0.0774
252	2	2	0	0.0313	0.0302

HmNS38

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
402	2	2	0	0.0323	0.0327
422	22	14	4	0.3548	0.3515
434	2	2	0	0.0323	0.0327
442	5	3	1	0.0806	0.0666
446	5	3	1	0.0806	0.0666
450	9	7	1	0.1452	0.1383
454	6	4	1	0.0968	0.084
458	3	3	0	0.0484	0.0495
462	4	4	0	0.0645	0.0666
466	3	3	0	0.0484	0.0495
474	1	1	0	0.0161	0.0162

HmNS58

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
233	2	2	0	0.0313	0.0317
236	2	2	0	0.0313	0.0317
239	2	0	1	0.0313	0.0157
242	10	6	2	0.1563	0.1338
245	11	7	2	0.1719	0.152
248	16	14	1	0.25	0.2708
251	9	9	0	0.1406	0.152
254	8	6	1	0.125	0.116
257	3	3	0	0.0469	0.048
272	1	1	0	0.0156	0.0157

HmNS31

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
238	1	1	0	0.0156	0.0132
240	1	1	0	0.0156	0.0132
262	2	0	1	0.0313	0.0132
264	2	0	1	0.0313	0.0132
266	7	1	3	0.1094	0.054
268	3	1	1	0.0469	0.0266
270	1	1	0	0.0156	0.0132
274	1	1	0	0.0156	0.0132
276	1	1	0	0.0156	0.0132
278	1	1	0	0.0156	0.0132
282	25	5	10	0.3906	0.2214
284	1	1	0	0.0156	0.0132
286	6	4	1	0.0938	0.0679
288	12	4	4	0.1875	0.1112

HmNS6

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
186	2	2	0	0.0313	0.0312
190	7	3	2	0.1094	0.0801
194	20	8	6	0.3125	0.2453
198	21	13	4	0.3281	0.3091
202	6	6	0	0.0938	0.0969
206	2	0	1	0.0313	0.0155
210	4	4	0	0.0625	0.0635
218	1	1	0	0.0156	0.0155
230	1	1	0	0.0156	0.0155

HmNS100

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
232	2	0	1	0.0313	0.0147
374	9	3	3	0.1406	0.0918
378	9	5	2	0.1406	0.108
382	7	5	1	0.1094	0.0918
386	6	4	1	0.0938	0.0758
390	8	6	1	0.125	0.108
394	6	2	2	0.0938	0.0602
406	3	1	1	0.0469	0.0296
414	2	2	0	0.0313	0.0296
422	3	1	1	0.0469	0.0296
426	2	2	0	0.0313	0.0296
430	1	1	0	0.0156	0.0147
434	4	2	1	0.0625	0.0448
454	2	0	1	0.0313	0.0147

HmNST7

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
228	3	3	0	0.0484	0.0496
236	3	3	0	0.0484	0.0496
244	2	2	0	0.0323	0.0328
248	10	8	1	0.1613	0.1576
252	4	4	0	0.0645	0.0667
256	5	5	0	0.0806	0.0842
260	4	4	0	0.0645	0.0667
264	2	2	0	0.0323	0.0328
268	4	4	0	0.0645	0.0667
272	2	0	1	0.0323	0.0163
276	4	4	0	0.0645	0.0667
278	1	1	0	0.0161	0.0163
280	2	2	0	0.0323	0.0328
284	2	2	0	0.0323	0.0328
288	5	5	0	0.0806	0.0842
292	3	3	0	0.0484	0.0496
296	2	2	0	0.0323	0.0328
304	1	1	0	0.0161	0.0163
306	1	1	0	0.0161	0.0163
316	1	1	0	0.0161	0.0163
328	1	1	0	0.0161	0.0163

HmNS56

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
211	1	1	0	0.0161	0.0163
215	1	1	0	0.0161	0.0163
219	5	5	0	0.0806	0.0842
221	1	1	0	0.0161	0.0163
223	19	13	3	0.3065	0.3044
225	13	11	1	0.2097	0.2171
227	4	2	1	0.0645	0.0496
231	2	2	0	0.0323	0.0328
233	2	2	0	0.0323	0.0328
235	1	1	0	0.0161	0.0163
237	3	3	0	0.0484	0.0496
239	3	3	0	0.0484	0.0496
241	1	1	0	0.0161	0.0163
245	2	2	0	0.0323	0.0328
247	2	2	0	0.0323	0.0328
253	2	2	0	0.0323	0.0328

HmNS14

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
252	2	0	1	0.0323	0.0084
255	4	0	2	0.0645	0.0169
258	55	1	27	0.8871	0.2721
261	1	1	0	0.0161	0.0084

HmNS28

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
123	1	1	0	0.0156	0.0157
125	1	1	0	0.0156	0.0157
127	5	5	0	0.0781	0.0814
129	7	7	0	0.1094	0.1161
131	4	2	1	0.0625	0.048
133	3	3	0	0.0469	0.048
135	4	2	1	0.0625	0.048
137	7	7	0	0.1094	0.1161
141	4	4	0	0.0625	0.0646
143	5	5	0	0.0781	0.0814
145	6	4	1	0.0938	0.0814
147	4	4	0	0.0625	0.0646
149	4	4	0	0.0625	0.0646
151	1	1	0	0.0156	0.0157
155	2	2	0	0.0313	0.0317
159	1	1	0	0.0156	0.0157
167	1	1	0	0.0156	0.0157
169	3	3	0	0.0469	0.048
185	1	1	0	0.0156	0.0157

HmNSa34

Allele	Count	Heterozygotes	Homozygotes	Freq.	Null Freq.
185	49	3	23	0.7903	0.3486
187	9	3	3	0.1452	0.0688
189	4	0	2	0.0645	0.0224

Appendix 11 Primer Sets Used in Parentage Analyses

Details of primer sets used for parentage analyses for both Abagold and HIK populations including testing conditions, labelling information and estimated allele size range

N. van den Berg

Name	Repeat	Size	Primer	Label	Accession No.	Testing Conditions	Size Range
HmNS28D F	$(CA)_{16}$	180bp	CAG TCA ATT TTC ATC GCA TT	FAM	EF033332	Standard PCR @ 55˚C	148 - 212
HmNS28D R			AGG TCG TTT TTC TCC TTC AG				
HmNS38T F	$(TCAC)_{10}$	454bp	CTG AGA CCC AAA GTT TTC TTT A	PET	EF 367113	Standard PCR @ 55˚C	414 - 494
HmNS38T R			ATC TAT GTT CAG GGT GTC AGT G				
HmNS58R F	$(GTT)_8$	246bp	TGC CAC TCA AAT GTT CCT TA	NED	EF 367119	Standard PCR @ 60˚C	222 - 270
HmNS58R R			CTA TTT CAG GTG TCC CCA GT				
HmNST7T F	$(CACT)_{26}$	313bp	CAC ATG GGT ACA ATG TGT GAA G	VIC	EF 455618	Standard PCR @ 60˚C	211 - 417
HmNST7T R			GGT AGC ACT GTT TCT CAC GA				
HmNS56D F	$(CA)_{20}$	259bp	TTC GGC AAG TGA ATG TCT AG	NED	EF 455619	Standard PCR @ 60˚C	219 - 299
HmNS56D R			CCG AGT TTG GAA TGT CTG AT				
N. Ruivo							
HmNR54H F	$(TTAGGG)_4$	359bp	5' TAA CAC TAA GTC CCT CAC CC 3'	VIC	EF063103	PCR @ 60 - 65°C	329-407
HmNR54H R			5' CAT TCT ACA TTC GAC ATT CG 3'				
HmNR120T F	$(TGAG)_{23}$	304bp	5' TTG AGC ATG AGT CGT TGA GC 3'	PET	EF121745	PCR @ 62°C	235-347
HmNR120T R			5' ACC TGC TCT TTA GCT CAG ATG G 3'				
HmNR136D F	$(CA)_{11}$	254bp	5' GAG TAA TAT GGG CAC CTC G 3'	VIC	DQ825710	PCR @ 60 - 65°C	211-309
HmNR136D R			5' GTT TGG AAT GTC TGA TTG GA 3'				
HmNR281P F	$(CTCAA)_{24}$	367bp	5' AAC CTT CAG TAA CCC ATG C 3'	FAM	EF512274	PCR @ 60°C	225-375
HmNR281P R			5' TGA ATA GGC ACC ATA AAG GG 3'				

Appendix 12 (a) CERVUS 2.0 Parentage Analyses Data for Abagold Population

All single parental pairs or candidate mother/father with a LOD score greater than 3.0 is assigned with 95% confidence. LOD scores less than 3.0 are indicated in red typeface

