An integrated linkage map of Perlemoen (*Haliotis midae*)

by
Juli-ann Hepple

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

Supervisor: Dr Rouvay Roodt-Wilding
Faculty of Science
Department of Genetics

December 2010

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2010

Summary

Haliotis midae, or Perlemoen, is the only cultured species of abalone in South Africa and is under great international demand. This species is considered endangered, making sustainable farming practises and law enforcement against poaching essential for maintaining wild stocks. A limited amount of broodstock animals are provided to each farm from which thousands of offspring are grown and exported. The prevention of inbreeding and preservation of genetic diversity within farmed stocks is necessary for future sustainable farming and production of genetically stable offspring. Further research into the genetic dynamics of Perlemoen will provide the knowledge for advanced management programs for optimal farming practises and essentially sustainable production. This study focuses on genetic linkage map development with the intention of future identification of markers associated with genes of economic importance, such as growth rate. Identification of markers linked to genes responsible for such phenotypic traits will ultimately allow farming practises to select naturally genetically superior animals for breeding, thereby enhancing production.

For the construction of a genetic linkage map of *H. midae*, microsatellite markers were developed using two strategies: FIASCO and screening of next generation sequence-by-synthesis contig data. The FIASCO-derived markers were characterised by genotype screening in 32 individuals from a full-sib family and analysed using Mendelian segregation expectations. The Illumina-derived markers were characterised by genotype screening in 32 individuals from wild populations and analysed against Hardy-Weinberg expectations. Forty four microsatellite-family combinations were obtained from FIASCO of which 28 provided informative genotype results (32% success). Twenty two markers were developed from sequence-by-synthesis screening. Fourteen provided reliable genotypes (37%) and six conformed to Hardy-Weinberg expectations.

These markers were used, in addition to 156 previously developed markers, to develop sex-specific and sex-average linkage maps in two full-sib families consisting of approximately 100 offspring each. One hundred and six polymorphic loci were used for linkage analysis (LOD>3) in both families. The number of linkage groups obtained from sex-specific maps ranged from 13-16. The average genome length ranged from 500 cM to 800 cM with an average marker spacing of 10 cM. The sex-average linkage map provided

18 linkage groups with an average genome length calculation of 1800 cM and average marker spacing of approximately 13 cM.

The linkage maps created in this study are preliminary but provide a stepping stone towards a high density map incorporating high throughput markers. This also provides a base for QTL mapping studies, in which phenotypic traits of interest can be identified and associated to specific locations in the *H. midae* genome for marker-assisted selection.

Opsomming

Haliotis midae, ook bekend as Perlemoen, is in groot internasionale aanvraag en is ook die enigste klipkous spesie waarmee in Suid Afrika geboer word. Hierdie spesie word as bedreig beskou en daarom is volhoubare boerdery bedrywe en wetstoepassing teen stroping noodsaaklik om wilde populasies te beskerm. Elke perlemoenplaas word met ‘n beperkte aantal broeidiere verskaf, waarvan die nageslag dan gekweek en uitgevoer word. Voorkoming van inteling en handhawing van genetiese diversiteit binne gekweekte populasies is noodsaakllik vir toekomstige volhoubare kweking en produksie van ń geneties stabiele nageslag. Verdere ondersoeke na die genetiese dinamika van Perlemoen sal die nodige kennis verskaf om sodoende gevorderde bestuursprogramme te ontwikkel, wat tot optimale kweek praktyke en effektiewe volhoubare produksie sal lei. Hierdie studie fokus op die ontwikkeling van ‘n genetiese koppelingskaart met die voorneme om toekomstige merkers te identifiseer wat met gene van ekonomiese belang, soos byvoorbeeld groei tempo geassosieerd is. Identifisering van merkers wat vir sulke fenotipiese eienskappe verantwoordelik is sal sodoende toelaat dat boerdery praktyke kan selekteer vir diere vir verbeterde teling en produksie.

Mikrosatelliet merkers is ontwikkel om die genetiese koppelingskaart saam te stel. Die volgende twee strategieë is benut: FIASCO en sifting van volgende generasie volgordebepaling-deur-sintese “contig” data. Die FIASCO-afgeleide merkers is gekarakteriseer deur genotipiese sifting in 32 individue van ‘n volsib familie en is deur Mendeliese segregasie verwagtinge ge-analiseer. Die Illumina-afgeleide merkers is gekarakteriseer deur genotipiese sifting in 32 individue van wilde populasies en is met Hardy-Weinberg ewewig ge-analiseer. Vier en veertig mikrosatelliet-familie kombinasies is deur FIASCO verky, waarvan 28 informatiewe genotipiese resultate gelewer het (32% sukses). Twee en twintig merkers is vanaf volgordebepaling-deur-sintese sifting ontwikkel. Veertien van hierdie merkers het betroubare genotipes (37%) verskaf en ses het aan Hardy-Weinberg verwagtinge voldoen.

Hierbenewens is 156 voorheen ontwikkelde merkers gebruik om geslagspesifieke en geslagsgemiddelde koppelingskaarte in twee volsib families saam te stel. Hierdie volsib families het uit ń naslag van 100 elk bestaan. Een honderd en ses polimorfiese lokusse is vir koppelingsanalise gebruik, waar ‘n LOD waarde groter as drie statisties betekenisvol geag was. Die aantal koppelingsgroepe verkry van geslagspesifieke kaarte het tussen 13

en 16 gewissel. Die gemiddelde genoom lengte het van 500 cM tot 800 cM met ‘n gemiddelde merker spasiëring van 10 cM. Die geslagsgemiddelde koppelingskaart het 18 koppelingsgroepe gehad met ‘n gemiddelde genoom lengte berekening van 1800 cM en ‘n gemiddelde merker spasiëring van ongeveer 13 cM.

Die koppelingskaarte wat in hierdie studie geskep is, is voorlopig en verskaf ‘n grondslag vir die ontwikkeling van ‘n hoër digtheidskaart, wat hoë deurset merkers inkorporeer. Dit verskaf ook ‘n basis vir kwantitatiewe kenmerk lokus karteringstudies. Hierdie karteringstudies kan fenotipiese eienskappe van belang identifiseer en assosieer met spesifieke posisies binne die *H. midae* genoom vir merker bemiddelde seleksie.

Acknowledgements

I would like to thank the following institutions: DNA Sequencing Facility, HIK Abalone Farm (Pty) Ltd., Innovation Fund, Roman Bay Sea Farm (Pty) Ltd. and Stellenbosch University. I must also sincerely thank the following people: Rouvay Roodt-Wilding for all her encouragement, patience and guidance provided throughout my time in the aqua lab, Aletta van der Merwe for always providing us with the supplies we need and Danie Brink for running the aquatic research team in Stellenbosch. I would like to thank all my fellow labrats in the Molecular aquatic research group (MARG), especially Lise, Clint and Ruhan Slabbert, especially Ruhan for all his patience and guidance. I would also like to acknowledge Prof. Johan Burger as well as all the people involved at the Department of Genetics for providing me with a home-away-from-home for the past 4 years. Michael-John, Veronique, Nats and Marika thanks for the strange, fun and random chats in the passage. I have made some amazing friends with some unforgettable memories. Rynhard, Cobus, Etienne, Ihan and Gerald you guys are amazing friends, thanks for all the laughs and coffee breaks. Arnold, Kate, Jess and Lise thank you for keeping me sane, or should I say insane, and for slapping me back to reality when science blurred the boundaries, you are friends for life. Kyle, Dale, Hercus, Bjorn and all my friends and family back home, you guys have been so supportive and I am so grateful. Ling, Julian and Tracey your support was everything I needed, thank you. Last but far from least dad and mom, your love, encouragement and support has made this thesis possible, I am eternally grateful!

Table of contents

Chapter one – Literature review

Table of contents

Chapter two – Microsatellite marker development

Table of contents

List of Figures

List of tables

List of abbreviations

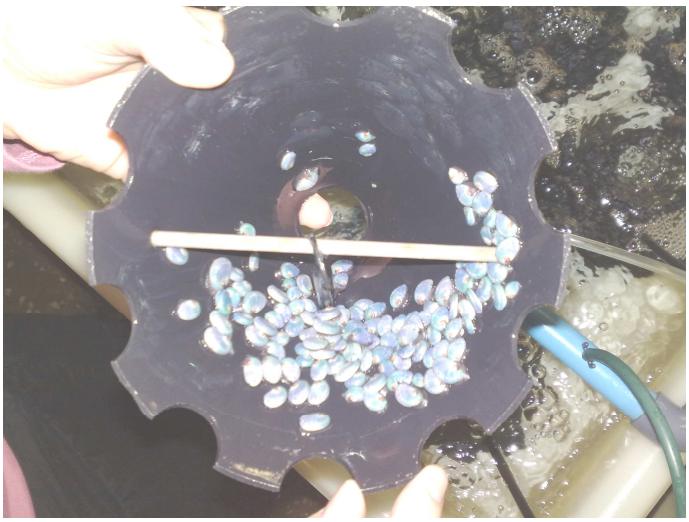
$	United States of America dollar
%	Percentage
(Pty) Ltd	Property limited
<	Less than
>	Greater than
®	Registered trademark
µg/ml	Micrograms per millilitre
µl	Microlitre
µM	Micromolar
3’	Three prime
5’	Five prime
A	Adenine
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
^{o}C	Degrees Celsius
cDNA	Complimentary DNA
CIM	Composite interval mapping
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
cM	CentiMorgan
CTAB	Cetyltrimethylammonium bromide
ddH_2O	Double distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediamine tetra-acetate ($C_{10}H_{16}N_2O_8$)
EST	Expressed sequence tag
FAO	Food and Agriculture Organisation of the United Nations
FIASCO	Fast isolation by AFLP of sequence containing repeats
G	Genome length
g	Grams
G	Guanine
gDNA	Genomic deoxyribonucleic acid
$G_{e\ ave}$	Estimated genome lengths’ average
G_e	Estimated genome length
G_o	Observed genome length

HWE	Hardy-Weinberg equilibrium
IAS	International Abalone Society
IBD	Identical by decent
kb	Kilobase pairs
KCl	Potassium chloride
LB	Luria-Bertani
LG	Linkage group
LOD	Logarithm of odds
LTP	Long term potential
m	Metre
M	Molar (Moles per Litre)
MALDI-TOF	Matrix-associated laser desorption ionisation-time of flight
MAS	Marker-assisted selection
mg/ml	Milligram per millilitre
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Millilitre
ML	Maximum likelihood
MLS	Minimal legal size
mm	Millimetre
mM	Millimolar
MML	Multipoint maximum likelihood
mRNA	Messenger ribonucleic acid
MtDNA	Mitochondrial deoxyribonucleic acid
NaAc	Sodium acetate
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanograms
ng/µl	Nanogram per microlitre
NGS	Next generation sequencing
NS	Non-stringency
p	Probability value (as a statistically significant limit)
pH	Concentration of hydrogen ions in a solution is expressed conventionally as its pH
PAGE	Poly-acrylamide gel electrophoresis
PCR	Polymerase chain reaction
PIC	Polymorphic information content
pmol	Picomole
pp.	Pages
PTP	Picotiter plate
QTL	Quantitative trait locus/loci
R	Rand (South African)

RAPD	Random amplified polymorphic DNA
RFID	Radio frequency identity
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Revolutions per minute
s	Marker spacing
S	Stringency
SDS	Sodium dodecyl sulfate
sec	Seconds
SIM	Simple interval mapping
SNP	Single nucleotide polymorphism
SSC	Standard saline citrate
ssDNA	Single-stranded deoxyribonucleic acid
SSR	Simple sequence repeat
STR	Short tandem repeat
T	Thymine
TAC	Total allowable catch
Taq	*Thermus aquaticus* DNA polymerase
TBE	Tris-Borate-EDTA buffer
TE	Tris-Ethylenediamine tetra-acetate
TEMED	N, N, N', N',-tetramethylenediamine
TEN	Tris- Ethylenediamine tetra-acetate, sodium chloride
TD-PCR	Touchdown polymerase chain reaction
T_m	Melting temperature
Tris-HCl	Tris- (hydroxymethyl) aminomethane hydrochloric acid
U	Units (enzyme)
U.S.A	United States of America
UV	Ultraviolet
v/v	Volume per volume
VNTR	Variable number of tandem repeat
w/v	Weight per volume

Chapter one

Literature review



1. *Haliotis midae* in general

1.1 Classification

Phylum: Mollusca
Class: Gastropoda
Super order: Vetigastropoda
Super family: Haliotoidea
Family: Haliotidae
Genus: *Haliotis*

(http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?lvl=0&id=36098 [accessed Jan 2010)

1.2 Structure

The protective shell of abalone is spirally shaped with a flattened anterior and ellipsoid perimeter. The outside of the shell tends to have a rough texture while the inner shell is more smooth and pearl-like. There is a row of holes found on the edge of the anterior known as tremata, which are essential for respiration and excretion (Fallu 1991; Landau 1992).

Abalone possess a muscular foot for attachment to rocky surfaces and for movement toward nourishment. This tissue is especially tough for protection against carnivores found in the ocean such as lobster, the main predator of adult Perlemoen. On the upper region of the abalone foot, a row of tentacles can be seen, which are used to sense predators (Tarr 1989; Fallu 1991).

1.3 Life cycle

Gametes of abalone species are spawned under appropriate conditions from female and male animals once or twice a year during the warmer months (Landau 1992). A small female releases approximately four million eggs per spawning and up to 15 million are released from the larger animals (Tarr 1989). This excessive egg release ensures the survival of many animals for successful proliferation and advancement of the species. The released eggs fuse with sperm and form a zygote approximately 0.2 mm in diameter which subsequently divides numerously until a trochophore larva is released from the egg (Figure 1.1). After a few days, a veliger larva develops which settles on the rocky sea bed for up to three weeks until spat stage (Tarr 1989; Fallu 1991; Landau 1992). The spat take approximately 8-10 years to develop from the juvenile stage to sexually mature adult and 30 years to a full adult size of approximately 200 mm in shell length (Sales and Britz 2001).

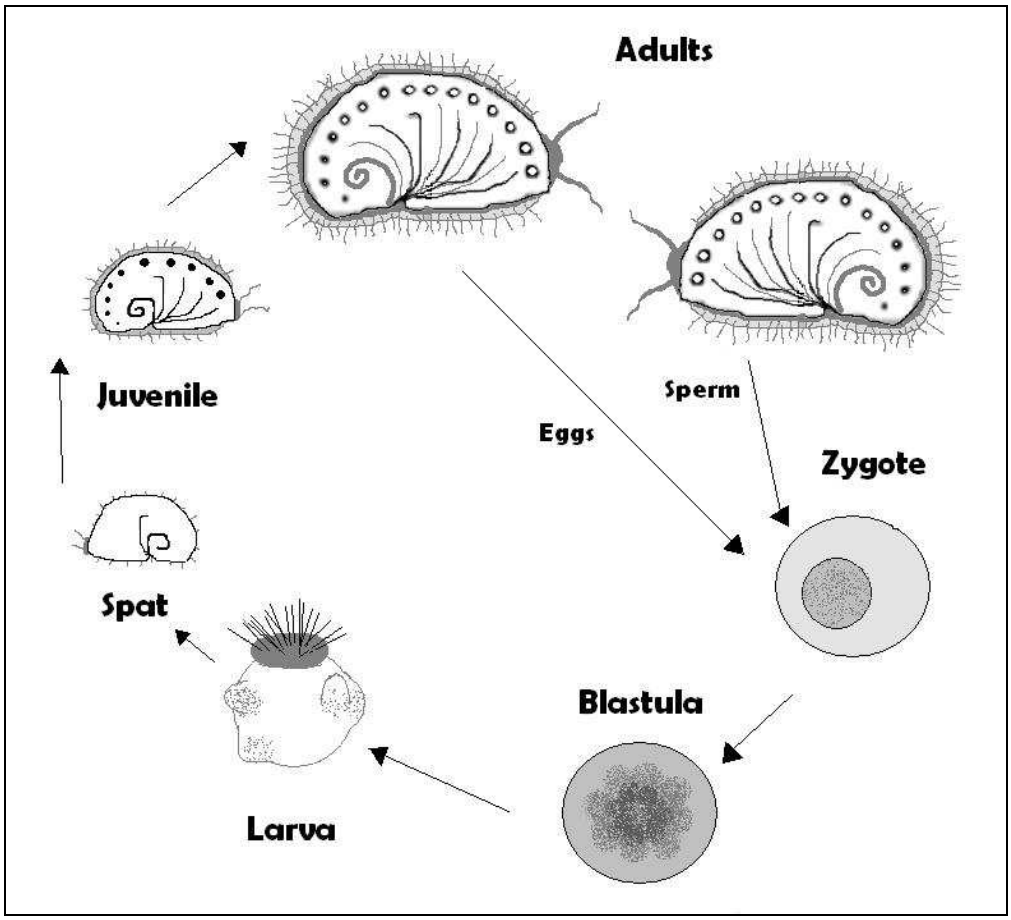


Figure 1.1: Abalone life cycle. The adult male and female animals release gametes which fuse to become a blastula. The larva grows and develops until settlement as a juvenile abalone before growth into a sexually mature animal [drawn by author].

1.4 Feeding

These herbivorous molluscs begin feeding on phytoplankton at the larval stage and progress to micro-algae and bacteria at spat stage, followed by seaweed grazing during adulthood. This change in feeding habits is a result of differentiation of mouth parts as they mature, allowing consumption of larger foods (Fallu 1991; Landau 1992).

1.5 Habitat and distribution

There are 56 known species of abalone worldwide forming part of the genus *Haliotis* (Geiger 2000), of which six are found on the Southern African coast. The southern coastal area of South Africa harbours the habitat of five native species of abalone. *Haliotis midae*, *H. spadicea* and *H. parva* are found along the West and East coasts with *H. spadicea* stretching up to Mozambique. *Haliotis queketti* and *H. alfredensis* (previously *H. speciosa*), however, are only found in the Eastern Cape region (Evans *et al*. 2004) (Figure 1.2). The sixth species, *H. pustulata*, is more prevalent on the upper East coast (Geiger 2000).

Haliotis midae has the second largest habitat range of the five species, found between the intertidal and subtidal zones up to a depth of 10m below sea level (Branch 2002).

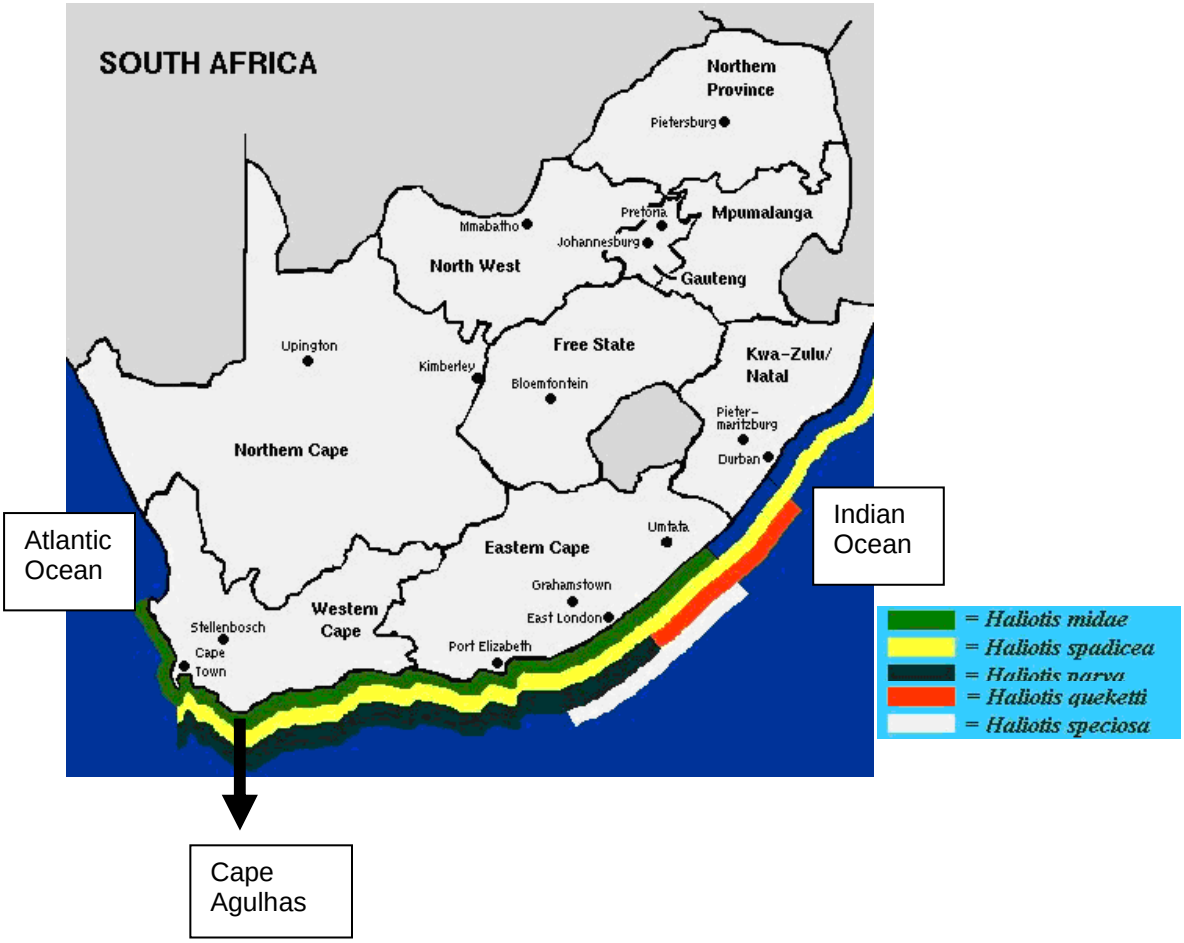


Figure 1.2: A map of South Africa displaying the regions in which *Haliotis* is found (http://web.uct.ac.za/depts/zoology/abnet/safrica.html [accessed Feb 2010]).

2. Exploitation of *Haliotis midae*

Different forms of abalone exploitation include recreational fishing, subsistence harvesting, commercial harvesting and illegal exploitation (Hauck and Sweijd 1999). Illegal exploitation or poaching, according to the Marine Living Resources Act of 1998, can include anything from harvesting an animal below the Minimum Legal Size (MLS) to harvesting more than the allowable catch as well as fishing in restricted areas.

Abalone harvesting for commercial as well as recreational activities was completely unregulated until 1970, with an annual catch of 2800 tonnes in 1965 (Tarr 1989). In 1995 the total allowable catch (TAC) was 615 tonnes and was reduced to 125 tonnes in the 2006/2007 season. In the 2007/2008 season the TAC reached a drastic low of 75 tonnes

before abalone harvesting in South Africa was completely prohibited from February 2008 by the Department of Environmental Affairs and Tourism (http://www.deat.gov.za/; “Court finds against the South African abalone industry association”, [online 2009]; Raemaekers and Britz 2009).

Despite this ban, massive crime syndicates continuously poach for the purpose of sales on the black market and continue to harvest this vulnerable species (http://www.deat.gov.za/; “Five alleged poachers in court, Vehicle and abalone confiscated after whole night stake-out” and “Two vehicles and over 3000 units of abalone confiscated, 2 arrests – Taiwanese vessel fined hefty fine of R1.5m” [accessed Aug 2009]).

The South African abalone was the first *Haliotis* species to be listed on ‘Convention on International Trade in Endangered Species of Wild Fauna and Flora’ Appendix III in May 2007 (CITES 2007; http://www.cites.org/eng/app/appendices.shtml [accessed Sept 2009]) as a means to reduce illegal exploitation and trade. The CITES documentation was therefore a prerequisite when trading in Perlemoen, thus reducing the black market sales to legitimate commercial institutions.

The ban on Perlemoen fishing and placement on the CITES endangered list was recently (1 July 2010) lifted, causing shock and utter disappointment by societies such as TRAFFIC (The wildlife trade monitoring network) international and the IAS (International Abalone Society). South Africa failed to implement the listing on CITES effectively by battling to ensure endorsement of CITES permits at ports of exit (http://www.internationalabalonesociety.org/africa_news.html; removal of trade controls signals bleak future for abalone [accessed Aug 2010]). This leaves the abalone industry in a difficult situation and calls for extremely well-planned trade management.

The necessity for improved crime-control and investigation is in great demand and should become a priority in fraudulent investigation units of the South African Police Service, especially considering the removal from the CITES list. The drastic plummet in abalone numbers, due to these abusive activities as well as disease and habitat loss, is one of the core drivers behind abalone aquaculture development in South Africa (Sales and Britz 2001). The development of successfully operative farmed supply of abalone to the

fisheries market will potentially allow wild stocks to recuperate with conservational input and effective law enforcement on illegal harvesting.

3. Abalone farming

Aquaculture is currently the fastest growing food-producing industry worldwide and provides approximately 45% of the worlds' marine food supply (Subasinghe *et al*. 2009). Aquaculture contribution to the world supply of fish, crustacean, mollusc and aquatic animals has improved radically, with an increase from 3.9% in 1970 to 36% in 2006 (FAO 2009). Many countries are involved in the aquaculture fisheries production industry of which China is by far the major producer of marine fish and shell fish, providing over 67% of the worlds' contribution to commercial aquatic animal products (Subasinghe *et al*. 2009). Approximately 51 million tonnes ($78.8 billion) worth of marine food was produced via aquaculture for human consumption in 2007, 27% of these included molluscs (FAO 2009).

Abalone are largely over exploited in all countries and the farming industry has seen TAC decreases in the majority of abalone fisheries, providing the incentive for strong management programs and conservation collaborations.

Although there are 56 known species of abalone not all are adequate for farming purposes due to size and growth rate limitations (Geiger 2000). To name a few economically renowned species: In Japan the commercial species are *H*. *discus*, *H*. *discus hannai*, *H*. *gigantea*, *H*. *sieboldii* and *H*. *diversicolor*; in Thailand *H*. *asinina*; in the U.S.A. *H*. *rufescens*, *H*. *corrugata* and *H*. *fulgens* are commercially farmed; in Europe *H*. *tuberculata* is of economic importance and in Australia *H*. *rubra* is a key species of commercial utilisation. *Haliotis midae*, the commercial species in South Africa, is locally referred to as Perlemoen or 'Paarlemoer' which is a name derived from the Dutch term 'mother-of-pearl' (Tarr 1989).

Abalone fisheries have existed in South Africa since 1949, however the first abalone hatchery rearing attempts were only developed in the 1980s (Sales and Britz 2001) afterwhich the first successful spawning and juvenile rearing took place in 1985 (Genade *et al*. 1988). Commercial abalone farms were established in 1990 with multiple collaborations leading to a burst in development within the industry. To date, there are 18

registered commercial abalone farms in South Africa (Britz and Lee 2009) which require continuous improvement of farming methods to remain competitive in this worldwide lucrative market. In South Africa, *H. midae* is the major saltwater aquaculture species and produced 508.14 tonnes worth R81,9 million out of a total produce (marine aquaculture species in South Africa) of 1052.19 tons worth R90,456 million in 2005 (Botes *et al*. 2006).

Haliotis midae is the largest of the five species of abalone found on the South African coast; hence it's commercial sought after status. The natural growth of *H*. *midae* to an adult size of 200 mm in 30 years in wild animals has been accelerated with aquaculture to a size of 100 mm in five years in hatchery reared animals (Sales and Britz 2001). Since many farmed abalone species are endangered, population structure as well as genetic history and management strategies must be investigated and inaugurated to enable sustainability and allow wild stocks to recover (Subasinghe *et al*. 2009).

4. Genetics and aquaculture

Genetic selection in aquaculture practices began over 2000 years ago in China by fish breeders. The domestication (adaptation of wild species to cultured species in idealistic environments) of fish species and selection of the "best looking" animals for breeding altered and affected gene frequencies and hence phenotypic traits within the cultured populations (FAO 2006). Farming practices and selective breeding has drastically improved since.

Mass selection for economically desirable traits has proven to be successful in several species such as salmonids (*Salmo salar*) (Gjoen and Bensten 1997; Quintin *et al*. 2005), Nile tilapia (*Oreochromis niloticus*) (Charo-Karisa *et al*. 2006), common carp (*Cyprinus carpio*) (Vandeputte *et al*. 2004) and Arctic charr (*Salvelinus alpinus*) (Ditlecadet *et al*. 2006). There have also been some advances in mass selection research of molluscs. The Eastern oyster (*Crassostrea virginica*) (Guo *et al*. 2003) and abalone *Haliotis asinina* (Lucas *et al*. 2006) have proved highly successful in phenotypic selective farming practices.

The use of genetic markers for assessment of variability between wild stocks and hatchery stocks, the identification of broodstock by observing allelic combinations and monitoring of

inbreeding provides accurate and essential knowledge for management in selective breeding programs. Molecular markers are used as a tagging method in which individuals and pedigrees are identified and linked through allelic variation (Pan *et al*. 2004; Hauser and Seeb 2008). This also allows the identification of parent contribution to mass spawning events by determining parentage through walk back selection (Sonesson 2005). The use of genetic markers for the identification of essential breeding traits has swiftly become a potential new method for selection of advantageous breeding qualities such as growth rate and disease resistance (Hulata 2001; Okumus and Ciftci 2003).

Selecting broodstock on the basis of the animals' genetic combination and inheritance patterns with information on allelic superiority is more powerful than selection on purely phenotypic traits, as seen in mass selection (Martinez *et al*. 2005). Thus, a future form of selection within abalone farms is the use of marker-assisted selection (MAS). This entails the use of dense linkage maps and association analysis of marker alleles closely linked to genes or traits of economic interest. These associated marker loci of interest for selective breeding programs are known as quantitative trait loci (QTL). Multiple QTL can be responsible for small increments of variation of a particular trait such as growth rate, spawning date, temperature tolerance or feed efficiency ratio (Beaumont and Hoare 2003; Baranski *et al*. 2008). In order to identify genes or gene regions which contribute significantly to a phenotypic trait of interest, genetic linkage maps are required and allow the positioning and identification of strong associations between markers and gene regions participating in a quantitative phenotype. The linkage between a marker locus and QTL should be at least 1 cM (centiMorgan) or less to be useful in selective breeding in an entire population with no recombinant division in impending generations. Fine QTL association mapping between selected quantitative phenotypic variation and specific allelic combinations within a linkage mapped pedigree is therefore necessary for marker assisted selective breeding (Okumus and Ciftci 2003; Hayes *et al*. 2006).

The choice of marker for a particular genetic question depends on a number of factors such as the rate of marker evolution, or the inheritance pattern and expression or even cost and expertise required to use the technology (Liu and Cordes 2004). Slow evolving markers, for example, are useless in inferring parentage and relatedness between populations or closely related species. Similarly, fast evolving loci are not informative in

phylogenetic history assessment (Okumus and Ciftci 2003). In linkage mapping and QTL association, more variable markers are required to allow the identification of allelic segregation patterns and identify recombination events (Hubert and Hedgecock 2004).

5. Molecular markers

5.1 General information – Tools for detecting DNA variation

Molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNP) and microsatellites (short tandem repeats - STRs) have found large scale application in aquaculture (Liu and Cordes 2004) (Table 1.1). These markers can be used in parentage assignment (Jerry *et al*. 2004; Castro *et al*. 2007; Ruivo 2007; Slabbert *et al*. 2009; Van den Berg and Roodt-Wilding 2010), measuring genetic variation between populations (Campbell *et al*. 2003; Evans *et al*. 2004; Hayes *et al*. 2006; Coibanu *et al*. 2009; Marchant *et al*. 2009), conservation studies (Ferguson *et al*. 1995; Gruenthal and Burton 2005) as well as the construction of linkage maps (Coimbra *et al*. 2003; Gilbey *et al*. 2004; Ohara *et al*. 2005; Baranski *et al*. 2006a; Moen *et al*. 2008; Du *et al*. 2009; Xia *et al*. 2010).

One of the first molecular markers used for genetic analyses was protein variants or allozymes. Allozymes were useful in that they provided information required to develop models for population structure (Hauser and Seeb 2008). This marker was however abandoned over time due to the tedious methodology and low levels of polymorphism within selective coding regions (Ferguson *et al*. 1995). In the late 1980s the first DNA markers were identified, mitochondrial DNA (mtDNA). This form of marker was used for population analysis due to the vast demographic information accessible (Grant and Bowen 2000; Lundy *et al*. 2000). The use of nucleotide DNA markers was thereafter initiated (RFLP, AFLP, RAPD, microsatellites and SNPs) (Table 1.1), upon the advent of polymerase chain reaction (PCR) (Mullis and Faloona 1987).

Table 1.1: Molecular markers used in aquaculture genetics (Liu and Cordes 2004)

Marker type	Prior information required?	Inheritance	Type	Polymorphic power	Predominant applications
Allozyme	Yes	Mendelian, Co-dominant	Type I	Low	Linkage mapping Population studies
mtDNA	No	Maternal inheritance	-	-	Maternal lineage
RFLP	Yes	Mendelian, Co-dominant	Type I or Type II	Low	Linkage mapping
RAPD	No	Mendelian, Dominant	Type II	Intermediate	Fingerprinting for population studies Hybrid identification
AFLP	No	Mendelian, Dominant	Type II	High	Linkage mapping Population studies
SNP	Yes	Mendelian, Co-dominant	Type I or Type II	High	Linkage mapping Population studies
Indels	Yes	Mendelian, Co-dominant	Type I or Type II	Low	Linkage mapping
STR	Yes	Mendelian, Co-dominant	Type I or type II	High	Linkage mapping Population studies Paternity analysis

Genetic markers can be categorised into the following groups: bi-allelic dominant markers including AFLPs and RAPDs, bi-allelic co-dominant markers including RFLPs and SNPs and lastly multi-allelic co-dominant markers such as microsatellites. Certain factors are required for reliable data analysis of a particular study; dominance, polymorphism per locus and independent segregation to name a few. When selecting markers for a specific study, these factors should be taken into account (Vignal *et al*. 2002).

Molecular markers are also considered as either Type I or Type II markers, depending on whether they are found in coding or non-coding regions (O' Brien 1991). Type II molecular markers are associated with anonymous genomic regions and can include markers such as microsatellites, AFLPs, RAPDs and SNPs (Zane *et al*. 2002). Type II markers are advantageous for framework linkage maps as they have higher prevalence of polymorphism. This is attributable to the fact that type II markers are developed from non-coding regions, which is generally more prone to mutations and thus more polymorphic and abundant (Chistiakov *et al*. 2006). These markers are, in addition, useful for population studies in which selective neutrality, due to the non-coding region locality, is assumed for the particular analytical tools used for data analysis (Vignal *et al*. 2002; Liu and Cordes 2004). Type I molecular markers are DNA sequences known to be associated with genes (Zane *et al*. 2002). These gene-linked markers are essential in linkage mapping and the identification of QTL for MAS in aquaculture (Liu and Cordes 2004;

Chistiakov *et al*. 2006). Type I markers can be developed from cDNA libraries by identifying STRs within expressed sequence tags (ESTs) which can then be mapped and used to identify QTL and particular genes (Dunham 2004; Wenne *et al*. 2007). EST-STR markers are characterised as type I as they are associated with an expressed region of the genomic sequence. The amount of STRs found in these regions is expected to be much lower than that of non-coding, anonymous genomic regions due to selective pressures (Chistiakov *et al*. 2006).

5.2 Restriction fragment length polymorphism (RFLP)

RFLP markers were initially identified and used in human genetic linkage mapping in 1980 (Botstein *et al*. 1980). Single base pair alterations were identified with the use of restriction endonuclease enzymes. Restriction endonucleases were isolated from bacteria and are essentially the key factor of RFLP discovery. These enzymes cleave specific recognition sequences found all along the genomes of all species. This is a defence mechanism in bacteria as an attack against foreign DNA. Any variation found between a certain region of an individual's genomic DNA (gDNA), such as indels (insertions/deletions) or rearrangements, will be seen as variable fragment sizes when separated by gel electrophoresis (Beaumont and Hoare 2003; Liu and Cordes 2004).

Results were traditionally visualised with Southern blot analysis (Birren *et al*. 1997). Recently PCR-based methods have become more customary if sequence information is available. RFLPs are co-dominant markers providing heterozygote information; however, these markers detect a low level of polymorphism (Okumus and Ciftci 2003).

5.3 Random amplified polymorphic DNA (RAPD)

RAPD markers were developed by Welsh and McClelland (1990) and Williams *et al*. (1990). It is a dominant marker developed from the amplification of random regions of the genome using short arbitrary primers. RAPDs are non-coding, type II markers and are therefore assumed to follow Mendelian segregation. Polymorphism is identified when a mutation occurs in the primer-binding site which will cause amplification in some individuals and none in others (Liu and Cordes 2004).

The advantage of RAPDs is the fast and uncomplicated use with relatively low cost (Elo *et al*. 1997). Standard primers are commercially available and polymorphism is easily

identified with agarose or poly-acrylamide gel electrophoresis (Liu and Cordes 2004). The disadvantages of RAPDs are found in analyses, which are largely based on estimation and assumption. This is a bi-allelic marker and although at times heterozygotes are seen, it is very rare and subjective (Hassanien 2008). Another challenge using RAPDs is reproducibility of data as slight changes in reaction conditions can cause variable banding patterns (Okumus and Ciftci 2003).

5.4 Amplified fragment length polymorphism (AFLP)

The AFLP marker is similar to RAPDs since it is a PCR-based, anonymous, dominant marker; however AFLP markers can identify numerous polymorphisms at once and is more reproducible than RAPDs. The AFLP was first developed by Vos *et al*. (1995) and makes use of known restriction sites for amplification.

Usually *Eco*RI and *Mse*I enzymes are used to perform a gDNA digestion, which would provide approximately 500 000 fragments. Known adaptor sequences are subsequently ligated to the ends of all the fragments by using the endonuclease cutting site sequence. Primers can be adapted to reduce the amount of fragments obtained by PCR through an addition of a base pair (A, T, C or G) to the 3' end, consequently reducing the fragments 4-fold for improved electrophoretic analysis. The specific fragments thus allow identification of mutations in primer-binding regions (Liu and Cordes 2004).

Although this is a limited marker in that it has dominant inheritance, it generates a large amount of loci with no prior knowledge of genome sequence information and it is possible, in full-sib family analysis, to identify co-dominant markers with the relevant computer programs (Liu and Cordes 2004).

5.5 Single nucleotide polymorphism (SNP)

SNPs are the most abundant marker found in all organisms, although they have a low polymorphic information content (PIC) as they are bi-allelic polymorphisms. Theoretically SNPs can be seen as multi-allelic as there are four nucleotide substitution possibilities; however in practice it is extremely rare to observe two or more substitutions at a SNP locus (Vignal *et al*. 2002).

Nucleotide substitutions occurring as purine to purine or pyrimidine to pyrimidine are referred to as transition substitutions. Alternatively, a purine to pyrimidine or pyrimidine to purine substitution is referred to as a transversion (Vignal *et al*. 2002).

The genotyping of SNPs was originally performed by direct sequencing in some form, either shot gun or with specifically designed primers. Currently high throughput methods such as matrix-associated laser desorption ionisation-time of flight (MALDI-TOF) mass spectroscopy, real time PCR, pyrosequencing and gene chips are implemented more routinely (Liu and Cordes 2004; Wenne *et al*. 2007).

5.6 Microsatellite markers

5.6.1 GENERAL BACKGROUND

Variable number of tandem repeats (VNTRs) are arrays of repeated nucleotides of various forms. These dynamic polymorphisms were identified in 1987 and used in conjunction with RFLP markers for construction of the first comprehensive linkage map in humans (Donis-Keller *et al*. 1987). VNTRs can be separated into two specific categories based on the size of the repeat unit: Minisatellites which are typically 9-65 bp repeat arrays and secondly, the shorter sequence-repeat units or microsatellites (Okumus and Ciftci 2003). The more simplified category of VNTRs (microsatellite) is referred to as short tandem repeats (STRs). STR sequences consist of 2-8 nucleotide repeats potentially recurring up to 100 times per locus (Goldstein *et al*. 1995; Chambers and MacAvoy 2000). The polymorphic nature of STRs makes this a prominent marker for PCR detection of variance and hence linkage mapping studies (Weber and May 1989).

These markers are found in perfect forms (Figure 1.3 a), compound forms (Figure 1.3 b), where tandem repeats are found adjacent to a dissimilar repeat sequence, and interrupted forms (Figure 1.3 c) in which mutations have occurred within the repeat unit (Goldstein and Schlötterer 1999). This can occur when point mutations arise within a repeat array (Ellegren 2000).

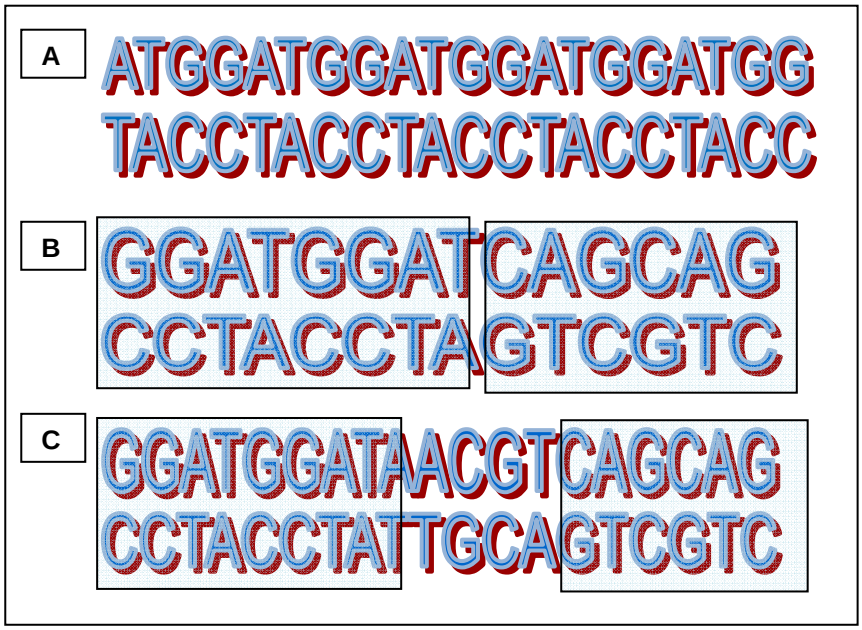


Figure 1.3: Microsatellite repeat types
A–Represents an $(ATGG)_5$ microsatellite array,
B–Represents a compound microsatellite type: $(GGAT)_2(CAG)_2$ and
C–Represents a complex or interrupted microsatellite: $(GGAT)_2 – (CAG)_2$.

Microsatellites are present in the genomes of all organisms and are found uniformly in intergenic and intronic regions and are less common in exonic regions (Chambers and MacAvoy 2000; Chistiakov *et al*. 2006). The “hotspot” prevalence of microsatellites within intronic regions and lack thereof in exonic regions, is a result of negative selection in which longer repeat sequences are detrimental to coding regions and are therefore excluded from inheritance to offspring (Chistiakov *et al*. 2006).

The abundance and average length of STRs vary between species, presumably due to species-specific genome dynamics (Lagercrantz *et al*. 1993). Microsatellites are relatively frequent in fish species, found approximately once in every 10 kbp (Wright 1993). Chistiakov *et al*. (2006) stated that in fish species the frequency of dinucleotides in general is approximately 30-67%. In studies of *Mytilus galloprovinialis*, $(AC)_n$ was found to be the most abundant repeat motif at a frequency of 1 motif every 429 kbp (Cruz *et al*. 2005) whereas in European flat oysters (*Ostrea edulis*), Naciri *et al*. (1995) also discovered an abundance of $(AC)_n$ motif’s with a motif found once every 139 bp. The most common microsatellite in vertebrates also follows a $(CA)_n$ repeat system and these sequences are generally longer and more frequent than in mollusc species (Cruz *et al*. 2005), occurring once every 15-50 kbp (Stallings 1995).

