

Linkage mapping in *Haliotis midae* using gene-linked markers

by Suzaan Jansen

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

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

December 2011

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 2011

Copyright©2011
Stellenbosch University
All rights reserved

Summary

Haliotis midae, or more commonly known as Perlemoen, is an abalone species found along the coast of South Africa. It is the only cultured abalone species in South Africa and has a high demand abroad. Due to its popularity as a seafood delicacy, illegal harvesting has taken its toll on Perlemoen numbers. This increases the need for sustainable farming efforts and efficient implementation of law enforcement practices against poachers. Abalone farms make use of a limited number of broodstock for breeding, so it is necessary to ensure that genetic effects such as inbreeding and bottlenecks do not interfere with the viability of the offspring. Research that focuses on the genetics of Perlemoen will greatly aid the farms to continue sustainable production of this species as well as enhance their breeding efficiency. This study focuses on the construction of a linkage map for *H. midae* that will allow the future identification of markers associated with genes important to production, such as growth and disease resistance. Identification of these genes will allow breeders to select genetically superior abalone that will be used for breeding programmes in which the phenotype of the offspring will be enhanced.

For the construction of a linkage map it is necessary to have enough informative markers for mapping. In this study, gene-linked microsatellite markers were developed by screening a contig assembly of *H. midae*'s transcriptome. Ninety-eight primer pairs could be developed from the contigs and 60 loci produced amplification products. Twenty-six microsatellites were found to be polymorphic (27% success rate).

In addition to these markers, 239 previously developed microsatellites and 48 gene-linked SNPs were used to develop sex-average and sex-specific linkage maps in four full-sib families consisting of approximately 100 offspring each. Of these markers 99 were informative in family DS1 (31% success rate), 81 in family DS2 (26%), 77 in family DS5 (24%) and 71 in family DS6 (23%). These markers were used for linkage analysis (LOD>3). The average number of linkage groups for the sex-average maps ranged from 17-19. The average genome length for these maps ranged from 700cM to 1100cM with an average marker spacing of 8cM. The sex-specific maps' linkage groups ranged from 13-17 with an average genome length of 600cM to 1500cM. The average marker spacing was approximately 16cM. The integrated map was constructed by merging the sex-average

maps. This map contained 25 linkage groups with an average genome length calculation of 1700cM and an average marker spacing of 9.3cM.

The linkage maps created in this study are the first to utilize SNPs in *H. midae*. Further incorporation of SNPs into linkage maps will enhance the density. The maps created in this study are of medium-density (65%) and provide a link to the development of high-density linkage maps to facilitate associations of phenotypic traits to certain markers, to so that QTL mapping can be performed. This information can be used for marker-assisted selection to produce genetically superior abalone.

Opsomming

Haliotis midae, of meer algemeen bekend as Perlemoen, is 'n klipkous spesie wat langs die kus van Suid-Afrika voorkom. Dit is die enigste gekweekte klipkous spesie in Suid-Afrika en het 'n hoë aanvraag in die buiteland. As gevolg van sy gewildheid as 'n seekos lekkerny, het onwettige stropery sy tol geneem op Perlemoen getalle. Hierdie verhoog die behoefte vir volhoubare boerdery pogings en doeltreffende implementering van wetstoepassing teen stropers. Perlemoenplase maak gebruik van 'n beperkte aantal broediere vir teling, dus is dit nodig om te verseker dat genetiese effekte soos inteling en genetiese bottelnekke nie inmeng met die lewensvatbaarheid van die nageslag nie. Navorsing wat fokus op die genetica van Perlemoen sal grootliks die plase steun om die volhoubare produksie van hierdie spesie voort te sit, sowel as hul teling doeltreffendheid te verbeter. Hierdie studie fokus op die ontwikkeling van 'n genetiese koppelingskaart vir *H. midae*, wat die toekomstige identifisering van die merkers wat verband hou met die gene wat belangrik is vir die produksie, soos groei en weerstand teen siektes sal verbeter. Identifisering van hierdie gene sal toelaat dat telers genetiese voortreflike Perlemoen kan kies vir teelprogramme waartydens die fenotipe van die nageslag sal verbeter word.

Vir die ontwikkeling van 'n genetiese koppelingskaart is dit nodig om genoeg informatiewe merkers vir die kartering te hê. In hierdie studie, is geen-gekoppelde mikrosatelliet-merkers ontwikkel deur 'contig' data van *H. midae* se transkriptoom te ondersoek. Agt en negentig inleier pare kon ontwikkel word uit die 'contigs' en 60 loki kon 'n amplifiseringsproduk lewer. Ses-en-twintig mikrosatelliete was polimorfies (27% suksessyfer).

Bykomend tot hierdie ontwikkelde merkers is 239 voorheen ontwikkelde mikrosatelliete en 48 geen-gekoppelde SNPs gebruik om geslagsgemiddelde en geslagspesifieke koppelingskaarte in vier volsib families, wat uit ongeveer 100 nageslag elk bestaan, te ontwikkel. Van hierdie merkers was 99 informatief in familie DS1 (31%), 81 in die familie DS2 (26%), 77 in die familie DS5 (24%) en 71 in die familie DS6 (23%). Hierdie merkers is gebruik vir 'n koppelingsanalise (LOD>3). Die gemiddelde aantal koppelingsgroepe vir die geslagsgemiddelde kaarte het gewissel van 17-19. Die gemiddelde genoom lengte vir hierdie kaarte het gewissel van 700cM tot 1100cM met 'n gemiddelde merker spasiëring van 8cm. Die koppelingsgroepe van die geslagspesifieke kaarte het gewissel van 13-17

met 'n gemiddelde genoom lengte van 600cM tot 1500cM. Die gemiddelde merker spasiëring was ongeveer 16cm. Die geïntegreerde kaart is saamgestel deur die samesmelting van die geslagsgemiddelde kaarte. Die kaart toon 25 koppelingsgroepe met 'n gemiddelde berekende genoom lengte van 1700cM en 'n gemiddelde merker spasiëring van 9.3cM.

Die genetiese koppelingskaarte wat in hierdie studie ontwikkel is, is die eerste om SNPs te gebruik in *H. midae*. Verdere insluiting van SNPs in koppelingskaarte sal die digtheid verhoog. Die kaarte wat in hierdie studie ontwikkel is, is van medium digtheid (65%) en bied 'n stap nader aan die ontwikkeling van hoë digtheid koppelingskaarte om fenotipiese eienskappe met sekere merkers te assosieer, vir kwantitatiewe kenmerk lokus kartering. Hierdie inligting kan gebruik word vir merker bemiddelde seleksie om geneties verbeterde Perlemoen te produseer.

Acknowledgements

I would like to thank the following institutions for their contributions to the study: Innovation Fund, Roman Bay Sea Farm (Pty) Ltd, HIK Abalone Farm (Pty) Ltd., Central Analytical Facility and Stellenbosch University. I would also like to thank the following people for their academic guidance and encouragement: my supervisor and study leader Dr. Rouvay Roodt-Wilding, our lab manager Dr. Aletta van der Merwe, our wonderful in house linkage mapping experts Juli Hepple and Jessica Vervalle, Dr. Ruhan Slabbert, and my fellow MARG students Clint Rhode, Lise Sandenbergh, Sonja Blaauw, Liana Swart and Jana du Plessis. Lastly I would like to acknowledge the non-academic support from friends and family that kept me sane when I wanted to crack: Nicolene, you were always there to share all the hardships over a cup of coffee, Rudi for being my rock in the toughest of times, my sister Corlé for picking up my slack at the flat when I had to work or write, and lastly to mom Surika and dad Jaco for your unmovable faith in my abilities.

Table of contents

Chapter one-Literature review

1. Abalone in South Africa	2
2. <i>Haliotis midae</i> in general.....	3
2.1 Classification	3
2.2 Biology of <i>H. midae</i>	3
2.2.1 Reproduction	3
2.2.2 Early development and settlement	4
2.2.3 Feeding and growth	5
3. Aquaculture.....	6
3.1 Overview of global aquaculture	6
3.2 Abalone aquaculture	6
3.3 Abalone aquaculture in South Africa	7
3.4 Abalone aquaculture genetic management	7
4. Molecular markers and their uses in aquaculture	8
4.1 General	8
4.2 Type 1 versus type 2 molecular markers	9
4.3 Microsatellite markers	10
4.3.1 General overview	10
4.3.2 Microsatellites in aquaculture	15
4.4 Single nucleotide polymorphisms	15
4.4.1 General overview	15
4.4.2 SNPs in aquaculture	16
4.4.3 Genotyping of the SNPs with the VeraCode GoldenGate Genotyping Assay of Illumina	17
5. Transcriptome sequencing as a valuable resource for marker development	19
5.1 Overview of transcriptome sequencing using next generation sequencing (NGS) platforms	19
5.2 Marker development using NGS platforms	21
6. Linkage mapping	22
7. Quantitative trait loci	24
8. Marker-assisted selection	24
9. Aims and objectives	25

Chapter two-Type 1 microsatellite development

1. Abstract.....	28
------------------	----

2. Introduction	29
3. Materials and methods:.....	30
3.1 Genomic DNA extractions	30
3.2 Microsatellite identification from the <i>H. midae</i> transcriptome and primer design	31
3.3 Contig homology search	32
3.4 Microsatellite amplification and analysis of polymorphism	32
3.5 Genotyping	33
4. Results.....	34
5. Discussion	38

Chapter three-Linkage mapping

1. Abstract.....	43
2. Introduction	44
3. Materials and methods.....	47
3.1 Mapping families	47
3.2 Genotyping of the gene-linked markers	47
3.2.1 Microsatellite markers	47
3.2.2 SNP markers	48
3.2.3 Genotype data	48
3.3 Linkage analysis	49
3.4 Linkage map integration	50
3.5 Genome coverage	51
3.5.1 Observed map length	51
3.5.2 Expected genome length	51
3.5.3 Genome coverage	51
4. Results.....	52
4.1 Gene-linked SNPs	52
4.2 Genotyping of the mapping families	52
4.3 Linkage mapping	54
4.3.1 Linkage map of family DS1	54
4.3.2 Linkage map of family DS2	61
4.3.3 Linkage map of family DS5	67
4.3.4 Linkage map of family DS6	75
4.3.5 Sex-average linkage group comparisons	81
4.3.6 Integrated map	98
5. Discussion	101
5.1 Segregation distortion and duplicated loci	102

5.2 Linkage mapping	104
5.3 Mapped microsatellites versus mapped SNPs	110
5.4 Conclusion	112

Chapter four-Conclusions and future applications

1. Microsatellite development	115
2. Linkage mapping in <i>H. midae</i>	116
3. Future studies and improvements.....	118
References	121
Appendices	I

List of figures

Figure 1.1: A map indicating the distribution of the 5 endemic abalone species found in South Africa.	2
Figure 1.2: Gonad colouration. A = greenish female gonad B = cream coloured male gonad (Roux 2011).	4
Figure 1.3: An illustration showing the life-cycle of abalone (Hepple 2010).	5
Figure 1.4: A - representation of the different forms of microsatellite repeats, where A indicates a perfect microsatellite (TACC), B - a compound microsatellite (GGAT) ₂ (CAG) ₂ and C - an interrupted or complex microsatellite repeat, (GGAT) ₂ ACGT(CAG) ₂	12
Figure 1.5: DNA replication slippage (Ellegren 2004).	13
Figure 1.6: A workflow of the VeraCode GoldenGate assay.	19
Figure 1.7: The workflow of the Illumina Solexa Genome Analyser.	21
Figure 2.1: 2% Agarose gel. A: optimised loci, B: no PCR product.	35
Figure 2.2: 12% PAGE gels. A: Polymorphic locus B: Monomorphic locus.	35
Figure 3.1: Sex-average map of family DS1 representing the 18 linkage groups.	56
Figure 3.2: Maternal map of family DS1 showing the 18 linkage groups.	58
Figure 3.3: Paternal map of family DS1.	59
Figure 3.4: Sex-average map of family DS2.	62
Figure 3.5: Maternal map of family DS2.	63
Figure 3.6: Paternal map of family DS2.	65
Figure 3.7: Sex-average map of family DS5.	69
Figure 3.8: Maternal map of family DS5.	71
Figure 3.9: Paternal map of family DS5.	73
Figure 3.10: Sex-average map of family DS6.	76
Figure 3.11: Maternal map of family DS6.	77
Figure 3.12: paternal map of family DS6.	79
Figure 3.13: Homology for sex-average linkage maps of family DS1 and DS2.	85
Figure 3.14: Homology for sex-average linkage maps of family DS5 and DS6.	87
Figure 3.15: Homology for sex-average linkage maps of family DS1 and DS5.	90
Figure 3.16: Homology for sex-average linkage maps of family DS1 and DS6.	93
Figure 3.17: Homology for sex-average linkage maps of family DS2 and DS5.	96
Figure 3.18: Homology for sex-average linkage maps of family DS2 and DS6.	98
Figure 3.19: The integrated map for <i>H. midae</i> , constructed by merging the sex-average maps of families DS1, DS2, DS5 and DS6.	101

List of tables

Table 1.1: Molecular markers used in aquaculture and their corresponding applications and polymorphic power (Liu and Cordes 2004).	10
Table 1.2 Linkage maps consisting mainly of microsatellites for some marine species.	23
Table 2.1: A summary of the microsatellite identified in <i>H. midae</i> 's transcriptome.	34
Figure 2.1: 2% Agarose gel. A: optimised loci, B: no PCR product.....	35
Table 2.2: Twenty-six polymorphic EST-STR marker loci.....	35
Table 2.3: BLAST results of the polymorphic microsatellites indicating the sequence description, organism, E-value and accession number for each contig that showed a positive hit.....	37
Table 3.1: The Joinmap [®] v.4 genotype data format for CP populations (Van Ooijen 2006).	48
Table 3.2 Genotyping success of the SNPs.....	52
Table 3.4: A summary of the informative markers obtained from inspecting the genotyping data of each mapping family	53
Table 3.5: Number of null alleles, duplicated and distorted loci for all the markers genotyped	53
Table 3.6: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS1.	60
Table 3.7: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS2.	67
Table 3.8: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal map of family DS5.	75
Table 3.9: Number of markers per linkage group, their corresponding lengths, average marker spacing and largest interval for the sex-average, maternal and paternal map of family DS6.	81
Table 3.10: Total number of informative microsatellite and SNP markers as well as, number of mapped microsatellite and SNP markers in the sex-average, maternal and paternal maps for each family.....	82
Table 3.11: Number of markers per linkage group, their corresponding lengths, average marker spacing and largest interval of the integrated map.	99

Table 3.12: Summary of abalone linkage maps indicating number and type of markers used for map construction, number of segregating families and the number of linkage groups..... 107

List of Abbreviations

%	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
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
C	Cytosine
cDNA	complimentary DNA
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
cm	Centimetre
cM	CentiMorgan
CTAB	Cetyltrimethylammonium bromide
ddH ₂ O	Double distilled water
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediamine Tetra-Acetic Acid (C ₁₀ H ₁₆ N ₂ O 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
kb	kilobase pairs
LOD	Logarithm of odds
m	Metre
M	Molar (Moles per Litre)
m/v	Mass per volume

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
mm	Millimetre
mM	Millimolar
MML	Multipoint maximum likelihood
mRNA	messenger ribonucleic acid
MtDNA	Mitochondrial Deoxyribonucleic Acid
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology information
ng	Nanograms
ng/μl	Nanogram per microlitre
NGS	Next generation sequencing
°C	Degrees Celsius
<i>p</i>	Probability value (as a statistically significant limit)
PAGE	Poly-acrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
pmol	Picomole
pp.	Pages
PTP	Picotiter plate
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic DNA
RFID	Radio Frequency Identity
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
rpm	revolutions per minute
SDS	Sodium Dodecyl Sulfate
sec	Seconds
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
T	Thymine
<i>Taq</i>	<i>Thermus aquaticus</i> DNA Polymerase
TBE	Tris-Borate-EDTA Buffer
TEMED	N, N, N', N',-tetramethylenediamine
T _m	Melting Temperature
Tris-HCl	Tris-Hydrochloric acid

™	Trademark
U	Units (enzyme)
v/v	Volume per Volume
VNTR	Variable Number Tandem Repeat
w/v	Weight per Volume

Chapter one

Literature Review

1. Abalone in South Africa

Abalone are marine herbivorous gastropods found worldwide along coastal areas. In total, there are 56 species of Haliotidae (Geiger 2000). In Southern Africa there are six abalone species with 5 found in South Africa. Three of these species (*Haliotis midae*, *H. parva*, *H. spadicea*) occur on the West and East coast of South Africa, whereas the other two species' (*H. queketti* and *H. alfredensis*) distribution is restricted to the East coast (Fig. 1.1) (Evans *et al.* 2004). *Haliotis midae*, more commonly known as Perlemoen, is the largest of these abalone and in conjunction with their non-cryptic lifestyle make them a suitable species for aquaculture (Roodt-Wilding and Slabbert 2006).

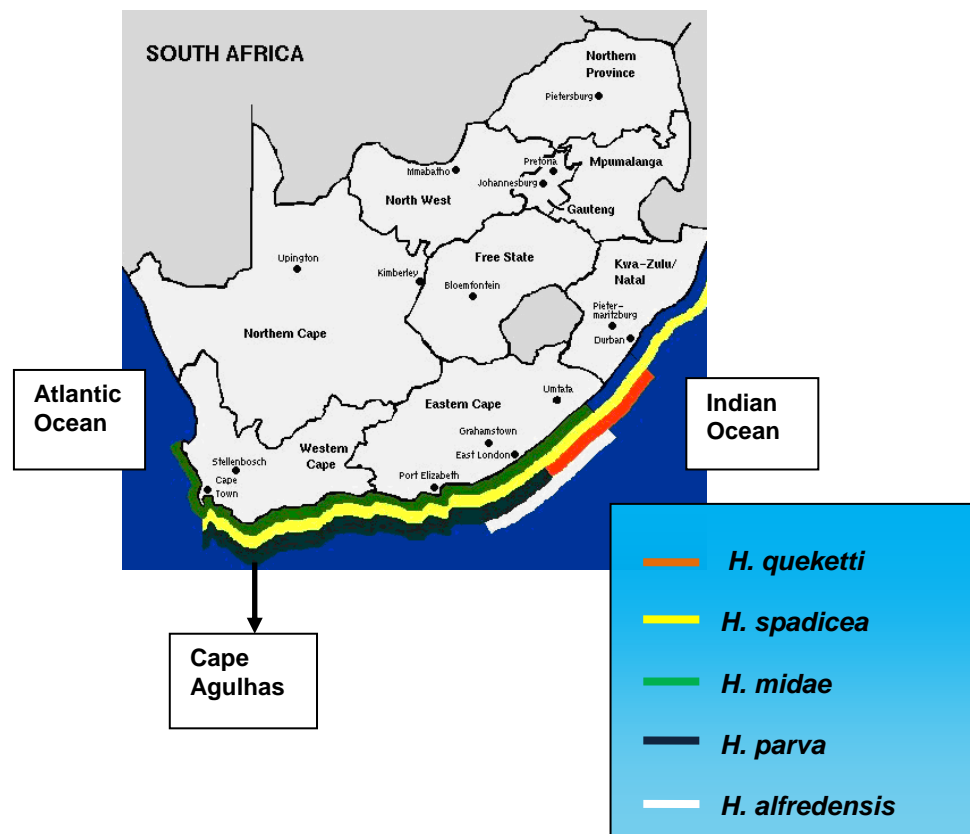


Figure 1.1: A map indicating the distribution of the five endemic abalone species found in South Africa (adapted from <http://web.uct.ac.za/depts/zoology/abnet/safrica.html>).

Abalone, however, is one of the most exploited marine resources in South Africa. This is mainly due to poaching and habitat loss as well as increased predation by the rock lobster (*Jasus lalandii*) (Mayfield and Branch 2000; Sales and Britz 2001; Steinberg 2005). Its rapid decrease in numbers led to the government's decision to ban all harvesting of wild abalone in South Africa for ten years in February 2008.

Perlemoen was subsequently put on the list of the 'Convention on International Trade in Endangered Species of Wild Flora and Fauna' Appendix III. This was in an effort to regenerate the wild population numbers and to reduce black market trade in Perlemoen (DEAT 2007). However, in May 2010 the CITES restrictions placed on wild harvesting of Perlemoen was lifted. Commercial fishing of Perlemoen was thus reinstated, but export permits are still required and the total allowable catch has been set at 150 tonnes (t) of abalone yearly as advised by scientists (Bürgener 2010).

The lifting of the ban was mainly due to the South African governments' inadequate implementation of the CITES permits at ports of exit. The wildlife trade monitoring network (TRAFFIC) has urged the South African government to re-evaluate their decision and to enlist Perlemoen on the CITES appendices once again, but before such a decision can be made serious issues with trade management have to be solved (Bürgener 2010).

2. *Haliotis midae* in general

2.1 Classification

Phylum: Mollusca

Super family: Haliotoidea

Class: Gastropoda

Family: Haliotidae

Super order: Vetigastropoda

Genus: *Haliotis*

(<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?lvl=0&id=36098>)

2.2 Biology of *H. midae*

2.2.1 Reproduction

Abalone are unisexual animals. The gender of an animal can easily be determined as illustrated in the figure below (Fig. 1.2). Female abalone have greenish gonads and males have a cream coloured gonad (Fig. 1.2).



Figure 1.2: Gonad colouration. A = greenish female gonad B = cream coloured male gonad (Roux 2011).

In males and females fecundity increases with size, for example a female with a shell size of 11.43cm will be able to produce 4.3 million eggs per spawning whereas a female with a shell size of 16cm could produce 16 million eggs. Sexual maturity is reached at approximately 7.2 years in the wild and about 3 years in a cultured environment and on the warmer East coast of South Africa (Wood and Buxton 1996). However, in a recent study by Roux (2011) it was found that *H. midae* males and females could reach sexually maturity as early as two years of age, implying that animals of this age could potentially be induced to spawn artificially.

Spawning depends on the water temperature, but usually occurs twice a year; in September to November and March to May for the South African abalone (Tarr 1989; 1995). Abalone have a growth spurt in winter, while in the summer and autumn months their growth slows down so that the gonads can recover from spawning (Tarr 1989). Once the water temperature is favourable, males start to spawn. This stimulates the females to release their eggs (Huchette *et al.* 2004). The sperm and eggs are simply released into the surrounding water (also known as broadcast spawning; Giese and Kanatani 1987), causing them to be potentially swept away by currents before fusing to form a zygote. This leads to a high number of sperm and eggs which is lost during each spawning event (Tarr 1989).

2.2.2 Early development and settlement

The fertilized eggs are dependent on currents to carry them to suitable environments to settle in. A fertilized egg is about 0.2mm in diameter. Approximately 20 hours after

fertilization the trochophore escapes from the egg after which it develops into a veliger larvae (Fig. 1.3). If a suitable substrate is found, the larvae settle. The encrusting coralline algae release a compound gamma-aminobutyric acid, GABA, which induces the larvae to settle. The juveniles shy away into crevices for protection against predation and storms until they are 5-6cm in diameter and only then do they occupy exposed rock (Tarr 1989).

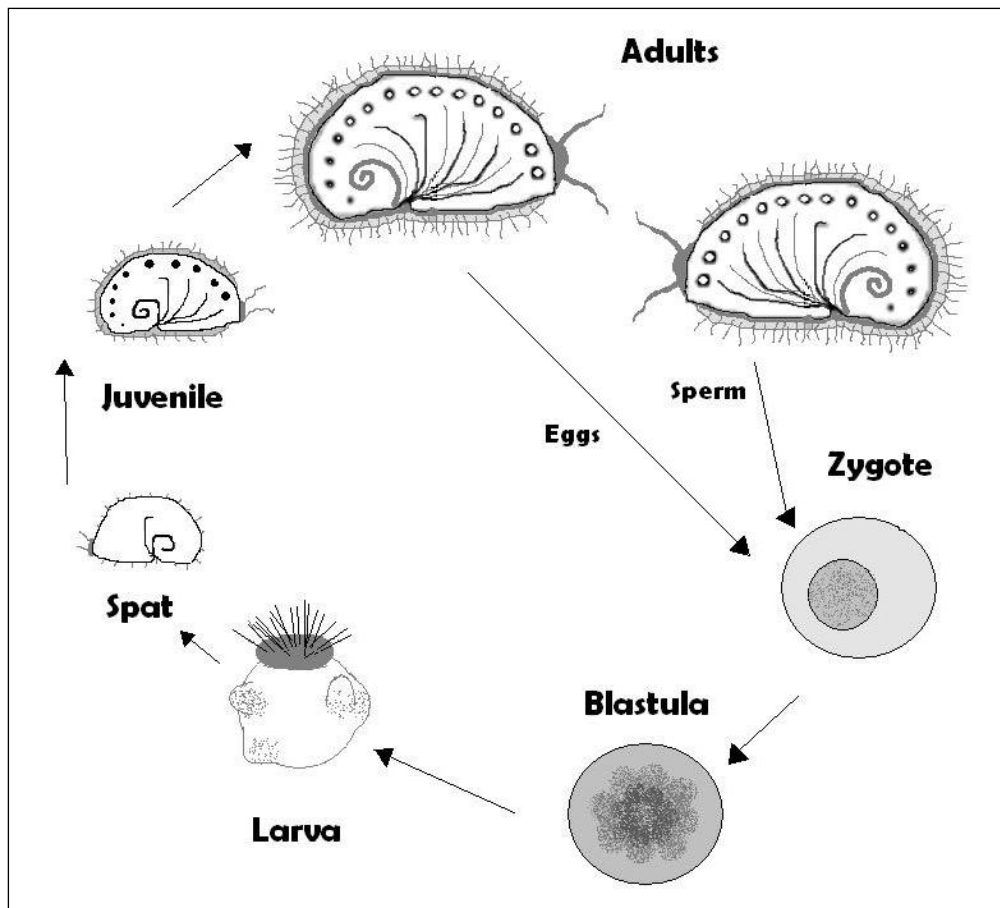


Figure 1.3: An illustration showing the life cycle of abalone (Hepple 2010).

2.2.3 Feeding and growth

Small abalone, with a shell length of about 5-10mm long, settle and graze on diatoms, that cover the alga *Lithothamnion*, found on the rock surfaces. At this stage the abalone's small shell is white in colour, but can have a dark red-brown colour as well as some green bands, depending on the different algal species that they feed on (Tarr 1989). When the animals reach maturity, usually 30mm long, their shells are mostly white. When the mature animals move to exposed rock surface their diets consist of drift kelp or overhanging kelp fronds (Tarr 1989). The change in diet from

micro-algae to seaweed is mainly due to differentiation of the abalones' mouth (Fallu 1991; Landau 1992).

3. Aquaculture

3.1 Overview of global aquaculture

Aquaculture has seen a rapid increase in production over the past four decades, accounting for 45% of the world's food fish and has increased from 1 million t in the 1950's to approximately 65 million t in 2008 (FAO 2010). The largest producers of farmed food fish in the world are the People's Republic of China with Sub-Saharan Africa still being one of the smallest producers of aquaculture species in the world, even though South Africa has the land space and water capabilities for aquaculture (Subasinghe *et al.* 2009).

3.2 Abalone aquaculture

The abalone aquaculture industry has increased considerably over the last decade from producing 3000 t in 2000 to over 40 000 t in 2008 (FAO 2009). This popular marine mollusc has been cultured in a variety of countries around the world including Japan, Thailand, South Korea, the USA, New Zealand, Australia, South Africa and Chile. China and Taiwan are currently the largest producer of farmed abalone in the world, producing approximately 33010 t of abalone on more than 300 farms (Troell *et al.* 2006; Allsopp *et al.* 2011). Outside of Asia, South Africa (together with Namibia) is the third largest producer of aquacultured abalone in the world (Allsopp *et al.* 2011).

Currently approximately 14 *Haliotis* species have commercial value. These include amongst others tropical abalone, *Haliotis asinina*; Pacific abalone, *H. discus hannai*; green abalone, *H. fulgens*; blackfoot abalone, *H. iris*; Australian abalone, *H. laevigata*; Perlemoen, *H. midae*; blacklip abalone, *H. rubra*; red abalone, *H. rufescens*; and European abalone, *H. tuberculata*. As poaching, habitat destruction and over-fishing has caused abalone species to reach dangerously low levels in the wild, abalone aquaculture has emerged as a means to supply the world demand for this sought after delicacy (Roodt-Wilding 2007).

3.3 Abalone aquaculture in South Africa

Commercial harvesting of *H. midae* started in 1949 and covered 580km of coastline from Cape Columbine to Quoin Point (Dichmont *et al.* 2000). The sustainability of this practice was not properly assessed and catches in the 1960s were much higher than what could be sustained. This led to the establishment of abalone aquaculture in the 1980s to relieve some pressure on wild stocks. Currently there are 18 registered abalone farms in South Africa ranging from Port Nolloth on the West coast to East London on the East coast. Cumulatively these farms generate about 900 t of abalone annually (934 t for 2008; Britz *et al.* 2009), making it a very valuable commodity for South Africa.

3.4 Abalone aquaculture genetic management

It is very important to genetically manage farmed abalone so that the commercial populations retain enough genetic variation to circumvent problems associated with bottlenecks because of the limited number of broodstock utilised on farms. This has previously been documented for the Pacific abalone as well as for blacklip abalone and Perlemoen (Evans *et al.* 2004; Li, Q *et al.* 2004). Genetic management is also vital for the ultimate genetic improvement of farmed abalone for traits of importance to production. Abalone farms experience extreme competition internationally and have to stay competitive to remain viable.

Genetic characterization, by making use of molecular markers such as allozymes, mitochondrial DNA, AFLPs (Amplified Fragment Length Polymorphisms), RAPDs (Random Amplified Polymorphic DNA), RFLPs (Restriction Fragment Length Polymorphisms), microsatellites (Short Tandem Repeats; STRs) and SNPs (Single Nucleotide Polymorphisms) represents one way of aiding the abalone farming industry. These markers can be used in a variety of applications in aquaculture. This includes parentage assignment (Jerry *et al.* 2004; Castro *et al.* 2007; Ruivo 2007; Slabbert *et al.* 2009; Van den Berg and Roodt-Wilding 2010), determining genetic variation between populations (Campbell *et al.* 2003; Evans *et al.* 2004; Hayes *et al.* 2006; Coibanu *et al.* 2009; Merchant *et al.* 2009), and 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). Identifying marker loci associated with economically important quantitative traits, including growth and

disease resistance (quantitative trait loci, QTL), can be used for selective breeding programs such as marker-assisted selection (MAS) (Roodt-Wilding and Slabbert 2006).

4. Molecular markers and their uses in aquaculture

4.1 General

Living organisms are all subject to mutations at DNA level. This occurs due to everyday cellular processes or even interactions between the organism and its environment. This in turn leads to different forms of the same marker loci seen in different individuals. These different forms, or alleles, cause a marker to be polymorphic. These polymorphisms together with genetic drift and selection, causes the genetic variation seen between individual organisms and species. Through the accumulation of point mutations, insertions and deletions, molecular markers are generated. When molecular markers are heritable and the polymorphism discernible, they are useful for research (Vignal *et al.* 2002; Liu and Cordes 2004).

Allozyme markers were the first molecular marker to find utility in aquaculture genetics in the early 1960's. They are different proteins produced by the same gene locus and thus represent polymorphisms of the genome as well as being a type 1 (associated with coding DNA) marker. These markers have been used in aquaculture for tracking inbreeding, stock identification, and parentage analysis. The disadvantages of this type of marker include null alleles (non-amplifying alleles), which cause heterozygote deficiency, and the high amounts of quality tissue samples required for analysis. Another disadvantage of this marker is the fact that polymorphism is investigated at protein level, which means that certain polymorphisms at DNA level can be masked by silent and synonymous peptide changes (Liu and Cordes 2004).

In the early 1980's the first DNA marker was identified, namely mitochondrial DNA (mtDNA). MtDNA, found in the mitochondria of cells, has been shown to accumulate more sequence divergence than nuclear DNA, probably due to a lack of DNA repair mechanisms. This, in combination with the maternal inheritance pattern of mtDNA, causes its fast mutation rate (Liu and Cordes 2004). In the past, allozyme and mtDNA were the markers of choice in aquaculture research. It is separate from the

nuclear genome and is easy to isolate (Okumus and Ciftci 2003). The high levels of polymorphism in mtDNA relative to allozyme markers made this non-nuclear marker the choice for population differentiation studies in aquaculture genetics (Liu and Cordes 2004). However with the invention of PCR, other types of markers including RAPDs, AFLPs, RFLPs, microsatellites and SNPs could be generated (Mullis and Faloona 1987). In aquaculture genetics microsatellites are the most widely used marker with SNPs fast approaching the same popularity status (Liu and Cordes 2004; Lo Presti *et al.* 2009).

With various marker types to choose from, care has to be taken when deciding which marker is most suited to the specific research aim. There are a few characteristics of molecular markers that have to be taken into account, including dominance, polymorphic information content (PIC), neutrality and independence of segregation before a choice of marker can be made (Vignal *et al.* 2002).

4.2 Type 1 versus type 2 molecular markers

Molecular markers can generally be divided into two categories, depending on where they are situated in the genome. Markers that are found to be located or associated with genic regions of the genome are termed type 1, or genic markers, and those markers that are found to be associated with anonymous regions of the genome are termed type 2 (O'Brien 1991). Microsatellite markers as well as SNP markers are generally type 2 markers, but if they are associated to genes of known functions, they are classified as type 1. This is also true if microsatellites and SNPs are developed from Expressed Sequence Tags (ESTs) as these represent transcribed segments of genes in a genome and are subsequently classified as type 1 markers (Liu and Cordes 2004).

The uses of type 1 markers are only now being fully appreciated. Their applications are wide spread and can assist aquaculture research in various applications (Table 1.1) (Liu and Cordes 2004). Microsatellites and SNPs that are identified in ESTs are for example preferable in the construction of genetic linkage maps. These functional maps have utility in comparative studies, candidate gene discovery as well as improved QTL identification (Vignal *et al.* 2002; Varshney *et al.* 2005). This makes ESTs a valuable resource for mining type 1 microsatellite and SNP markers (Serapion *et al.* 2004). However, it should be stated that gene-linked markers are

usually less polymorphic, which has implications for studies dependant on the polymorphic nature of markers such as microsatellites, including linkage and pedigree analysis (Fraser *et al.* 2005).

Table 1.1: Molecular markers used in aquaculture and their corresponding applications and polymorphic power (Liu and Cordes 2004).

Marker type	Prior information required?	Inheritance	Type	Polymorphic power	Predominant applications
Allozyme	Yes	Mendelian, Co-dominant	Type 1	Low	Linkage mapping Population studies
mtDNA	No	Maternal inheritance	-	-	Maternal lineage
RFLP	Yes	Mendelian, Co-dominant	Type 1 or Type 2	Low	Linkage mapping
RAPD	No	Mendelian, Dominant	Type 2	Intermediate	Fingerprinting for population studies Hybrid identification
AFLP	No	Mendelian, Dominant	Type 2	High	Linkage mapping Population studies
SNP	Yes	Mendelian, Co-dominant	Type 1 or Type 2	High	Linkage mapping Population studies
Indels	Yes	Mendelian, Co-dominant	Type 1 or Type 2	Low	Linkage mapping
STR	Yes	Mendelian, Co-dominant	Type 1 or Type 2	High	Linkage mapping Population studies Paternity analysis

4.3 Microsatellite markers

4.3.1 General overview

Microsatellite markers belong to a class of genomic sequences termed variable number tandem repeats (VNTRs) and are made up of simple sequence repeats that are about 1-6bp long and occur in tandem (Lit and Luty 1989; Tautz 1989).

Chapter one-Literature review

Minisatellites, which is the other type of repeat found in the class VNTRs, have longer repeat units of 10-100bp (Buschiazzi and Gemell 2006).

Microsatellites are evenly, but non-randomly, spaced throughout the genome and are located in genomic as well as coding DNA. They are abundant in all species and have been indicated to occur about every 1.87kb in fish (Chistiakov *et al.* 2006) with a mutation rate of 10^{-2} - 10^{-6} per locus per generation. Compared to the mutational rate of non-repetitive DNA, 10^{-9} , microsatellites mutate at a much higher rate, leading to the high polymorphic abundance of this marker (Weber and Wong 1993). They are small enough to be amplified by PCR, which is important for genotyping. The number of repeats of a given microsatellite can vary considerably, making it very polymorphic and thus useful in an array of different studies including linkage mapping (Weber and May 1989; Chistiakov *et al.* 2006). The size difference of the repeats (the alleles) that contribute to the polymorphic nature of microsatellite markers can be genotyped by techniques including polyacrylamide gel electrophoresis (PAGE) or analyses of fluorescent peaks obtained from the labelled PCR products on a genetic analyser, enabling visualisation of the size differences using software such as Genemapper.

These markers can occur in different forms; perfect, compound or interrupted (Fig. 1.4). Compound forms occur when repeat segments are found next to different repeat segments and interrupted microsatellites occur when mutations accumulate in the repeat segment (Goldstein and Schlötterer 1999).

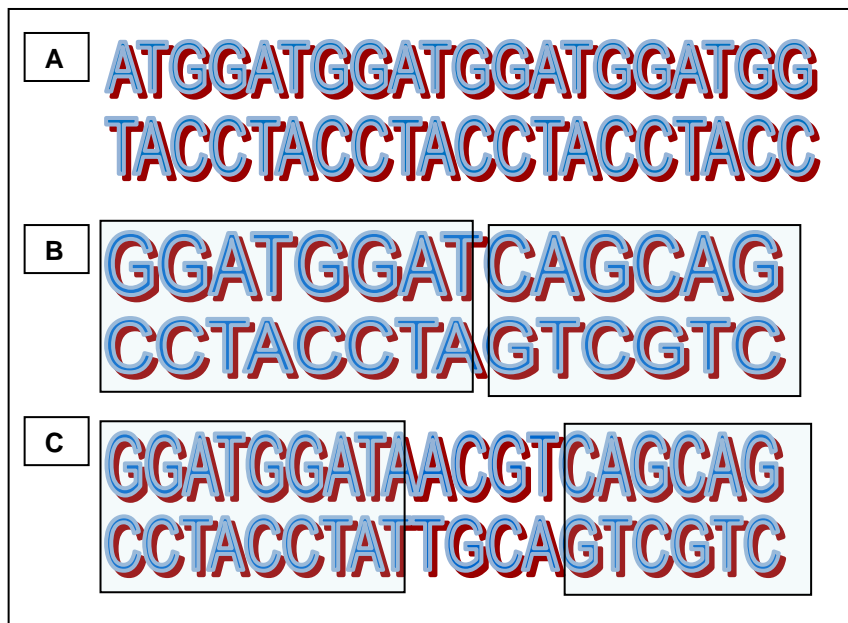


Figure 1.4: A - representation of the different forms of microsatellite repeats, where A indicates a perfect microsatellite (ATGG), B - a compound microsatellite (GGAT)₂(CAG)₂ and C - an interrupted or complex microsatellite repeat (GGAT)₂ACGT(CAG)₂ (Hepple 2010).