No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score	No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score
1	AbA01F	AbG06	AbG66	7.81E+00	51	AqA06F	AbG09	AbG76	3.30E+00
2	AbA01S	AbG20	AbG38	4.44E+00	52	AqA06S	AbG46	AbG84	2.74E+00
3	AbA02F	AbG93	AbG66	5.12E+00	53	AqA07F	AbG48	AbG90	7.18E+00
4	AbA02S	AbG03	AbG75	3.01E+00	54	AqA07S	AbG92	AbG75	6.23E+00
5	AbA03F	AbG55	AbG23	6.14E+00	55	AqA08F	AbG17	AbG67	3.98E+00
6	AbA03S	AbG02	AbG25	3.71E+00	56	AqA08S	AbG92	AbG66	6.35E+00
7	AbA04F	AbG02	AbG78	5.30E+00	57	AqA09F	AbG50	AbG64	7.00E+00
8	AbA04S	AbG02	AbG76	4.06E+00	58	AqA09S	AbG50	AbG90	4.97E+00
9	AbA05F	AbG06	AbG90	3.81E+00	59	AqA10F	AbG94	AbG90	7.24E+00
10	AbA05S	AbG02	AbG28	2.82E+00	60	AqA10S	AbG46	AbG64	3.78E+00
11	AbA06F	AbG94	AbG90	5.96E+00	61	AqA11F	AbG94	AbG75	6.95E+00
12	AbA06S	AbG92	AbG65	4.82E+00	62	AqA11S	AbG48	AbG67	5.90E+00
13	AbA07F	AbG02	AbG64	3.83E+00	63	AqA12F	AbG55	AbG63	3.92E+00
14	AbA07S	AbG49	AbG24	3.40E+00	64	AqA12S	AbG56	AbG90	3.47E+00
15	AbA08F	AbG06	AbG85	4.11E+00	65	AqA13F	AbG92	AbG78	5.74E+00
16	AbA08S	AbG08	AbG30	3.89E+00	66	AqA13S	AbG50	AbG78	6.92E+00
17	AbA09F	AbG47	AbG65	3.86E+00	67	AqA14F	AbG09	AbG75	5.71E+00
18	AbA09S	AbG92	AbG28	9.49E+00	68	AqA14S	AbG93	AbG90	9.12E+00
19	AbA10F	AbG08	AbG91	4.41E+00	69	AqA15F	AbG50	AbG84	7.59E+00
20	AbA10S	AbG19	AbG25	4.68E+00	70	AqA15S	AbG06	AbG26	5.94E+00
21	AbA11F	AbG03	AbG60	2.99E+00	71	AqA16F	AbG18	AbG64	3.79E+00
22	AbA11S	AbG48	AbG67	3.80E+00	72	AqA16S	AbG09	AbG36	4.31E+00
23	AbA12F	AbG05	AbG60	5.65E+00	73	AqA17F	AbG02	AbG90	7.82E+00
24	AbA12S	AbG58	AbG75	4.77E+00	74	AqA17S	AbG49	AbG29	4.03E+00
25	AbA13F	AbG04	AbG28	8.71E+00	75	AqA18F	AbG05	AbG30	3.51E+00
26	AbA13S	AbG95	AbG60	3.19E+00	76	AqA18S	AbG09	AbG85	3.41E+00
27	AbA14F	AbG47	AbG28	4.24E+00	77	AqA19F	AbG92	AbG90	4.51E+00
28	AbA14S	AbG48	AbG90	7.60E+00	78	AqA19S	AbG50	AbG78	5.38E+00
29	AbA15F	AbG45	AbG60	4.97E+00	79	AqA20F	AbG09	AbG87	4.37E+00
30	AbA15S	AbG01	AbG90	4.31E+00	80	AqA20S	AbG13	AbG68	4.68E+00
31	AbA16F	AbG06	AbG65	6.10E+00	81	HA01F	AbG17	AbG30	4.55E+00
32	AbA16S	AbG93	AbG88	8.13E+00	82	HA01S	AbG92	AbG60	5.65E+00
33	AbA17F	AbG93	AbG90	6.71E+00	83	HA02F	AbG02	AbG66	6.78E+00
34	AbA17S	AbG02	AbG63	3.95E+00	84	HA02S	AbG95	AbG90	2.79E+00
35	AbA18F	AbG50	AbG66	5.47E+00	85	HA03F	AbG06	AbG66	3.13E+00
36	AbA18S	AbG16	AbG87	5.67E+00	86	HA03S	AbG92	AbG63	3.04E+00
37	AbA19F	AbG04	AbG67	2.96E+00	87	HA04F	AbG09	AbG80	4.58E+00
38	AbA19S	AbG02	AbG67	4.77E+00	88	HA04S	AbG45	AbG42	3.89E+00
39	AbA20F	AbG16	AbG67	3.53E+00	89	HA05F	AbG08	AbG69	5.57E+00
40	AbA20S	AbG18	AbG60	4.88E+00	90	HA05S	AbG19	AbG77	3.33E+00
41	AqA01F	AbG93	AbG66	5.10E+00	91	HA06F	AbG06	AbG66	3.80E+00
42	AqA01S	AbG50	AbG24	5.03E+00	92	HA06S	AbG95	AbG65	2.33E+00
43	AqA02F	AbG02	AbG66	6.73E+00	93	HA07F	AbG45	AbG27	4.38E+00
44	AqA02S	AbG19	AbG66	3.23E+00	94	HA07S	AbG11	AbG72	3.76E+00
45	AqA03F	AbG58	-	3.98E+00	95	HA08F	AbG06	AbG28	3.54E+00
46	AqA03S	AbG05	AbG41	4.84E+00	96	HA08S	AbG92	AbG62	3.42E+00
47	AqA04F	AbG19	AbG82	3.59E+00	97	HA09F	AbG45	AbG26	3.64E+00
48	AqA04S	AbG50	AbG80	4.31E+00	98	HA09S	AbG10	AbG28	8.21E+00
49	AqA05F	AbG15	AbG90	3.64E+00	99	HA10F	AbG95	AbG28	3.59E+00
50	AqA05S	AbG08	AbG73	4.88E+00	100	HA10S	AbG92	AbG60	4.53E+00

No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score	No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score
101	HA11F	AbG52	AbG85	4.87E+00	**151**	IJA16F	AbG53	AbG91	4.79E+00
102	HA11S	AbG49	AbG96	5.82E+00	**152**	IJA16S	AbG52	AbG74	4.89E+00
103	HA12F	AbG92	AbG66	5.05E+00	**153**	IJA17F	AbG01	AbG72	4.84E+00
104	HA12S	AbG20	AbG72	2.70E+00	**154**	IJA17S	AbG49	AbG90	5.25E+00
105	HA13F	AbG46	AbG72	3.84E+00	**155**	IJA18F	AbG92	AbG66	4.37E+00
106	HA13S	AbG48	AbG64	6.51E+00	**156**	IJA18S	AbG92	AbG69	3.27E+00
107	HA14F	AbG10	AbG28	9.35E+00	**157**	IJA19F	AbG45	AbG34	3.35E+00
108	HA14S	AbG92	AbG75	4.72E+00	**158**	IJA19S	AbG02	AbG90	5.29E+00
109	HA15F	AbG48	AbG66	8.30E+00	**159**	IJA20F	AbG49	AbG78	2.86E+00
110	HA15S	-	AbG78	4.57E+00	**160**	IJA20S	AbG95	AbG28	7.05E+00
111	HA16F	AbG07	AbG97	5.51E+00	**161**	RBA01F	AbG17	AbG85	5.66E+00
112	HA16S	AbG08	AbG25	5.53E+00	**162**	RBA01S	AbG92	AbG75	4.52E+00
113	HA17F	AbG50	AbG78	3.73E+00	**163**	RBA02F	AbG50	AbG28	3.44E+00
114	HA17S	AbG50	AbG66	6.67E+00	**164**	RBA02S	AbG17	AbG64	2.66E+00
115	HA18F	AbG45	AbG68	6.77E+00	**165**	RBA03F	AbG08	AbG23	4.50E+00
116	HA18S	AbG07	AbG67	3.34E+00	**166**	RBA03S	AbG08	AbG83	4.08E+00
117	HA19F	AbG95	AbG64	3.08E+00	**167**	RBA04F	AbG10	AbG75	4.87E+00
118	HA19S	AbG50	AbG34	3.88E+00	**168**	RBA04S	AbG50	AbG78	7.11E+00
119	HA20F	AbG92	AbG28	8.74E+00	**169**	RBA05F	AbG92	AbG78	6.04E+00
120	HA20S	AbG50	AbG43	4.76E+00	**170**	RBA05S	AbG20	AbG97	2.86E+00
121	IJA01F	AbG53	AbG64	3.22E+00	**171**	RBA06F	AbG14	AbG80	3.27E+00
122	IJA01S	AbG92	AbG78	5.93E+00	**172**	RBA06S	AbG11	AbG28	9.49E+00
123	IJA02F	AbG94	AbG31	6.96E+00	**173**	RBA07F	AbG06	AbG66	3.11E+00
124	IJA02S	AbG56	AbG88	5.14E+00	**174**	RBA07S	AbG08	AbG65	6.10E+00
125	IJA03F	AbG06	AbG67	3.76E+00	**175**	RBA08F	AbG19	AbG90	4.18E+00
126	IJA03S	AbG09	-	4.34E+00	**176**	RBA08S	AbG50	-	3.48E+00
127	IJA04F	AbG56	AbG72	5.24E+00	**177**	RBA09F	AbG93	AbG66	8.91E+00
128	IJA04S	AbG02	AbG72	7.85E+00	**178**	RBA09S	AbG04	AbG28	7.20E+00
129	IJA05F	AbG95	AbG89	5.35E+00	**179**	RBA10F	AbG06	AbG74	6.34E+00
130	IJA05S	AbG46	AbG60	4.86E+00	**180**	RBA10S	AbG07	AbG27	3.95E+00
131	IJA06F	AbG92	AbG64	2.18E+00	**181**	RBA11F	AbG06	AbG78	4.45E+00
132	IJA06S	AbG07	AbG67	2.59E+00	**182**	RBA11S	AbG11	AbG75	3.66E+00
133	IJA07F	AbG56	AbG31	4.09E+00	**183**	RBA12F	AbG53	AbG75	5.23E+00
134	IJA07S	AbG53	AbG66	3.45E+00	**184**	RBA12S	AbG09	AbG90	5.19E+00
135	IJA08F	AbG05	AbG32	4.19E+00	**185**	RBA13F	AbG05	AbG73	5.86E+00
136	IJA08S	AbG06	AbG60	3.79E+00	**186**	RBA13S	AbG48	AbG60	6.26E+00
137	IJA09F	AbG56	AbG66	2.58E+00	**187**	RBA14F	AbG93	AbG75	3.17E+00
138	IJA09S	AbG06	AbG60	5.09E+00	**188**	RBA14S	AbG07	AbG75	3.45E+00
139	IJA10F	AbG50	AbG75	4.23E+00	**189**	RBA15F	AbG19	AbG73	1.83E+00
140	IJA10S	AbG05	AbG60	5.65E+00	**190**	RBA15S	AbG56	AbG64	4.24E+00
141	IJA11F	AbG44	AbG63	2.21E+00	**191**	RBA16F	AbG93	AbG65	5.97E+00
142	IJA11S	AbG45	AbG80	5.95E+00	**192**	RBA16S	AbG92	AbG35	4.11E+00
143	IJA12F	AbG13	AbG74	2.37E+00	**193**	RBA17F	AbG58	AbG75	4.48E+00
144	IJA12S	AbG02	AbG75	4.40E+00	**194**	RBA17S	AbG49	AbG71	3.12E+00
145	IJA13F	AbG52	AbG74	4.58E+00	**195**	RBA18F	AbG04	AbG32	3.90E+00
146	IJA13S	AbG45	AbG82	3.82E+00	**196**	RBA18S	AbG01	AbG90	3.17E+00
147	IJA14F	AbG92	AbG66	4.95E+00	**197**	RBA19F	AbG56	AbG64	3.07E+00
148	IJA14S	AbG06	AbG67	4.69E+00	**198**	RBA19S	AbG52	AbG33	2.47E+00
149	IJA15F	AbG92	AbG73	2.24E+00	**199**	RBA20F	AbG92	AbG64	4.05E+00
150	IJA15S	AbG14	AbG77	3.63E+00	**200**	RBA20S	AbG20	AbG23	4.02E+00

Appendix 12 (b) CERVUS 2.0 Parentage Analyses Data for HIK Population

All single parental pairs or candidate mother/father that was assigned with a LOD score of greater than 3.0 is unambiguous at 95% confidence. Assignments with LOD scores less than 3.0 are indicated in red typeface.