Microsatellites are thought to emerge from “regions of cryptic simplicity” or in other words, regions in the genome presenting potential expansion of repeat arrays through relatively few mutational events (Tautz *et al*. 1986; Buschiazzo and Gemmell 2006). New microsatellite alleles are formed by means of “slipped-strand mispairing” through which a repeat unit is added or lost due to misalignment of the template and nascent strand by the DNA polymerase enzyme (Figure 1.4) (Levinson and Gutman 1987; Chambers and MacAvoy 2000; Ellegren 2000). This structure must be stabilised, possibly through the formation of a hairpin loop or tri/quadruplex configuration. The fixation of an STR through strand mispairing and stabilisation complexes is dependent on the type of repeat as well as the position in the genome (Ellegren 2000; Tóth *et al*. 2000).

The second proposed theory for microsatellite evolution is that of gene conversion. Gene conversion includes unequal crossover or misalignment of chromosomal DNA during meiotic recombination (Figure 1.5 a) (Hancock 1999) and unidirectional genetic transfer (Figure 1.5 b) (Jeffreys *et al*. 1994). The concept of gene conversion as a mechanism of microsatellite evolution has been questioned and is more accepted for minisatellite expansion (Ellegren 2004). There have however been studies which link gene conversion to microsatellite expansion (Chistiakov *et al*. 2006).

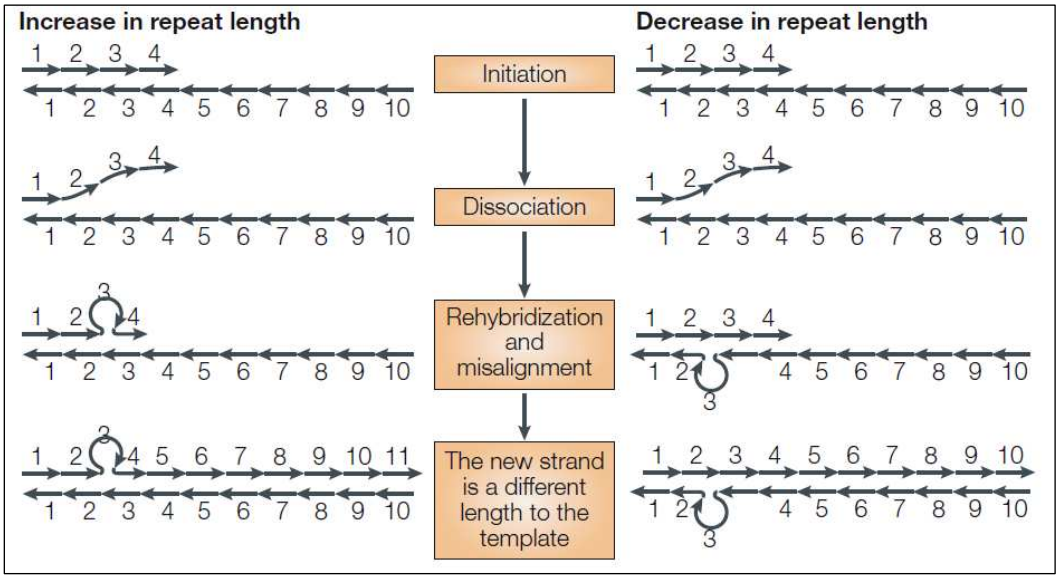


Figure 1.4: Slipped strand mispairing of microsatellites [Taken from Ellegren 2004].

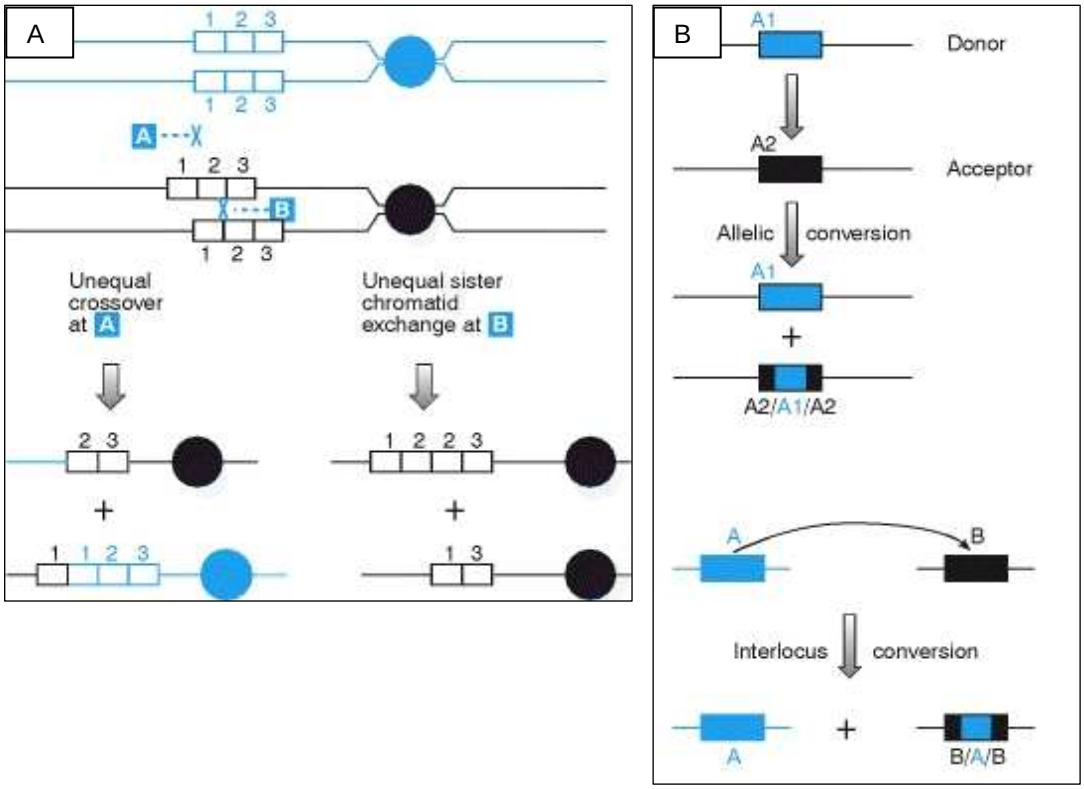


Figure 1.5: Gene conversion as a possible source of microsatellite origin and expansion. A-Shows unequal cross over, which can occur during meiotic replication and recombination. B-Gene conversion through unidirectional transfer of information, either through locus conversion or transfer between loci [Taken from Strachan and Read 1999].

The evolution of microsatellites is thought of as a dynamic "life cycle" of which STRs are born, grow and die, all depending on the direction of mutational events i.e. whether repeat units are added or removed from a repeat array, so the length of the allele changes with every generation (Chambers and MacAvoy 2000). The form of repeat can also be interrupted (Figure 1.3 c) when SNPs arise within an array. This is thought to be an initiator for microsatellite degradation or "death" as the repeat unit can be further interrupted and revert to a random sequence (Ellegren 2000; Buschiazzo and Gemmell 2006).

Mutation rates of microsatellites vary from 10^{-2}-10^{-6} events per locus per generation, however rates will differ between species and between loci within a species (Ellegren 2000). The average mutation rate seen in humans is 10^{-3} events per locus per generation (Brinkman *et al*. 1998; Ellegren 2004; Leclercq *et al*. 2010). Schug *et al.* (1997) identified a much lower mutation rate of $6x10^{-6}$ events per locus per generation in *Drosophila.*

Variable mutation rates between STR loci is attributed to numerous factors such as the length of a repeat, mutation 'hot spots', point mutations as well as sex and age biases. Microsatellites containing longer repeat arrays are prone to a higher frequency of replication slippage and thus a higher mutation rate (Zhu *et al*. 2000; Brandström *et al*. 2008); although larger motif units (nucleotide repeat length) have less chance of polymerase slippage (Buschiazzo and Gemmell 2006). The balance between point mutations and STR evolution is also a factor that is not well documented although it does play a major role in STR evolutionary dynamics (Ellegren 2004).

Microsatellites are assumed to be neutral as the anonymous nature of these markers suggests that they are not under selection pressures. Some STRs are however found in centromeric and telomeric regions and are thought to be involved in chromosome compacting in fish species, resulting in selective superiority. These molecular markers have also been associated with termination during transcription of certain human genes and are found in promoter regions, which also promote selective pressures (Chistiakov *et al*. 2006).

5.6.2 MICROSATELLITES AND USES IN AQUACULTURE

Microsatellites are used in various genetic research projects for a range of applications as they are easy to detect and analyse in addition to being the most informative individual marker available (Tautz 1989; Reece *et al*. 2004). The polymorphic (multi-allelic) nature of STRs makes this marker an ideal genotyping tool for identification of genetic differences between individuals (Tautz 1989).

Microsatellite markers have previously been used in aquaculture genetics for parentage assignment (Chambers and McAvoy 2000; Jerry *et al*. 2004; Ruivo 2007; Slabbert *et al*. 2009); wild and farm population studies (Garcia de Leon *et al*. 1998; Jackson *et al*. 2003; Li *et al*. 2006) as well as linkage mapping in Tilapia (*Oreochromis niloticus*) (Kocher *et al*. 1998); Japanese flounder (*Paralichthys olivaceus*) (Coimbra *et al*. 2003); Pacific oyster (*Crassostrea gigas*) (Hubert and Hedgecock 2004); European sea bass (*Dicentrarchus labrax*) (Chistiakov *et al*. 2005) and Japanese abalone (*Haliotis discus hannai*) (Sekino *et al*. 2006).

When selecting markers for linkage mapping, there are certain factors to consider, for example; the number of markers required for statistical analysis and the amount of information given by a single locus. When using full-sib families for the development of a linkage map, polymorphic markers are particularly useful as the phase of alleles can be easily identified (Hauser and Seeb 2008). Microsatellite PIC (polymorphic information content) per locus and relative abundance make these markers ideal for framework linkage mapping.

5.6.3 MICROSATELLITE MARKER DEVELOPMENT STRATEGIES

5.6.3.1 Genomic library enrichment and hybridisation

Microsatellite flanking regions are particularly variable as these markers are predominantly found in non-coding regions of all organisms. This makes cross-species amplification rather difficult in many instances (Zhu *et al*. 2000). There are various methods for developing microsatellite markers of which subtractive hybridisation is one, used extensively in non-model species such as abalone. Zane *et al*. (2002) reviewed the advantages and disadvantages of this method and improved the tedious steps through simplification and optimisation.

In the early 1990s, enrichment protocols were developed for microsatellite isolation in which cloning from partial genomic libraries was performed and radioactive labelled probes used to screen for repeats; however this did not prove to be highly successful and was especially problematic in organisms with low levels of repeat arrays (Tautz 1989; Walderbieser 1995). Enrichment and hybridisation selection techniques were subsequently developed to simplify the isolation steps and provide a means of scanning sequences prior to cloning (Armour *et al*. 1994; Kandpal *et al*. 1994). Ostrander *et al*. (1992) and Paetkau (1999) also developed a primer extension enrichment protocol to increase the amount of repeats in the library for an increased probability of marker isolation. Zane *et al*. (2002) later used all the reported slight improvements and further established a well optimised, rapid protocol for maximum marker capture through enrichment and hybridisation using numerous PCR amplification steps (Figure 1.6), also known as the Fast isolation by AFLP of sequences containing repeats (FIASCO) technique.

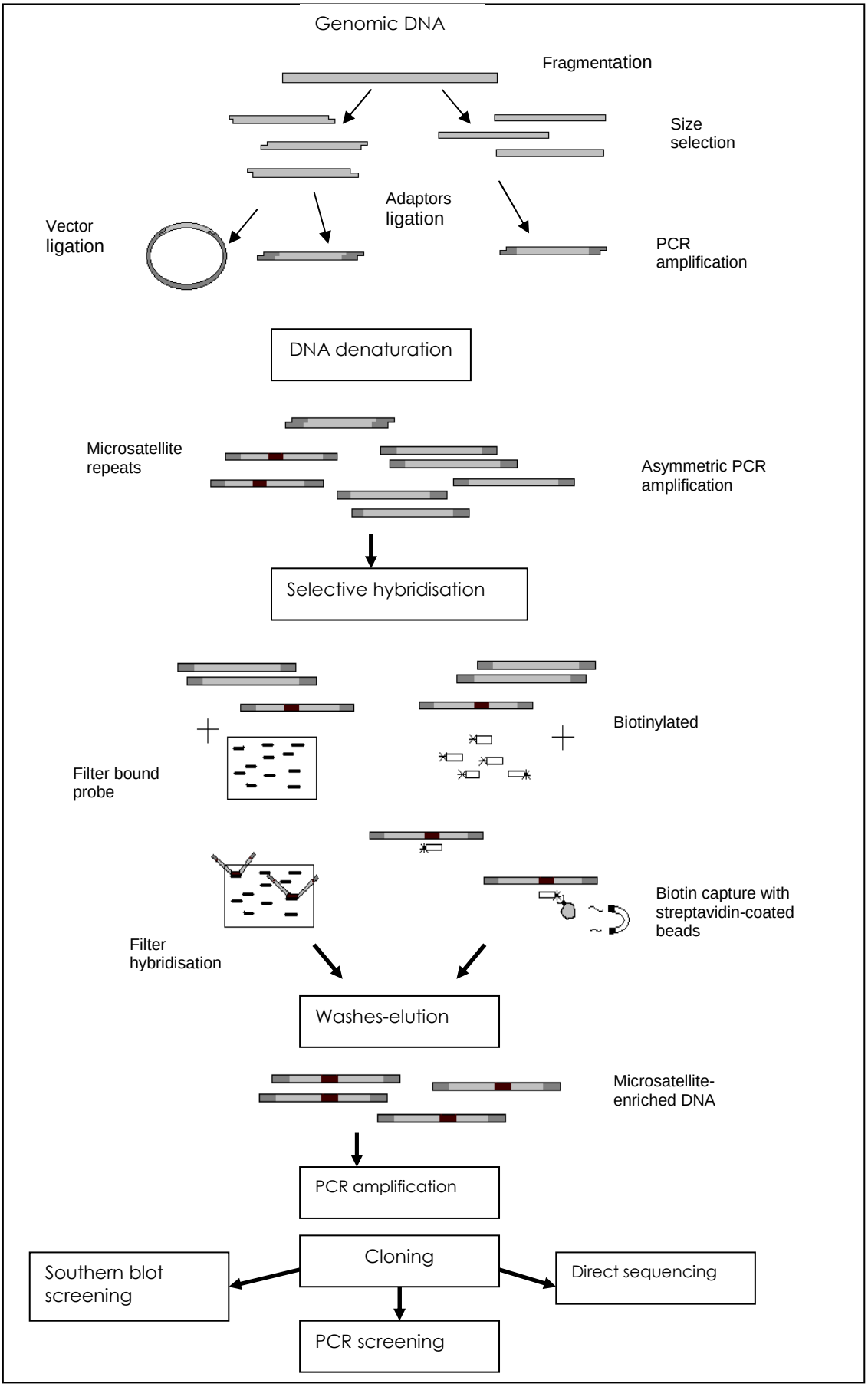


Figure 1.6: The FIASCO marker development technique [Figure adapted from Zane *et al*. 2002].

The difference between a normal hybridisation protocol and the FIASCO technique ultimately lies in the library enrichment and PCR amplification using AFLP adaptors (Figure 1.6). These steps provide numerous copies of microsatellites and improve the isolation success rate immensely. An improvement from 4-5% to 50-95% was seen, with variance depending on the species (Zane *et al*. 2002). The FIASCO technique is widely used to develop anonymous type II molecular markers.

There have been alternative attempts to improve the early enrichment protocols by altering restriction endonucleases and probes used for a more cost-effective, simplified and successful method. Hamilton *et al*. (1999) developed a SNX linker protocol which allows fragment screening prior to cloning and allows the use of multiple and various restriction enzymes for different species. The SNX protocol was further modified and improved by Glenn and Schable (2005) through the use of a super SNX linker, which provides a more efficient TA cloning step due to the addition of a "GTTT pig tail" to the linker.

A second technique for STR development, used in this study, is the detection of microsatellite repeat arrays from direct sequence information. This has rapidly become the method of choice due to the advances in direct sequencing, in which masses of sequences data can be obtained and used as a basis for microsatellite development.

5.6.3.2 Next generation sequencing and microsatellite marker development

There have been major advances in high throughput sequencing techniques in the past five years (Shendure and Ji 2008). Two systems, namely the GS-FLX (Roche) and the Illumina Genome Analyser both use massive parallel sequence-by-synthesis methods as opposed to the 96 samples per run in the traditional capillary based "cycle sequencing" of the Sanger sequencing system (Sanger 1988). The next generation sequencing also excludes cloning and hence the bias associated with cloning; however much shorter reads are produced (Mardis 2008). Such advances will have a major influence on cost-effective marker development strategies and large-scale genome and transcriptome sequence analysis (Wall *et al*. 2009), especially in non-model organisms such as abalone.

The GS-FLX system (http://www.454.com), also known as pyrosequencing or 454 sequencing, has been available for commercial use since 2004. This system employs

pyrophosphate detection upon annealing of a particular base pair through a series of enzymatic reactions (workflow information: http://www.454.com/enabling-technology/the-technology.asp [Feb 2010]). A library of template DNA fragments are developed with the use of nebulisation (Figure 1.7 a). An emulsion PCR (Figure 1.7 b) is subsequently performed in which numerous single fragments are amplified in parallel within emulsion micro reactors. The beads containing the amplified products are placed into single wells of a picotiter plate (PTP) combined with enzyme and packing beads for the sequencing initiation and optimal bead placing (Figure 1.7 c). The sequencing reaction uses sequential cyclic flow of a single nucleotide type which is fluorescently labelled. Upon hybridisation with the complimentary template strand within each well, enzymatic reactions occur which will ultimately release a light signal directly proportional to the quantity of bases bound at each template strand in each well, to take into account homoploymers (Figure 1.7 d). The single fragments obtained are approximately 250 bp in length (400 bp for the latest titanium series) which are then compiled into contig fragments using software programs such as Cap3 (Huang and Madan 1999), Eagleview (Huang and Marth 2008), SCARF (Barker *et al*. 2009) and Gap5 (Bonfield and Whitwham 2010). The sequencing facilities also provide specific packages including software programs such as GS *de novo* assembler (http://www.454.com) and CASAVA 1.6 (http://www.illumina.com), which also align fragments into contigs.

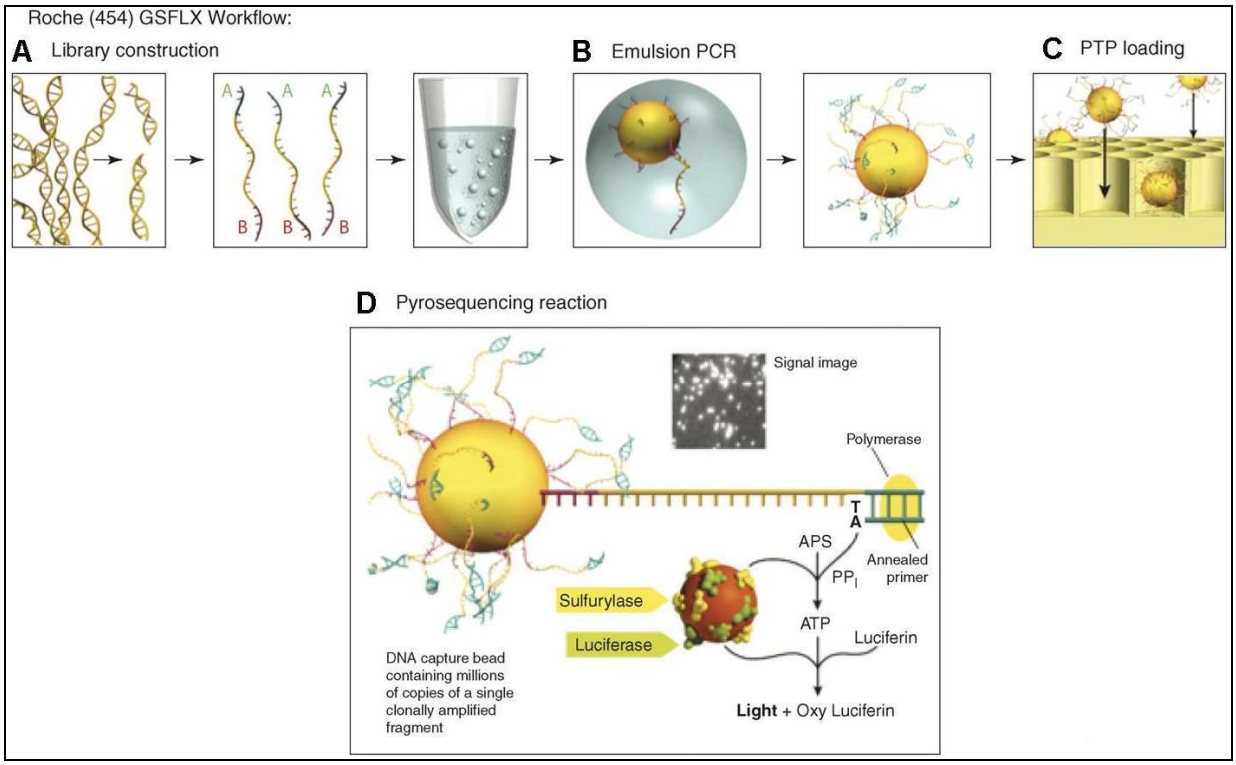


Figure 1.7: The work flow of the GS-FLX (Roche) sequencing system. A-The library construction of ssDNA through *in vitro* nebulisation, adaptor ligation and micro reactor placement steps. B-Emulsion PCR of single fragment-containing agarose beads. C-The beads containing clonally amplified fragments are placed into single wells of a picotiter plate (PTP) along with enzyme and packing beads. D-The pyrosequencing reaction occurs in parallel across the plate, detecting complimentary annealing through enzymatic cleavage from pyrophosphate to release oxy Luciferin and light [Taken from Mardis 2008].

The Illumina Genome Analyser (Illumina GA) (http://www.illumina.com), uses a similar methodology as the GS-FLX, however, the PCR amplification is performed on an 8-lane glass flow cell platform in which groups of amplified clusters are sequenced (work flow information: http://www.illumina.com/pages.ilmn?ID=203 [Feb 2010]). The Illumina was available for commercial use in 2006 and has been used in numerous high throughput projects since (Frio *et al*. 2010; Graham *et al*. 2010; Turner *et al*. 2010). The process is initiated with the development of an *in vitro* single-stranded oligo-adaptor ligated library which is placed on the flow cell surface using a micro fluidic cluster station for the sequence by synthesis. Cluster PCR is performed in these isolated areas as both primers are present on the flow cell and the template simply folds over to form a bridge. Approximately 1000 copies of each template cluster are obtained after PCR amplification (Figure 1.8). Sequencing these templates commences with 3'-OH deactivated, fluorescently labelled dNTPs ensuring single base incorporation per cycle. The resultant image is processed and the dNTP de-blocked for the subsequent cycle of base

incorporation. This process takes approximately four days and produces single fragment reads of 34-40 bp or 40-50 million sequences overall (Mardis 2008).

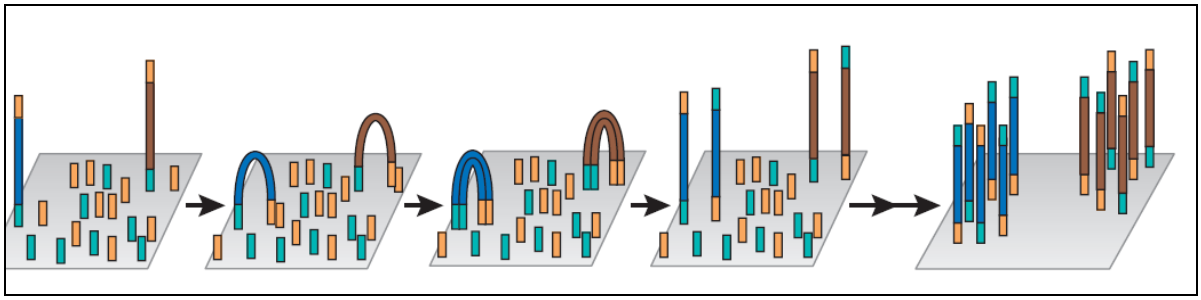


Figure 1.8: The workflow of the Illumina Solexa Genome Analyser. An *in vitro*–constructed adaptor-flanked shotgun library is attached to the solid surface of the flow cell. Cluster PCR is performed within the area of the original library as the surface is covered with both primers. Approximately 1000 copies of a single template library is created in clusters [Taken from Shendure and Ji 2008].

Data from both 454 pyrosequencing and the Illumina can be subjected to post sequence processing with either development of contigs from overlapping single reads or alternatively resequencing with comparisons to a reference genome. The Illumina GA sequence-by-synthesis technique has been used in several studies to develop SNP markers (Hillier *et al*. 2008; Novaes *et al*. 2008; Van Tassell *et al*. 2008) as well as microsatellite markers (Dempewolf *et al*. 2010); and when using transcriptome information, it can be used to specifically develop EST markers for comparative analysis and mapping. Pyrosequencing has also been used to develop microsatellite markers (Abdelkrim *et al*. 2009; Allentoft *et al*. 2009; Santana *et al*. 2009; Slabbert 2010) as well as SNP markers (Barbazuk *et al*. 2007; Maughan *et al*. 2009; Sanchez *et al*. 2009).

5.6.4 MICROSATELLITE GENOTYPING CHALLENGES

When using microsatellites, in some cases, allele calling can be extremely subjective and imprecise. There are five main sources of allele calling errors: biochemical artefacts, DNA quality, sequence alteration (mutation), human error (Pompanon *et al*. 2005) and genome dynamics (Baranski 2006). An allele which does not follow the traditional peak form can easily be misinterpreted due to these abnormalities. When genotyping in a full-sib family, however, the alleles can be identified in the parental genotypes initially and the expected peaks should thus be observable in the offspring.

PCR artefacts can affect genotyping, for example, Go*Taq* polymerase often adds an extra Adenosine (A) nucleotide to the end of the PCR product (Pompanon *et al*. 2005). An

"extra" peak is thus seen and the misinterpretation probability is augmented (Figure 1.9). Another phenomenon often seen with automated genotyping is the "stutter band" (Figure 1.10), seen when slippage of the *Taq* polymerase takes place due to the repetitive nature of microsatellite markers (Ginot *et al*. 1996).

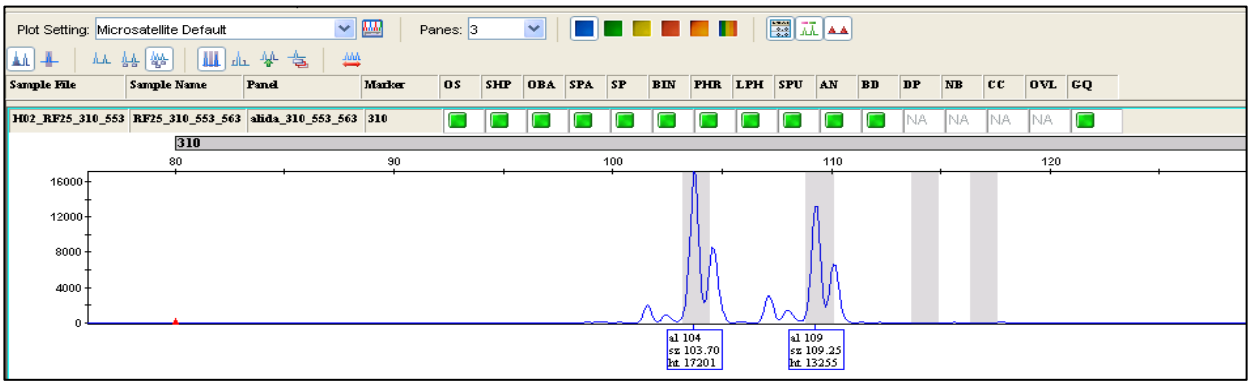


Figure 1.9: Electropherogram of double peaks (addition of dATP by the polymerase enzyme). The additional peak of each allele has a lower intensity as there is a higher frequency of molecules that do not possess extra dATP.

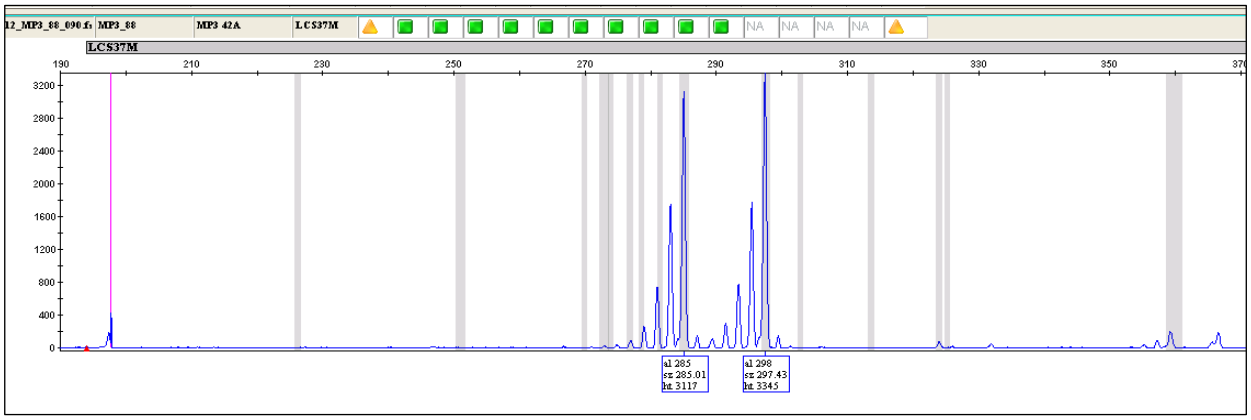


Figure 1.10: Electropherogram of stutter peaks. This is due to the slippage nature of polymerase enzymes and results in an array of alleles building up to the actual allele size.

Null alleles occur when a specific allele does not amplify due to the primer binding sites differing from the oligonucleotide designed. This is mostly due to mutations of some sort occurring in the flanking region of the locus and can be rectified by redesigning the primer (Callen *et al*. 1993; Pemberton *et al*. 1995). Null alleles are a major concern when it comes to microsatellite genotyping as the error rate can drastically affect genetic diversity, population structure or parentage assignment studies. In population studies, for example, this may cause an over estimation of the number of homozygotes and an under estimation

of the number of heterozygotes (Callen *et al*. 1993; Panova *et al*. 2008). Null alleles are generally detected once Hardy-Weinberg equilibrium tests are performed on the data and an excess in homozygotes is noted (Hoffman and Amos 2005). In pedigree analysis it is easy to detect a null allele in most cases (Figure 1.11).

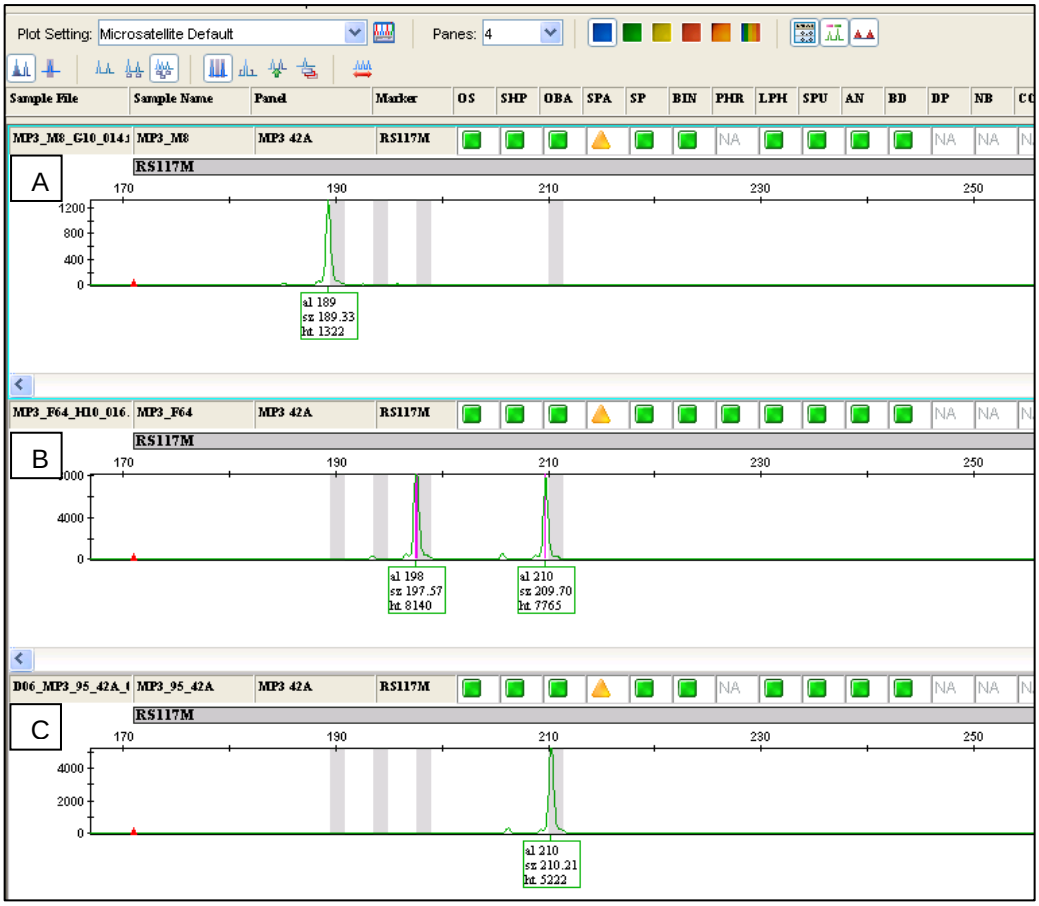


Figure 1.11: Electropherogram of a null allele segregation
A-Parent 1 B-Parent 2 C-Offspring
Parent 1 possesses only one allele and appears homozygous whereas parent 2 is clearly heterozygous. The null allele from parent 1 can be confirmed when offspring demonstrate one of the alleles from parent 2 but not parent 1.

Allele drop-out is another genotype phenomenon which can only be detected in pedigree-specific studies and is basically a stochastic PCR "struggle" between the reactions (especially in a multiplex) preventing a larger allele from amplification. The individual therefore appears homozygous but is in fact heterozygous with a non-amplifying or much lower intensity allele (Figure 1.12).

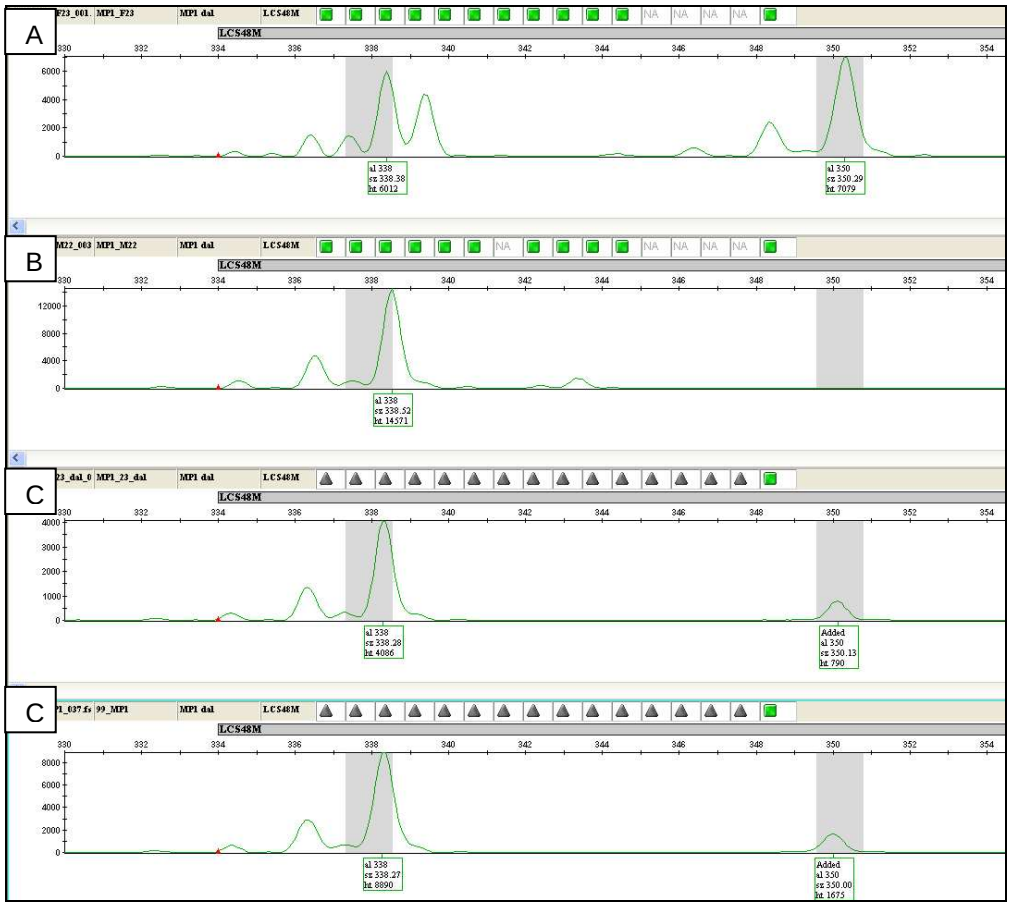


Figure 1.12: Electropherogram of allele drop-out during electrophoresis
A-Parent 1 B-Parent 2 C-Offspring
This figure displays the effects of allele drop-out, which might be problematic when looking at population data. In a pedigree it is easy to identify.

Allele "shifts" can occur when using automated genotyping strategies during the capillary electrophoresis, providing an opportunity for erroneous allele calling (Figure 1.13). This is a problem mostly associated with parentage or population genetic studies rather than pedigree analysis such as with linkage mapping. Figure 1.14 depicts another example of possible misinterpretation of alleles. Parent one could be miscalled as homozygous and the genotype of parent two could also be misleading. However, when observing the allele segregation within the offspring, it is clear which alleles are inherited from which parent.

Figure 1.13: Electropherogram of allele shifts during electrophoresis
A-Parent 1 B-Parent 2 C-Offspring
This figure displays the effects seen in automated genotyping in which in some cases allele shifts can occur. In a pedigree it can be easily corrected as the parental alleles are known, however in population studies it could be misinterpreted as separate alleles.

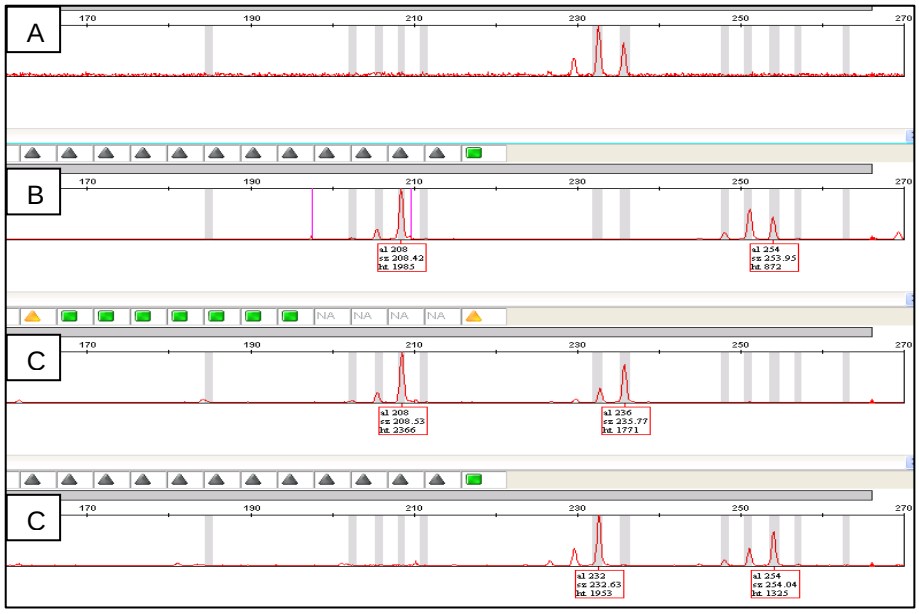


Figure 1.14: Electropherogram of ambiguous genotypes
A-Parent 1 B-Parent 2 C-Offspring
This figure of an electropherogram shows some difficulties in determining whether an individual is homozygous or heterozygous. Parent 1 is heterozygous. Parent 2 has one smaller allele (208) and a controversial larger allele (254) which can only be confirmed as a single allele when assessing the offspring. The larger allele of parent 2 possesses a stutter peak as only the 254 allele segregates in the offspring.

Baranski (2006) noted that microsatellite loci with more than two alleles per individual are generally regarded as non-reliable markers and consequently discarded. Figure 1.15 shows offspring possessing three alleles per individual. There have been proposed hypotheses to explain the presence of multiple alleles such as genome duplication and aneuploidy, tandem duplication and locus transfer by mobile elements (Baranski 2006). In some cases two of the alleles are inherited as a unit, providing grounds for the tandem duplication premise. Such a marker can still be useful in linkage mapping studies provided the genotypes are adjusted accordingly by assigning the two alleles as one.

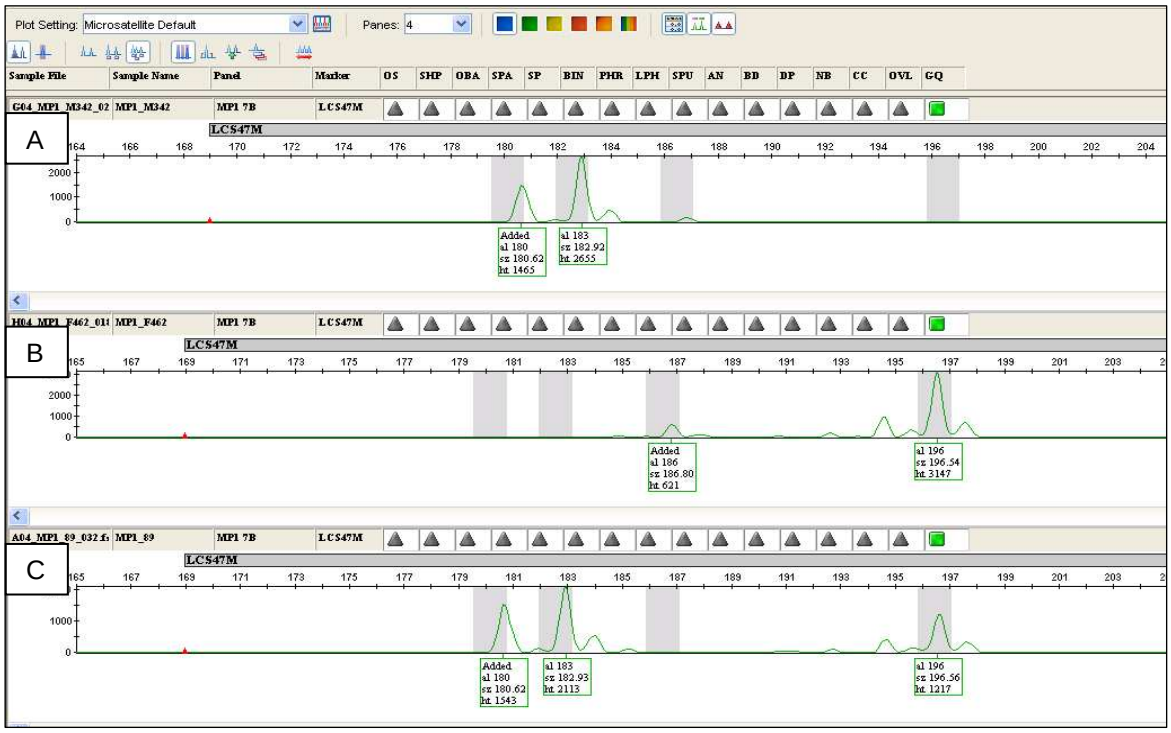


Figure 1.15: Electropherogram of a duplicated locus
A-Parent 1 B-Parent 2 C-Offspring
It is clearly seen in this figure that both the alleles from parent 1 are segregating as a unit in the offspring. This is an indication of a duplicated locus.