Microsatellites can also be classified in terms of the length of the repeat unit for example, a repeat unit constituting two nucleotides will be referred to as a dinucleotide and a repeat unit made up of 3 nucleotides will be a trinucleotide etc. In vertebrates, dinucleotides occur most frequently, whereas trinucleotides are much more prevalent in exonic regions (Li, Y-C *et al.* 2004).

The mechanisms, which propagate microsatellites have been described, but are still not fully understood. One such a model is the process of DNA replication slippage (Fig. 1.5) (Levinson and Gutman 1987; Tautz 1989). The slippage rate is correlated to the microsatellite length, indicating that longer microsatellites have a higher degree of polymorphism (Primmer and Ellegren 1998; Whittaker *et al.* 2003; Sainudiin *et al.* 2004; Leclerq *et al.* 2010). It has been postulated that there must be a threshold repeat value for propagation of microsatellites through DNA replication slippage, as short microsatellites, with only a few repeat units, do not expand through this process based on certain models (Meisser *et al.* 1996; Rose and Falush 1998). A proposed hypothesis for the generation of very short microsatellites (also called proto-microsatellites) states that they could arise from random point mutations (Jarne *et al.* 1998; Leclerq *et al.* 2010). Leclerq *et al.* (2010) however argued that no

minimum threshold is required for microsatellite propagation through DNA replication slippage and that it can occur at a minimum length of two repeats, which is the minimum requirement for DNA polymerase to slip during DNA replication.

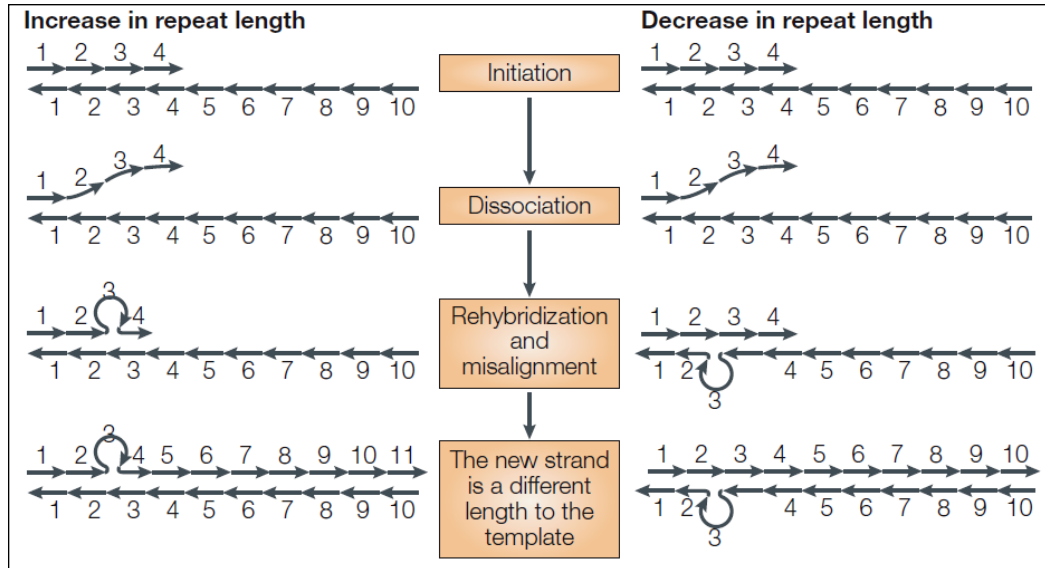


Figure 1.5: DNA replication slippage (Ellegren 2004).

Gene conversion, or non-reciprocal recombination, is another way by which microsatellites length can be altered (Sekar *et al.* 2009). This mechanism includes the unequal cross-over of chromosomal sections during meiotic (and mitotic) recombination (Hancock 1999; Li *et al.* 2002). Studies on human trinucleotide diseases, such as fragile X syndrome, and in *E. coli* indicated that gene conversion could lead to the instability of tandem repeats, especially trinucleotides (Dere *et al.* 2006). In members of the *Salmonidae* family, which are known to undergo tetraploidisation, gene conversion was found to be the mechanism involved in the differentiation and evolution of duplicated loci (Chistiakov *et al.* 2006).

Microsatellite markers have the highest PIC value compared to other markers due to the number of alleles that can be present at a specific locus and the mode of inheritance (co-dominant markers), which means that both allelic forms can be detected (Liu and Cordes 2004). Although this type of molecular marker has advantages over the older generation markers including RFLPs and AFLPs, it still has some drawbacks. To design primers for the amplification of microsatellite loci, the sequences flanking the microsatellite has to be known, which is not the case for markers such as RAPDs and AFLPs. If the sequence is not known, genomic libraries

have to be constructed and sequenced before primer design can take place, which is time consuming (Sekar *et al.* 2009). Problems associated with genotyping microsatellites further complicate matters. One such limitation of microsatellites is genotyping errors, resulting from the size-based nature of these markers. Genotyping of microsatellites are often complicated by stutter bands, which occurs due to the polymerase that slips during the PCR. These peaks can have the same intensity as the true peak, making allele scoring difficult and creating genotyping errors. This makes it very hard to compare data between laboratories as the genotyping data largely depend on the particular researcher's method of scoring and standardization of the alleles (Liu and Cordes 2004). Null alleles constitute another problematic phenomenon. This is a common occurrence in microsatellite markers and occurs when a specific flanking region of a microsatellite, has undergone a mutation so that the primer can no longer bind and produce a PCR fragment. This can become a major problem when genotyping microsatellites, as a null allele cannot be scored and implies that such individuals can only be included as missing data or homozygotes. This can become a problem for diversity studies, but it is possible in some instances to confirm the presence of a null-allele in a progeny, making it useful in mapping studies. However, when a null-allele cannot be traced in a progeny it has to be excluded. A specific drawback with microsatellite development is non-specific amplification. This means a researcher can spend quite a long time optimising a PCR. There thus needs to be a standardized method of scoring alleles when laboratories are collaborating for consistency and comparison (Hauser and Seeb 2008; Sekar *et al.* 2009).

The uses of type 1 markers have been stressed in a previous section, but type 1 microsatellites have numerous advantages when developed from ESTs or transcriptome sequences. In a study where cross-species amplification of microsatellites within the genus *Actinidia* was evaluated, only type 1 microsatellites were chosen. The authors stated that type 1 microsatellites had a greater transfer rate because they were anchored to ESTs or genes due to their sequence conservation (Fraser *et al.* 2005). Microsatellites that were developed from ESTs in *Meretrix meretrix* (hard clam) could be used to identify genes and was used in further population genetic analysis (Li, H *et al.* 2010). When transferred microsatellites are mapped to a species' linkage map, comparative studies between species can be

conducted. This can potentially elucidate certain genomic features, especially in instances where markers are transferred from a model organism to a non-model organism.

4.3.2 Microsatellites in aquaculture

Applications of microsatellites in aquaculture include genome mapping, parentage, kinship stock structure determination and genetic variability estimation (Merchant *et al.* 2009; for review see McAndrew and Napier 2011). Microsatellites have been isolated for a variety of marine species including amongst others, giant tiger prawn, *Penaeus monodon* (Xu *et al.* 1999); Atlantic salmon, *Salmo salar* (Vasimaggi *et al.* 2005); silver crucian carp, *Carassius auratus gibelio* (Yue *et al.* 2004); rock carp, *Procypris rabaudi* (Yue *et al.* 2009); and Mozambique tilapia, *Oreochromis mossambicus* (Sanju *et al.* 2010).

Over the years microsatellites have been identified in a variety of abalone species. These include Pacific abalone (Huang and Hanna 1998; An and Han 2006; Sekino *et al.* 2006; Zhan *et al.* 2008b; Li, Q *et al.* 2010), blacklip abalone (Evans *et al.* 2000; Baranski *et al.* 2006b), green abalone (Cruz *et al.* 2005), pink abalone, *Haliotis corrugata* (Díaz-Viloria *et al.* 2008); as well as Perlemoen (Bester *et al.* 2004; Slabbert *et al.* 2008; Hepple 2010). The trend in all of these studies is that ESTs are being used increasingly as resources for microsatellite mining, generating gene-linked microsatellites that can be mapped.

4.4 Single nucleotide polymorphisms

4.4.1 General overview

Single nucleotide polymorphisms or SNPs are polymorphisms that are caused by point mutations at a specific locus resulting in different alleles. In theory SNPs can have four alleles, but usually only have two and are thus bi-allelic markers. This leads to lower PIC values than for example microsatellite markers. This drawback is easily overcome as SNPs are abundant across the whole genome and, like microsatellites; they are inherited as co-dominant markers (Vignal *et al.* 2002; Liu and Cordes 2004). Gupta *et al.* (2001) reported a frequency of one SNP every 100-300bp in any given genome and in humans, it was found that one SNP occurs every 500-1000bp (Cooper *et al.* 1985; Li and Sadler 1991; Syvanen 2001). Studies on molluscs, such as the Pacific and eastern oyster (*Crassostrea gigas* and

Crassostrea virginica, respectively) have found that one SNP can occur as frequently as once every 40-60bp (Curole and Hedgecock 2005; Quilang *et al.* 2007). In the abalone, *H. discus hannai*, it has been reported that one SNP is present for every 100bp of DNA, while previous studies on SNP prevalence in *H. midae* indicated one SNP every 113-185bp (Bester *et al.* 2008; Rhode *et al.* 2008) and more recently every 150bp (Rhode 2010). This proposed frequency of SNPs in *H. midae*'s genome makes it possible to construct dense genetic linkage maps that are needed for QTL analysis in this aquaculture species.

The popularity of SNP markers in molecular studies is due to its abundance in all organisms, capacity of genotyping by high-throughput platforms and the nucleotide level at which this marker reveals polymorphisms, which other markers cannot (Liu and Cordes 2004). These markers also have a number of advantages, which makes their utility in molecular studies more profound. Firstly, SNPs are often responsible for the genetic variation between individuals that could possibly be a casual variant for a specific disease or trait. This makes the mapping of potential causative SNPs a priority for aquaculture species (Rafalski 2002; Butcher *et al.* 2007). Microsatellite genotyping using genetic analysers is still costly, but because SNP genotyping can be conducted using high-throughput techniques, genotyping costs can be lowered (Fan *et al.* 2003; Shen *et al.* 2005; Barbazuk *et al.* 2007).

The most accurate and most popular technique for SNP discovery is DNA sequencing, in particular EST-sequencing. ESTs have been used in species including half-smooth tongue sole, *Cynoglossus semilaevis* (Sha *et al.* 2010); and an important tree species, lodgepole pine, *Pinus contorta* (Parchman *et al.* 2010) as well as Perlemoen (Blaauw 2011). Due to the fact that SNPs are considered gene tagged markers when developed from ESTs, they can be used in comparative genome studies between different species (Moreno-Vazquez *et al.* 2003; Lindbald *et al.* 2005). These markers are also useful in population studies as they are more stably inherited than other markers with higher mutational rates (Hastbacka *et al.* 1992; Marshall *et al.* 1993).

4.4.2 SNPs in aquaculture

SNPs have a variety of uses in aquaculture. They can be used for traceability of aquaculture species (Hayes *et al.* 2006; Maretto *et al.* 2010), estimating genetic

variability between wild and cultured stocks (Rengmark *et al.* 2006; Ciobanu *et al.* 2010), linkage analysis (Kongchum *et al.* 2010; Du *et al.* 2010) and QTL identification and mapping (Liu and Cordes 2004; Malosetti *et al.* 2011; Palti *et al.* 2011). SNPs have been developed for numerous aquaculture species including Atlantic salmon (Renmark *et al.* 2006); Atlantic cod, *Gadus morhua* (Moen *et al.* 2008); Japanese flounder, *Paralichthys olivaceus* (He *et al.* 2008); turbot, *Scophthalmus maximus* (Vera *et al.* 2011), grass carp, *Ctenopharyngodon idella* (Xia *et al.* 2010), common carp *Cyprinus carpio* (Zheng *et al.* 2011) and Pacific oyster *Crassostrea gigas* (Guo *et al.* 2011).

In recent years, SNP resources have increased in several abalone species. A total of 137 SNPs has been identified in Pacific abalone (Qi *et al.* 2008; 2009; 2010; Zhang *et al.* 2010). These were all developed using either an EST or gene-targeted approach. In Perlemoen various methods have been investigated for the development of SNPs. These include construction of cDNA libraries to screen ESTs for SNPs (Bester *et al.* 2008), mining SNPs from ESTs of various *Haliotidae* for transfer to Perlemoen (Rhode 2010) and more recently using the sequenced transcriptome of *H. midae* for the development of gene-linked SNPs (Blaauw 2011). A limited number of SNP have also been developed for *Haliotis leavigata* (30), *H. rubra* (28), *H. rufescens* (24), *H. fulgens* (17) and *H. iris* (18) (Kang *et al.* 2010).

4.4.3 Genotyping of the SNPs with the VeraCode GoldenGate Genotyping Assay of Illumina

SNPs can be genotyped through a variety of techniques including traditional bi-directional sequencing, MALDI-TOF (Matrix Assisted Laser Desorption Ionization - Time of Flight), high-resolution melt analysis, pyrosequencing and SNP chips (Liu and Cordes 2004; Wenne *et al.* 2007). All of these have been shown to be successful in various SNP genotyping studies, but their high-throughput capabilities and associated costs differ. These differences and availability of technologies are to be considered when choosing the best genotyping platform for the associated study.

Next generation sequencing has created an avenue for large scale SNP development. New SNP genotyping platforms, such as SNP chips, have been developed to genotype these large number of SNPs in a fast and efficient manner in a large number of individuals (Syvanen *et al.* 2005). However, non-model organisms,

which have limited numbers of SNPs available, cannot make use of these high-throughput systems. In these instances a medium-throughput genotyping platform will be more appropriate. One such a platform is the VeraCode GoldenGate Genotyping Assay of Illumina that can multiplex 48, 96, 144, 192 and 384 SNP loci in a single reaction for up to 480 individuals per assay in a cost-effective way (Fan *et al.* 2003).

GoldenGate genotyping has been successfully used in a variety of species including wheat *Triticum* spp (Akhunov *et al.* 2009), soybean, *Glycine max* (Hyten *et al.* 2008); turkey, *Meleagris gallopavo* (Kerstens *et al.* 2009); cod, *Gadus morhua* (Hubert *et al.* 2010); catfish, *Ictalurus punctatus* (Wang *et al.* 2008) and white and black spruce *Picea glauca* and *P. mariana* (Pavy *et al.* 2008).

This technology incorporates a microbead-based array that uses an optical fibre bundle as substrate for the microarray. The fibre bundle in turn consists of 50 000 individual fibres that are etched to create a well that holds a specific microbead type that genotypes a specific SNP (Oliphant *et al.* 2002). Each type of microbead is covalently attached to an oligonucleotide sequence, which is specific for a particular SNP (Oliphant *et al.* 2002). Genomic DNA is attached to a solid support and mixed with oligonucleotide probes labelled with two different fluorescent dyes, Cy3 and Cy5 that are allele-specific (ASO). A third locus-specific probe (LSO) binds downstream of the SNP site and any unbound probe is washed away. Enzymatic extension of ASO to LSO and ligation is performed followed by PCR amplification with primers specific for the ASO and LSO. The ASO primer carries a fluorescent tag that is used for allele calling. The PCR products are hybridized to the microbead array via the complementary oligonucleotides on the beads (Fig. 1.6). The array is then analysed on a specialist bead station (BeadXpress) through analysing the Cy3 or Cy5 intensities at a given SNP site. If equal signal intensities are received with an approximate value of 1:1, then a heterozygous genotype is scored for that specific SNP, but if the signal intensity for only Cy3 is seen (1:0) then a homozygous genotype is scored and vice versa for Cy5 (Shen *et al.* 2005).

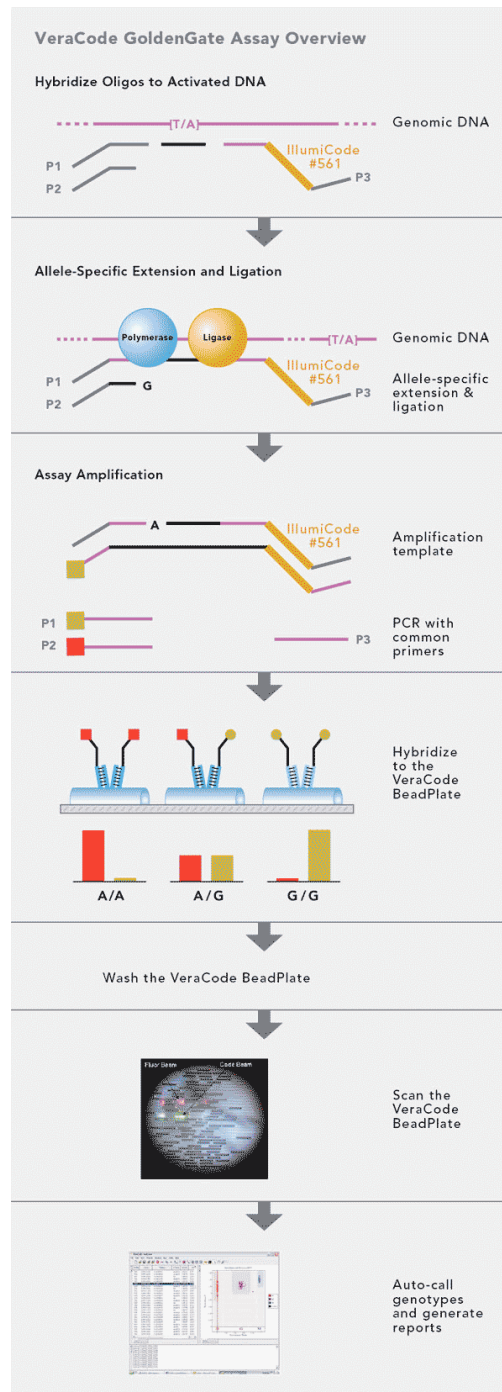


Figure 1.6: A workflow of the VeraCode GoldenGate assay.

(http://www.illumina.com/technology/veracode_goldengate_assay.ilmn)

5. Transcriptome sequencing as a valuable resource for marker development

5.1 Overview of transcriptome sequencing using next generation sequencing (NGS) platforms

Chapter one-Literature review

With the advent of NGS it has become possible to sequence transcriptomes or generate ESTs of non-model organisms for which limited genomic resources are available. This is especially true for various species where whole genome sequencing is still impractical. These functional sequences provides a number of benefits including: lack of introns and non-coding DNA, which makes interpretation of data much easier as well as highly functional information enclosed in the sequence, as it corresponds to sequences of genes. Thus transcriptome sequencing is a very useful tool for gene discovery and annotation, marker discovery and population studies dealing with genetic variation such as adaptive traits (Parchman *et al.* 2010).

There are several sequencing platforms available for transcriptome sequencing. One of these is the Illumina Genome Analyser II, which utilizes sequence-by-synthesis technology. This sequencing technology removes several time-consuming steps associated with traditional Sanger sequencing as well as being more cost- and time efficient (Margulies *et al.* 2005; Ellegren 2008; Hudson 2008; Vera *et al.* 2008; Parchman *et al.* 2010).

Currently the Illumina Solexa Genome Analyzer II produces hundreds of millions of sequences of 2x150 bp long. For non-model species such as *H. midae*, *de novo* assembly of a whole genome sequencing run is a daunting task, which makes transcriptome sequencing a better option as the sequence template is devoid of introns and intergenic DNA that complicates *de novo* sequence assembly without a reference genome. The coverage depth is also higher, when looking at the amount of data generated in a transcriptome sequencing run, because of the smaller size of the transcriptome compared to its corresponding genome (Emrich *et al.* 2007; Pop and Salzberg 2008; Wall *et al.* 2009; Parchman *et al.* 2010). The longer reads (150bp), as opposed to previous versions' shorter reads (50bp), produced by the Illumina Solexa Genome Analyser II, also enables longer contig assemblies, making *de novo* sequencing increasingly easier for organisms with no reference genome (Available at http://www.illumina.com/Documents/products/technotes/technote_denovo_assembly.pdf, accessed June 2011).

The Illumina Solexa Genome Analyser II sequencing process makes use of an 8-lane glass flow plate which has an *in-vitro* single-stranded oligo-adapter ligated

library attached to it. Cluster PCR amplification is conducted on a cluster station and is possible because both primers are available on the glass flow cell. Each cluster library is amplified and the template folds over to form a bridge. After PCR, approximately a thousand copies of each cluster are obtained, which are then sequenced (Fig. 1.7). These cluster templates are sequenced by starting with a 3' OH deactivated, fluorescently labelled dNTP ensuring that only a single base is incorporated. The resultant image is captured and the dNTPs de-blocked for the following cycle of base incorporation. The whole process should take about 4 days and the sequence reads obtained are 100-200bp (Available at http://www.illumina.com/Documents/products/technotes/technote_denovo_assembly.pdf, accessed June 2011). This sequence technology has been available since 2006 and has been used in many high-throughput studies (Celton *et al.* 2010; Frio *et al.* 2010; Graham *et al.* 2010; Hyten *et al.* 2010; Turner *et al.* 2010; Gunnarsdóttir *et al.* 2011; Xu *et al.* 2011).

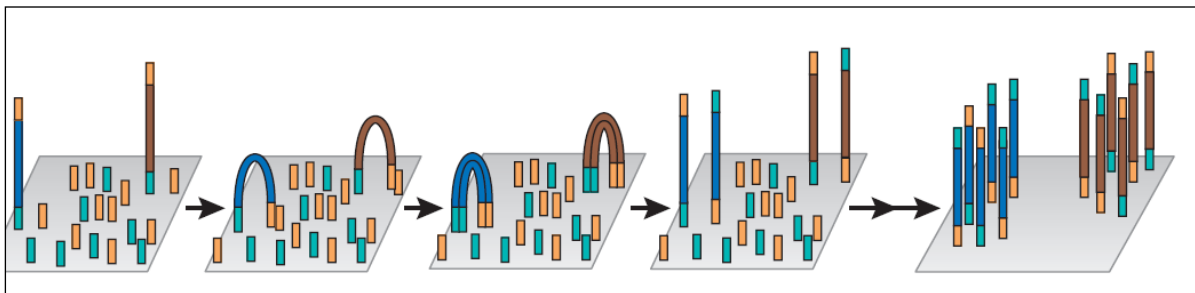


Figure 1.7: The workflow of the Illumina Solexa Genome Analyser. An *in vitro*-constructed adaptor-flanked shotgun library is attached to the solid surface of the flow cell. Cluster PCR is performed within the area of the original library as the surface is covered with both primers. Approximately 1000 copies of a single template library are created in clusters (Shendure and Ji 2008).

5.2 Marker development using NGS platforms

A very accurate and popular technique for SNP discovery is direct sequencing of DNA (in particular ESTs or transcriptome sequencing as these generate type 1 SNPs). NGS has sped up the process of developing SNPs in this way as these sequencing technologies generate sequences in a relative short amount of time, compared to traditional Sanger sequencing and potentially contain thousands of SNPs. SNP identification using NGS has been successfully utilised in a variety of

species, including half-smooth tongue sole (Sha *et al.* 2010); an important tree species, *Pinus contorta* (Parchman *et al.* 2010); Sydney blue gum, *Eucalyptus grandis* (Novaes *et al.* 2008); round worm, *Caenorhabditis elegans* (Hillier *et al.* 2008); catfish (He *et al.* 2003); maize, *Zea mays* (Barbazuk *et al.* 2007); as well as cattle, *Bos taurus* (Van Tassell *et al.* 2008). The transcriptome of *H. midae* has also been sequenced via NGS (Franchini *et al.* 2011) and used for SNP identification (Blaauw 2011).

EST sequencing or transcriptome sequencing by NGS platforms has also proven to be successful for identifying and developing microsatellites (Karsi *et al.* 2002; Zhan *et al.* 2008a; Zhan *et al.* 2008b; Dempewolf *et al.* 2010; Li, H. *et al.* 2010; Parchman *et al.* 2010; Sha *et al.* 2010; Vogiatzi *et al.* 2011). These gene-linked markers have several benefits over genomic markers including possible linkage to functional genes allowing for the mapping of gene-associated markers and comparative genomics (Sarropoulou *et al.* 2008). NGS will provide the opportunity of discovering thousands of gene-linked markers that can be mapped to a linkage map for species with little or no genomic information. This will increase the chances of discovering QTLs, which can further on be used in MAS for economically important traits (Jalving *et al.* 2004; McAndrew and Napier 2010).

6. Linkage mapping

Microsatellites are currently the marker of choice for genetic map construction and numerous microsatellite-based maps have been constructed for aquaculture species (Table 1.2). The reason for this is the generalisation of PCR, the co-dominant inheritance and multi-allelic nature of microsatellites. The multiple alleles lead to high heterozygosity values, lowering the number of reference families needed for building the map. Genotyping is made easier by simple PCR and allele sizing on polyacrylamide gels, followed by sequencing on ABI sequencing systems, as confirmation of the polymorphism, and automated genotyping (Vignal *et al.* 2002).

According to Liu and Cordes (2004), if linkage mapping is the primary goal of a research project, it is advised to develop type 1 microsatellites from the start from either EST-libraries or a sequenced transcriptome. This would facilitate candidate

gene discovery, as the development of microsatellites can be laborious. Mapping these markers will then have a dual function.

It has to be noted that single nucleotide polymorphisms (SNPs) are fast gaining in popularity and supplementing microsatellite markers as the primary markers for mapping. The reason for this shift is that single base changes may be responsible for variations between individuals and are more frequently associated with QTLs. Furthermore, these markers can be used in high-throughput genotyping platforms with fewer genotyping errors occurring than in microsatellites, resulting in lowered costs and improved genotyping data. Lastly, as these markers occur more frequently in genomes than microsatellites, their inclusion could lead to greater saturation of genetic maps (Beuzen *et al.* 2000).

Marker development for abalone has rapidly increased with large numbers of different markers that have been developed and used for amongst others linkage map construction. However, these linkage maps have been limited to only a few commercially important species. These include maps for *H. discus hannai* (Liu *et al.* 2006; Sekino and Hara 2007), *H. rubra* (Baranski *et al.* 2006a) and *H. diversicolor* (Shi *et al.* 2010; Zhan *et al.* 2011).

Table 1.2 Linkage maps consisting mainly of microsatellites for some marine species.

Species	Number of mapped markers*	Map length (cM)*	Linkage groups*	Reference
Arctic charr	327	390/992	46	Woram <i>et al.</i> 2004
Blacklip abalone	102/98	621/766	17/20	Baranski <i>et al.</i> 2006a
Blue mussel	116/121	825/863	14	Lallius <i>et al.</i> 2007
Brown trout	288	346/912	37	Gharbi <i>et al.</i> 2006
Common carp	268	4111	50	Sun and Liang 2004
Japanese flounder	231/304	741.1/670.4	25/27	Coimbra <i>et al.</i> 2003
Pacific abalone	94/119	1366/1774	19/22	Liu <i>et al.</i> 2006
Pacific oyster	119	1031	11	Li and Guo 2004
Pacific oyster	102	616/770	22	Hubert and Hedgecock 2004
Rainbow trout	903	2750	31	Guyomard <i>et al.</i> 2006
Sea bass	162	567/906	25	Chistiakov <i>et al.</i> 2005
Sea bream	204	1242	26	Franch <i>et al.</i> 2006
South China abalone	233/179	2817.1/2773	18/17	Shi <i>et al.</i> 2010
Tilapia	546	1311	24	Lee <i>et al.</i> 2005
Yellow tail	175/122	548/473	21/25	Ohara <i>et al.</i> 2005

* If sex-specific maps were created, the respective male value is given before the female value.

7. Quantitative trait loci

A QTL can be defined as a chromosomal section containing DNA polymorphism that has a significant effect on a specific phenotype of an organism. One QTL will typically not explain all the phenotypic variance seen for a particular trait and the relative contribution of the QTL has to be calculated. The number of QTLs that affect a specific trait will elucidate information about whether the trait is controlled by a large number of genes, each contributing a small effect on the phenotype, or a few major genes, each contributing a large effect (Davie and Hertzler 2000).

Once a high-density genetic linkage map is constructed, markers that are closely linked to a particular QTL can be identified and its position determined on the linkage map due to co-segregation with a molecular marker, such as SNPs or microsatellites (Lo Presti *et al.* 2009). QTLs have been identified and mapped in various aquaculture species, including rainbow trout, *Oncorhynchus mykiss* (Jackson *et al.* 1998; Danzmann *et al.* 1999; Sakamoto *et al.* 1999; Robinson *et al.* 2001; Perry *et al.* 2001; Reid *et al.* 2005); and Nile tilapia, *Oreochromis niloticus* (Agresti *et al.* 2000; Shirak *et al.* 2002; Howe and Kocher 2003).

QTLs have been identified in only two abalone species: *H. rubra* (Baranski *et al.* 2008) and *H. discus hannai* (Liu *et al.* 2007). In the study by Baranski *et al.* (2008), a genome-wide search was conducted to detect QTL for growth rate in *H. rubra*. Ten putative QTLs could be identified with the phenotypical variance explained by the QTL ranging from 3.60% to 22.28%. In the study by Liu *et al.* (2007), growth-related characteristics were surveyed for QTL analysis. These included amongst others shell length, total weight, shell width and shell weight. The QTL detected for each trait varied from one to three with variance explained by the QTL ranging from 8.0% to 35.9%.

8. Marker-assisted selection

Marker-assisted selection is the final step in the molecular breeding of aquaculture species with specific desirable traits. It is defined as a selection process in which future broodstock are chosen based on genotypes of molecular markers and not on phenotype alone as would be the case with traditional selective breeding (Liu and

Cordes 2004). To perform MAS, QTLs or genes involved in the expression of certain traits should be identified (Lo Presti *et al.* 2009).

The construction of a linkage map based on a large number of markers (gene-linked markers being ultimately the most informative) is the first step towards MAS. Once the map density is sufficient, QTL identification can commence. QTL mapping involves accumulating phenotypic information for the mapping families as well as typing the segregation patterns of markers in the corresponding families to identify QTL associated with a particular phenotype, such as growth. The number of QTLs affecting the specific trait should also be determined. This information will aid the aquaculture industry by identifying strains that can be crossed to yield enhanced animals, which have an enhanced capacity for certain traits, such as growth (Liu and Cordes 2004). The mapped QTLs, together with the gene-linked marker-based linkage maps, will facilitate the identification of candidate genes through comparative and gene expression studies so that MAS can take place for traits important to industry.

9. Aims and objectives

This study can be divided into two main sections:

1. Microsatellite marker development
2. Linkage mapping of gene-linked markers

1. MICROSATELLITE MARKER DEVELOPMENT USING THE SEQUENCED TRANSCRIPTOME

Aim: To develop gene-linked microsatellites from the sequenced transcriptome of *H. midae*

Objectives:

- To identify gene-linked microsatellite repeats from the assembled contigs of the *H. midae* transcriptome sequencing data set
- Develop primers for the repeats which exhibited sufficient flanking regions.
- To amplify these primers by PCR to determine if the microsatellites are present in *H. midae*'s genome.

Chapter one-Literature review

- To conduct polymorphism screening by PAGE with the optimised microsatellite markers.
- Sequencing of the polymorphic loci to validate the polymorphism.
- Genotype the polymorphic loci in four mapping families to determine level of polymorphism and segregation of these markers.

2. LINKAGE MAPPING OF GENE-LINKED MARKERS

Aim: Use gene-linked microsatellite markers developed in this study and previously developed EST-derived and cross-species microsatellites as well as previously developed type 1 SNP markers to create a linkage map for *H. midae* using 4 full-sib families containing 100 offspring individuals each.

Objectives:

- To conduct segregational analysis and determine if Mendelian inheritance is followed.
- To inspect segregational patterns of alleles that will be used to calculate recombination values by means of odds ratios.
- To group markers according to linkage of odds (LOD) analysis and to order the markers using the regression mapping algorithm found in Joinmap[®]
- To convert the recombination frequencies into genetic distance, centimorgan (cM) using the Kosambi mapping function.
- Construct sex-specific and sex-average maps separately and compare the maternal and paternal maps.
- Calculate genome length to determine the degree of genome coverage of the linkage map.

Chapter two

Type 1 microsatellite development

Chapter two-Type 1 microsatellite development

1. Abstract

Haliotis midae is an important economical aquaculture commodity for South Africa and generates enormous revenue annually. The industry has shifted its focus to the genetic improvement of this species to increase production. The genetic improvement program makes use of molecular markers for various applications including pedigree analysis, linkage mapping, QTL identification and marker-assisted selection. Microsatellites are a very popular marker type for use in aquaculture as it has a range of applications. Currently two-hundred and sixty-four microsatellite markers, including the microsatellites developed in this study, have been developed for *H. midae*. In a previous study, the transcriptome of *H. midae* was sequenced using the Illumina Genome Analyser II platform. The assembled contigs were subsequently screened for microsatellite repeat motifs. Out of the 24341 contigs, 1.3% contained microsatellite repeat motifs. Ninety-eight primer pairs could be designed for PCR amplification. An amplification success rate of 61% was achieved and 25 of these microsatellites were found to be polymorphic. A BLASTX search was performed with the contigs containing the microsatellite repeats to determine which proteins these genic markers are associated with. The development of microsatellite markers using the assembled contigs of *H. midae*'s transcriptome provided gene-linked markers that can be used in future studies including linkage mapping.

Keywords: *Haliotis midae*, microsatellite markers, marker development, gene-linked markers

2. Introduction

Molecular markers can accumulate mutations, which cause them to be polymorphic. When exacerbated by genetic drift and selection, these markers show genetic variation between individuals and species. The markers, which are heritable, can be used in a variety of applications in aquaculture, including population studies, linkage mapping, maternal lineage and hybrid identification (for review see Liu and Cordes 2004). Molecular markers that have had wide utility in aquaculture include allozymes, mtDNA, RFLPs, RAPDs and AFLPs, with microsatellites being the most popular of these marker types. Recently SNPs have begun to supplement microsatellites as the predominant marker for aquaculture studies. The uses of SNPs are only now being fully appreciated and it is speculated that future aquaculture studies will be using SNPs to a larger extent (Liu and Cordes 2004; Dahle *et al.* 2008).

Microsatellite markers are made up of simple sequence repeats that are about 1-6bp long and many copies of repeats can occur in tandem (Litt and Luty 1989; Tautz 1989). Microsatellites are evenly spaced throughout the genome and are located in genic and anonymous DNA. They are small enough to be amplified by PCR, which is important for genotyping. The number of repeats of a given microsatellite can vary considerably, making it very polymorphic and thus useful in an array of different studies. The polymorphism of microsatellites is based on the size difference of alleles that can be visualised by techniques such as PAGE. This size difference is propagated by a process termed replication slippage that occurs when the DNA polymerase slips during DNA replication (Levinson and Gutman 1987; Tautz 1989).

This type of marker has the highest PIC value compared to other markers due to the number of alleles that can be present at a specific locus. Microsatellites are also co-dominant markers allowing for easy identification of different alleles. Applications of microsatellites in aquaculture include genome mapping, parentage, kinship and stock structure analyses (O'Connell and Wright 1997; Liu and Cordes 2004; Chistiakov *et al.* 2006). Microsatellites have been isolated for a variety of marine species including black tiger prawn *Penaeus monodon* (Xu *et al.* 1999), New Zealand snapper *Pagrus auratus* (Hauser *et al.* 2002), goldfish *Carassius auratus gibelio* (Yue *et al.* 2004) and Atlantic salmon (Vasemägi *et al.* 2005).

Chapter two-Type 1 microsatellite development

Microsatellites have also been developed for a variety of abalone species, including Pacific abalone *H. discus hannai* (Sekino *et al.* 2005; An and Han 2006), blacklip abalone *H. rubra* (Evans *et al.* 2000; Baranski *et al.* 2006b), green abalone *H. fulgens* (Cruz *et al.* 2005), pink abalone *H. corrugata* (Díaz-Viloria *et al.* 2008), as well as Perlemoen *H. midae* (Bester *et al.* 2004; Slabbert *et al.* 2008; 2010).

With the advent of next-generation sequencing it has become possible to sequence whole genomes of organisms. For non-model organisms, which have little or no genomic information available, this still largely remains impractical. Sequencing of transcriptomes or ESTs may represent a more appropriate approach for generating sequences for these non-model organisms, such as *H. midae*. These sequences contain functional information and are devoid of introns and intergenic DNA, which could complicate *de novo* sequence assembly (Andersen and Lübberstedt 2003; Bouck *et al.* 2007). Thus, transcriptome sequencing is a very useful tool for gene discovery and annotation, and population studies dealing with genetic variation such as adaptive traits (Parchman *et al.* 2010).

ESTs or sequenced transcriptomes can also be used as a very effective tool for marker discovery. These regions can be used in the identification of polymorphic type 1 or genic molecular markers, which can be used to saturate and add functional information to linkage maps. These gene-linked markers can also be used for QTL identification, gene discovery, and comparative mapping as well as population studies. In this study microsatellite markers were developed from the sequenced transcriptome of *H. midae* in order to generate polymorphic type 1 markers to saturate the preliminary linkage map of *H. midae*. A denser linkage map will provide a step forward towards candidate gene discovery and QTL identification studies, currently underway in this economically important species.

3. Materials and methods:

3.1 Genomic DNA extractions

Randomly selected wild *H. midae* individuals, four each from Saldanha and Witsand, were selected for the PCR optimisation and polymorphism analysis of the microsatellite loci.