No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score	No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score
1	AbH01F	Hik19	Hik81	3.90E+00	51	AqH06F	Hik04	Hik76	5.50E+00
2	AbH01S	Hik47	Hik40	5.85E+00	52	AqH06S	Hik52	Hik94	9.38E+00
3	AbH02F	Hik68	Hik29	7.03E+00	53	AqH07F	Hik15	Hik35	5.87E+00
4	AbH02S	Hik69	Hik82	6.04E+00	54	AqH07S	Hik45	Hik85	4.45E+00
5	AbH03F	Hik67	Hik76	5.80E+00	55	AqH08F	Hik55	Hik82	9.69E+00
6	AbH03S	Hik67	Hik84	2.33E+00	56	AqH08S	Hik20	Hik40	8.12E+00
7	AbH04F	Hik48	Hik99	2.43E+00	57	AqH09F	Hik16	Hik27	5.12E+00
8	AbH04S	Hik55	Hik100	6.03E+00	58	AqH09S	Hik45	Hik32	4.05E+00
9	AbH05F	Hik52	Hik36	5.12E+00	59	AqH10F	Hik21	Hik35	6.77E+00
10	AbH05S	Hik55	Hik38	3.90E+00	60	AqH10S	Hik01	Hik30	4.60E+00
11	AbH06F	Hik72	Hik76	6.98E+00	61	AqH11F	Hik20	Hik23	5.91E+00
12	AbH06S	Hik75	Hik100	4.49E+00	62	AqH11S	Hik04	Hik32	3.98E+00
13	AbH07F	Hik62	Hik38	2.87E+00	63	AqH12F	Hik68	Hik86	8.90E+00
14	AbH07S	Hik67	Hik38	2.65E+00	64	AqH12S	Hik46	Hik33	2.91E+00
15	AbH08F	Hik17	Hik36	8.97E+00	65	AqH13F	Hik51	Hik29	4.22E+00
16	AbH08S	Hik55	Hik84	8.51E+00	66	AqH13S	Hik12	Hik29	5.18E+00
17	AbH09F	Hik15	Hik83	4.79E+00	67	AqH14F	Hik57	Hik26	4.54E+00
18	AbH09S	Hik04	Hik94	5.76E+00	68	AqH14S	Hik47	Hik44	3.65E+00
19	AbH10F	Hik18	Hik78	4.59E+00	69	AqH15F	Hik02	Hik100	2.37E+00
20	AbH10S	Hik45	Hik94	7.38E+00	70	AqH15S	Hik70	Hik94	3.29E+00
21	AbH11F	Hik10	Hik28	4.88E+00	71	AqH16F	Hik62	Hik82	2.73E+00
22	AbH11S	Hik10	Hik37	6.93E+00	72	AqH16S	Hik10	Hik90	2.67E+00
23	AbH12F	Hik68	Hik76	7.03E+00	73	AqH17F	Hik17	Hik37	7.20E+00
24	AbH12S	Hik57	Hik100	3.55E+00	74	AqH17S	Hik01	Hik37	8.91E+00
25	AbH13F	Hik11	Hik79	2.11E+00	75	AqH18F	Hik04	Hik23	4.60E+00
26	AbH13S	Hik07	Hik29	5.17E+00	76	AqH18S	Hik61	Hik23	5.30E+00
27	AbH14F	Hik13	Hik76	3.02E+00	77	AqH19F	Hik15	Hik37	8.95E+00
28	AbH14S	Hik74	Hik76	7.70E+00	78	AqH19S	Hik61	Hik100	3.36E+00
29	AbH15F	Hik55	Hik94	6.77E+00	79	AqH20F	Hik14	Hik100	3.80E+00
30	AbH15S	Hik17	Hik35	6.87E+00	80	AqH20S	Hik74	Hik76	8.67E+00
31	AbH16F	Hik74	Hik94	7.72E+00	81	HH01F	Hik68	Hik82	4.55E+00
32	AbH16S	Hik16	Hik35	9.70E+00	82	HH01S	Hik55	Hik89	5.27E+00
33	AbH17F	Hik49	Hik94	7.60E+00	83	HH02F	Hik46	Hik77	6.63E+00
34	AbH17S	Hik67	Hik76	6.57E+00	84	HH02S	Hik46	Hik92	4.26E+00
35	AbH18F	Hik52	Hik94	4.23E+00	85	HH03F	Hik08	Hik95	2.50E+00
36	AbH18S	Hik70	Hik23	3.10E+00	86	HH03S	Hik57	Hik94	5.30E+00
37	AbH19F	Hik65	Hik82	3.74E+00	87	HH04F	Hik20	Hik24	9.47E+00
38	AbH19S	Hik50	Hik94	5.15E+00	88	HH04S	Hik53	Hik37	3.11E+00
39	AbH20F	Hik06	Hik78	3.22E+00	89	HH05F	Hik55	Hik85	8.38E+00
40	AbH20S	Hik53	Hik97	3.11E+00	90	HH05S	Hik69	Hik85	6.23E+00
41	AqH01F	Hik72	Hik29	3.24E+00	91	HH06F	Hik45	Hik87	7.49E+00
42	AqH01S	Hik53	Hik82	4.16E+00	92	HH06S	Hik57	Hik97	4.15E+00
43	AqH02F	Hik45	Hik38	3.66E+00	93	HH07F	Hik56	Hik39	3.31E+00
44	AqH02S	Hik67	Hik23	5.07E+00	94	HH07S	Hik05	Hik42	3.50E+00
45	AqH03F	Hik74	Hik35	5.72E+00	95	HH08F	Hik02	Hik37	3.97E+00
46	AqH03S	Hik15	Hik37	7.12E+00	96	HH08S	Hik45	Hik76	7.38E+00
47	AqH04F	Hik57	Hik82	3.41E+00	97	HH09F	Hik17	Hik29	7.57E+00
48	AqH04S	Hik68	Hik94	3.80E+00	98	HH09S	Hik72	Hik91	3.49E+00
49	AqH05F	Hik72	Hik26	4.00E+00	99	HH10F	Hik20	Hik78	6.59E+00
50	AqH05S	Hik10	Hik40	6.28E+00	100	HH10S	Hik70	Hik100	4.85E+00

No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score	No.	Offspring ID	Candidate MALE	Candidate FEMALE	LOD Score
101	HH11F	Hik20	-	8.50E+00	151	IJH16F	Hik72	Hik29	5.15E+00
102	HH11S	Hik65	Hik43	4.10E+00	152	IJH16S	Hik55	Hik94	6.13E+00
103	HH12F	Hik74	Hik85	2.93E+00	153	IJH17F	Hik04	Hik76	4.86E+00
104	HH12S	Hik45	Hik94	9.32E+00	154	IJH17S	Hik68	Hik94	5.78E+00
105	HH13F	Hik65	Hik84	4.07E+00	155	IJH18F	Hik20	Hik29	7.58E+00
106	HH13S	Hik63	Hik29	1.76E+00	156	IJH18S	-	Hik36	7.08E+00
107	HH14F	Hik12	Hik35	6.13E+00	157	IJH19F	Hik51	Hik76	6.11E+00
108	HH14S	-	Hik35	6.05E+00	158	IJH19S	Hik03	Hik40	3.45E+00
109	HH15F	Hik68	Hik82	4.67E+00	159	IJH20F	Hik22	Hik23	3.06E+00
110	HH15S	Hik69	Hik93	3.69E+00	160	IJH20S	Hik61	Hik37	3.40E+00
111	HH16F	Hik68	Hik94	6.47E+00	161	RBH01F	Hik68	Hik76	6.11E+00
112	HH16S	Hik01	-	6.53E+00	162	RBH01S	Hik74	Hik84	5.13E+00
113	HH17F	Hik65	Hik27	6.49E+00	163	RBH02F	Hik13	Hik76	3.35E+00
114	HH17S	Hik14	Hik31	3.23E+00	164	RBH02S	Hik11	Hik94	4.01E+00
115	HH18F	Hik17	Hik82	8.16E+00	165	RBH03F	Hik20	Hik36	9.49E+00
116	HH18S	Hik67	Hik94	7.56E+00	166	RBH03S	Hik19	Hik85	5.79E+00
117	HH19F	Hik67	Hik30	3.64E+00	167	RBH04F	Hik22	Hik30	4.14E+00
118	HH19S	Hik55	Hik24	9.75E+00	168	RBH04S	Hik65	Hik94	5.58E+00
119	HH20F	Hik75	Hik85	3.22E+00	169	RBH05F	Hik10	Hik35	6.52E+00
120	HH20S	Hik52	Hik37	5.99E+00	170	RBH05S	Hik07	Hik94	4.75E+00
121	IJH01F	Hik12	Hik42	2.76E+00	171	RBH06F	Hik70	Hik92	4.68E+00
122	IJH01S	Hik17	Hik86	5.32E+00	172	RBH06S	Hik19	Hik39	3.93E+00
123	IJH02F	Hik67	Hik86	5.79E+00	173	RBH07F	Hik10	Hik40	7.24E+00
124	IJH02S	Hik15	Hik40	5.22E+00	174	RBH07S	Hik12	Hik29	6.82E+00
125	IJH03F	Hik47	Hik36	5.94E+00	175	RBH08F	Hik11	Hik35	6.71E+00
126	IJH03S	Hik52	Hik98	4.66E+00	176	RBH08S	Hik55	Hik43	4.92E+00
127	IJH04F	Hik68	Hik76	7.46E+00	177	RBH09F	Hik48	Hik84	4.99E+00
128	IJH04S	Hik68	Hik24	2.57E+00	178	RBH09S	Hik64	Hik40	4.37E+00
129	IJH05F	Hik67	Hik76	6.62E+00	179	RBH10F	Hik66	Hik27	2.94E+00
130	IJH05S	Hik01	Hik27	4.85E+00	180	RBH10S	Hik70	Hik76	4.28E+00
131	IJH06F	Hik64	Hik76	7.73E+00	181	RBH11F	Hik68	Hik84	8.00E+00
132	IJH06S	Hik69	Hik38	3.15E+00	182	RBH11S	Hik74	Hik33	4.39E+00
133	IJH07F	Hik14	Hik40	6.48E+00	183	RBH12F	Hik15	Hik99	4.73E+00
134	IJH07S	Hik46	Hik43	3.64E+00	184	RBH12S	Hik46	Hik40	4.63E+00
135	IJH08F	Hik66	Hik40	4.43E+00	185	RBH13F	Hik72	Hik76	7.14E+00
136	IJH08S	Hik68	Hik100	4.21E+00	186	RBH13S	Hik61	Hik24	2.26E+00
137	IJH09F	Hik15	Hik29	9.83E+00	187	RBH14F	Hik71	Hik76	6.52E+00
138	IJH09S	Hik17	Hik29	8.52E+00	188	RBH14S	Hik50	Hik29	2.87E+00
139	IJH10F	Hik68	Hik76	7.49E+00	189	RBH15F	Hik57	Hik42	4.16E+00
140	IJH10S	Hik17	Hik97	4.60E+00	190	RBH15S	Hik16	Hik40	4.98E+00
141	IJH11F	Hik02	Hik78	2.50E+00	191	RBH16F	Hik11	Hik35	6.00E+00
142	IJH11S	Hik10	Hik36	7.01E+00	192	RBH16S	Hik52	Hik94	8.08E+00
143	IJH12F	Hik04	Hik97	2.79E+00	193	RBH17F	Hik46	Hik31	4.59E+00
144	IJH12S	Hik68	Hik94	5.07E+00	194	RBH17S	Hik46	Hik40	3.70E+00
145	IJH13F	Hik68	Hik76	6.25E+00	195	RBH18F	Hik14	Hik36	8.88E+00
146	IJH13S	Hik09	Hik82	4.51E+00	196	RBH18S	Hik64	Hik31	2.66E+00
147	IJH14F	Hik15	Hik98	4.55E+00	197	RBH19F	Hik10	Hik29	8.44E+00
148	IJH14S	Hik14	Hik40	5.23E+00	198	RBH19S	Hik22	Hik77	3.45E+00
149	IJH15F	Hik72	Hik78	3.85E+00	199	RBH20F	Hik10	Hik40	5.99E+00
150	IJH15S	Hik52	Hik100	4.84E+00	200	RBH20S	Hik09	Hik28	7.62E+00