6. Linkage mapping

6.1 Overview of linkage mapping in aquaculture

Linkage maps are developed using recombination frequencies between genetic markers and identification of close grouping through linkage disequilibrium. The concept is based on the fact that markers physically close to one another are inherited as a unit depending on whether recombination interrupts this association. Markers with a 50% (or larger) recombination frequency are assigned as unlinked as the large amount of crossing over

between them permits the assumption that they are found on separate chromosomes (Collard *et al*. 2005). Linkage between paired markers is calculated by the use of logarithm of odds (LOD) scores, which is a statistical ratio of linkage versus no linkage. In the majority of mapping studies a threshold LOD value of 3 is employed, which indicates a probability of linkage at 1000:1 favouring linkage. Reducing this threshold will allow additional markers to be assigned to a linkage group if a particular marker was not significantly linked at LOD 3, and vice versa. Mapping functions are used to convert recombination frequencies to mapping distance (centiMorgan) as the relationship between these two variables is not linear. The two most frequently used mapping functions are known as Haldane (Haldane 1919) and Kosambi (Kosambi 1944). The Haldane mapping function assumes equally probable cross over events between two loci during meiosis, whereas the Kosambi mapping function incorporates interference making the latter a more sensitive algorithm for linkage map construction (Danzmann and Gharbi 2007; Huehn 2010).

A number of genetic linkage maps have been developed in domesticated animals including swine (Ellegren *et al*. 1994), sheep (Crawford *et al*. 1995), bovine (Kappes *et al*. 1997) and chickens (Maddox *et al*. 2001). The first linkage map for an aquaculture species was developed in the late 1990s for tilapia, *Oreochromis niloticus* (Kocher *et al*. 1998; Agresti *et al*. 2000). Multiple linkage maps have since been developed for aquaculture species including the Japanese flounder, *Paralichthys olivaceus* (Coimbra *et al*. 2003), the white shrimp, *Penaeus vannamei* (Pérez *et al*. 2004; Zhang *et al*. 2007) and numerous others (Table 1.2 a). The development of linkage maps for abalone has only been a recent occurrence and has thus far been developed for *Haliotis rubra* (Baranski *et al*. 2006a), *H. discus hannai* (Liu *et al*. 2006) and more recently, *H. diversicolor* Reeve (Yaohua *et al*. 2010). There have also been QTL mapping studies for some fish species (Table 1.2 b) as well as Blicklip abalone (Baranski *et al*. 2008).

A preliminary linkage map has also been developed for *H*. *midae* (Badenhorst 2008) using AFLP and microsatellite markers. The haploid chromosome number of *H*. *midae* (18) has been recently determined by Van der Merwe and Roodt-Wilding (2008). Eighteen linkage groups are thus the aim for a linkage map of *H*. *midae* with the assumption that each linkage group represents a chromosome.

Table 1.2: A list of aquaculture species for which linkage maps and QTL maps have been created. A-Examples of aquaculture species for which linkage maps have been developed using microsatellites, including the number of markers used, map length, linkage group number and average distance between markers. B-Examples of aquaculture species for which QTL maps have been developed.

A Species	Number of markers*	Map length (cM)*	Linkage groups*	Distance between markers*	Reference
Sea bream	204	1242	26	6.1	Franch *et al*. 2006
Sea bass	162	567/906	25	3.5/5.6	Chistiakov *et al*. 2005
Rainbow trout	903	2750	31	3.0	Guyomard *et al*. 2006
Brown trout	288	346/912	37	1.2/3.2	Gharbi *et al*. 2006
Arctic charr	327	390/992	46	1.2/3.0	Woram *et al*. 2004
Tilapia	546	1311	24	2.4	Lee *et al*. 2005
Common carp	268	4111	50	15.3	Sun and Liang 2004
Pacific oyster	119	1031	11	9.5	Li and Guo 2004
Pacific oyster	102	616/770	11/22	6.0/7.6	Hubert and Hedgecock 2004
Blue mussel	116/121	825/863	14	8.0/8.1	Lallias *et al*. 2007
Yellow tail	175/122	548/473	21/25	4.8	Ohara *et al*. 2005
Blacklip abalone	102/98	621/766	17/20	7.3/9.8	Baranski *et al*. 2006a
Pacific abalone	94/119	1366/1774	19/22	18.2/18.3	Liu *et al*. 2006
South China abalone	233/179	2817.1/2773	18/17	25/25.7	Yaohua *et al*. 2010

B Species	Trait	Reference
Rainbow trout	Embryonic development	Martinez *et al*. 2005
Rainbow trout	Body length, thermotolerance	Perry *et al*. 2005
Rainbow trout	Thermotolerance	Danzmann *et al*. 1999
Rainbow trout	Growth, conditioning factor and maturity age	Martyniuk *et al*. 2003
Rainbow trout	Infectious hematopoietic necrosis	Rodriquez *et al*. 2004
Atlantic salmon	Body weight, condition factor	Reid *et al*. 2005
Arctic charr	Growth rate	Tao and Boulding 2003
Tilapia	Thermotolerance	Moen *et al*. 2004
Black-lip abalone	Growth	Baranski *et al*. 2008

*If sex-specific maps were created: The marker number, map length, linkage group and distance between markers for the male map is displayed before the respective female map.

In genetic studies, linkage maps are essential for QTL identification and mapping of economically significant traits (Hubert and Hedgecock 2004; Reece *et al*. 2004; Baranski *et al*. 2006a) as well as lethal genes (Launey and Hedgecock 2001) and ultimately MAS strategies (where only genetically superior animals are used for breeding as they possess favourable traits for farming) (Hayes *et al*. 2006). Comparative mapping, in which multiple linkage maps from an economically commercial species is merged, is also a useful tool for identification of advantageous traits in selective breeding (Birren *et al*. 1997).

Informative molecular markers such as microsatellites are useful for genetic linkage map construction (Hubert and Hedgecock 2004; Chistiakov *et al*. 2006) due to hypervariability, dispersal along the genome, co-dominant inheritance and the ability to multiplex for automated genotyping (Chistiakov *et al*. 2006). An additional advantage of microsatellite markers in large-scale linkage mapping is that fewer reference families are required for construction of a map due to the amount of information retrieved per marker locus (Vignal *et al*. 2002). Microsatellites are therefore the most preferred markers for linkage mapping in aquaculture, allowing the allelic variation and transferability between populations to improve the identification of QTL and advance MAS (Chambers and MacAvoy 2000). Some examples of aquaculture species for which QTL maps have been developed for various economically important traits are found in Table 1.2 b.

6.2 Linkage mapping – what programs are available

It is easy to infer genotype segregation patterns when using full-sib pedigrees in linkage map development as a result of predictable Mendelian segregation patterns. When using single pedigree data for linkage map development, however, is not optimal as recombination frequencies differ slightly between families of a population, resulting in erroneous map distances. This effect can be reduced by using genotype data from multiple families for linkage mapping, with subsequent merging to compensate for any recombination differences (Ferguson and Danzmann 1998).

The lengths of sex-specific linkage maps vary between species and between sexes within a species. This phenomenon is due to the contrasting levels of recombination between the sexes and can also have an effect on map distances. Maps developed from female molluscs are generally longer than those seen using data from the males (Hubert and Hedgecock 2004; Chistiakov *et al*. 2006). For this reason linkage analysis is performed in more than one pedigree on male and female segregation patterns, separately, using various software programs.

The most commonly used linkage mapping software available includes: MapMaker (Lander *et al*. 1987) or JoinMap® v3 (Van Ooijen and Voorrips 2001) and more recently, JoinMap® v4 (Van Ooijen 2006). MapMaker and JoinMap® v4 differ in that JoinMap® v4 can integrate separate sex-specific maps and allow additional linkage data to be added with recalculations. JoinMap® v4 uses two algorithms for linkage mapping; regression

mapping algorithm (Stam 1993) and multipoint maximum likelihood (MML) algorithm (Jansen *et al*. 2001). The ML algorithm was developed to allow the fast, efficient mapping of extreme numbers of marker data as the regression model became time-consuming (Van Ooijen 2006).

6.2.1 JOINMAP® v4 - REGRESSION MAPPING ALGORITHM (STAM 1993)

The regression model places loci into various positions in a pair-wise, goodness-of-fit calculation in order to establish the best possible order. It is initiated with the most informative pair of loci and expands as markers are searched for the best position. Pairwise recombination frequencies and LOD scores are used for this procedure and up to three "rounds" of ordering is performed to ensure the maximum amount of loci is included in a single map. Markers with a recombination frequency lower than the threshold and LOD score higher than the LOD threshold are used for these calculations. A "ripple function" is also used subsequent to goodness-of-fit positioning to ensure no local bias positioning occurs, providing an overall map positioning strategy (Stam 1993). The recombination frequencies are ultimately converted to distance with the use of either Haldane or Kosambi's mapping algorithms (Van Ooijen 2006), prior to drawing the map with MapChart (Voorrips 2002).

6.2.2 JOINMAP® v4 – MAXIMUM LIKELIHOOD MAPPING ALGORITHM (JANSEN *et al*. 2001)

The maximum likelihood algorithm was developed to overcome the complexities of mapping with extremely large data sets. The linear regression algorithm (Stam 1993) cannot effectively analyse data sets with over 50 marker loci mapped to a single linkage group; hence the development of a MML based algorithm by Jansen *et al*. (2001). This MML algorithm consists of three techniques which will successfully order the markers appropriately and determine relative distances between them. The three techniques are: simulated annealing, Gibbs sampling and spatial sampling. Simulated annealing (Kirkpatrick *et al*. 1983) uses Monte Carlo optimisation in which an acceptance probability is used to search maximum likelihood of positioning for an estimated map construction. Gibbs sampling (a Monte Carlo expectation maximisation algorithm) is then used to optimise the order (Van Ooijen 2006). Gibbs sampling assumes that adjacent chromosomal segments recombine independently and are therefore not influenced by another crossover event nearby, as with Haldane's mapping function (Van Ooijen 2006). Spatial sampling prevents "trapped local positioning" by allowing the map to be built

gradually, halting at each of 5 pre-set thresholds (most stringent to least stringent). These thresholds are instated to allow the markers to be randomly placed in a list to compare each locus with the next and exclude any recombination frequencies that are below the threshold. This provides a spatial sample of loci where the markers that are too close together are excluded and the process is performed at each decreasing threshold that is set (Van Ooijen 2006).

7. QTL mapping

Economically attractive phenotypic traits such as growth rates, food conversion rates, cold tolerance and meat quality in commercial aquatic species are mostly due to polygenic effects. Multiple genes or specific gene regions are responsible, in a quantitative manner, for variation of a particular trait between individuals of a species (Ferguson and Danzmann 1998; Danzmann *et al*. 1999). Variation of these loci between individuals or population groups gives an indication of adaptation to certain evolutionary pressures resulting in superior genetic haplotypes for certain traits (Slate 2005). QTL are discovered via linkage mapping of molecular markers with comparison to associations with the particular trait of interest (Ferguson and Danzmann 1998; Danzmann *et al*. 1999). If a marker is closely linked to a QTL and is not separated by recombination, the marker is said to be in linkage disequilibrium within that family (Slate 2005).

Dense linkage maps are necessary for the identification of QTL as this will allow accurate estimates of linkage probability and reduce type II errors (falsely predicted significant associations). A wide range of mapped, highly informative molecular markers, which are easily genotyped, is the key to accurate predictions of linked segregation with designated phenotypic traits (Baranski 2006). A coarse QTL map is initially created, where the distance between markers is approximately 10-30 cM. A fine map, densely occupied with markers, is subsequently created in which the distance between markers is approximately 1-5 cM.

There are two main mapping family designs, depending on whether inbred or outbred mapping populations are available. Inbred lines are advantageous as it provides either an F_2 interbred population, in which two F_1 homozygote lines are crossed, or a backcross

population, in which F_1 and a parent is crossed, all possessing heterozygous F_1 parents for all segregating loci, making it more informative than outcrossed populations. Outbred populations make use of sib-ships, either half- or full-sib, that require the use of highly polymorphic markers for an improved power of QTL detection as not all mapping families are guaranteed to be informative at every locus (Slate 2005). When looking at the type of mapping families used for bivalve molluscs in recent mapping studies, it is occasionally seen that inbred lines are created such as the map constructed for the Pacific oyster (Hubert and Hedgecock 2004). In most instances however, half- or full-sib families are used for such projects (Lallias *et al*. 2007; Du *et al*. 2009; Yaohua *et al*. 2010). The highly fecund nature of molluscs such as abalone (Evans *et al*. 2004; Lucas *et al*. 2006), allows the production (in farming environments) of extremely large families, ideal for linkage mapping and QTL projects using the half/full-sib family design.

There are three possible routes to follow when embarking upon QTL mapping: Single-point analysis, Simple interval mapping and Composite interval mapping. Single-point or single-marker analysis does not make use of a linkage map but alternatively makes use of t-tests, ANOVA and linear regression models for estimation of significant linkage. This method is not entirely satisfactory as the strength of QTL-linked markers is largely underestimated. SIM makes use of dense linkage maps and provides a much more accurate result. CIM is the most informative, accurate method of the three mentioned. CIM encompasses interval mapping and linear regression in conjunction with permutation tests to determine threshold significance of QTL-marker linkage. The flanking markers of a particular QTL are identified by means of this method and are used to further uncover the chromosomal region between them (Collard *et al*. 2005).

The focus on creating genetic maps and saturating these maps has become a priority in all cultured *Haliotis* species in order to develop maps with sufficient markers for future QTL mapping and MAS.

8. Marker-assisted selection

Markers closely linked to a QTL of interest and segregating in a population can be used in MAS. This is an incredibly useful method of breeding animals with regard to traits that are economically favourable but difficult to monitor, such as disease resistance and feed efficiency (Tanksley 1993; Ferguson and Danzmann 1998). Ten to 20 QTL are generally required to monitor a particular economically desirable trait and utilise in MAS (Hayes *et al*. 2006), however a minimum of 15 QTL is preferred (Wenne *et al*. 2007). Identification of marker alleles specifically linked to QTL can only be performed in large populations. This is because the accuracy of measuring genetic distance and marker order placement is directly proportional to the number of individuals used in the study. Markers should be at least 1 cM from the particular QTL to ensure dependable segregation alongside the trait in every population (Collard *et al*. 2005).

Once markers have been successfully associated with a phenotype with significant certainty, the marker must be validated. Validation of markers is basically conformation of effectiveness of trait prediction in various populations from various genetic backgrounds (Collard *et al*. 2005).

The development of an extensive genetic linkage map as well as a dense QTL map, using type I and type II markers, will assist in future MAS. With the use of these well developed maps, economically favourable genes can be identified through tight associations with markers. This information will, ultimately, be at the disposal of farmers and improve practices by promoting selection of broodstock animals that contain the marker alleles associated with preferred traits for optimal breeding (Chistiakov *et al*. 2006).

9. Aims and objectives

This study consists of two main sections:

1) Microsatellite marker development using two strategies and,
2) Linkage mapping

1) Microsatellite marker development using two strategies –

Aim: Two separate methods of microsatellite marker development (FIASCO for type II microsatellites and directly from sequence-by-synthesis data for type I microsatellites) are used and compared. The success and difficulty of both techniques are evaluated and compared.

Objectives: The first technique, FIASCO, is a hybrisdisation method based on a repeat enrichment library and AFLP adaptor-specific primers. The method focuses on capturing the repeat containing fragments from the enrichment library using streptavidin-coated magnetic beads and PCR amplification. The positively cloned sequences are put through various editing strategies and optimised for PCR amplification with sequence-specific primers. Since the primers are designed from non-coding random regions of the genome, type II anonymous microsatellites are developed.

The second method of microsatellite identification focuses on searching through sequences directly obtained from Illumina next generation sequencing. The sequences are obtained from cDNA and are therefore all type I EST-STR markers.

Markers obtained from these two methods and the level of technicality, time and costs of the two methods are compared.

2) Linkage mapping –

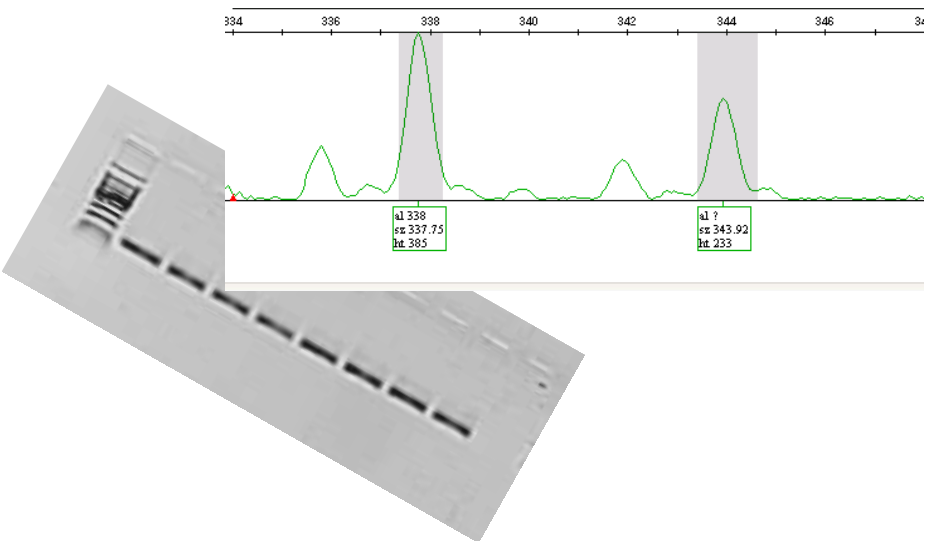
Aim: Use 201 previously developed and 14 additional (developed in this study) microsatellite markers for sex-specific linkage analysis using 100 offspring from two full-sib *H*. *midae* families and integration thereof using JoinMap® v4 (Van Ooijen 2006).

Objectives: The microsatellite markers which are found to be polymorphic in the full-sib family are genotyped using fluorescently labelled markers. The genotype data is then used for the construction of a framework linkage map.

The segregation patterns of marker alleles are used to identify recombination rates by means of odds ratios. The distance between markers as well as marker order is calculated based on maximum likelihood and mapping functions, all incorporated in JoinMap® v4. The final map is drawn up as separate female and male maps as the recombination rates and genome sizes vary. Comparisons and complete analysis of the markers and positions are evaluated.

Chapter two

Microsatellite marker development



1. Abstract

Haliotis midae **is an important commercial aquaculture resource in South Africa and the focus of the industry has thus shifted towards a genetic improvement program. Molecular markers are an important component of such a program and can be used for various applications including pedigree analysis, linkage mapping, QTL mapping and marker-assisted selection. Microsatellites are currently the marker of choice for such studies of abalone species. This study contributed to a larger study in which 44 family-specific microsatellite markers were developed using adapted FIASCO methods (Slabbert *et al*. 2010/Appendix A). A total of 38 out of 88 (44 x 2) marker-family combinations were informative. Of these 38 combinations, 10 did not conform to Mendelian segregation, but could be explained by null alleles, possible PCR errors or allele scoring difficulties. Fourteen microsatellites were also developed using Illumina sequence-by-synthesis contigs. These markers were subjected to various statistical analyses such as allele frequencies, F-statistics and Hardy-Weinberg probability testing. Only six of these fourteen markers conformed to Hardy-Weinberg expectations. The average observed and expected heterozygosities were 0.366 and 0.453, respectively, with an average F_{IS} displaying a heterozygote deficiency (0.178). The overall result of sequence-by-synthesis proved to be a more cost effective and time saving method for microsatellite marker development. Both strategies provided markers which, due to null alleles and error rates, should be analysed in pedigrees for observation of inheritance patterns prior to any future studies incorporating them.**

Keywords: Perlemoen, inheritance, marker development, Mendelian segregation, microsatellite markers, null alleles

[Work in this chapter was divided into two main sections: FIASCO marker development and Next Generation (Illumina, 454 Roche) marker development. The markers developed by the FIASCO development strategy was combined with data generated by other members of the laboratory for publication as a Brief note [Slabbert *et al*. 2010 (See Appendix A)].

2. Introduction

Haliotis midae is the only one of six southern African abalone species that is considered as an economic resource. Farming with *H*. *midae* was initiated due to successful spawning of abalone in 1981 (Genade *et al.* 1988), the lowering of TAC (Cook 1998) as a consequence of massive wild population decline due to illegal harvesting, habitat destruction by the rock lobster *Jasus lalandii* (Day and Branch 2002) and the economic viability of this industry. A genetic improvement program was implemented in South Africa in 2005 with the aim of protecting the wild abalone stocks while simultaneously developing improved farming strategies (Brink *et al.* 2009). Various disciplines are involved in such a management program, including molecular genetics. Molecular markers such as RFLPs, RAPDs, AFLPs and more recently microsatellites and SNPs, are extremely useful for identifying variation between individuals and populations for these management strategies. Genetic markers are therefore a crucial means of assisting the industry in reaching their goals of increased and sustainable production. These markers are useful for applications such as parentage analysis (Ruivo 2007; Van den Berg and Roodt-Wilding 2010), monitoring of genetic diversity (Slabbert *et al*. 2009), linkage mapping (Baranski *et al.* 2006a, Badenhorst 2008), QTL-mapping (Baranski *et al*. 2008) and ultimately MAS (Hayes *et al.* 2007).

Microsatellites or STRs have become the marker of choice for the majority of these applications. These polymorphic markers are especially popular as they are multi-allelic, co-dominant, fairly abundant and well dispersed within most genomes (Goldstein and Schlötterer 1999). FIASCO has been the microsatellite marker development strategy of choice in recent years as it provides an enrichment step to increase the number of isolated microsatellite loci. It is based on a partial enrichment library construction followed by repeat motif hybridisation and fragment capture with magnetic beads. A cloning and screening step is subsequently performed for sequencing and microsatellite identification (Zane *et al*. 2002).

Next generation sequencing (NGS) such as 454 pyrosequencing (Roche) and sequence-by-synthesis (Solexa Genome Analyzer 2; Illumina Inc.) now provide a major platform for fast, accurate and cost-effective high throughput large-scale marker development (Mardis 2008; Abdelkrim *et al*. 2009; Allentroft *et al*. 2009; Santana *et al*. 2009; Slabbert 2010) in which raw sequence data is generated by massive parallel sequencing and data collection

(GS-FLX Pyrosequencing: http://www.454.com; Illumina Genome Analyser: http://www.illumina.com). The starting material can be from either genomic or cytoplasmic origin. If messenger RNA (mRNA) is used as the template strands for sequencing and converted to cDNA, it provides information specifically from coding regions. The single reads are thereafter compiled into unigene contigs and subjected to further bioinformatic analyses. This information is thus coding transcriptome sequence (i.e. EST) data and is extremely useful for expression profiling (Karsi *et al*. 2002; Lang *et al*. 2009), analysis of alternative splice forms (Zha *et al*. 2005; Menon *et al*. 2009) and comparative genome analysis (Leu *et al*. 2007; Kucuktas *et al*. 2009).

Unlike FIASCO, NGS excludes the tedious cloning, enrichment and screening steps which make microsatellite marker development a complex and costly affair. Various studies {Abdelkrim *et al.* (2009) [markers developed from the endangered Blue Duck]; Allentroft *et al.* (2009) [used ancient DNA for marker development]; Slabbert (2010) [marker development for preliminary linkage and QTL mapping]; Santana *et al.* (2009) [marker development for cross-species transfer]} have thus far effectively developed microsatellites with the 454 NGS method with extremely promising results. Microsatellite development with sequence-by-synthesis technology (Illumina) is limited thus far (Dempewolf *et al*. 2010); however several SNPs have been developed for several species (Hillier *et al*. 2008; Wall *et al*. 2009; Hyten *et al*. 2010). The reason for this is the short read lengths obtained via Illumina sequence-by-synthesis which hinders the amount of possible primer designing, despite the formation of contig alignments which provides read lengths of approximately 34-40 bp (Mardis 2008).

Pyrosequencing and sequence-by-synthesis can both provide EST information if the sequence data is obtained directly from mRNA (Van Tassell *et al*. 2008). ESTs provide a window of information into expression profiling, comparative analysis and transcriptome structure with the use of EST-STR markers. These markers are regarded as type I markers as they are associated with known genic regions and are more transferable between species than type II anonymous markers. EST-STRs generally have a lower level of polymorphism, however, are valuable for QTL analysis as they provide information of putative proteins (Bouck and Vision 2007; Slabbert 2010).

The repetitive, highly mutable nature and random distribution of microsatellites led to the assumption that these genetic loci are purely accumulated "junk" DNA. It has since been proven, however, that such loci are in fact extremely functional in some instances as regulatory elements and activation protein binding sites. EST-STRs have been found in promoter regions of numerous genes, for example the Insulin *sp*1 protein binding (Araki *et al*. 1987) and c-*KI-ras* in *Drosophila* (Hoffman *et al*. 1990), acting as activators or enhancers; thus resulting in repeat conservation in these instances. The reviews by Li *et al.* (2002; 2004) discuss the functionality of microsatellites providing numerous examples of their importance in most genomes. STRs have been linked to mechanisms such as chromatin organisation, centromere and telomere structure, enzyme activity, gene expression and recombination. The presence of STRs within coding regions causes direct selective influences as a result of shift mutations, notorious in microsatellites, which could result in frame shifts of vital genes. Microsatellites involved in protein coding and genome dynamics are therefore conserved (Li *et al*. 2002; 2004).

Microsatellite markers, irrespective of the development strategy, should be used with caution. Although genotyping and analyses are theoretically simple, STRs are notorious for abnormal behaviour in PCR and electrophoretic separation. Technical challenges and errors such as null alleles (Callen *et al.* 1993; Pemberton *et al.* 1995), large allele drop-out (Jones and Ardren 2003) and PCR artefacts are often seen, leading to the misinterpretation of data which will affect the outcomes of parental assignments as well as population analysis, linkage mapping and QTL mapping. Characterisation of these markers is thus necessary to identify problematic markers as well as those most suited to a specific application. Currently there are 226 microsatellite markers for *H. midae* developed by three main methods namely FIASCO (Bester *et al.* 2004; Slabbert *et al.* 2008; 2010), data mining (Rhode 2010) and next generation sequence screening (Slabbert 2010).

Characterisation of microsatellite markers can be performed by analysing segregation patterns within full-sib families. The inheritance of alleles should follow specific patterns depending on the parental genotypes. For example, if both parent genotypes are heterozygous for four alternative alleles [Parent 1 (A_1A_2) and Parent 2 (A_3A_4)], the offspring will inherit fourpossible genotypes; A_1A_3, A_1A_4, A_2A_3 or A_2A_4 (i.e. 1:1:1:1 segregation pattern). If both parents are heterozygous for two specific alleles [Parent 1

(A_1A_2) and Parent 2 (A_1A_2)], then a 1:2:1 segregation pattern should be observed; A_1A_1, A_1A_2 or A_2A_2. Lastly, if one parent is homozygous and the other heterozygous [Parent 1 (A_1A_1) and Parent 2 (A_2A_3)], a 1:1 segregation pattern should be observed; A_1A_2 or A_1A_3.

Hardy-Weinberg equilibrium (HWE) analysis can also be used to assess the usefulness of a marker. HWE tests use the fixed relationship between allele and genotype frequencies to determine whether the equilibrium or stable state of these two frequencies is reached under certain assumptions over generations. These assumptions include an infinitely large, randomly mating population in which no selection, mutation or migration takes place and where no population stratification is present. The heterozygosity values retrieved from statistical analysis can be useful in obtaining information regarding the informative nature of a marker (Ziegler and König 2006).

Microsatellite marker development was performed using the FIASCO technique and the markers were characterised within a full-sib family using Mendelian segregation chi-squared tests. NGS (sequencing-by-synthesis; Solexa Genome Analyser, Illumina Inc.) was also used to develop microsatellite markers and characterised using Hardy-Weinberg analysis in wild populations.

3. Materials and methods

3.1 Genomic DNA extractions

Various *Haliotis midae* broodstock were individually spawned and crossed during the 2006 season to create a number of full-sib and half-sib families on Roman Bay Sea Farm (Gansbaai, South Africa) and HIK Abalone Farm (Hermanus, South Africa). Thirty two offspring from two of these full-sib families, consisting of >1000 individuals [Family 7B (Roman Bay) and Family 42A (HIK)], were selected for FIASCO marker characterisation. Thirty two unrelated, randomly selected individuals from two separate wild populations, namely Witsand and Saldanha were used for characterisation of the markers developed via sequence-by-synthesis Illumina technology.

The extraction of DNA was performed using the cetyltrimethyl ammonium bromide or CTAB extraction method (Saghai-Maroof *et al.* 1984), from epipodial tentacles (Slabbert and Roodt-Wilding 2006). The tissue was homogenised in 300 µl CTAB lysis buffer (1.4 M

NaCl; 20 mM Ethylenediamine tetra-acetate (EDTA [pH 8]); 2% (w/v) CTAB; 100 mM Tris-HCl [pH 6.8] and 0.2% (w/v) β-mercapto-ethanol) as well as 0.5 mg/ml Proteinase K. The tissue was subsequently incubated overnight in a water bath set at 60℃. Equal volumes (300 µl) of chloroform: isoamylalcohol (24:1) were added to the homogenised solution and shaken on a vortex for 5 min at a low speed. The samples were centrifuged at 12000 rpm for 5 min using an Eppendorf Centrifuge 5415D set at 4℃. The supernatant was carefully removed and placed in a new eppendorf tube for subsequent DNA precipitation via the addition of 2/3 volume isopropanol and incubation at -20℃ overnight. The samples were centrifuged at 12000 rpm for 20 min and the pellet washed with 70% (v/v) ethanol, followed by a second centrifugation step. The alcohol was removed and the pellet dried in an oven at 55℃. One hundred microlitres double dis tilled water (ddH_2O) was added to re-suspend the extracted DNA. The DNA was stored at -20℃ until required.

3.2 Microsatellite isolation: The Fast Isolation by AFLP of Sequences COntaining Repeats (FIASCO)

3.2.1 DIGESTION - LIGATION

An enrichment partial genomic library was constructed by performing a digestion-ligation reaction simultaneously in a 25µl reaction containing 250ng gDNA, 1X One-Phor-All Buffer, 5mM DTT (Dithiotreitol), 50µg/ml BSA (Bovine serum albumin), 1 µM *Mse*I-N adaptor, 200 µM Adenosine triphosphate (ATP), 2.5 U *Mse*I restriction enzyme and 1U *T4*-ligase. The reaction was incubated at 37℃ for 3 ho urs in a water bath to allow DNA digestion with *Mse*I restriction enzyme to take place and subsequent ligation of *Mse*I-N adaptors. The reaction product was then diluted (1:10) for further isolation procedures.

3.2.2 AFLP AMPLIFICATION ONE

Amplification of repeat sequences was performed using AFLP-adaptor specific primers (*Mse*I-N: 5'-GAT GAG TCC TGA GTA AN-3') in a PCR (Zane *et al*. 2002). These primers are a 1:1:1:1 preparation of four AFLP specific primers namely *Mse*I-1, *Mse*I-2, *Mse*I-3 and *Mse*I-4. A 20 µl PCR reaction was performed (5 µl digestion/ligation solution, 1X $MgCl_2$-free Buffer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 120 ng *Mse*-IN and 0.4 U *Taq* polymerase) using the following thermocycler conditions: 94℃ for 5mins; 94℃ for 30s, 53℃ for 1 min, 72℃ for 1 min (for 16, 18, 20, 22 and 24 cycles); and an elongation at 72℃ for 5 min. The annealing steps were performed at five varying cycles to identify which produced the most optimal amplification. The PCR was performed on a GeneAmp® PCR system 2700 thermocycler and the product was visualised on a 2% (w/v) ethidium

bromide-stained agarose gel using a UV light on a MultiGenius BioImaging System (Syngene) at the Genetics Department (Stellenbosch University).

3.2.3 HYBRIDISATION OF gDNA AND THE CAPTURING THEREOF

Biotinylated $(AC)_{12}$ and $(GATC)_6$ probes were used in separate hybridisation reactions and repeat sequences were isolated. The hybridisation reaction was performed in a 100µl volume reaction containing 250-500 ng DNA from the PCR product, 50-80pmol $(AC)_{12}$ or $(GATC)_6$ probe, 4.2X SSC (Standard Saline Citrate) and 0.07% (w/v) SDS (Sodium dodecyl sulfate). An initial denaturation step was performed by placing the reaction at 95℃ for 3 min and the probe subsequently hybridised at 25℃ for 15 min. Three hundred microlitre TEN_{100} buffer (10 mM Tris-HCl [pH 8], 1 mM EDTA [pH 8], 100 mM NaCl [pH 8]) was used to dilute the hybridisation mixture in order to capture the annealed DNA fragments.

One hundred milligram (100 µl) streptavidin coated magnetic beads (Sigma) were used for selective capturing of hybridised DNA fragments due to the attractive binding abilities between biotin and streptavidin. The beads were washed four times using an equal volume of TEN_{100} which was thereafter re-suspended in 40 µl of the aforementioned buffer. An addition of unrelated (non-abalone) DNA (10 µl) was required for reduction of non-specific binding of gDNA. Beads containing unrelated DNA were added to the hybridisation solution which was incubated at room temperature for 30 min while shaking on a vortex at low speed. The hybridisation buffer was removed by using a magnetic particle collector to fix the bead-probe complex and the remaining liquid subsequently discarded. Three non-stringency (NS) and three stringency (S) washes were performed for removal of non-specific binding. The NS washes were performed by adding 400 µl TEN_{1000} (10 mM Tris-HCl [pH8], 1 mM EDTA [pH 8] and 1 M NaCl [pH 8]) to the bead complex and mixing on a vortex at low speed for 5 min. A magnetic particle collector (Sigma) was used to fix the bead complex while the supernatant was discarded. This step was repeated three times, however on the last wash the supernatant was kept in a clean eppendorf tube and stored at -20℃ for later use. The S washes wer e performed using the same protocol as the NS washes, however 400 µl of 0.2X SSC/ 0.1% (w/v) SDS was used as opposed to TEN_{1000}.

Two denaturation steps were carried out (see Zane *et al.* 2002). The first (D1) was performed in 50 µl of TE (10 mM Tris-HCl [pH8], 1 mM EDTA [pH8]) solution and was placed at 95℃ for 5 min and thereafter immediately on ice. The supernatant was removed using a magnetic particle collector (Sigma) to fix the beads and kept at -20℃ for later use. The second denaturation step (D2) was performed using 15 µl of 0.15 M NaOH. The supernatant of the second denaturation step was removed and neutralised on ice with 1µl of 0.1667 M CH_2COOH (Acetic acid) and stored at -20℃.

One volume of isopropanol and 0.15 M Sodium acetate (NaAc) was added to the supernatant obtained from the four washes (NS, S, D1 and D2). Tubes were placed at -20℃ overnight for DNA precipitation. Centrifugatio n at 12000 rpm for 30 min was used to pellet the DNA followed by drying and resuspension in 50 µl of ddH_2O. The samples were electrophoretically separated on an agarose gel for confirmation of DNA fragments present.

3.2.4 AFLP AMPLIFICATION TWO

The second AFLP amplification reaction was performed using 2 µl of the captured DNA (D2) in a 20 µl PCR reaction containing 1X $MgCl_2$-free Buffer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 120 ng *Mse*I-N and 0.4 U *Taq* polymerase. The following thermocycler conditions were used: 94℃ for 5 mins; 94℃ for 30s , 53℃ for 1min, 72℃ for 1 min for 30 cycles; and an elongation step at 72℃ for 5mins. P CR reactions were performed on a GeneAmp® PCR system 2700 thermocycler and results analysed on a 2% (w/v), 1X Tris-Borate-EDTA buffer (TBE), ethidium bromide gel with exposure to UV light.

3.2.5 CLONING OF ISOLATED MICROSATELLITES

Fragments were cloned into a TOPO®TA cloning vector and the ampicillin resistant, recombinant *E. coli* cells selected. The cloning reaction was performed containing 0.5 - 4.0µl fresh PCR product, 1 µl salt solution (1.2 M NaCl and 0.06 M $MgCl_2$), 1 µl TOPO®TA cloning vector and the volume brought to 6 µl using sterile water. The reaction was placed at room temperature for 30 min and subsequently placed on ice. The competent cells were chemically transformed by the addition of 2 µl cloning reaction, mixing gently and incubating on ice for 30 min. The cells were heat shocked for plasmid uptake at 42℃ for 30 s and placed immediately on ice. Two hundred and fifty microlitre of the transformed cells were added to SOC-medium [2% (w/v) Bacto-tryptone, 0.5% (w/v) Yeast Extract, 0.05% (w/v) NaCl, 10 ml 250 mM KCl, 18 ml 20% (w/v) sterile Glucose and 5 ml 2M sterile

$MgCl_2$ at pH 7] and shaken at 175 rpm for 1 hour at room temperature. Ten microlitres, 20µl and 50µl of transformed cells/SOC-medium solution was then spread onto respective LB (Luria-Bertani) ampicillin agar plates and incubated at 37℃ overnight.

3.2.6 COLONY SCREENING WITH PCR

Ten to twenty single colonies were picked from the LB Ampicillin agar plates and used in a colony PCR reaction with vector-specific primers, M13 forward (5'-GGTTTTCCCAGTCACGAC-3') and M13 reverse (5'-GGAAACAGCTATGACCATG-3') in order to identify cells containing a DNA insert. Single colonies were placed into a PCR mix containing 1X Go*Taq* Buffer, 1.5 mM $MgCl_2$, 0.8 mM dNTPs, 5 µM M13 forward and reverse primer and 0.2 U *Taq* polymerase in a 20 µl reaction volume. The PCR cycle consisted of a 10 min denaturation step at 94℃ fol lowed by 25 cycles of 94℃ for 1min, 55℃ for 1 min and 72℃ for 1 min and a final elong ation step of 72℃ for 10 min. The result was visualised on a 2% (w/v), 1X TBE, ethidium bromide gel under UV light.

3.2.7 SEQUENCING AND ANALYSIS

The colonies showing a positive PCR amplification were sequenced in a 10 µl reaction containing 9 ng DNA, 2 µl BigDye v3 terminator and 1.6 pmol primer (M13 Forward or M13 Reverse). The sequencing products were electrophoretically separated on an ABI 3730xl DNA Analyser (Applied Biosystems) with the M13-forward (5'-GTAAAACGACGGCCAGT-3') vector-specific primer.

Sequence scanner v1 (Applied Biosystems) was used to identify and remove vector and primer sequences. Sequences were subsequently analysed with BatchPrimer v3 (You *et al.* 2008) to search for specific microsatellite repeats and simultaneously design primers.

Once microsatellites were identified, all the sequences containing repeats were aligned using a BLASTn search in BioEdit v7.0.9.0 (Hall 1999) against National Center for Biotechnology information (NCBI) as well as the local laboratory microsatellite database for *H*. *midae*. This was carried out to identify whether any homology between previously developed markers was seen. Any redundant markers were excluded from further analyses and development.

3.3 Microsatellite development using sequence-by-synthesis

Twenty four thousand three hundred and sixty assembled contig sequences were developed from the Illumina data by Van der Merwe (2010). The contigs were formed using Velvet v0.7.57 (Zerbino and Birney 2008) with parameters set as per Van der Merwe (2010). These sequences were placed into the internet based program BatchPrimer v3 (You *et al.* 2008) after conversion to a single FASTA format file containing all 24360 sequences. Contig sequences less than 100 bp were removed from the file as primer designing becomes challenging and not optimal when limited flanking regions are present. BatchPrimer v3 allows specific parameters to be set to ensure that the most optimal primers are designed (Table 2.1).

Table 2.1: Parameters set for BatchPrimer v3.

Parameter	**Value**
Primer length	16 – 25 (21 as optimal)
GC-content	30 – 70% (50% as optimal)
Annealing temperature	50 – 70 (60 as optimal)
Product size	90 – 350 (120 as optimal)
Other parameters	Default

Tetra-, penta- and hexanucleotide sequences with at least three contiguous repeats were selected for further characterisation as longer repeat units generally provide a higher level of polymorphism (Amos 1999).

The selected sequences containing microsatellite repeats were queried against the NCBI GenBank database (www.ncbi.nlm.nih.gov) using BLASTn to identify possible homology to expressed regions of related species. A megaBLAST (highly similar) alignment was used against the nucleotide collection database and the default BLAST algorithm settings were applied. A significant alignment was accepted when an S (score) value >50 and an E value <1e-04 was obtained (Farber and Medrano 2004).

3.4 Microsatellite amplification and analysis of polymorphism (FIASCO and sequence-by-synthesis primers)

3.4.1 PRIMER DESIGN AND OPTIMISATION OF NOVEL MICROSATELLITE LOCI

Primers were designed for amplification of a particular microsatellite locus if sufficient flanking region was available from the sequenced fragment. Oligonucleotides were developed using BatchPrimer v3 software and chosen based on lowest primer dimer probability, an optimal length of 20 bp and an optimal annealing temperature of 60℃ (Table 2.1).

Touchdown PCR (TD-PCR) with varying $MgCl_2$ and primer concentrations was initially attempted for all microsatellite loci. The initial optimisation was attempted in a 10 µl reaction containing 20 ng DNA, 1X *Taq* Buffer, 2 mM $MgCl_2$, 200 µl each dNTP, 0.25 U Go*Taq* polymerase and 2 µM of each primer. The primer, $MgCl_2$ and DNA concentrations as well as annealing temperatures were altered according to the initial result seen on a 2% (w/v) agarose gel. The PCR cycle consisted of an initial denaturing step of 94℃ for 5 min followed by 10 cycles of 30 s at 94℃, 30s at 65℃ and 30s at 72℃. The annealing temperature was lowered by 1℃ in consecutive cycles, until an annealing temperature of 55℃ was reached. This was then followed by 30 cycles of 30 s at 94℃, 30s at 55℃ and 30 s at 72℃, with a final extension for 7 min at 72℃. Thereafter, gradient PCR was performed to identify the most optimal annealing temperature of each primer pair (Rahman *et al*. 2000).

The Kapa2G FAST PCR cycle was used for the challenging EST-STR primers with the following cycle: The initial denaturation was performed at 95℃ for 2 min followed by 30 cycles of 94℃ for 1 s and 62℃ (annealing) for 5 s . The final extension step was performed at 72℃ for 10 s.

The optimised microsatellite markers were tested for polymorphism. This was performed using the DNA of 8 randomly chosen individuals using poly-acrylamide gel electrophoresis (PAGE). The poly-acrylamide gels (12% w/v) were prepared as follows: 3 ml Acrylamide/Bis-acrylamide (49:1 ratio [Sigma]) was added to 2 ml 5X TBE and 5 ml distilled water. Eight hundred microlitres of 10% (w/v) APS was subsequently added to the solution followed by 16 µl TEMED (Sigma). The gel was assembled in a Mighty Small apparatus (BioRad).

3.4.2 PRIMER LABELLING AND OPTIMISATION

Microsatellite repeats which showed polymorphism were labelled with different fluorescent labels (VIC, FAM, NED, PET) for genotype analysis. Labelling was dependant on the size of the product and the distance of the primer from the repeat unit. Primers which had products with overlapping sizes were assigned different labels for observable differentiation in multiplex genotype analysis. The primer chosen for labelling was based on three criteria; 1) the primer closest to the repeat sequence to reduce the amount of stutter, 2) the longest primer of the two and three) the primer containing a GC clamp on the 3' end to ensure secure binding.

The same reaction conditions were used for the fluorescently labelled primers as for the unlabelled counterpart; however, additional optimisation was necessary in some cases (see section 3.4.1; current chapter). PCR amplifications were performed using a GeneAmp® PCR system 2700 thermocycler.