Chapter two-Type 1 microsatellite development

DNA extractions were done using the CTAB extraction method (Saghai-Marooif *et al.* 1984). Epipodial tentacles were used for DNA extractions (Slabbert and Roodt-Wilding 2006). Three hundred microliters of CTAB lysis buffer (1.4M NaCl; 20mM Ethylene Diamine Tetra-Acetate (EDTA [pH 8]; 2% (m/v) CTAB; 100mM Tris-HCl [pH 6.8]), 0.2% (m/v) β -mercapto-ethanol) and 0.5mg/ml Proteinase K was added to the tissue. This was then incubated overnight in a water bath at 60°C. After incubation, equal volumes of chloroform: isoamylalcohol (24:1) was added to the homogenised solution and shaken on a vortex for 5 min at low speed. The samples were subsequently centrifuged at high speed (12000 rpm) for 5 mins using an Eppendorf Centrifuge 5415D at 4°C. The supernatant was carefully transferred with pipette to a new eppendorf tube. If some of the interphase or lower phase was transferred with the supernatant, the chloroform: isoamylalcohol (24:1) step with vortex and centrifugation was repeated. After 2/3 volume isopropanol was added, the samples were incubated at -20°C to precipitate the DNA overnight. Following this incubation, the samples were centrifuged at 12000 rpm for 20 mins and the pellet washed with 70% (v/v) ethanol, followed by a second centrifugation step. Alcohol was carefully removed and the pellet dried in an oven at 55°C, where after 100 μ l of ddH₂O was added to the dried pellet to resuspend it. The resuspended DNA was stored at -20°C in the freezer until further use.

3.2 Microsatellite identification from the *H. midae* transcriptome and primer design

Previously, the transcriptome of *H. midae* was sequenced using 19 animals on the Illumina Genome Analyser II and 24341 contigs assembled using this data (Van der Merwe *et al.* 2011). The contigs were assembled using the Velvet v0.7.57 software (Zerbino and Birney 2008) with parameters set as per Van der Merwe *et al.* (2011). The sequences were then converted into a single FASTA format file using the web-based program BatchPrimer v.3 (You *et al.* 2008). Contigs of less than 100bp were excluded from the FASTA file generation (Hepple 2010).

This FASTA file was used for marker identification using BatchPrimer3 v1.0 (You *et al.* 2008). Parameters were set at six repeat units for dinucleotides, four for tri- or tetranucleotides and for penta- and hexanucleotide markers, three contiguous repeats were used to identify microsatellite repeats from the data file.

Chapter two-Type 1 microsatellite development

Microsatellite repeats that had sufficient flanking sites were used to design primers using BatchPrimer3 v1.0. Parameters were set at: primer length 16-25bp with 21bp as optimal, GC-content 30-70% (optimal 50%), annealing temperature 50-70°C (optimal 60°C), product size 90-350bp (optimal 120bp). Primers with an optimal length of 20bp and an optimal annealing temperature of 60°C were chosen for this study.

The contig sequences containing the microsatellites were searched for similarity to the Molecular Aquatic Research Group (MARG) database to check their redundancy. The microsatellites in the contig and database sequences were masked using the program RepeatMasker (Smit *et al.* 2004) and a BLAST similarity search conducted to identify significant hits. Hits with an E-value of <1E-03 were taken as positive hits and those microsatellites were excluded from the study.

3.3 Contig homology search

The contigs containing SSR repeats were used to search the NCBI non-redundant database using Blast2GO v2.5.0 software with the BLASTX functionality. Positive hits were those which had an E-value of <1E-03.

3.4 Microsatellite amplification and analysis of polymorphism

The same initial PCR program was used to amplify all microsatellite loci. Optimisation was done in a final volume of 10µl containing 20ng of genomic DNA, 0.5X Springbok buffer containing 1.5mM MgCl₂, 0.8mM dNTPs, 0.4mM of each primer and 0.5U Springbok *Taq* polymerase (Molecular Diagnostic Services). The PCR cycle consisted of an initial denaturing step at 94°C for 5 mins followed by 25 cycles of 45 s at 94°C, 45 s at 50°C and 45 s at 72°C, with a final extension step at 72°C for 10 mins. According to the initial PCR product of the microsatellite loci [visualised on a 2% (m/v) agarose gel (Appendix 1C)], the annealing temperature was either lowered (where no band was visualised) or increased (where non-specific bands were visualised). The PCR amplifications were done in a GeneAmp® PCR system 2700 thermocycler.

The microsatellites that could be optimised were analysed for polymorphism level. Eight randomly chosen individuals were used to screen for polymorphism of the microsatellite loci on a 12% (m/v) polyacrylamide gel (Appendix 1D).

Chapter two-Type 1 microsatellite development

Polymorphic microsatellites' PCR products were first cleaned using a SigmaSpin™ Post-Reaction Clean-Up Column Kit (SIGMA-ALDRICH™) and then sequenced using the BigDye version 3.1 ready mix (Applied Biosystems) according to manufacturer's specifications (for PCR program see Appendix 3 cycle-program 4). Four individuals' PCR products were chosen for sequencing, based on the quality and polymorphic appearance on the polyacrylamide gel. Sequence products were sent for capillary electrophoresis at the Stellenbosch University Central Analytical Facility. Sequencing analyses was performed on BioEdit (Hall 1999) in conjunction with multiple alignments of the sequences with the appropriate contig sequences using ClustalW (Thomson *et al.* 1994). This was done to verify that the amplified products were indeed the sequence originally targeted by the microsatellite locus.

3.5 Genotyping

Microsatellite repeats that were polymorphic were labelled with fluorescent dyes (VIC, FAM, NED, and PET [Life Technologies®]) for visualisation of peaks during genotyping. The primer (forward or reverse) chosen to label was dependent on which was closest to the repeat to reduce the amount of stutter as well as the longest of the two primers (Rampling *et al.* 2001). These primers were then amplified using the same PCR conditions as for the unlabelled counterpart. Some optimisation was necessary in some cases (see above section on PCR optimisation).

The microsatellites were subsequently sorted into multiplexes (Table 2.2) for genotyping. The colour of the labelled dyes and the respective lengths of the expected products were taken into account as not to include two microsatellites with the same length and the same colour dye. A QIAGEN® Multiplex kit was used to amplify the target loci in the multiplex according to the manufacturer's instructions. A final volume of 10µl containing 10ng of template DNA was added to 3.5µl 2X QIAGEN Multiplex PCR master mix (containing HotStart *Taq*® DNA Polymerase, Multiplex PCR Buffer with 6mM MgCl₂ and dNTP Mix) (QIAGEN®), 0.9µl primer mix (20µM of each primer) and dH₂O. The PCR cycle followed to amplify the target locus was: an initial 10 min denaturing step at 95°C, followed by 35 cycles of 94°C for 30 s, 57°C for 90 s and 72°C for 1 min. The PCR completed with an elongation step of 60°C for 30 mins.

Chapter two-Type 1 microsatellite development

4. Results

A total of 328 microsatellites were identified from *H. midae*'s transcriptome (Table 2.1). Of all the repeat types searched, trinucleotides was the most abundant (61%), with dinucleotides the second most abundant (22%) and tetranucleotides, the third most abundant (15%) microsatellite repeat type. Only three penta- and hexanucleotides were found (Table 2.1). Of the 312 sequences that contained microsatellites, 104 contained sufficient flanking regions for primer design (Table 2.1). Some of the microsatellites were shown to be redundant after a BLAST similarity search against the MARG database (6 microsatellite markers) and were excluded from the study.

Table 2.1: A summary of the microsatellites identified in *H. midae*'s transcriptome.

Number of sequences searched:	24341
Number of sequences containing repeats:	312
Number of sequences with sufficient flanking regions for primer design:	104
Number of primer pairs:	98
Total number of microsatellites found:	328
Dinucleotide:	72
Trinucleotide:	201
Tetranucleotide:	49
Pentanucleotide:	3
Hexanucleotide:	3

Of these 98 primer pairs (Appendix 2, Table S1), 62 (63%) could be successfully optimised with Springbok *Taq* (Fig. 2.1). After amplification of eight randomly chosen individuals with these optimised microsatellite loci (Appendix 2, S2), the markers were analysed for polymorphism by conducting PAGE analyses (Fig. 2.2). Twenty six fluorescently labelled primers could be designed, based on the level of polymorphism observed. The colours of the labels were chosen as to facilitate grouping markers together for multiplexes (see section 3.5). After optimisation of the labelled primers, using the standard optimised PCR conditions, 25 primer pairs remained for genotyping as locus *ILL2.98293* could not be re-optimised with the fluorescently labelled primer (Table 2.2).

Chapter two-Type 1 microsatellite development

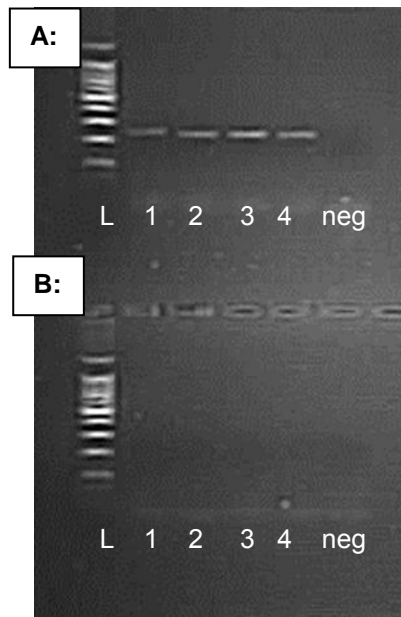


Figure 2.1: 2% Agarose gel. A: optimised loci. B: no PCR product. L=1000bp ladder, 1-4=individuals, neg=negative control.

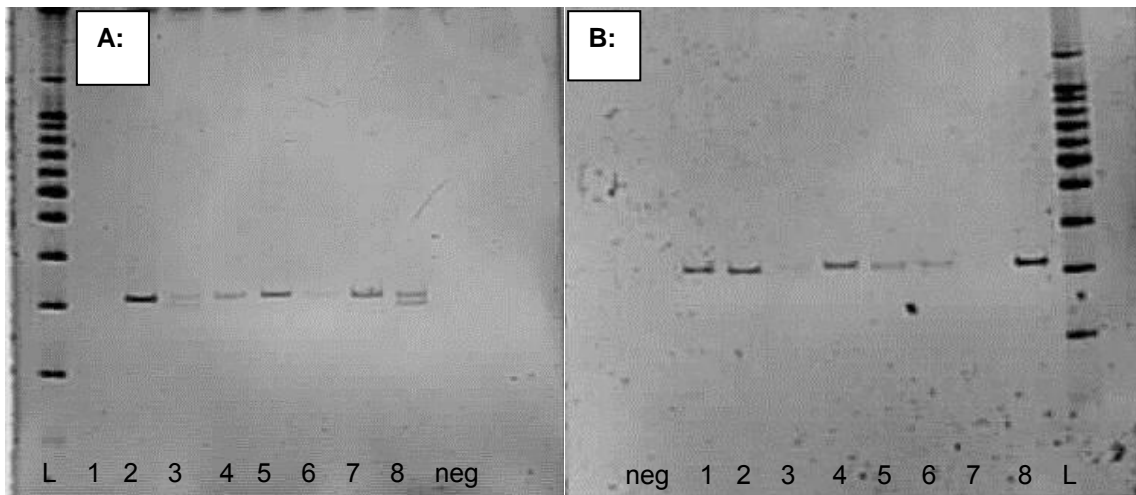


Figure 2.2: 12% PAGE gels. A: Polymorphic locus. B: Monomorphic locus. L=1000bp ladder, 1-8=individuals, neg=negative control

Table 2.2: Twenty-six polymorphic EST-STR marker loci

Marker name	T _a (°C)	Motif	Multiplex	Primers
<i>ILL2.140027</i>	50	TTC	ESTMP10	F: NED -TCCATTCCTTCATCTCTTACACTCT R: AACAAAGAAGAACCCAGAACACTCTA
<i>ILL2.64307</i>	50	AAG	ESTMP10	F: ATGAGGAAGAGAATGAGGATAAACC R: FAM -GATAAACACACGTCTCACATACAGC
<i>ILL2.134845</i>	50	AGG	ESTMP10	F: VIC -GATAGATTTTGGAAACGAAGACAGG R: GTAGTGCAGGGGTACTTTACCAG
<i>ILL2.38396</i>	50	ATT	ESTMP10	F: PET -ATATCCCAATGTGACGGAGATACTA

Chapter two-Type 1 microsatellite development

				R: GGAGAAACCTATACCGATAAGGAAC
<i>ILL2.70036</i>	50	TG	ESTMP11	F: TCTACATCTAAACTGGCATCTACAGC R: NED -CATAAAACTGGTTTAACTGCTGGAC
<i>ILL2.76149</i>	50	AT	ESTMP11	F: FAM -GGAATGGGCTAGTAAGTAGTTCAAA R: ACTAAATGTTTCAGATGTGTGGTGAC
<i>ILL2.118779</i>	50	AT	ESTMP11	F: VIC -AAGATTCCCTCCCTGAATAGAATGTC R: GTGTATCTCAATAATGAGCGAATCC
<i>ILL2.29450</i>	50	TG	ESTMP11	F: ATCTGCAGGAGTATTTTCAGTGATTC R: PET -TATCACATAGAAGCGTCTCTGACC
<i>ILL2.104411</i>	50	CA	ESTMP13	F: FAM -TTGTCTGTGTTGCACATATATACCG R: TTGTCTGTGTTGCACATATATACCG
<i>ILL2.126949</i>	50	CA	ESTMP14	F: GGATACACACCTACACACACTCGAT R: VIC -GTATGTGTTTCGTACGTGTTTGCTT
<i>ILL2.6458</i>	50	CAT	ESTMP12	F: NED -TCATTCTTCCAGTCAGATATCAGGT R: TGATGTTACTAGTCTTGCTCGATGA
<i>ILL2.39227</i>	50	CCT	ESTMP12	F: CCCTACCCCTTATCTAGTTGGTG R: FAM -AGACTACAAAGTGACAGGTGGATCT
<i>ILL2.60863</i>	50	TCA	ESTMP12	F: VIC -TGTCAACATCAGCTAGTTTGTGCATC R: TATGATCAAGGAGAGCATAGAGGAG
<i>ILL2.64121</i>	50	CCT	ESTMP12	F: GTGTTGGTTGTATTAATGTGTGTGC R: PET -AAATAGAAAGTAGGAGGGAGGGAAG
<i>ILL2.66010a</i>	50	ATT	ESTMP13	F: NED-GAGTACCAAAGGGAGATAACCAAAC R: AAGAATGCCTACGATACTTCAACAC
<i>ILL2.93709b</i>	50	CCG	ESTMP13	F: VIC -AATGCTAACAACCTTGAGAGGAAG R: TCTCTCTTCTCAGAGCCTCCAC
<i>ILL2.97931a</i>	50	TGA	ESTMP13	F: GAACAGAGACATTTGACCCTAAAGA R: PET -TATAGACTTGCAAAGAGTTGTCTCG
<i>ILL2.112066</i>	50	GAA	ESTMP14	F: AAGATTGACCTTCCCTCACTTTCTG R: NED -CAATTTCTTCTTTCAGTCCAACCTCTC
<i>ILL2.128607</i>	50	GCA	ESTMP14	F: FAM -ATTACATTATTGACCAGCCAGAC R: CATCACAGATGTGTAGCTAAATTGG
<i>ILL2.71359</i>	50	GCA	ESTMP9	F: GGTTTGCAATGACCCCATAC R: NED -TGTGCATCTATGCGTGTGTG
<i>ILL2.87955</i>	50	GTGA	ESTMP9	F: FAM -ATTCTCAGCTGCATGGTTCC R: CGAGCGTGCCTCTATTCTGT
<i>ILL2.128551</i>	50	TGGA	ESTMP9	F: VIC -CTGAGCTCTGCAGCTGTGAC R: GCTCTGTTCCGTTGTATGCTG
<i>ILL2.140858</i>	50	GTGA	ESTMP9	F: PET -TTCTAGATGTGCGCATGACCT R: GGTGTCTTTGGCCAGGTAGT
<i>ILL2.8738</i>	50	(TGT) _n (TGC)	ESTMP15	F: NED -ACATTCCATAAGCTGGTCTTTACACG R: AAGTGAATGCCTCATGTTATTAGC
<i>ILL2.47613</i>	45	AC	None	F: FAM -GTGTTTACAAGGCGTCATATCAGTA R: CCTGCGCCTACTTACAATAAATG
<i>ILL2.98293b</i>	50	CA	ESTMP15	F: AATATCTGCGCATCACTCACAC R: VIC -GGTCGTAGTAGTCGATTTCTGAATG

Thirteen of the 26 contigs, containing polymorphic microsatellite repeats, exhibited sequence similarities to protein sequences in the NCBI non-redundant database when searched with the BLASTX functionality of Blast2GO v5.2.0 software (Table 2.3). Two of the loci (*ILL2.38396* and *ILL2.64307*) blasted to another abalone species *Haliotis discus hannai* (Pacific abalone) of Japanese origin. Another

Chapter two-Type 1 microsatellite development

gastropod (*Helix aspersa*), the garden snail, produced a hit to microsatellite loci *ILL2.60863*.

Table 2.3: BLAST results of the polymorphic microsatellites indicating the sequence description, organism, E-value and accession number for each contig that showed a positive hit.

Microsatellite loci	Sequence description	Organism	E-value	Genbank Accession number
<i>ILL2.6458</i>	Oxidoreductase, zinc-binding dehydrogenase family	<i>Trichinella spiralis</i>	5.16115E-68	NP_502269.1
<i>ILL2.39277</i>	Neuro precursor cell developmentally down-regulated 8	<i>Aedes aegypti</i>	1.13745E-6	XP_001650246.1
<i>ILL2.60863</i>	Eukaryotic translation initiation factor 2 alpha subunit	<i>Helix aspersa</i>	2.00889E-61	AAO20109.1
<i>ILL2.76149</i>	S-adenosylhomocysteine hydrolase-like 1	<i>Danio rerio</i>	1.1097E-65	CAN87933.1
<i>ILL2.70036</i>	Transaldolase-like protein	<i>Anopheles darlingi</i>	9.39243E-15	EFR24452.1
<i>ILL2.93706</i>	Muscle myosin heavy chain	<i>Loligo bleekeri</i>	0.0	ACD68201.1
<i>ILL2.112066</i>	Calmodulin	<i>Renilla reniformis</i>	5.33662E-20	P62184.2
<i>ILL2.128607</i>	2-oxoglutarate dehydrogenase	<i>Aedes aegypti</i>	1.00252E-105	XP_001652168.
<i>ILL2.38396</i>	secreted protein, acidic and rich in cysteine	<i>Haliothis discus discus</i>	6.60009E-15	BAK22657.1
<i>ILL2.140027</i>	Ras-related protein Rap 1-b precursor	<i>Branchiostoma floridae</i>	1.17162E-62	XP_002594348.1
<i>ILL2.8738</i>	B-cell translocation gene 1	<i>Crassostrea gigas</i>	9.76656E-45	ACH92125.1
<i>ILL2.64307</i>	Heat shock protein 90	<i>Haliothis discus hannai</i>	2.61824E-180	ACX94847. 1

Chapter two-Type 1 microsatellite development

<i>ILL2.29450</i>	X-box binding protein 1	<i>Mytilus edulis</i>	1.84592E-31	ABA43316.1
-------------------	-------------------------	-----------------------	-------------	------------

5. Discussion

The method of choice for developing microsatellite markers for various non-model species has been the FIASCO technique (Zane *et al.* 2002; Reece *et al.* 2004; Baranski *et al.* 2006b; Slabbert *et al.* 2008; Zhan *et al.* 2008a; Hepple 2010). However, this is a very laborious method for microsatellite development as it includes the construction of clones. It has been speculated that the time and funding that is invested in this avenue of microsatellite development is not worth the number of microsatellites that can be used in further analysis (Slabbert *et al.* 2010). Hepple (2010) demonstrated that developing microsatellites from transcriptome contigs provided a 37% success rate, whereas the FIASCO technique had a 32% rate. Therefore, although the efficiency of both strategies for developing microsatellite markers is comparatively similar, SSRs developed using ESTs have an enhanced utility as they contain functional information. The relative ease associated with developing EST-SSRs also makes this avenue more attractive for marker development for non-model organisms.

The sequenced transcriptome of *H. midae* proved to be a valuable resource for developing type 1 microsatellite markers. More than 300 microsatellite repeats could be identified from the contigs and 98 primer pairs could be developed. The percentage (1.35%) of repeats identified from the 24341 contigs are relatively low compared to other aquaculture species [4.95% for Pacific abalone (Li, Q *et al.* 2010a.); 2.2% for fleshy prawn, *Fenneropenaeus chinensis* (Wang *et al.* 2005); 13.7% for black tiger shrimp (Maneeruttanarungroi *et al.* 2006); and 11.2% for channel catfish, *Ictalurus punctatus* (Serapion *et al.* 2004)]. This is probably due to partial coverage of the transcriptome by the assembled contigs (Van der Merwe *et al.* 2011). Evidence for this is the low match of microsatellites from *H. midae* (genomic and EST-SSRs) that could be located in the assembled contigs, indicating that some previously developed EST-SSR markers could be located in sequences that were not covered by the contigs (Franchini *et al.* 2011).

Chapter two-Type 1 microsatellite development

Type 1 microsatellites generally have a higher amplification rate with less PCR optimisation than genomic microsatellites and show a higher degree of cross-species amplification due to their conserved nature (Yu and Li 2007; Parchman *et al.* 2010; Qi *et al.* 2010; Franchini *et al.* 2011). This could explain the high amplification rate (63%) of the microsatellites developed in this study. Similar studies in Perlemoen (Hepple 2010); Pacific abalone (Zhan *et al.* 2008b; Li, Q *et al.* 2010); Pacific oyster, *Crassostrea gigas* (Yu and Li 2007, 2008); Chinese scallop, *Chlamys farreri* (Zhan *et al.* 2008a); and tongue sole (Sha *et al.* 2010) obtained similar amplification rates with microsatellites developed from ESTs.

Microsatellites of all expected types (di-, tri-, tetra-, penta- and hexanucleotides) could be identified from the contigs. Trinucleotides were the most abundant form of microsatellites identified in this study. This was expected as these microsatellites occur in coding regions with greater mutational constraints (Conner and Hughes 2003). Mutations, which will cause changes in the open reading frames, can possibly cause serious phenotypical changes. When a trinucleotide gains or loses a repeat unit, no change will occur in the open reading frame, but if for example, a dinucleotide gains or loses a repeat unit the open reading frame will change, resulting in a frameshift mutation, which could have a biological consequence (Sha *et al.* 2010).

Polymorphic microsatellites form the basis for linkage mapping studies. The development of these microsatellites have been very difficult for *H. midae* as only a very small number of microsatellites, developed for this species, are polymorphic (Bester *et al.* 2004; Slabbert *et al.* 2008; Rhode 2010; Hepple 2010; Slabbert *et al.* 2010). This study developed 25 new polymorphic microsatellite markers, which can be used for linkage mapping. In the current study, only 42% of the successfully amplified microsatellites were found to be polymorphic after polymorphism screening in eight wild individuals and sequencing of the PCR products. Type 1 microsatellites have been noted to be less polymorphic than genomic microsatellites as these regions are more conserved between individuals as well as between different species (Cho *et al.* 2000; Gupta *et al.* 2003). This is in accordance with the results obtained in the current study, although some studies have shown that high polymorphism rates could be obtained by EST-SSRs (Zhan *et al.* 2008a; Li, Q *et al.* 2010), whereas other studies show similar polymorphic rates to microsatellites

Chapter two-Type 1 microsatellite development

developed from genomic DNA (Yu and Li 2007; 2008; Zhan *et al.* 2008b; Li, H *et al.* 2010).

The microsatellites that were shown to have an association with known protein sequences (Table 2.3) provide a valuable resource for comparative studies as well as candidate gene identification in *H. midae*. Two organisms of the class gastropoda, *H. discus hannai* and *Helix aspersa* (garden snail) showed sequence similarity to some of the developed markers in *H. midae*. *Crassostrea gigas* (Pacific oyster), an important aquaculture species, was also amongst the organisms exhibiting sequence similarity to the microsatellite developed from *H. midae*'s transcriptome. Pacific abalone and Pacific oyster both have linkage maps, and recently both have developed more EST-SSRs (Yu and Li 2007; Zhan *et al.* 2008b; Li, Q *et al.* 2010). These species have a large amount of genomic information available, in comparison to Perlemoen, which is seen as a genome 'poor' species. The genomic information of these genome 'rich' species can be used to compare to *H. midae*'s genome and so transfer genome information from the genome rich species to the genome poor species. If the EST-SSRs are found to be associated to the same genes as in this study, and they can be mapped to the species linkage map, the genome locations and organisations can be investigated as well as sequence variation between these genic areas of the various genomes. The garden snail does not have large scale molecular genomic resources, including molecular markers or linkage maps available yet, which makes comparative genomics not currently that plausible. These comparative studies could however elucidate evolutionary relationships and different genomic arrangements between different species (Xia *et al.* 2010).

The microsatellite markers, which were found to be closely associated with known proteins, could serve as a starting point for candidate gene identification for traits important to production such as growth specifically marker *ILL2.8738*, which showed sequence homology to the B-cell translocation gene 1 in the Pacific oyster. This gene is part of an anti-proliferative gene family and has been shown to regulate cell growth and differentiation in humans (Lee *et al.* 2003) and could potentially be involved in growth of abalone as well. Marker *ILL2.29450* was shown to be associated to X-box binding protein 1 (Table 2.3). This protein is a DNA-binding transcription factor involved in B-cells maturation and is thus involved in the immune

Chapter two-Type 1 microsatellite development

response in humans (Liou *et al.* 1990; Smith *et al.* 2008). If this is the case for abalone, this protein may be very important for disease resistance.

These EST-STRs are also very useful in linkage mapping as their positions can provide insights to the locations of genic DNA. If the microsatellites that showed associations to known proteins can be mapped on the linkage map of *H. midae*, QTL identification would be enhanced. The microsatellites, which did not show any sequence homologies to proteins, could possibly indicate that these markers are associated to novel *H. midae* genes. As more sequences are annotated for related species, the more useful this information will become.

The genic microsatellite markers developed in this study can therefore be used in various genetic applications. These include genetic characterisation of wild and cultured *H. midae* populations in population studies, linkage mapping when analysed for segregation in structured families (see chapter 3), pedigree analysis, as well as testing the transferability of these microsatellites to other abalone species (Yu and Li 2008; Li, Q *et al.* 2010).

Chapter three

Linkage mapping

Chapter two-Type 1 microsatellite development

1. Abstract

Constructing a linkage map for a species with little genomic information available, such as Perlemoen, provides a stepping stone towards marker-assisted selection. It is these breeding programs that could lead to genetically superior abalone and thus sustainable farming practices. A linkage map densely populated with markers allows for the segregation of these markers alongside traits important to production, such as disease resistance and growth. This study provides a fundamental step towards future identification of QTLs and genes. In this study, four full-sib families were used for map construction, each consisting of approximately 100 offspring. Two-hundred and ninety five markers were available for mapping (239 previously developed microsatellites, 31 previously developed polymorphic gene-linked SNPs, and 25 polymorphic gene-linked microsatellite markers developed during the current study) and subjected to segregation analysis and linkage analysis for sex-specific and sex-average map construction using Joinmap[®] v.4. Fifty-nine markers could be mapped to family DS1's maternal map with 18 linkage groups and an average genome length of 1482.5cM. For the paternal derived map of family DS1 64 markers could be linked to 18 linkage groups with an average genome length of 1105.5cM. Family DS2's maternal map had 49 markers ordered on 15 linkage groups with an average genome length of 1292cM. The paternal map of family DS2 had 51 markers that could be linked to 17 linkage groups with an average genome length of 1358.5cM. The sex-specific maps of family DS5 had 44 markers placed on 15 linkage groups on the maternal map and 37 markers with 13 linkage groups on the paternal map with average genome lengths of 992.75cM and 602.5cM, respectively. Forty-three markers could be placed on 14 linkage groups of family DS6's maternal map. The genome length of this map was 1383.5cM. The paternal derived map of family DS6 had 44 markers that could similarly be mapped to 14 linkage groups, resulting in an average genome length of 952cM. The sex-average maps of the four families were merged where possible to form an integrated map. This integrated map contained 109 markers that could be mapped to 25 linkage groups. The resulting average genome length of the map was calculated at 1684cM with an average marker spacing of 9.3cM. This is the first linkage map constructed for *Haliotis midae* to contain SNPs. These markers are ideal for constructing high-density linkage maps. The linkage map created in this study's map density is still however relatively low with a genome coverage of 65%. The number of SNPs used in this study was very low (48) and future studies should focus on incorporating more SNPs into the linkage map to increase the map density. The use of genic-derived markers in linkage map construction is very effective and more maps should focus on including these markers into linkage maps.

Keywords: *Haliotis midae*, linkage maps, genome coverage

Chapter two-Type 1 microsatellite development

2. Introduction

In 2005 Stellenbosch University, in conjunction with five abalone aquaculture farms (HIK, Abagold, Aquafarm, Roman Bay and I&J abalone) initiated a genetic program for the enhancement of South African abalone (Brink *et al.* 2009). A number of different genetic projects have been undertaken to reach this goal. These include diversity assessments (Evans *et al.* 2004; Slabbert *et al.* 2009), pedigree analysis (Ruivo 2007; Van den Berg and Roodt-Wilding 2010), gene expression (Franchini *et al.* 2011; Van der Merwe *et al.* 2011), triploidy induction (Brink *et al.* 2009) and linkage mapping (Badenhorst 2008; Hepple 2010).

Genetic maps have been generated for a number of aquaculture species. The first of these consisted of maps constructed using AFLP markers. Aquaculture species with AFLP-based linkage maps include Nile tilapia, *Oreochromis niloticus* (Kocher *et al.* 1998); walking catfish, *Clarias macrocephalus* (Poompuang and Na-Nakorn 2004); Pacific oyster (Li and Guo 2004); blue mussel, *Mytilus edulis* (Lallias *et al.* 2007); and European flat oyster, *Ostrea edulis* (Lallias *et al.* 2007). AFLPs were also initially used as genetic markers for map construction in abalone (Liu *et al.* 2006; Badenhorst 2008). AFLPs, however, have a major shortcoming in that these markers have low utility in other populations and species. They are also labour intensive to generate and require expensive laboratory equipment (Liu and Cordes 2004).

This makes markers that show a stable pattern of inheritance, are co-dominant, informative and that have utility in other populations and species, a better best choice for constructing dense linkage maps. This came in the form of microsatellite markers, which quickly become the most popular marker choice for linkage mapping projects. Unlike AFLPs, these PCR-based markers are reproducible and are transferable between species and populations (Liu and Cordes 2004). Aquaculture species with genetic maps based on microsatellites include, amongst others, Pacific oyster (Hubert and Hedgecock 2004); Arctic charr, *Salvelinus alpinus* (Woram *et al.* 2004); tilapia, *Oreochromis* spp. (Lee *et al.* 2005); yellowtail, *Seriola quinqueradiata* and *S. lalandi* (Ohara *et al.* 2005); rainbow trout (Guyomard *et al.* 2006); turbot (Bouza *et al.* 2007); brown trout, *Salmo trutta* (Gharbi *et al.* 2006); Pacific white shrimp, *Litopenaeus vannamei* (Zhang *et al.* 2007); and large yellow croaker, *Pseudosciaena crocea* (Ning *et al.* 2007). Microsatellite-based linkage maps have

Chapter two-Type 1 microsatellite development

also been utilised in abalone research and have been constructed for Pacific abalone, *Haliotis discus hannai* (Liu *et al.* 2006); blacklip abalone, *H. rubra* (Baranski *et al.* 2006a); small abalone, *H. diversicolor* (Shi *et al.* 2010); as well as Perlemoen, *H. midae* (Hepple 2010).

Dense linkage maps are required for QTL identification to complement marker assisted breeding programs (Liu and Cordes 2004) and this has led to the investigation of alternative markers for generating dense genetic maps. Single nucleotide polymorphisms have been identified as a suitable marker type for construction of dense linkage maps due to their abundance in genomes and the relative ease with which they can be genotyped using automated platforms. SNP-based linkage maps have been constructed for important aquaculture species including Atlantic salmon (Moen *et al.* 2008); and Pacific white shrimp (Du *et al.* 2009). Microsatellites together with SNP markers are also being utilised more in aquaculture species. Examples of these are linkage maps of the channel and blue catfish, *Ictalurus punctatus* and *I. furcatus* (Kucuktas *et al.* 2009); and grass carp (Xia *et al.* 2010). EST-associated markers are also very useful markers for mapping when considering QTL analysis in downstream projects as these are type 1 markers, they could be linked to a possible QTL (Yue *et al.* 2004).

According to Wu *et al.* (2007) linkage analysis is based on the co-segregation of adjacent markers and their co-transmission to the next progeny generation. The linkage of markers can be determined by their recombination fractions, which in turn can be converted to genetic map distances using a mapping function. Before linkage analysis, the data has to be inspected for Mendelian segregation; i.e. if they segregate independently. The segregation of markers, and their recombination events, are investigated in the mapping populations (backcross, recombinant inbred, F2 intercross, outcross mapping panels and pedigreed multigenerational line information) (Danzmann and Gharbi 2007)

Backcross or F2 inbred lines are the most informative mapping populations for linkage mapping as it allows for easy identification of non-recombinant and recombinant allelic haplotypes, due to the fact that the parents are homologous for every gene. However, most aquaculture populations are outbred, because breeding has usually not progressed to produce inbred populations. In these instances, a full-

Chapter two-Type 1 microsatellite development

sib pedigree will be used in a pseudo-test backcross analysis. A marker segregating in such a family will be derived from two outbred parents and up to four marker alleles could possibly be segregating at a single locus. The offspring can thus inherit two of the possible four alleles, which allow the segregation of the parental alleles to be traced (Wu *et al.* 2007). For linkage mapping projects to be successful when using full-sib families, the number of offspring has to be very large. In abalone, as in many fish species, this problem is overcome as they are highly fecund species (Launey and Hedgecock 2001; Evans *et al.* 2004; Lucas *et al.* 2006).

Markers that are segregating in the mapping panel are subsequently placed into appropriate linkage groups. Chi-square analysis is done on each marker with different recombination ratios depending on whether the parents are both heterozygous or if only one is heterozygous (Wu *et al.* 2007). Linkage analysis can then commence by grouping certain linked loci together. After the loci have been grouped, the mapping order is determined by two algorithms specifically developed for this purpose. Both are incorporated into the Joinmap[®] v4 software package; the regression mapping algorithm (Stam 1993) and the maximum likelihood calculations (Jansen *et al.* 2001). After the markers have been placed in a specific order, the relative mapping distances can be determined.

Double crossovers and interference are two complex phenomena that have an effect on calculating mapping distances. Double crossovers can cause erroneous mapping distances if its occurrence is not taken into account. When a double cross-over occurs it can be misinterpreted as no recombination between two loci; leading to the underestimation of the mapping distance between these two loci. Interference on the other hand is a process whereby chiasma formation along the length of a chromosome inhibits other chiasma formation on the same chromosome. This could lead to the misinterpretation that certain markers are linked when they are in fact unlinked, and can undergo recombination. Mapping functions are utilised to account for these phenomena when converting recombination frequencies into genetic distance (Danzmann and Gharbi 2007). The two most widely used mapping functions are the Haldane (Haldane 1919) and the Kosambi (Kosambi 1944) mapping functions.

Chapter two-Type 1 microsatellite development

In this chapter, microsatellite and SNP markers were subjected to segregation analysis and linkage analysis in four full-sib families using Joinmap[®] v.4 software to construct an integrated, medium-density linkage map for *H. midae*.

3. Materials and methods

3.1 Mapping families

Four full-sib mapping families (DS1, DS2, DS5, and DS6) were used for linkage analysis with the informative microsatellites being genotyped in parents and offspring of all families. The families consisted of a 103 offspring for family DS1, 98 for family DS2 and DS6, and 100 for family DS5. These families originate from two different abalone aquaculture farms: DS1 and DS2 were sampled from Roman Bay Sea Farm (Pty) Ltd in Gansbaai and DS5 and DS6 were sampled from HIK Abalone Farm (Pty) Ltd in Hermanus.

DNA extractions of parents and offspring were conducted as described in chapter 2 (section 2.2).

3.2 Genotyping of the gene-linked markers

3.2.1 Microsatellite markers

The parents of the families were screened with the 25 gene-linked microsatellites developed in this study as well as 239 previously developed microsatellites: 11 from Bester *et al.* (2004) [markers 1-11], 63 from Slabbert *et al.* (2008) [markers 12-74], 25 EST-derived and cross-species microsatellites from Rhode (2010) [markers 75-99] (Appendix 3, Table S3 and S4), 44 from Slabbert *et al.* (2010) [markers 100-143], 14 from Hepple (2010) [markers 144-157] and 82 from Slabbert (2010) [markers 158-239]. *[The PCR amplification and genotyping of markers 1-74 and 100-239 were performed by Miss. J. Vervalle. The 25 gene-linked microsatellites developed in this study and the 25 previously developed EST-derived and cross-species microsatellites (marker 75-99) developed by Rhode (2010) were amplified by PCR and genotyped by the author].*

Microsatellites, which were polymorphic in the parents, were genotyped in the offspring individuals of each family via multiplex reactions. The multiplex reactions were performed as follow: 10ng of template DNA was added to 5µl 2X QIAGEN Multiplex PCR master mix (containing HotStart *Taq* DNA polymerase, Multiplex PCR Buffer, 6mM MgCl₂ and dNTP Mix)(QIAGEN), 0.9µl Primer mix (20µM of each

Chapter two-Type 1 microsatellite development

primer) and dH₂O to a final volume of 10µl. The PCR reaction for the multiplex was as follow: The cycle started with an initial denaturing step of 10 mins at 95°C, followed by 35 cycles of 94°C for 30 s, 57°C for 90 s and 72°C for 1 min. The PCR program is completed with a final elongation step of 60°C for 30 mins. The amplified fragments were sent to the Central Analytical Facility (CAF) for genotyping on the ABI 3730xl DNA Analyser (Applied Biosystems) with a GeneScan™ 600 LIZ® Size Standard (Applied Biosystems) for fragment length determination. Genemapper v4.1 software was used to visualise the fluorescent peaks. This data could then be used to interpret genotypes for the offspring for each marker (Applied Biosystems).

3.2.2 SNP markers

Forty-eight SNPs were genotyped in the four mapping populations. These comprised of 24 SNPs (*in vitro* SNPs; Blaauw 2011), four from Bester *et al.* (2008), eight from Rhode (2010) and 12 novel SNPs developed *in silico* (Blaauw 2011) using the SNP discovery application available on the CLC Workbench v4.5. These SNPS were included on a 48-plex VeraCode GoldenGate BeadXpress assay. The genotyping plots produced by the assay were analysed with GenomeStudio™ Genotyping Module v1.0 software to obtain the SNP genotypes. SNPs and individuals, which did not score reliable genotypes (call rate lower than 80%) or failed in the genotyping run (gencall rate ≤ 0.25), were excluded from further analysis. SNPs that illustrated ambiguous clustering were also removed prior to further analysis.