References

Abiola O., Angel J.M., Avner P., Bachmanov A., Belknap J.K., Bennett B., Blankenhorn E.P., Blizard D.A., Bolivar V., Brockmann G.A., Buck K.J., Bureau J.F., Casley W.L., Chesler E.J., Cheverud J.M., Churchill G.A., Cook M., Crabbe J.C., Crusio W.E., Darvasi A., de Haan G., Demant P., Doerge R.W., Elliot R.W., Farber C.R., Flaherty L., Flint J., Gershenfield H., Gibson J.P., Gu J., Gu W., Himmelbauer H., Hitzemann R., Hsu H.C., Hunter K., Iraqi F.A., Jansen R.C., Johnson T.E., Jones B.C., Kempermann G., Lammert F., Lu L., Manly K.F., Mathews D.B., Medrano J.F., Mehrabian M., Mittleman G., Mock B.A., Mogil J.S, Montagutelli X., Morahan G., Mountz J.D., Nagase H., Nowakowski R.S., O'Hara B.F., Osadchuk A.V., Paigen B., Palmer A.A., Peirce J.L., Pomp D., Rosemann M., Rosen G.D., Schalkwyk L.C., Seltzer Z., Settle S., Shimomura K., Shou S., Sikela J.M., Siracusa L.D., Spearow J.L., Teuscher C., Threadgill D.W., Toth L.A., Toye A.A., Vadasz C., Van Zant G., Wakeland E., Williams R.W. Zhang H.G. and Zou F. (2003) The nature and identification of quantitative trait loci: a community's view. *Nature Genetics Reviews*, 4: 911-915

Aho T., Rönn J., Piironen J. and Björklund M. (2006) Impacts of effective population size on genetic diversity in hatchery reared Brown trout (*Salmo trutta* L.) populations. *Aquaculture*, 253: 244-248

Amos W., Hoffman J.I., Frodsham A., Zhang L., Best S. and Hill V.S. (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, 7: 10-14

Anne C. (2006) Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects. *Genetica*, 127: 101-120

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

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

Beaumont A.R. and Hoare K. (2003) Biotechnology and Genetics in Fisheries and Aquaculture (Publishers: *Blackwell Publishing*), p13

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

Bentsen H.B. and Olesen I. (2002) Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture*, 204: 349-359

Bernatchez L. and Duchesne P. (2000) Individual-based genotype analysis studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences*, 57: 1-12

Bester A.E., Slabbert R. and D'Amato M.E. (2004) Isolation and characterisation of microsatellite markers in the South African abalone (*Haliotis midae*). *Molecular Ecology Notes*, 4: 618-619

BIOEDIT Sequence Alignment Editor [online] available: http://www.mbio.ncsu.edu/BioEdit/bioedit.html, 16 February 2008

Boudry P., Collet B., Cornette F., Hervouet V. and Bonhomme F. (2002) High variance in reproductive success of Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture*, 204: 283-296

Brown R.C., Woolliams J.A. and McAndrew B.J. (2005) Factors influencing population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture*, 247: 219-225Buckleton J. and Triggs C. (2006) The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. *Forensic Science International*, 160: 193-199

Buckleton J. and Triggs C. (2006) The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. *Forensic Science International*, 160: 193-199

Campbell D., Duchesne P. and Bernatchez L. (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Molecular Ecology*, 12: 1979-1991

Castro J., Bouza C., Presa P., Pino-Querido A., Riaza A., Ferreiro I., Sáchez L. and Martínez P. (2004) Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture*, 242: 119-135

Castro J., Pino A., Hermida M., Bouza C., Riaza A., Ferreiro I., Sáchez L. and Martínez P. (2006) A microsatellite marker tool for parentage analysis in Senegal sole (*Solea senegalensis*): Genotyping errors, null alleles and conformance to theoretical assumptions. *Aquaculture*, 261: 1194-1203

Cercueil A., Bellermain E. and Manel S. (2002) PARENTE: Computer program for parentage analysis. *Journal of Heredity*, 93: 458-459

Chambers G.K. and MacAvoy E.S. (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B*, 126: 455-476

Chistiakov D.A., Hellemans B. and Volckaert F.A.M. (2005) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, 255: 1-29

CHROMAS 1.45 [online] available: http://www.technelysium.com.au/chromas14x.html, 16 February 2008

Coastal and Marine Life [online] Available: http://www.aquarium.co.za/pdf/kids/educational/3B_abalone.pdf, 20 February 2008

Coimbra M.R.M., Kobayashi K., Koretsugu S., Hasegawa O., Ohara E., Ozaki A., Sakamoto T., Naruse K. and Okamoto N. (2003) A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture*, 220: 203-218

Cook P. (1998) The current status of abalone farming in South Africa. *Journal of Shellfish Research*, 17: 601-602

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

Dakin E.E. and Avise J.C. (2004) Microsatellite null alleles in parentage analysis. *Journal of Heredity*, 93: 504-509

Davis G.P. and Hetzel D.J.S. (2000) Integrating molecular genetic technology with traditional approaches for genetic improvement in aquaculture species. *Aquaculture Research*, 31: 3-10

Day E. and Branch G.M. (2002) Effects of sea urchins (*Parechinus angulosus*) on recruits and juveniles of abalone (*Haliotis midae*). *Ecological Monographs*, 72: 133-149

Day R.W. and Fleming A.E. (1992) *The determinants and measurement of abalone growth*. Abalone of the World, Biology, Fisheries and Culture: Proceedings of the 1st International Symposium on Abalone, 1989, La Paz, Mexico. [Shepherd S.A, Tegner M.J. and Guzman Del Proo S.A.], Fishing News Books

Degnan S.M., Imron., Geiger D.L. and Degnan B.M. (2006) Evolution in temperate and tropical seas: Disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae). *Molecular Phylogenetics and Evolution*, 41: 249-256

Deka R., Guangyun S., Smelser D., Ahong Y., Kimmel M. and Chakraborty R. (1999) Rate and directionality of mutations and effects of allele size constraints at anonymous, gene-associated and disease-causing trinucleotide loci. *Molecular Biology and Evolution*, 16: 1166-1177

Dekkers J.C.M. and Hospital F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature*, 3: 22-32

Díaz N.F. and Neira R. (2005) Biotechnology Applied to Aquaculture I. Classic Biotechnologies Applied to the Reproduction of Cultivated Species. *Ciencias Investigación Agronómicas*, 31: 39-52

Dodds K.G., Tate M.L. and Sise J.A. (2005) Genetic evaluation using parentage information from genetic markers. *Journal of Animal Science*, 83: 2271-2279

Dong S., Kong J., Zhang T., Meng X. and Wang R. (2006) Parentage determination of Chinese shrimp (*Fenneropenaeus chinensis*) based on microsatellite DNA markers. *Aquaculture*, 258: 283-288

Doyle R.W. and Herbinger C.M. (1995) Broodstock improvement strategies based on DNA fingerprinting: examples and cost-benefit analyses. *Aquaculture*, 137: 271-284

Duchesne P. and Bernatchez L. (2002) An analytical investigation of the dynamics of inbreeding in multi-generation supportive breeding. *Conservation Genetics*, 3: 47-60

Duchesne P., Godbout M.H. and Bernatchez L. (2002) PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes*, 2: 191-193

du Plessis A. (2006) A brief overview of the abalone industry in South Africa. *VI International Abalone Symposium*, Puerto Veras, Chile, 19-24 February

Edwards K.J., Barker J.H.A., Daly A., Jones C. and Karp A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques*, 20: 758-760

Elliot N.G. (2000) Genetic improvement programmes in abalone: what is the future? *Aquaculture Research*, 31: 51-59

Evans B., Bertlett J., Sweijd N., Cook P. and Elliot N.G. (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

Excoffier L. and Heckel G. (2006) Computer programs for population genetics analysis: a survival guide. *Nature*, 7: 745-758

FISHTECH™ INC [online] Available: http://www.fishtech.com/facts.html, 16 February 2008

Garcia de Leon F.J., Canonne M., Quillet E., Bonhomme F. and Chatain B. (1998) The application of microsatellite markers to breeding programmes in the sea bass, *Dicentrarchus labrax*. *Aquaculture*, 159: 303-316

Gaudeul M., Bottraud I.T., Barjon F. and Manel S. (2004) Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): comparison of AFLP and microsatellite markers. *Journal of Heredity*, 92:508-518

Geiger D.L. (2000) Distribution and biogeography of the recent Haliotidae (Gastropoda: Vetigastropoda) world-wide. *Bollettino Malacolgico*, 35: 57-120

Gerber S., Mariett S., Streiff R., Bodénès C. and Kremer A. (2000) Comparison of microsatellite and amplified length polymorphism markers for parentage analysis. *Molecular Ecology*, 9: 1037-1048

Gilbey J., Knox D., O'Sullivan M. and Verspoor E. (2005) Novel DNA makers for rapid, accurate and cost effective discrimination of the continental origin of Atlantic salmon (*Salmo salar* L.). *ICES Journal of Marine Science*, 62: 1609-1616

Goodnight K.F. and Queller D.C. (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*, 8: 1231-1234

Gordon H.R. and Cook P.A. (2001) World abalone supply, markets and pricing: historical, current and future. *Journal of Shellfish Research*, 20: 567-570

Gyapay G., Ginot F., Nguyen S., Vignal A. and Weiseenback J. (1996) Genotyping procedures in linkage mapping. *A Companion to Methods in Enzymology*, 9: 91-97

Hahn K.O. (1989). CRC *Handbook of culture of abalone and other marine gastropods*. Florida: CRC Press

Hamilton M.B., Pincus E.L., Di Fiore A. and Fleischer R.C. (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques*, 27: 500-507

Hayes B., Baranski M., Goddard M.E. and Robinson N. (2007) Optimisation of marker assisted selection for abalone breeding programs. *Aquaculture*, 265: 61-69

Herbinger C.M., Doyle R.W., Pitman E.R., Paquet D., Mesa K.A., Morris D.B., Wright J.M. and Cook D. (1995) DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture*, 137: 245-256

Hoffman J.I. and Amos W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, 14: 599-612