3.4.3 PCR MULTIPLEXES

Microsatellite markers were placed into multiplex PCR reactions, taking into account the particular dye used to label each primer and the size range of the expected products (Figure 2.1). A QIAGEN® Multiplex kit was used to amplify the target loci following the instructions provided by the manufacturer, with slight modifications to conserve and extend the use of the kit. The reactions were performed as follows: 10 ng of template DNA was added to 3.5 µl 2X QIAGEN Multiplex PCR master mix (containing HotStart*Taq*® DNA Polymerase, Multiplex PCR Buffer with 6 mM $MgCl_2$ and dNTP Mix) (QIAGEN®), 0.9 µl Primer mix (20 µM of each primer) and ddH_2O to a final volume of 7 µl. The following PCR cycle was used to amplify the target locus: The cycle is initiated with a 10 min denaturing step at 95℃ for 15 min, followed by 35 cycles of 9 4℃ for 30s, 57℃ for 90 min and 72℃ for 1min. The PCR completed with an elongation step of 60℃ for 30 min.

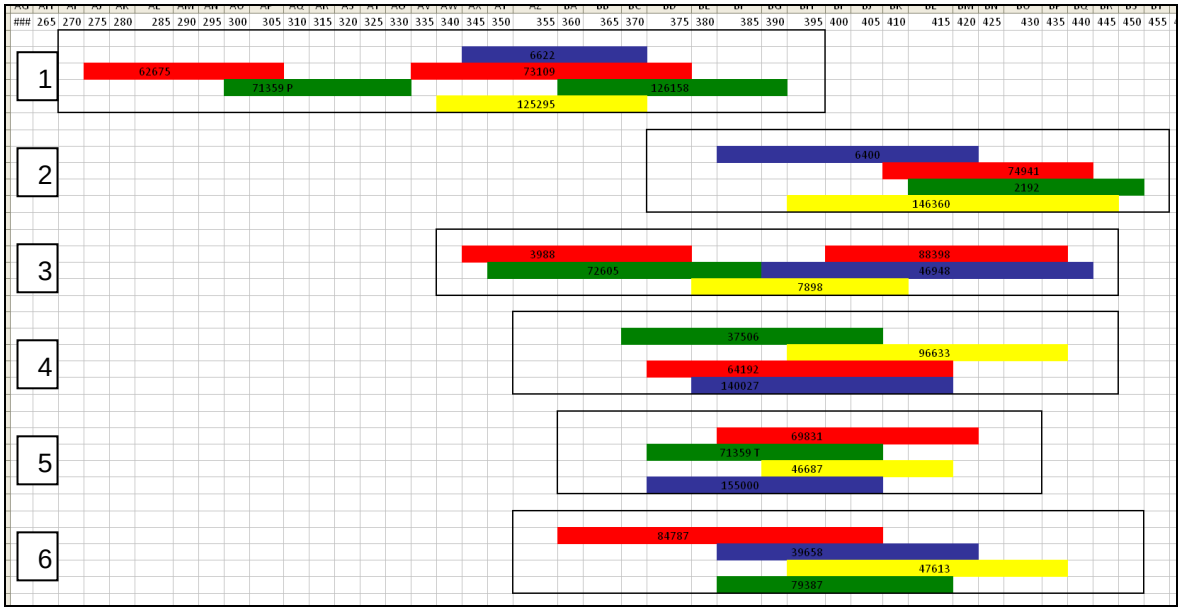


Figure 2.1: Preparation of microsatellite multiplexes based on size range and dye colour.

1- *Hmid*ILL1-006622, -062675, -073109, -071359P, -126158, -125295
2- *Hmid*ILL1-006400, -074941, -002192, -146360
3- *Hmid*ILL1-003988, -088398, -072605, -046948, -007898
4- *Hmid*ILL1-037506, -096633, -064192, -140027
5- *Hmid*ILL1-069831, -071359T,-046687,-155000
6- *Hmid*ILL1-084787, -039658, -047613, -079387

3.4.4 GENOTYPING OF MICROSATELLITE MARKERS

The FIASCO markers were subjected to segregation analysis using 32 full-sib offspring from the two families, 7B and 42A, sampled at Roman Bay (Gansbaai, South Africa) and HIK (Hermanus, South Africa) respectively. Markers obtained from the Illumina sequence-by-synthesis method were analysed in 32 unrelated, randomly selected individuals from two separate wild populations namely Witsand and Saldanha.

The markers were genotyped in PCR multiplexes on the ABI 3730xl DNA Analyser (Applied Biosystems) and the lengths determined by comparison to the GeneScan™ 600 LIZ® Size Standard (Applied Biosystems). Analysis of genotypes was performed using GeneMapper v4.1 software (Applied Biosystems).

3.4.5 STATISTICAL ANALYSIS

Characterisation of FIASCO developed microsatellite markers was performed by analysing segregation patterns within the full-sib families. Segregation analysis was performed for each locus in both families by comparing the observed ratios to the expected Mendelian segregation ratios (1:1:1:1, 1:2:1 or 1:1) using a chi-square test.

The EST-STR loci which provided reliable genotypes were subjected to numerous statistical tests. HWE exact tests were performed using Genepop v4 (Rousset 2008). If less than four alleles were present in the dataset, the complete enumeration method was used to calculate *p* (probability) values. If more than four alleles were detected, a Markov Chain algorithm (20 batches with 1000 dememorisations per batch) proved to be more accurate as it includes a standard error. Allele frequencies, observed and expected heterozygosities, as well as F_{IS}-statistics (heterozygous excess or deficiency) were calculated. Null allele frequencies were calculated in Genepop v4 (Rousset 2008) using an adapted ML method, developed by Brookfield (1996), which distinguishes between true null alleles and potential genotype failure.

4. Results

Two separate methods of microsatellite marker development were employed to assess the success rates and identify various advantages and disadvantages of both strategies. The two methods include the traditional enrichment clone-based FIASCO method versus marker development via screening of sequence-by-synthesis (Illumina GA) contigs.

4.1 Marker Development using FIASCO – (Slabbert *et al*. 2010)

The genomic DNA extractions were performed using the CTAB method as explained in section 3.1. The concentration obtained from tentacle samples was approximately 60-100 ng/µl for each sample.

The FIASCO marker development technique was performed to obtain non-redundant microsatellite-containing sequences. An AFLP amplification step was carried out subsequent to the digestion-ligation step. The annealing temperature of this PCR step was performed at five varying cycles (16, 18, 20, 22, 24) to identify which produced the most optimal amplification for a range of fragment sizes. Figure 2.2 depicts the “smears” obtained at the various annealing temperatures and in this case, 16 or 18 cycles were used for further amplification steps. The success of the hybridisation of biotinylated probes to fragmented DNA and the subsequent capturing thereof (via four washing steps) was visualised by gel electrophoresis (as seen in Figure 2.3).

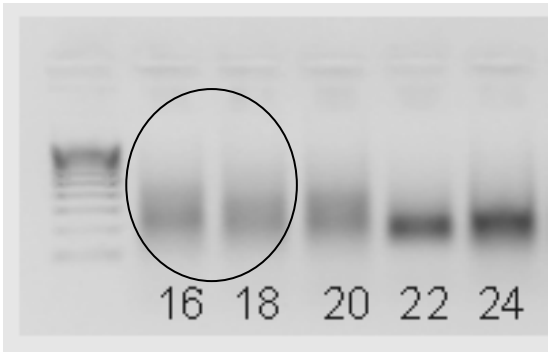


Figure 2.2: AFLP amplification 1 result (FIASCO). Variable annealing cycle lengths used to obtain a range of fragment sizes (16 or 18 cycles provided the best amplification in this example).

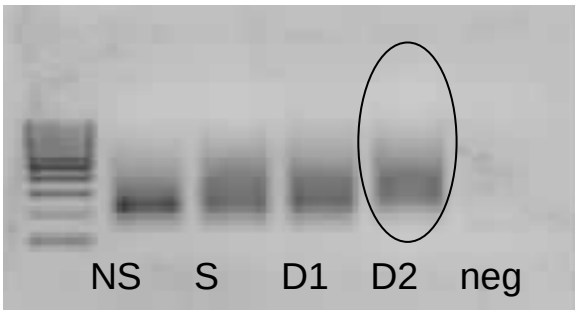


Figure 2.3: Gel electrophoresis result subsequent to the washing steps (FIASCO). NS-Non stringent (TEN_{1000}); S-Stringent (0.2X SSC/ 0.1% SDS); D1-Denaturing step 1; D2-Denaturing step 2; neg-negative control with no DNA present.

Once the repeat-containing fragments were isolated, they were cloned into TOPO®TA cloning vectors and consequently transformed into ampicillin-resistant *E*. *coli* cells. This was effectively observed using ampicillin-containing agar medium for selective growth of transformed cells.

A total of 978 recombinant clones were selected from the ampicillin-containing LB plates (Figure 2.4) and sequenced. Of these clones, 49% contained repeat motifs. The fragments were edited with Sequence Scanner v1 (Applied Biosystems), leaving a total of 222 (23%) primer pairs which could be designed. Of these designed primer pairs, 44 were found to be polymorphic (20%).

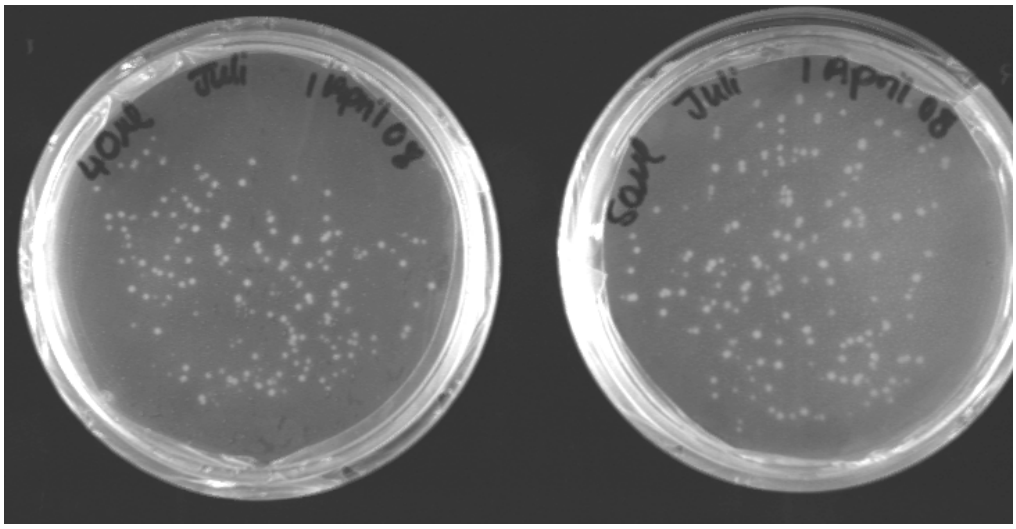


Figure 2.4: *E. coli* cells growing under ampicillin selection, prior to sequencing. These colonies were selected for PCR screening.

These 44 polymorphic markers were characterised within 32 offspring of the two full-sib families (7B - Roman Bay; 42A - HIK). From a total of 88 marker-family combinations (44 x 2), 38 proved to be informative, 18 combinations had only one allele type for both parents (monomorphic), 8 combinations had three or more alleles while 24 combinations could not be reliably scored or amplified and were subsequently discarded from further analyses. From the 38 informative marker-family combinations, 28 (32%) conformed to expected Mendelian segregation patterns (1:1:1:1, 1:2:1, 1:1; $p>0.05$) using a chi-square test [see section 3.5.1; current chapter], while 10 combinations did not (Appendix B).

The majority of markers providing reliable genotypes gave segregating patterns that were easily discernable, except for three (*Hmid*2044T, *Hmid*0006M and *Hm*LCS147T) which displayed null allele segregation patterns. Figure 2.5 depicts the observable null allele segregation in *Hmid*2044T and *Hmid*0006M.

Eight family-marker combinations displayed duplicated loci (i.e. an individual will possess more than two alleles): *Hmid*0166R in Family 7B and 42A, *Hm*NSS1H in Family 42A, *Hm*NS18M in Family 7B, *Hm*LCS152M in family 42A, *Hm*LCS71T in Family 42A, *Hm*LCS58M in Family 7B and *Hm*NS32M in Family 42A.

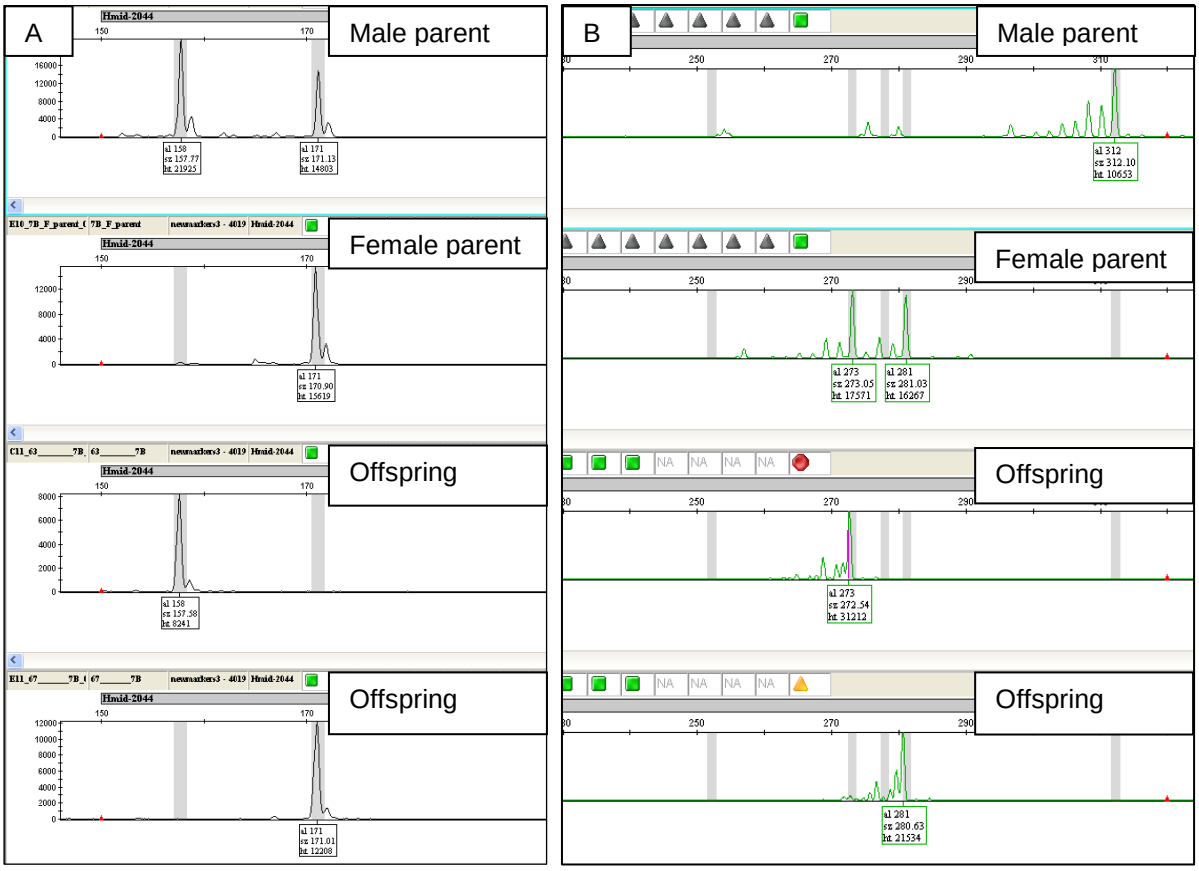


Figure 2.5: The segregation of null alleles in markers *Hmid*2044T and *Hmid*0006M
A-Null allele segregation of marker *Hmid*2044T. The offspring genotypes 171/171 and 171/null cannot be distinguished. B-Null allele segregation of marker *Hmid*0006M. The male parent possesses a null allele, which can easily be identified by observing the segregation pattern of the offspring genotypes.

4.2 EST microsatellite marker development using next generation sequencing

Twenty four thousand three hundred and sixty sequences from the Illumina contig data were subjected to microsatellite motif searches with the use of BatchPrimer v3 (Hall 1999). Eight hundred and sixty nine STRs were identified, however, only 307 [35%] (including di- and trinucleotide motifs) of these could be used to design primers due to the restricted flanking region sequence of these loci. Ten hexanulcleotide, 19 pentanucleotide and 24 tetranucleotide microsatellite loci were selected from the 307 sequences after elimination of fragments which were extremely short (less than 100 bp). After removal of redundant loci with the use of BLASTn alignments against the nucleotide NCBI database, seven hexanucleotide, 15 pentanucleotide and 10 tetranucleotide microsatellite sequences (60%) remained for further characterisation.

From the 32 aforementioned expressed sequence tag (EST)-STRs, 30 could be successfully optimised with either Go*Taq* Polymerase (Promega) using a general

Touchdown PCR or alternatively, Kapa2G *Taq* polymerase (KAPA Biosystems) using a Fast Touchdown PCR cycle, as suggested by the manufacturer (Table 2.2). These primer pairs were subjected to PAGE subsequent to PCR amplification, using eight randomly selected individuals in order to identify potential polymorphism. Twenty seven fluorescently labelled primers were designed after selecting candidate polymorphic loci from the PAGE results. The primer labels were chosen according to possible multiplex groupings (see section 2.2.6 and Figure 2.1; current chapter), of which 22 labelled primers could be successfully optimised (Table 2.2) using the previously established optimised PCR conditions.

Fourteen EST-STR loci provided reliable genotype results after amplification in 32 randomly selected individuals from the wild population (Saldanha and Witsand) sample groups. These were subjected to statistical analyses: Hardy-Weinberg probability tests, maximum likelihood estimation of null allele frequencies [EM algorithm; Dempster *et al*. 1977] and F_{IS} values [Weir and Cockerham 1984] (Table 2.3). Four markers (*Hmid*ILL1-006400, *Hmid*ILL1-126158, *Hmid*ILL1-96633 and *Hmid*ILL1-003988) were monomorphic. Of the ten EST-STRs remaining, four [*Hmid*ILL1-006622, *Hmid*ILL1-046948, *Hmid*ILL1-064192 and *Hmid*ILL1-146360] did not conform to HWE ($p<0.05$). The remaining six loci [*Hmid*ILL1-007898, *Hmid*ILL1-047613, *Hmid*ILL1-002192, *Hmid*ILL1-088398, *Hmid*ILL1-140027 and *Hmid*ILL1-037506] did conform to HWE expectations ($p>0.05$), even though *Hmid*ILL1-002192, *Hmid*ILL1-088398, *Hmid*ILL1-140027 and *Hmid*ILL1-037506 displayed high null allele frequencies. The average observed and expected heterozygosities were 0.366 and 0.453, respectively. The average F_{IS} displayed a positive value of 0.178.

Table 2.2: Primer information of twenty two optimised EST-STR markers. Optimised PCR conditions, the type of repeat as well as primer sequences (labelled primers also shown)

Marker name	Optimised conditions	Motif	Primers
*Hmid*ILL1-006400H	56℃ (G)	$(GCACA)_3$	F: GGCAACTAATGGAGAAGCTG R: TATCTCGAAGGCTCGTAAGC#
*Hmid*ILL1-007898P	58℃ (G)	$(AATTT)_3$	F: CCAGACACCACACAAAGCTA+ R: GTGAGCATCTGCCTTCCTT
*Hmid*ILL1-037506P	56℃ (K2G)	$(AGGTC)_3$	F: TGTACACATAGCCTGCTTGC R: TTCCAGTGCTGGTTGTGTAG^
*Hmid*ILL1-046948H	61℃ (K2G)	$(TCACCA)_4$	F: CTGGCTGGTCATGGTTATG R: CAGGGATGCAGAGATAATGC#
*Hmid*ILL1-062675P	58℃ (G)	$(TGAGA)_3$	F: CACTCAGCTACGTTCTGCAT* R: GTAGTGGGACCAAGCTCACT
*Hmid*ILL1-064192P	61℃ (K2G) F	$(AAATA)_3$	F: GGCTACTTCACCGCCTTTA R: TCCATGCAAGTAGGCTAGGT*
*Hmid*ILL1-069831H	58℃ (G)	$(CACCGA)_3$	F: CTACACATCACTGAGCAGCA* R: GACAGAGTCGGTCTTCACG
*Hmid*ILL1-072605P	58℃ (G)	$(AGGTG)_4$	F: TTCGGTGTCACCCTTGAT R: CCAATTTCACCTCACCTCAC^
*Hmid*ILL1-073109H	60℃ (K2G) F	$(TTGTTC)_3$	F: CACCTGATGACGGAGTAACA R: AGGGAAGGCTGAAGTGAAC*
*Hmid*ILL1-088398P	58℃ (G)	$(TTTTG)_3$	F: CCCTCTTACAAAACGTCTCC R: TGTAGACTTGTCCCTTGCAG*
*Hmid*ILL1-126158P	58℃ (G)	$(CTTCA)_3$	F: GACTTGGTAACTCCGACAGG R: AGTAGAACGCACGCATTACC^
*Hmid*ILL1-146360P	60℃ (G)	$(TTCTT)_5$	F: GTTAGTGTACGCCGTTAGGG R: GACCAATTCCAACCCAGA+
*Hmid*ILL1-155000P	58℃ (K2G) F	$(ATTTG)_3$	F: ACGAGGAGCAGCTTCAAA R: GCGGTATAGAAAGCTCCACT#
*Hmid*ILL1-002192T	58℃ (K2G) F	$(ATAC)_4$	F: CATCAGAAGAGCACAGTTGG^ R: TCCCCCTGTAATCTCAGAAC
*Hmid*ILL1-003988T	TD (K2G) F	$(GAGT)_4$	F: ATGGAGCTCTTCCTCTTCCT* R: GGCCAGTTAAGTTCTTCTGC
*Hmid*ILL1-006622T	61℃ (G)	$(TTAT)_4$	F: ACGATCAATGCTCCATCC# R: CACAATCTGCCTCTTACCTG
*Hmid*ILL1-039658T	61℃ (K2G) F	$(CTGC)_4$	F: CGCCTCAATTCCTTACTCTC R: CTTTGGATCTGGCTATGGAG#
*Hmid*ILL1-046687T	60℃ (G)	$(TGAG)_4$	F: CGCCAAAGATTACCATGC R: ATACGCCGTGATACATCTGG+
*Hmid*ILL1-047613T	58℃ (G)	$(ACAG)_5$	F: GTGGTGTTTACAAGGCGTCA+ R: CCCGGACTTTAACCATGTCT
*Hmid*ILL1-074941T	TD (K2G) F	$(GGCA)_4$	F: ACCAACTGGCTTTCAGAGG* R: CCAATCCACACACTAGCTGA
*Hmid*ILL1-079387T	62℃ (K2G)	$(AAGA)_4$	F: GGAGATTCTGCGAGTCAAGT R: TAGCTTCCTCCACATCCATC^
*Hmid*ILL1-084787T	56℃ (K2G)	$(GCCA)_5$	F: GAGGTTCGGCTCATACAAAG* R: CACCTGGGCAAGTATTCATC

- Blue fluorescent dye (FAM), * - Red fluorescent dye (PET), ^ - Green fluorescent dye (VIC), + - Yellow fluorescent dye (NED), K2G – Kapa 2G *Taq* Polymerase, G – Go*Taq* Polymerase, F – Fast TD PCR: Initial denaturation at 95℃ for 2min; followed by 30 cycles of 94℃ for 5s, X℃ gradient for 10s and 72℃ for 5s; and lastly an elongation at 72℃ for 2min, TD – Touch down PCR: annealing temp range: 55℃-65℃.

Table 2.3: EST-STR statistical analysis results. Allele frequencies, observed (H_o) and expected heterozygosities (H_e) as well as F_{IS} statistics, null allele frequencies and Hardy-Weinberg equilibrium (HWE) *p* values are displayed

Primer name	Allele frequency	H_o	H_e	F_{IS}	Null allele frequency	*p* value (0.05) HWE * *P*<0.05: distorted
*Hmid*ILL1-006622	0.0338-0.3663	0.438	0.498	0.1241	0.6000	0.0440*
*Hmid*ILL1-007898	0.1667-0.8333	0.281	0.239	-0.1818	0.0000	1.0000
*Hmid*ILL1-046948	0.1250-0.5417	0.188	0.444	0.5831	0.4812	0.0000*
*Hmid*ILL1-047613	0.0161-0.2258	0.875	0.855	-0.0244	0.0633	0.6844
*Hmid*ILL1-064192	0.1935-0.8065	0.125	0.307	0.5973	0.8614	0.0042*
*Hmid*ILL1-002192	0.1607-0.8393	0.219	0.240	0.0913	0.8452	0.5332
*Hmid*ILL1-146360	0.0161-0.1613	0.625	0.898	0.3072	0.1768	0.0000*
*Hmid*ILL1-088398	0.1522-0.6087	0.313	0.403	0.2294	0.6604	0.1130
*Hmid*ILL1-140027	0.1207-0.8793	0.219	0.196	-0.1200	0.8710	1.0000
*Hmid*ILL1-037506	0.0217-0.8913	0.125	0.145	0.1415	0.8882	0.0864

Sixteen of the 32 EST-STRs (50%) were homologous to previously identified genic sequences (Table 2.4). Eleven of these loci were optimised and labelled for further analysis, of which six provided reliable genotype data.

Table 2.4: BLASTn alignments of EST-STR loci to genic regions of related species. E value, Coverage, Identity and Score values are displayed in which S>50 and E<1e-04 were regarded as significantly homologous.

Sequence ID	Homology-BLASTn	E-value	Coverage	Identity	Score
Seq index:2779 NODE 34335*	*Haliotis rufescens* vitelline envelope zona pellucida domain protein 18 (VEZP18) mRNA, complete cds	0.00E+00	83%	81%	351
Seq index:3203 NODE 37506^	*Haliotis discus discus* Ras-related protein Rab-1A mRNA	0.00E+00	83%	95%	1832
Seq index:4471 NODE 45685*	*Haliotis discus discus* calcineurin B mRNA, complete cds	3.00E-171	81%	98%	330
Seq index:4671 NODE 46948	*Haliotis asinina* HasCL444Contig1, mRNA sequence	0.00E+00	66%	92%	382
Seq index:7631 NODE 62675	*Haliotis asinina* HasCL213Contig1, mRNA sequence	2.00E-81	35%	87%	169
Seq index:9530 NODE 72605	*Haliotis rubra* clone Hrub9.H03 microsatellite sequence	2.00E-23	11%	95%	64
Seq index:12430 NODE 88398^	*Haliotis asinina* HasCL262Contig1, mRNA sequence	0.00E+00	23%	93%	743
Seq index:13822 NODE 96633*	*Haliotis diversicolor supertexta* beta-1-3 glucan recognition protein mRNA	0.00E+00	70%	91%	473
Seq index:18386 NODE 125295*	*Haliotis asinina* HasCL146Contig1, mRNA sequence	2.00E-110	12%	86%	222
Seq index:20401 NODE 140027^	*Danio rerio* RAP1B, member of RAS oncogene family (RAP1B) mRNA	3.00E-28	3%	85%	163
Seq index:124 NODE 2192^	TSA: *Haliotis asinina* HasCL469Contig1, mRNA sequence	0.00E+00	11%	97%	363
Seq index:404 NODE 6622^	HtEST186 subtracted libraries from abalone exposed to bacterial challenge *Haliotis tuberculata* cDNA similar to LIM domain-binding protein, mRNA sequence	0.00E+00	29%	90%	463
Seq index:3496 NODE 39658	*Haliotis asinina* developmental microarray library *Haliotis asinina* cDNA clone 788 5', mRNA sequence	5.00E-83	14%	91%	172
Seq index:9987 NODE 74941	CCAG22710.b1 CCAG *Petrolisthes cinctipes* heart, gills, whole crab (N) *Petrolisthes cinctipes* cDNA	3.00E-74	54%	75%	156
	abalone2-050716_D07_d7_07.ab1 *Haliotis discus hannai* hyposalinity subtracted cDNA library *Haliotis discus hannai* cDNA	5.00E-62	13%	93%	134
Seq index:10753 NODE 79387	hasininaP0011N11_224 Adult tropical abalone mantle (*Haliotis asinina*)	4.00E-96	24%	95%	195
Seq index:11767 NODE 84787	hasininaP0023I03_504 Adult tropical abalone mantle (*Haliotis asinina*)	0.00E+00	13%	91%	356

* Microsatellites that did not amplify after optimisation attempts ^ Markers which were optimised and labelled for further analysis
__ Markers which were mapped to a linkage group (Chapter 3)

5. Discussion

5.1 Marker development using the FIASCO technique

Microsatellite marker development has always been a tedious, prolonged and costly affair. The FIASCO technique, used in this study, has been the most optimal and appealing method to use in recent years as it provides an enrichment step to increase the probability of isolating unique microsatellite loci (Zane *et al*. 2002; Reece *et al*. 2004; Baranski *et al*. 2006b; Slabbert *et al*. 2008). Although the enrichment step has improved the technique, the overall yield is still generally low considering the amount of effort and funding put towards the process (Slabbert *et al*. 2010). Another factor to consider in the FIASCO procedure is the use of cloning and live culture amplification, which notoriously permits biases. The technique is however useful and does provide a relatively significant amount of markers, particularly for a non-model species for which genome information is limited (Reece *et al*. 2004; Baranski *et al*. 2006b; Slabbert *et al*. 2008).

Two hundred and twenty two primers were designed from 978 recombinant clones of which 88 polymorphic family-marker combinations (44x2) were developed and 38 of these combinations provided reliable genotypes. Eighteen family-marker combinations were monomorphic, 24 combinations non-reliable and eight were duplicated. Microsatellite markers are co-dominant and should thus follow expected Mendelian segregation patterns. Of the 38 informative marker-family combinations, 28 conformed to expected Mendelian segregation patterns (1:1:1:1, 1:2:1, 1:1, p>0.05) and were informative, while 10 combinations did not (Appendix B). Informative, in this case, implies that the segregation phase could be clearly discerned in the offspring of a full-sib family.

Null alleles are the primary cause of departure from Mendelian segregation as microsatellite flanking regions have high mutation rates (Callen *et al.* 1993; Pompanon *et al.* 2005). Three (30%) [*Hmid*0006M in Family 42A, *Hmid*2044T in Family 7B and *Hm*LCS147T in Family 7B] of the 10 combinations which did not conform to Mendelian segregation could be explained by the presence of null alleles. This is comparable to the study done by Baranski *et al*. (2006b) which presented a null allele frequency of 27%.

It is often easy to identify null alleles when using pedigree data as the alleles can be predicted by observing parental and offspring segregation patterns. Marker *Hmid*0006M,

for example, clearly shows the familial segregation patterns of a null allele present in the male parent of family 42A. The parent presents a genotype of a 'homozygote', however when observing the offspring a segregating allele from the male parent is visibly absent (Figure 2.5). The offspring were reassigned as heterozygotes for segregation analysis. These markers then conformed to the expected segregation patterns (1:1:1:1; $p>0.05$). In other words, if one parent is heterozygous A_1/A_2 and the other homozygous A_3/A_3, the offspring should contain the genotypes A_1/A_3 and A_2/A_3, with an expected segregation ratio of 1:1. However, if the second parent possesses a null allele and therefore only appears to be homozygous, the offspring should contain: A_1/null, A_2/null, A_1/A_3 and A_2/A_3 with an expected segregation pattern of 1:1:1:1. The null allele for *Hmid*2044T was a little more complicated as the heterozygous 171/null genotypes could not be distinguished from the homozygous 171/171 genotypes (Figure 2.5). Therefore the genotypes of this marker could not be altered to test segregation patterns against a more appropriate expected hypothesis and primer re-design should thus be the rectification strategy.

Distortion of the other combinations (*Hmid*0053D, *Hmid*0310D, *Hmid*4018D, *Hm*LCS71T and *Hm*NSp31M for Family 42A and *Hmid*0310D and *Hmid*0065M for Family 7B) could possibly be explained by gametic selection (Sekino *et al.* 2006) or zygotic selection (Reece *et al*. 2004), in which deleterious alleles are eliminated resulting in imbalanced segregation patterns. This is common in bivalve species such as oysters due to their highly fecund nature (Launey and Hedgecock 2001). Abalone are also highly fecund (Evans *et al*. 2004; Lucas *et al*. 2006) and are thus possibly affected by these mechanisms of selection; however there are relatively few studies concerning this. Differences in segregation distortion were observed when comparing the two families. This is seen specifically with markers *Hmid*0053D and *Hmid*0065C, where the marker follows the expected Mendelian segregation in one family but not in the other (Appendix B). This trend has also been observed in other studies and could possibly result from differences in lethal gene heterozygosity between families (Reece *et al*. 2004; Baranski 2006). Germline mutation (Brohede *et al.* 2002), selection pressure against markers linked to recessive deleterious alleles (Launey and Hedgecock 2001) and PCR errors or allele scoring difficulties (Jones and Ardren 2003) are other possible explanations for the distortion seen in these loci.

The presence duplicated loci (eight marker-family combinations) is not an uncommon phenomenon. There have been proposed hypotheses to explain the occurrence of multiple alleles such as genome duplication and aneuploidy, tandem duplication and locus transfer

by mobile elements (Baranski 2006). In some cases two of the alleles are inherited as a unit, providing grounds for the tandem duplication premise. Such a marker can still be useful in linkage mapping studies provided the genotypes are adjusted accordingly by assigning the two alleles as one. Duplicated markers in this study were excluded from further analysis, however, to avoid complication.

5.2 Marker development using Illumina sequence-by-synthesis

The primer development using Illumina sequence-by-synthesis provided a total of 14 primers, of which 10 were polymorphic i.e. a success rate of 37%. Optimisation was less problematic than with those designed from FIASCO and redundancy was also less of a hindrance. The bias associated with FIASCO marker development strategies can cause a higher frequency of primer mismatch than from direct sequencing, making optimisation more challenging and hence less successful.

The longer repeat motifs (tetranucleotide, pentanucleotide and hexanucleotide) as well as longer repeat arrays, were chosen for this study as longer microsatellites tend to provide a higher level of variation (Buschiazzo and Gemmell 2006). Type I EST-STRs are however not expected to be found in excess or have a high level of polymorphism as they could be under selective pressures. Microsatellites are predominantly found in non-coding regions of the genome in all species, with only 10-15% of these loci found in coding regions. Negative selection against frame shift mutations plays a major role in this trend (Metzgar *et al*. 2000).

The departure from HWE seen in locus *Hmid*ILL1-064192 could be explained by the presence of null alleles, with regard to the high expected null allele frequency determined (Table 2.3). The remaining three loci (*Hmid*ILL1-006622, *Hmid*ILL1-046948 and *Hmid*ILL1-146360) not conforming to HWE could also be explained by the presence of null alleles, considering the positive F_{IS} values denoting a heterozygote deficiency. An alternative hypothesis for HWE departure is linkage disequilibrium which could cause unexpected allele frequencies. The locus *Hmid*ILL1-146360 could be a candidate for such genome dynamics as the null allele frequency is rather low (Table 2.3).

The average observed (0.366) and expected (0.453) heterozygosities were somewhat comparable. The average F_{IS} value of 0.178 suggested that there is an overall heterozygote deficiency in the marker alleles. These are EST-STR markers, derived from

genic coding regions and are thus expected to present a lower level of polymorphism due to selective pressures, as previously mentioned.

There were eight primer pairs that did not amplify or provided non-reliable genotype data. This could be attributed to factors such as primer designing over intron/exon boundaries (Wall *et al*. 2009) or the inclusion of an intron region too large to amplify with PCR (Pérez *et al*. 2005), as these EST-STR primer pairs are designed from cDNA (reverse transcribed from modified mRNA sequences i.e. spliced). Studies using genic-derived markers can usually identify the intron/exon boundaries prior to primer design by comparison of the locus to the reference genome sequence (Wall *et al*. 2009), unlike studies on species such as *H*. *midae*, which does not have such information available. PCR artefacts (Jones and Ardren 2003) and null alleles (Pompanon *et al*. 2005) are additional reasons for failed amplification of markers.

The EST-derived microsatellite markers were aligned to the NCBI database using BLASTn alignments, prior to PCR optimisation, to identify any redundancy. Sixteen of the marker flanking regions aligned to mRNA sequences of closely related species (Table 2.4). Five of these sequences did not amplify after numerous attempts at optimisation. Eleven of these loci were used to develop labelled primers and six of these (*Hmid*ILL1-037506, *Hmid*ILL1-046948, *Hmid*ILL1-088398, *Hmid*ILL1-140027, *Hmid*ILL1-002192 and *HmidILL1*-006622) provided reliable genotype data. All six loci aligned to *Haliotis* species-derived sequences except *Hmid*ILL1-140027, which showed 85% similarity (3% coverage) to a RAt Sarcoma (RAS) oncogene family protein from Zebra fish [*Danio rerio*]. *Hmid*ILL1-037506 also aligned to a RAS mRNA sequence of *H. discus discus* with 93% similarity (E=0; 83% coverage). The RAS-related protein is involved with long-term potential (LTP) pathways in Zebra fish (Sato *et al*. 2009). LTP pathways are usually associated with mammals as it involves memory via long lasting signal transduction between neurons (Shors and Matzel 1997). Lastly, *Hmid*ILL1-006622 aligned significantly (90% similarity; E=0; 29% coverage) to a domain binding protein expressed in abalone exposed to *Haliotis tuberculata*, possibly inferring information on immune system pathways.

Although EST-STRs are generally not highly polymorphic, the significance of these type I markers is in the association to expressed regions of the genome, providing valuable information of genic positioning when incorporated into linkage maps and QTL analysis.

5.3 Comparing FIASCO and Illumina strategies for marker development

The FIASCO isolation technique provided a success rate of 32%, with 28 marker-family combinations out of 88 optimised and useful, which is comparable to similar studies (Rhode 2010). Marker development with the use of sequence-by-synthesis data provided a success rate of 37% with 22 primer pairs optimised and 10 providing reliable genotypes.

This is one of the first attempts at microsatellite marker development using the Illumina GA sequence-by-synthesis contigs, another study was done by Dempewolf *et al*. (2010). Numerous SNP development studies have been successfully executed using this technology (Hillier *et al*. 2008; Van Tassell *et al*. 2008; Hyten *et al*. 2010).

Both these strategies have advantages and disadvantages but both ultimately provide a comparable amount of polymorphic markers. In the study done by Slabbert (2010), 454 pyrosequencing (which currently provides longer sequence reads than Illumina) was performed for marker isolation and 82 markers were obtained. Abdelkrim *et al*. (2009) and Santana *et al*. (2009) compared STR marker development strategies using 454 (Roche) sequence results and enrichment techniques. In the study performed by Abdelkrim *et al*. (2009), 17215 reads were used to detect microsatellites and after selecting 203 di-, tri- and tetranucleotide repeats, 24 markers were chosen of which 13 were ultimately usable. Another study using 454 pyrosequencing was performed by Santana *et al*. (2009), in which 1692 contig sequences were used to isolate microsatellite markers. The authors found that 97% of the contigs contained repeat units; however after primer design and optimisation, 14% were developed into usable markers. This shows that although the efficiency is not always exceptionally higher than that of FIASCO, when comparing cost of sequencing and the time required for development, it is in fact a vast improvement.

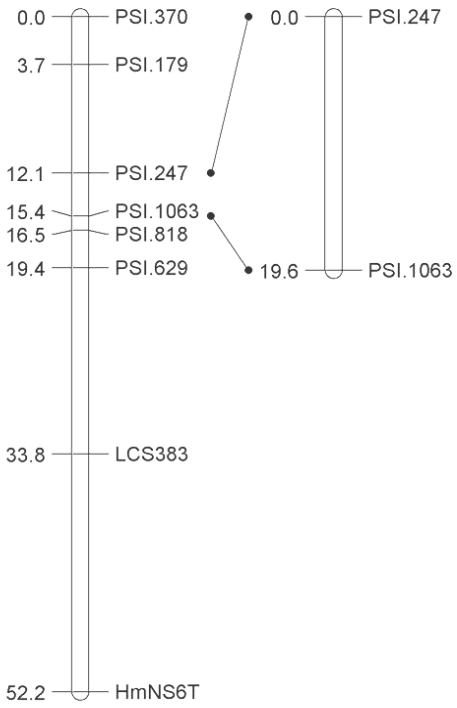
It is often difficult to design primers for amplification of microsatellites using either method if the marker is found at the terminal end of the sequences and no flanking region is available. The Illumina sequences tend to have shorter read lengths and have ambiguous error prone ends, making primer designing even more challenging. FIASCO-derived loci, however, seem to be vastly more prone to redundancy. The primary advantage of marker development using sequence-by-synthesis methods is the extreme reduction in cost and time required to develop equivalent or a larger amount of polymorphic loci compared to enrichment techniques. The cost reduction between FIASCO and NGS marker

development, reported in the study by Abdelkrim *et al*. (2009), was 3-5 times favouring the latter. A second, extremely appealing factor to consider is that type I microsatellites (EST-STRs) can be developed via this method if transcriptome sequencing is performed, providing direct information regarding putative genes.

Characterised microsatellite markers can be used in applications such as linkage mapping. Informative loci, i.e. markers which provide traceable allelic segregation patterns within a particular mapping family, are required for linkage analysis. The uninformative loci in this study should not be completely discarded from future applications such as pedigree analysis, since polymorphic content differs between pedigrees. Loci containing null alleles, loci which are duplicated within the *H. midae* genome or loci which are difficult to score or show unreliable amplification should be used with caution and it is recommended they be assessed for their suitability within a specific application.

Chapter three

Linkage mapping in Haliotis midae



1. Abstract

Development of a linkage map in a non-model species, such as Perlemoen, creates the building blocks towards sustainable farming practices through MAS as well as identification and characterisation of genic regions, possibly associated with economically viable traits. The idea is to create a map densely populated with markers for application in breeding strategies in which certain markers are inherited alongside an economically important trait, such as growth rate or disease resistance. This study produces a small step toward a foundation for future QTL association studies and further investigation into genes and their trait-linked inheritance. Two full-sib families were used for segregation and pairwise recombination analysis of 214 markers (156 previously developed as well as the 44 FIASCO and 14 EST-STRs from this study) in the program JoinMap® v4, to develop sex-specific and sex-average linkage maps. Fifty two markers were mapped to 15 linkage groups in the paternal derived map of family 7B with an average genome length of 574 cM and an average marker spacing of 9.87 cM. Similarly, 41 markers were linked in the maternal derived map placed into 16 linkage groups with an average genome length of 766 cM and an average marker spacing of 14.24 cM. The paternal derived map of Family 42A included 47 mapped markers within 14 linkage groups, an average genome length of 505 cM and an average marker spacing of 10.38 cM. Finally the maternal sex-specific map for Family 42A provided a map of 38 markers mapped into 13 linkage groups with an average genome length of 575 cM and an average spacing of 12.45 cM. Seven linkage groups between the paternal maps could be joined with at least one concurrent marker and only three between the maternal maps. The sex-average linkage maps were constructed into 18 linkage groups for both families with average genome lengths of 945 cM and 868 cM for Family 7B and 42A, respectively. Thirteen of these linkage groups could be joined with MapChart with at least one concurrent marker. Although this is the second linkage map created for *Haliotis midae*, it is still preliminary and only represents coverage of approximately 50%. Microsatellites are extremely useful markers for creating a framework map; however future projects should contain more genic derived markers as well as more abundant markers such as SNPs.