SNPs that could be mapped to the sex-average and sex-specific maps were searched for similarity in the non-redundant protein database of NCBI using the BLASTX functionality (E-value <1E-03) in the Blast2GO software package v.2.5.0.

3.2.3 Genotype data

Before data analysis, the genotype data of both SNP and microsatellite markers were converted into the appropriate Joinmap® v.4 format for outcross, CP, populations (Van Ooijen 2006). This format consists of five different potential crosses of parental alleles (Table 3.1).

Table 3.1: The Joinmap® v.4 genotype data format for CP populations (Van Ooijen 2006).

Code	Description	Possible genotypes
<abxcd>	Heterozygous in both parents, 4 alleles	ac, ad, bc, bd, --

Chapter two-Type 1 microsatellite development

<efxeg>	Heterozygous in both parents, 3 alleles	ee, ef, eg, fg, --
<hkhk>	Heterozygous in both parents, 2 alleles	hh, hk, kk, --
<lmxl>	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 Linkage analysis

Linkage analysis was performed on all of the informative gene-linked markers by observing their segregational patterns in each family's offspring as well as the maternal and paternal patterns separately by Joinmap[®] v.4 software (Van Ooijen 2006).

Segregational analysis was performed on all of the informative gene-linked markers by a Chi-square goodness-of-fit test to determine which of the markers showed segregation distortion as this could influence linkage analysis. Distorted markers were not excluded from locus grouping, but rather noted. Mendelian observed and expected ratios were tested by Chi-square analysis and tested at seven different significance levels (0.1-0.0001) provided by Joinmap[®] v.4.

Joinmap[®] v.4 has different statistical tests to group loci into linkage groups namely: the independence LOD score test statistic, p-value of the test for independence, recombination frequency and the linkage LOD. The test used in this particular study was the 'independence using a LOD score for significant testing'. The markers' recombination frequencies were used for this. If the recombination frequencies of certain markers were less than a set threshold and had a LOD score higher than the set threshold, they were considered as linked (Van Ooijen 2006). Choosing the LOD threshold value for linkage is dependent on the mapping families' size. For smaller families a more stringent value (i.e. a LOD score higher than 4) should be used for inclusion into certain linkage groups whereas for larger mapping panels of approximately 100 individuals, a LOD score of 3 could be used (Danzmann and Gharbi 2007). Previous studies indicated that a LOD score of 3 was stringent enough to group markers into their specific linkage groups for *H. midae* (Hepple 2010).

Map order can be determined by two mapping algorithms incorporated in Joinmap[®] v.4. The regression mapping algorithm is the default calculation option for

Chapter two-Type 1 microsatellite development

determining map order in Joinmap® v.4, but this algorithm struggles to handle data with more than 50 loci in a given linkage group. In such instances the maximum likelihood calculations option is used. The regression mapping algorithm handles data by adding loci one by one. This procedure is started by inspecting which pair of loci is the most informative. The next locus is added by searching for its best position in the linkage group relative to the informative pair by using goodness-of-fit calculations. If the goodness-of-fit values differ with too much between two loci or when this locus has negative distance values it is removed from further analysis and the next locus is analysed in the same manner until all loci in the linkage group have been through this process once. After the first round of adding the loci in mapping orders, a second and third round could be carried out to possibly add additional loci and position the loci more accurately (Van Ooijen, 2006). In this study the regression mapping algorithm was used.

Recombination frequency is converted into map distances using either the Haldane or Kosambi map functions. There are slight differences between the two functions; the Haldane mapping function corrects for double crossovers, whereas the Kosambi mapping function corrects for both double crossovers and interference (Danzmann and Gharbi 2007; Huehn 2010). In this study, the Kosambi mapping function was used to convert recombination frequencies into map distances (cM).

3.4 Linkage map integration

The sex-specific linkage groups of each family were inspected by eye for marker grouping homology. Those sex-specific linkage groups, which were shown to have the same markers, were combined to form a merged map of both parents, which could then be compared to the homologous sex-average map to verify marker placement and order. In instances where marker order differed between the sex-specific and the sex-average maps, Chi-square values obtained from the goodness-of-fit calculation of comparisons of recombination frequencies, were checked and markers with high values were subsequently excluded from the linkage groups to achieve a more reliable marker order.

The different families' sex-average maps were then integrated using the 'combine groups for map integration' function in Joinmap® v.4 to form a single integrated map, which contained all the mapping information of the four families where possible.

Chapter two-Type 1 microsatellite development

Maps were drawn using MapChart software (Voorrips 2002).

3.5 Genome coverage

3.5.1 Observed map length

To calculate map length, the telomeric regions from the beginning of the linkage group to the first marker of the linkage group must not be excluded. This is achieved by multiplying the length of each linkage group (LG) by twice the average length of the end marker to the end of the linkage group (Ohara *et al.* 2005). Most mapping programmes do this calculation automatically.

3.5.2 Expected genome length

The expected genome length was calculated by two equations, A and B (see below). The average of the two equations was then used to obtain $G_{e\text{ ave}}$.

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

Where G_{oi} represents the observed length of the linkage group i and k_i represents the number of markers at linkage group i (Chakravarti *et al.* 1991).

$$\mathbf{B: } G_e = G_o + [(2tG_o) / n]$$

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

3.5.3 Genome coverage

Genome coverage could then subsequently be calculated by equation C.

$$\mathbf{C: } GC = G_o / G_{e\text{ ave}}$$

Where GC is the genome coverage, G_o is the observed map length and $G_{e\text{ ave}}$ is the average expected genome length.

Chapter two-Type 1 microsatellite development

4. Results

4.1 Gene-linked SNPs

After analysis of the 48 SNPs genotyped by the Illumina VeraCode GoldenGate BeadXpress platform it was shown that seven of the SNPs failed to produce reliable genotypes and that ten were monomorphic (Table 3.2). This resulted in 31 polymorphic SNPs that were available for mapping.

Table 3.2 Genotyping success of the SNPs.

SNPs genotyped	Successful genotypes	Polymorphic SNPs	Monomorphic SNPs	Failed SNPs
48	41	31	10	7

The SNPs that showed significant (E-value <1E-03) similarity to proteins in the non-redundant database are shown in the table below (Table 3.3).

Table 3.3: Blast results for mapped SNP markers.

SNP name	Development approach	SNP polymorphism	Homology	Organism	Genbank Accession number
3D10_1	EST-mining	G>A	Haemocyanin	<i>Haliotis midae</i>	EU135917
3B4_2	EST-mining	C>T	Ribosomal protein L8e	<i>Haliotis midae</i>	EU135916
SNP4691_183	<i>In vitro</i>	G>A	Heat shock protein 70	<i>Stenella coeruleoalba</i>	ACZ54254.1
SNP1834_464	<i>In vitro</i>	G>A	Alpha tubulin	<i>Chlamys farreri</i>	AAR39411.1
SNP149.2_165	<i>In vitro</i>	G>A	Heat shock protein 70	<i>Scylla paramamosain</i>	ACN54681.1
SNP3129_923	<i>In silico</i>	G>A	Arginine kinase	<i>Haliotis diversicolor</i>	ADH59421.1

4.2 Genotyping of the mapping families

Chapter two-Type 1 microsatellite development

The four families were genotyped with the gene-linked markers and those that were found to be polymorphic were inspected for correct allelic combinations of the markers using the parental genotypes. If the polymorphic markers' parent genotype combinations did not include one of the five allelic combinations for CP populations then these were excluded from further analysis (Table 3.4). In family DS1, 99 out of the 295 markers were informative with family DS2 having 81 informative markers, family DS5 having 77 informative markers and family DS6 having 71 informative markers (Table 3.4).

Table 3.4: A summary of the informative markers obtained from inspecting the genotyping data of each mapping family.

Family	Gene-linked microsatellites [*]	Gene-linked SNPs	EST-derived and cross species microsatellites	Other microsatellites [#]	Total
DS1	6	10	5	78	99
DS2	11	0	8	62	81
DS5	4	10	7	56	77
DS6	9	11	7	44	71

* Developed in this study

Genotyped by Miss J. Vervalle

Null alleles occur frequently in microsatellite genotyping data. When the segregation pattern of the null allele can be traced in the pedigree, the null allele can be used in segregation analysis. For all the microsatellites genotyped (by the author as well as Miss J. Vervalle), 44 null alleles were visualised across all four families. In family DS1 all 13 null alleles could be used for segregation analysis. Family DS2 had 16 null alleles and 14 of these could be traced in the pedigree and was thus used for segregation analysis. Eight out of the 11 null alleles found in family DS5 were usable and all ten of family DS6's null alleles were usable. Forty-six of the microsatellite markers were duplicated and were excluded from further analysis (Table 3.4). Markers were inspected for Mendelian segregation. Those markers that did not confirm to correct segregation ratios were deemed distorted. These markers were however not excluded from mapping (Table 3.5).

Table 3.5: Number of null alleles, duplicated and distorted loci for all the markers genotyped.

Family	Null alleles	Duplicated loci	Distorted loci
--------	--------------	-----------------	----------------

Chapter two-Type 1 microsatellite development

DS1	13	21	38
DS2	15	14	32
DS5	11	5	28
DS6	10	6	16
Total	49	46	114

4.3 Linkage mapping

Sex-average maps were created in Joinmap[®] v.4 by selecting the 'create population node' option. Sex-specific linkage maps were created separately in Joinmap[®] v.4 using the 'create maternal and paternal population node' function for each family.

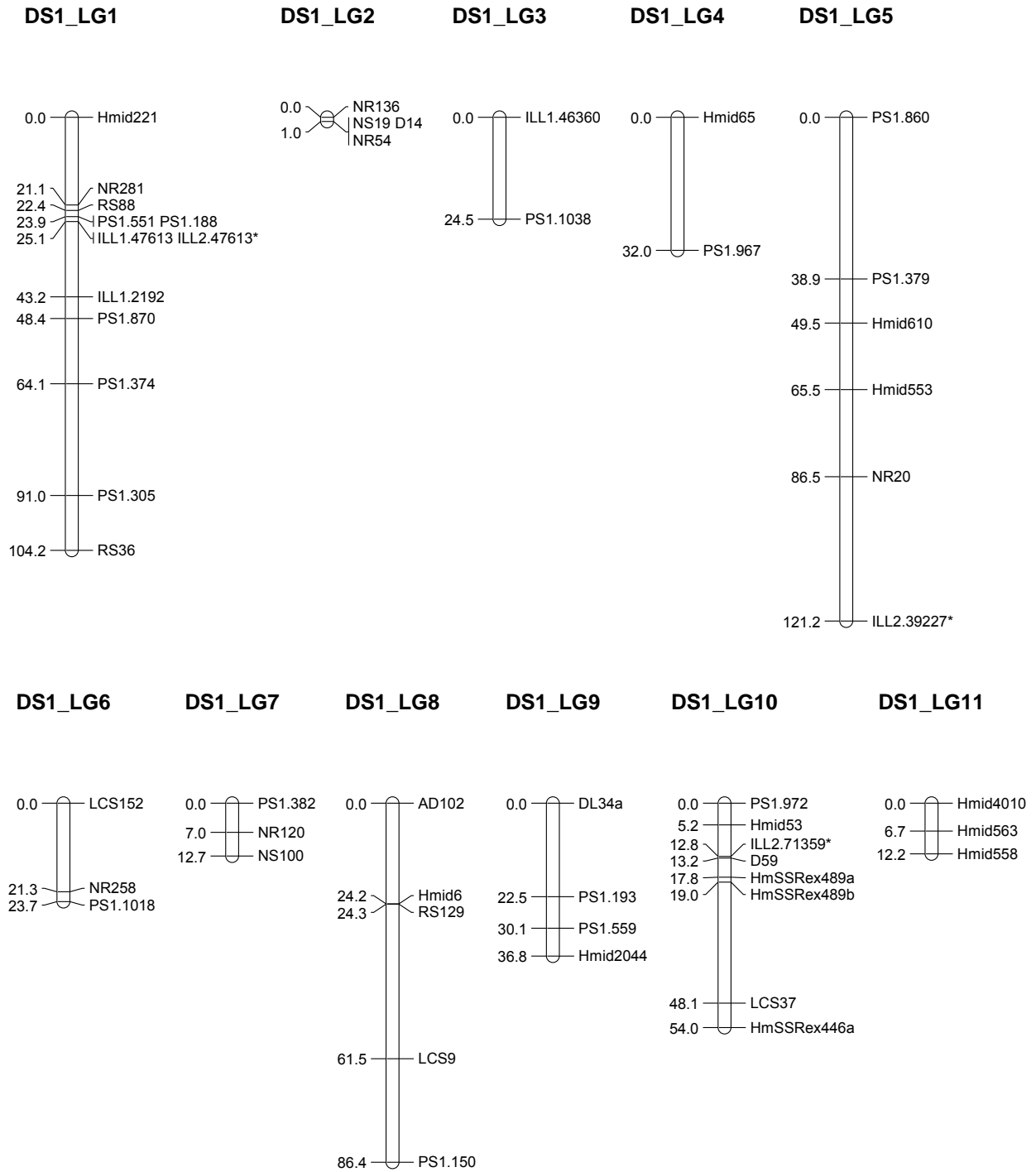
4.3.1 Linkage map of family DS1

4.3.1.1 Sex average linkage map of family DS1

From the 99 markers that were found to be informative in family DS1, from the farm Roman Bay, 74 (75%) could be mapped to 18 linkage groups of the sex-average map (Fig. 3.1). No SNPs could be mapped for this family (Table 3.10). The length of each linkage group ranged from 1.0cM to 121.2cM with an average marker spacing of 9.5cM. The number of markers per linkage group ranged from 2-12 (Table 3.6). The genome length calculated with equation A was 1114.4cM and that of equation B was 1246.32cM. The genome coverage (equation C) of the map was approximately 62%.

Four of the 99 markers could not reliably be grouped to a specific linkage group (LOD 3) and 21 of the grouped markers could not be mapped due to either insufficient linkage or Chi-square values higher than 1.0.

Chapter two-Type 1 microsatellite development



Chapter two-Type 1 microsatellite development

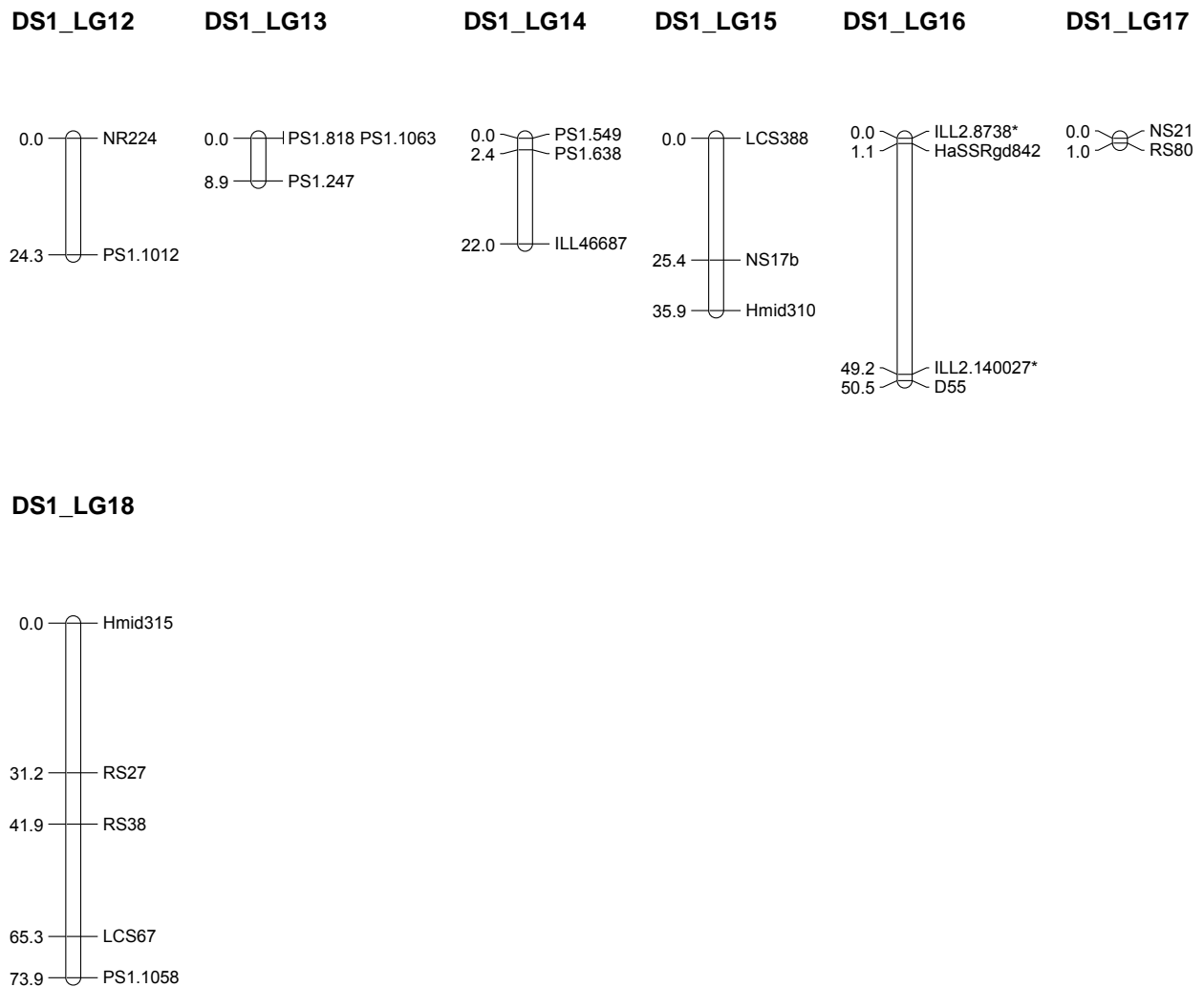


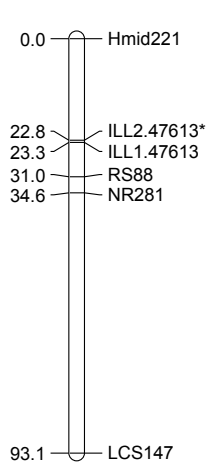
Figure 3.1: Sex-average map of family DS1 representing the 18 linkage groups. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in the current study.

4.3.1.2 Maternal map of DS1

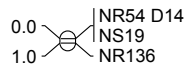
Fifty-nine of the 99 (60%) informative markers could be mapped to the maternal map of family DS1 to 18 linkage groups (Fig. 3.2). The linkage groups ranged in length from 1.0cM to 136.7cM with an average marker spacing of 13.0cM. The number of markers ranged from 2-7 per linkage group (Table 3.6). The expected genome coverage as calculated by equation A provided a genome length estimation of 1481.46cM and equation B, a genome length of 1483.7cM. The genome coverage of the map, as calculated by equation C, was approximately 51%.

Chapter two-Type 1 microsatellite development

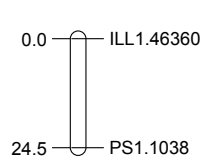
DS1P1_LG1



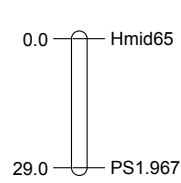
DS1P1_LG2



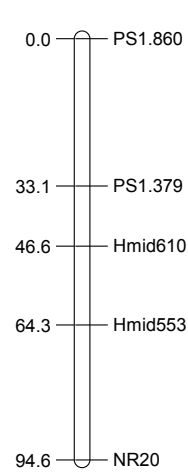
DS1P1_LG3



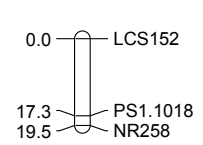
DS1P1_LG4



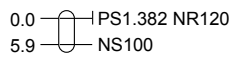
DS1P1_LG5



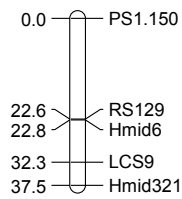
DS1P1_LG6



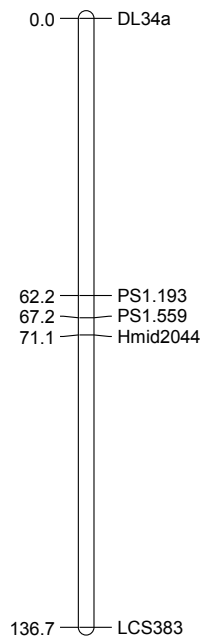
DS1P1_LG7



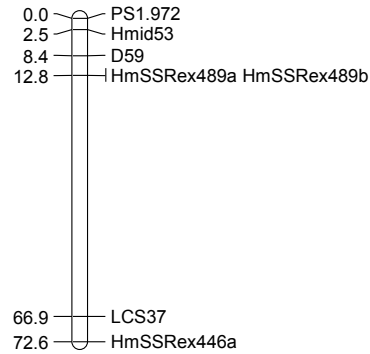
DS1P1_LG8



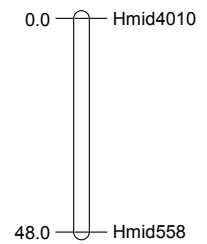
DS1P1_LG9



DS1P1_LG10



DS1P1_LG11



Chapter two-Type 1 microsatellite development

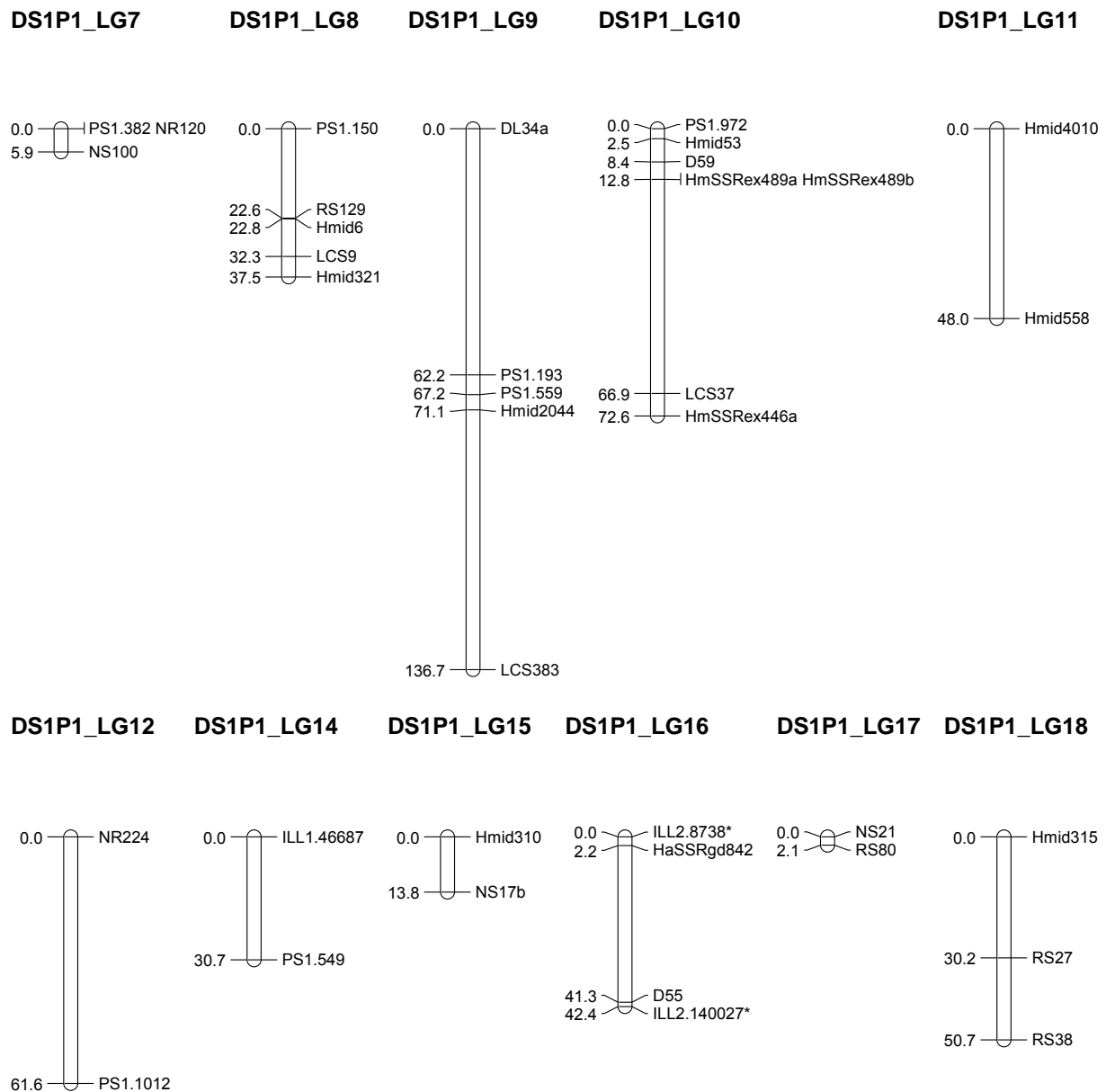


Figure 3.2: Maternal map of family DS1 showing the 18 linkage groups. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

4.3.1.3 Paternal map of family DS1

The paternal map of family DS1 consisted of 64 out of the 99 informative markers, which could be mapped to 18 linkage groups (Fig. 3.3). The length of the linkage groups ranged from 3.3cM to 93.1cM with an average marker spacing of 10.1cM, which is lower than in the maternal map. The number of markers per linkage group ranged from 2-12 (Table 3.6). The expected genome length of the map was calculated from equation A and B with values 1105.88cM and 1106.57cM, respectively. The genome coverage was approximately 56%.

Chapter two-Type 1 microsatellite development

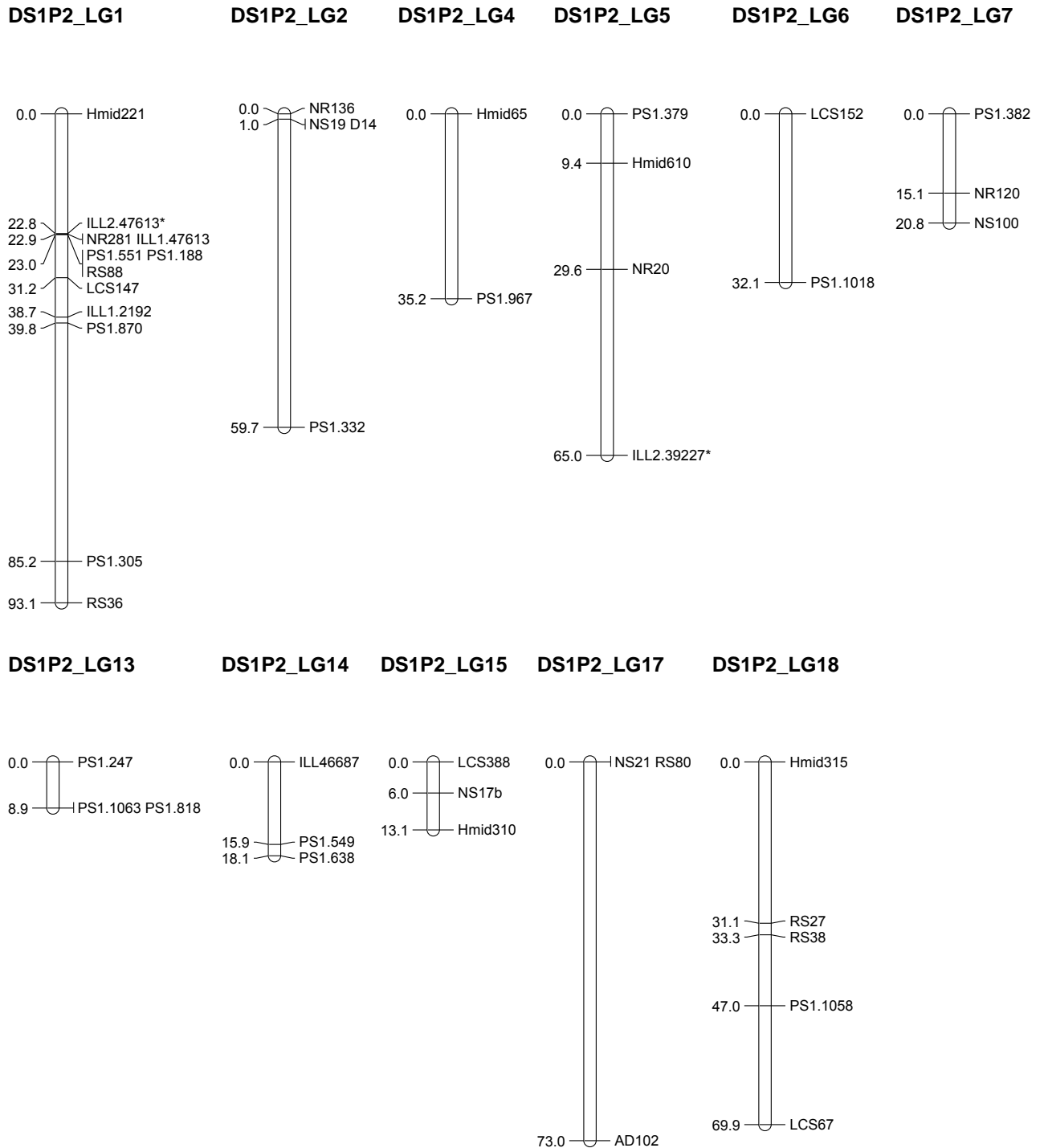


Figure 3.3: Paternal map of family DS1. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

Table 3.6: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS1.

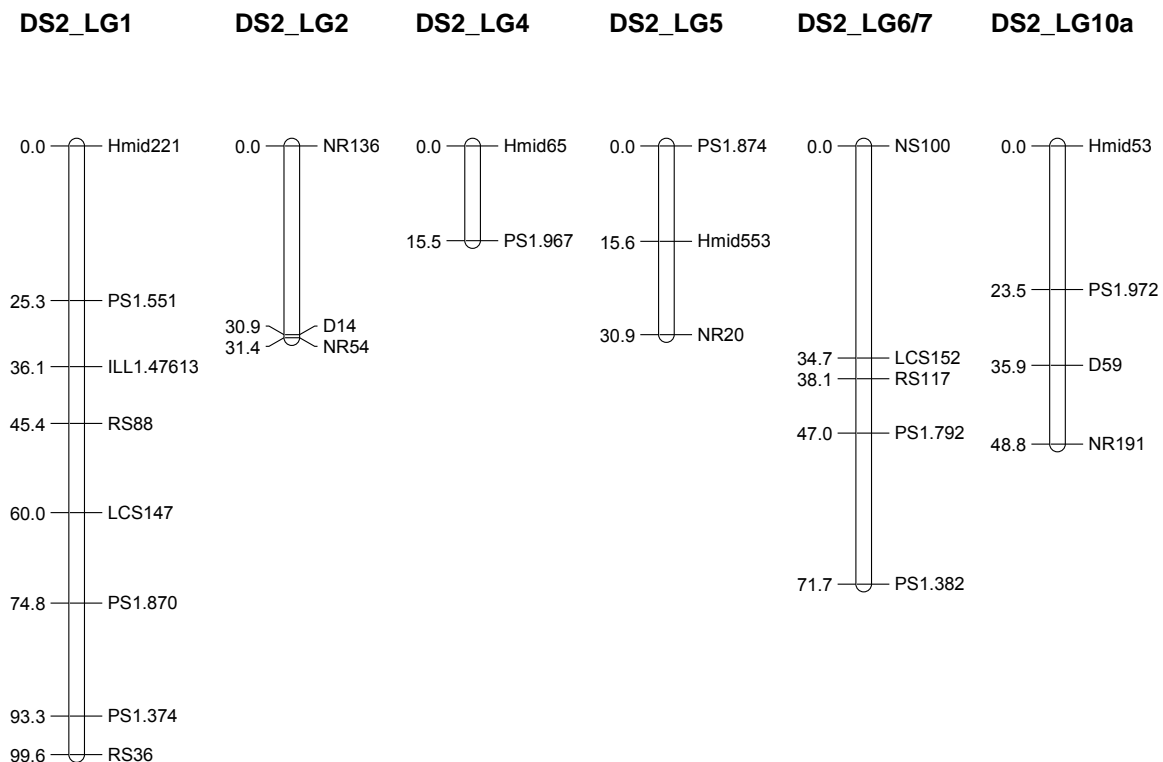
Sex-average map:					Maternal map:					Paternal map:				
Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)
1	12	104.2	8.7	26	1	6	91.3	15.2	58.5	1	12	93.1	7.8	45.4
2	4	1	0.3	1	2	4	1	0.3	1	2	4	59.7	14.9	58.7
3	2	24.5	12.3	24.5	3	2	24.5	12.3	24.5	4	2	35.2	17.6	35.2
4	2	32	16	32	4	2	29	14.5	29	5	4	65	16.3	35.4
5	6	121.2	20.2	38.9	5	5	94.6	19	33.1	6	2	32.1	16.1	32.1
6	3	23.7	7.9	21.3	6	3	19.5	6.5	17.3	7	3	20.8	6.9	15.1
7	3	12.7	4.2	7	7	3	5.9	2	5.9	8	5	56.6	11.3	34.8
8	5	86.4	17.3	37.2	8	5	37.5	7.5	22.6	9	3	29.6	9.9	25.2
9	4	36.8	9.2	22.5	9	5	136.7	27.3	65.6	10	7	39.9	5.7	15
10	8	54	6.8	29.1	10	7	72.6	10.4	54.1	11	3	3.3	1.1	2.1
11	3	12.2	4.1	6.7	11	2	48	24	48	12	2	3.2	1.6	3.2
12	2	24.3	12.2	24.3	12	2	61.6	30.8	61.6	13	3	8.9	3	8.9
13	3	8.9	3	8.9	14	2	30.7	15.4	30.7	14	3	18.1	6	15.9
14	3	22	7.3	19.6	15	2	13.8	6.9	13.8	15	3	13.1	4.4	7.1
15	3	35.9	12	25.4	16	4	42.4	10.6	39.1	17	3	73	24.3	73
16	4	50.5	12.6	48	17	2	2.1	1.1	2.1	18	5	69.9	14	31.1
17	2	1	0.5	1	18	3	50.7	16.9	30.2					
18	5	79.3	15.86	31.2										
Total	74	730.6	170.5	404.6		59	761.9	220.7	537.1		64	621.5	160.8	438.2
Average	4.1	40.6	9.5	22.5		3.5	44.8	13	31.6		4	38.8	10.1	27.4

4.3.2 Linkage map of family DS2

4.3.2.1 Sex average linkage map of family DS2

Family DS2, also originating from Roman Bay Sea Farm, had 81 markers that were informative. Fifty-four of these informative markers could be mapped to 17 linkage groups (Fig. 3.4). Genotyping of the SNPs in family DS2 failed and no SNPs were available for mapping (Table 3.10). The length of the linkage groups ranged from 0.5cM to 99.6cM with an average marker spacing of 11.3cM. The number of markers in each linkage group ranged from 2-8 (Table 3.7). The genome length as calculated by equation A was 1103.31cM and the genome length as calculated by equation B was 1151.34cM. The genome coverage of the map was calculated to be approximately 55%.

Four of the informative markers could not be grouped to a specific linkage group and a further 23 markers could not be mapped after being grouped to a specific linkage group.



Chapter two-Type 1 microsatellite development

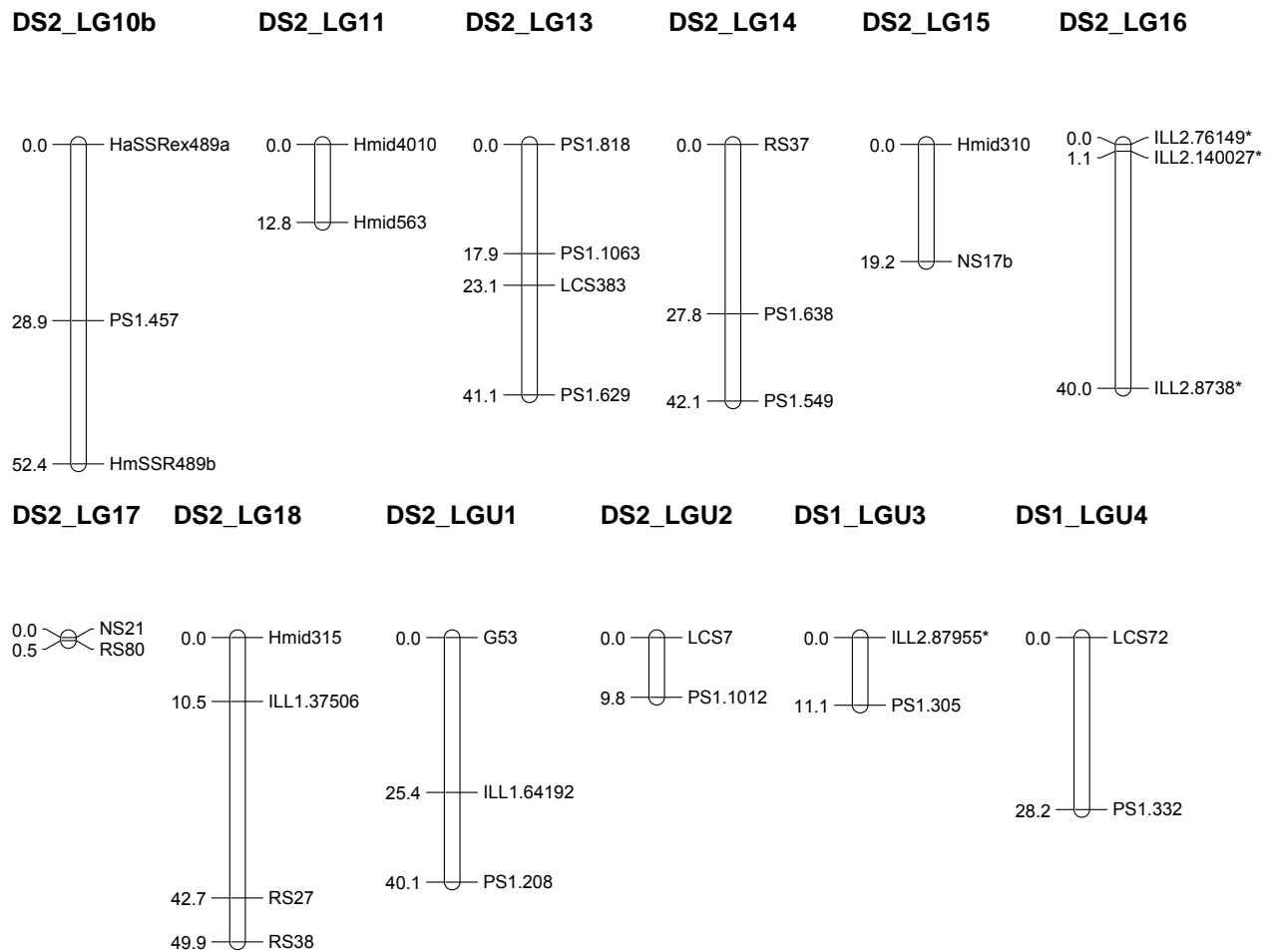


Figure 3.4: Sex-average map of family DS2. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

4.3.2.2 Maternal map of family DS2

The maternal map of family DS2 had 49 out of 81 (60.5%) markers, which could be mapped to 15 linkage groups (Fig. 3.5). The length of the linkage groups ranged from 8.2cM to 103.6cM with an average marker spacing of 14.6cM. The number of markers per linkage group ranged from 2-8 (Table 3.7). The genome length as calculated by equation A gave a result of 1292.12cM and equation B, a result of 1292.67cM. The genome coverage of the maternal map of family DS2 was approximately 53%.