Hoff-Olsen P., Mevag B., Staalstrøm E., Hovde B., Egeland T. and Olaisen B. (1999) Extraction of DNA from decomposed human tissue An evaluation of five extraction methods for short tandem repeat typing. *Forensic Science International*, 105: 171-183

Holst P.J. (1999) Recording and on-farm evaluations and monitoring: breeding and selection. *Small Ruminant Research*, 34: 197-202

Hospital F., Goldringer. and Openshaw S. (2000) Efficient marker-based recurrent selection for multiple quantitative trait loci. *Genetics Research*, 75: 357-368

Huang B.X., Peakall R. and Hanna P.J. (2000) Analysis of genetic structure of blacklip abalone (*Haliotis rubra*) populations using RAPD, minisatellite and microsatellite markers. *Marine Biology*, 136: 207-216

Isberg S.R., Chen Y., Barker S.G. and Moran C. (2004) Analysis of microsatellites and parentage testing in saltwater crocodiles. *Journal of Heredity*, 95: 445-449

Jarne P. and Lagoda P.J.L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11: 424-429

Jerry D.R., Preston N.P., Crocos P.J., Keys S., Meadows J.R.S. and Li Y. (2004) Parentage determination of Kuruma shrimp *Penaeus* (*Marsupenaeus*) *japonicas* using microsatellite markers (Bate). *Aquaculture*, 235: 237-247

Jerry D.R., Preston N.P., Crocos P.J., Keys S., Meadows J.R.S. and Li Y. (2006a) Application of DNA parentage analyses for determining relative growth rates of *Panaeus japonicus* families reared in commercial ponds. *Aquaculture*, 254: 171-181

Jerry D.R., Evans B.S., Kenway M. and Wilson K. (2006b). Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon*. *Aquaculture*, 255: 542-547

Jones A.G. and Ardren W.R. (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, 12: 2511-2523

Jones E.S., Dupal M.P., Kölliker R., Drayton M.C. and Forster J.W. (2001) Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.) *Theoretical and Applied Genetics*, 102: 405-415

Kirby V.L., Villa R. and Powers D.A. (1998) Identification of microsatellites in the Californian red abalone, *Haliotis rufescens*. *Journal of Shellfish Research*, 17: 801-804

Lai Y. and Sun F. (2003) Microsatellite mutations during the polymerase chain reaction: mean field approximations and their applications. *Journal of Theoretical Biology*, 224: 127-137

Lai Y. and Sun F. (2004) Sampling distribution for microsatellites amplified by PCR: mean field approximation and its applications to genotyping. *Journal of Theoretical Biology*, 228: 185-194

Lagercrantz U., Ellegren H., and Anderson L. (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research*, 21: 1111-1115

Li Q. and Kijima A. (2007) Sequences characterisation of microsatellite DNA sequences in Pacific abalone (*Haliotis discus hannai*). *Journal of Ocean University of China*, 6: 47-52

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

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

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

Liu Z.J. and Cordes J.F. (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238: 1-37

Leighton D.L. (1989) Abalone (genus *Haliotis*) mariculture on the North American Pacific coast. *Fishery Bulletin*, 87: 689-702

Leighton D.L. (2000) *The Biology and Culture of the California Abalones*. Pittsburgh, P.A. Dorrance Publishing Co., Inc. 216pp

Loubser N. (2005) Abalone farming in South Africa: a regional success story. *7th Aquaculture Association of Southern Africa Conference*, Grahamstown, 12-16 September

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

Manel S., Gaggiotti O.E. and Waples R.S. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20: 136-142

Marine and Coastal Management [online] available: http://www.mcm-deat.gov.za, 16 February 2008

Marshall T.C., Slate J., Kruuk L.E.B. and Pemberton J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7: 639-655

Martinez V. (2006) Importance and implementation of molecular markers in selective breeding programs for aquaculture species. *8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, MG, Brazil, 13-18 August

Mayfield S. and Branch G.M. (2000) Interrelations among rock lobsters, sea urchins, and juvenile abalone: implications for community management. *Canadian Journal of Fisheries and Aquatic Sciences*, 57: 2175-2185

McBride S. and Conte F.S. (1996) California Abalone Aquaculture. University of California, Davis Outreach Program Report

McShane P.E. (1992) *Early life history of abalone: a review*. Abalone of the World, Biology, Fisheries and Culture: Proceedings of the 1st International Symposium on Abalone, 1989, La Paz, Mexico. [Shepherd S.A, Tegner M.J. and Guzman Del Proo S.A.], Fishing News Books

Melamed P., Gong Z., Fletcher G. and Hew C.L. (2002) The potential impact of modern biotechnology on fish aquaculture. *Aquaculture*, 204: 255-269

Miller M.J. and Yuan B.Z. (1997) Semi-automated resolution of overlapping stutter patterns in genomic microsatellite analysis. *Analytical Biochemistry*, 251: 50-56

Morin P.A., Luikart G. and Wayne R.K. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19: 210-216

Mueller U.G. and Wolfenbarger L.L. (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*, 14: 389-394

NCBI BLAST [online] available: http://www.ncbi.nih.gov/BLAST/, 16 February 2008

Neff B.D., Fu P. and Gross M.R. (2000a) Microsatellite multiplexing in fish. *Transactions of the American Fisheries Society*, 129: 584-593

Neff B.D., Repka J. and Gross M.R. (2000b) Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? *Molecular Ecology*, 9: 529-539

Norris A.T., Bradley D.G. and Cunningham E.P. (2000) Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture*, 182: 73-83

Oakes F.R. and Ponte R.D. (1996) The abalone market: Opportunities for cultured abalone. *Aquaculture*, 140: 187-195

Paetkau D. (1999) Microsatellites obtained using strand extension: an enrichment protocol. *BioTechniques*, 26: 690-697

Pemberton J.M., Slate J., Bancroft D.R. and Barrett J.A. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, 4: 249-252

Pérez M., Cruz F and Presa P. (2005) Distribution properties of polymononucleotide repeats in Molluscan genomes. *Journal of Heredity*, 96: 40-51

Piggott M.P., Banks S.C. and Beheregaray L.B. (2006) Use of SSCP to improve the efficiency of microsatellite identification from microsatellite-enriched libraries. *Molecular Ecology Notes*, 6: 613-615

Primer 3 [online] available: https://sourceforge.net/project/showfiles.php, 16 February 2008

Primmer C.R. Raudsepp T., Chowdhary B.P. Møller A.P. and Ellegren H. (1997) Low frequency of microsatellites in the avian genome. *Genome Research*, 7: 471-482

QIAGEN® PCR Cloning Handbook. (2001) For QIAGEN® supplied PCR Cloning Kits. [online] Available: http://www1.qiagen.com/literature/, 20 February 2008

Rahman M.H., Jaquish B. and Khasa P.D. (2000) Optimization of PCR protocol in microsatellite analysis with silver and SYBR stains. *Plant Molecular Biology Reporter*, 18: 339-348

Refseth U.H., Fangan B.M. and Jakobsen K.S. (1997) Hybridisation capture of microsatellites directly from genomic DNA. *Electrophoresis*, 18: 1519-1523

Rengmark A.H., Slettan A., Skaala O., Lie O. and Lingaas F. (2005) Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture*, 253: 229-237

Richards R.I. and Sutherland G.R. (1994) Simple repeat DNA is not replicated simply. *Nature Genetics*, 6: 114-116

Robinson N., Baranski M., Rourke M. and Hayes B. (2003). Application of advanced genetic techniques to abalone breeding. *10th Abalone Aquaculture Workshop*, Abalone Aquaculture Subprogram, Fisheries Research and Development Corporation, Canberra, Port Lincoln, Australia, 19-21 November

Roodt-Wilding R. and Slabbert R. (2006) Molecular markers to assist the South African abalone industry. *South African Journal of Science*, 102: 99-102

Ross C.L., Dyer K.A., Erez T., Miller S.J., Jaenike J. and Markow T.A. (2003) Rapid divergence of microsatellite abundance among species of Drosophila. *Molecular Biology and Evolution*, 20: 1143-1157

Rychlik W. and Rhoads R.E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Research*, 17: 8543–8551

W. Rychlik, Oligo 4.1 Primer Analysis Software, 1992.

Saghai-Maroof M.A., Solima K.M., Jorgenson R.A. and Allard R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proceedings of the National Academy of Sciences* (USA), 81: 8014-8018

Sambrook J., Fritsch E.F. and Maniatis T. (1989) Molecular cloning: A laboratory Manual. *Cold Spring Harbor*, Laboratory Press, New York

Schlötterer C. and Tautz D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20: 211-215

Sekino M. and Hara M. (2006) Linkage maps for the Pacific Abalone (genus *Haliotis*) based on microsatellite DNA markers. *Genetics*, 175: 945-954

Selvamani M.J.P., Degnan S.M. and Degnan B.M. (2001) Microsatellite genotyping of individual larvae: parentage assignment in aquaculture. *Marine Biotechnology*, 3: 478-485

Slabbert R. and Roodt-Wilding R. (2006) Non-destructive sampling of juvenile abalone using epipodial tentacles and mucus: method and application. *African Journal of Marine Science*, 28: 719-721

Slate J., Marshall T. and Pemberton J. (2000) A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Molecular Ecology*, 9: 801-808

Smith P.J. and Conroy A.M. (1992) Loss of genetic variation in hatchery produced abalone, *Haliotis iris*. *New Zealand Journal of Marine and Freshwater Research*, 26: 81-85

Sonesson A.K. (2003) Possibilities for marker-assisted selection in fish breeding schemes. *Paper presented at MAS workshop held prior to conference*. Marker Assisted Selection: A fast track to increase genetic gain in plant and animal breeding, Turin, Italy, 17-18 October. Molecular marker assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries, *Conference 10*, *FAO Biotechnology Electronic Forum*, 17 November -12 December

Sonesson A.K. (2005) A combination of walk-back selection and optimum contribution selection in fish: a simulation study. *Genetics Selection and Evolution*, 37: 587-599

Steven C.R., Hill J., Masters B. and Place A.R. (2005) Genetic markers in blue crabs (*Callinectes sapidus*) I: Isolation and characterization of microsatellite markers. *Journal of Experimental Marine Biology and Ecology*, 319: 3-14

Stevens M.M. (2003) Cultured Abalone (spp. *Haliotis*). *Seafood Watch Seafood Report*, Monterey Bay Aquarium, 1-12

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

Taggart J.B. (2007) FAP: an exclusion-based parental assignment program with enhanced predictive functions. *Molecular Ecology Notes*, 7: 412-415

Tarr R.J.Q., Williams P.V.G. and Mackenzie A.J. (1996) Abalone, sea urchins and rock lobster: a possible ecological shift may affect traditional fisheries. *South African Journal of Marine Sciences*, 17: 319-323

Tarr R. (2000) Perlemoen: The South African abalone. Dept of Environmental Affairs and Tourism, 1-9

Tegner M.J. (1989) The feasibility of enhancing red sea urchin, *Strongylocentrotus franciscanus*, stocks in California: an analysis of the options. *Marine Fisheries Review*, 51: 1-22

Thomas M., Möller F., Wiehe F. and Tautz D. (2007) A pooling approach to detect signatures of selective sweeps in genome scans using microsatellites. *Molecular Ecology Notes*, 7: 400-403

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

Van Oosterhout C., Weetman D. and Hutchinson W.F. (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology*, 6: 255-256

van Schalkwyk M. (2007) Marine Living Resources Act (18/1998): Policy for the development of a sustainable marine aquaculture sector in South Africa. *Government Gazette Staatskoerant*, No 30263, 507: 1-20, 7 September, Pretoria.