Keywords: *Haliotis midae*, linkage map, genome length, coverage marker spacing

2. Introduction

The first generation of linkage maps for most aquaculture species predominantly made use of AFLP markers ([Tilapia] Kocher *et al*. 1998; [Pacific oyster] Hubert and Hedgecock 2004; [Pacific oyster] Li and Guo 2004; [Pacific abalone] Liu *et al*. 2006; [Blue mussel] Lallias *et al*. 2007; [Perlemoen] Badenhorst 2008). It became evident, however, that more transferable markers are necessary for utilisation of these maps in other populations or species. Microsatellites quickly became the marker of choice for more dense linkage mapping projects ([Tilapia] Lee *et al*. 2005; [Yellowtails] Ohara *et al*. 2005; [Turbot] Bouza *et al*. 2007; [grass carp] Xia *et al*. 2010) as these are co-dominant, highly variable and PCR-based making them easily reproducible (Liu and Cordes 2004). Currently, as linkage maps are becoming more dense and QTL mapping a priority, EST-associated microsatellites are becoming more popular, as well as SNP markers due to their abundance and high throughput analysis ([Atlantic salmon] Moen *et al*. 2008; [Atlantic cod] Hubert *et al*. 2010; [grass carp] Xia *et al*. 2010).

Linkage mapping is based on the observation of DNA marker co-segregation to determine relative distance between the markers and the order in which they are positioned. The segregation patterns and recombination events are discerned using a mapping population, typically a backcross or F_2 pedigree that represents homologous inbred parent lines. F_2 lines are the most informative for linkage mapping since it allows easy identification of non-recombinant and recombinant allelic haplotypes, as the parental types are homologous for every gene.

It is, however, exceptionally challenging to create inbred F_2 or backcross lines in outbred populations which are found in most aquaculture species, including abalone. In this case a full-sib pedigree will be used in a pseudo-test backcross analysis method in which a maximum of four possible alleles per locus will be observed. The offspring will thus inherit two (one from each parent) of the four possible alleles (Wu *et al*. 2007). Fortunately mollusc species are highly fecund (Launey and Hedgecock 2001; Evans *et al*. 2004; Lucas *et al*. 2006) resulting in the production of extremely large full-sib families, ideal for linkage analysis.

In order to keep track of the mapping family individuals, numerous tagging methods have been used in various aquaculture species depending on the purpose, cost and retention time required. Reliable tagging of individual juvenile abalone is challenging. Wild adult individuals are usually tagged using wire or fishing line for attachment of a tag to the respiratory pores (Newman 1966). Prince (1991) used nylon rivets to attach numbered discs to the trema of abalone. Animals as small as 50 mm were tagged in this way. Commercial broodstock are often tagged using a numbered Perspex piece imbedded in quick-set putty, which is seen in South African farms. Breen *et al*. (2003) glued plastic tags to the shell of *Haliotis iris* using Superglue. Gallardo *et al.* (2003) fed *H. asinina* artificial diets to give the shell, which is naturally a brownish colour, a blue growth band and reported it as a successful method for this species. This method may unfortunately depend on the biology of the abalone species being tagged. Kube *et al*. (2007) used coloured beads to identify families by glueing them to the shells of 10 month (12 mm) old juveniles. Tag loss for larger wild animals was less, at 4-35% per annum (Prince 1991) and 3% for 27 months and older commercial animals compared to 55% for 10 - 21 month old animals (Kube *et al*. 2007). Kube *et al*. (2007) reported a total tag loss of 89% over a 28 month period. The most recent development in abalone tagging in which individual animals can be identified is the Radio frequency identity (RFID) tags, which are attached using either superglue or quick-set putty (Figure 3.1).

Figure 3.1: Radio frequency identity tagging on *H. midae* using Pratley's quickset putty

The genotyped alleles of the population are expected to follow an independent assortment pattern following Mendelian segregation. In other words, unlinked markers show a maximum of 50% recombination between two loci. The segregation patterns of each locus are analysed separately prior to linkage mapping to ensure the assumptions are followed. There are four possible expected segregation ratio groupings for chi-square analysis. The

first three groups include both parents which are heterozygous following a 1:1:1:1, 1:2:1 or 3:1 segregation pattern including distinguishable null alleles. The last group incorporates test cross configuration and a single heterozygote parent leading to a 1:1 expected segregation pattern (Wu *et al*. 2007). After segregation analysis, linkage can be performed.

For linkage map construction, JoinMap® v4 (Van Ooijen 2006) provides four grouping statistics to select from, namely, a test for independence with the LOD score as a test statistic, a test for independence using *p* values, recombination frequencies and linkage LOD (Van Ooijen 2006). The pair wise recombination of markers is used to place markers into groups of linked loci based on a LOD test statistic. The LOD score is a ratio of observed recombination, used to determine linkage between loci with a default threshold of 3. Mapping programs use the linear order of the markers in each individual to identify the break points at which recombination occurred by observing allelic variation within the individual (Table 3.1 shows an extremely simplified example). The selection of LOD threshold values is dependent on mapping family size and should not place pair wise loci with a recombination greater than 0.2 in a linkage group as this will result in underestimation of the true map distance. A smaller sample size will hence require a more stringent threshold to avoid false linkages and prevent obscurities (Danzmann and Gharbi 2007). The accepted threshold for linkage mapping with a family size of approximately 100 is a LOD score of 3. Once the markers are placed into linkage groups, marker order is subsequently elucidated through a ranking of LOD scores and ML calculations (Ott 1999; Van Ooijen 2006). There are two algorithms provided by JoinMap® v4 specifically created for marker ordering, the regression mapping algorithm (Stam 1993) and the MML mapping algorithm (Jansen *et al*. 2001).

Table 3.1: Basic representation of linkage mapping data to identify recombination events. The left column illustrates markers (M) 1-7 and the top row displays the individuals (I) 1-9. Between markers 2 and 3 there is one recombination event, i.e. the observed recombination fraction is 1/9 = 0.111

	I1	I2	I3	I4	I5	I6	I7	I8	I9
M1	H	A	H	H	A	H	A	H	A
M2	H	A	H	H	A	H	A	H	A
M3	H	H	H	H	A	H	A	H	A
M4	A	H	H	A	A	A	H	A	A
M5	A	H	H	A	H	A	H	A	H
M6	A	H	H	A	H	H	H	A	H
M7	A	H	A	A	H	H	H	H	H

Two complex factors should be considered when mapping distance between two loci, as not all recombination events are equally probable across all areas of a chromosome. Double crossovers and interference can drastically alter the calculated distances between two points when equating recombination to physical distance as a directly proportional relationship. Double crossovers occurring between two points would not alter the parental-phase genotype observed in the progeny, which can result in underrepresentation of true genetic distance. The second phenomenon with regards chromosome structure and size is the presence of interference. If, for example, two loci are located far apart but the chromosomal structure does not accommodate chiasmata, the loci appear physically close together when in fact they are not. Recombination “hot spots” provides another explanation for variable recombination along a chromosome (Bhargava and Fuestes 2010). Mapping functions are therefore created to surmount the effects of such trends (Danzmann and Gharbi 2007).

Comprehensive and dense linkage maps are usually required for accurate calculations of genome length (*G*) using mapped interval summation. In first generation linkage maps where genome coverage is low and molecular markers limited, a preliminary estimate can only be determined. There are various methods for calculating genome lengths, such as the method-of-moments estimator (Hulbert *et al*. 1988), the ML estimator (Chakravarti *et al*. 1991) and the method described by Fishman *et al*. (2001). Chakravarti *et al*. (1991) revealed, in a comparative study between the methods-of-moments estimator and the ML

methods, that when low density linkage maps are used to calculate genome lengths the ML method is superior. The Fishman *et al*. (2001) method uses an estimated marker spacing (s) which is added to the length of each linkage group and totalled. This estimated value allows incorporation of the chromosome ends, before and after the first and last marker, for a more accurate estimation of genome length.

In this study, sex-specific and sex-average linkage maps were constructed using JoinMap® v4 (Van Ooijen 2006) within two full-sib *H*. *midae* families using microsatellite markers. The linkage groups obtained were merged where homologous markers were identified between paternal and maternal maps as well as sex-average family maps. The genome length for *H. midae* was also estimated using these framework linkage maps.

3. Materials and methods

3.1 Tagging the mapping family

Three different tagging methods were tested for efficiency and retention time. The methods include plastic aerials placed in the tremata, bee-tags glued to the shell and RFID tags attached to the shell using glue or quickset putty (Quickset Putty from Pratley, South Africa) and scanned with a RFID reader (Slabbert *et al*. [in prep.]). The RFID tagging system was used in this study to identify the mapping family animals.

One hundred offspring from two full-sib families, 7B (sampled at Roman Bay) and 42A (sampled at HIK) were sampled and subsequently tagged with RFID tags using Quickset Putty (Pratley, South Africa). DNA extractions were performed as described in chapter 2 (section 3.1).

3.2 Genotyping and data preparation

The parents of two full-sib families, 7B and 42A, were screened using previously developed STR markers: 11 from Bester *et al*. 2004 [markers **1-11**], 63 from Slabbert *et al*. (2008) [markers **12-74**], 44 from Slabbert *et al*. (2010) [markers **75-118**] and 82 from Slabbert *et al*. ([in prep.]) [markers **119-200**]. The 14 reliable EST-STRs developed during this study were also included. [*PCR amplification and genotyping of markers 17-92 and 119-200 was performed by Dr P. Wang. PCR amplification and genotyping of markers 1-16, 93-118 and the 14 EST-STRs was performed by the author*]. One hundred individuals from both families were

genotyped with the informative polymorphic loci (marker genotypes from which allele phase could be determined from segregation patterns) in multiplex reactions. The multiplex reactions were performed as follows: 10ng of template DNA was added to 3.5 µl 2X QIAGEN® Multiplex PCR master mix (containing HotStart*Taq*® DNA Polymerase, Multiplex PCR Buffer with 6 mM $MgCl_2$ and dNTP Mix) (QIAGEN®), 0.9 µl Primer mix (20 µM of each primer) and ddH_2O to a final volume of 7 µl. The following PCR cycle was used to amplify the target locus: The cycle was initiated with a 10 min denaturing step at 95℃ for 15 min, followed by 35 cycles of 94℃ for 30 s, 57℃ for 90 s and 72℃ for 1min. The PCR completed with an elongation step of 60℃ for 3 0 min. The markers were genotyped on the ABI 3730xl DNA Analyser (Applied Biosystems) and the lengths determined by comparison to the GeneScan™ 600 LIZ® Size Standard (Applied Biosystems). The analysis of genotypes was performed using GeneMapper v4.1 software (Applied Biosystems).

The genotype data obtained for each of the polymorphic markers in each family was converted to the appropriate format for a CP (outcross) population suggested by Van Ooijen (2006) (Table 3.2). Sex-specific maps were created by converting abxcd and efxeg genotypes to either hkxhk and lmxll or hkxhk and nnxnp genotypes for maternal and paternal maps, respectively.

Table 3.2: Genotype coding requirements JoinMap® v4 [Van Ooijen 2006]

Code	Description	Possible genotypes
<abxcd>	Heterozygous in both parents, 4 alleles	ac, ad, bc, bd, --
<efxeg>	Heterozygous in both parents, 3 alleles	ee, ef, eg, fg, --
<hkxhk>	Heterozygous in both parents, 2 alleles	hh, hk, kk, --
<lmxll>	Heterozygous in first parent*	ll, lm, --
<nnxnp>	Heterozygous in second parent#	nn, np, --

* Male parent was chosen as the first parent in both families
# Female parent was chosen as the second parent in both families
-- missing data

3.3 Analysis of linkage between microsatellite markers

Linkage analysis was performed using all polymorphic markers by observing paternal and maternal segregation patterns separately using JoinMap® v4 software (Van Ooijen 2006). Prior to the linkage grouping a chi-squared goodness-of-fit test was performed to allow drastically distorted markers (p<0.005) to be excluded from further analysis. The segregation analysis of marker genotypes was performed by JoinMap® v4 with the expected Mendelian ratios of each locus elucidated by the program. The genotype frequencies for each locus are determined and used to identify any segregation distortion with the use of a chi-square probability test. The expected ratios, to which the observed ratios are compared, were determined using the information given in Table 3.3. Deviation from the expected ratios in any locus was provided by JoinMap® v4, with seven default significance thresholds (0.1-0.0001). Once the distorted markers had been excluded, locus grouping could commence with selected parameters.

The test for 'independence using a LOD score for significance testing' was selected for locus grouping in this study. Markers presenting a recombination frequency less than the set threshold and a LOD score greater than the set threshold were considered as linked (Van Ooijen 2006). The threshold was set at an initial LOD score of two with an increasing stringency of one increment to a final threshold of 10 with a default set recombination threshold of 0.4. Linkage groups within a LOD threshold of three were chosen for further analysis (framework markers). The regression mapping algorithm was used to calculate map order.

Table 3.3: Mendelian segregation patterns of genotype codes for JoinMap® v4 [Van Ooijen 2006]

Code	Ratio	Classification into genotypes
ac, ad, bc, bd	1:1:1:1	ac, ad, bc, bd
ee, ef, eg, fg	1:1:1:1	ee, ef, eg, fg
hh, k-	1:3	hh and k-; hk and kk included in class k-
h-, kk	3:1	h- and kk; hh and hk included in class h-
hh, hk, kk	1:2:1	hh, hk and kk
ll, lm	1:1	ll and lm
nn, np	1:1	nn and np

A goodness-of-fit calculation was used to compare direct recombination and mapping function recombination frequencies and was expressed as a chi-squared measure. A G^2 likelihood ratio statistic was used for calculating the likelihood ratio between direct and derived recombination frequencies, with the equation seen below:

$G^2 = 2 \sum O \log (O/E)$

Where O denotes observed number of individuals and E represents expected number of individuals. Log represents the natural logarithm and the sum over all cells is represented by ∑ (Stam 1993).

The Kosambi mapping function was used to convert recombination frequencies between markers into centiMorgan map distances (Kosambi 1944). Maps for each population were graphically represented using MapChart software (Voorrips 2002).

3.4 Linkage map integration

Linkage groups from each sex-specific map were scanned for homology of marker placement by eye. The groups within the male maps and female maps were then combined where possible to display the homology between the families. Sex-average maps were also created where possible in which linkage groups from the paternal and maternal map from the same family were merged.

3.5 Marker distribution and genome coverage

When estimating map length, one must take into account the distance from the ends of the chromosome to the first marker which includes the telomeric region. This is performed by multiplying the length of each linkage group by two times the average distance from the end marker to the end of the chromosome (Ohara *et al*. 2005).

The genome length was calculated using two methods displayed below (referred to as equation **A** and **B**).

A: $G_e = \sum G_{oi} [(k_i+1) / (k_i-1)]$

Where G_oi represents the observed length of the linkage group i and ki represents the number of markers at linkage group i (Chakravarti *et al*. 1991).

B: $G_e = G_o + (2tG_o) / n$

Where G_o represents the total observed length, t represents the number of linkage groups and n denotes the number of intervals between markers (Fishman *et al*. 2001).

The average marker spacing [s] per linkage group can be calculated by dividing the total length of the group by the number of intervals (Liu *et al*. 2006; Lallias *et al*. 2007). The genome coverage was subsequently determined using the equation $G_o/G_{e\ ave}$, where $G_{e\ ave}$ is the average expected genome length between the A and B equation results.

4. Results

4.1 Tagging

The bee-tags and RFID tags proved to be the most reliable tagging method out of the three proposed techniques. The RFID tags were the most successful, presenting 100% tag retention over the three months; however it is costly and more time-consuming than bee-tags. This method is thus only applicable when tagging few animals (less than 200) for individual identification (Slabbert *et al*. [in prep.]). The RFID tags were hence chosen for this particular study.

4.2 Marker segregation

Out of the 214 markers available for this study, 106 markers were polymorphic in family 7B and 106 in family 42A. Four null alleles were identified in Family 7B, of which two were usable and seven in Family 42A with five that were useful. Finally, 19 marker-family combinations displayed duplicated alleles and were removed from further analysis. In some cases such as with marker *Hm*LCS71T in family 7B, the duplication was surmountable by selecting a single parent specific allele (Figure 3.2).

Figure 3.2: Electropherogram showing duplication in locus *Hm*LCS71T. The male and female parent possess two alleles, however the offspring tend to inherit three or four alleles. To surmount such genotypes for use in linkage mapping, one allele from each parent was chosen which was distinguishable within the offspring and an "lmxll" or "nnxnp" coding was used. Offspring one will thus be "nn" and offspring two "np".

4.3 Linkage mapping

Separate sex-specific linkage groups were created for each family (Figures 3.3-3.7), using the "create maternal and paternal population node" function in JoinMap® v4. A sex-average linkage map was also created per family by merging linkage groups with homologous marker placement (Figures 3.8 and 3.9). The linkage groups were arbitrarily ordered from longest to shortest in each case.

4.3.1 LINKAGE MAP OF FAMILY 7B FROM ROMAN BAY

The linkage map of the male parent for the Roman Bay (7B) family consisted of 52 mapped microsatellite markers (LOD 3), placed into 15 linkage groups. Fourteen markers were ungrouped and six could not be assigned to a linkage group. Two markers were extremely distorted (p<0.005), and thus removed from further linkage assignment. The remaning 32 marker data had either too much missing data or was not reliable. The genome length calculation using equation A was 599 cM and alternatively, equation B provided a length estimation of 549 cM. The genome coverage ($G_o/G_{e\ ave}$) was

approximately 55%. The linkage groups ranged in length from 2.5 cM to 38.7 cM with an average marker spacing of 9.87 cM. The number of markers per linkage group ranged from 2-9 (Table 3.4).

Table 3.4: Sex-specific linkage map details for family 7B paternal (7B-p1).

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	9	38.70	4.84	15.72
2	3	38.00	19.00	21.54
3	7	33.80	5.63	14.38
4	3	27.70	13.85	17.84
5	5	26.20	6.55	26.22
6	2	22.80	22.80	22.81
7	2	18.30	18.30	18.27
8	3	17.70	8.85	11.67
9	3	17.30	8.65	10.10
10	3	16.20	8.10	13.88
11	3	15.30	7.65	10.29
12	3	14.10	7.05	14.06
13	2	8.40	8.40	8.41
14	2	5.90	5.90	5.91
15	2	2.50	2.50	2.50
TOTAL	52	302.85	148.07	213.60
AVERAGE	3.47	20.19	9.87	14.24

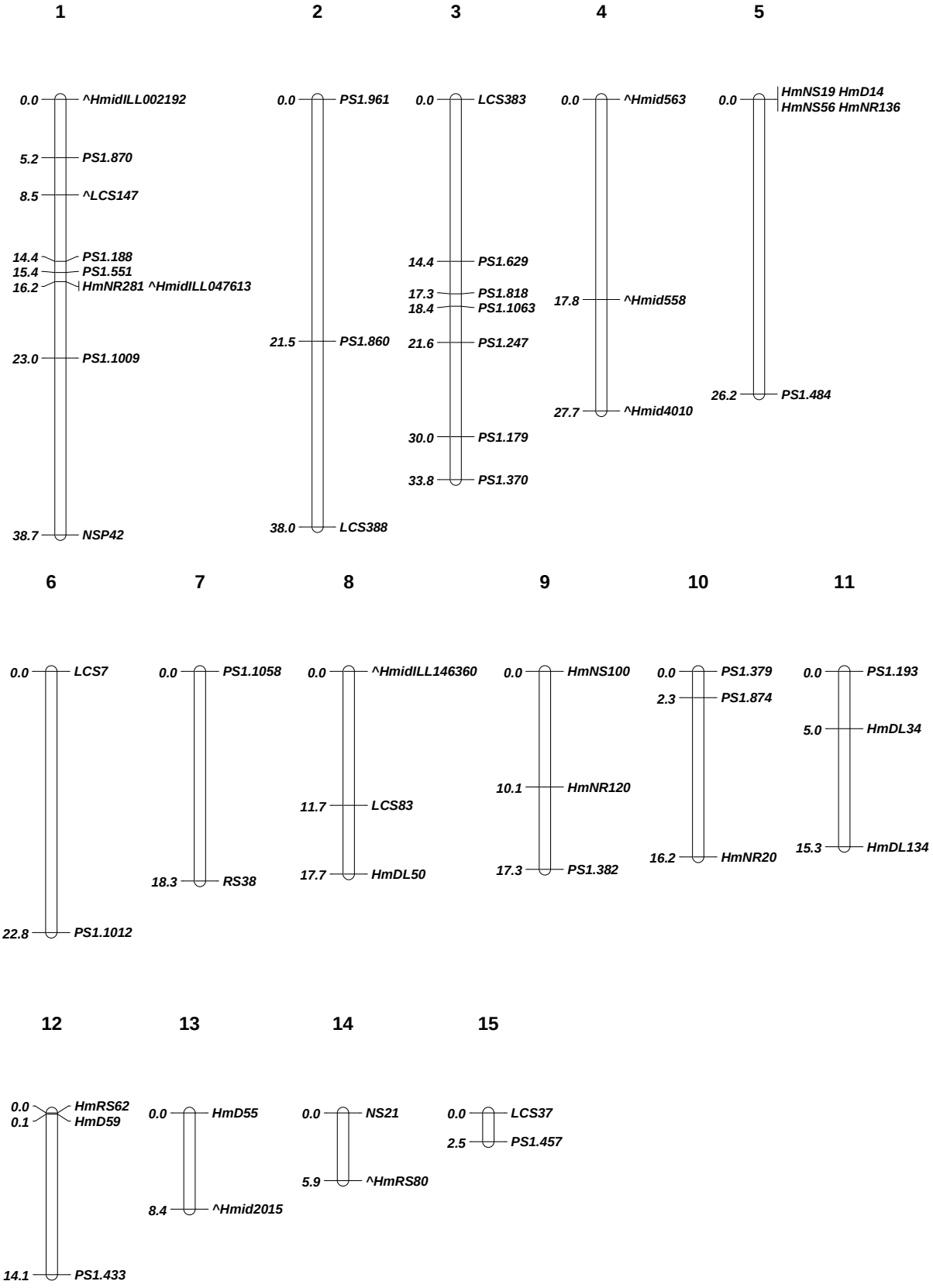


Figure 3.3: The sex-specific linkage map of family 7B paternal (p1) [MapChart]. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Markers with a ^ in front of the name denotes the markers which were PCR amplified and/or developed by the author. Linkage groups were placed arbitrarily from the longest to shortest in length.

The female map consisted of 41 mapped microsatellite markers which were placed into 16 linkage groups. Fourteen markers were ungrouped and five could not be mapped in the assigned group linked. Two markers were distorted (p<0.005). The remaning 44 marker data had either too much missing data or was not reliable. The genome length calculated using equation A gave a result of 794 cM and 738 cM was obtained using equation B. Genome coverage of approximately 44% was seen. Linkage groups ranged in length from 1.3cM to 56.9 cM with overall average marker spacing of 14.24 cM, which is higher than in the male map, as expected. The amount of markers per linkage group ranged from 2-7 (Table 3.5).

Table 3.5: Sex-specific linkage map details for family 7B maternal (7B-p2)

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	7	56.90	9.48	19.90
2	3	40.90	20.45	21.07
3	2	33.20	33.20	33.23
4	3	29.90	14.95	29.86
5	2	27.90	27.90	27.89
6	2	27.10	27.10	27.08
7	3	22.10	11.05	14.28
8	2	20.10	20.10	20.06
9	2	18.30	18.30	18.27
10	2	11.60	11.60	11.57
11	2	11.60	11.60	11.60
12	2	7.40	7.40	7.37
13	2	7.10	7.10	7.11
14	2	4.80	4.80	4.83
15	3	3.70	1.85	2.49
16	2	1.30	1.30	1.25
TOTAL	41	323.9	227.88	257.89
AVERAGE	2.56	20.24	14.24	16.12

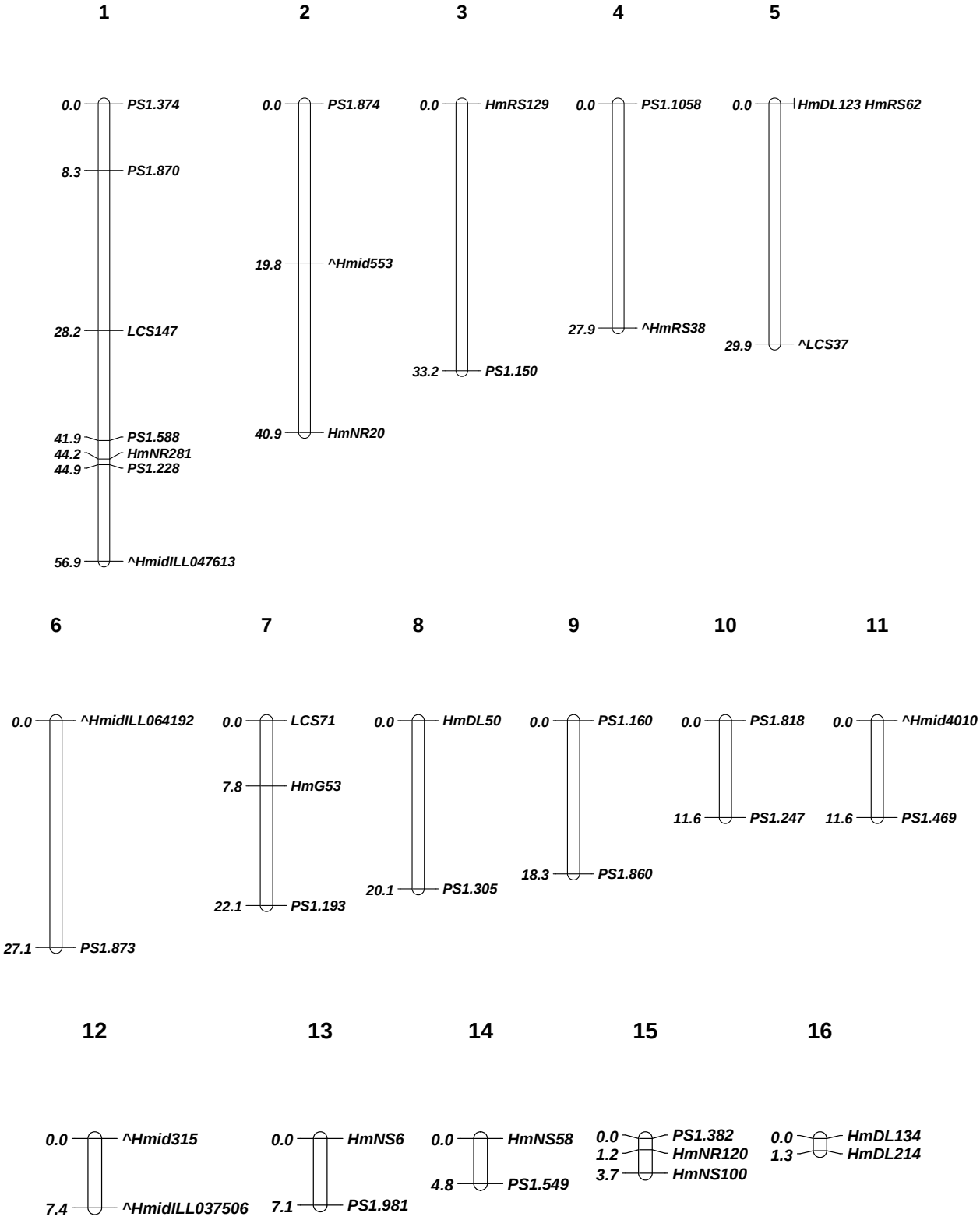


Figure 3.4: The sex-specific linkage map of family 7B maternal (p2) [MapChart]. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Markers with a ^ in front of the name denotes the markers which were PCR amplified and/or developed by the author. Linkage groups were placed arbitrarily from the longest to shortest in length.

4.3.2 LINKAGE MAP OF FAMILY 42A FROM HIK

The linkage map of the male parent for the Roman Bay family consisted of 47 mapped microsatellite markers (LOD 3). These markers were placed into 14 linkage groups, however 17 markers could not be assigned to a linkage group and five were distorted (p<0.005). The remaning 37 marker data had either too much missing data or was not reliable. The genome length calculation using equation A was 518 cM and alternatively, equation B provided a length average of 491 cM. The genome coverage was approximately 54%. The linkage groups ranged in length from 1.10 cM to 34.0 cM with average marker spacing of 10.38 cM. The amount of markers per linkage group ranged from 2-7 (Table 3.6).

Table 3.6: Sex-specific linkage map details for family 42A paternal (42A-p1)

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	6	34.00	6.80	19.70
2	4	33.40	11.13	20.49
3	4	30.80	10.26	23.49
4	3	30.20	15.10	16.31
5	4	30.00	10.00	27.83
6	7	23.00	3.80	9.65
7	2	19.90	19.90	19.87
8	2	19.60	19.60	19.63
9	3	17.30	8.65	14.05
10	2	8.90	8.90	8.88
11	3	7.30	3.65	5.68
12	3	7.00	3.50	5.05
13	2	3.30	3.30	3.27
14	2	1.10	1.10	1.08
TOTAL	47	265.8	145.29	194.98
AVERAGE	3.36	18.99	10.38	13.93

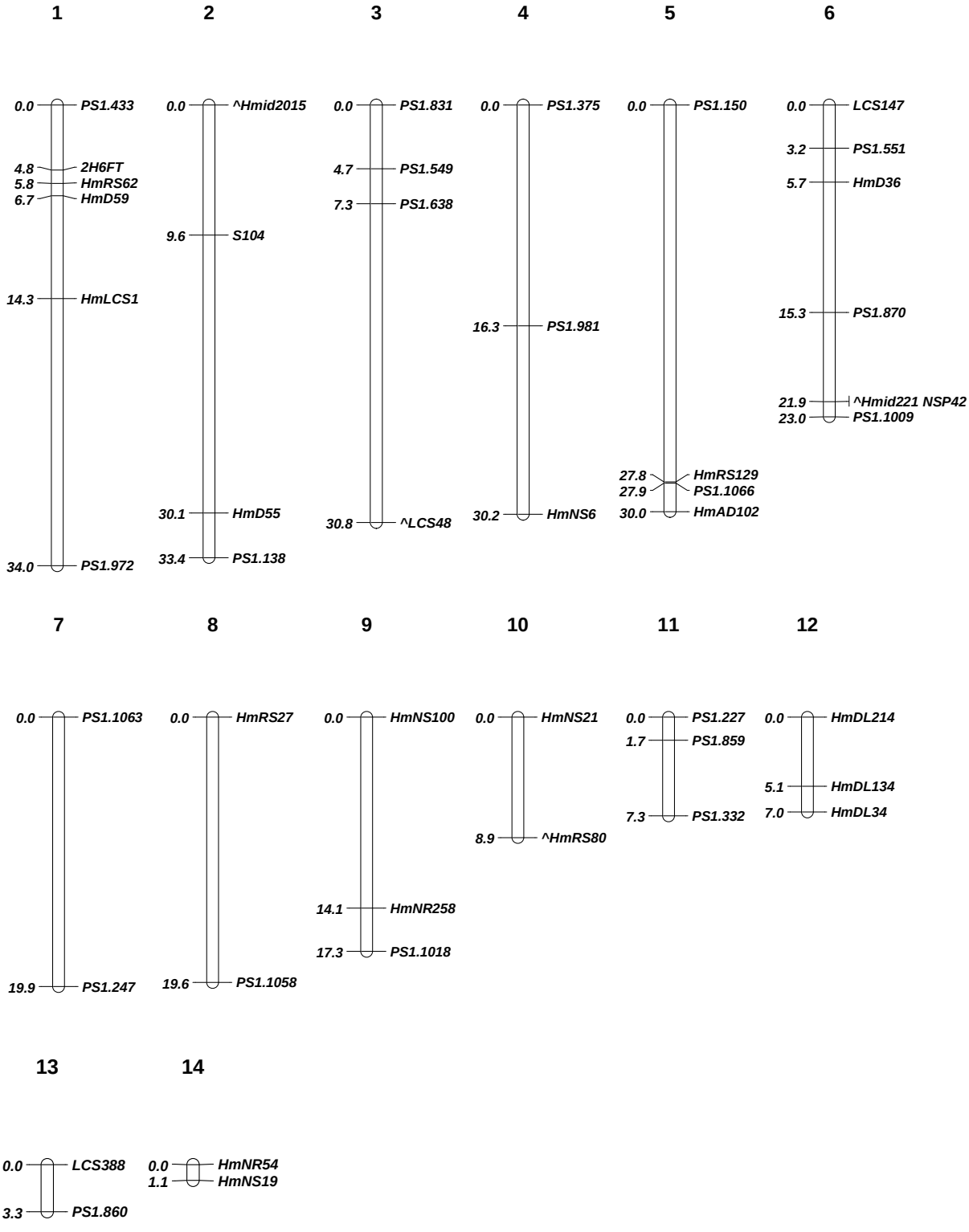


Figure 3.5: The sex-specific linkage map of family 42A paternal (p1) [MapChart]. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Markers with a ^ in front of the name denotes the markers which were PCR amplified and/or developed by the author. Linkage groups were placed arbitrarily from the longest to shortest in length.

The female map consisted of 38 mapped microsatellite markers which were placed into 13 linkage groups. Fourteen markers could not be linked and one was distorted (p<0.005). The remaning 53 marker data had either too much missing data or was not reliable. The genome length calculated using equation A gave a result of 595 cM and 554 cM was obtained using equation B. Genome coverage of approximately 49% was seen. The linkage groups ranged in length from 3.3 cM to 43.8 cM with an overall average distance of 12.45 cM, which is also higher than in the male map. This indicated that the recombination rate in *H*. *midae* differed between the sexes. The number of markers per linkage group ranged from 2-5 (Table 3.7).

Table 3.7: Sex-specific linkage map details for family 42A maternal (42A-p2)

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	5	43.80	10.95	19.29
2	3	34.80	17.40	33.69
3	3	32.60	16.30	26.87
4	2	31.30	31.30	31.32
5	5	24.80	6.20	13.52
6	2	23.30	23.30	23.29
7	4	15.70	5.23	11.21
8	3	14.80	7.40	8.06
9	3	13.10	6.55	8.16
10	2	12.80	12.80	12.77
11	2	12.30	12.30	12.33
12	2	8.80	8.80	8.79
13	2	3.30	3.30	3.27
TOTAL	38	271.4	161.83	212.57
AVERAGE	2.9	20.87	12.45	16.35

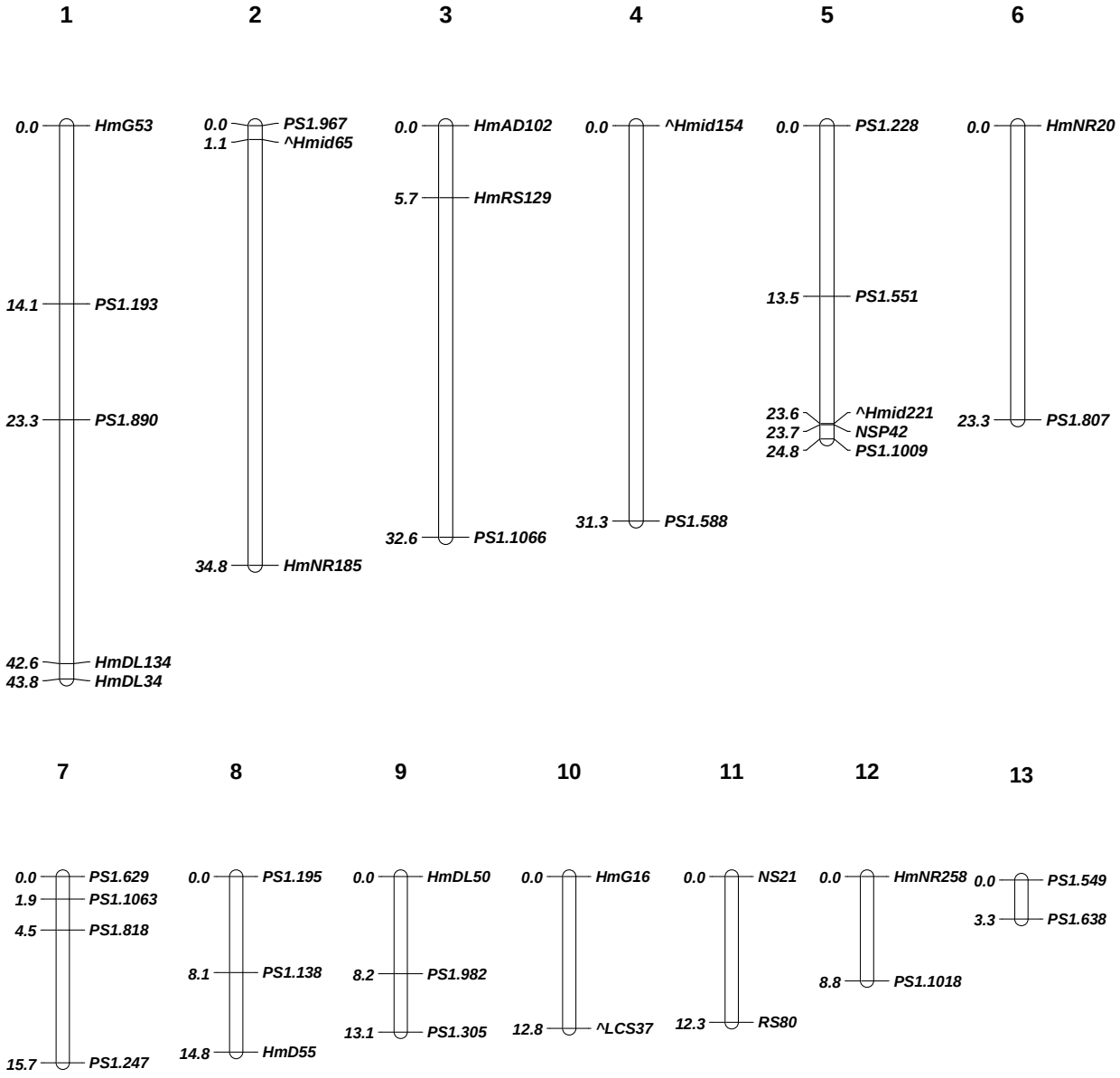


Figure 3.6: The sex-specific linkage map of family 42A maternal (p2) [MapChart]. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Markers with a ^ in front of the name denotes the markers which were PCR amplified and/or developed by the author. Linkage groups were placed arbitrarily from the longest to shortest in length.

4.3.3 COMPARISON OF SEX-SPECIFIC LINKAGE MAPS

The linkage groups drawn from the paternal and maternal maps of the two mapping families were compared and aligned where possible. Seven linkage groups from the paternal maps between family 7B and 42A were joined with at least one concurrent marker (Figure 3.7). Four corresponding markers were identified between LG1_7B and LG6_42A; however the marker order was not co-linear. The two markers concurrent between LG3_7B and LG7_42A were in the same order; although the interval distances varied by

17 cM. The remaining two joined linkage groups had one marker in common. When comparing the maternal derived maps, three linkage group comparisons could be constructed. It can be seen in Figure 3.7 C that two linkage groups (7 and 16) from family 7B aligned to LG1 from family 42A in the maternal map comparisons.

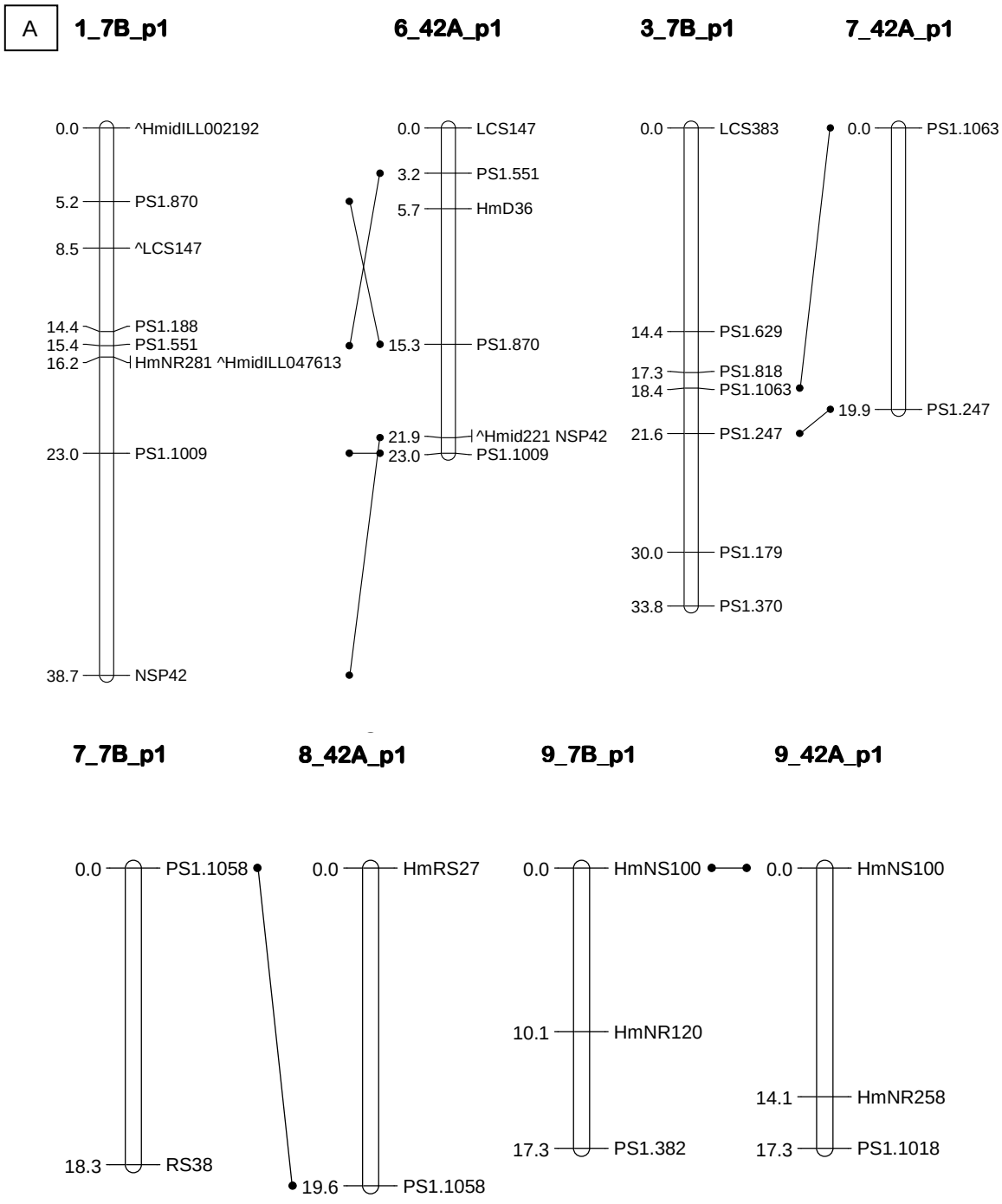


Figure 3.7: The sex-specific linkage groups joined by homologous markers. The first four comparisons of the paternal linkage maps; LG1_7B with homology to LG6_42A, LG8_7B with homology to LG7_42A, LG7_7B with homology to LG8_42A and lastly LG9_7B with homology to LG9_42A.