Chapter two-Type 1 microsatellite development

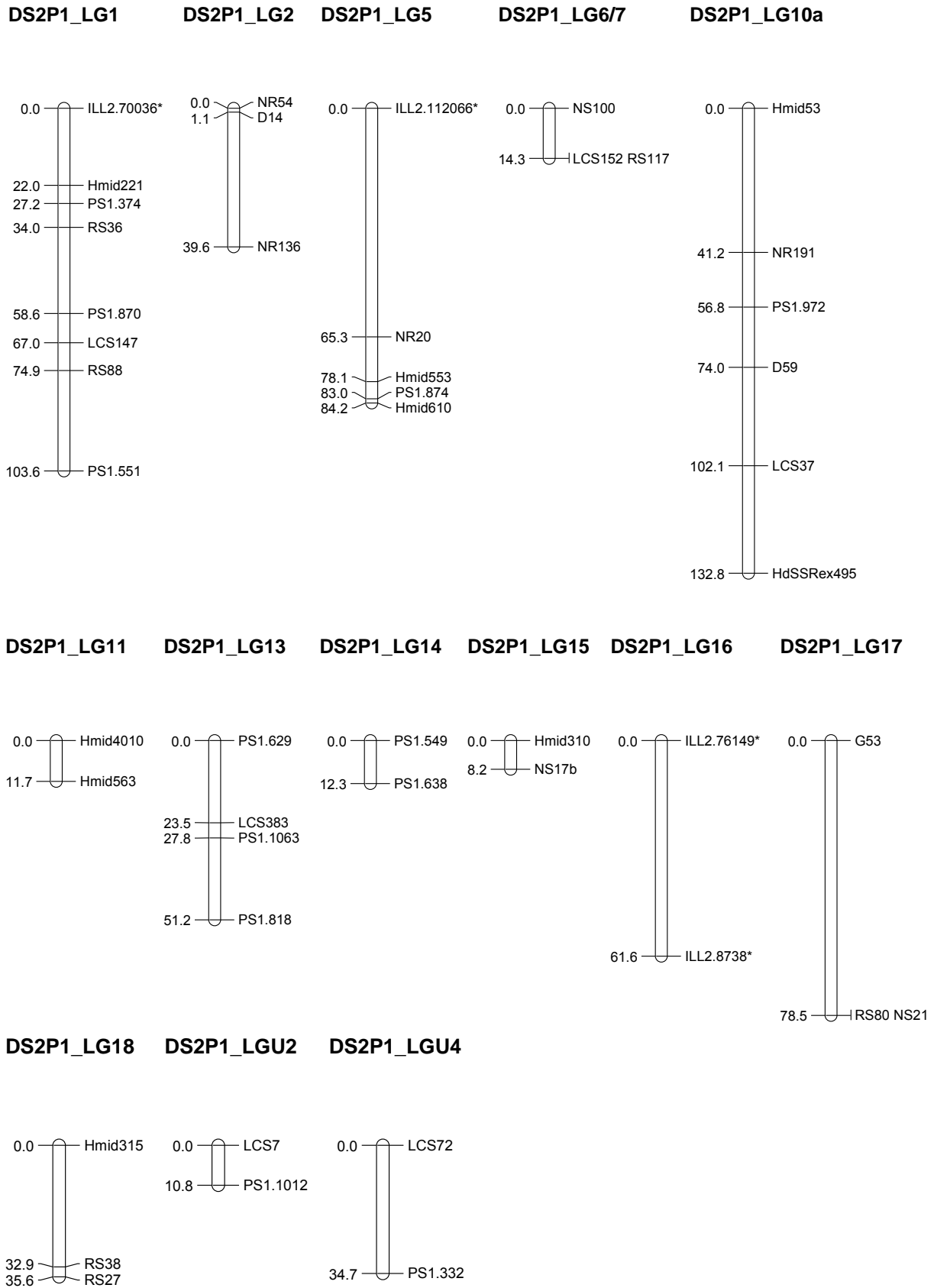
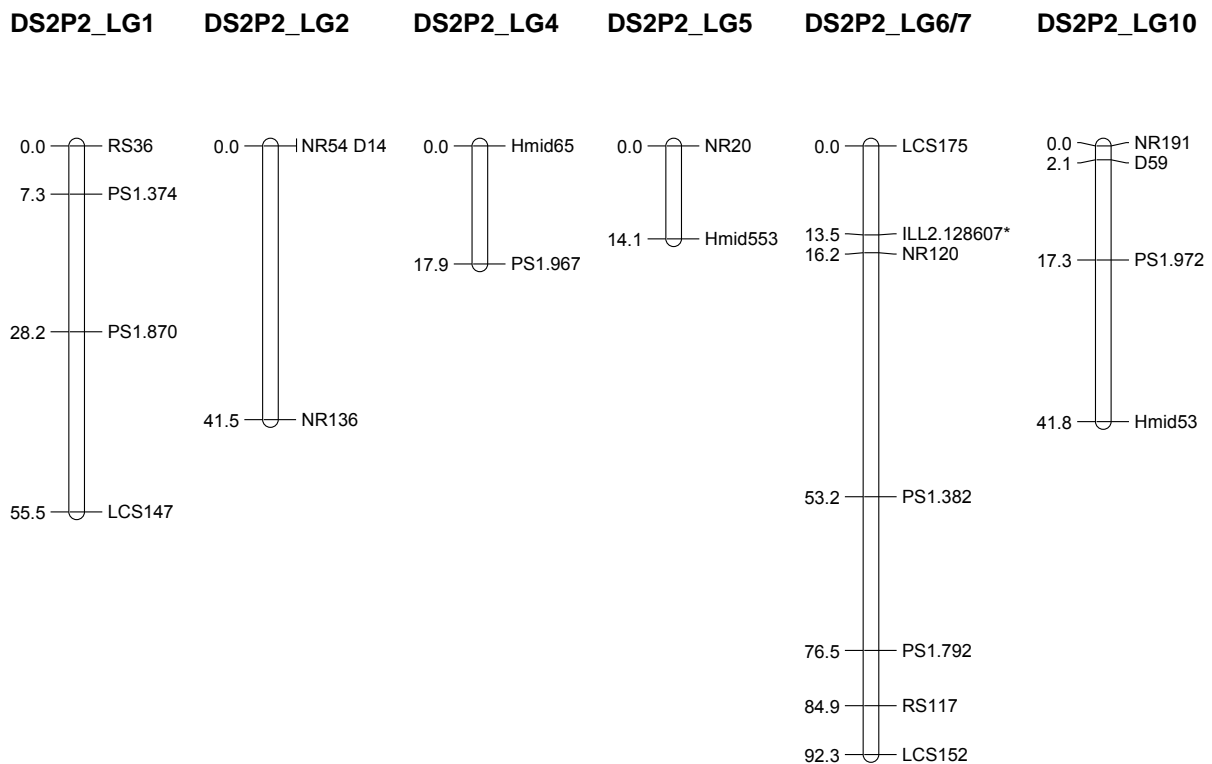


Figure 3.5: Maternal map of family DS2. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

Chapter two-Type 1 microsatellite development

4.3.2.3 Paternal map of family DS2

The paternal linkage map of family DS2 consisted of 51 out of the 81 (63%) informative markers that could be mapped to 17 linkage groups (Fig. 3.6). The length of the linkage groups ranged from 1cM to 92.3cM with an average marker spacing of 12.8cM, which is lower than the maternal map. The number of markers per linkage group ranged from 2-7 (Table 3.7). The genome length, as calculated by equation A, gave a value of 1373.61cM, whereas equation B resulted in a value of 1342.60cM. Genome coverage of the paternal map for family DS5 was approximately 50%.



Chapter two-Type 1 microsatellite development

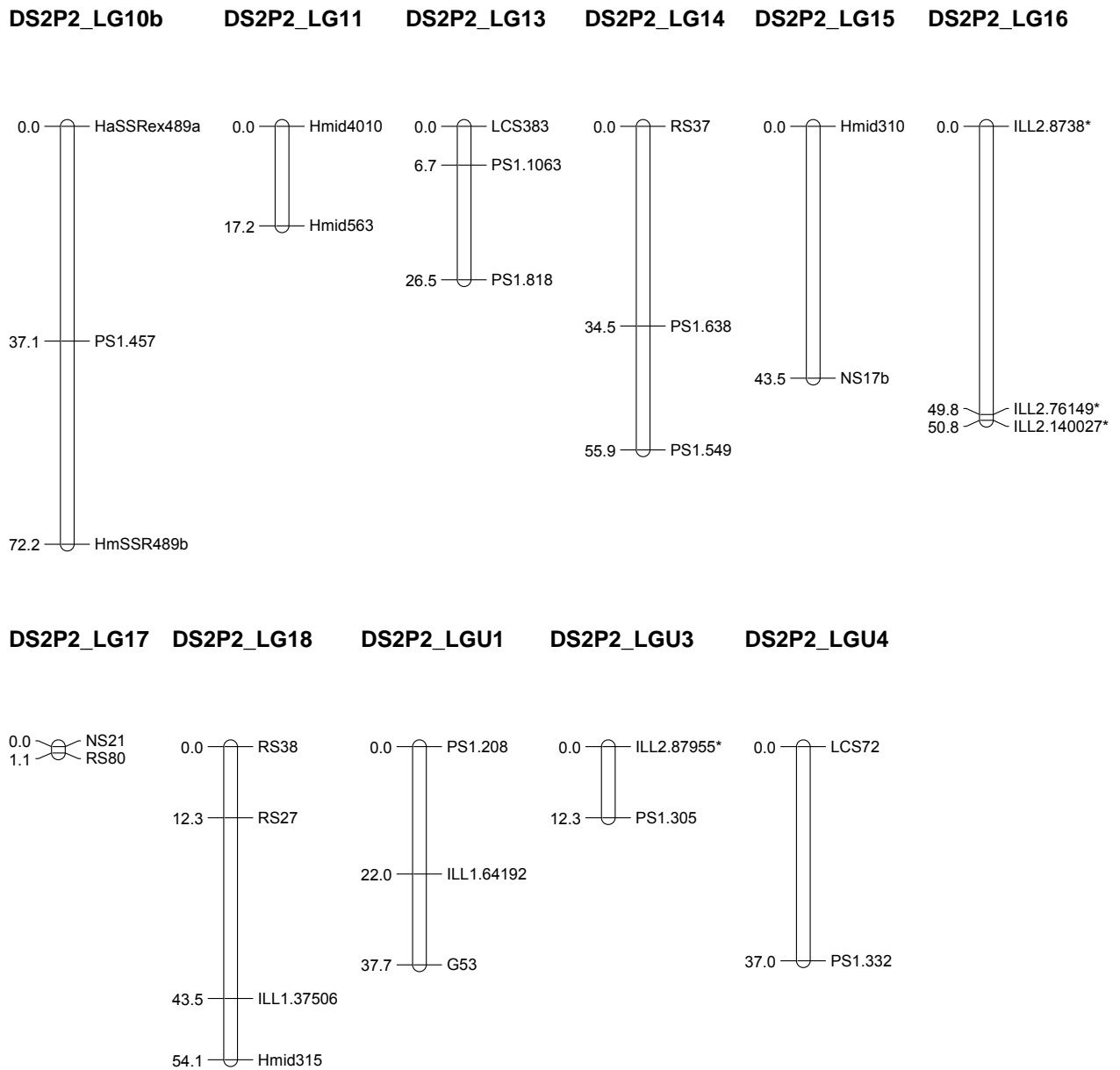


Figure 3.6: Paternal map of family DS2. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study

Table 3.7: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS2.

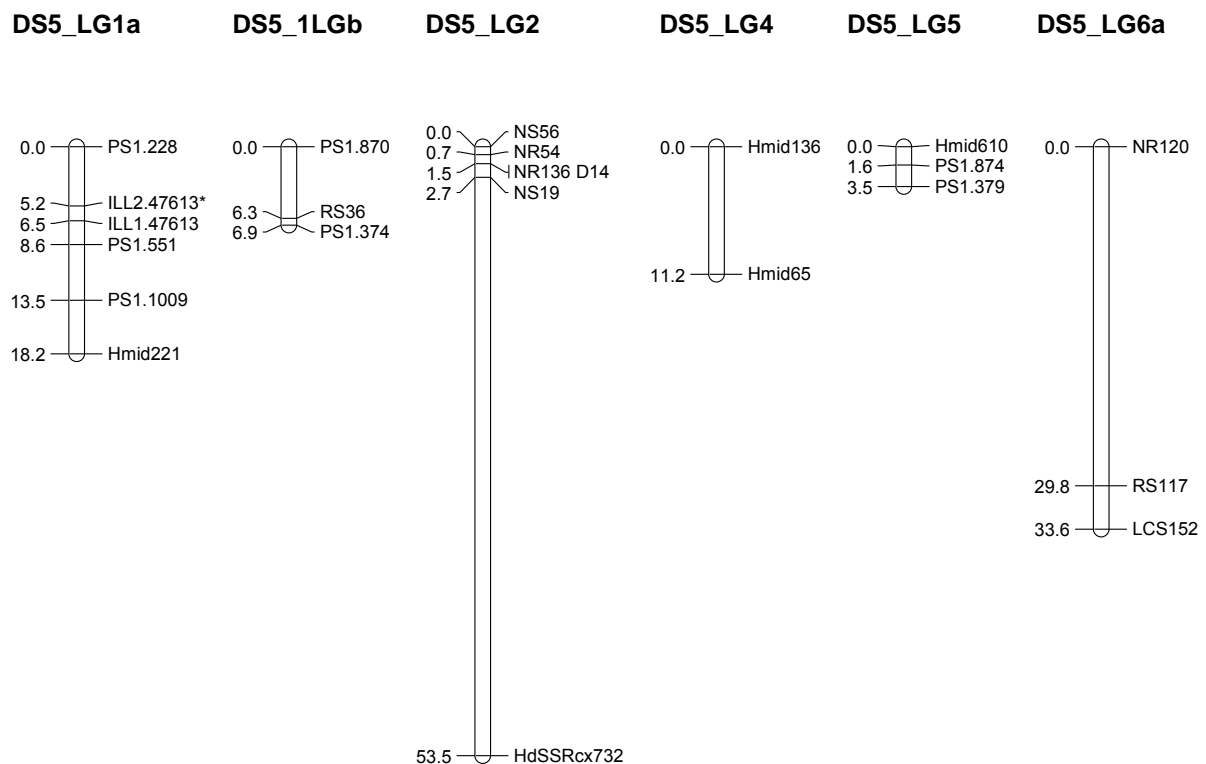
Sex-average map:					Maternal map:					Paternal map:				
Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)
1	8	99.6	12.45	25.3	1	8	103.6	13	28.7	1	4	55.5	13.9	27.3
2	3	31.4	10.5	30.9	2	3	39.6	13.2	38.5	2	3	41.5	13.8	41.5
4	2	15.5	7.75	15.5	5	5	84.2	16.9	65.3	4	2	17.9	9	17.9
5	3	30.9	10.3	15.6	6/7	3	14.4	4.8	14.3	5	2	14.1	7.05	14.1
6/7	5	71.7	14.34	34.7	10	6	132.8	22.1	41.2	6/7	7	92.3	13.2	37
10a	4	58.8	14.7	23.5	11	2	11.7	5.9	11.7	10a	4	41.8	10.5	24.5
10b	3	52.4	17.5	28.9	13	4	51.2	12.8	23.5	10b	3	72.2	24.1	37.1
11	2	12.8	6.4	12.8	14	2	12.3	6.2	12.3	11	2	17.2	8.6	17.2
13	4	41.1	10.3	18	15	2	8.2	4.1	8.2	13	3	26.5	8.8	19.8
14	3	42.1	14	27.8	16	2	61.6	30.8	61.6	14	3	55.9	18.6	34.5
15	2	19.2	9.6	19.2	17	2	78.5	39.3	78.5	15	2	43.5	21.8	43.5
17	2	0.5	0.25	0.5	18	3	35.6	11.9	32.9	16	3	50.8	16.9	49.8
18	4	49.9	12.5	32.2	U2	2	10.8	5.4	10.8	17	2	1	0.5	1.1
U1	3	40.1	13.4	25.4	U4	2	34.7	17.4	34.7	18	4	54.1	13.5	31.2
U2	2	9.8	4.9	4.9						U1	3	37.7	12.6	22
U3	2	11.1	5.55	11.1						U3	2	12.3	6.2	12.3
U4	2	28.2	14.1	28.8						U4	2	37	18.5	37
Total	54	615.1	178.4	355.1		46	679.2	203.5	462.2		51	671.3	217.4	467.8
Average	3.2	36.2	11.3	20.9		3.3	48.5	14.6	33		3	39.5	12.8	27.5

4.3.3 Linkage map of family DS5

4.3.3.1 Sex-average linkage map of family DS5

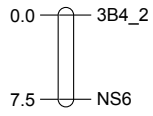
For family DS5 from the farm HIK, 52 out of 77 markers (68%) could be mapped to 18 linkage groups (Fig. 3.7). Of the 52 markers that could be mapped to family DS5's sex-average map, six were SNPs (Table 3.10). The length of the linkage groups ranged from 1.9cM to 53.5cM with an average marker spacing of 6.5cM. The number of markers per linkage group ranged from 2-6 (Table 3.8). The genome lengths, as calculated from equations A and B, were 707.18cM and 720.53cM, respectively with genome coverage of approximately 48%.

Family DS5 had 2 markers, which could not be mapped to a specific linkage group. Twenty-five of the informative markers that could be grouped to a linkage group (LOD 3) could not be mapped due to problems including insufficient linkage and Chi-square values higher than 1.0.



Chapter two-Type 1 microsatellite development

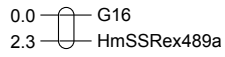
DS5_LG6b



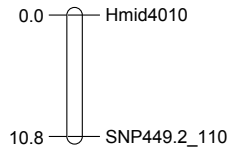
DS5_LG9



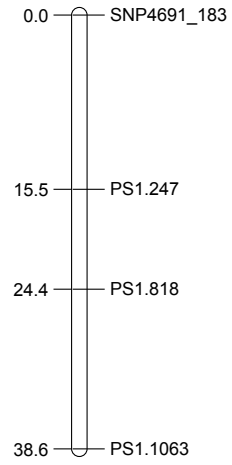
DS5_LG10



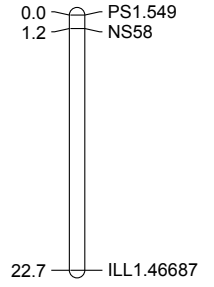
DS5_LG11



DS5_LG13



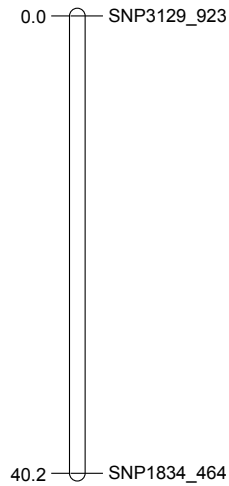
DS5_LG14



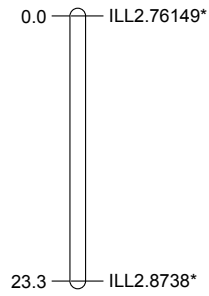
DS5_LG15



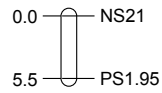
DS5_LG16a



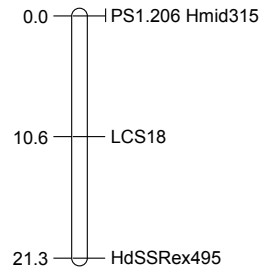
DS5_LG16b



DS5_LG17



DS5_LG18



Chapter two-Type 1 microsatellite development

DS5_LGU3

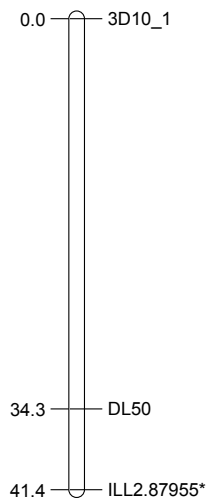


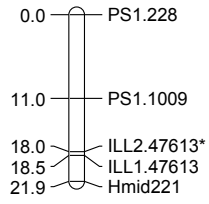
Figure 3.7: Sex-average map of family DS5. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

4.3.3.2 Maternal map of family DS5

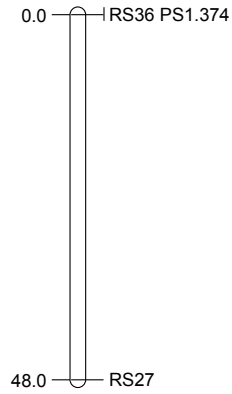
The maternal linkage map of family DS5 consisted of 44 mapped informative markers from the original 77 informative markers (Fig. 3.8). Two of the 44 mapped markers consisted of SNPs (Table 3.10). The length of the linkage groups ranged from 7.7cM to 80.0cM with an average marker spacing of 10.1cM. The number of markers per linkage group ranged from 2-5 (Table 3.8). Equation A resulted in an expected genome length of 1014.75cM and equation B, 970.45cM. The approximate genome coverage of family DS5 was calculated to be 47%.

Chapter two-Type 1 microsatellite development

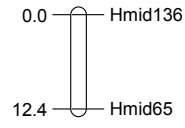
DS5P1_LG1a



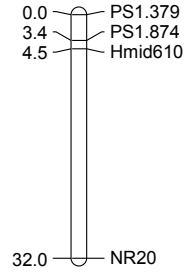
DS5P1_LG1b



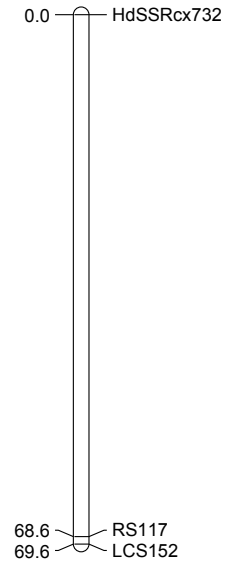
DS5P1_LG4



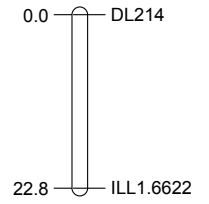
DS5P1_LG5



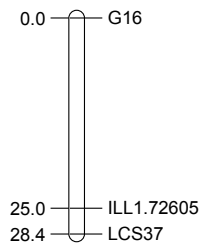
DS5P1_LG6a



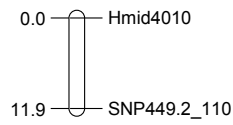
DS5P1_LG9



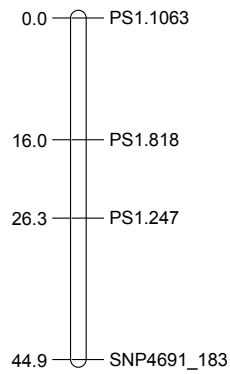
DS5P1_LG10



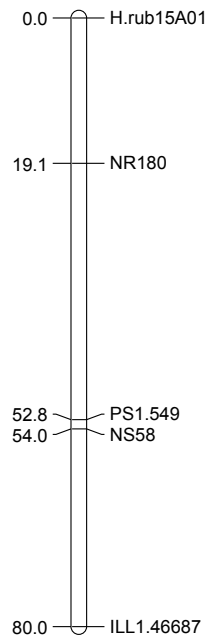
DS5P1_LG11



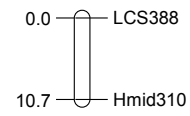
DS5P1_LG13



DS5P1_LG14



DS5P1_LG15



Chapter two-Type 1 microsatellite development

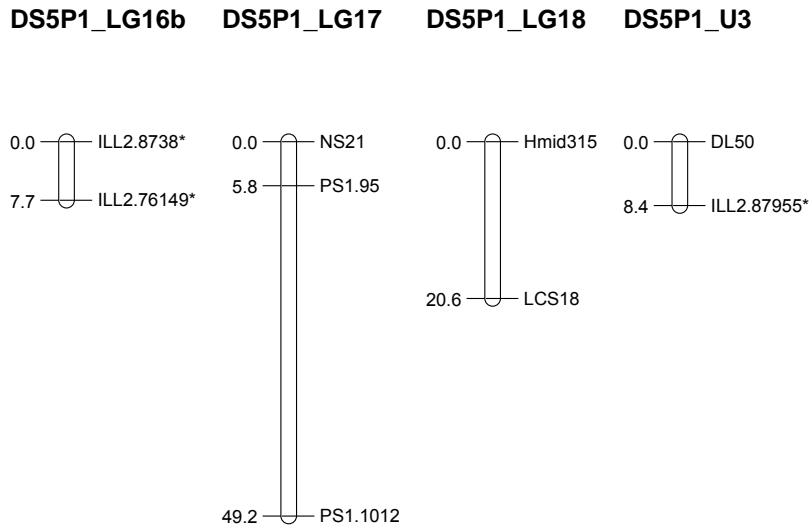


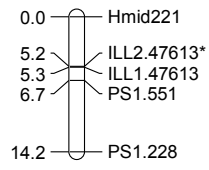
Figure 3.8: Maternal map of family DS5. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

4.3.3.3 Paternal map of family DS5

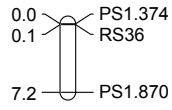
The paternal linkage map of family DS5 consisted of 37 out of the 77 (40%) informative markers that could be mapped to 13 linkage groups (Fig. 3.9). No SNPs could be mapped to family DS5's paternal map (Table 3.10). The length of the linkage groups ranged from 2cM to 50.7cM with an average marker spacing of 7.4cM. The number of markers per linkage group ranged from 2-5 (Table 3.8). Genome lengths of family DS5 were 584.03cM with equation A and 620.85cM with equation B. The genome coverage was calculated to be approximately 46%.

Chapter two-Type 1 microsatellite development

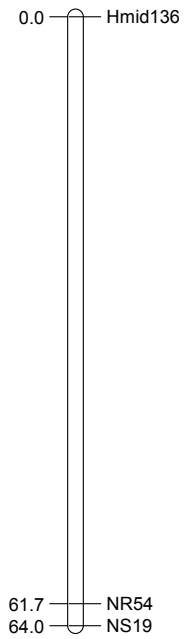
DS5P2_LG1a



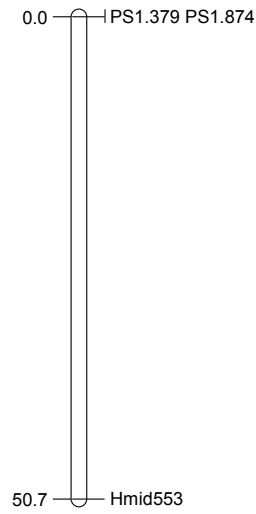
DS5P2_LG1b



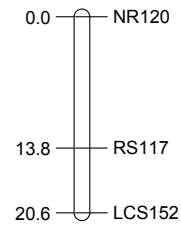
DS5P2_LG2



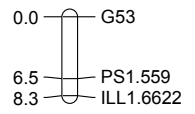
DS5P2_LG5



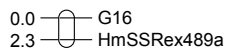
DS5P2_LG6a



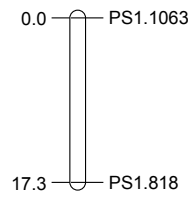
DS5P2_LG9



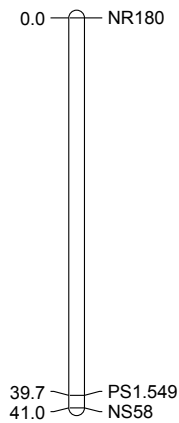
DS5P2_LG10



DS5P2_LG13



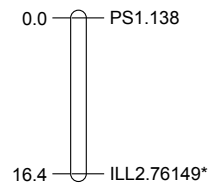
DS5P2_LG14



DS5P2_LG15

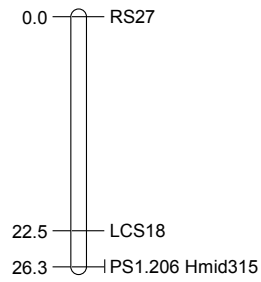


DS5P2_LG16a



Chapter two-Type 1 microsatellite development

DS5P2_LG18



DS5P2_U3

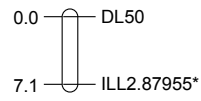


Figure 3.9: Paternal map of family DS5. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

Table 3.8: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS5.

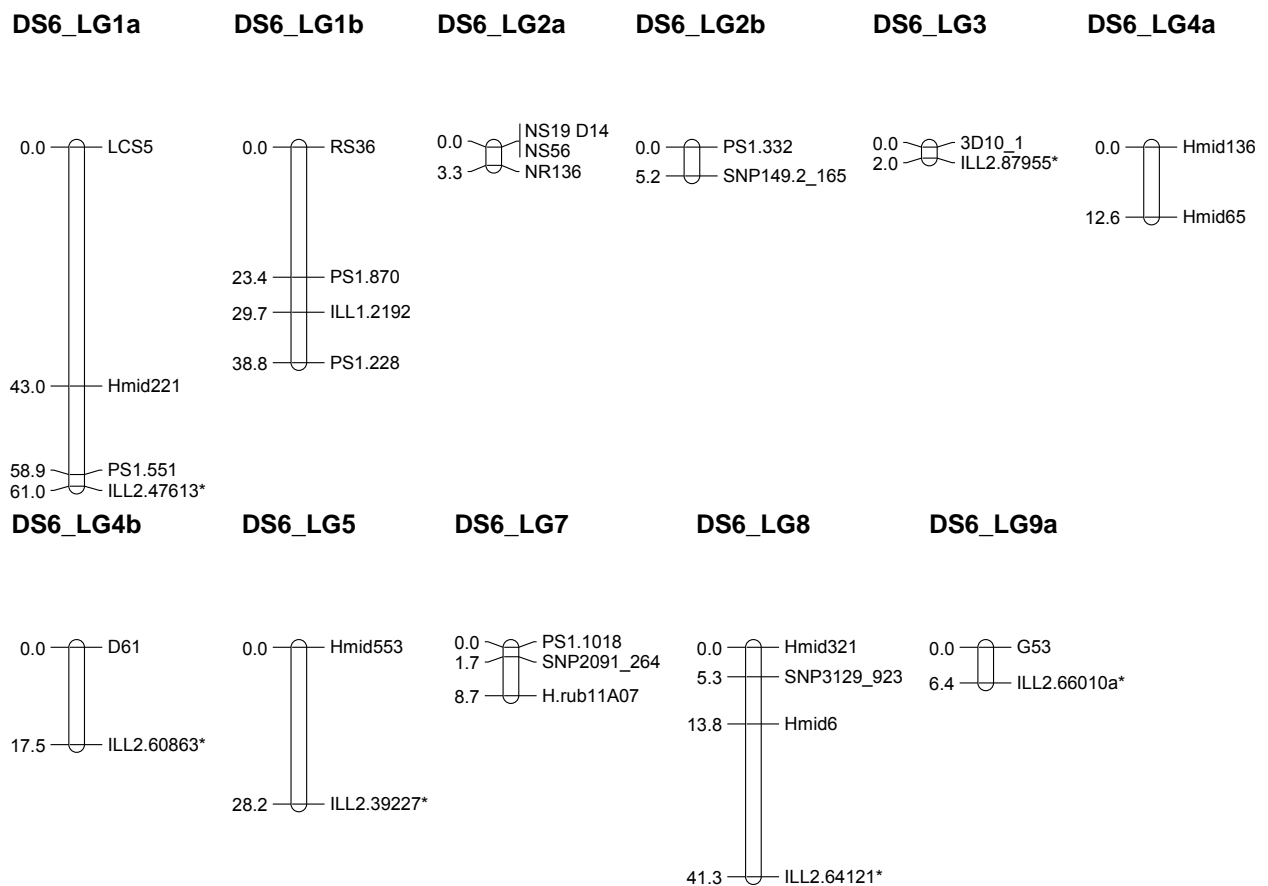
Sex-average map:					Maternal map:					Paternal map:				
Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)
1a	6	18.2	3	5.2	1a	5	21.9	4.4	11	1a	5	14.2	2.8	7.5
1b	3	6.9	2.3	6.3	1b	3	48	16	48	1b	3	7.2	2.4	7.1
2	6	53.5	8.9	50.8	4	2	12.4	6.2	12.4	2	3	64	21.3	61.7
4	2	11.2	5.6	11.2	5	4	32	8	27.5	5	3	50.7	16.9	50.7
5	3	3.5	1.2	1.9	6a	3	69.6	23.2	68.6	6a	3	20.6	6.9	13.8
6a	3	33.6	11.2	29.8	9	2	22.8	11.4	22.8	9	3	8.3	2.8	6.5
6b	2	7.5	3.8	7.5	10	3	28.4	9.5	25	10	2	2.3	1.2	2.3
9	2	2.2	1.1	2.2	11	2	11.9	6	11.9	13	2	17.3	8.7	17.3
10	2	2.3	1.2	2.3	13	4	44.9	11.2	18.6	14	3	41	13.7	39.7
11	2	10.8	5.4	10.8	14	5	80	16	33.7	15	2	2	1	2
13	4	38.6	9.7	15.5	15	2	10.7	5.4	10.7	16	2	16.4	8.2	16.4
14	3	22.7	7.6	21.5	16b	2	7.7	3.9	7.7	18	4	26.3	6.6	22.5
15	2	1.9	0.95	1.9	17	3	49.2	16.4	43.4	U3	2	7.1	3.6	7.1
16a	2	40.2	20.1	40.2	18	2	20.6	10.3	20.6					
16b	2	23.3	11.65	23.3	U3	2	8.4	4.2	8.4					
17	2	5.5	2.75	5.5										
18	3	21.3	7.1	11										
U3	3	41.4	13.8	34.3										
Total	52	344.6	117.4	281.2		44	468.5	151.9	370.3		37	277.4	95.9	254.6
Average	2.9	19.1	6.5	15.6		2.9	31.2	10.1	24.7		2.8	21.3	7.4	19.6

4.3.4 Linkage map of family DS6

4.3.4.1 Sex-average map of family DS6

Family DS6, also originating from HIK, had 58 out of the 71 (82%) informative genic markers that were mapped to the sex-average linkage map (Fig. 3.10). Seven of these mapped markers were SNPs (Table 3.10). The lengths of the linkage groups ranged from 1.2cM to 109.5cM with an average marker spacing of 6.48cM. On average, three markers occurred on each linkage group (Table 3.9). The expected genome length, as calculated by equation A, resulted in a genome length of 818.67cM and for equation B, the result was 844.42cM. Using equation C, the genome coverage for DS6's sex-average map was 49%.

One of the informative markers could not be grouped into a specific linkage group (LOD 3) and 12 of the grouped informative markers could not be mapped in their specific linkage groups.



Chapter two-Type 1 microsatellite development

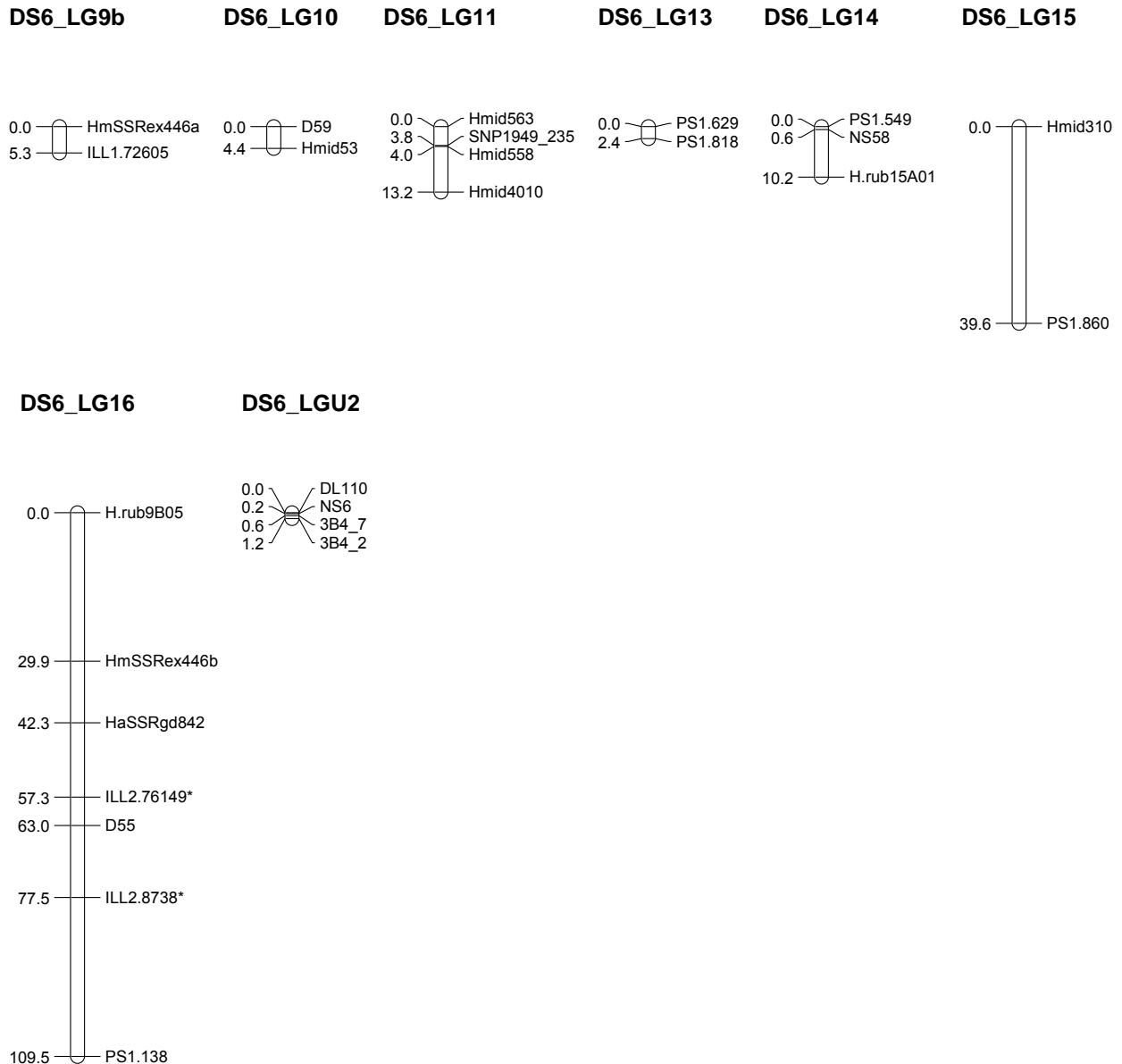


Figure 3.10: Sex-average map of family DS6. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

4.3.4.2 Maternal map of family DS6

The maternal map of family DS6 had 43 markers that could be mapped to 14 linkage groups with approximately three markers occurring on each linkage group (Fig. 3.11). The maternal map of family DS6 contained three SNP markers (Table 3.10). The lengths of the linkage groups ranged from 2.5cM to 169cM with an average marker spacing of 14.95cM (Table 3.9). The expected genome length calculated by equation A was 1380.51cM and for equation B, it was 1385.9cM. This accounts for a genome coverage of approximately 51%.