Villanueva B., Verspoor E. and Visscher P.M. (2002) Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Animal Genetics*, 33: 33-41

Walsh P.S., Metzger D.A. and Higuchi R. (1991) Chelex®100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10: 506-513

Wang J. (2004) Sibship reconstruction from genetic data with typing errors. *Genetics*, 166: 1963-1979

Weber J.L. and Wang C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, 2: 1123-1128

Webster M.S. and Reichart L. (2005) Use of microsatellites for parentage and kinship analyses in animals. *Methods in Enzymology*, 395: 222-238

Wesmajervi M.S., Westgaard J.I. and Delghandi M. (2006) Evaluation of a novel pentaplex microsatellite marker system for paternity studies in Atlantic cod (*Gadus morhua* L.). *Aquaculture Research*, 37: 1195-1201

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

Zilberman N., Reikhav S., Hulata G. and Ron M. (2005) High-throughput genomic DNA extraction protocol from tilapia's fin tissue. *Aquaculture*, 255: 597-599

Addendum A

Inbreeding and Genetic Diversity

Abstract

A preliminary assessment of allelic richness within a wild population sourced from the east coast of South Africa, hatchery populations at Abagold and HIK and F1 progeny from both these farms, revealed no apparently significant declines in number of alleles at each of nine microsatellite loci investigated. Between Abagold and HIK, a slight difference in the mean number of alleles was observed with a higher frequency of null alleles evident in the Abagold population (5.3%) when compared to HIK (4.7%). No significance was established for any of these observed differences as this was essentially an initial view of allele differences and not a complete assessment of genetic diversity.

Introduction

Several studies emphasise the importance of maintaining genetic diversity within a confined breeding population (Ward *et al*. 2000; Bentsen & Olesen 2002; Appleyard & Ward 2006) when a selective breeding program is to be undertaken. Thus despite this not being an objective of this study, the output from the parentage analyses provided the means for a basic comparison of the allelic richness at each locus and observed heterozygosity levels within the two hatchery populations investigated for each farm and a wild population sourced from the east coast of South Africa.

Keeping in mind the longer term objectives of the project and the need to maximise the results from the various individual studies in speeding the industry towards the implementation of MAS, further understanding of the inherent genetic diversity within the breeding populations may aid the reassessment of elements of the experimental design in an effort to take full advantage of the potential genetic gain to be realised from such an endeavour. This is by no means a replacement for a full population study, one which is currently underway within the research group and due for completion early 2008. However, the population study encompasses wild populations only and the results presented here in this exercise may prompt a more in depth study into the variation and relationships existing within and between the current hatchery populations and their next generation offspring as successive generations are established by means of walk-back selection.

The output from the parentage analyses in this study were used to make three comparisons of genetic variation: (1) A comparison of allele data for a wild population (Black Rock), Abagold hatchery animals and Abagold PRS F1 offspring; (2) A similar comparison was made for the HIK population; (3) The slight (although not significant) size advantage noted in both PRS growth groups for HIK F1 progeny when compared to Abagold offspring of the same generation prompted a third comparison of allelic richness; that between HIK and Abagold breeding populations.

It is important to note that the results presented in this section should by no means be considered a full analysis of genetic variance but are merely a comparison of the allelic richness and allele frequency observed for each locus within each population to ascertain any changes therein.

Literature Review

Genetic Diversity

Conflicting opinions exist regarding genetic diversity in wild versus hatchery stocks. While various studies (Cross & King 1983: as cited by Mgaya *et al*. 1995; Sbordoni *et al*. 1986: as cited by Mgaya *et al*. 1995; Verspoor 1988: as cited by Mgaya *et al*. 1995; Hedgecock & Sly 1990; Smith & Conroy 1992) have shown that a notable loss of genetic diversity succeeds hatchery culture within species of fish and molluscs; Gosling (1982: as cited by Mgaya *et al*. 1995) and Dillon and Manzi (1987: as cited by Mgaya *et al*. 1995), on the other hand, observed no significant differences between wild and hatchery populations.

More recent studies of marine molluscs have revealed loss of genetic variation in hatchery populations and mass selection lines when compared to wild populations (Evans *et al*. 2004; Li *et al*. 2004; Appleyard & Ward 2006). This loss in diversity was evident as a reduction in microsatellite allele numbers (35-62%) while heterozygosity was less affected, if at all. Loss of alleles as opposed to loss in heterozygosity, however, is considered a more meaningful measure of whether genetic variation is being preserved since heterozygosity is particularly insensitive to large genetic changes that may occur in cultivated aquaculture stocks within the first few generations following isolation (Hedgecock & Sly 1990). Non-amplifying or null alleles, genotype scoring errors, sampling effects and inbreeding may all offer suitable explanations for the small changes observed in Ho although are likely to be minor contributors (Appleyard & Ward 2006).

The genetic differences observed between hatchery and wild stocks can be ascribed to a variety of factors: degree of inbreeding, dissimilar larval survival rates under hatchery conditions and the extent of genetic drift in relation to the effective breeding population (Mgaya *et al*. 1995). These differences are likely to be most obvious between wild populations and hatchery stocks where both original breeding stock and successive generations are being used concurrently as breeders, without knowledge of underlying pedigrees; hence the risk of inbreeding is heightened.

Koehn *et al*. (1988: as cited in Evans *et al*. 2004) and Danzmann *et al*. (1989: as cited in Evans *et al*. 2004) show that these reductions in genetic diversity have a longer term detrimental effect on valuable commercial traits such as growth rate and fitness since it is not the loss of neutral microsatellite alleles that is the major concern but rather the mirrored loss of variation at coding regions within the genome (Evans *et al*. 2004).

Some practical suggestions for preserving genetic diversity within a closed breeding population have been made by several authors based on their observations. Smith and Conroy (1992) suggest a minimum of a 1:1 sex ratio as good spawning practice while Gosling (1982: as cited in Mgaya *et al*. 1995) expands on this theme, recommending a minimum of forty-five breeding individuals in a sex ratio of 1:1. Theory, in general, suggests that greater than 99% genetic variability can be upheld with an effective population size (Ne) for breeding stock of greater than fifty individuals (Allendorf & Ryman 1987: as cited in Mgaya *et al*. 1995; Gall 1987 as cited in Mgaya *et al*. 1995).

Relatedness and Inbreeding

Inbreeding has been reported in some molluscs (McGoldrick & Hedgecock 1997) and hence the effect of inbreeding depression on growth, an essential component of fitness in molluscs, is due some careful consideration (Naciri-Graven *et al*. 2000). These authors proved that life-history traits are more prone to experience considerable inbreeding depression in comparison to morphological traits.

Increased homozygosity within small, isolated breeding populations may arise as a consequence of genetic drift or chance mating of related individuals; which in turn could lead to reduction in mean phenotypic value of fitness associated traits (Naciri-Graven *et al*. 2000). Mating between relatives invariably produces offspring which are less viable, smaller or less fertile than the average population – the occurrence of which is known as inbreeding depression.

The high fecundity of abalone ultimately leads to the utilization of a limited number of breeders in the production of a large number of offspring, the genetic variability of which is directly related to the number of parents used, unbalanced contributions of parents to spawning and relatedness of breeders (Evans *et al*. 2004; Brown *et al*. 2005). Since some of these offspring are invariably used as parents for the next generation when incorporated into selective breeding programs, inbreeding is likely to occur resulting in decreased performance (Hedgecock *et al*. 1995; Boudry *et al*. 2002). This reduced genetic variability limits the potential of future genetic improvement programs when employing selective breeding (Boudry *et al*. 2002).

Falconer (1989) and Lynch and Walsh (1998) claim that this negative effect occurs when inbreeding increases the chances of an individual being (1) homozygous for a segregating, harmful recessive allele or (2) homozygous at loci showing over dominance for traits having reduced fitness value (as cited in Slate *et al*. 2004). Estimation of these inbreeding levels within closed hatchery populations is crucial since these populations tend to be small and consequently inbreeding may affect the response to artificial selection pressure and ultimately jeopardize the future of a MAS breeding program (Naciri-Graven *et al*. 2000).

Genetic and Environmental Effects on Relatives

Very often, relatives exhibit similar phenotypic characteristics but this physical resemblance may be due as much to the effects of a shared environment as to the effects of hereditary genetic composition. Quantifying the environmental influences is fundamental to an understanding of evolutionary response to selection and for quantitative phenotypic traits, estimating the additive genotypic elements or, in other words, the heritability of a trait, is usually done from the starting point of physical similarity between related individuals (Kruuk & Hadfield 2007).

Environmental effects are found to be most confounding and ubiquitous within a wild population where the environment is unmanageable; but within a controlled, properly designed experiment, the scale of the environmental impacts can be minimized (Kruuk & Hadfield 2007). Garant and Kruuk (2005: as cited in Kruuk & Hadfield 2007) further emphasise that uncontrolled environmental effects will impact on the estimation of heritability when using marker based methods, an important consideration when evaluating the outputs of marker assisted selection.

Materials and Methods

No additional laboratory work was completed for this exercise as existing genotypic data was available for the four populations involved in the original parentage assignment study (Chapter Five). In addition, allele frequency data was available for a wild population from Black Rock (2003) which was used when screening the molecular markers for suitability for parentage analyses (Chapter Four).

The program CERVUS 2.0 was employed to calculate the following statistics for each locus within each population: allele number (Na), observed heterozygosity (Ho), expected heterozygosity (He), PIC, exclusion probabilities, HWE and null allele frequency (Null Freq).

For both Abagold and HIK data comparisons were then constructed between the wild population from Black Rock, the current hatchery population and the PRS F1 generation. The slight, although not significant, size advantage (weight and shell length) noted in both PRS growth groups for HIK F1 progeny when compared to Abagold offspring of the same generation prompted a third comparison of allelic richness between HIK and Abagold breeding populations. These small growth advantages may be as a result of environmental effects or a combination of environmental influences and inherited genetic benefits. Significance levels of any changes observed were not established given that insufficient time was available to conduct any further analyses.

Results

Allelic Richness in Wild, Abagold Hatchery and Next Generation Farm Populations

Domesticated hatchery populations may show less genetic diversity to wild populations and evidence of this reduced variability may be a lower number of alleles per locus and/or lower expected heterozygosity (Li *et al*. 2004). A comparison was made between allele frequency data from a wild population from the east coast of South Africa, hatchery stock at Abagold and the first generation progeny produced as Phase I of the PRS and from whence, based on the performance of these individuals, the breeding of specific family lines will commence through a walk-back selection process. The mean data values of these three populations are shown in Table 23 while more comprehensive data for each locus within each of the populations is provided in Appendix A.