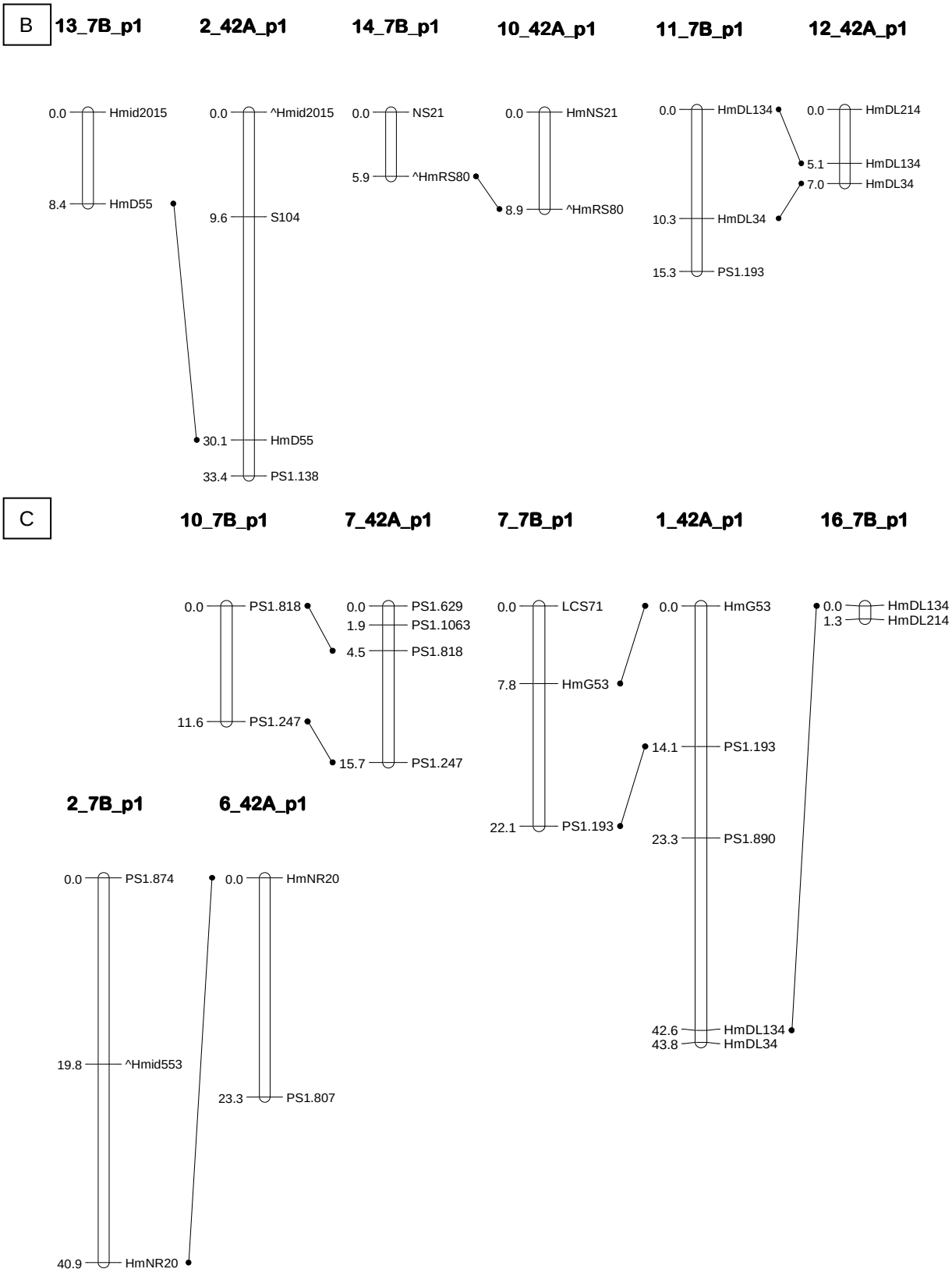


Figure 3.7 continued: The sex-specific linkage groups joined by homologous markers. The first 4 comparisons of the paternal linkage maps; LG1_7B with homology to LG6_42A, LG8_7B with homology to LG7_42A, LG7_7B with homology to LG8_42A and lastly LG9_7B with homology to LG9_42A.

4.3.4 SEX-AVERAGE LINKAGE MAPS

The linkage groups ranged in length from 1.60 cM to 55.80 cM (Family 7B) with overall average marker spacing of 14.41 cM. The number of markers per linkage group ranged from 2-12 (Table 3.8). The linkage groups for Family 42A ranged in length from 0.50 cM to 63.10 cM with overall average marker spacing of 12.57 cM. The number of markers per linkage group ranged from 2-10 (Table 3.9).

Table 3.8: Sex-average linkage map details for family 7B

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	12	55.80	5.07	10.80
2	4	52.50	17.50	38.60
3	3	45.10	22.55	26.19
4	6	44.10	8.82	23.44
5	7	41.80	6.97	18.16
6	3	35.60	17.8	20.54
7	3	32.90	16.45	20.36
8	2	27.90	27.90	27.90
9	2	25.90	25.90	25.90
10	4	25.00	8.33	16.76
11	2	23.70	23.70	23.70
12	2	22.10	22.10	22.10
13	2	21.60	21.60	21.60
14	2	13.80	13.80	13.80
15	3	10.80	5.40	6.45
16	2	7.80	7.80	7.80
17	2	6.10	6.10	6.10
18	3	1.60	1.60	1.39
TOTAL	64	494	259.39	331.60
AVERAGE	3.5	27.45	14.41	18.42

Table 3.9: Sex-average linkage map details for family 42A

Linkage group	No. of markers	Length (cM)	Ave. spacing [s] (cM)	Largest interval (cM)
1	10	63.10	7.01	18.15
2	4	36.80	12.26	20.08
3	6	34.80	6.96	20.23
4	2	34.10	34.10	34.10
5	3	34.10	17.05	28.09
6	4	34.10	11.36	20.27
7	2	30.60	30.60	30.60
8	4	30.30	10.10	22.53
9	2	28.20	28.20	28.20
10	5	24.50	6.13	11.63
11	3	23.10	1.05	13.94
12	2	22.50	22.50	22.50
13	2	15.00	15.00	15.00
14	3	12.30	6.15	8.23
15	2	10.80	10.80	10.80
16	3	6.40	3.20	4.69
17	2	3.30	3.30	3.30
18	2	0.50	0.50	0.50
TOTAL	61	444.50	226.27	312.81
AVERAGE	3.39	24.69	12.57	17.38

Linkage groups with homologous markers of each family were merged to create sex-average linkage maps. Linkage groups were ordered from longest to shortest, as with the sex-specific maps. The maps of family 7B and 42A both assembled into 18 linkage groups; however the LG lengths and marker orders differed quite significantly (Figures 3.8 and 3.9). The genome length calculated using equation A gave a result of 1009 cM and 881cM was obtained using equation B for family 7B. The genome length calculated using equation A gave a result of 918 cM and 817 cM was obtained using equation B for family 42A. Figure 3.8 displays the sex-average linkage map of Family 7B and Figure 3.9 displays the sex-average linkage map of Family 42A.

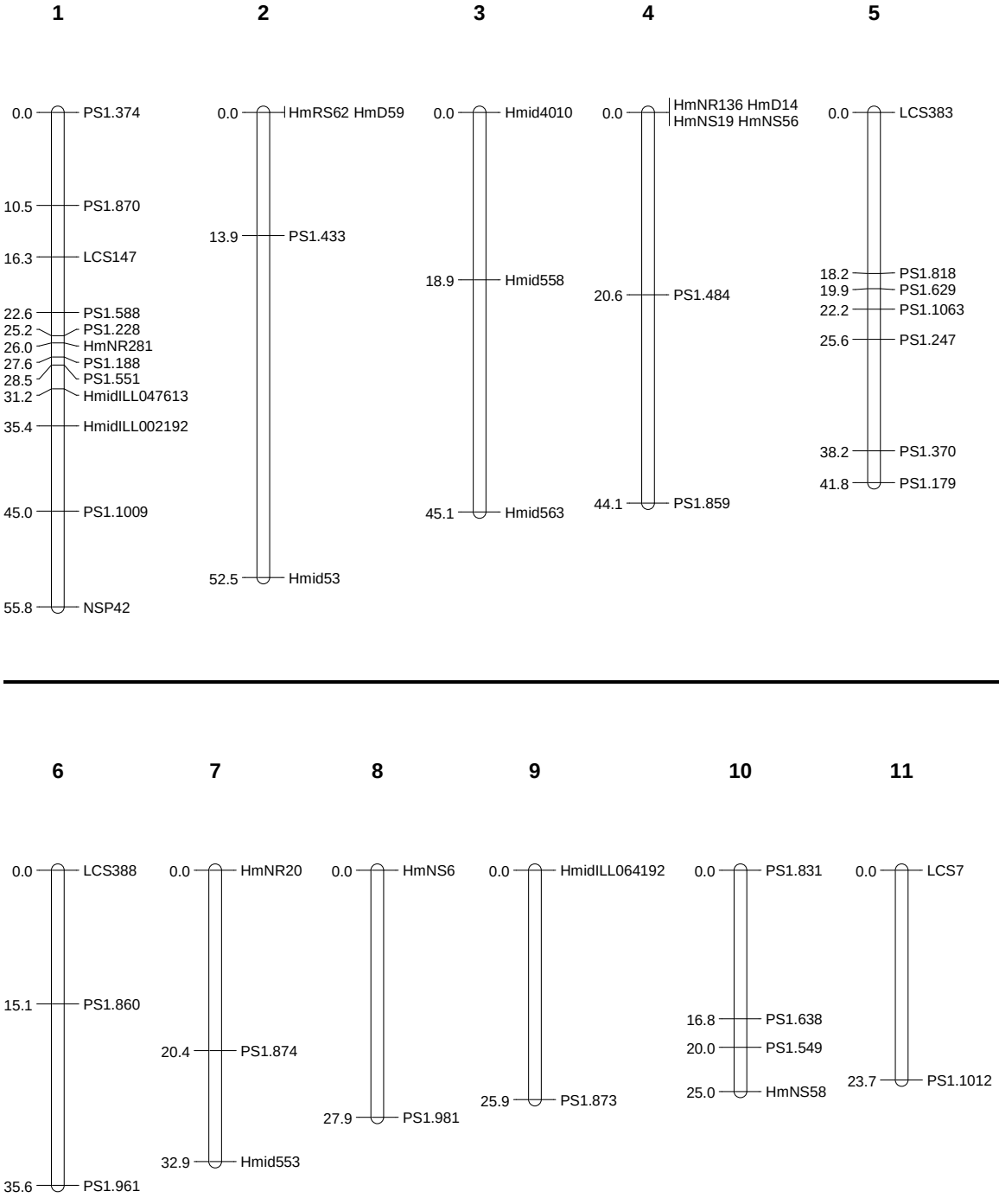


Figure 3.8: Sex-average linkage map of family 7B. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Linkage groups were placed arbitrarily from the longest to shortest in length.

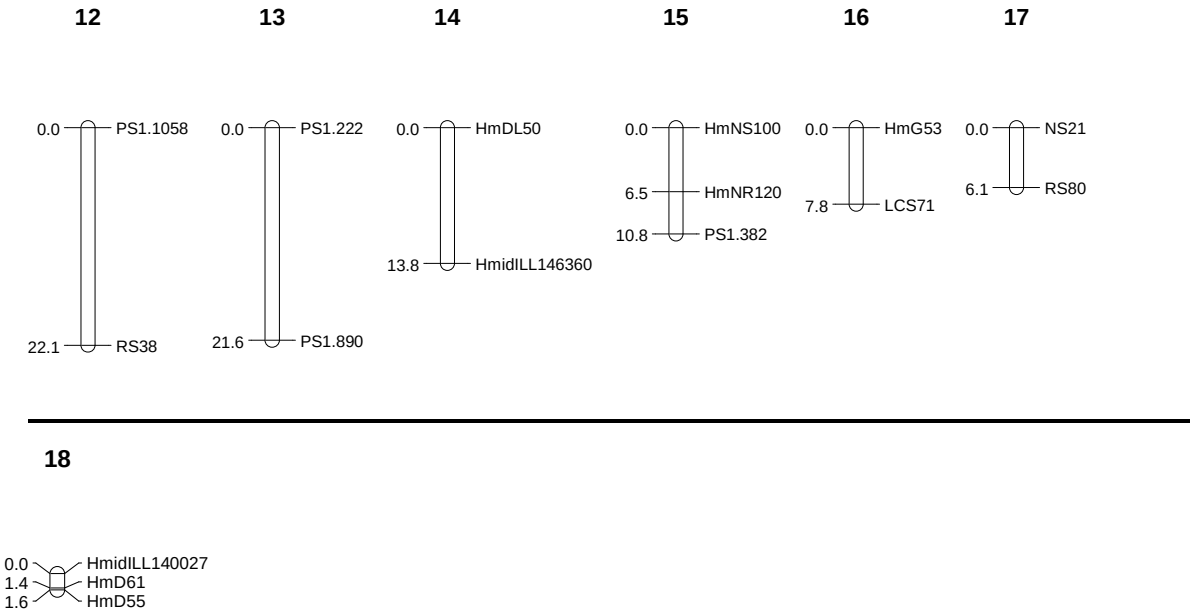


Figure 3.8 continued: Sex-average linkage map of family 7B. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Linkage groups were placed arbitrarily from the longest to shortest in length.

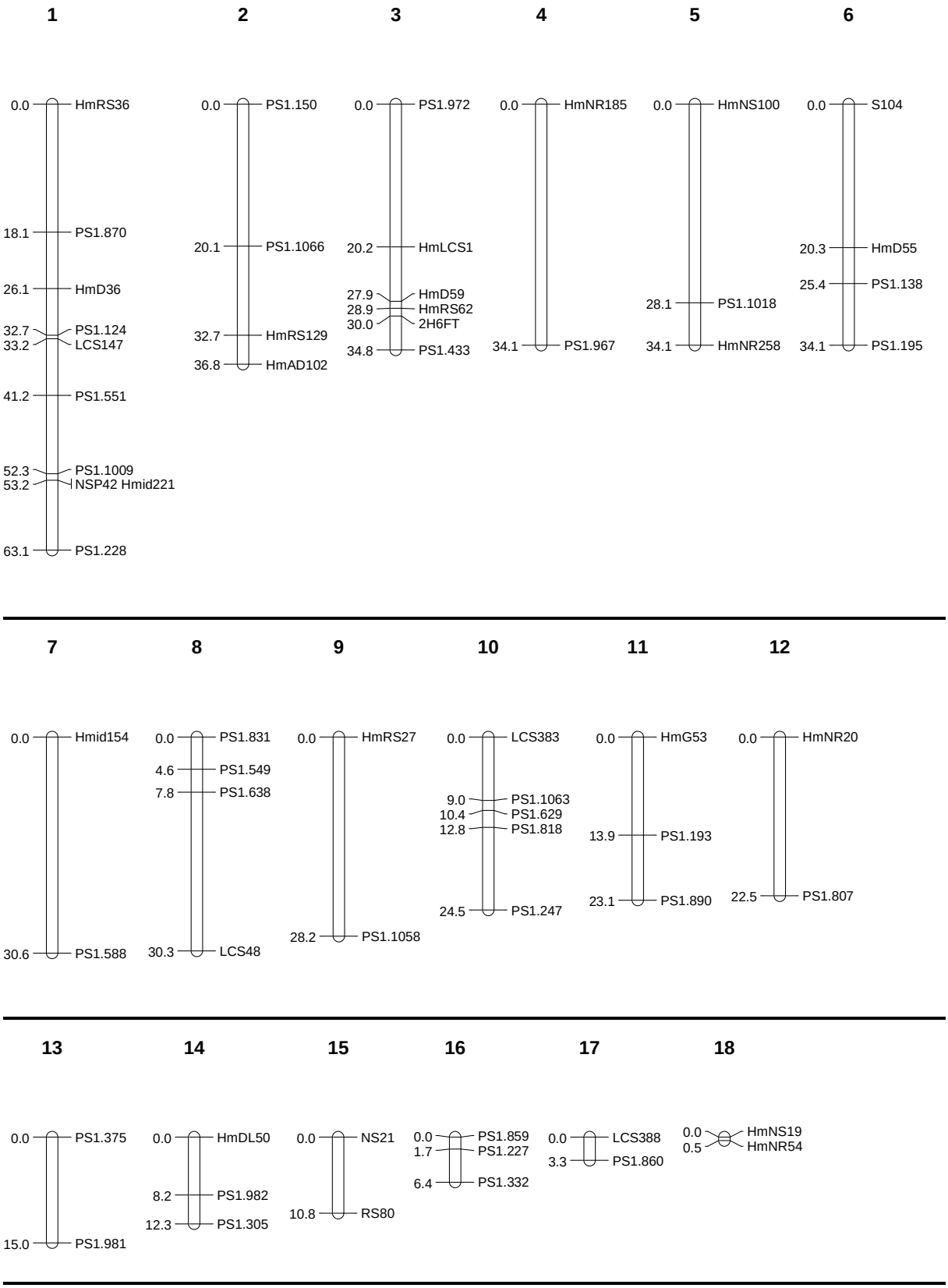


Figure 3.9: Sex-average linkage map of family 42A. Markers are displayed on the right and the cumulative mapping distance (Kosambi cM) on the left. Linkage groups were placed arbitrarily from the longest to shortest in length.

Thirteen of the linkage groups between the two families could be joined with at least one concurrent marker (Figure 3.10). The co-linearity of the marker orders was not always maintained as seen between LG10_7B and LG8_42A as well as LG5_7B and LG10_42A. There were also two cases in which two linkage groups from one family seemed to in fact to be found on one chromosome: LG13_7B and LG16_7B can both be joined to LG11_42A and LG16_42A and LG18_42A can both be joined to LG4_7B.

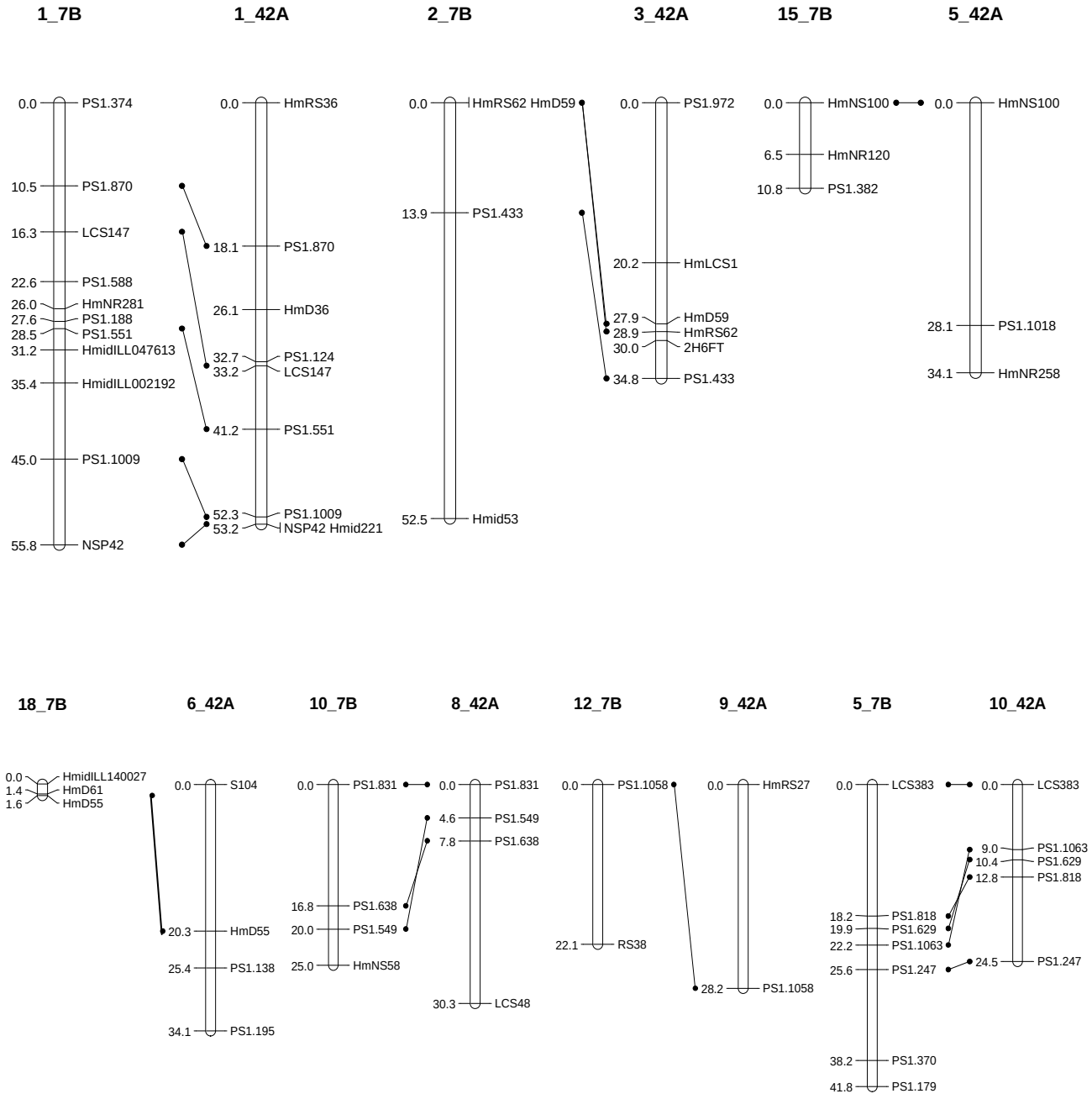


Figure 3.10: The sex-average linkage groups joined by homologous markers. The homologies are seen between LG1_7B and LG1_42A, LG2_7B and LG3_42A, LG15_7B and LG5_42A, LG18_7B and LG6_42A, LG10_7B and LG8_42A, LG12_7B and LG9_42A, LG5_7B and LG10_42A, LG13_7B and LG11_42A and LG16_7B, LG7_7B and LG12_42A, LG14_7B and LG14_42A, LG17_7B and LG15_42A, LG4_7B and LG18_42A and LG16_42A and finally LG6_7B and LG17_42A

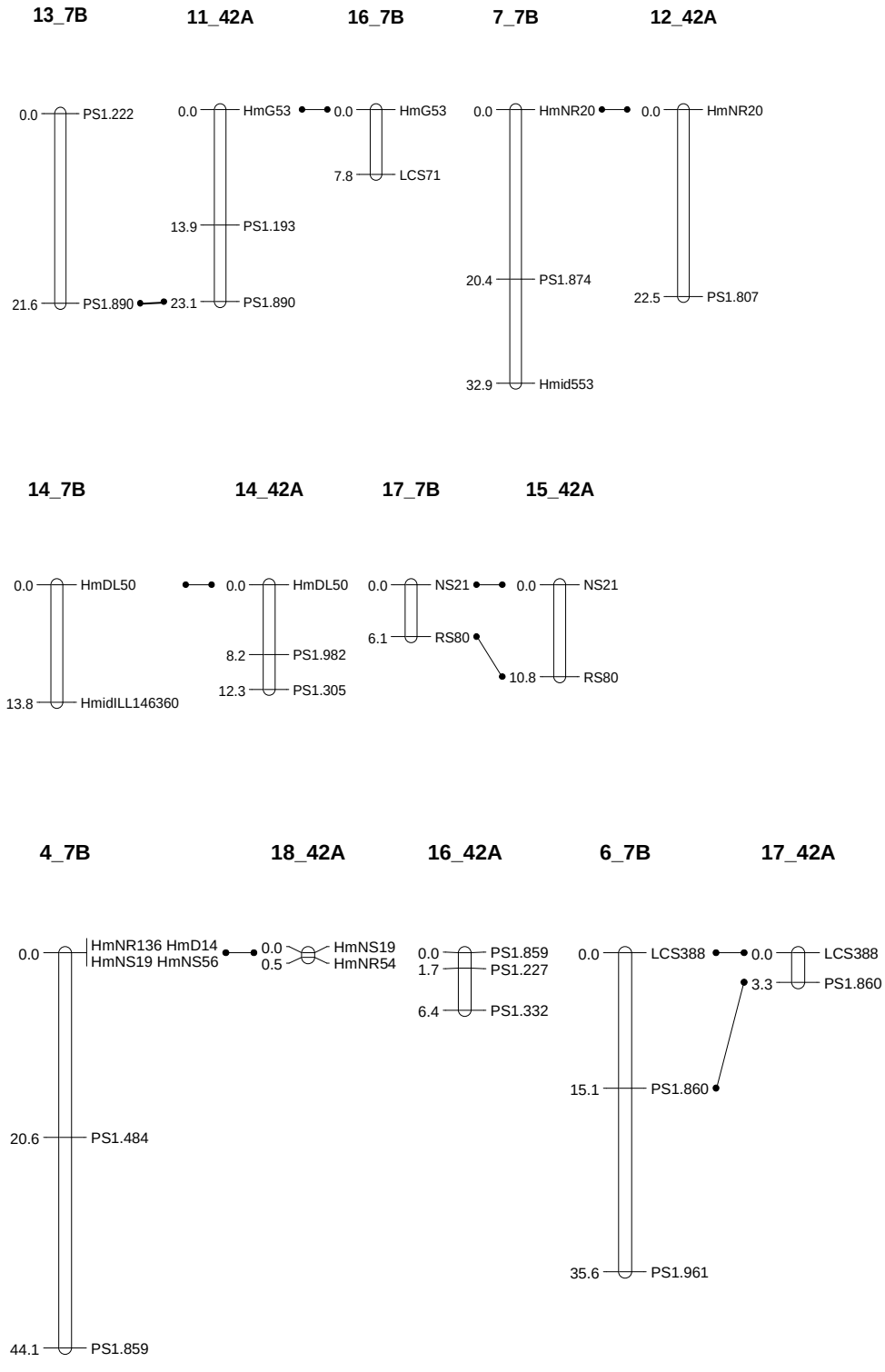


Figure 3.10 continued: The homologies are seen between LG1_7B and LG1_42A, LG2_7B and LG3_42A, LG15_7B and LG5_42A, LG18_7B and LG6_42A, LG10_7B and LG8_42A, LG12_7B and LG9_42A, LG5_7B and LG10_42A, LG13_7B and LG11_42A and LG16_7B, LG7_7B and LG12_42A, LG14_7B and LG14_42A, LG17_7B and LG15_42A, LG4_7B and LG18_42A and LG16_42A and finally LG6_7B and LG17_42A

5. Discussion

This study focused on developing a framework linkage map for Perlemoen using microsatellite markers, by joining linkage groups from sex-specific as well as sex-average linkage groups using JoinMap® v4. This is the second linkage map created for *H. midae*. The first linkage map (Badenhorst 2008) contained mostly AFLP markers which are extremely useful for map density, although the reproducibility and transferability is low compared to sequence-specific markers such as microsatellites (Hubert and Hedgecock 2004). The challenge faced with using microsatellites is the error rates that are associated with these markers. Genotyping errors can drastically inflate recombination estimates and map lengths as well as causing marker order ambiguity. An error rate as low as 5% has been shown in simulation studies to cause map inflation of up to 50% (Ball *et al*. 2010).

5.1 Linkage mapping

The haploid chromosome number of *H*. *midae* has recently been confirmed as 18 (Van der Merwe and Roodt-Wilding 2008) and is thus the expected number of linkage groups in map construction. The sex-specific linkage maps in this study all provided a linkage group number of less than 18 (*7B_P1*- **15LG**/ *7B_P2*- **16LG**/ *42A_P1*- **14LG**/ *42A_P2*- **13LG**); however, this is a common occurrence in preliminary linkage maps (Baranski *et al.* 2006a; Liu *et al*. 2006). The sex-average maps both assembled into 18 linkage groups. This does not necessarily imply that the markers are spread across all chromosomes as the coverage and density is still rather low and groups could potentially merge when more information is obtained.

The lengths of the male maps (599 cM and 518 cM for family 7B and 42A, respectively), calculated using equation A, were shorter than the female map lengths (794 cM and 595 cM for family 7B and 42A respectively) as expected. These lengths are rather low compared to the study done by Yaohua *et al*. (2010) in *H. diversicolor*, in which approximately 200 markers were used (male/female) and lengths of 1896.5 cM and 1875.2 cM were obtained for the male and female map, respectively. It is more comparable to the linkage map developed for Pacific abalone, *H. discus hannai* (167/160 markers, 702 cM male and 888 cM female; Sekino and Hara 2007) and the blacklip abalone, *H. rubra* (102/98 markers, 621 cM male and 766 cM female; Baranski *et al*. 2006a) in which fewer

markers than the former examples, were used. Sex-specific recombination frequencies in species such as abalone and fish are frequently observed, although this concept is not entirely understood (Danzmann and Gharbi 2001; Sekino and Hara 2007).

A marker interval of approximately 10 cM was calculated for each sex-specific map which is surprising with the level of marker density seen. A marker interval of 20 cM is sufficient for QTL mapping, although 10-15 cM or less is more appropriate in aquaculture species due the fact that fewer markers are informative in outbred populations (Massault *et al*. 2008).

The linkage groups of the sex-specific maps from the two families were compared and joined where homologous markers were found. Ten groups were joined with 1-4 homologous markers; however, marker order was not always conserved. In the first group comparison between the paternal maps (p1) in which linkage group 1 from family 7B_p1 and linkage group 6 from family 42A_p1 were joined (Figure 3.7), markers PS1.551 and PS1.870 as well as Nsp41 and PS1.1009 were reversed. It also seems, when comparing linkage groups 7 and 16 in family 7B_p2 to linkage 1 of family 42A_p2 (Figure 3.7), that the two groups are in fact found on one chromosome but were separated due to limited map density. The sex-average maps (Figures 3.8 and 3.9) both consisted of 18 LGs, which is equivalent to the number of haploid chromosomes found in *H*. *midae*. The map lengths were calculated using the two equations A and B, providing estimated map lengths of 1009 cM (A) and 881 cM (B) for family 7B as well as 918 cM (A) and 817 cM (B) for family 42A. When considering the sex-averaged linkage maps for the aforementioned discrepancy between marker placement of PS1.551, PS1.870, Nsp41 and PS1.1009, the orders concur with that of family 7B.

The sex-average linkage map of family 7B displays a separation of markers *Hm*G53-*Hm*LCS7 (LG16) and PS1.890 (LG13) which is not concurrent with family 42A (LG11), and seems as though these markers were found on different chromosomes. The fact that marker PS1.193 is present in family 42A and absent from 7B could be the reason for such division. This reinforces the significance of using more than one family for linkage map construction as missing data can cause major discrepancies and affect QTL mapping.

Map coverage of the two sex-average maps were calculated giving 52% coverage in family 7B and 51% coverage in family 42A. When comparing the sex-average linkage maps, 13 links could be made in which two cases were seen where marker order was inverted. It should also be noted that two linkage groups from family 7B (13 and 16) were linked to one group (11) from family 42A, as seen in the sex-specific maps. There was another link seen between linkage group 4 in 7B and linkage group 16 and 18 from family 42A.

The low coverage is corroborated by the low marker density with an average spacing of 10 cM, which is more comparable to previous studies done by Sekino and Hara (2007) and Baranski *et al.* (2006a) in other species of abalone. The marker order varied quite largely between families and between the sexes, most likely due to the level of density. The differences in markers that were polymorphic between families were expected as abalone are highly fecund and diverse.

5.2 Segregation distortion

Distorted markers found in this study were mostly restricted to one of the sexes and families. The markers were not located to a single region of a linkage group which could indicate possible deleterious fitness alleles (Launey and Hedgecock 2001). It is interesting to note that markers *Hm*DL134 and *Hm*DL214 both displayed distorted segregation patterns in the paternal map (P1) of family 42A, however when placed on the linkage map both mapped to the same LG which corresponds with the female map in family 7B. The two markers were linked to a third marker in 42A_p1, *Hm*DL34, which also corresponded with 7B_p1 and 42A_p2. The patterns of marker placement can thus be appropriately compared when using multiple families for linkage mapping analysis.

Segregation distortion has frequently been reported in bivalves (Hubert and Hedgecock 2004; Sekino and Hara 2007; Slabbert *et al*. 2010) suggesting that, due to high levels of fecundity and thus genetic load, recessive deleterious mutations and a deficiency of identical by decent (IBD) homozygotes is prevalent (Launey and Hedgecock 2001). Segregation distortion is also often a result of errors and difficulty in microsatellite genotyping (Pompanon *et al*. 2005; Ball *et al*. 2010). Dubious makers as well as samples with low DNA quality were excluded from this linkage mapping study to reduce the error rate. Mendelian segregation analysis was used as an attempt to identify marker data with

errors although in some cases it cannot be detected; for example if a parent is homozygous A_1A_1 and the other heterozygous A_1A_2 the offspring could either be A_1A_1 or A_1A_2. If some offspring are incorrectly typed as A_1A_2 then it will not necessarily be detected in segregation analysis but will cause significant effects on the recombination patterns of those individuals (Ball *et al*. 2010).

5.3 Duplicated markers

Duplicated markers are often observed where microsatellite genotyping is concerned and, as seen with those containing null alleles, markers are generally excluded from further analysis. Although the cause and mechanisms of duplicated loci is not well understood, genotypes can in some cases be manipulated for linkage mapping analysis. For example, in the case of marker NS32 in family 7B in which both parents (P1 and P2) contained four alleles namely P1: 152/155/174/176 and P2: 152/155/164/174, there are only two alleles that can distinguish between P1 and P2, 164 and 176. The offspring contain a combination of these alleles and various amounts of alleles are present in each individual: 152/155/164/176, 152/155/164/174, 152/155/174, 152/155/174/176, 155/174, 152/164/176 and 155/164/176. A method of using these genotypes would be to look at the inheritance pattern of the two alleles 164 and 176 only. In other words, does the individual contain either 164 or 176 or both? If allele 164 is present in the individual but 176 is not, then the JoinMap® v4 genotype coding will be **eg**. If allele 176 is present and 164 is not, then the genotype will be **ef**. If the individual contains 164 as well as 176 then **fg** will be the appropriate code. Lastly, if neither 164 nor 176 are present then the genotype code would be **ee**. In this way **f** can be said to segregate from P1 whereas **g** denotes an allele segregating from P2 and **e** denotes the remainder of the genotypes which are insignificant in this case. This is obviously not ideal and can only work for smaller pedigree-specific projects as it takes time and effort.

5.4 EST microsatellites

Sixteen EST-STR markers were BLASTn aligned on the NCBI database (www.ncbi.nlm.nih.gov/BLAST) to identify significant similarity to gene regions of closely related species (see Chapter 2). Two of the EST-STR markers (*Hmid*ILL1-002192 and *Hmid*ILL1-140027) segregated in family 7B and were mapped onto linkage groups 1 and 18 of the sex-average linkage map. *Hmid*ILL1-002192 segregated in the paternal map of family 7B on linkage group 1 although this marker did not segregate in the maternal map.

*Hmid*ILL1-140027 did not link to any group in the sex-specific linkage maps; however when all the data was used for a sex-average map, the marker was placed alongside *Hm*DL50 on linkage group 14 in family 7B (Figure 3.10). These two markers were not polymorphic in family 42A and were not incorporated in this sex-average map.

*Hmid*ILL1-140027 aligned to a Zebra fish RAS oncogene protein family and *Hmid*ILL1-002192 to an *H. asinina* cDNA fragment. The use of these EST-STRs specifically and ESTs in general in linkage mapping can thus provide further insight into locations and associations of genic regions to markers, making QTL analysis more significant. In other words, if markers linked to economically important traits can be identified, MAS would become a more easily obtainable goal.

5.5 Conclusions and future applications

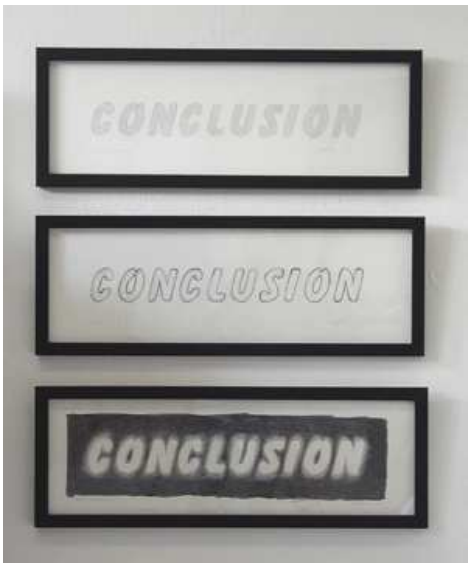
The preferred population type for linkage analysis are inbred or F_2 constructed lines, however, it is extremely challenging in aquaculture species such as *H*. *midae* due to their extended generation time and method of reproduction. The advantage of this species is the availability of large family sizes which makes it possible to develop a useful linkage map. The linkage map created in this study is a preliminary map and provides a small but significant step toward the ultimate goal of QTL mapping and MAS. The genome coverage is still only approximately 50% and the change in marker order between maps illustrates a lack of marker density.

The effects of missing data, genotyping errors and segregation distortion on linkage map construction has been reported in several studies from several species (Dodds *et al*. 2004; Lehmensiek *et al*. 2005; Pompanon *et al*. 2005; Cartwright *et al*. 2007). Although it is often unavoidable, reporting such effects, for example the differences in map lengths and marker orders, is essential in progression to improvement. Lehmensiek *et al*. (2005) carried out a curation study, in which previously developed linkage maps of wheat were re-examined to identify errors and improve the accuracy of marker placement. Missing data seemed to have more of an effect than genotyping errors, and translocations also had a large influence on marker orders. An essential finding in these comparisons was the fact that QTL mapping was improved significantly when using the revised linkage map even though simulations studies (Dodds *et al*. 2004) disagree. The impending linkage mapping projects should thus be well curated and compared to previously developed maps for

improved accuracy and ultimately developing a dense framework for QTL identification. At present, the map is still at a low density and the differences in marker orders display the need for more polymorphic markers and the addition of more abundant markers such as SNPs.

Chapter four

Conclusions and future applications



1. Microsatellite marker development

Microsatellite markers are useful for genetic linkage map construction (Hubert and Hedgecock 2004; Chistiakov *et al*. 2006) due to hypervariability, dispersal along the genome, co-dominant inheritance and the ability to multiplex for automated genotyping (Chistiakov *et al*. 2006). An additional advantage of microsatellite markers is that fewer reference families are required for construction of a map due to the amount of information retrieved per marker locus (Vignal *et al*. 2002). Microsatellites are therefore preferred markers for linkage mapping in aquaculture (Chambers and MacAvoy 2000).

Microsatellites were previously developed using hybridisation and cloning methods, such as FIASCO (Zane *et al*. 2002) and SNX (Hamilton *et al*. 1999). These are tedious and time consuming methods. More recent advances in sequencing allow faster, cheaper methods of producing markers. FIASCO and Illumina GA sequence-by-synthesis methods were used in this study for developing microsatellite markers.

Screening of colonies during the FIASCO method provided fragments (49%) containing repeats. Twenty three percent of these sequences were used to design primers after redundancy and primer designing practicality were assessed. Twenty percent were found to be useful and polymorphic, of which 32% conformed to Mendelian segregation. The Illumina sequence-by-synthesis marker development strategy provided 35% marker design success rate. Forty four percent of these markers provided reliable genotype data, of which 43% conformed to Hardy-Weinberg expectations. It can clearly be seen that the FIASCO and sequence-by-synthesis methods of marker development both have advantages and disadvantages in the process but essentially provide a very similar success rate.

PCR optimisation of markers seemed to be significantly simpler for the Illumina (sequence-by-synthesis)-derived markers than the FIASCO-derived markers. This could be attributed to the fact that primers for the FIASCO-derived markers were developed from sequences of cloned fragments and not directly from genomic sequences, as with the Illumina-derived primers. Both strategies presented genotype data in which non-reliable, duplicated, and null alleles were present, making these phenomena more of an expected microsatellite

characteristic than a development strategy flaw. Microsatellite markers, irrespective of the development strategy, should thus be used with caution. Although the genotyping and analyses is theoretically simple, STRs are notorious for abnormal behaviour in PCR and electrophoretic separation. Technical challenges and errors such as null alleles (Callen *et al.* 1993; Pemberton *et al.* 1995), large allele drop-out (Jones and Ardren 2003) and PCR artefacts are often seen; leading to misinterpretation of data which will affect the outcomes of parental assignments as well as population analysis, linkage mapping and QTL mapping. Characterisation of these markers is thus necessary to identify problematic markers as well as those most suited to a specific application.

Through the use of Mendelian segregation analysis, one can easily identify markers with abnormal inheritance patterns or peak formations. Often null alleles and allele drop-outs are missed when analysis is performed in wild populations as the inheritance cannot be monitored. Twenty seven percent null alleles were identified in the segregation analysis of FIASCO-derived markers, which is comparable to the study done by Baranski *et al*. (2006b). HWE probability tests can also be useful in identifying null alleles when analysing marker data, however, allele drop-outs and other abnormal segregation patterns are often missed.

Type I EST-STRs have been associated with functionality within coding regions as well as promotor regions and can be essential in the expression of certain genes. This changes the perception of microsatellites as neutral markers, although, the association with coding regions allows enhanced identification of QTL. Sequence-by-synthesis methods can be extremely useful as a type I marker development strategy, however, it is not a high throughput method of marker discovery and still requires time for optimisation.

The aim was to develop microsatellite markers via two methods namely; FIASCO and sequence-by-synthesis (Illumina GA) data screening, and this was achieved. The marker development using FIASCO was a collaborative effort and after numerous attempts, 44 polymorphic markers were obtained. This process is an extremely tedious and costly method causing challenges in optimisation and genotype scoring. A sharp eye for allele scoring and an understanding of segregation is necessary to keep genotype errors to a minimum. On the other hand, the markers obtained with FIASCO provided highly

polymorphic genotypes with four segregating alleles, in some cases. The markers developed from Illumina GA sequence-by-synthesis data was significantly less tedious and costly and provided type I markers. These markers have a low level of polymorphism, however, rendering them less informative in linkage mapping. Both methods have been useful although using next generation sequencing, which is continuously improving in accuracy and reducing in cost, will be the best route for future marker development projects.

2. Linkage mapping in *Haliotis midae*

The tagging system of abalone in a farming environment has proved to be rather complicated and although the retention time of bee-tags has been sufficient for mass identifications (family baskets), the success rate (60%) is not sensitive enough for single animal identification. After testing a number of strategies for tagging, RFID tags seem to provide the most effective system for single animal identification. The RFID method is much more costly than bee-tags and is far more time consuming, but does provide 100% retention, which is ideal for individual identification. It is not necessary to track individual animals in preliminary framework linkage map development; however, when more dense linkage maps are created and QTL mapping initiated, specific animal phenotypes must be tracked in order to follow certain trends. In this study, the offspring were tagged with RFID tags to assess the system for tagging of future linkage mapping families taking part in QTL studies. The success rate was high, with 100% of the tags being retained after three months. Certain precautions must be taken when tagging animals, as careless methodology can result in tag loss. The animals' shells need to be dried and the tags fully submerged in the putty, as these tags possess glass exterior and can be broken if rubbed against other animals' shells persistently for long periods of time. Animals should not be placed in water until the quickset putty is completely dry. If the tagging method is followed to the specifications in Slabbert *et al*. (in prep.), the individual identifications of animals for phenotypic monitoring will be significantly simpler than with bee-tags.

The number of haploid chromosomes in *H. midae* has recently been confirmed to be 18; so for optimal map coverage, 18 LGs should be obtained in the linkage map. It is unusual for a preliminary framework linkage maps in non-model species to cover all the chromosomes as too few markers are available. This is the second linkage map developed

for *H. midae* and the first using microsatellites only. The linkage map developed by Badenhorst (2008) was performed using one full-sib family and incorporated mostly AFLP markers and some microsatellites that were available. Badenhorst (2008) obtained 10 and 12 linkage groups (male and female), a genome length of 743 cM and 1470 cM (male and female) and an average marker spacing of 20 cM.

Two full-sib families were used in this study in which 106 polymorphic microsatellite loci were available for analysis. The number of linkage groups obtained from sex-specific maps ranged from 13-16, which is a slight improvement from the original map. The average genome length ranged from 500 cM to 800 cM with an average marker spacing of 10 cM. The genome length calculation was lower than the calculation in the map constructed by Badenhorst (2008), however the spacing between markers is significantly smaller. Approximately half the amount of polymorphic markers were eventually linked to a LG in each sex-specific map after removal of distorted markers and exclusion of ungrouped and unlinked loci (LOD>3). This is a major loss of markers and could possibly be attributed to factors such as limited recombination data available for positioning of markers by the mapping algorithms. This can be overcome with the incorporation of many more markers and by using additional mapping families to improve the number of informative markers and homology across families.