Chapter two-Type 1 microsatellite development

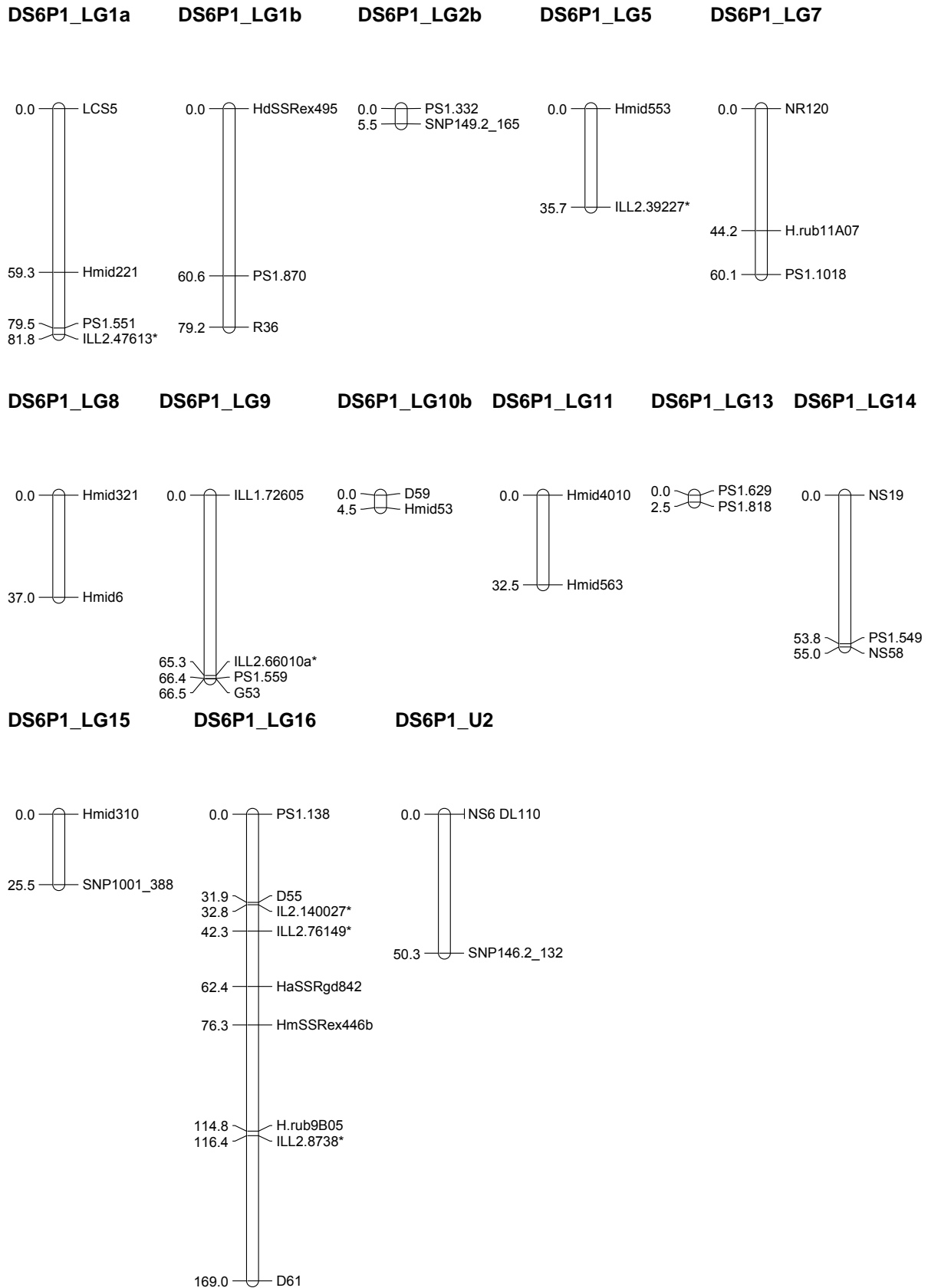
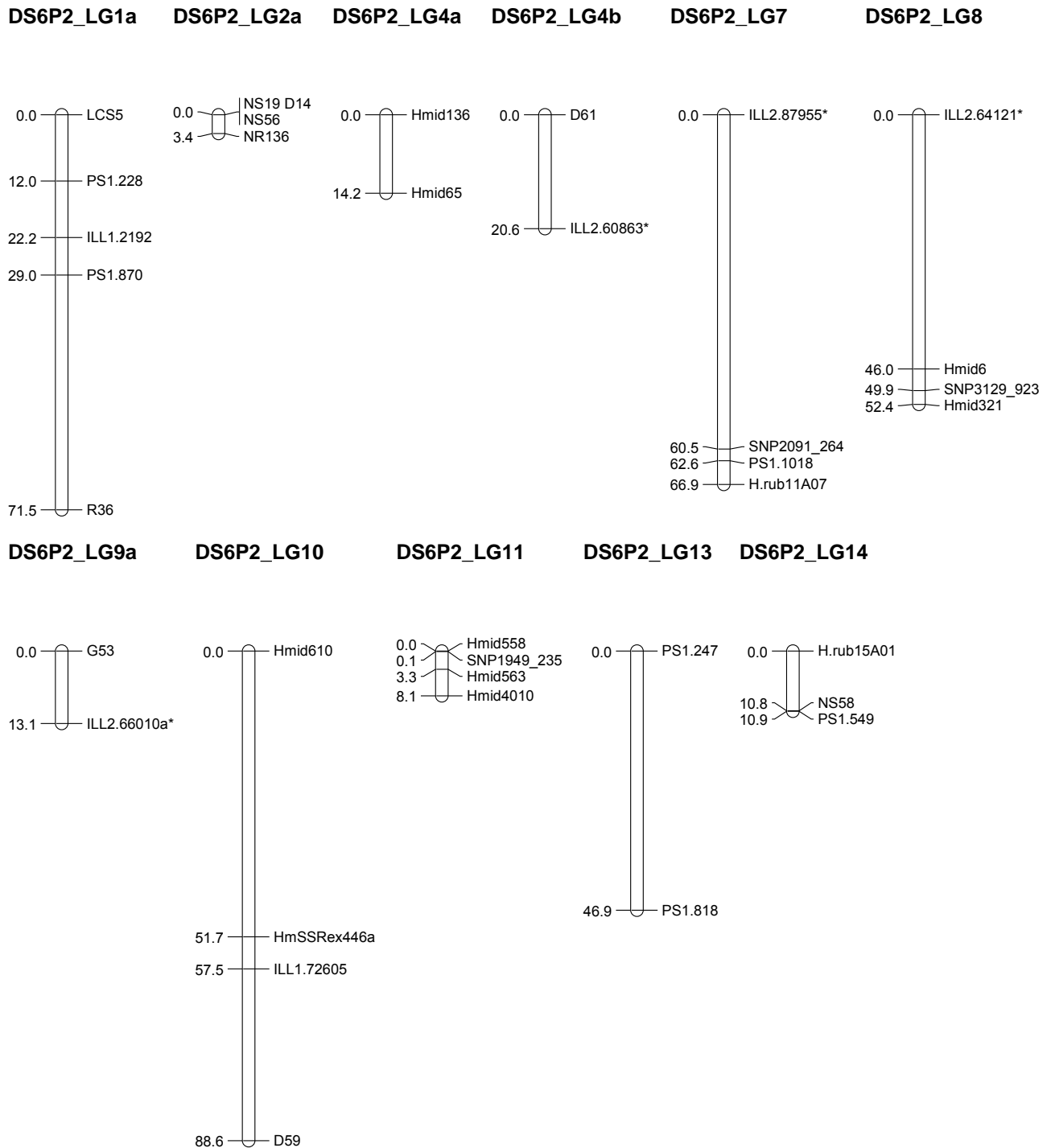


Figure 3.11: Maternal map of family DS6. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

Chapter two-Type 1 microsatellite development

4.3.4.3 Paternal map of family DS6

Forty-four markers could be mapped to 14 linkage groups on DS6's paternal map. On average, three markers occurred on each linkage group (Fig.3.12). DS6's paternal map had five SNPs that could be mapped (Table 3.10). The lengths of the linkage groups ranged from 3cM to 88.6cM with an average marker spacing of 10.88cM (Table 3.9). Equation A for genome length resulted in a value of 971.82cM and 932.06cM for equation B. Equation C resulted in 51% genome coverage of the map.



Chapter two-Type 1 microsatellite development

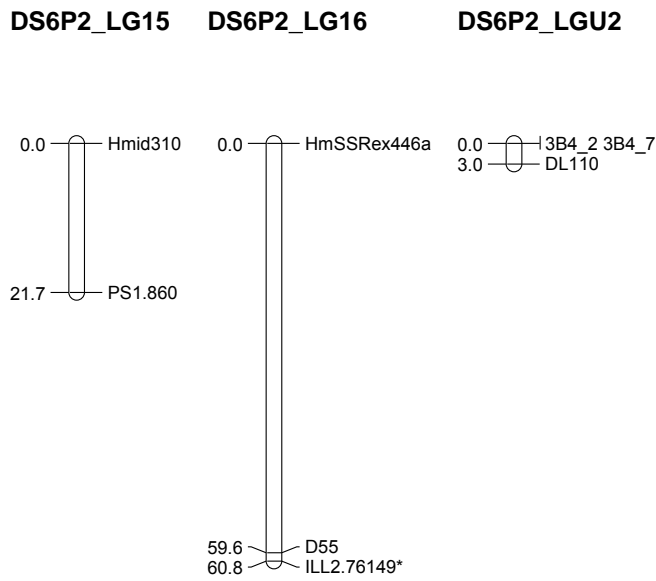


Figure 3.12: paternal map of family DS6. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centiMorgan; cM) of the markers are shown on the left. * Markers developed in this study.

Table 3.9: Number of markers per linkage group, their corresponding lengths, average markers spacing and largest interval for the sex-average, maternal and paternal maps of family DS6.

Sex-average map:					Maternal map:					Paternal map:				
Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)	Linkage groups:	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)
1a	4	61	15.3	43	1a	4	81.8	20.45	59.3	1	5	71.5	14.3	42.5
1b	4	38.8	9.7	24.4	1b	3	79.2	26.4	60.6	2a	4	3.4	0.85	3.4
2a	4	3.3	0.8	3.3	2b	2	5.5	2.8	5.5	4a	2	14.2	7.1	14.2
2b	2	5.2	2.6	5.2	5	2	35.7	17.9	35.7	4b	2	20.6	10.3	20.6
3	3	2	0.6	2	7	3	60.1	20	44.2	7	4	66.9	16.73	60.5
4a	2	12.6	6.3	12.6	8	2	37	18.5	37	8	4	52.4	13.1	46
4b	2	17.5	8.75	17.5	9	4	66.5	16.6	65.3	9a	2	13.1	6.55	13.1
5	2	28.2	14.1	28.2	10	2	4.5	2.3	4.5	10	4	88.6	22.15	51.7
7	3	8.7	2.9	7	11	2	32.5	16.3	32.5	11	4	8.1	2.03	4.8
8	4	41.3	10.3	27.5	13	2	2.5	1.3	2.5	13	2	46.9	23.45	46.9
9a	2	6.4	3.2	6.4	14	3	55	18.3	53.8	14	3	10.9	3.6	10.8
9b	2	5.3	2.7	5.3	15	2	25.5	12.8	25.5	15	2	21.7	10.85	21.7
10	2	4.4	2.2	4.4	16	9	169	18.8	52.6	16	3	60.8	20.3	59.6
11	4	13.2	3.3	9.2	U2	3	50.3	16.8	50.3	U2	3	3	1	3
13	2	2.4	1.2	2.4										
14	3	10.2	3.4	9.6										
15	2	39.6	19.8	39.6										
16	7	109.5	15.6	32										
U2	4	1.2	0.3	0.6										
Total	58	410.8	123.05	280.2		43	705.1	209.25	529.3		44	482.1	152.31	398.8
Average	3	21.62	6.48	14.75		3	50.36	14.95	37.81		3	34.44	10.88	28.49

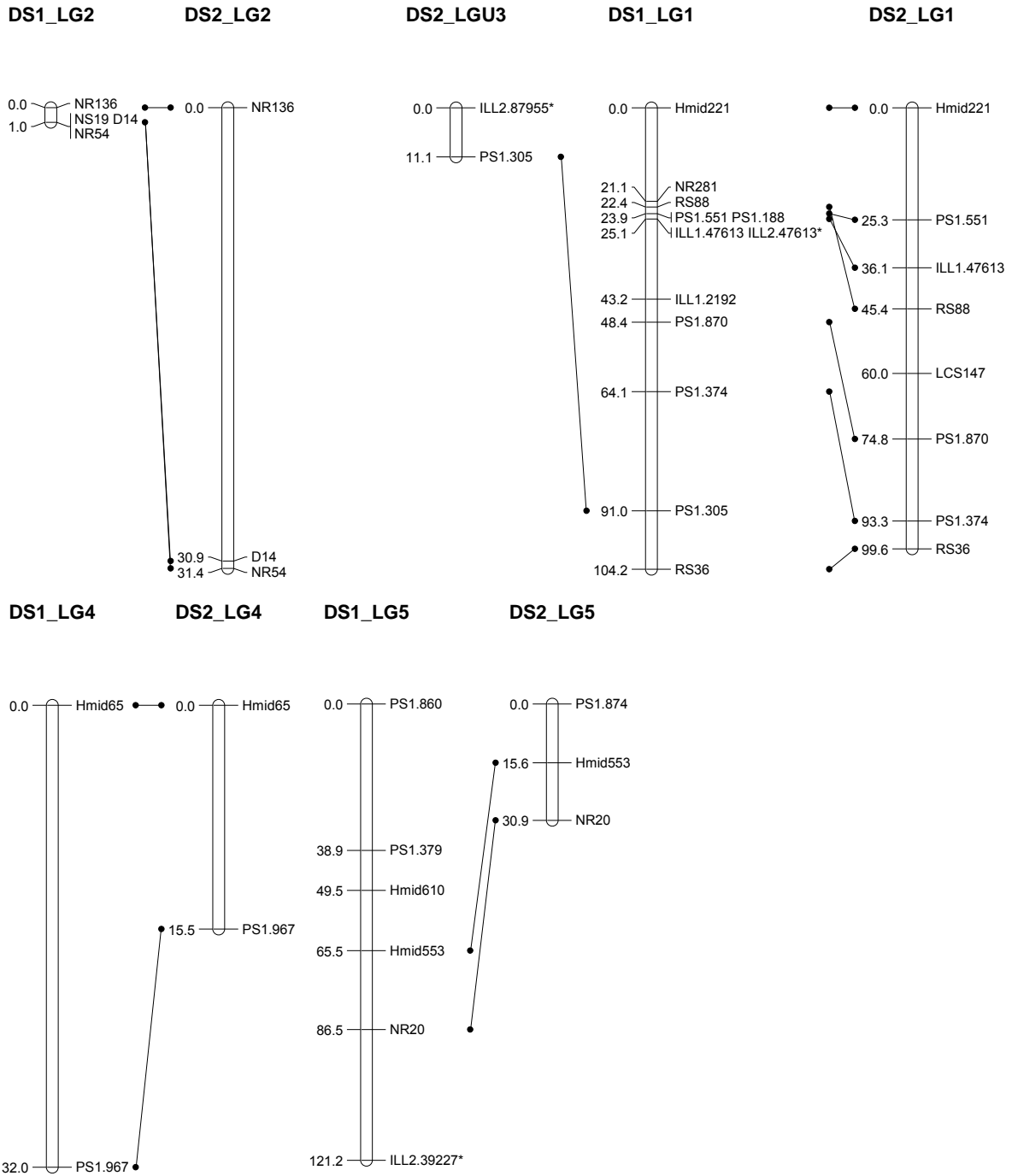
Table 3.10: Total number of informative microsatellite and SNP markers as well as, number of mapped microsatellite and SNP markers in the sex-average, maternal and paternal maps for each family.

	Total nr. informative markers		Nr. mapped markers in sex-average map		Nr. mapped markers in maternal map		Nr. mapped markers in paternal map	
	Microsatellite	SNP	Microsatellite	SNP	Microsatellite	SNP	Microsatellite	SNP
DS1	88	11	74	0	59	0	64	0
DS2	81	0	54	0	49	0	51	0
DS5	67	10	47	6	42	2	37	0
DS6	60	11	51	7	40	3	39	5

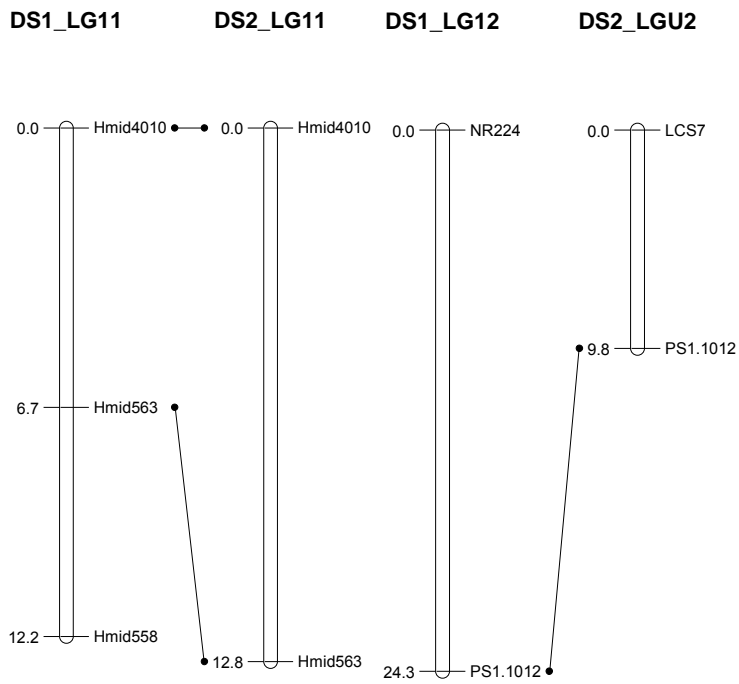
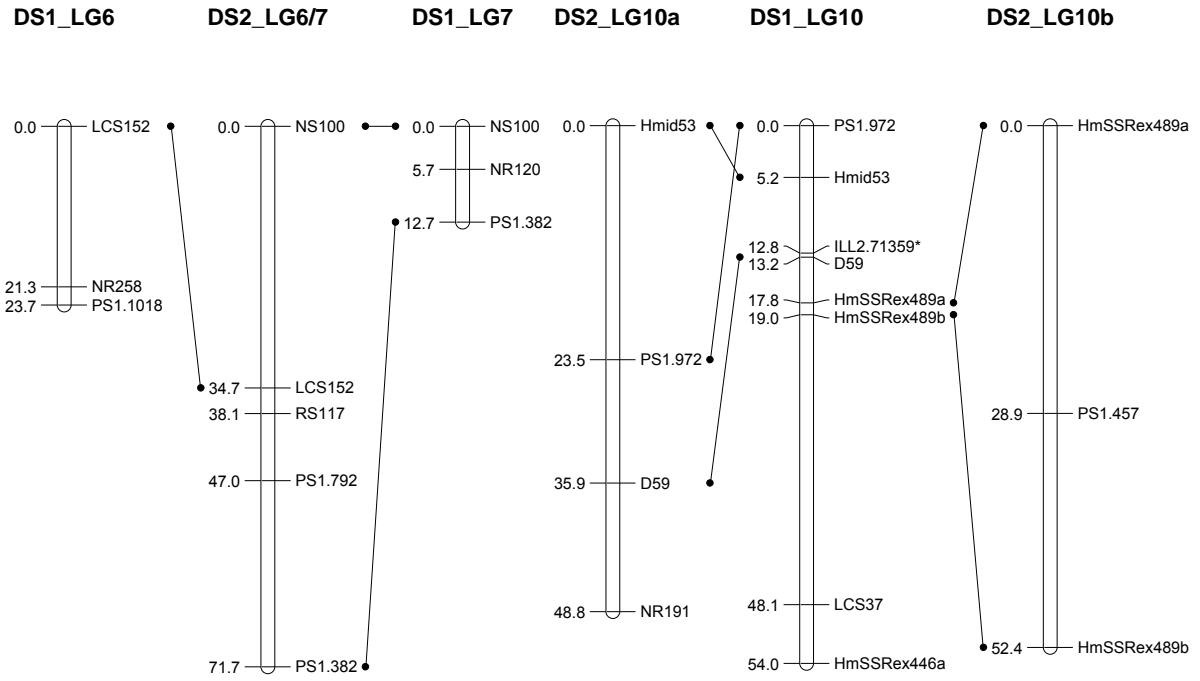
4.3.5 Sex-average linkage group comparisons

Thirty-eight links could be made between family DS1 and DS2. Fifteen of family DS1's linkage groups were linked to 16 of family DS2's linkage groups with at least one concurrent marker. Marker order placement was not always maintained. For example, for marker *RS88* the placement on DS1_LG1 was before marker *PS1.551*, whereas on DS2_LG1 it was after marker *PS1.551* (Fig. 3.13). For linkage group DS2_LG10a, the order of markers *Hmid35* and *PS1.972* was inverted to that found on DS1_LG10. Some linkage groups showed homology to more than one group such as DS1_LG10 to DS2_LG10a and DS2_LG10b (Fig. 3.13). Two of DS2's unknown linkage groups (DS2_LGU3 and DS2_LGU2) showed homology to linkage groups of DS1. DS2_LGU3 shared one marker, *PS1.305*, with DS1_LG1 and DS2_U2 shared one marker (*PS1.1013*).

Chapter two-Type 1 microsatellite development

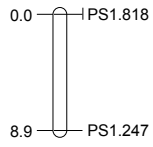


Chapter two-Type 1 microsatellite development

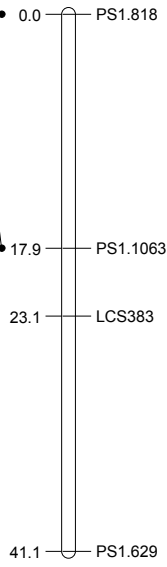


Chapter two-Type 1 microsatellite development

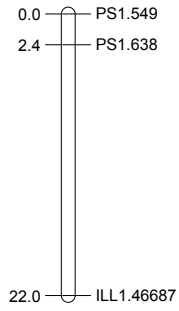
DS1_LG13



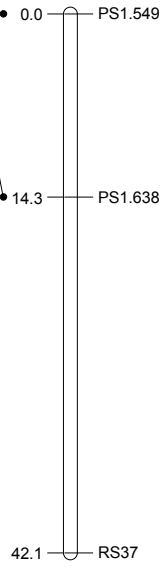
DS2_LG13



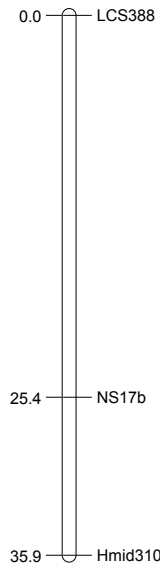
DS1_LG14



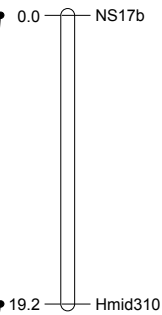
DS2_LG14



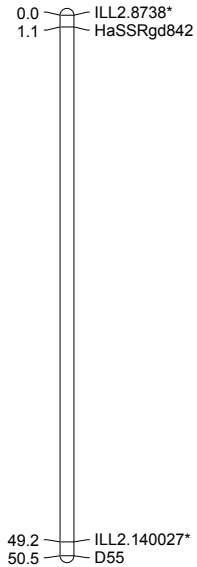
DS1_LG15



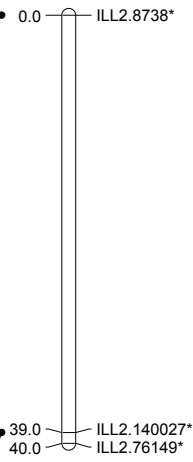
DS2_LG15



DS1_LG16



DS2_LG16



Chapter two-Type 1 microsatellite development

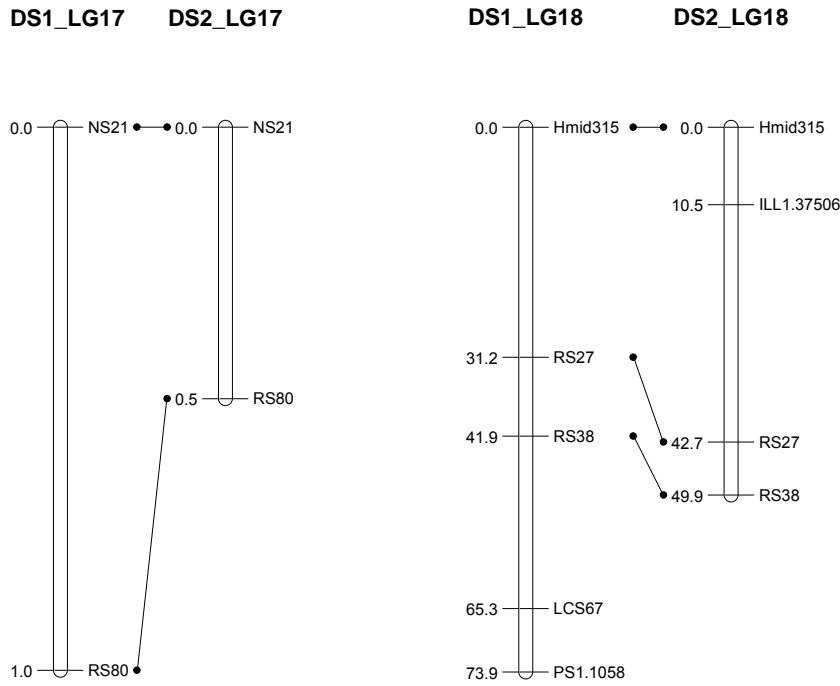
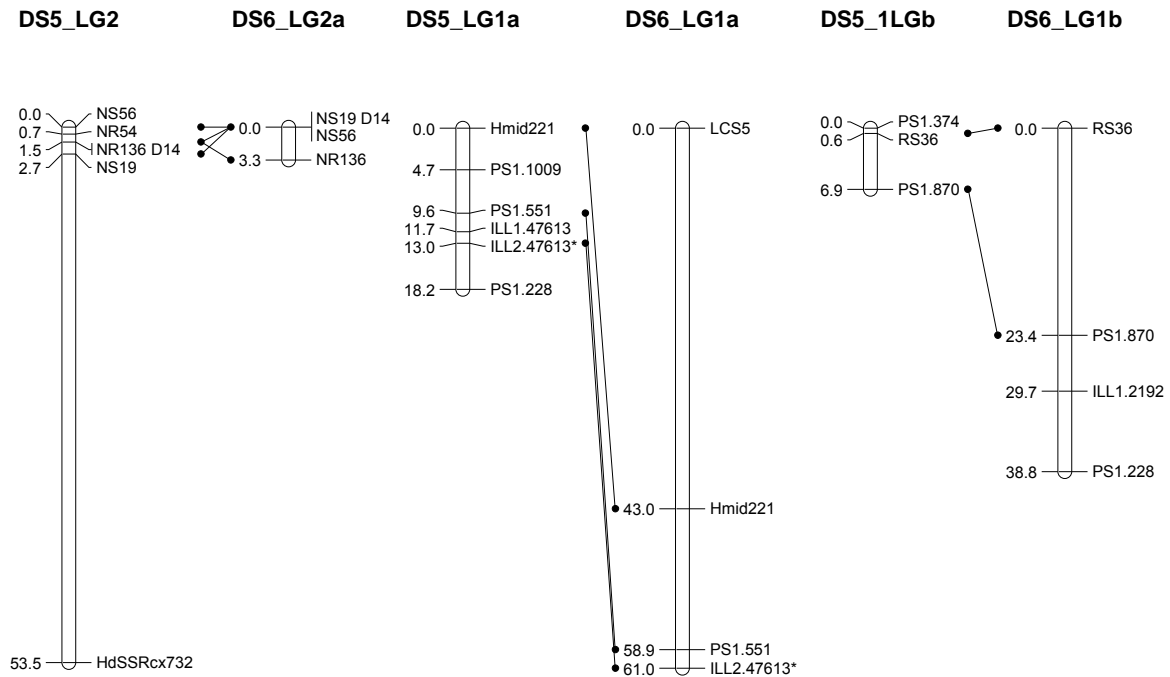
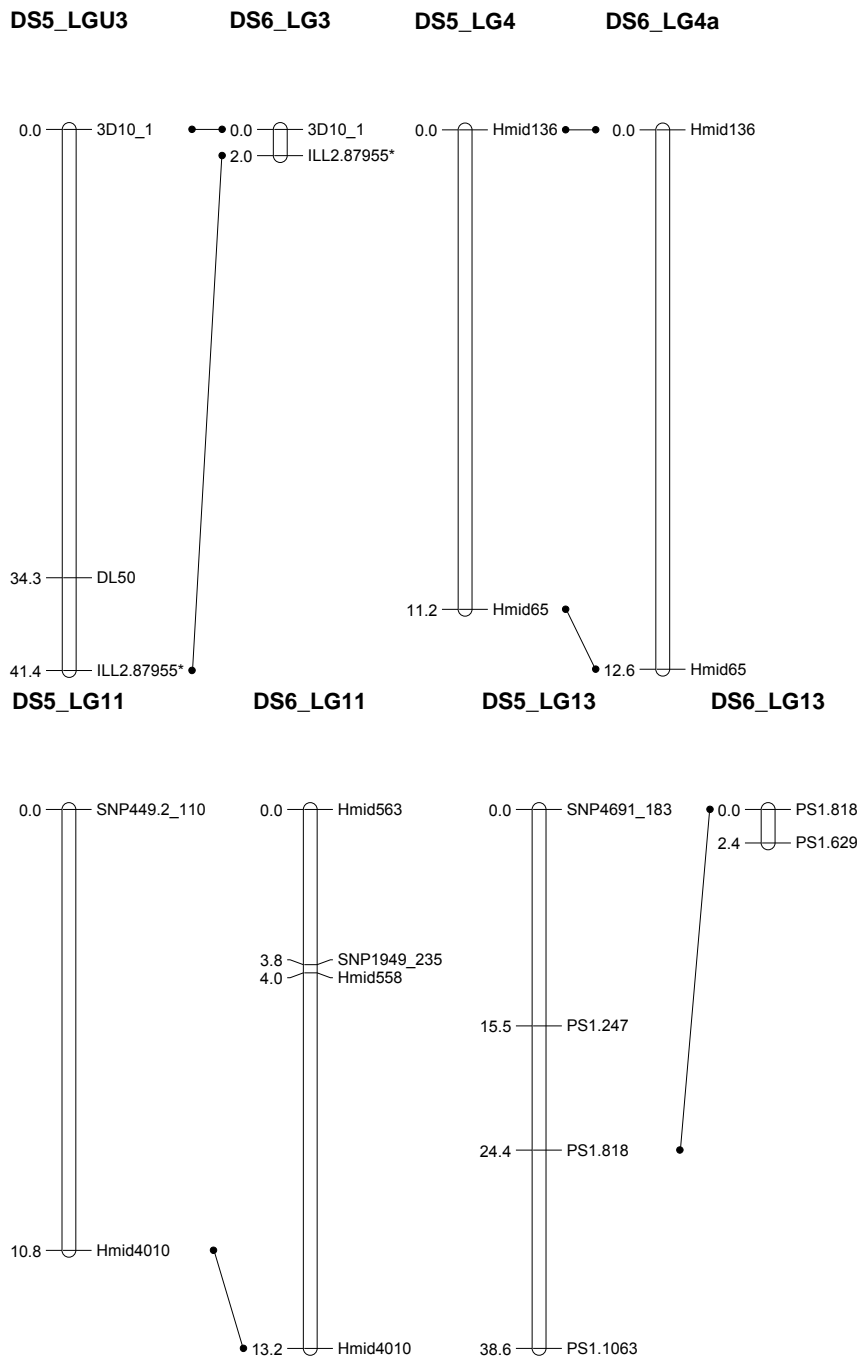


Figure 3.13: Homology for sex-average linkage maps of family DS1 and DS2. * Markers developed in this study.

Eleven of family DS5's 18 linkage groups could be joined to family DS6's 19 linkage groups with at least one concurrent marker (Fig. 3.14). Marker co-linearity was mostly retained, except between DS5_LG2 and DS6_LG2a.



Chapter two-Type 1 microsatellite development



Chapter two-Type 1 microsatellite development

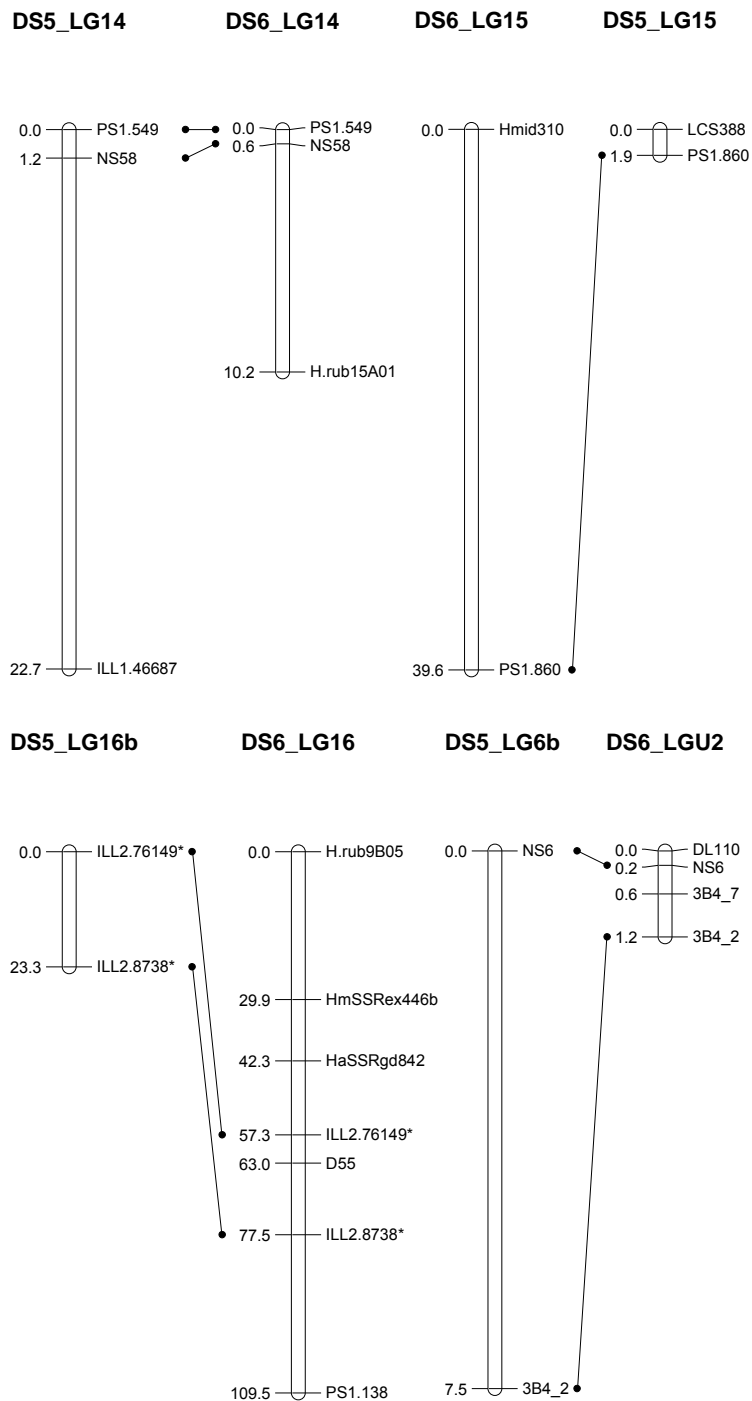
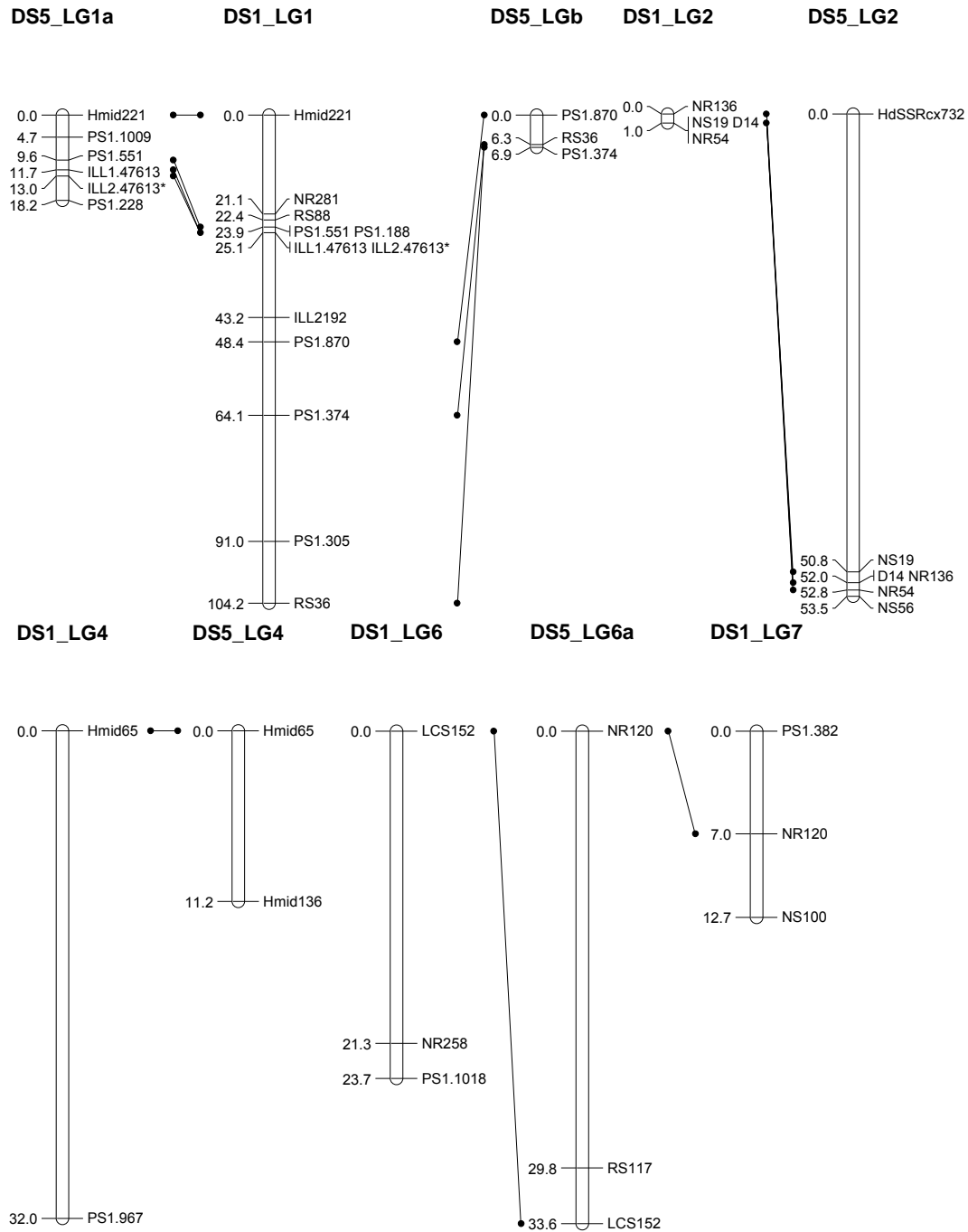


Figure 3.14: Homology for sex-average linkage maps of family DS5 and DS6. * Markers developed in this study.