Table 23 Comparison of allele frequency data from a wild abalone population, hatchery breeders from Abagold and F1 progeny spawned through mass spawning practises for the PRS

Mean Values	Wild Abalone Population	Hatchery Population	F1 Generation Offspring
n	32	98	200
Na	16.889	20.444	19.778
Ho	0.816	0.706	0.670
He	0.872	0.799	0.801
PIC	0.856	0.782	0.787
Null Freq	0.026	0.053	0.096

Key

N = number of individuals examined in each population

Na = mean number of alleles/locus; Ho = observed heterozygosity

He = expected heterozygosity, PIC = polymorphic information content

Null Freq = mean null allele frequency

Na values highlighted in green show that the average number of alleles per locus was higher in the two hatchery populations than in the wild population; values indicated in red for Ho and PIC show where mean observed heterozygosity levels and PIC of the marker suite was lower than for the wild population; while for null allele frequency, higher values (highlighted in red) for hatchery stock and the F1 generation indicate that the estimated average number of null alleles segregating with a locus has increased quite significantly within these two populations.

Allelic Richness in Wild, HIK Hatchery and Next Generation Farm Populations

Given the potential for inbreeding to occur and hence loss of genetic variation within subsequent generations when implementing mass spawning practices together with mass selection programs, this occurrence needs to be assessed and controlled. The mean allele number per locus and He within the populations under investigation for HIK, namely the breeders and F1 progeny, were compared to the data obtained for the wild abalone population sourced from Black Rock. These results are shown in Table 24 with details for each locus specifically being provided in Appendix B.

Table 24 Comparison of mean allele frequency data for a wild abalone population, hatchery breeders for HIK and F1 progeny produced for the PRS

Mean Values	Wild Abalone Population	Hatchery Population	F1 Generation Offspring
n	32	100	200
Na	16.889	21.778	21.111
Ho	0.816	**0.761**	**0.670**
He	0.872	0.841	0.816
PIC	0.856	0.826	0.803
Null Freq	0.026	**0.047**	**0.104**

Key

N = number of individuals examined in each population

Na = mean number of alleles/locus; Ho = observed heterozygosity

He = expected heterozygosity, PIC = polymorphic information content

Null Freq = mean null allele frequency

Similarly for HIK, values for Na indicated in green reveal a higher average number of alleles per locus. Mean Ho values have decreased for hatchery and F1 generation offspring (highlighted in red), possibly as a result of the higher estimated frequency of null alleles within these populations (highlighted in red).

Allelic Richness Present in Abagold and HIK Hatchery Populations

When juveniles were assessed for fast or slow growth, it was observed that, across the five rearing locations, HIK offspring appeared to perform better; having, on average, higher weights and shell lengths at one year of age when compared to Abagold offspring at these same locations. This prompted a comparison between the hatchery breeding stocks at these two farms to ascertain whether there might be any underlying difference in overall genetic diversity and whether this could be rectified by supplementing current breeding stocks with new breeders for which a high degree of genetic diversity was pre-determined.

Table 25 highlights the results of a comparison between these two breeding populations with regards to allele number (Na), observed heterozygosity (Ho) and null allele frequency for each of the nine loci tested. Figures highlighted in green indicate where the value was the higher of the two values compared (except in the instance of null allele frequency where a lower value is considered preferably); and blue indicates where the two values were equivalent.

Table 25 Comparison between Abagold and HIK hatchery populations with respect to locus specific allele number (Na), observed heterozygosity (Ho), expected heterozygosity (He), Polymorphic Information Content (PIC) and null allele frequency (Null freq)

HIK HATCHERY BREEDERS					
Locus	Na	Ho	He	PIC	Null freq
HmNR120	24	0.812	0.940	0.932	0.0703
HmNR136	25	0.717	0.687	0.672	-0.0370
HmNR281	33	0.703	0.946	0.939	0.1447
HmNR54	12	0.729	0.760	0.722	0.0181
HmNS28	27	0.891	0.939	0.930	0.0236
HmNS38	14	0.667	0.833	0.813	0.1092
HmNS56	16	0.525	0.610	0.592	0.0714
HmNS58	20	0.881	0.912	0.900	0.0135
HmNST7	25	0.920	0.943	0.935	0.0098
Mean values:	**21.8**	**0.761**	**0.841**	**0.826**	**0.0471**

ABAGOLD HATCHERY BREEDERS					
Locus	Na	Ho	He	PIC	Null freq
HmNR120	26	0.859	0.952	0.945	0.0495
HmNR136	19	0.505	0.514	0.498	-0.0106
HmNR281	27	0.691	0.950	0.942	0.1556
HmNR54	9	0.722	0.710	0.662	-0.0068
HmNS28	27	0.726	0.932	0.922	0.1237
HmNS38	12	0.67	0.822	0.797	0.0983
HmNS56	19	0.443	0.468	0.452	0.0195
HmNS58	19	0.854	0.908	0.896	0.0272
HmNST7	26	0.88	0.937	0.928	0.0250
Mean values:	**20.4**	**0.706**	**0.799**	**0.782**	**0.0535**

Higher value of the two corresponding values compared for a locus

Equivalent values

Discussion

As has already been touched upon, an indication of differences in genetic diversity would be detectable as a shift in the frequency of common alleles, in other words, seen as a change in the allele number at microsatellite loci and to a lesser extent, a reduction in observed mean heterozygosity (Smith & Conroy 1992; Evans *et al*. 2004; Li *et al*. 2004).

All populations in this study were analysed using the same microsatellite marker suite of nine loci (refer Chapter Five Table 17). Results for Abagold indicate that the hatchery population shows even greater genetic diversity than the wild population at Black Rock, evident in the mean number of alleles per locus (Na) of 20.4 versus the 16.9 present in the latter population (see: Page 166 Table 23) although this may be a factor of the smaller sample size of the Black Rock population examined. Mean Ho is lower for the hatchery population but according to Appleyard and Ward (2006), heterozygosity is less reflective of a change in diversity since non-amplifying alleles, genotyping errors, sampling effects and inbreeding may all be making small contributions to the change observed in this value. The higher null allele frequency observed in the hatchery population provides support to this statement. The inherent diversity in the hatchery population is high and appears to well reflect the diversity of the wild population.

The F1 generation although comparable, is showing a slight loss in genetic diversity evident by the marginally lower mean Na value of 19.8. This drop in Na, however, does not appear to be significant, possibly given the high degree of genetic diversity in the founder population. However, any changes in allele frequency should be monitored in successive generations and managed to prevent a loss of variation becoming a barrier to realising the full potential of genetic gain possible through MAS. PIC values of the marker suite for the hatchery stock and F1 progeny are comparable while being slightly lower than for the wild population. This could be due to a higher frequency of null alleles segregating within the former populations and a higher degree of genotyping errors given the larger data sets being processed.

Four loci differed significantly (at the 1% level) from HWE within the F1 generation. CERVUS 2.0 did not calculated HWE for several loci within both the hatchery and F1 progeny groups due to insufficient data being available for the program to carry out the analysis; therefore, direct comparisons were not possible. However, the fact that multiple loci were out of HWE in the first generation suggests potential substructure in the form of related individuals among breeders or the possible presence of inbreeding (Dong *et al*. 2006). The higher frequency of null alleles segregating in this generation may also have impacted on HWE (approximately 5% for hatchery animals and 10% for F1 progeny).

Within HIK populations, the hatchery population and the F1 population exhibit greater genetic variation, evident by the mean number of alleles per locus of 21.8 and 21.1 respectively (see: Page 167 Table 24), when compared to the Na value for the wild population of 16.8, again this may be due to the smaller sample size of the wild population examined. These populations appear to have remained well representative of the natural population.

Both the HIK and Abagold hatchery populations were harvested from the west coast of South Africa in the vicinity of Hermanus, while the Black Rock wild population was sourced from along the east coast close to Port Elizabeth. This may explain the slight difference in inherent genetic diversity within these two populations, given the different environmental conditions under which they lived; however, this difference is not statistically significant. A decline in Ho is seen across the populations which may possibly be as a result of the increasing null allele frequency from 2% (wild population) up to 10% (F1 population). Two loci were significantly out of HWE and similar to Abagold, these findings could either be as a result of related individuals and/or inbreeding or due to the higher frequency of null alleles (5% for hatchery population and 10% for F1 population).

In a study by Ward *et al*. (2000) on farmed (hatchery), naturalised and endemic Pacific oyster populations using seventeen allozymes markers, no significant differences in variation were found. In contrast, Evans *et al*. (2004) in a study of genetic diversity in hatchery and wild populations of *H. midae* and *H. rubra* using three microsatellite markers, revealed loss of genetic diversity within both populations but without associated loss in overall heterozygosity. In the latter study, hatchery and first generation populations were investigated. A later study on Pacific oyster populations (four successive mass selection lines) was conducted by Appleyard and Ward (2006) and encompassed both allozymes and microsatellite loci. Diversity at microsatellite loci was found to decrease across successive mass selected generations while no loss was observed for allozyme loci. Heterozygosity values remained constant across generations for both these marker types. Li *et al*. (2004) in a study on *H. discus hannai* strains compared to wild populations, observed similar results to those of Appleyard and Ward (2006).

All the mentioned studies encompassed a more holistic approach to assessing genetic diversity taking cognisance of: relatedness and estimations of inbreeding, estimations of effective population sizes, potential bottlenecks, pair-wise FST tests and analysis of variance of allele frequencies to name a few of the investigations performed.

In the comparison between Abagold and HIK hatchery populations, results indicate that for the majority of the microsatellite loci (five out of nine loci), HIK breeders revealed greater genetic diversity leading to a slightly higher mean allele number per locus. The higher mean null allele frequency observed in the Abagold population may have contributed to the slightly lower Ho values

in this population when compared to HIK. None of these differences, however, are statistically significant.

It may be discovered for the Abagold population, upon comparison at the individual level, that only a few breeders lack sufficient genetic variation, which in turn affects the average values. Should these individuals be shown to possess significantly lower diversity levels, consideration should then be given to replacing these breeders with alternative adults for which a higher level of variation has been established. This is particularly relevant should these individuals be found to be contributing appreciably to spawning as their continued contribution may result in reduced genetic diversity in future generations in the long term.

The results from the parentage assignment study also highlighted the need to remove some of the current non-value adding breeders from the hatcheries at both farms. The option exists to potentially replace the breeders that are currently adding no value to spawning with new wild stock prior to the commencement of controlled family-breeding, in an effort to maintain the effective breeding population. Alternatively, managing effective population size (Ne) through the pooling of larvae from smaller spawning groups, as recommended by Gaffney *et al*. (1992: as cited in Mgaya *et al*. 1995), is another basic approach and is the better choice to sustaining genetic variation within a closed population.