The sex-average linkage map provided 18 linkage groups with an average genome length calculation of 1800 cM and average marker spacing of approximately 13 cM, which is slightly more promising. A marker interval of 20 cM is sufficient for QTL mapping, although 10-15 cM or less is more appropriate in aquaculture species due the fact that fewer markers are informative in outbred populations (Massault *et al*. 2008). The coverage (50%) would definitely need to be improved to ensure dependable segregation alongside the trait in every population (Collard *et al*. 2005). The addition of informative markers will also allow easier merging of linkage groups as well as linkage maps, i.e. there will be more anchor markers prevalent between families. This will allow the merging of future linkage maps to the maps created in this studay as well as the first framework map developed by Badenhorst (2008).

The aim in this chapter was to develop sex-specific and sex-average linkage maps of *H. midae*, from two unrelated families using 100 full-sib offspring, and join the maps where homologous markers were found. This aim was achieved and was performed in JoinMap® v4, which is an extremely user-friendly mapping program.

3. Future applications and improvements

Microsatellite markers are indeed extremely useful markers for construction of genetic linkage maps as they are not only dispersed along the genome of most organisms, but are highly polymorphic as well. The markers found in non-coding regions are especially polymorphic making segregation analysis in a full-sib family effortless, in most cases. Microsatellites do harbour some complications though, as all markers seem to. The theory of allele segregation and observation of segregation can often be contradictory such as those seen with null alleles, allele drop-outs and stutter bands. These markers should thus be characterised in pedigree data in which parental genotypes can be used as a reference, so that these complications can be identified and either rectified by redesigning primers or used with caution. Microsatellite markers (type I and type II) provide an extremely useful framework for linkage maps, however, when looking at the effects of errors associated with these markers, it would be even more beneficial to add high throughput markers such as SNPs (Ball *et al*. 2010). SNPs are more abundant and can be genotyped with high throughput methods, which will facilitate the development of a high-density linkage map.

The advantage of abalone as an outcross population is the extremely high fecundity of the species, providing large families for linkage analysis. Tagging methodology for these animals when performing non-destructive sampling is essential for future QTL studies so that individuals can be monitored. The RFID tagging system proved to be most beneficial, however, an even more optimal system would be to keep the mapping family animals separate from all other animals and use genetic barcodes in addition to RFID tags. In this study, 100 offspring from two unrelated full-sib families were used for linkage map construction. Two families is an improvement from one, as it provides double the probability of obtaining informative marker genotypes. When using EST-derived markers, which generally possess a low level of polymorphism, using more families would be even more beneficial. Increasing the number of families, and thereby increasing the number of

informative marker data, will also improve the ability to join linkage groups and merge maps from different families.

This study provides a step toward creating a saturated map for the South African commercial abalone, *H. midae*. This has become a priority not only for this species but more generally in all genetic research on cultured *Haliotis* species in order to develop maps with sufficient markers for future QTL mapping and ultimately MAS.

References

References

Abdelkrim, J., Robertson, B.C., Stanton, J.-A.S., and Gemmell, N.J. (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques* **46**: 185-192

Agresti, J.J., Seki, S., Cnaani, A., Poompuang, S., Hallerman, E.M., Umiel, N., Hulata, G., Gall, G.A.E., and May, B. (2000) Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. *Aquaculture* **185**: 43-56

Allentoft, M.E., Schuster, S.C., Holdaway, R.N., Hale, M.L., McLay, E., Oskam, C., Gilbert, M.T.P., Spencer, P., Willerslev, E., and Bunce, M. (2009) Identification of microsatellites from an extinct moa species using high throughput (454) sequence data. *BioTechniques* **46**: 195-200

Amos, W. (1999) A comparative approach to the study of microsatellite evolution. In: *Microsatellites, Evolution and Applications*. Goldstein, D.B., and Schlötterer, C. (eds). Oxford University Press. pp 66-79

Araki, E., Shimada, F., Uzawa, H., Mori, M., and Ebina, Y. (1987) Characterisation of the promoter region of the human insulin receptor gene. Evidence for promoter activity. *The Journal of Biological Chemistry* **262**: 16186-16191

Armour, J.A., Neumann, R., Gobert, S., and Jeffereys, A.J. (1994) Isolation of human simple repeat loci by hybridisation selection. *Human Molecular Genetics* **3**: 599-605

Badenhorst, D. (2008) Development of AFLP markers for *Haliotis midae* for linkage mapping. [Unpublished Master of Science thesis] Stellenbosch University, South Africa.

Ball, A.D., Stapley, J., Dawson, D.A., Birkhead, T.R., Burke, T., and Slate, J. (2010) A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the Zebra finch (*Taeniopygia guttata*). *BMC Genomics* **11**: 218-232

Baranski, M.D. (2006) Molecular genetic markers for abalone aquaculture. [Unpublished Doctor of Philosophy thesis] Deakin University, Victoria, Australia.

References

Baranski, M., Loughnan, S., Austin, C., and Robinson N. (2006a) Linkage maps of microsatellite DNA markers for the blacklip abalone *Haliotis rubra*. *Animal Genetics* **37**: 563-570

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

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

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L., and Schnable, P.S. (2007) SNP discovery via 454 transcriptome sequencing. *Plant Journal* **51**: 910-918

Barker, M.S., Dlugosch, K.M., Reddy, A.C.C., Amyotte, S.N., and Rieseberg, L.H. (2009) SCARF: maximizing next-generation EST assemblies for evolutionary and population genomic analysis. *Bioinformatics* **25**: 535-536

Beaumont, A.R., and Hoare, H. (2003) *Biotechnology and genetics in fisheries and aquaculture*. Oxford: Blackwell Science Ltd. pp 47-88

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

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

Birren, B., Green, E.D., Hieter, P., Klapholz, S., Myers, R.M., Riethman, H., and Roskams, J. (1997) *Genome Analysis. Mapping Genomes.* Cold Springs Harbour Laboratory Press. pp 33-48

Bonfield, J.K., and Whitwham, A. (2010) Gap5-editing the billion fragment sequence assembly. *Bioinformatics* doi:10.1093/bioinformatics/btq268

References

Botes, L., Thompson, G., and Louw, R. (2006) Benchmarking survey of the South African aquaculture (marine and fresh water) sector. AISA (Aquaculture Institute of South Africa) report [Available at: http://www.ai-sa.org.za/newsletter/documents/ASIABenchmarkingSurvey2006_001.pdf; accessed Jan 2010] pp 11-13

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **34**: 314-331

Bouck, A., and Vision, T. (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907-924

Bouza, C., Hermida, M., Pardo, B.G., Fernáandez, C., Fortes, G.G., Castro, J., Sánchez, L., Presa, P., Pérez, M., Sanjuán, A., de Carlos, A., Álvarez-Dios, J.A., Ezcurra, S., Cal, R.M., Piferrer, F., and Martínez, P. (2007) A microsatellite genetic map of the Turbot (*Scophthalmus maximus*). *Genetics* **177**: 2457-2467

Brandström, M., Bagshaw, A.T., Gemmell, N.J., and Ellegrans, H. (2008) The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Molecular Biology and Evolution* **25**: 2579-2587

Breen, P.A., Kim, S.W., and Andrew, N.L. (2003) A length-based Bayesian stock assessment model for the New Zealand abalone *Haliotis iris*. *Marine and Freshwater Research* **54**: 619-634

Brink, D., Vlok, A.C., Sadie, A., Roodt-Wilding, R., and Slabbert, R. (2009) Genetic improvement of the South African abalone (*Haliotis midae*) through selective breeding. Conference presentation, International Abalone Symposium, Pattaya, Thailand, June 2009

Brinkman, B., Klintschar, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998) Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. *American Journal of Genetics* **62**: 1408-1415

Britz, P.J., and Lee, B. (2009) AISA Aquaculture Benchmarking Survey. Volume 1: Primary Production. A report for the Aquaculture Institute of South Africa and Swiss Contact, produced by Enviro-Fish Africa (Pty.) Ltd. pp 73

References

Brohede, J., Primmer, C.R., Møller, A., and Ellegren, H. (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Research* **30**: 1997-2003

Brookfield, J.F.Y. (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**: 453-455

Buschiazzo, E., and Gemmell, N.J. (2006) The rise, fall and renaissance of microsatellites in eukaryote genomes. *BioEssays* **28**: 1040-1050

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C., and Sutherland, G.R. (1993) Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. *American Journal of Human Genetics* **52**: 922-927

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

Cartwright, D.A., Troggio, M., Velasco, R., and Gutin, A. (2007) Genetic mapping in the presence of genotyping. *Genetics* **176**: 2521-2527

Castro, J., Pino, A., Hermida, M., Bouza, C., Chavarrías, D., Merino, P., Sánchez, L., and Martínez, P. (2007) A microsatellite marker tool for parentage assessment in gilthead seabream (*Sparus aurata*). *Aquaculture* **272**: 210-216

Chakravarti, A., Lasher, L.K., and Reefer, J.E. (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* **128**: 175-182

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

Charo-Karisa, H., Komen, H., Rezk, M.A., Ponzoni, R.W., Van Arendonk, J.A.M., and Bovenhuis, H. (2006) Heritability estimates and response to selection for growth of Nile tilapia (*Oreochromis niloticus*) in low-input earthen ponds. *Aquaculture* **261**: 479-486

References

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

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A., and Volckaert, F.A.M. (2005) A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. *Genetics* **170:** 1821-1826

Coibanu, D.C., Bastiaansen, J.W.M., Magrin, J., Rocha, J.L., Jiang, D-H., Yu, N., Gieger, B., Deeb, N., Rocha, D., Gong, H., Kinghorn, B.P., Plastow, G.S., Van der Steen., H.A.M., and Mileham, A.J. (2009) A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*) *Animal Genetics* **41**: 39-47

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*. *Aquacuture* **220**: 203-218

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., and Pang, E.C.K. (2005) An introduction to markers, quantitative trait loci (QTLs) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**: 169-196

Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, 2007) *Fourteenth meeting of the Conference of the Parties, The Hague, Netherlands*, 3-15 June 2007

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

Crawford, A.M., Dodds, K.G., Ede, A.J., Pierson, C.A., Montgomery, G.W., Garmomway, H.G., Beattie, A.E., Davies, K., Maddox, J.F., Kappes, S.W., Stone, R.T., Nguyen,T.C., Penty, J.M., Lord, E.A., Broom, J.E., Buitkamp, J., Schwaiger, W. Epplen, J.T., Matthew, P., Matthew, M.E., Hulme, D.J., Beh, K.J., McGraw R.A., and Beattie, C.W. (1995) An autosomal genetic linkage map of the sheep genome. *Genetics* **140**: 703-724

References

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

Danzmann, R.G., and Gharbi, K. (2001) Gene mapping in fishes: A means to an end. *Genetica* **111**: 3-23

Danzmann, R.G., and Gharbi, K. (2007) Linkage mapping in aquaculture species. In: *Aquaculture genome technologies*. Liu, Z. (ed). Blackwell Publishing Ltd. USA. pp 139-169

Danzmann, R.G., Jackson, T.R., and Ferguson, M.M. (1999) Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. *Aquaculture* **173**: 45-58

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

Dempewolf, H., Kane, N.C., Ostevik, K.L., Geleta, M., Barker, M.S., Lai, Z., Stewart, M.L., Bekele, E., Engels, J.M.M., Cronk, Q.C.B., and Rieseberg, L.H. (2010) Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass. The development of a library of expressed sequence tags, microsatellite loci and the sequence of its chloroplast genome. *Molecular Ecology Resources*, no. doi:10.1111/j.1755-0998.2010.02859.x

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* **39**: 1-38

Ditlecadet, D., Dufresne, F., Le François, N.R., and Blier, P.U. (2006) Applying microsatellites in two commercial strains of Arctic char (*Salvelinus alpinus*): Potential for a selective breeding program. *Aquaculture* **257**: 37-43

Dodds, K.G., Ball, R., Djorovic, N., and Carson, S.D. (2004) The effect of an imprecise map on interval mapping QTL. *Genetical Research* **84**: 47-55

Donis-Keller, H., Green, P., Helms, C., Carinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowen, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G., Rediker,

K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J., and Abrahamson, J. (1987) A genetic linkage map of the human genome. *Cell* **51:** 319-337

Du, Z-Q., Ciobanu, D.C., Onteru, S.K., Gorbach, D., Mileham, A.J., Jaramillo, G., and Rothschild, M.F. (2009) A gene-based SNP linkage map for Pacific white shrimp, *Litopenaeus vannamei*. *Animal Genetics* **41**: 286-294

Dunham, R.A. (2004) Gene mapping, QTL mapping and marker assisted selection. In: *Aquaculture and fisheries biotechnology*: *Genetic approaches*. CABI Publishing, USA. pp 122-140

Ellegren, H. (2000) Microsatellite mutations in the germline: Implication for evolutionary interference. *Trends in Genetics* **16**: 551-558

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

Ellegren, H., Chowdhary, B.P., Jahoansson, M., Marklund, L., Fredholm, M., Gustavsson, I., and Andersson, L. (1994) A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics* **137**: 1089-1100

Elo, K., Ivanoff, S., Vuorinen, J.A., and Piironen, J. (1997) Inheritance of RAPD markers and detection of interspecific hybridisation with brown trout and Atlantic salmon. *Aquaculture* **152**: 55-65

Evans, B., Bartlett, 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

Fallu, R. (1991) *Abalone farming*. Oxford: Fishing News Books. A division of Blackwell Scientific Publications Ltd. pp 1-120

References

FAO (2006) The State of World Fisheries and Aquaculture 2004 (SOFIA). FAO Fisheries and Aquaculture Department. Food and Agriculture Organisation of the United Nations.

FAO (2009) The State of World Fisheries and Aquaculture 2008 (SOFIA). FAO Fisheries and Aquaculture Department. Food and Agriculture Organisation of the United Nations.

Farber, C.R., and Medrano, J.F. (2004) Identification of putative homology between horse microsatellite flanking sequences and cross-species ESTs, mRNAs and genomic sequences. *Animal Genetics* **35**: 28-33

Ferguson, M.M., and Danzmann, R.G. (1998) Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Science* **55**: 1553-1563

Ferguson, A., Taggart, J.B., Prodöhl, P.A., McMeel, O., Thompson, C., Stone, C., McGinnity, P., and Hynes, R.A. (1995) The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo. Journal of Fish Biology* **47**: 103-126

Fishman, L., Kelly, A.J., Morgan, E., and Willis, J.H. (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701-1716

Franch, R., Louro, B., Tsalavouta, M., Chatziplis, D., Tsigenopoulos, C.S., Sarropoulou, E., Antonello, J., Magoulas, A., Mylonas, C.C., Babbucci, M., Patarnello, T., Power, D.M., Kotoulas, G., and Bargelloni, L. (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics* **174**: 851-861

Frio, T.R., Panek, S., Iseli, C., Di Gioia, S.A., Kumar, A., Gal, A., and Rivolta, C. (2010) Ultra high throughput sequencing excludes MDH1 as candidate gene for RP-28-linked retinitis pigmentosa. *Molecular Vision* **15**: 2627-2633

Gallardo, W.G., Bautista-Teruel, M.N., Fermin, A.C., and Maite, C.L. (2003) Shell marking by artificial feeding of the tropical abalone *Haliotis asinina* Linne juveniles for sea ranching and stock enhancement. *Aquaculture Research* **34**: 839-842

References

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

Geiger, D.L. (2000) Distribution and biogeography of the Haliotidae (Gastropoda: Vestigastrodpoda) world-wide. *Bolletino Malacologia* **35**: 57-120

Genade, A.B., Hirst, A.L., and Smit C.J. (1988) Observations on the spawning, development and rearing of the South African abalone *Haliotis midae* Linn. *South African Journal of Marine Science* **6**: 3-12

Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Hoyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., Okamoto, N., Ferguson, M.M., Holm, L.E., and Guyomard, R., (2006) A linkage map for brown trout (*Salmo trutta*): Chromosome homeologies and comparative genome organization with other Salmonid fish. *Genetics* **172**: 2405-2419

Gilbey, J., Verspoor, E., McLay, A., and Houlihan, D. (2004) A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Animal Genetics* **35**: 98-105

Ginot, F., Bordelais, I., Nguyen, S., and Gyapay, G (1996) Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research* **24**: 540-541

Gjoen, H.M., and Bensten, H.B. (1997) Past, present and future of genetic improvement in Salmon aquaculture. *Journal of Marine Science* **54**: 1009-1014

Glenn, T.C., and Schable, N.A. (2005) Isolating microsatellite DNA loci. *Methods in Enzymology* **395**: 202-222

Goldstein, D.B., and Schlötterer, C. (1999) *Microsatellites: Evolution and Applications*. Oxford University Press, New York. pp 1-22

Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463-471

References

Graham, I.A., Besser, K., Blumer, S., Branigan, C.A., Czechowski, T., Elias, L., Guterman, I., Harvey, D., Isaac, P.G., Khan, A.M., Larson, T.R., Pawson, T., Penfield, T., Rae, A.M., Rathbone, D.A., Ross, J., Smallwood, M.F., Segura, V., Towsend, T., Vyas, D., Winzer, T., and Bowels, D. (2010) The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**: 328-331

Grant, W.S., and Bowen, B.W. (2000) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society* **88**: 673-689

Gruenthal K.M., and Burton R.S. (2005) Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conservation Genetics* **6**: 929-939

Guo, X., Ford, S.E., De Brosse, G., Smolowitz, R., and Sunila, I. (2003) Breeding and evaluation of eastern oyster strains selected for MSX, DERMO and JOD resistance (abstract). *Journal of Shellfish Research* **22**: 353-354

Guyomard, R., Mauger, S., Tabet-Canale, K., Martineau, S., Genet, C., Krieg, F., and Quillet, E. (2006) A type I and type II microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms. *BMC Genomics* **7**: 302-314

Haldane, J.B.S. (1919) The combination of linkage values and the calculation of distance between the loci of the linked factors. *Journal of Genetics* **8**: 299-309

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

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

Hancock, J.M. (1999) Microsatellites and other simple sequences: Genomic context and mutational mechanisms. In: *Microsatellites: Evolution and Applications*. Goldstein, D.B., and Schlötterer, C. (eds). Oxford University Press, England. pp 1-9

References

Hassanien, H.A. (2008) Use of randomly amplified polymorphic DNA (RAPD) analysis to detect genetic variation in sea bass (*Dicentrarchus labrax*). *Journal of Fisheries and Aquatic Science* **3**: 39-46

Hauck, M., and Sweijd, N.A. (1999) A case study of poaching in South Africa and its impact on fisheries management. *Journal of Marine Science* **56**: 1024-1032

Haung, W., and Marth, G. (2008) Eagleview: A genome assembly viewer for next generation sequencing technologies. *Genome Research* **9**: 1538-1543

Haung, X., and Madan, A. (1999) Cap3: A DNA sequence assembly program. *Genome Research* **9**: 868-877

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

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

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

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Stewart, D.A., Stromberg, M., Tsung, E.F., Wylie, T., Schedl, T., Wilson, R.K., and Mardis, E.R. (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* **5**: 183-188

Hoffman, A.A., Turelli, M., and Harshman, L.G. (1990) Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans. Genetics* **126**: 933-948

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

References

Hubert, S., and Hedgecock, D. (2004) Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* **168:** 351-362

Hubert, S., Higgins, B., Borza, T., and Bowman, S. (2010) Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics* **11**: 191-204

Huehn, M. (2010) Random variability of map distances based on Kosambi's and Haldane's mapping functions. *Journal of Applied Genetics* **51**: 27-31

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

Hulbert, S.H., Terrence, W.I., Legg, E.J., Lincoln, S.E., Lander, E.S., and Michelmore, R.W. (1988) Genetic analysis of the fungus, *Bremia lactucae*, using Restriction Fragment Length Polymorphisms. *Genetics* **120**: 947-958

Hyten, D.L., Cannon, S.B., Song, Q., Weeks, N., Fickus, A.D., Shoemaker, R.C., Specht, J.E., Farmer, A.D., May, G.D., and Gregan, P.B. (2010) High throughput SNP discovery through deep resequencing of a reduced representation library to anchor orient scaffolds in the soybean whole genome sequence. *BMC Genomics* **11**: 38-45

Jackson, T.R., Martin-Robichauda, D.J., and Reith, M.E. (2003) Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. *Aquaculture* **220**: 245-259

Jansen, J., De Jong, A.G., and Van Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theoretical and Applied Genetics* **102**: 1113-1122

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., and Armour, J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* **6**: 136-145

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

References

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

Kandpal, R.P., Kandpal, G., and Weissman, S.M. (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers*. Proceedings of the National Academy of Sciences* (*USA*) **91**: 88-92

Kappes, S.M., Keele, J.W., Stone, R.T., McGraw, R.A., Sonstegard, T.S., Smith, T.P., Lopez-Corrales, N.L., and Beattie, C.W. (1997) A second-generation linkage map of the bovine genome. *Genome Research* **7**: 235-249

Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., Ju, Z., Mickett, K.D., and Liu, Z. (2002) Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene* **285**: 157-168

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983) Optimisation by simulated annealing. *Science* **220**: 671-680

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D., and McAndrew, B. (1998) A genetic linkage map of a cichlid fish, the tilapia *Oreochromis niloticus*. *Genetics* **148**: 1225-1232

Kosambi, D.D. (1944) The estimation of map distances from recombination values. *Annals of Eugenics* **12**: 172-175

Kube, P.D., Appleyard, S.A., and Elliot, N.G. (2007) Selective breeding greenlip abalone (*Haliotis laevigata*): Preliminary results and issues. *Journal of Shellfish Research* **26**: 821-824

Kucuktas, H., Wang, S., Li, P., He, C., Xu, P., Sha, Z., Liu, H., Jiang, Y., Baoprasertkul, P., Somridhivej, B., Wang, Y., Abernathy, J., Guo, X., Liu, L., Muir W., and Liu, Z. (2009) Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics* **181**: 1649-1660

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

Lallias, D., Lapegue, S., Hecquet, C., Boudry, P., and Beaumont, A.R. (2007) AFLP-based genetic linkage maps of the blue mussel (*Mytilus edulis*). *Animal Genetics* **38**: 340-349

Landau, M. (1992) *Introduction to aquaculture*. Stockton State College. John Whiley and Sons, Inc. pp 181-186

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newbury, L. (1987) MAPMAKER, an interactive computer package for constructing primary genetic linkage map of experimental and natural populations. *Genomics* **1**: 174-180

Lang, R.P., Bayne, C.J., Camara, M.D., Cunningham, C., Jenny, M.J., and Langdon, C.J. (2009) Transcriptome profiling of selectively bred pacific oyster *Crassostrea gigas* families that differ in tolerance of heat shock. *Marine Biotechnology* **11**: 650-668

Launey, S., and Hedgecock, D. (2001) High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* **159**: 255-265

Leclercq, S., Rivals, E., and Jarne, P. (2010) DNA slippage occurs at microsatellite loci without minimal threshold lengths in humans: A comparative genomic approach. *Genome Biology and Evolution* **2**: 325-335

Lee, B.Y., Lee, W.J., Streelman, J.T., Carleton, K.L., Howe, A.E., Hulata, G., Slettan, A., Stern, J.E., Terai, Y., and Kocher, T.D. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* **170**: 237-244

Lehmensiek, A., Eckermann, P.J., Verbyla, A.P., Appels, R., Sutherland, M.W., and Daggard, G.E. (2005) Curation of wheat maps to improve map accuracy and QTL detection. *Australian Journal of Agriculture Research* **56**: 1347-1354

Leu, J.-H., Chang, C.-C., Wu, J.-L., Hsu, C.-W., Hironon, I., Aoki, T., Juan, H.-F., Lo, C.-F., Kou, G.-H., and Huang, H.-C. (2007) Comparative analysis of differentially

expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*. *BMC Genomics* **8**: 120-133

Levinson, G., and Gutman, G.A. (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* **15**: 5323-5338

Li, L., and Guo, X. (2004) AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology* **6**: 26-36

Li, Q., Yu, H., and Yu, R. (2006) Genetic variability assessed by microsatellites in cultured populations of the Pacific oyster (*Crassostrea gigas*) in China. *Aquaculture* **259**: 95-102

Li, Y.-C., Korol, A.B., Fahima, T., and Nevo, E. (2004) Microsatellites within genes: Structure, function and evolution. *Molecular Biology and Evolution* **21**: 991-1007

Li, Y.-C., Korol, A.B., Fahima, T., Beiles, A., and Nevo, E. (2002) Microsatellites: Genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* **11**: 2453-2465

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

Lucas, T., Macbeth, M., Degnan, S.M., Knibb, W., and Degnan, B.M. (2006) Heritability estimates for growth in tropical abalone *Haliotis asinina* using microsatellites to assign parentage. *Aquaculture* **256**: 146-152

Lundy, C.J., Rico, C., and Hewitt, G.M. (2000) Temporal and spatial genetic variation in spawning grounds of European hake (*Merluccius merluccius*) in the Bay of Biscay. *Molecular Ecology* **9**: 2067-2079

References

Maddox, J.F., Davies, K.P., Crawford, A.M., Hulme, D.J., Vaiman, D., Cribiu, E.P., Freking, B.A., Beh, K.J., Cockett, N.E., Kang, N., Riffkin, C.D., Drinkwater, R., Moore, S.S., Dodds, K.G., Lumsden, J.M., van Stijn, T.C., Phua, S.H., Adelson, D.L., Burkin, H.R., Broom, J.E., Buitkamp, J., Cambridge, L., Cushwa, W.T., Gerard, E., Galloway, S.M., Harrison, B., Hawken, R.J., Hiendleder, S., Henry, H.M., Medrano, J.F., Paterson, K.A., Schibler, L., Stone, R.T., and van Hest, B. (2001) An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research* **11**: 1275-1289

Marchant, S., Hayes, P.A., Marin, S.A., and Winkler, F.M. (2009) Genetic variability revealed with microsatellite markers in an introduced population of the abalone *Haliotis discus hannai* Ino. *Aquaculture Research* **40**: 298-304

Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133-141

Martinez, V., Thorgaard, G., Robinson, B., and Sillanpää, M.J. (2005) An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects. *Genetical Research* **86**: 209-221

Martyniuk, C.J., Perry, G.M.L., Mogahadam, H.K., Ferguson, M.M., and Danzmann, R.G. (2003) The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout. *Journal of Fish Biology* **63**: 746-764

Massault, C., Bovenhuis, H., Haley, C., and De Koning, D. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285**: 23-29

Maughan, P.J., Yourstone, S.M., Jellen, E.N., and Udall, J.A. (2009) SNP discovery via genomic reduction, barcoding and 454 pyrosequencing in amaranth. *Plant Genome* **2**: 260-270

Menon, R., Zhang, Q., Zhang, Y., Fermin, D., Bardeesy, N., DePinho, R.A., Lu, C., Hanash, S.M., Omenn, G.S., and States, D.J. (2009) Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. *Cancer Research* **69**: 300-309

References

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

Moen, T., Hayes, B., Baranski, M., Berg, P.R., Kjøglum, S., Koop, B.F., Davidson, W.S., Omholt, S.W., and Lien, S. (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* **9**: 223-237

Moen, T., Hoyheim, B., Munck, H., and Gomez-Raya, L. (2004) A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Animal Genetics* **35**: 81-92

Mullis, K.B., and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction*. Methods in Enzymology* **155**: 335-350

Naciri, Y., Vigouroux, Y., Dallas, J., Desmarais, E., Delsert, C., and Bonhomme, F. (1995) Identification and inheritance of $(GA/TC)_n$ and $(AC/GT)_n$ repeats in the European flat oyster *Ostrea edulis* (L.). *Molecular Marine Biology and Biotechnology* **4**: 83-89

Newman, G.G. (1966) Movements of the South African abalone *Haliotis midae*. *Division of Sea Fisheries of South Africa Investigational Report* **56**: 1-20

Novaes, E., Drost, D.R., Farmerie, W.G., Pappas Jr, G.J., Grattapaglia, D., Sederoff, R.R., and Kirst, M. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterised genome. *BMC Genomics* **9**: 312

O'Brien, S.J. (1991) Molecular genome mapping lessons and prospects. *Current Opinion in Genetics and Development* **1**: 105-111

Ohara, E., Nishimura, T., Nagakura, Y., Sakamoto, T., Mushiake, K., and Okamoto, N. (2005) Genetic linkage maps of two yellowtails. *Aquaculture* **244**: 41-48

Okumus, I., and Ciftci, Y. (2003) Fish population genetics and molecular markers: II-Molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquatic Sciences* **3**: 51-79

References

Ostrander, E.A., Jong, P.M., Rine, J., and Duyk, G. (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences* (*USA*) **89**: 3419-3423

Ott, J. (1999) Aspects of statistical inference. In: *Analysis of human genetic linkage*. The John Hopkins University Press. pp. 37-41

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

Pan, Y.-W., Chou, H.-H., You, E.-M., and Yu, H.-T. (2004) Isolation and characterisation of 23 polymorphic microsatellite markers for diversity and stock analysis in tiger shrimp (*Penaeus monodon*). *Molecular Ecology Notes* **4**: 345-347

Panova, M., Kinen, T.M., Fokin, M., André, C., and Johannesson, K. (2008) Microsatellite cross-species amplification in the genus *Littorina* and detection of null alleles in *Littorina saxatilis*. *Journal of Molluscan Studies* **74**: 111-117

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, F., Erazo, C., Zhinaula, M., Volckaert, F., and Calderón, J. (2004) A sex-specific linkage map of the white shrimp *Penaeus* (*Litopenaeus*) *vannamei* based on AFLP markers. *Aquaculture* **242**: 105-118

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

Perry, G.M.L., Ferguson, M.M., Sakamoto, T., and Danzmann, R.G. (2005) Sex-linked quantitative trait locus for thermotolerance and length in the rainbow trout. *The Journal of Heredity* **96**: 97-107

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* **6**: 847-859

References

Prince, J.D. (1991) A new technique for tagging abalone. *Australian Journal of Marine and Freshwater Research* **42**:101–106

Quintin, C.D., McMillan, I., and Glebe, B.D. (2005) Development of an Atlantic salmon (*Salmo salar*) genetic improvement program: Genetic parameters of harvest body weight and carcass quality estimates with animal models. *Aquaculture* **247**: 211-217

Raemaekers, S., and Britz, P. (2009) Profile of the illegal abalone fishery (*Haliotis midae*) in the Eastern Cape Province, South Africa: Organised pillage and management failure. *Fisheries Research* **97**: 183-195

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

Reece, K.S., Ribeiro, W.L., Gaffney, P.M., Carnegie, R.B., and Allen, S.K. Jr. (2004) Microsatellite marker development and analysis in the Eastern Oyster (*Crassostrea virginica*): Confirmation of null alleles and non-mendelian segregation ratios. *Journal of Heredity* **95**: 346-352

Reid, D.P., Szanto, A., Glebe, B., Danzmann, R.G., and Ferguson, M.M. (2005) QTL for body weight and condition factor in the Atlantic Salmon (*Salmo salar*): Comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* **94**: 166-172

Report of the Working Group on the Application of Genetics in Fisheries and Mariculture (WGAGFM), 19-23 March 2007, Ispra, Italy. pp 70

Rhode, C. (2010) Development of gene-linked molecular markers in South African abalone (*Haliotis midae*) using an *in silico* mining approach. [Unpublished Master of Science thesis] Stellenbosch University, South Africa.

Rodrigues, M.F., Lapatra, S., Williams, S., Famula, T., and May, B. (2004) Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* **241**: 93-115

References

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Research* **8**: 103-106

Ruivo, N. (2007) Microsatellite genotyping of contributing broodstock and selected offspring of *Haliotis midae* submitted to a growth performance recording scheme. [Unpublished Master of Science thesis] Stellenbosch University, South Africa

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

Sales, J., and Britz, P.J. (2001) Research on abalone (*Haliotis midae* L.) cultivation in South Africa. *Aquaculture Research* **32**: 863-874

Sanchez, C.C., Smith, T.P.L., Weidmann, R.T., Vallejo, R.L., Salem, M., Yao, J., and Rexroad III, C.E. (2009) Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* **10**: 559-566

Sanger, F. (1988) Sequences, sequences and sequences. *Annual Review of Biochemistry* **57**: 1-28

Santana, Q.C., Coetzee, M.P.A., Steenkamp, E.T., Mlonyeni, O.X., Hammond, G.N.A., Wingfield, M.J., and Wingfield, B.D. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* **46**: 217-223

Sato, Y., Hashiguchi, Y., and Nishid, M. (2009) Temporal pattern of loss/persistence of duplicate gene involved in signal transduction and metabolic pathways after teleost-specific genome duplication. *BMC Evolutionary Biology* **9**: 127-140

Schug, M.D., Mackay, T.F.G., and Aquadro, C.F. (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genetics* **15**: 99-102

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

References

Sekino, M., Kobayashi, T., and Hara, M. (2006) Segregation and linkage analysis of 75 novel microsatellite DNA markers in pair crosses of Japanese abalone (*Haliotis discus hannai*) using the 50-tailed primer method. *Marine Biotechnology* **8**: 453-466

Shendure, J., and Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**: 1135-1145

Shors, T.J., and Matzel, L.D. (1997) Long-term potential: What's learning got to do with it? *Behavioral and Brain Sciences* **20**: 597-614

Slabbert, R. (2010) Identification of Growth Related Quantitative Trait Loci within the Abalone *Haliotis midae*, Using Comparative Microsatellite Bulked Segregant Analysis. [Unpublished Doctor of Philosophy thesis] Stellenbosch University, South Africa.

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

Slabbert, R., Bester, A.E., and D'Amato, M.E. (2009) Analyses of genetic diversity and parentage within South African Hatchery of the abalone *Haliotis midae* Linnaeus using microsatellite markers. *Journal of Shellfish Research* **28**: 369-375

Slabbert, R., Hepple, J., Venter, A., Nel, S., Swart, L., Van den Berg, N.C., and Roodt-Wilding, R. (2010) Isolation and inheritance of 44 microsatellite loci in the South African abalone *Haliotis midae* L. *Animal Genetics* **41**: 332-333

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

Slate, J. (2005) Quantitative trait locus mapping in natural population: progress, caveats and future directions. *Molecular Ecology* **14**: 363-379

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

References

Stallings, R.L. (1995) Conservation and Evolution of (CT)/(GA), Microsatellite sequences at orthologous positions in diverse mammalian genomes. *Genomics* **25**: 107-113

Stam, P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* **3**: 739-744

Strachan, T., and Read, A.P. (1999) Instability of the human genome: mutation and DNA repair. In: *Human Molecular Genetics 2*. BIOS Scientific publishers. pp 209-235

Subasinghe, R., Soto, D., and Jia, J. (2009) Global aquaculture development and its role in sustainable development. *Reviews in Aquaculture* **1**: 2-9

Sun, X., and Liang, L. (2004) A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* **238**: 165-172

Tanksley, S.D. (1993) Mapping polygenes. *Annual Review of Genetics* **27**: 205-229

Tao, W.J., and Boulding, E.G. (2003) Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.) *Heredity* **91**: 60-70

Tarr, R.J.Q. (1989) Abalone. In: *Oceans of life off southern Africa*. Payne, A.I.L., and R. Crawford, J. M. (eds). Vlaeberg Publishers, Cape Town. pp. 62-69

Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acid Research* **17**: 6463-6471

Tautz, D., Trick, M., and Dover, G.A. (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**: 652-656

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

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T., and Nuzhdin, S.V. (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260-263

References

Van den Berg, N-C., and Roodt-Wilding, R. (2010) Parentage assignment in *Haliotis midae* L.: a precursor to future genetic enhancement programmes for South African abalone. *Aquaculture Research* **41**: 1387-1395

Van der Merwe, M. (2010) Growth-related gene expression in Haliotis *midae.* [Unpublished Doctor of Philosophy] Stellenbosch University, South Africa

Van der Merwe, M., and Roodt-Wilding, R. (2008) Chromosome number of the South African aballone *Haliotis midae*. *African Journal of Marine Science* **30**: 195-198

Van Ooijen, J.W. (2006) JoinMap® v4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma, B.V. Wageningen, The Netherlands

Van Ooijen, J.W., and Voorrips, R.E. (2001) *JoinMap 3.0 software for the calculation of genetic linkage maps*. Plant Research International, Wageningen, The Netherlands

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., and Sonstegard, T.S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**: 247-252

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Gela, D., Vallod, D., Chevassus, B., and Linhart, O. (2004) Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio L*.). *Aquaculture* **235**: 223-236

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

Voorrips, R.E. (2002) MapChart: software for the graphical presentation of linkage maps and QTL. *The Journal of Heredity* **93**: 77-78

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

Walderbieser, G.C. (1995) PCR-based identification of AT-rich tri and tetranucleotide repeat loci in an enriched plasmid library. *Biotechniques* **19**: 742-744

Wall, P.K., Leebens-Mack, J., Chanderbali, A.S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L.P., Hu, Y., Carlson, J.E., Ma, H., Schuster, S.C., Soltis, D.E., Soltis, P.S., Altman, N., and dePamphilis, C.W. (2009) Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**: 347-367

Weber, J.L., and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*. **44**: 388-396

Weir, B.S., and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370

Welsh, J., and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research* **18**: 7213-7218

Wenne, R., Boudry P., Hemmer-Hansen, J., Lubienieki, K.P., Was, A., and Kause, A. (2007) What role for genomics in fisheries management and aquaculture? *Aquatic Living Resources* **20**: 241-255

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535

Woram, R.A., McGowan, C., Stout, J.A., Gharbi, K., Ferguson, M.M., Hoyheim, B., Davidson, E.A., Davidson, W.S., Rexroad, C., and Danzmann, R.G. (2004) A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* **47**: 304-315

Wright, J.M. (1993) DNA fingerprinting of fishes. In: *Biochemistry and molecular biology of fishes*. Hochachka, P.W., and Mommsen, T. (eds.). Elsevier Press, New York. pp 57-91

References

Wu, R., Ma, C.-X., and Casella, G. (2007) *Statistical Genetics of Quantitative Traits*. Springer, U.S.A. pp 25-71

Xia, J.H., Liu, F., Zhu, Z.Y., Fu, J., Feng, J., Li, J., and Yue, G.H. (2010) A consensus linkage map of the grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs. *BMC Genomics* **11**: 135-151

Yaohua, S., Ximing, G., Zhifeng, G., Aimin, W., and Yan, W. (2010) Preliminary genetic linkage map of the abalone *Haliotis diversicolor* Reeve. *Chinese Journal of Oceanology and Limnology* **28**: 549-557

You, F.M., Huo, N., Gu, Y.Q., Luo, M.C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J. and Anderson, O.D. (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**: 253

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

Zha, X.F., Xia, Q.Y., Zhao, P., Li, J., Duan, J., Wang, Z.L., Qian, J.F., and Xiang, Z.H. (2005) Detection and analysis of alternative splicing in the silkworm by aligning expressed sequence tags with the genomic sequence. *Insect Molecular Biology* **14**: 113-119

Zhang, L., Yang, C., Zang, Y., Li, L., Zhang, X., Zhang, Q., and Xiang, J. (2007) A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. *Genetica* **131**: 37-49

Zerbino, D.R., and Birney, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*: **18:** 821-829

Zhu, Y., Queller, D.C., and Strassmann, J.E. (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. *Journal of Molecular Evolution* **50**: 324-338

Ziegler, A., and König, I.R. (2006) A statistical approach to genetic epidemiology. Sensen, C.W., Elston, R.C., Olson, J.M., and Palmer, L. (eds.) WILEY-VCH, Germany. pp 17-20

Appendices

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

BRIEF NOTES

doi:10.1111/j.1365-2052.2009.02001.x

Identification of positively selected sites in the goat *kappa casein* (*CSN3*) gene

A. Clop*, A. Zidi† and M. Amills†

*Genesis Faraday Partnership, Roslin BioCentre, Roslin, Midlothian, EH25 9PS, UK. †Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

Accepted for publication 13 October 2009

Background: Analysis of the rates of non-synonymous vs. synonymous substitutions ($w = d_N/d_S$) in the *kappa casein* (*CSN3*) genes of Bovidae has provided evidence that positive selection has accelerated amino acid divergence of this protein.[1] The goat *CSN3* gene is extraordinarily polymorphic, with sixteen alleles identified to date.[2] A significant part of this genetic variation is non-synonymous, suggesting that positive selection might have been a driving force shaping the allelic repertoire of this gene. Moreover, there is evidence that these *CSN3* variants have differential effects on dairy traits.[3] Herewith, we have used codon-based substitution models to identify and characterize with high resolution the footprint of positive selection in the goat *CSN3* gene.

Analysis of positive selection: Sequence data employed in our analysis are listed in Table S1. Positively selected codons were identified by using the CODEML program, which is integrated in the PAML 3.15 package.[4] A detailed description of the models and test of significance is reported in Appendix S1. The global w value of the goat *CSN3* gene was significantly greater than one ($P < 0.01$), a feature indicative of positive selection. In contrast, a distance-based analysis with the MEGA 3.1 software did not yield significant results (data not shown). Analysis with PAML 3.15 identified three codons (p.Gln65Arg, p.Val86Ile and Asp111Gly, reference sequence: CAA43174) displaying suggestive evidence of positive selection (posterior probability > 0.90, Table S2). Large standard deviations in mean estimates of w might be likely attributable to the fact that the sequence dataset was rather small. Moreover, chance variation in the mutation pattern across sites can produce false positives in this kind of test,[5] so further studies will be needed to confirm our results. Amongst the three identified codons, p.Gln65Arg appeared to be particularly interesting because both PMUT and Polyphen indicated that this substitution might alter protein function (Table S2). The biological meaning of positive selection at the *CSN3* loci of Bovidae has not been fully elucidated.[1] Interestingly, CSN3 bioactive peptides play an important role in gastrointestinal mucosal immunity and they have antithrombotic and antihypertensive properties.[6,7] In consequence, positive selection at *CSN3* might affect biological fitness in the context of any of these physiological processes.

References

1 Ward T. J. *et al.* (1997) *Genetics* **147**, 1863–72.
2 Prinzenberg E. M. *et al.* (2005) *J Dairy Sci* **88**, 1490–8.
3 Caravaca F. *et al.* (2009) *J Dairy Sci* **92**, 2960–4.
4 Yang Z. (1997) *Comput Appl Biosci* **13**, 555–6.
5 Hughes A. L. & Friedman R. (2008) *Immunogenetics* **60**, 495–506.
6 Baldi A. *et al.* (2005) *J Dairy Res* **72**, 66–72.
7 Silva S. V. & Malcata F. X. (2005) *Int Dairy J* **15**, 1–15.

Correspondence: M. Amills (Marcel.Amills@uab.cat)

Supporting Information

Additional supporting information may be found in the online version of this article.

Appendix S1 Supplementary materials and methods.

Table S1 Sequences employed in the CODEML analysis of goat *CSN3* variability.[1]

Table S2 Identification of positively selected codons in the goat *CSN3* gene and in silico prediction of amino acid substitution effects.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

---

doi:10.1111/j.1365-2052.2009.02003.x

Isolation and segregation of 44 microsatellite loci in the South African abalone *Haliotis midae* L.

R. Slabbert, J. Hepple, A. Venter, S. Nel, L. Swart, N. C. van den Berg and R. Roodt-Wilding

Molecular Aquatic Research Group, Department of Genetics, Stellenbosch University, Stellenbosch, 7600, South Africa

Accepted for publication 14 October 2009

Source/description: Microsatellite loci were isolated from *Haliotis midae* genomic DNA using either the FIASCO method[1] or the SNX Unilinker method.[2] The FIASCO method was modified to accommodate additional restriction enzymes, namely MseI, EcoRI and MspI (New England Biolabs). New adaptor and primer sequences were designed for EcoRI and MspI (Table S1). AluI and RsaI were used for the SNX method. The DNA libraries were enriched using an AC-rich probe and cloned using the QIAGEN PCR Cloning kit. Sequencing was performed on an ABI 3730 × l DNA analyser. Sequences were edited using SEQUENCE SCANNER version 1 (Applied Biosystems). Primers were designed for repeat containing sequences using BATCHPRIMER version 1[3] and labelled using fluorescent dyes.