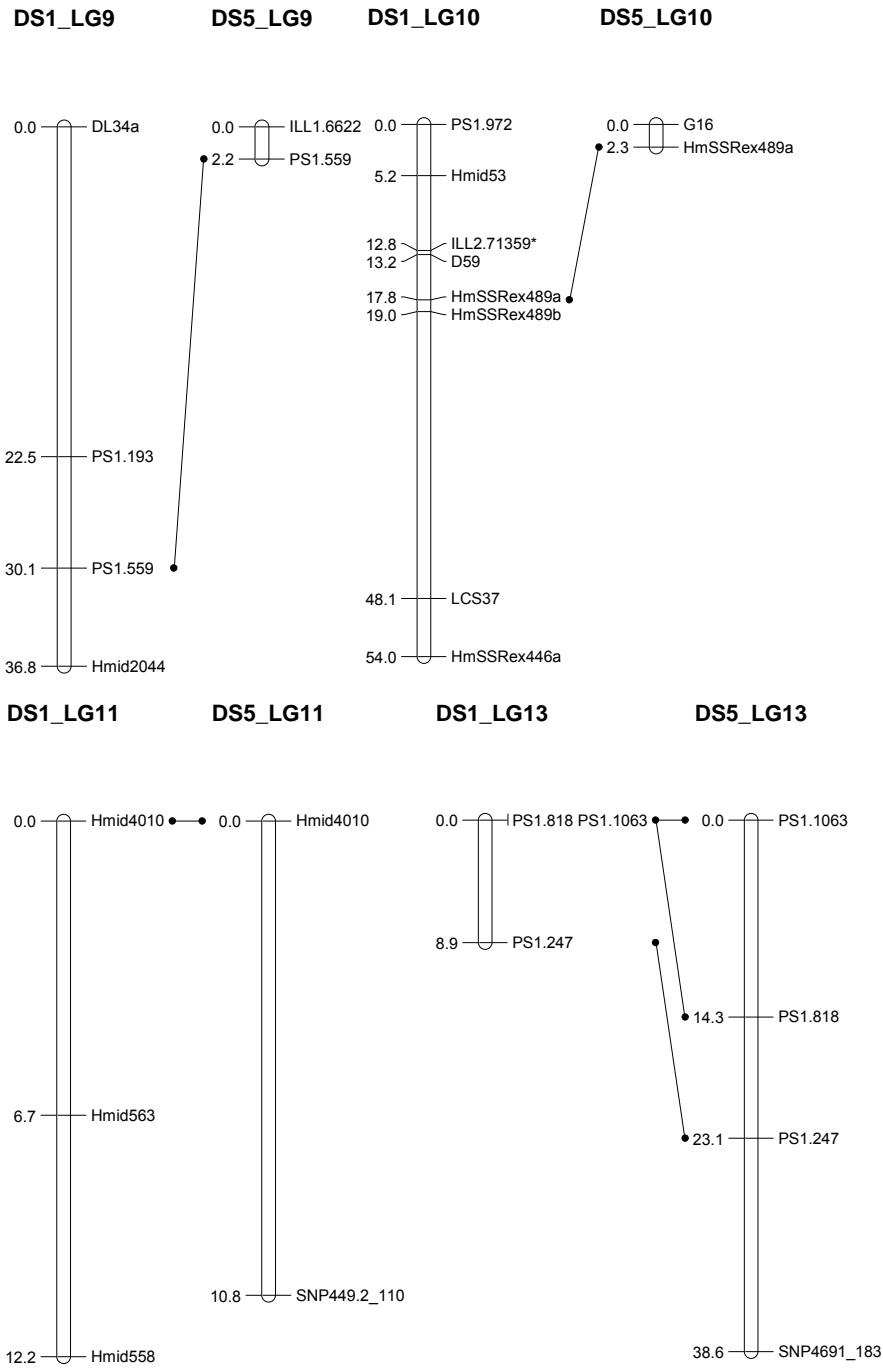
Fifteen of the 18 linkage groups in family DS1 could be joined to 15 of family DS5's 18 linkage groups. Co-linearity of marker order was retained except for the marker orders of *NR136* and *NS19* on linkage groups DS1_LG2 and DS5_LG2 (Fig. 3.15). DS5_LG1a and DS5_LG1b show homology with the same linkage group, DS1_LG1.

Chapter two-Type 1 microsatellite development

DS5_LG6a shows homology with both DS1_LG6 and DS1_LG7. The same applies to DS5_LG15, which showed homology to DS1_LG15 as well as to LG1_LG5.



Chapter two-Type 1 microsatellite development



Chapter two-Type 1 microsatellite development

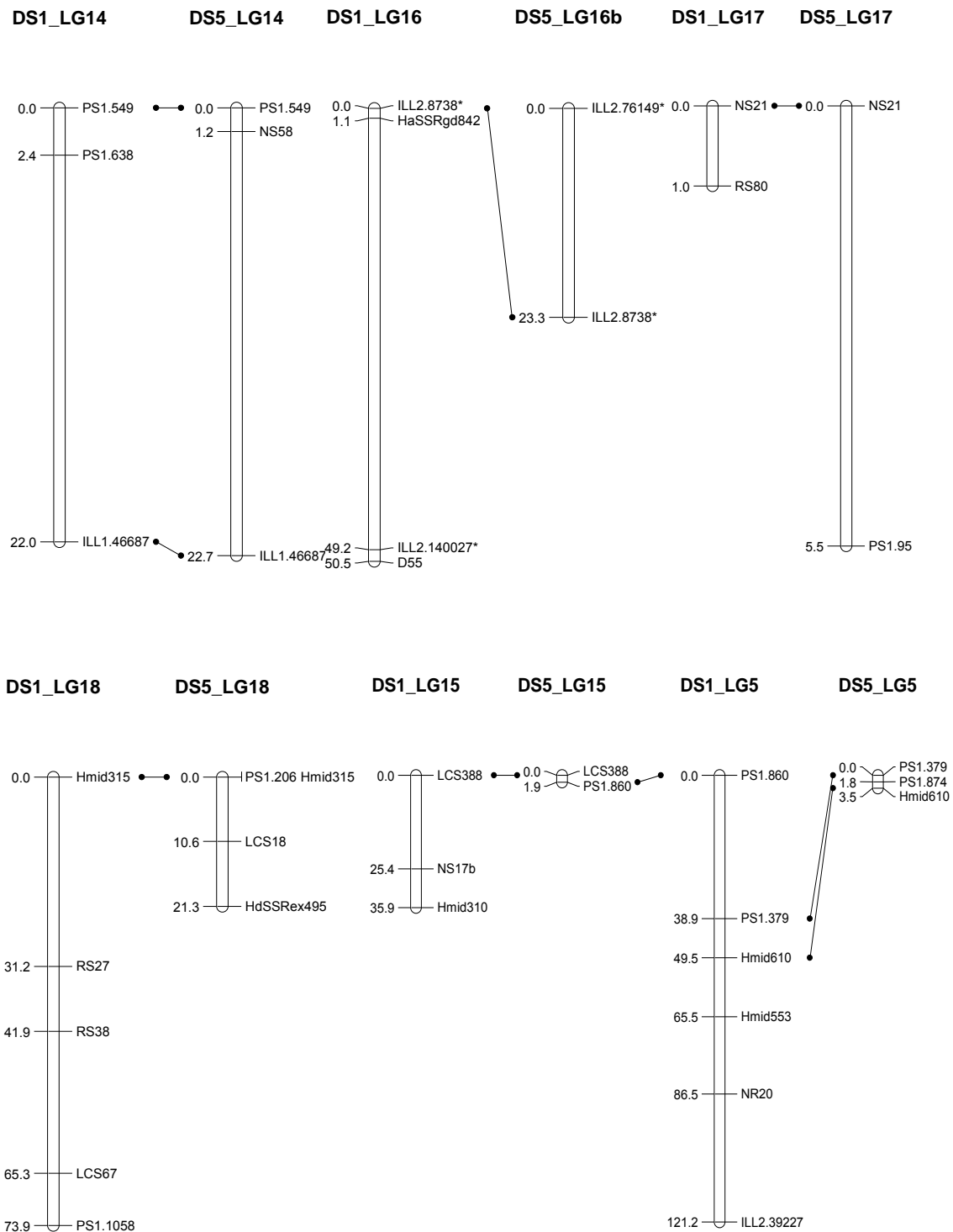
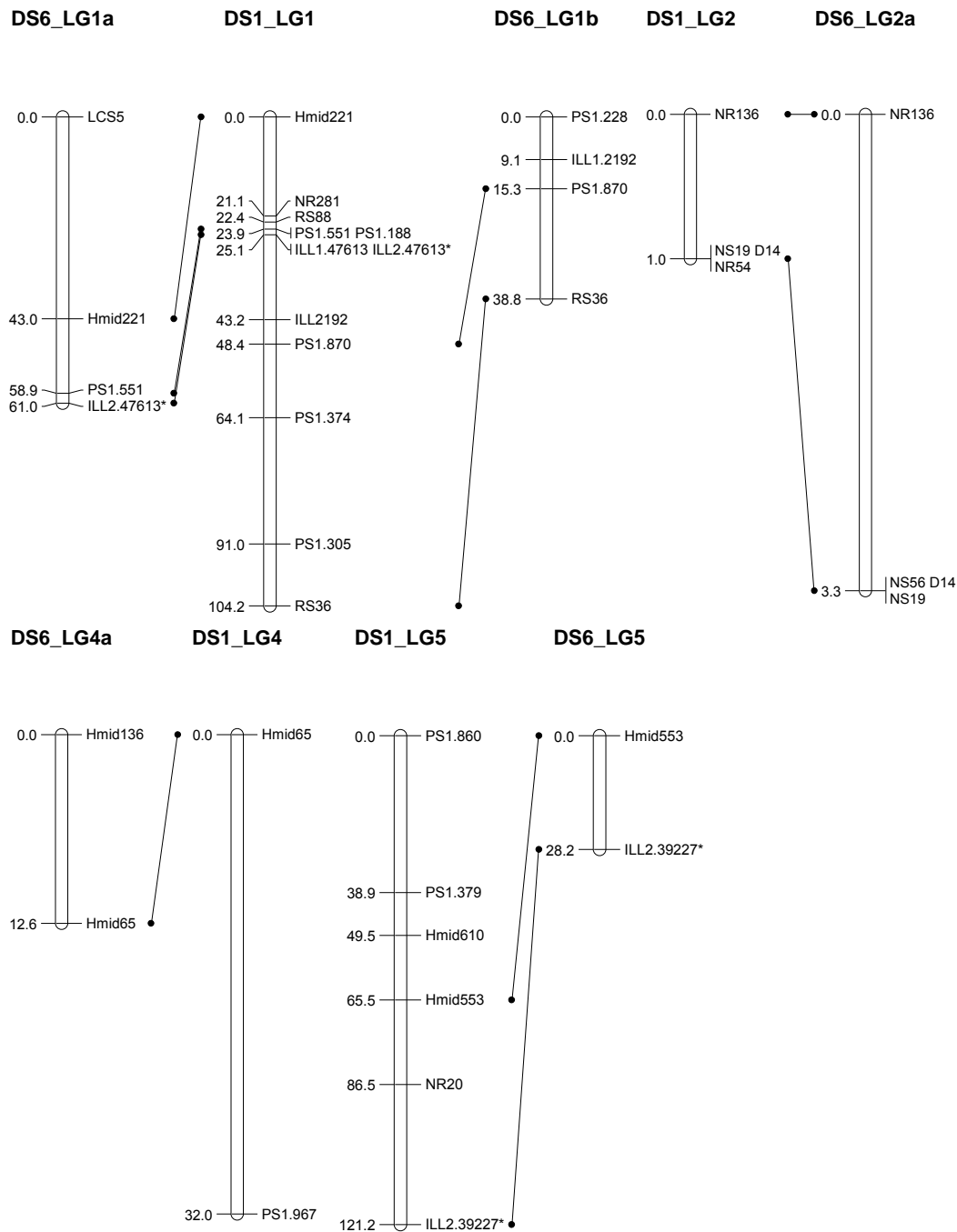


Figure 3.15: Homology for sex-average linkage maps of family DS1 and DS5. * Markers developed in this study.

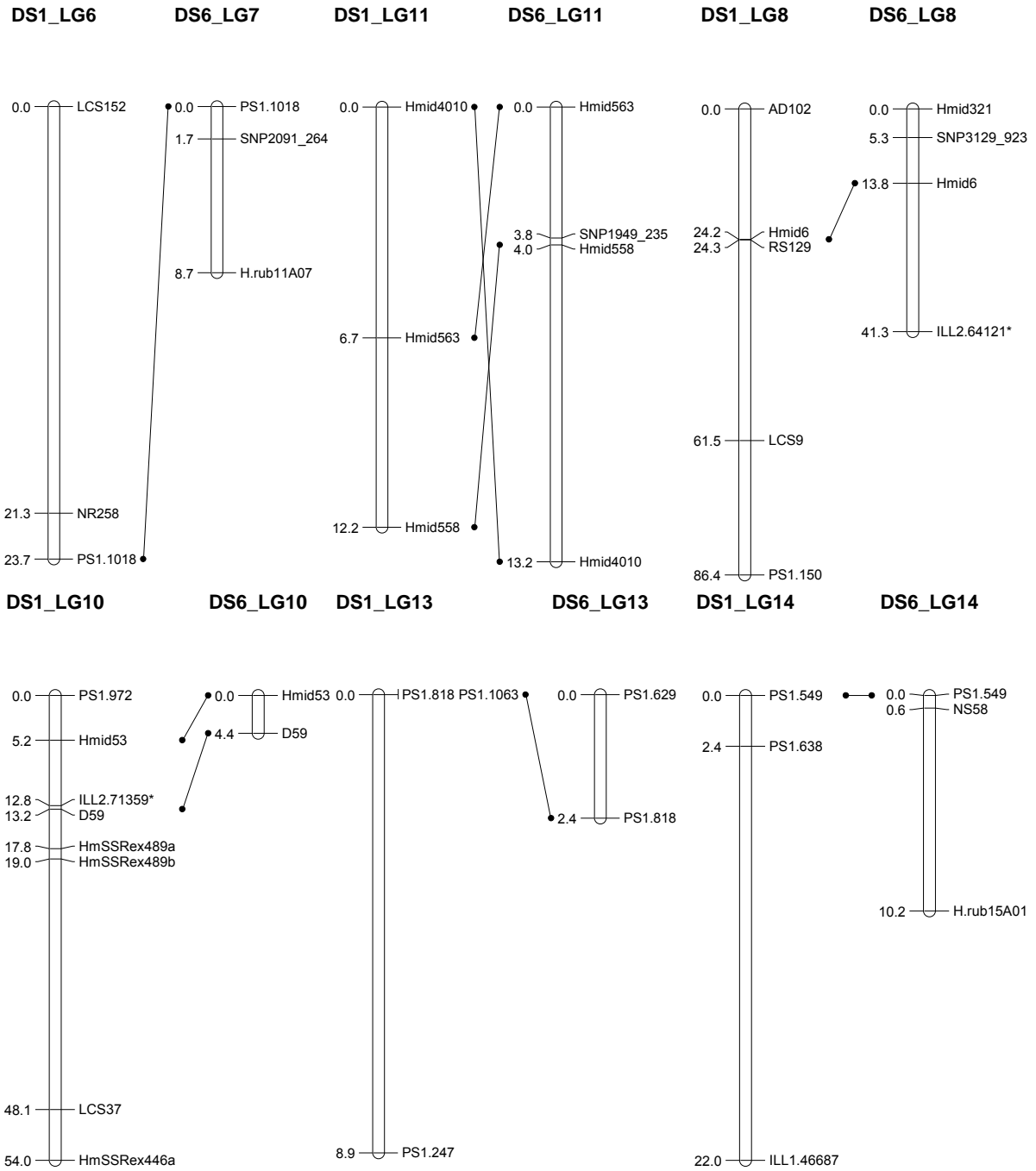
Thirteen of family DS1's 18 linkage groups showed homology to 13 of family DS6's 19 linkage groups (Fig. 3.16). Marker order was not conserved between LGs DS1_LG11 and DS6_LG11, and DS1_LG16 and DS6_LG16. For the first mentioned comparison, it was found that marker *Hmid4010* on DS1_LG11 was located on the

Chapter two-Type 1 microsatellite development

opposite end of the LG for DS6_LG11. The same situation was observed for the marker order discrepancy between DS1_LG16 and DS6_LG16. Marker *ILL2.8738* was more distally located on DS6_LG16 than on DS1_LG16. Certain LGs showed homology with more than one LG. These were DS1_LG1, which showed homology with DS6_LG1a and DS6_LG1b; and DS6_LG15, which had homology with LGs DS1_LG15 and DS1_LG5.



Chapter two-Type 1 microsatellite development



Chapter two-Type 1 microsatellite development

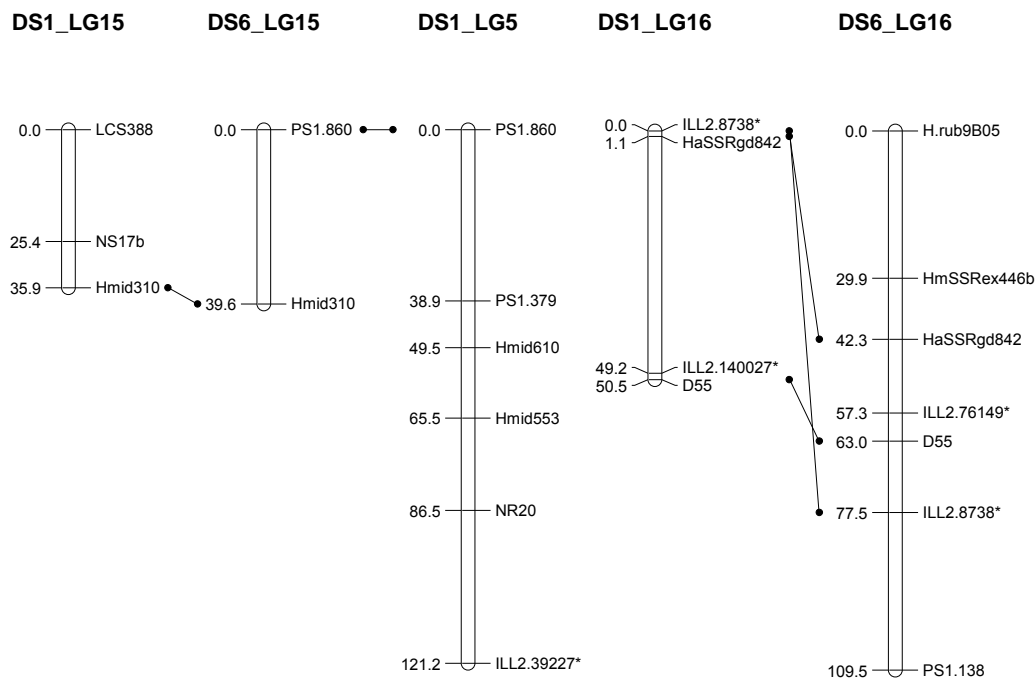
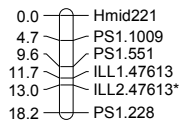


Figure 3.16: Homology for sex-average linkage maps of family DS1 and DS6. * Markers developed in this study.

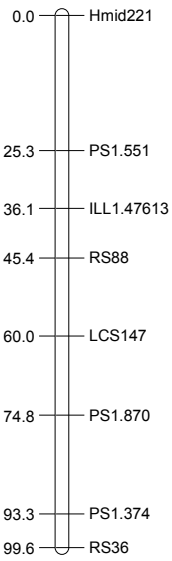
Eleven of the 17 linkage groups of family DS2 could be linked to 12 of the 18 linkage groups of family DS5's linkage map. On LG DS5_LG1b, the marker order of *RS36* and *PS1.374* was inverted to that on LG DS1_LG1 (Fig 3.17). For the homologous markers on LGs DS2_LG2 and DS5_LG2, the marker order may be conserved, but the genetic distance of marker *NR136*, differs approximately 52cM between the two families.

Chapter two-Type 1 microsatellite development

DS5_LG1a



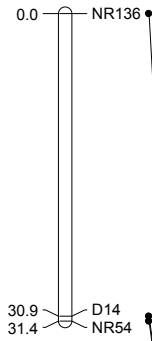
DS2_LG1



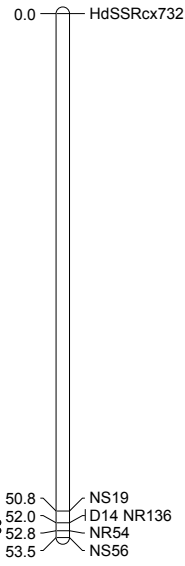
DS5_1LGb



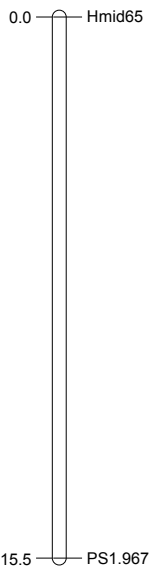
DS2_LG2



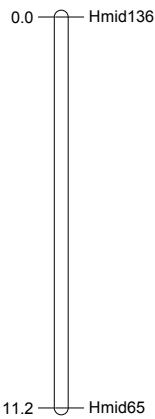
DS5_LG2



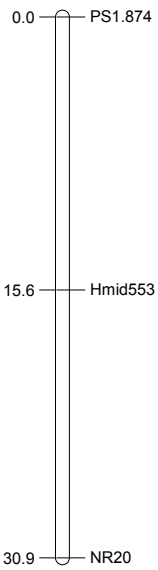
DS2_LG4



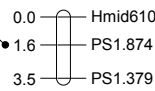
DS5_LG4



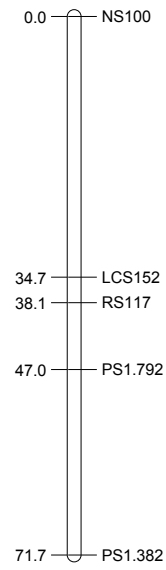
DS2_LG5



DS5_LG5



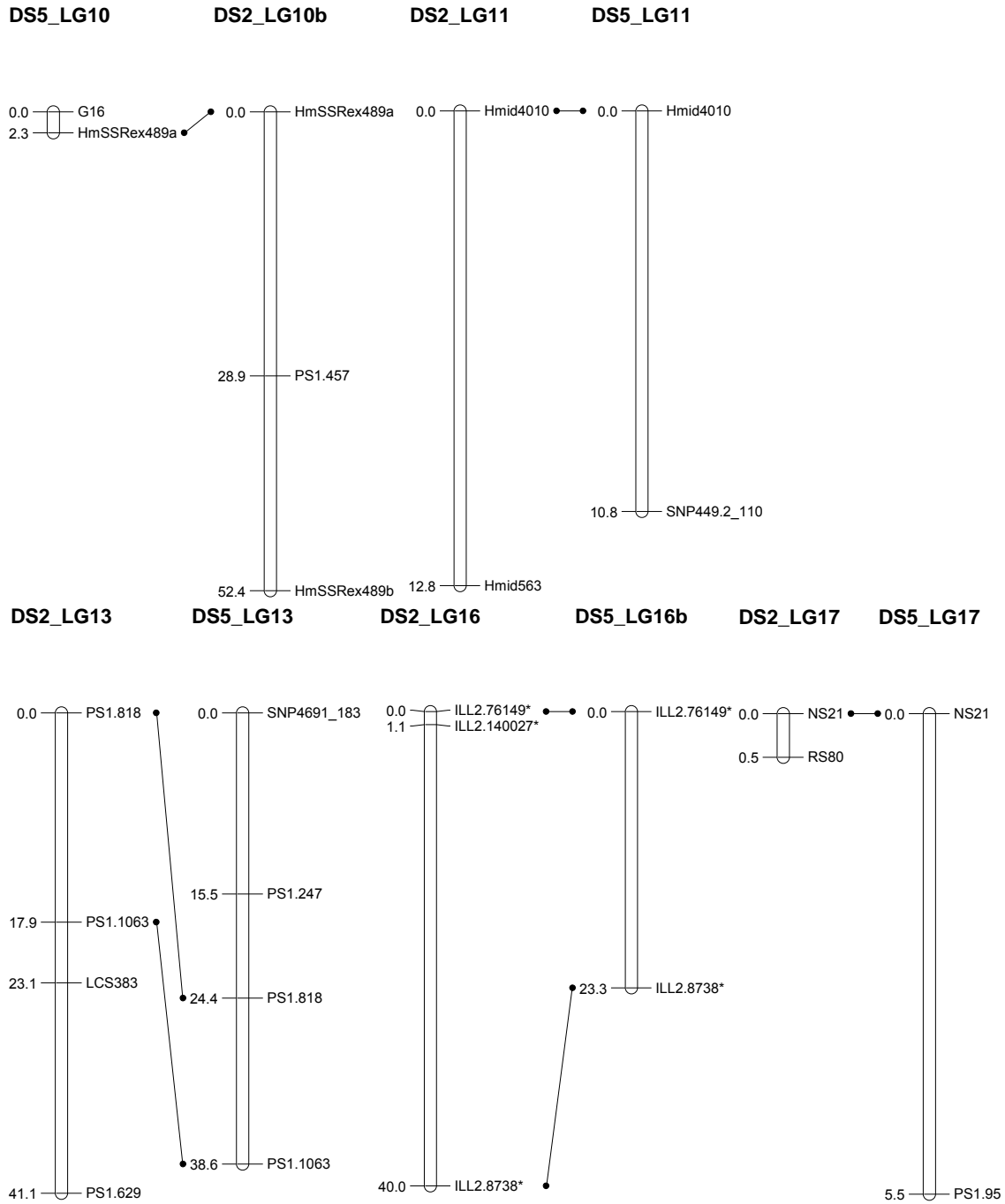
DS2_LG6/7



DS5_LG6a



Chapter two-Type 1 microsatellite development



Chapter two-Type 1 microsatellite development

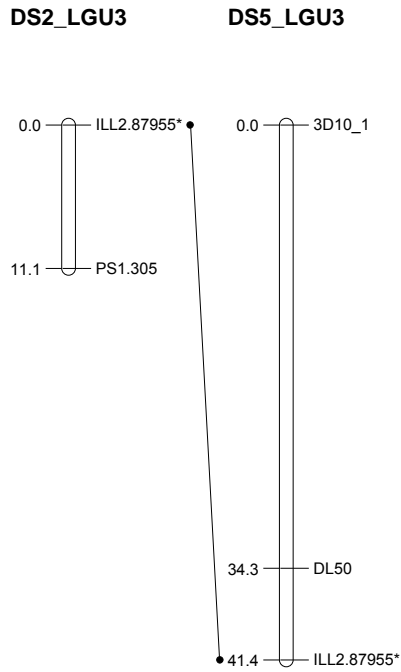
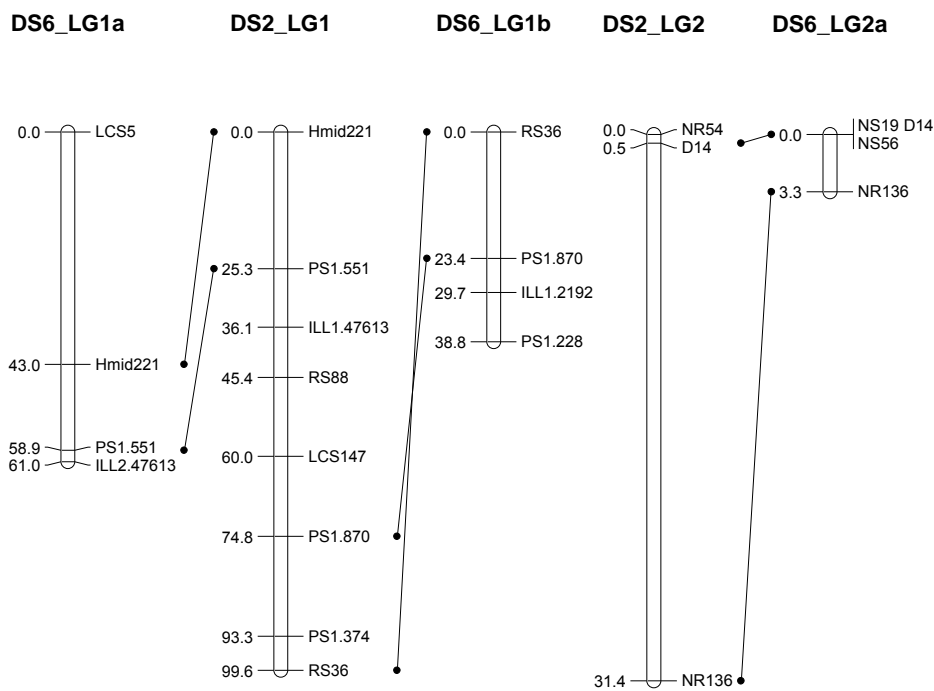


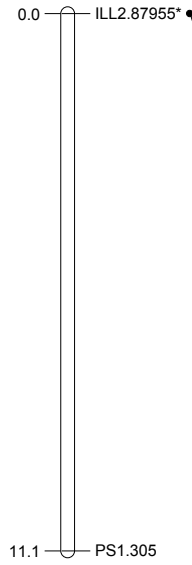
Figure 3.17: Homology for sex-average linkage maps of family DS2 and DS5. * Markers developed in this study.

Nine of family DS2's LGs could be linked to 10 of family DS6's 19 linkage groups with at least one concurrent marker (Fig. 3.18). There is one marker order discrepancy between LG2_LG1 and DS6_LG1b with marker *PS1.870* and *RS36*. DS6_LG1a and DS6_LG1b both show homology with DS2_LG1.

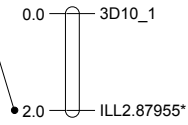


Chapter two-Type 1 microsatellite development

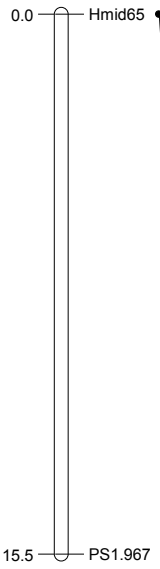
DS2_LGU3



DS6_LG3



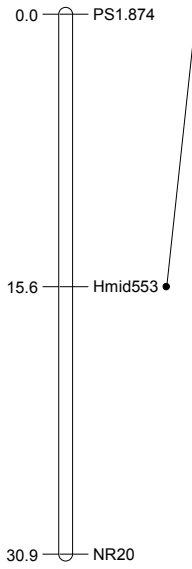
DS2_LG4



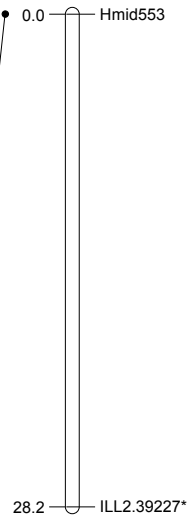
DS6_LG4a



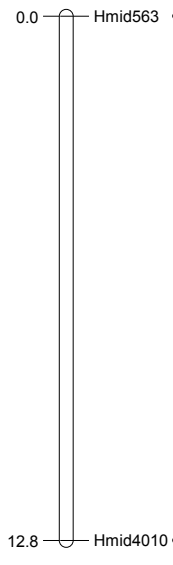
DS2_LG5



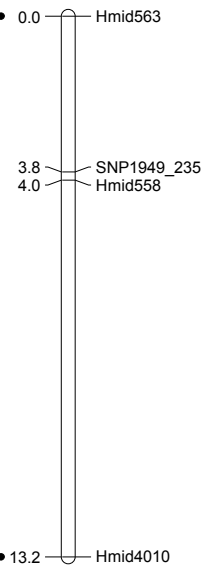
DS6_LG5



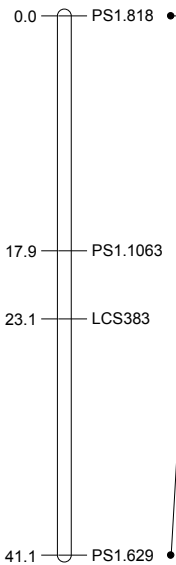
DS2_LG11



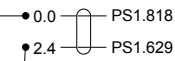
DS6_LG11



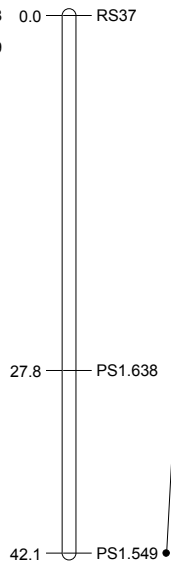
DS2_LG13



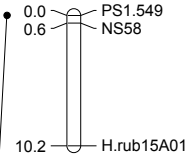
DS6_LG13



DS2_LG14



DS6_LG14



Chapter two-Type 1 microsatellite development

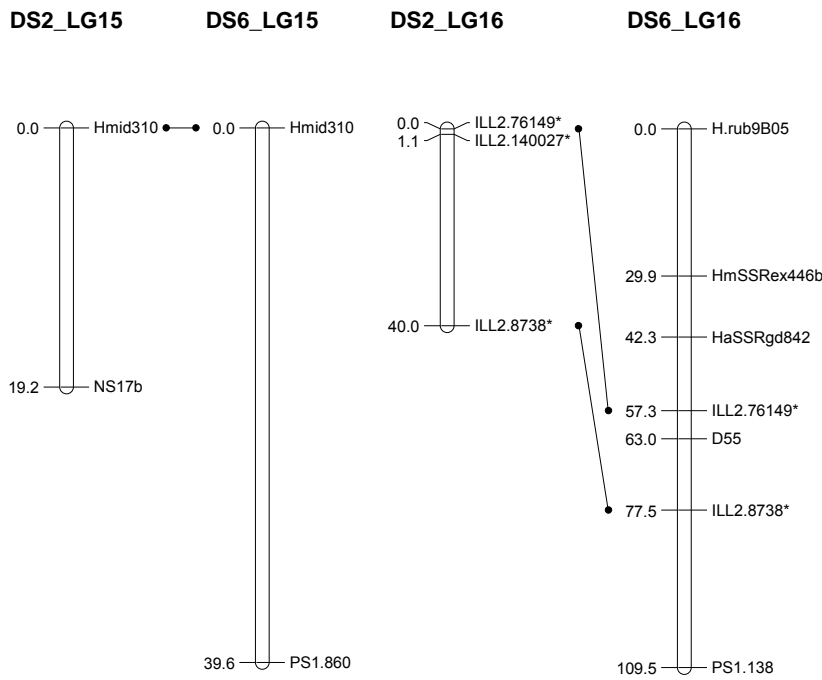


Figure 3.18: Homology for sex-average linkage maps of family DS2 and DS6. * Markers developed in this study.

4.3.6 Integrated map

The integrated map, formed by joining the sex-average maps for the different families in Joinmap® v.4, had 25 linkage groups, which ranged from 4.5cM to 121.5cM with an average marker spacing of 9.3cM (Fig 3.20). The number of markers mapped to the linkage groups ranged from 2-15 (Table 3.11). A total of 109 markers could be placed on this map with two of the mapped markers being SNPs. Genome length as calculated by equation A gave a result of 1703.39cM and 1663.84cM for equation B. The genome coverage, calculated by equation C, was 65%.

Table 3.11: Number of markers per linkage group, their corresponding lengths, average marker spacing and largest interval of the integrated map.