The recruitment of wild breeding stocks to boost hatchery stocks will, however, cause an associated loss in accumulated genetic gain which is to be avoided when a selective breeding program is already in progress (Brown *et al*. 2005). In saying that though, loss of genetic variation is a cumulative process, and the effects thereof are more often seen in the later stages of a breeding scheme while the consequences are less likely to be observed in shorter term selection experiments (Bentsen & Olesen 2002).

Family-based selective breeding has not yet been established within the South African abalone industry, so replenishment of abalone breeders within the hatcheries is still worth consideration. However, conditioning takes time and since the performance of these animals will not have been evaluated through parentage analyses, they will need to be individually monitored for a visual assessment of their potential, preferably in 1:1 spawning practices. Since selection programs are invariably established for long-term genetic gain, loss of genetic variation within the hatchery stocks need to be effectively managed so as to minimise the potential impact of this process coming to the fore just when genetic gains are beginning to be realised within the population.

Appendix A Allele Frequency Data for a Black Rock Wild Population and Two Abagold Hatchery Populations

Comparison of allelic richness and heterozygosity at each locus as observed for Black Rock wild population, Abagold hatchery breeders and F1 generation juveniles

BLACK ROCK WILD ABALONE POPULATION

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	32	24	0.897	0.943	NA	NP	NP	NS	0.024
HmNR136	31	20	0.750	0.797	NA	NP	NP	NS	0.026
HmNR281	31	21	0.714	0.919	NA	NP	NP	*	0.107
HmNR54	32	10	0.781	0.757	NA	NP	NP	NS	-0.014
HmNS28	32	19	0.906	0.943	0.924	0.747	0.855	NA	0.012
HmNS38	31	11	0.742	0.833	0.803	0.495	0.667	NA	0.046
HmNS56	31	16	0.839	0.854	0.826	0.536	0.700	NA	0.000
HmNS58	31	10	0.781	0.856	0.803	0.520	0.688	NA	0.032
HmNST7	31	21	0.935	0.945	0.926	0.754	0.859	NA	-0.003
Mean values:	**31.3**	**16.9**	**0.816**	**0.872**	**0.856**	**0.610**	**0.754**		**0.026**

HATCHERY BREEDERS

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	92	26	0.859	0.952	0.945	0.807	0.893	NA	0.050
HmNR136	95	19	0.505	0.514	0.498	0.159	0.339	NS	-0.011
HmNR281	94	27	0.691	0.95	0.942	0.801	0.889	NA	0.156
HmNR54	90	9	0.722	0.71	0.662	0.299	0.473	NS	-0.007
HmNS28	95	27	0.726	0.932	0.922	0.744	0.853	NA	0.124
HmNS38	91	12	0.67	0.822	0.797	0.479	0.652	NS	0.098
HmNS56	97	19	0.443	0.468	0.452	0.128	0.298	NS	0.020
HmNS58	96	19	0.854	0.908	0.896	0.676	0.807	NA	0.027
HmNST7	92	26	0.88	0.937	0.928	0.762	0.865	NA	0.025
Mean values:	**93.6**	**20.4**	**0.706**	**0.799**	**0.782**	**0.539**	**0.674**		**0.053**

F1 GENERATION JUVENILES

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	198	26	0.833	0.939	0.933	0.775	0.873	NA	0.059
HmNR136	200	17	0.415	0.480	0.469	0.139	0.315	**	0.097
HmNR281	200	28	0.740	0.942	0.936	0.784	0.878	NA	0.119
HmNR54	197	10	0.736	0.741	0.703	0.345	0.524	NS	-0.006
HmNS28	192	25	0.724	0.923	0.916	0.725	0.841	NA	0.120
HmNS38	197	11	0.624	0.840	0.819	0.515	0.683	**	0.145
HmNS56	200	18	0.390	0.520	0.504	0.163	0.343	**	0.181
HmNS58	195	19	0.713	0.892	0.880	0.640	0.781	**	0.110
HmNST7	199	24	0.859	0.930	0.923	0.747	0.855	NA	0.038
Mean values:	**197.6**	**19.8**	**0.670**	**0.801**	**0.787**	**0.537**	**0.677**		**0.096**

Key

n	Number of individuals tested in the population
NS	No significant deviation from Hardy Weinberg equilibrium
NA	Hardy Weinberg not calculated
**	Indicates significant deviation from Hardy Weinberg equilbrium
Na	Number of alleles segregating at a locus
Ho	Observed Heterozyosity

He	Expected Heterozygosity
PIC	Polymorphic Information Content
Excl1	Exclusion Probability (First Parent)
Excl2	Exclusion Probability (Both Parents)
HWE	Hardy Weinberg Equilibrium
Null Freq	Null Allele Frequency

Appendix B Allele Frequency Data for a Black Rock Wild Population and Two HIK Hatchery Populations

Comparison of mean allele frequency data for each microsatellite loci within the Black Rock wild population, HIK hatchery breeders and F1 generation juveniles

BLACK ROCK WILD ABALONE POPULATION

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	32	24	0.897	0.943	NA	NP	NP	NS	0.024
HmNR136	31	20	0.750	0.797	NA	NP	NP	NS	0.026
HmNR281	31	21	0.714	0.919	NA	NP	NP	*	0.107
HmNR54	32	10	0.781	0.757	NA	NP	NP	NS	-0.014
HmNS28	32	19	0.906	0.943	0.924	0.747	0.855	NA	0.012
HmNS38	31	11	0.742	0.833	0.803	0.495	0.667	NA	0.046
HmNS56	31	16	0.839	0.854	0.826	0.536	0.700	NA	0.000
HmNS58	31	10	0.781	0.856	0.803	0.520	0.688	NA	0.032
HmNST7	31	21	0.935	0.945	0.926	0.754	0.859	NA	-0.003
Mean values:	**31.3**	**16.9**	**0.816**	**0.872**	**0.856**	**0.610**	**0.754**		**0.026**

HATCHERY BREEDERS

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	101	24	0.812	0.940	0.932	0.770	0.870	NA	0.070
HmNR136	99	25	0.717	0.687	0.672	0.323	0.518	NS	-0.037
HmNR281	101	33	0.703	0.946	0.939	0.791	0.883	NA	0.145
HmNR54	96	12	0.729	0.760	0.722	0.372	0.550	NS	0.018
HmNS28	101	27	0.891	0.939	0.930	0.765	0.867	NA	0.024
HmNS38	99	14	0.667	0.833	0.813	0.509	0.679	**	0.109
HmNS56	99	16	0.525	0.610	0.592	0.237	0.429	NS	0.071
HmNS58	101	20	0.881	0.912	0.900	0.685	0.813	NA	0.014
HmNST7	100	25	0.920	0.943	0.935	0.779	0.875	NA	0.010
Mean values:	**99.7**	**21.8**	**0.761**	**0.841**	**0.826**	**0.581**	**0.720**		**0.047**

F1 GENERATION JUVENILES

Locus	n	Na	Ho	He	PIC	Excl1	Excl2	HWE	Null freq
HmNR120	198	26	0.788	0.936	0.929	0.765	0.866	NA	0.086
HmNR136	198	18	0.500	0.596	0.581	0.227	0.420	**	0.093
HmNR281	200	31	0.705	0.937	0.931	0.769	0.869	NA	0.139
HmNR54	197	11	0.680	0.774	0.742	0.396	0.575	NS	0.061
HmNS28	192	24	0.682	0.939	0.933	0.773	0.872	NA	0.158
HmNS38	198	14	0.626	0.764	0.742	0.403	0.587	NS	0.094
HmNS56	199	18	0.407	0.571	0.555	0.203	0.392	**	0.194
HmNS58	196	20	0.765	0.892	0.881	0.644	0.783	NS	0.075
HmNST7	197	28	0.873	0.937	0.930	0.765	0.867	NA	0.033
Mean values:	**197.2**	**21.1**	**0.670**	**0.816**	**0.803**	**0.549**	**0.692**		**0.104**

Key

n	Number of individuals tested in the population
NS	No significant deviation from Hardy Weinberg equilibrium
NA	Hardy Weinberg not calculated
**	Indicates significant deviation from Hardy Weinberg equilbrium
Na	Number of alleles segregating at a locus
Ho	Observed Heterozyosity

He	Expected Heterozygosity
PIC	Polymorphic Information Content
Excl1	Exclusion Probability (First Parent)
Excl2	Exclusion Probability (Both Parents)
HWE	Hardy Weinberg Equilibrium
Null Freq	Null Allele Frequency

References

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

Bentsen H.B. and Olesen I. (2002) Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture*, 204: 349-359

Boudry P., Collet B., Cornette F., Hervouet V. and Bonhomme F. (2002) High variance in reproductive success of Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture*, 204: 283-296

Brown R.C., Woolliams J.A. and McAndrew B.J. (2005) Factors influencing population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture*, 247: 219-225

Dong S., Kong J., Zhang T., Meng X. and Wang R. (2006) Parentage determination of Chinese shrimp (*Fenneropenaeus chinensis*) based on microsatellite DNA markers. *Aquaculture*, 258: 283-288

Evans B., Bertlett J., Sweijd N., Cook P. and Elliot N.G. (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

Hedgecock D. and Sly F. (1990) Genetic drift and effective population sizes of hatchery-propagated stocks of Pacific oyster, *Crassostrea gigas*. *Aquaculture*, 88: 21-28

Hegdecock D., McGoldrick D.J. and Bayne B.L. (1995) Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines. *Aquaculture*, 137: 285-298

Kruuk L.E.B. and Hadfield J.D. (2007) How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, pre-publication

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

McGoldrick D. and Hedgecock D. (1997) Fixation, segregation and linkage of allozyme loci in inbreed families of Pacific oyster *Crassostrea gigas* (Thunberg): implication for the cause of inbreeding depression. *Genetics*, 146: 321-334

Mgaya Y.D., Gosling.E.M., Mercer J.P. and Donlon J. (1995) Genetic variation at three polymorphic loci in wild and hatchery stocks of abalone, *Haliotis tuberculata* Linnaeus. *Aquaculture*, 136: 71-80

Naciri-Graven Y., Launey S., Lebayon N., Gerard A. and Baud J.P. (2000) Influence of parentage upon growth in *Ostrea edulis*: evidence for inbreeding depression. *Genetic Research*, 76: 159-168

Slate J., David P., Dodds.K.G., Veenvliet B.A., Glass B.C., Broad T.E. and McEwan J.C. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Journal of Heredity*, 93: 255-265

Smith P.J. and Conroy A.M. (1992) Loss of genetic variation in hatchery produced abalone, *Haliotis iris*. *New Zealand Journal of Marine and Freshwater Research*, 26: 81-85

Ward R.D., English L.J., McGoldrick D.J., Maguire G.B., Nell J.A. and Thompson P.A. (2000) Genetic improvement of the Pacific oyster *Crassostrea gigas* (Thunberg) in Australia. *Aquaculture Research*, 31: 35-44