PCR conditions: PCR reactions were carried out in a final volume of 10 μl containing 20 ng DNA, 0.2 μM of each primer, 200 μM deoxyribonucleotide triphosphates (dNTPs), 0.1 unit of GoTaq polymerase (Promega), 1× GoTaq Clear Flexi Buffer (Promega) and 2 mM $MgCl_2$. PCR programs are shown with Table S2. Genotyping was performed using an ABI 3730 × l DNA analyser and GENEMAPPER version 4 (Applied Biosystems).

Isolation results: A total of 978 recombinant clones were sequenced, and 49% of these contained repeat motifs. A total of 222 primer pairs were designed of which 44 were found to be polymorphic (Table S2).

Polymorphism and segregation analysis: DNA from 32 individuals from each of two full-sib families (Family7B and Family42A) was extracted. The Mendelian segregation patterns (1:1:1:1, 1:2:1, 1:1) of the 44 markers were examined in these families using the chi-square test. Among a total of 88 marker-family combinations (44 × 2), 38 were informative, 18 combinations were monomorphic, eight combinations had three or more alleles and 24 combinations could not be reliably scored or amplified (Table S2). Of the 38 informative marker-family combinations, 10 did not conform to expected Mendelian segregation patterns ($P < 0.05$). Of these 10 combinations, three (*Hmid0006M* for Family42A and *Hmid2044T* and *HmLCS147* for Family7B) could be explained by the presence of null alleles. *Hmid0006M* and *HmLCS147T* conformed to Mendelian segregation after we corrected for the null alleles (Table S2). The distortion of the other combinations (*Hmid0053D*, *Hmid0310D*, *Hmid4018D*, *HmLCS71T* and *HmNSp31M* for Family42A and *Hmid0310D* and *Hmid0065M* for Family7B) could possibly be explained by PCR errors or scoring difficulties.[4] These markers should be used with caution.

Acknowledgements: We thank Belinda Swart and Peizheng Wang for technical assistance, Aletta van der Merwe for critical evaluation of the manuscript, Stellenbosch University for facilities, Roman Bay Sea Farm and HIK Abalone Farm for samples and The Innovation Fund for funding.

References

1 Zane L. *et al.* (2002) *Mol Ecol* **11**, 1–16.
2 Hamilton M. B. *et al.* (1999) *BioTechniques* **27**, 500–7.
3 You F. M. *et al.* (2008) *BMC Bioinf* **9**, 253.
4 Jones A. G. & Ardren W. R. (2003) *Mol Ecol* **12**, 2511–23.

Correspondence: R. Slabbert (rslabbrt@sun.ac.za)

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Primer and adaptor sequences of EcoRI and MspI restriction enzymes which were used for the FIASCO method.

Table S2 Marker information, PCR conditions, segregation analyses and accession numbers of 44 novel microsatellite loci for *Haliotis midae*.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

doi:10.1111/j.1365-2052.2009.02005.x

Multifocal retinal dysplasia in the German Spitz (Klein and Mittel) is not caused by mutations in *BEST1*

J. Aguirre-Hernández and D. R. Sargan

Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK

Accepted for publication 20 October 2009

Source/description: Two mutations in the *bestrophin 1* gene (*BEST1*) have been identified as the cause of canine multifocal retinopathy. A nonsense mutation in exon 2 (c.182C>T, GenBank accession number EF110978.1; p.Arg25X) in the Great Pyrenees, English mastiff and Bullmastiff breeds, and a missense mutation in exon 5 (c.591G>A, p.Gly161Asp) in the Coton de Tulear were identified.[1]

We identified 11 cases of multifocal retinal dysplasia (MRD) in the German Spitz Klein breed (GSK) and two in the German Spitz Mittel (GSM). Using Pedhunter,[2] a common ancestor was found for all affected GSK (3–6 generations back) (see Table S1 and Fig. S1). To determine whether *BEST1* was implicated, exons and flanking intronic sequences were sequenced and microsatellite allele distributions were studied.

BEST1 PCR and sequencing: Published primer sequences were used[1] with MegaMix Blue (Microzone, West Sussex, UK) and touchdown PCR (initial annealing temperature 72 °C, decreasing 1 °C per cycle to 58 °C; for exon 4 and the combined exons 3 and 4, initial and final annealing temperatures were 72 and 60 °C respectively).

Exons 2 and 5, which harbour the published mutations and nearby intronic sequences, were sequenced in 12 GSK: seven cases and five controls (parents or full-sibs of the cases) and in two GSM parents of a case. No non-synonymous polymorphisms were found in these exons and no polymorphisms segregated with the disease state. The rest of the exons and intronic sequences were screened in two GSK cases; intragenic SNPs showed that they do not share the same *BEST1* allele (see Table S1 and Fig. S1). Both dogs showed heterozygosity in one or more SNPs.

Microsatellite analysis: Microsatellites were searched within the gene and along 10 kb on both flanks.[3] Low-complexity sequences precluded the study of an intragenic microsatellite (CFA18, 57508605 Mb; Ensembl CanFam2.0, Assembly May 2006, Genebuild June 2009).[4] A microsatellite 8.431 kb downstream of the gene (primers CTGCTTCAACAGAGCTT GGA and ACGGCTGTGGCTAAAGTTGA) had a small number of alleles and proved difficult to score. The microsatellite *AHT103* (409.7 kb downstream) was amplified using a D2-labelled forward primer (Proligo) and touchdown PCR (initial and final annealing temperatures of 66 °C and 54 °C); products were run on a CEQ8000 sequencer. Twenty two GSK were studied: 7 cases and 15 controls; allele distributions between the two groups were not significantly different (T2 chi-square test as implemented in clump; $P = 1.0$, 10 000 simulations).[5]

Appendix B

Table 2.5: Marker information, PCR conditions, segregation analyses and accession numbers of 44 novel FIASCO based microsatellite loci for *Haliotis midae*.

Locus Name	Repeat Sequence (5' – 3')	Primer Sequence (5' – 3')	PCR Program Tm	Family 42A			Family 7B			GenBank Accession Number
				Cross (male x female)	Expected Segregation	*p*-value[1]	Cross (male x female)	Expected Segregation	*p*-value[1]	
*Hmid*0006M	$(ACAT)_n(AC)_n(ACAT)_n$	F: ATGAATTAGTGAACGCCTT R: TTTGTTAGATATTCGCCAT	P2 Touch-down	312/312[2] x 273/281	1:1[2]	***[2]	Non-reliable Genotypes			GQ927108
				312/U[3] x 273/281	1:1:1:1[3]	0.771[3]				
*Hmid*0007C	$(AC)_n...(ACGC)_n$	F: AAATTATTCGGCTAAATGC R: AGGCCACACACCCATTTCC	P2 Touch-down	Non-reliable Genotypes			359/359 x 349/359	1:1	0.080	GQ927109
*Hmid*0053D	$(GT)_n$	F: AAACTGACCATCGCACTT R: AGACACATGCTCACGCAC	P2 Touch-down	096/096 x 102/108	1:1	***	102/106 x 094/096	1:1:1:1	0.372	GQ927110
*Hmid*0065C	$(CT)_n...(AC)_n...(AC)_n...(AC)_n$	F: TATGAGCACCAATCAACGC R: TTTCCTGTCTCTTCCGCAC	P2 Touch-down	246/253 x 242/253	1:1:1:1	0.054	234/251 x 209/234	1:1:1:1	***	GQ927111
*Hmid*0136T	$(ACTC)_n$	F: GATATCGTTATAAGCGGTG R: ATATCGGGTAGCATTGTCA	P2 Touch-down	085/089 x 085/085	1:1	0.160	Monomorphic			GQ927112

Appendix B

*Hmid*0154C	$(GAT)_n...(GAT)_n...(GAT)_n$	F: GTAACAGTGTTTAGCACCA	P2	243/247	1:1:1:1	0.068	214/229	1:1:1:1	0.289	GQ927113
		R: CATGTTTTGTATCATAAGG	Touch-down	x			x			
				227/243			214/243			
*Hmid*0166R	$(GAT)_n$	F: GCATGTGTTTATGACTGATGGC	P2	Duplicated Locus			Duplicated Locus			GQ927114
		R: ACACTCAAGCGACTTTGGC	Touch-down							
*Hmid*0221T	$(ACAG)_n$	F: GGTAGATCAACTGTCGCAGG	P2	156/168	1:1:1:1	0.321	164/172	1:1:1:1	0.741	GQ927115
		R: ATATTATCAAATCTGTTGCC	Touch-down	x			x			
				156/172			156/168			
*Hmid*0310D	$(GT)_n$	F: GATATGTTGAAATGGGGTT	P2	111/114	1:1:1:1	***	110/118	1:1:1:1	***	GQ927119
		R: AACAAACACAGACACTCAG	Touch-down	x			x			
				118/124			105/116			
*Hmid*0315M	$(GCGT)_n(GT)_n$	F: CGGTACTATGGTGTTCGCTT	P2	113/128	1:1	0.263	113/117	1:1:1:1	0.431	GQ927120
		R: AAAAACTGTTAGGGAGGCAA	Touch-down	x			x			
				122/122			113/128			
*Hmid*0321C	$(GT)_n...(GT)_n$	F: AATATGTGGAGCACGACGC	P2	147/155	1:1	1.00	Monomorphic			GQ927121
		R: AGGGCTACCCCACTGGTC	Touch-down	x						
				155/155						
*Hmid*0553D	$(GT)_n$	F: ACTAAGTTTCATAACGAACGCAC	P2	Non-reliable Genotypes			160/160	1:1:1:1	0.576	GQ927122
		R: TGTTCAGGACACACCCCAC	Touch-down				x			
							160/162			
*Hmid*0558D	$(GT)_n$	F: TGGAGCGTGTGAGTGAGAG	P2	111/117	1:1:1:1	0.626	113/117	1:1:1:1	0.654	GQ927123
		R: GATTTGGGCGCATGGACAG	Touch-down	x			x			
				111/113			107/111			

Appendix B

*Hmid*0561D	$(CG)_n$	F: ACAACATGCACACAGACACG	P2	Monomorphic			Monomorphic			GQ927116
		R: ACATGTGCACGCCTTTTG	Touch-down							
*Hmid*0563M	$(AG)_n(ACAT)(AG)_n(ACAGA$	F: AGCGTGTGTGTGTTGTGTG	P2	Monomorphic			221/225	1:1:1:1	0.080	GQ927117
	$G)_n(ACAG)_n(ACTG)(ACAG)_n$	R: TGTCCGTGGTTCGCATAAAG	Touch-down				x			
							202/221			
*Hmid*0610D	$(GT)_n$	F: AGCAGTAAAGTCCTAGGGTGTG	P2	121/121	1:1	0.077	119/123	1:1	0.479	GQ927118
		R: ACTTTTACATCTAGACACAGGGC	Touch-down	x			x			
				106/125			119/119			
*Hmid*2009AD	$(TG)_n$	F: TGTAGTAGACGGTGCAAGGAT	P2	Monomorphic			Monomorphic			GQ927144
		R: GGTGCCAACTGTTCACAATA	Touch-down							
*Hmid*2015M	$(TG)_n(TGTC)_n$	F: GCCTGTGTCGGTCTATCTGT	P2	Non-reliable Genotypes			Non-reliable Genotypes			GQ927124
		R: ATCGGAGACTCAACATTTGC	Touch-down							
*Hmid*2032C	$(GA)_n...(GA)_n...(GA)_n...(GA)_n$.	F: CAGTGTGTTTGTGTGTTGCTC	P2	Non-reliable Genotypes			Non-reliable Genotypes			GQ927125
	$..(AG)_n...(GT)_n...(GT)_n...(GA)_n$	R: CACCTGTTGTTGCTGCTCT	Touch-down							
	$...(GA)_n...(GA)_n...(GAGG)_n$									
*Hmid*2044T	$(GAGT)_n$	F: AATTTAGACGAGTGGCTTGTG	P2	Monomorphic			158/171	1:1	***	GQ927126
		R: CACATGGGTACATTGTGTGAG	Touch-down				x			
							171/171			
*Hmid*2047BD	$(GT)_n$	F: GTACGTTCAACACACCAGTCA	P2	Monomorphic			Monomorphic			GQ927127
		R: ACCTTCTACGCATTGAACATC	Touch-down							
*Hmid*4009C	$(TG)_n...(TG)_n...(TG)_n...$	F: GTGTGCGTCGAATACTTTCA	P1	Monomorphic			Monomorphic			GQ927128
	$(TG)_n...(TG)_n...(TG)_n$	R: GTATTCGAAGCACACCAACA	53℃							
*Hmid*4010D	$(AC)_n$	F: TCCTGATCATGAAAGCAAAAC	P2	Non reliable genotypes			189/203	1:1:1:1	0.741	GQ927131
		R: CTGAACATGGTGAGCAACTG	Touch-down				x			
							199/201			

Appendix B

*Hmid*4018D	$(GT)_n$	F: ACATATCCACGCTCGTAAGC	P2	212/225	1:1:1:1	***	Monomorphic		GQ927129
		R: CTACACCTGCACCTACACCA	Touch-down	x					
				200/225					
*Hmid*4019D	$(GT)_n$	F: TCAGCACAATGCTACGTCAT	P2	Monomorphic			Monomorphic		GQ927132
		R: CCCACCACACAGACATACAC	Touch-down						
*Hmid*4022C	$(TG)_n$...$(TG)_n$...$(TG)_n$...	F: TGTGAATTATCATGGGCATCT	P2	Monomorphic			Monomorphic		GQ927130
	$(TG)_n$...$(TG)_n$...$(TG)_n$...	R: CCCTCCTGTAACTTGTCCTG	Touch-down						
	$(GT)_n$...$(TGA)_n$								
*Hm*NSS1H	$(GGGTTA)_n$	F: TAA TTC CAG CAG CTG AAA AA	P3	Non-reliable Genotypes			Duplicated Locus		EF033331
		R: AAC AAC GAC CCT AAA CCA TC	56℃						
*Hm*NS18M	(ACCA)$(AGG)_n(ACC)_n$AG	F: AATTGTCTCCTTTGTTCTTCTTT	P3	Duplicated Locus			200/208	1:1	0.367 EF367120
	$(ACC)_n$AG(AAC)	R: TATTTGTGACTTTAGGTGAGGAC	60℃				x		
							179/179		
*Hm*LCS152M	$(CAA)_n(CCA)_n(CTA)_n(CCA)_n$	F: TGATAAATGTCACTTGGGATCA	P2	144/149	1:1	0.114	Duplicated Locus		GQ927139
	$(CTA)_n(CCA)(CTA)_n$	R: ACACGTGCAAACTGGTCTTC	Touch-down	x					
				149/149					
*Hm*LCS34M	$(TCAC)_n(TCAA)(TCAC)$	F: TAAAGTTGGCGTAAAATCTCA	P1	Non-reliable Genotypes			Non-reliable Genotypes		DQ993230
	$(TCAA)_n$	R: AACCAAGATGTAGCCAGGAT	58℃						
*Hm*R16T	$(CAGT)_n$	F: TGGGTACAATATGTGAAGCC	P2	Non-reliable Genotypes			Non-reliable Genotypes		GQ927138
		R: GCAAATTCAAAATGCTCAAA	Touch-down						
*Hm*LCS71T	$(GGAT)_n$	F: GTCACTGTATGTCTCGCACAC	P2	186/231	1:1	***	Duplicated Locus		DQ993221
		R: ATTTTGACGTTACTGCACGTA	Touch-down	x					
				186/186					
*Hm*LCS388M	$(GCGT)_n(GTT)_n(GT)_n$	F: GACAACCGGGATTCAAAC	P2	Non-reliable Genotypes			144/149	1:1	0.317 GQ927140
		R: ACTTCTTCAGGCAATATTACTAAG	Touch-down				x		

Appendix B

							144/144			
*Hm*NSp31M	$(CAA)_n(CAG)_n(CAA)_n$	F: CTCGGGTTCAGTTACCTACA	P3	288/307	1:1:1:1	***	Monomorphic			EU126856
		R: CAAGTCAGGGTGGTCGTCTTTCC	60℃	x						
				288/291						
*Hm*NSp42C	$(AGT)_n...(AGT)_n...(AGT)_n...$	F: CCTGAATTTATAGTAGTA	P3	286/311	1:1:1:1	0.411	293/305	1:1:1:1	0.431	EU126858
	$(TGT)_n$	R: CTCACTCAAACACACCCAAA	60℃	x			x			
				286/323			281/305			
*Hm*LCS58M	$(TG)_n(CG)(TG)_n(CGTG)$	F: GACAATTGAGAACATGTTTTTG	P1	Duplicated Locus			Non-reliable Genotypes			DQ993225
	$(TGCG)_n(TG)_n$	R: AACACTGATATGGTCAGTCCA	58℃							
*Hm*LCS383T	$(GTGA)_n$	F: TAAACTGCAAAATACCCACC	P2	Non-reliable Genotypes			402/405	1:1:1:1	0.289	GQ927141
		R: TCTTTACATCAAAGCTCCGT	Touch-down				x			
							409/412			
*Hm*S104T	$(GAGT)_n$	F: GTCGTCAGAAGCAGTTTGAA	P1	Non-reliable Genotypes			Non-reliable Genotypes			GQ927137
		R: AAATCAATTTCTAGTGCCCC	51℃							
*Hm*2H6FT	$(CACT)_n$	F: GTACTCATGCTCAACTGCGT	P2	Non-reliable Genotypes			Non-reliable Genotypes			GQ927136
		R: TGTCAAGTTACCGATCAAGG	Touch-down							
*Hm*3A11FM	$(CA)_n(CG)(CA)_n$	F: TGCTCGTCCTATAGCAATGT	P2	Non-reliable Genotypes			Non-reliable Genotypes			GQ927135
		R: ATAGAGCAACGTGCATTCAC	Touch-down							
*Hm*LCS147T	$(GAGT)_n$	F: CTTCTGTCATCCACAAGAGC	P2	Non-reliable Genotypes			187/187[2]	1:1[2]	***[2]	GQ927134
		R: GCTGGAATATTGTTGAAAGC	Touch-down				x			
							126/110			
							187/U[3]	1:1:1:1[3]	0.154[3]	
							x			
							126/110			

Appendix B

*Hm*NS32M	$(CAA)_n(GTC)(GTT)_n(GTC)$	F: TTCCGGATAAAGTAAATCGTC	P3	Non-reliable Genotypes			Duplicated Locus			GQ927142
	$(GTT)_n$	R: ACACTATAGAATACAGCGGCC	52℃							
*Hm*LCS175M	$(GT)_n(GA)_n$	F: TTCACACTCTACAGGGTTGG	P2	278/278	1:1	0.479	247/282	1:1	0.576	GQ927133
		R: AACACATTCAAAACCCACC	Touch-down	x			x			
				276/280			280/280			
*Hm*NS21M	$(CT)_n(C)_n(CT)_n(T)_n(CT)_n$	F: ACAAACTCACGCACACATTC	P2	235/240	1:1:1:1	0.355	221/229	1:1:1:1	0.089	GQ927143
		R: AGGTTGAAGAATGAGTGCGT	Touch-down	x			x			
				215/227			217/221			

[1] Departure from Mendelian segregation, *** $p<0.05$
[2] Genotype, segregation ratio and *p*-value without correction for null allele.
[3] Genotype, segregation ratio and *p*-value with correction for null allele.

Appendix C

Title:

A comparison of three tagging methods for abalones under culture conditions.

Authors:

Ruhan Slabbert [a]

Arnoldus C. Vlok [a]

Juli-ann Hepple [a]

Mathilde Van der Merwe [a]

Danie Brink [a]

[a] Aquaculture Division, Department of Genetics, Stellenbosch University, Private Bag X1, Matieland, 7602, SOUTH AFRICA

Corresponding author:

R. Slabbert

E-mail: rslabbrt@sun.ac.za

Tel nr.: +27 (0)21 8085837

Fax nr.: +27 (0)21 8085833

Postal address:

J.C.Smuts Building

Department of Genetics

Stellenbosch University

Private Bag X1

Matieland

7602

SOUTH AFRICA

Appendix C

Abstract - Three methods used during a genetic improvement programme to tag individual abalones (*Haliotis midae*) were evaluated to assess tag retention and practicality within a commercial environment. The methods entailed the use of 1) plastic tubes (aerial method), 2) numerical bee-tags and 3) radio frequency identification (RFID) tags. RFID and bee-tags showed the greatest tag retention under culture conditions (98 % and 75 % respectively) compared to plastic tube (aerial) tags (55 %). The aerial method is inexpensive and is recommended for small, short-term experiments. Both RFID and bee-tags are suitable for long-term, large scale experiments, but the cost of RFID tags and equipment will be a limiting factor in its application.

Keywords: *Haliotis*, abalone, tagging, bee-tag, RFID, plastic tubing

1 Introduction

The tagging of aquaculture species within a breeding programme or a restoration programme is of utmost importance as specific individual identification is needed for pedigree reconstruction or data collection. The method of tagging would depend on the biology of the species being used, i.e. invasive or non-invasive. Tagging methods used include passive integrated transponder (PIT) tags (Prentice et al. 1990; Gibbons and Andrews 2004) for sea urchins (Hagen 1996); internal numerical or fluorescent elastomer tags for crayfish (Jerry et al. 2001), lobster (Uglem et al. 1996) and fish (Bruyndoncx et al. 2002); coded wire tags for eels (Thomassen et al. 2000), aluminium tags and enamel paint for venerid clams (Stewart and Creese 2000) and coloured bee tags for scallop (Barbeau et al. 1996) and Zebra mussel (Thorpe et al. 2002).

A few tagging methods have been used for abalones to date. Wild adult individuals were tagged using wire or fishing line to attach a tag to the trema (Newman 1966). Prince (1991) used nylon rivets to attach numbered discs to the trema of an abalone, tagging animals as small as 50 mm. Commercial broodstock were tagged using quick-set putty resin with a numbered Perspex piece imbedded (personal communication: Louise Jansen, HIK Abalone Farm, Hermanus, RSA). Breen et al. (2003) glued plastic tags to the shell of *Haliotis iris* using super glue. Gallardo et al. (2003) fed *Haliotis asinina* artificial diets to give the shell, which is naturally a brownish colour, a blue growth band and reported it as a successful method for this species. This method may unfortunately depend on the biology of the abalone species being tagged. Kube et al. (2007) used coloured beads to identify families by gluing them to the shells of 10 month (shell length = 12 mm) old juveniles. Total tag retention of 45 % was recorded after an eleven month period, while a total tag loss after 28

months was 89 %. Tag loss for older abalones (2 years old to adult) was less as reported by Prince (1991) at 4 – 35 % per annum for wild samples and Kube et al. (2007) at 3 % at age 27 months. This study evaluates and compares three tagging methods used during the genetic improvement programme for the South African abalone, *Haliotis midae*. Tag retention, practicality and difficulties are discussed. The methods include plastic tubing (aerial method) placed in a trema, bee-tags glued to the shell and radio frequency identification tags (RFID tags; same as PIT tags) attached to the shell using quickset putty.

2 Materials and Methods

2.1 Data Collection

Counting of tags was performed during routine sampling (10 month intervals) activities on the abalone farms and form part of a larger genetics research programme. All tagging methods were evaluated under culturing conditions and tagged abalones underwent all activities associated with abalone farming, such as grading. Mortalities were calculated by counting all empty shells within a production basket.

2.2 Aerial method (Plastic tubing; Figure 1a)

One thousand juvenile abalones (average shell length = 16.2 ± 2.1 mm) were tagged using different coloured plastic tubing (available from Gasket & Shim Industries, Cape Town, South Africa) to facilitate visual identification of family groups. No numerical identification was added to the tubing. The tubing was cut into 1 cm lengths and used for all sized abalones. The tag could only be applied when the animals were anaesthetised since contractions of the foot muscle would make tagging more difficult (personal observation). The abalones were therefore anaesthetised using magnesium sulphate ($MgSO_4$; White et al. 1996) prior to tagging. The foot of the abalone is then lifted away to reveal the trema (Figure 2) at the posterior (Figure 2) of the shell. The tubing was then pushed through the most posterior trema of adequate size and pulled through as far as possible using tweezers, leaving a short enough length on the inside of the shell so as not to agitate the animal or cause the tubing to fall out. Each animal was returned to the production basket immediately after tagging, but since all 1000 were removed from the basket at the same time some animals may be out of the water until tagging was completed (2 hours with 5 researchers tagging).

2.3 Bee-tag method (Figure 1b)

Appendix C

Nine hundred and thirty juvenile abalones (average shell length = 36.7 ± 3.5 mm) were tagged using colour coded bee-tags with numerical identifications (available from Swienty, Denmark). The different colour and numerical combinations will be used for visual identification of individual abalones or family groups. Abalones were anaesthetised ($MgSO_4$) prior to tagging to facilitate the handling of this large sample size. Shell surfaces were first cleaned and dried using normal paper towels (Twinsaver, South Africa). The tag was attached to the posterior region, near the shell whorl (Figure 1b and 2). A drop of super glue (Blits Stik from Bostik, South Africa), approximately the size of the tag, was applied to the dry clean surface area and the bee-tag was immediately placed using a thin pointed instrument (such as a dissection needle). Each individual abalone was tagged with only one bee-tag so that each had a unique colour and number combination, however a total of 52 animals had to be tagged with 2 or 3 tags due to the limited unique numbers for each coloured bee-tag (1 to 99). The abalones were continuously returned (10 minutes after tag application to allow drying of glue) to the production baskets during the tagging process, but since all 1000 were removed from the basket at the same time some animals may be out of the water until tagging was completed (2 hours with 1 researcher tagging).

2.4 RFID method (PIT tags; Figure 1c)

One hundred and twenty three juvenile abalones (350.9 ± 33.5 mm) were tagged using 12 mm long radio frequency identification tags (RFID; Destron Fearing Corporation, available from Identipet, South Africa). Tag identification is performed using a Mini Portable Reader HS5900L F (Destron Fearing Corporation, available from Identipet, South Africa) which reads and stores a 16 character code used for identification of individual abalones. Abalones were anaesthetised ($MgSO_4$) prior to tagging to facilitate the handling of this large sample size. The shell surface was cleared of loose debris and feed remnants using a soft brush and dried using normal paper towel (Twinsaver, South Africa). Quickset putty resin (Quickset Putty from Pratley, South Africa) was prepared according to the instructions of the manufacturer. A piece of putty big enough to cover the RFID tag (600 mm^3) was attached to the posterior end of the shell, above the shell whorl (Figure 1c and 2). The putty was spread into the shell ridges (Figure 2) by pressing down gently to avoid shell breakage. The RFID tag was then embedded within the putty and covered as far as possible. The RFID tags are covered with a waterproof glass casing, making it ideal for the marine environment, because it is protected from corrosion. By covering the tag with the putty it will also protect it from breakage. The abalones were

continuously returned (10 minutes after tag application to allow setting) to the production baskets during the tagging process, but since all 200 were removed from the basket at the same time some animals may be out of the water until tagging was completed (2 hours with 1 researcher tagging).

3 Results and Discussion

The effect on mortalities on tag loss data was negligible and was comparable to normal production data (results not shown). The tagging methods are summarised in Table 1 and reports tag position, methodologies, visibility, suitability, risks and tag retentions.

3.1 Aerial method

The application of the plastic tubing was time consuming and technically difficult. Animals with an average size of 16.2 ± 2.1 mm could be successfully tagged using this method. Smaller abalones were either too difficult to handle or the shells were too brittle to allow tag application using the current protocol, making it a technically difficult method as well. A total of 1000 animals could be tagged within 2 hours by 5 researchers, making it a labour intensive exercise requiring a large working area. This method also requires a practised and delicate hand since the shells of small abalones are brittle and could break during the tagging process and if a too long part of the tubing are left on the inside of the shell it may irritate the animals, causing possible stress and growth retardation. Tag retention was 55 % after a 10 month period. This did not compare well to other trema based methods which reported 70 – 80 % retention after a 9 month period (Prince 1991). There could be several reasons for tag loss: 1) tags can be lost when animals are handled during grading operations; 2) the movement of the foot muscle could push the tag all the way through the trema; 3) the increase in the diameter of the posterior trema prior to closing could cause the tubing to fall out. The different coloured tubing was highly visible and made the identification of family groups assigned to a colour easy. No identification of individuals could be made, since no numbers were written on individual tags. Numbers may be written on tags using a waterproof pen, but this is difficult due to the thickness of the tubing used in this study (1 mm diameter). Another possibility is machine engraving during manufacturing of tubing, but this may increase costs. The low tag retention and labour intensiveness of this method makes it more suitable to small, short-term and low-cost experiments.

3.2 Bee-tag method

The application of the bee-tags was fast and easy. A total of 1000 abalones were tagged by one researcher within a 2 hour period. The animals were anaesthetised during this study in order to

handle the large sample size. Anaesthetisation will also keep the animals from walking over each other during the glue drying process, thus stopping animals from getting into contact with the glue or the accidental sticking of tags to other animals or surfaces. The only immediate risk to the animals during tagging was direct contact to the super glue. The numbers on the bee-tags enabled individual identification, while the different colours and colour combinations enabled group identification. Tag retention after a 10 month period was 75 %. When tag retention was measured in the readability of the tag number, retention dropped to 61 %. The retention level was higher compared to other methods using super glue for attaching tags to abalone shells which reported tag retention of 45 % after 11 months (Kube et al. 2007). The tag loss and number loss may be caused by several things: 1) animals rubbing against one another may dislodge a tag; 2) the corrosive effect of sea water or water movement may corrode the glue or the number of the tag; 3) the type and quality of the super glue used could limit the ability of a tag to bind to the shell (personal communication: Steven Ashlin, Roman Bay Sea Farms, Gansbay, RSA); 4) small *Spirobus* spp. on the shells may obscure the tags and their numbers, but these were successfully removed using a sharp object. One solution to the loss of tag numbers was discovered by accident, because of the limited numerical combinations of bee-tags 2 or 3 tags were applied to some individuals. Almost 93 % of two or three tag individuals were successfully recovered and identified. Another solution to the number retention challenge is to use more durable tags or treat the tags with waterproof, non-toxic substances to protect them from corrosion. An important detail to consider when using bee-tags is to decide on the orientation of a tag on an animal to ensure that one can discriminate between the numbers 6 and 9, and 66 and 99. The high tag retention, low cost (approximately € 0.03 per tag) and ease of tag application makes this an ideal method for long-term experiments with large sample sizes requiring data collection for the same individual animals.

3.3 RFID method

The application of the RFID tags was fast and easy. A total of 200 animals were tagged within a two hour period by one researcher. The animals were anaesthetised during this study in order to handle the large sample size, but it is not necessary since the movement of the animals will not influence the application of either putty or tag. The biggest risk to the abalones was the potential for shell damage during the application of the putty. Tag retention was 98 % after a 10 month period and was slightly better compared to other studies using epoxy resins as attachment medium (90 % after 7 months;

Stewart and Creese 2000). A high tag retention was expected, since broodstock animals on farms tagged using putty showed no tag loss at all (personal communication: Louise Jansen, HIK Abalone Farm, Hermanus, RSA). Possible routes of tag loss include: 1) the wrong type of putty was used (personal communication: Steven Ashlin, Roman Bay Sea Farms, Gansbay, RSA) causing the putty-tag complex to dislodge from the shell; 2) the putty was not spread well enough in between the shell ridges, causing ineffective binding between putty and shell surface; 3) the RFID tag was not properly covered with putty, causing it to dislodge from the putty or get damaged during the handling of the animals. The unique electronic identification numbers of each tag allows for the identification of individual animals. The identification of groups is more labour intensive since each tagged animal must be identified first by ID number before it can be sorted. A possible solution to this will be to use different coloured putty for each group, thus facilitating rapid sorting. Another restriction of RFID tags were their size (12 mm). Animals should be large enough to allow for an adequate shell surface size so as not to inhibit growth. The researchers recommend that abalones of at least 60 mm in shell length should be used for this tag type since an adequate amount of putty should be used to attach and cover the tag (600 mm^3) and smaller animals will not have a sufficient shell area. When using RFID tags are applied to animals where weight are to be measured it is recommended that animals are weighed before and after tag application so the tag weight can be deducted from weight data collected in future studies. The high tag retention and ease of tag application makes this an ideal method for long-term experiments which require large sample sizes and data collection for the same individual animals. This method is expensive (€ 5.30 per tag) and tag readers are also costly, but if data collection for all (or most) individuals within a sample is critical RFID tags would be invaluable. For example: this tag type was used within our current project for extremely valuable animals that need to be monitored over a period of a few years.

Acknowledgements. We are also indebted to Bruce Finnemore from Identipet, South Africa for donating RFID tags and for lending a reader for use in this study. Thanks also go to all the hatchery managers and students involved in the logistics and tagging during this study. This study was funded by The Innovation Fund, Abagold, Aquafarm, HIK Abalone, I&J Abalone Farm and Roman Bay Sea Farms.

References

Appendix C

Barbeau M.A., Hatcher B.G., Scheibling R.E., Hennigar A.W., Taylor L.H., Risk A.C., 1996; Dynamics of juvenile sea scallop (*Placopecten magellanicus*) and their predators in bottom seeding trials in Lunenburg Bay, Nova Scotia. Can. J. Fish. Aquat. Sci. 53, 2494–2512.

Breen P.A., Kim S.W., Andrew N.L., 2003, A length-based Bayesian stock assessment model for the New Zealand abalone *Haliotis iris*. Mar. Freshwater Res. 54, 619–634.

Bruyndonx L., Knaepkens G., Meeus W., Bervoets L., Eens M., 2002, The evaluation of passive integrated transponder (PIT) tags and visible implant elastomer (VIE) marks as new marking techniques for the bullhead. J. Fish Biol. 60, 260–262.

Gallardo W.G., Bautista-Teruel M.N., Fermin A.C., Marte C.L., 2003, Shell marking by artificial feeding of the tropical abalone *Haliotis asinina* Linne juveniles for sea ranching and stock enhancement. Aquac. Res. 34, 839–842.

Gibbons J.W., Andrews K.M., 2004, PIT tagging: simple technology at its best. Bioscience 54, 447–454.

Hagen N.T., 1996, Tagging sea urchins: a new technique for individual identification. Aquaculture 139, 271–284.

Jerry D.R., Stewart T., Purvis I.W., Piper L.R., 2001, Evaluation of visual implant elastomer and alphanumeric internal tags as a method to identify juveniles of the freshwater crayfish, *Cherax destructor*. Aquaculture 193, 149–154.

Kube P.D., Appleyard S.A., Elliott N.G., 2007, Selective breeding Greenlip abalone (*Haliotis laevigata*): preliminary results and issues. J. Shellfish Res. 26, 821–824.

Newman G.G., 1966, Movements of the South African abalone *Haliotis midae*. South African Division of Sea Fisheries Investigational Report No 56.

Prentice E.F., Flagg T.A., McCutcheon C.S., Brastow D.F., 1990, PIT-tag monitoring systems for the hydroelectric dams and fish hatcheries. Am. Fish. Soc. Sym. 7, 317–322.

Prince J.D., 1991, A new technique for tagging abalone. Aust J Mar Freshwat Res. 42, 101–106.

Stewart M.J., Creese R.G., 2000, Evaluation of a new tagging technique for monitoring restoration success. J. Shellfish Res. 19, 487–491.

Thomassen S., Pedersen M.I., Holdensgaard G., 2000, Tagging the European eel Anguilla Anguilla (L.) with coded wire tags. Aquaculture 185, 57–61.

Appendix C

232 Thorpe J.H., Alexander J.E., Cobbs G.A., 2002, Coping with warmer, large rivers: a field experiment
233 on potential range expansion of northern quagga mussel (*Dreissena bugensis*). Freshwater Biol. 47,
234 1779–1790.
235 Uglem I., Nœss H., Farestveit E., Jørstad K.E., 1996, Tagging of juvenile lobsters (*Homarus*
236 *gammarus* (L.)) with visible implant fluorescent elastomer tags. Aquacult. Eng. 15, 499–501.
237 White H.I., Hecht T., Potgieter B., 1996, The effect of four anaesthetics on Haliotis midae and their
238 suitability for application in commercial abalone culture. Aquaculture 140, 145–151.

239 **Figure captions**

240 **Figure 1: a)** Schematic figure of aerial (plastic tubing) method. Abalones are anaesthetised (1).
241 Tubing 1 cm in length (2) are cut and inserted into the posterior trema (3, 4 and 5) using tweezers. **b)**
242 Schematic figure of bee-tag method. Abalones are anaesthetised and the shell surface is cleaned
243 (1). Super glue is applied to the shell (2) and the bee-tag is attached using a needle (3 and 4). Tags
244 are distinguished by various colours and numbers (5). **c)** Schematic figure of RFID method.
245 Abalones are anaesthetised and the shell surface is cleaned (1). The quickset putty (resin) is
246 prepared (2) and attached to the shell (3). The RFID tag is then placed into the putty (4). Tag
247 numbers are read using a hand-held reading device (5).

248 **Figure 2:** A schematic representation of an abalone showing the posterior and anterior of the shell as
249 well as other morphological information.
250

Appendix C

Table 1: A summary of the three tagging methods used for abalones.

Method	Tag position	Attachment method	Anaesthetisation required	Risk to animals	Recommended size of animal at tagging	Group / family identification	Individual identification	Suitability	Tag retention
Aerial	Trema	Pliers / tweezers	Yes	Tissue and shell damage	15 mm +	Yes	No	Small, short-term, low-cost experiments. Inexpensive.	55 % [a]
Bee-tag	Shell	Super glue	Optional	Chemical exposure	10 mm +	Yes	Yes	Long-term experiments with small or large sample sizes. Inexpensive.	75 % [a] 61 % [b] 93 % [c]
RFID	Shell	Quickset putty	Optional	Shell damage	60 mm +	Yes	Yes	Long-term experiments with small or large sample sizes. Expensive.	98 % [a]

[a] Tag retention after 10 months.

[b] Tag retention after 10 months in terms of number durability (dependant on tag quality).

[c] Tag retention after 10 months when abalones are tagged with 2 or more bee-tags.

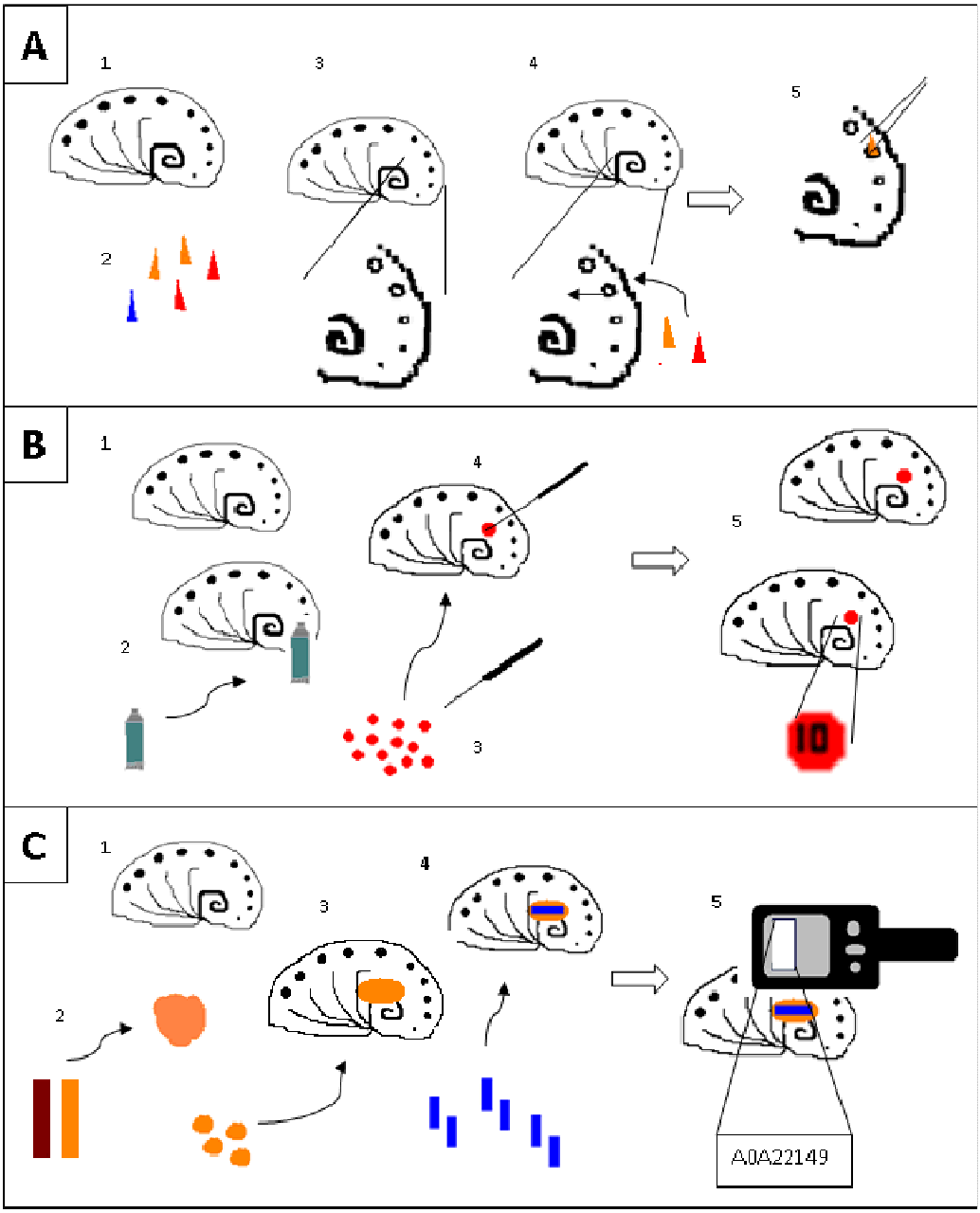


Figure 1

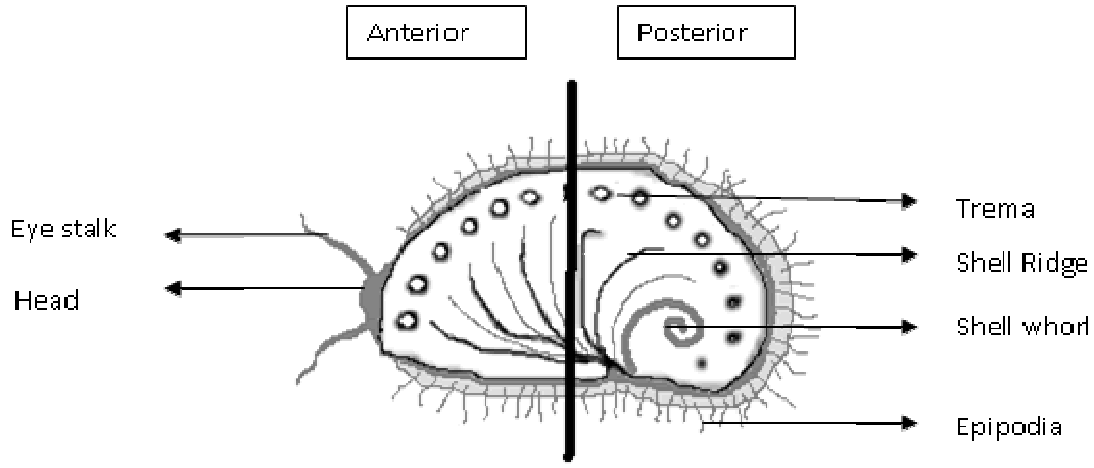


Figure 2