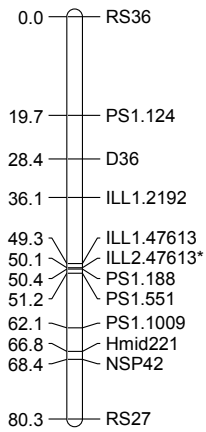
Linkage group	Markers per linkage group	Length of linkage group (cM)	Average marker spacing (cM)	Largest interval (cM)
1	12	80.3	6.7	19.7
2a	5	66	13.2	25.7
2b	7	55	7.9	38.9

Chapter two-Type 1 microsatellite development

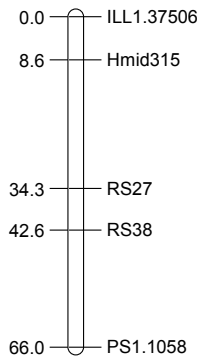
3a	7	63.7	9.1	15.9
3b	2	36	18	36
3c	2	23.7	11.9	23.7
3d	2	10.9	5.5	10.9
4a	3	39.9	13.3	23.2
4b	3	35	11.7	27.3
5a	3	28.8	9.6	15.6
5b	3	35.3	11.8	17.8
6/7	8	90	11.3	48.3
8	4	41.4	10.3	19.9
9a	4	34.1	8.5	21.8
9b	4	49.6	12.4	40.6
9c	3	6.5	2.2	4.3
9d	2	12.1	6.1	12.1
10	15	121.5	8.1	34.7
11	3	15.2	5.1	10.9
13	4	31.1	7.8	12.7
14	7	64.7	9.2	16.8
15	5	64.5	12.9	25.7
16	5	40.4	8.1	20.3
17	2	4.5	2.3	4.5
18	4	39.9	9.9	23.8
Total	119	1090.1	232.6	551.1
Average	4.8	43.6	9.3	22.0

Chapter two-Type 1 microsatellite development

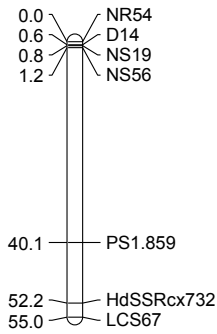
INT_LG1



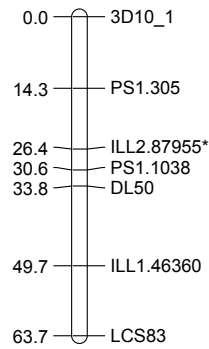
INT_LG2a



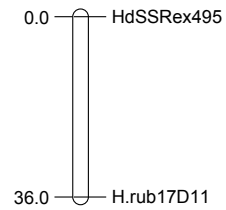
INT_LG2b



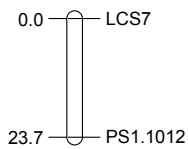
INT_LG3a



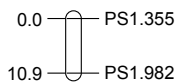
INT_LG3b



INT_LG3c



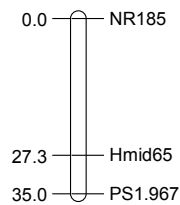
INT_LG3d



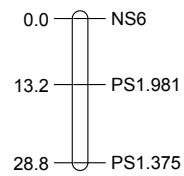
INT_LG4a



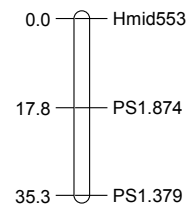
INT_LG4b



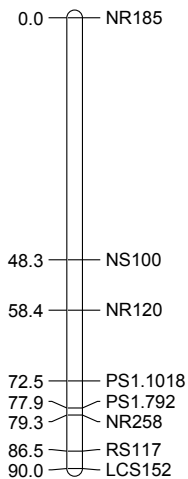
INT_LG5a



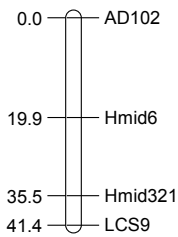
INT_LG5b



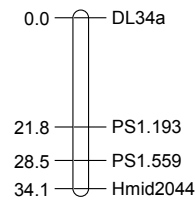
INT_LG6/7



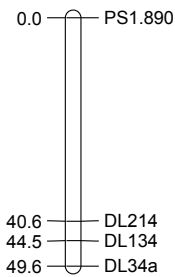
INT_LG8



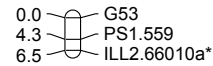
INT_LG9a



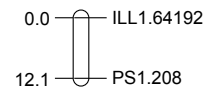
INT_LG9b



INT_LG9c



INT_LG9d



Chapter two-Type 1 microsatellite development

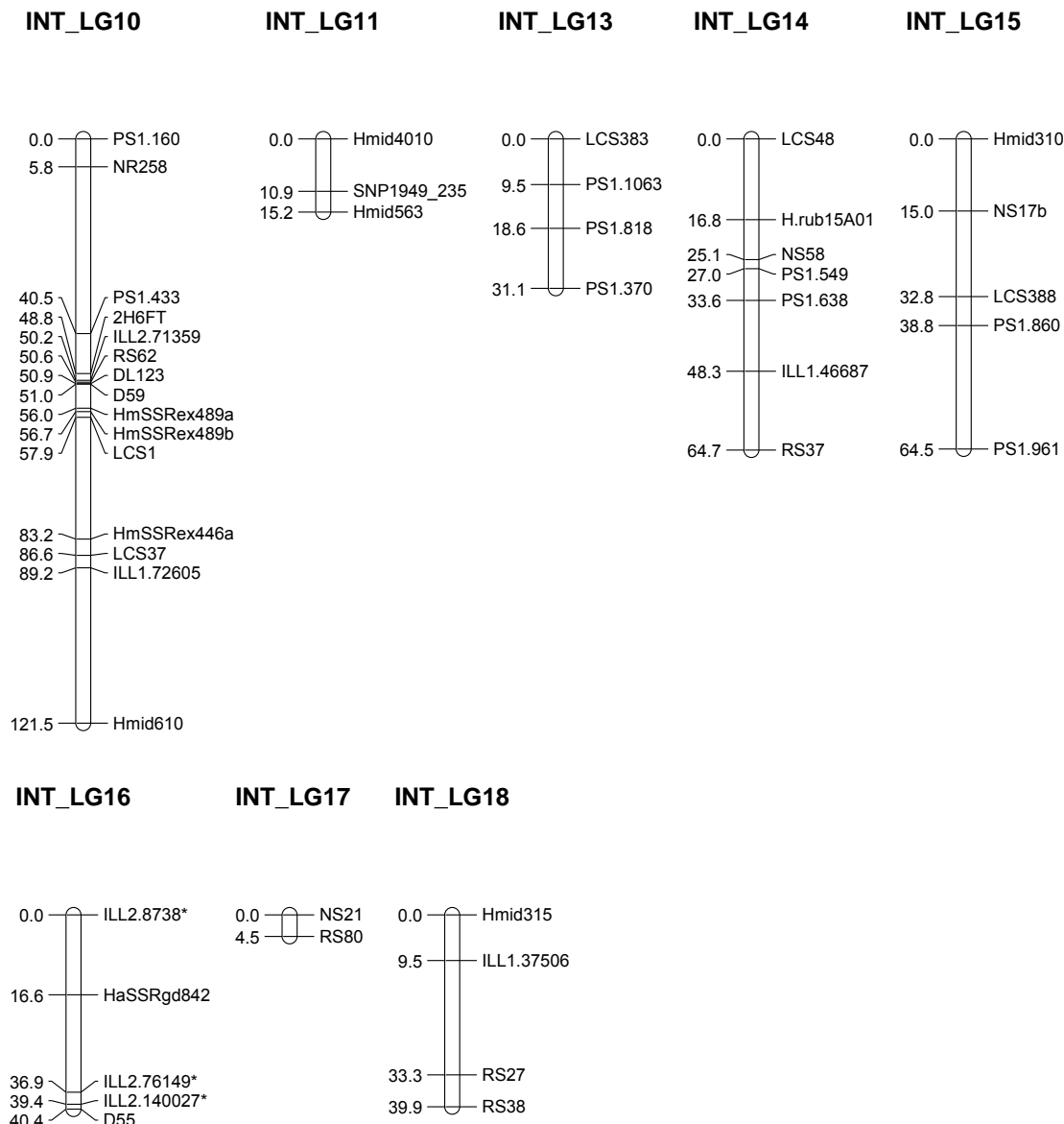


Figure 3.19: The integrated map for *H. midae*, constructed by merging the sex-average maps of families DS1, DS2, DS5 and DS6. The markers, which are mapped, are displayed on the right of the linkage group and the cumulative distances (in centimorgan; cM) of the markers are shown on the left. * Markers developed in this study.

5. Discussion

In this study, genic markers were used to develop a genetic linkage map for *H. midae*. This is the third linkage map developed for *H. midae* and the second map using microsatellite markers. However, this is the first map to use microsatellite as well as SNP markers for map construction in *H. midae* and lays the groundwork for future SNP mapping projects. The first linkage map (Badenhorst 2008) was constructed using mostly AFLP markers. These markers are extremely useful when

Chapter two-Type 1 microsatellite development

constructing linkage maps for species with little genomic information as no prior sequence information is needed when developing these markers. They also generate linkage maps with high marker densities, but their low reproducibility and transferability between laboratories make these markers less useful than genomic markers (Hubert and Hedgecock 2004). The second linkage map (Hepple 2010) was constructed using microsatellite markers to create a preliminary linkage map for *H. midae*. This map was only mapped in two families and did not contain any SNP markers. In the current study, genomic- and genic-SNPs were included in the data set as well as newly developed genic microsatellite markers (see chapter 2). These markers in, conjunction with microsatellites used for mapping by Hepple (2010), were mapped to four different families and is the first linkage map of *H. midae* to contain SNPs.

5.1 Segregation distortion and duplicated loci

The segregating markers were inspected for segregation distortion using Chi-square analysis (Table 3.3). These markers were mostly distorted in only a single family, except one SNP (*SNP1834_464*) that was found to be distorted in three families (DS1, DS5, DS6) (Family DS2 had no informative SNPs as the genotyping failed in this specific family and thus no markers were available to be mapped in family DS2). The parental alleles for this marker were **hkxhk**, meaning that three genotypic classes (**hh**, **hk** and **kk**) should be observed. However, the genotype data for this SNP only had two genotypic classes (**hk** and **kk**) with no 'hh' homozygote genotype. This SNP is associated with the alpha tubulin gene (Table 3.3), which is part of a large globular family. In eukaryotes many isoforms can exist (Joshi and Cleveland 1990). One possible explanation for the homozygote deficiency could be as a result of zygotic viability selection (Li *et al.* 2009), meaning that this SNP could potentially be linked to a deleterious recessive mutation that could affect tubulin formation. As tubulin is responsible for the movement of chromatids to the separate poles of the cell during mitosis and meiosis, defective tubulin can result in major genetic defects.

Segregation distortion is not uncommon in highly fecund species such as abalone (Hubert and Hedgecock 2004; Sekino and Hara 2007; Slabbert *et al.* 2010). It is believed that the high fecundity in these animals, resulting in many meiotic events, leads to a potential higher mutation rates in gametes; which could result in

Chapter two-Type 1 microsatellite development

deleterious recessive mutations leading to a high genetic load (Launey and Hedgecock 2001). These high mutation rates can also lead to identical by descent homozygotes, which can cause the overrepresentation of homozygote alleles in the offspring (Launey and Hedgecock 2001). These mechanisms contribute to segregation distortion in abalone as they influence the genotype of the offspring, which will have an effect on the expected Mendelian segregation ratios.

Microsatellite markers, which constitute the majority of markers in this study, are prone to genotyping errors due to their size-based nature, as well as null alleles and duplicated loci. Genotyping errors also include errors such as human error and using low quality DNA. A genotyping error occurs when an observed genotype is not scored as the actual genotype, leading to errors in the data set that will be used for subsequent linkage mapping. A 5% error rate in genotype data has been shown to cause up to 50% inflation in the map distance of linkage mapping studies (Pampanon *et al.* 2005; Ball *et al.* 2010). Markers, which exhibited unreliable genotypes even after re-typing, were excluded from further mapping in this study to reduce errors introduced due to erroneous genotypes.

In the four families, approximately 18% of the loci contained null alleles, with 3% being gene-linked markers and 15% being genomic markers. In a study conducted on Pacific abalone (Li, Q *et al.* 2010), also using genic microsatellites, 17% of the loci had null alleles. The study by Li, Q *et al.* (2010) also noted that the frequency of null alleles were lower than for microsatellites previously developed from genomic DNA (57%). This is due to the fact that genic microsatellites are developed from conserved segments of DNA and are less likely to undergo mutations in primer flanking regions that cause null alleles. Null alleles are frequently observed in highly fecund species such as abalone and occur when a mutation in the primer binding site causes failure of the primer to anneal to the template DNA during PCR amplification. This will lead to a genotype of a heterozygote individual being incorrectly scored as a homozygote and a null allele homozygote to be scored as missing data. This causes errors in the genotype data causing distorted loci. However, if the segregation of the null alleles can be traced, the marker with the segregating null allele can then be subjected to segregational analysis and shown not to be distorted after all. These null alleles are coded in Joinmap[®] v.4 as **abxcd**. For some genotypic combinations this will not be possible for example: if the parental

Chapter two-Type 1 microsatellite development

alleles are A_1NxA_2B (**efxeg**), it will be impossible to determine if a genotype is either A_2N or A_1A_2 , which will lead to wrong assignment of a genotype to a certain genotype class. These types of markers should be excluded from further linkage analysis or primer redesign can be carried out. Due to the large number of microsatellites needed for mapping projects, redesign has to be investigated to determine if such an endeavour is practical (Li *et al.* 2009).

Some microsatellites showed more than two alleles when genotyped (Table 3.3). This indicates that these microsatellites had undergone duplication and are referred to as duplicated loci. Microsatellites can be duplicated by a variety of mechanisms including genome duplication, tandem duplication or locus transfer by mobile elements (Baranski *et al.* 2006b). In some instances, two of the alleles are inherited in a disomic pattern if tandem duplication is taken as the model of locus duplication. Such markers can still be useful for mapping; however, duplicated markers were excluded in this study to avoid the complicating process of assigning alleles to duplicated loci. Such loci can also become a problem in future studies including candidate gene discovery and synteny mapping.

5.2 Linkage mapping

In a previous study by Van der Merwe and Roodt-Wilding (2008) the haploid number of *H. midae* was confirmed to be 18, thus the study aimed to obtain 18 linkage groups in map construction. The sex-average linkage maps of DS1 and DS5 had 18 linkage groups with families DS2 and DS6 having 17 and 19 linkage groups, respectively. Even though the linkage groups are similar to the haploid number of *H. midae*, caution must be taken as the marker density of the map is still quite low and as more markers are added to the map, linkage groups might be merged. The sex-specific maps' linkage groups ranged from 13 to 17 linkage groups. This lower number of linkage groups observed than expected is frequently seen in low-density preliminary linkage maps (map coverage of integrated map in the current study: 65%) (Baranski *et al.* 2006a; Liu *et al.* 2006).

The linkage group naming in the current study was based on comparing groups (before map construction) with homologous markers across the four families and then naming the linkage groups accordingly where consistent homologous markers were observed. Family DS1 served as the linkage group naming template. All other

Chapter two-Type 1 microsatellite development

families' linkage groups were compared to DS1 so that linkage groups with more than two homologous markers could be characterized as the same linkage group in a different family. However, for certain linkage groups in the other three families no homologous markers could be observed and these were then given an unknown (U) number for example: DS2_LGU1. Some unknown linkage groups were shared across families and they were given the same unknown number. In family DS2 there were four unknown linkage groups. Interestingly, linkage groups 3, 8, 9 and 12 are not present in DS2 according to comparisons to family DS1. The unknown linkage groups probably represent the four 'missing' linkage groups, but because these groups don't show any homologous markers to family DS1's linkage groups 3, 8, 9, or 12, appropriate names could not be given. However, if more families in conjunction with more markers could be mapped together with the markers that are present on the unknown (U1, U2, U3 and U4) as well as the known linkage groups (LG3, LG8, LG9 and LG12), the unknown groups could be renamed to the appropriate linkage group in the future. With comparisons conducted with these unknown groups, it seems as though DS2_LGU3 and DS5_LGU3 could possibly be DS2_LG3 and DS5_LG3 as both showed homology with DS6_LG3, but further investigation has to be performed in order to confirm this.

The genome lengths of the maternal and paternal maps differed. The average expected genome lengths for maternal maps were 1482.6cM, 1292.4cM, 992.6cM, and 1383.21cM for family DS1, DS2, DS5 and DS6, respectively. The paternal maps are expected to be shorter than the maternal maps as sex-specific recombination rates have been observed in abalone and other fish species (Singer *et al.* 2002; Moen *et al.* 2004; Sekino and Hara 2007; Zhan *et al.* 2011). The mechanism responsible for sex-specific recombination differences is not well understood and remains a topic of much discussion (Danzmann and Gharbi 2001; Singer *et al.* 2002; Sekino and Hara 2007). However, only family DS1, DS5 and DS6 (1106.4cM, 602.4cM and 951.93cM, respectively) conformed to this expected result. Family DS2's paternal map (1358.2cM) was slightly longer than the female map. This difference between the two parents' map lengths can possibly be explained by the fact that even though the observed length of the maternal map is indeed longer than the paternal map, the number of linkage groups in the maternal map (14) is lower than that in the paternal map (17). This will have an effect on genome length when

Chapter two-Type 1 microsatellite development

calculating the expected genome length using equation A and B. The sex-specific map lengths in the current study are longer than those obtained for *H. rubra* and *H. discus hannai*, but shorter than for *H. diversicolor* (Baranski *et al.* 2006a; Sekino and Hara 2007; Shi *et al.* 2010) (Table 3.12).

Table 3.12: Summary of abalone linkage maps indicating number and type of markers used for map construction, number of segregating families and the number of linkage groups.

Species	Markers	Segregation families	Linkage groups	Average marker interval (cM)	Mapped markers	Expected map length (cM)	Map coverage	Reference
<i>H. diversicolor</i>	412 AFLP markers	1 family (76 offspring)	16-18	25.7 (female) 25.0 (male)	90 (female) 94 (male)	2773.0 (female) 2817.1 (male)	67.6% (female) 67.3% (male)	Shi <i>et al.</i> 2010
<i>H. diversicolor</i>	182 Microsatellites	1 family (96 offspring)	16	4.6 (integrated map)	175 (integrated map)	943.8 (integrated map)	80.7% (integrated map)	Zhan <i>et al.</i> 2011
<i>H. rubra</i>	122 Microsatellites	1 family (95 offspring)	17-20	9.8 (female) 7.3 (male)	98 (female) 102 (male)	1586.2 (female) 940.5 (male)	64% (female) 80% (male)	Baranski <i>et al.</i> 2006a
<i>H. discus hannai</i>	365 AFLP markers 10 RAPD markers 9 Microsatellites	1 family (106 offspring)	19-22	18.3 (female) 18.2 (male)	119 (female) 147 (male)	2584.4 (female) 2054.8 (male)	68.6% (female) 66.5% (male)	Liu <i>et al.</i> 2006
<i>H. discus hannai</i>	180 Microsatellites	3 families (60-96 offspring)	18-19	6.3 (female) 4.7 (male)	160 (female) 167 (male)	1156.7 (female) 899.1 (male)	76.6% (female) 78.4% (male)	Sekino and Hara 2007

Chapter two-Type 1 microsatellite development

The average marker interval was found to be between 10cM and 15cM, which is the required interval needed for QTL mapping in aquaculture species where outbred populations are used (Massault *et al.* 2008). However, it has to be mentioned that the interval has to be constant across the whole map. The linkage map constructed in this study still has many gaps of very low marker density, which is not suited for QTL identification. These gaps first have to be filled before QTLs can be unambiguously identified. Marker density is very low compared to other studies in abalone (Baranski *et al.* 2006a; Sekino and Hara 2007; Zhan *et al.* 2011). In the study by Sekino and Hara (2007) 160 and 167 markers could be mapped to the maternal and paternal maps respectively, which are much higher in comparison with the current study's number of markers that could be mapped in the sex-specific maps. The studies by Baranski *et al.* (2006a) and Zhan *et al.* (2011) also mapped more markers to their respective maps (98 markers in the paternal and 102 in the maternal map of *H. rubra* and 175 markers in the integrated map of *H. diversicolor*), which accounts for the denser maps obtained in those studies. However, similar results were obtained by Hepple (2010) and QTL could successfully be identified by Slabbert (2010) from the preliminary linkage map constructed for *H. midae* by Hepple (2010).

Linkage groups of the sex-average maps were compared across all four families and joined where marker homology was found. Comparisons between families DS1 and DS2's sex-average maps exhibited the greatest similarity, with 25 links that could be made between the two maps (Fig. 3.13). The lowest map similarity was between the sex-average maps of families DS2 and DS6, with 17 links of homologous markers between them. The potential reason for this could be because DS2 is from a different farm than DS6 and it seems that families from the same farms showed similar results across all maps (Figure 3.18).

When DS2 and DS5's LG1 was inspected for marker homology, it was found that DS2 shared three markers with DS5_LG1a and three markers with DS5_LG1b. This division of DS5's LG1 into two separate groups could be due to the fact that DS2_LG1 has additional markers (*RS88* and *LCS147*), which link the two groups into one group in family DS2. Marker order between the homologous markers on DS2_LG1 and DS5_LG1a were conserved, but there was a marker order discrepancy between the homologous markers on DS2_LG1 and DS5_LG1b.

Chapter two-Type 1 microsatellite development

DS5_LG1b's marker order for the homologous markers were *PS1.870*, *RS36*, *PS1.374*, but for DS2_LG1 *RS36* and *PS1.374* were inverted. DS1_LG1 could be used to further verify which marker order could be the most probable by checking it against DS2_LG1 and DS5_LG1b. DS1_LG1's marker order for these three markers are in concurrence with DS2_LG1's. DS6_LG1b could not be used for the verification process, as it does not contain marker *PS1.374*. Thus, the reason for DS5_LG1a being on separate linkage groups could be as a result of the markers *RS88*, *LCS147* and *ILL1.21921b* not being mapped to these two linkage groups. These mentioned markers could represent the 'link' that is needed to join these two groups together (as is seen in DS1_LG1 and DS2_LG1). This shows the usefulness of using more than one family for linkage map construction so that a broader perspective of marker order and linkage group composition can be obtained. If only family DS5 was used for map construction, DS5_LG1a and DS5_LG1b would be concluded to be two separate linkage groups and that the marker orders of DS5_LG1b was the only correct order. These errors could seriously affect QTL mapping.

Linkage groups 6 and 7 showed homology across all four families (Fig. 3.13-3.18). Homologous markers were not restricted to corresponding linkage groups. Some markers that were mapped to linkage group 6 in one family, showed homology to linkage group 7 in another family, and *vice versa*. DS2_LG6/7 was so named because markers on DS1_LG6 and markers on DS1_LG7 were mapped to one group in DS2. When linkage group DS1_LG6 was compared to DS2_LG6/7, one link could be made (*LCS152*). The distal markers (*NS100* and *PS1.382*) of DS1_LG7 were mapped to similar positions on DS2_LG6/7. Linkage group DS5_LG6a of family DS5 had two links, markers *LCS152* and *RS117*, to DS2_LG6/7 of family DS2 with conserved map order. DS5_LG6a also showed homology to DS1_LG6 (*LCS152*), but also had another link to DS1_LG7 (*NR120*). This again shows the probability that markers of linkage groups 6 and 7 could be shared between these two groups and are probably not separate, but one group that cannot be linked (in families DS1, DS5 and DS6) due to low marker density. Other support for this is that marker *PS1.1018*, on linkage group DS6_LG7 of family DS6 had a link to DS1_LG6 of family DS1. The linkage group DS2_LG6/7 of family DS2 is the only linkage group across all four families, which could map the markers from both linkage groups 6 and 7 onto one group. Unfortunately, the marker densities of the DS5 and DS6 maps were very low

Chapter two-Type 1 microsatellite development

and did not contain enough markers to map both linkage groups 6 and 7. In family DS5 only linkage group 6 could be mapped and in family DS6, only linkage group 7. However, family DS1, which had the highest genome coverage (62%) and highest marker density (4 markers per linkage group on average) had two separate linkage groups for linkage groups 6 and 7. In the integrated map, most markers from both linkage groups 6 and 7 were joined into one group (Int_LG6/7). If more families can be mapped with the anchor markers of linkage groups 6 and 7 together with other marker types that can increase marker density, such SNPs, the question of whether linkage groups 6 and 7 are indeed one group can possibly be resolved.

Linkage groups DS5_LG15 and DS5_LG15 had marker *PS1.860* mapped to them. This marker could also be mapped to DS1_LG5. This could mean that these LGs could be located on the same chromosome. However, greater marker densities are needed to resolve this matter as only one concurrent marker was shared between these LGs. Another possible explanation could be that the markers present on DS1_LG15 did not show strong enough linkage to *PS1.860* and subsequently could not be mapped to LG DS1_LG15. Marker *PS1.860* could however be mapped to another linkage group, which had a marker present and exhibited strong cross linkage with marker *PS1.860*. The marker could thus be mapped to DS1_LG5. It has to be mentioned, however, that just because a marker can be mapped to a certain linkage group does not mean that this was the correct placement for it. The same situation was observed with marker *PS1.305* on LG DS2_LGU3, which showed homology to LG DS1_LG1. This marker was only present in families DS1 and DS2 so that it's very difficult to determine which marker placement of this marker is the correct one.

The number of informative markers in each separate family is low compared to the overall markers available for mapping (295) (Table 3.2). This low number of informative markers makes it very difficult to construct dense linkage maps with even high numbers of polymorphic microsatellites. Increasing the number of mapping families provides an avenue for increasing the number of informative markers available for map construction. For example: if some markers are polymorphic, but are not segregating in the mapping families, they cannot be used for map construction. However, if additional mapping families are used to genotype these markers in, they could potentially be informative in these families and thus be

Chapter two-Type 1 microsatellite development

included for map construction. By increasing the number of mapping families from two (Hepple 2010) to four in the current study, more markers was made available for map construction, leading to a higher map density (65% vs. 50%) and more markers that could be placed on the integrated map (109 vs.70).

After the sex-average maps were inspected for marker homology, they were merged where possible to form an integrated map (Fig. 3.19). The resulting map had 25 linkage groups. There are more linkage groups in the integrated map than the haploid number for abalone ($x=18$). The reason for this phenomenon was because some linkage groups in certain families did not share any anchor markers to any other linkage groups across the families or they did not have enough linkage information to be joined to other already merged linkage groups. These linkage groups could potentially be joined in the future to their appropriate linkage groups as more anchor markers become available in subsequent mapping projects. There are four different presumable LGs for LG 3 and 9 (Fig. 3.19). Linkage group INT_LG3a had the most markers mapped to it, whereas linkage group INT_LG3b, INT_LG3c and INT_LG3d only contained two markers each. The markers on these three linkage groups could not be joined to the larger INT_LG3a. However, if more shared markers are found between these four linkage groups they could eventually be joined to form one integrated linkage group, INT_LG3. For linkage groups INT_LG9a, INT_LG9b, INT_LG9c and INT_LG9d, some markers were homologous between the four linkage groups. Marker *DL34a* was shared between linkage group INT_LG9a and INT_LG9b, and marker *PS1.559* was shared between INT_LG9a and INT_LG9c. The reason for these linkage groups not being joined together in the integrated map was because the linkage information contained in these linkages were not significant enough for Joinmap[®] v4.0 (LOD<3) to allow merging and subsequent linkage group integration between the aforementioned groups.

Certain LGs, such as INT_LG2b, contain markers that are not present in any families' linkage maps. The markers mapped to this LG, did not have sufficient linkage separately to be mapped to the sex-average maps separately, but together these markers could be mapped. This again illustrates the benefit of using more mapping families.

5.3 Mapped microsatellites versus mapped SNPs

Chapter two-Type 1 microsatellite development

Altogether 264 polymorphic microsatellites and 31 polymorphic SNPs were available for mapping. After inspecting the markers for informativeness these numbers decreased (Table 3.2). The SNPs that could be mapped in this study were low (Table 3.10). This is due to SNPs' low PIC values. They are bi-allelic and can therefore only have two alleles, resulting in these markers containing half the polymorphic information in comparison to a marker with four alleles and can thus only be encoded with **hkxhk**, **lmxll** and **nnxnp**. In a study conducted in Atlantic salmon (Moen *et al.* 2008), microsatellites and SNPs were used to construct a linkage map. They achieved a very successful rate (99%) for the mapping of the SNPs. The reason for this is that the number of mapping families were sufficiently high (10) as to allow maximum incorporation of informative SNPs in the mapping project. In addition, the number of SNPs relative to microsatellites was almost three times higher, thus increasing the PIC value of the SNPs to be closer to that of the microsatellites. The marker type with the higher PIC value can drown out, so to speak, the effect of the lower informative marker type (SNPs), which can lead to them to be not mapped as readily as the microsatellites. By increasing the number of SNPs in mapping projects and increasing the number of reference families this problem can be overcome in future mapping projects in *H. midae*.

Of the seven SNPs, that could be mapped to the sex-average and sex-specific maps, six provided significant hits to known proteins (Table 3.3). These SNPs could provide insights into SNPs that could be responsible for different gene variants as well as identifying genes in *H. midae*. Arginine kinase was one of the proteins that were shown to be associated to *SNP3129_923*. This protein is involved in the maintenance of ATP levels in invertebrate muscle. It has been shown that arginine kinase binds actin with a reasonable affinity in molluscs (Reddy *et al.* 1992). This protein could be potentially valuable if it was shown that arginine kinase influences the mass of the epipodium, thus leading to greater abalone meat yield. The organism that the SNP was identified in was another abalone species *H. diversicolor* (Table 3.3), for which recently a microsatellite linkage map was developed (Zhan *et al.* 2011). The SNP resources for this species is quite limited (Kang *et al.* 2010), but as more SNPs are identified for this species, more comparative work can be conducted between these two abalone species. *SNP4691_183* and *SNP149.2_165* both showed similarity to heat shock protein 70 (HSP70). This protein is shown to be

Chapter two-Type 1 microsatellite development

responsible for adaptive changes in organisms exposed to stress, by causing refolding of proteins. HSP70 could be a valuable protein for *H. midae* as certain environmental stressors can effect abalone reproduction and growth (Roux 2011).

SNPs usually contain less genotyping errors as their genotyping is simplified by being bi-allelic and not size-based. The numbers of SNPs that can potentially be developed for *H. midae* will contain much more information than microsatellites. Research focusing on developing a denser linkage map for *H. midae* incorporating gene-linked SNPs is currently underway. This map will be constructed to facilitate QTL identification as it is hoped that the integrated microsatellite and SNP map containing a higher density of SNP markers will have a map density sufficient for this goal. It is furthermore advised to first map SNPs and then to sequentially add the microsatellites to the map as adding the error prone microsatellites first will introduce errors into the map from the beginning (Groenen *et al.* 2009).

5.4 Conclusion

Abalone's highly fecund nature leads to the generation of large numbers of offspring, making the use of large full-sib populations possible for linkage mapping. However, these full-sib families are also outbred populations, which are not the most desirable populations when it comes to linkage mapping as the segregation of alleles are not as predictable as inbred lines. Inbred (for example backcross or F2 populations) constructed lines are more suitable for linkage analysis as the genotypes of these lines are more homozygous, making segregation analysis easier as certain genotypes are expected in the offspring. It has to be noted that these types of populations are very challenging to construct for aquaculture species due to their method of reproduction, as well as their long generation time. Even though this is a second-generation linkage map, concerning microsatellite markers, this is the first linkage map to include SNPs for *H. midae*. The map coverage remains low at approximately 65%, even though the map coverage was improved from the previous linkage map for *H. midae* (50%; Hepple 2010) with the addition of more informative markers. However, different families were used for the first microsatellite-based linkage map (Hepple 2010) and thus these two maps are not directly comparable as fewer markers in Hepple (2010) could have been informative in the four new families.

Chapter two-Type 1 microsatellite development

Lehmensiek *et al.* (2005) carried out a curation study, in which the linkage map for wheat was reassessed to address issues with marker order and placement. They found that although many factors may play a role in marker order discrepancies, it was the effects of missing data on marker order and placement that was most worrying. It was advised that linkage maps be thoroughly curated as this can improve QTL mapping (Lehmensiek *et al.* 2005). The effects of missing data as well as genotyping errors and segregation distortion on linkage mapping has been described for several species (Dodds *et al.* 2004; Lehmensiek *et al.* 2005; Pompanon *et al.* 2005; Cartwright *et al.* 2007).

The low genome coverage and marker order inconsistencies found in this linkage map shows the need for more polymorphic markers, which spans a greater area of the genome to be developed for *H. midae*. In this instance, SNPs will play a greater role in future mapping projects for this species. A denser linkage map containing more SNPs will greatly improve the marker density of the map so that QTL identification can be facilitated.

Chapter four

Conclusions and future applications

Chapter four- Conclusions and future applications

1. Microsatellite development

The development of microsatellites has until recently been mostly performed by clone-based methods such as FIASCO (Zane *et al.* 2002), which is a time consuming and expensive method. Recent technological advances in sequencing have made it possible to create large sequence data sets for many species. The use of next generation sequencing, such as the Illumina Genome Analyser II, produces thousands of sequences that can be used for example in marker mining. This provides an avenue for faster and cheaper marker development.

Markers developed from genic sequences such as ESTs or transcriptomes provide an advantage over type 2 markers in that they are linked to a potential functional aspect of the genome such as a gene, and can be used in comparative studies, candidate gene discovery and gene expression studies. An advantage of EST-SSRs is that these markers enhance QTL identification by being linked to genes. This makes transcriptome sequencing a vital tool for developing type 1 markers, including microsatellite and SNP markers (Franchini *et al.* 2011).

Microsatellites possess a range of properties that make them useful for a variety of genetic applications. They are highly polymorphic and have co-dominant inheritance as well as being dispersed throughout the entire genome. They can also be used in multiplex reactions for automated genotyping (Chistiakov *et al.* 2006). This makes microsatellite markers an ideal marker type for linkage mapping, with genetic maps available for a variety of aquaculture species (Liu and Cordes 2004). Even though microsatellites are well suited for linkage mapping, certain errors can occur while genotyping these loci. Their size-based nature makes them prone to stutter bands, which complicate the genotype scoring process. They are also known for null alleles and duplicated loci, which introduce errors into the data set and cannot always be identified. Segregation analysis of the markers identifies which markers give non-reliable genotypes or contain null alleles. This information can be useful when characterizing the markers in wild populations as null-alleles and abnormal inheritance patterns are often missed in these situations (Hepple 2010).

Chapter four- Conclusions and future applications

SNPs also follow Mendelian inheritance and have a lower genotype error rate associated to them than microsatellites. However, they have lower polymorphic information content due to their bi-allelic nature, which means that the equivalent number of SNPs will yield a much lower number of information than microsatellites. This is easily solved as SNPs occur more frequently in genomes than microsatellite markers. Their sheer numbers will also lead to higher density maps than those seen for microsatellite markers. As with microsatellite markers, SNPs can be developed from ESTs or transcriptome data to generate type 1 markers (Wu *et al* 2008).

The aim of this study to develop polymorphic microsatellite markers from the sequenced transcriptome of *H. midae* was achieved with the development of 26 gene-linked polymorphic microsatellites *via* this method.

2. Linkage mapping in *H. midae*

The haploid chromosome number for *H. midae* has been confirmed to be $x=18$. A sufficiently dense linkage map would comprise of 18 linkage groups when linkage analysis is performed. Maps comprising of an insufficient number of molecular markers would comprise of more or less than the expected number of linkage groups. This is because the markers would not span all the chromosomes or that the gaps between the markers are too large to map certain markers on one linkage group. This is the third linkage map constructed for *H. midae*. The first map was constructed with mainly AFLPs and the second with microsatellites exclusively. This is the first linkage map to incorporate SNPs with microsatellites into a linkage map for *H. midae*. The map constructed in the current study aimed to increase the density of the available microsatellite map by including more mapping families as well as more markers for linkage analysis. Map coverage was improved from 50% to 65%.

The linkage map in the current study was constructed using four full-sib families and 295 polymorphic markers that were available for analysis. Many of these markers were excluded as they had multiple alleles, null alleles, non-reliable genotypes or many missing data that caused errors in the data set. About 65% of the remaining markers could be mapped to a specific linkage group as some markers remained ungrouped and unlinked ($LOD < 3$). This loss in markers, due to limited recombination

Chapter four- Conclusions and future applications

information, was noted by Hepple (2010) and it was suggested that more families be used for map construction. The use of more families for mapping increases the number of markers that can potentially be mapped as not all markers are informative across all families.

The sex-specific maps of the preliminary AFLP map were 12 for the maternal map and 10 for the paternal map (Badenhorst 2008). This is much lower than the haploid chromosome number of *H. midae*. Linkage groups of the sex-specific maps in the current study ranged from 13-17 linkage groups and for the sex-average maps it was 17-19 linkage groups. This is more or less the same as for the linkage map constructed by Hepple (2010), indicating that the maps are starting to converge to the correct linkage group number for *H. midae* (18). The number of markers per linkage group has improved from the previous linkage maps from 3.5 in the preliminary microsatellite map (Hepple 2010) and 3.1 in the preliminary AFLP map (Badenhorst 2008) to 4.5. This indicates that the marker density has increased in this study compared to the other linkage maps constructed for *H. midae*.

The genome length of the integrated map was 1683.62cM with an average marker spacing of 9.3cM. The genome length was a little bit lower than that calculated for the map constructed by Hepple (2010), but the average marker spacing was improved. This marker density is sufficient for QTL analysis as the recommended marker spacing for such projects is advised to range from 10-15cM for outbred populations (Massault *et al.* 2008). It has to be noted, however, that this marker spacing is localised and that regions in the map exist with much lower marker densities. The genome coverage of the integrated map in the current study was improved from the preliminary microsatellite map. The integrated map had a genome coverage of 65%, whereas the preliminary microsatellite and preliminary AFLP maps had a coverage of 50%. This increase in density is very promising for future QTL studies, but a much denser map is needed for correct marker-trait segregation analysis to be done accurately (Collard *et al.* 2005). More markers are still needed for linkage group- as well as linkage map integration with anchor markers present in all or most mapping families. It will also help to elucidate marker order discrepancies still seen between sex-average maps. This will help to integrate future maps with the

Chapter four- Conclusions and future applications

integrated map so that linkage groups can be joined to obtain the correct haploid number.

The aim of this study to develop sex-specific and sex-average maps for *H. midae* using four unrelated full-sib families comprising of approximately 100 offspring each and to merge the sex-average maps to produce an integrated map containing all of the four mapping families' mapping information was therefore achieved using the user-friendly mapping software, Joinmap[®] v4.

3. Future studies and improvements

Genetic linkage maps are valuable tools for studies focusing on the genetic improvement of aquaculture species as they facilitate QTL analysis and subsequent MAS. Microsatellite markers have been shown to be the best marker for map construction due to their multi-allelic nature and abundance, but SNPs are starting to supplement and even replace microsatellites as the marker of choice for map construction. This is mainly due to their relative abundance in comparison with microsatellites. However, the information contained in this type of marker is much lower compared to microsatellites due to the bi-allelic nature of SNPs. This is easily solved by using more markers and increasing the number of reference families used for map construction. Microsatellite markers also harbour some complications concerning genotyping as the size differences between the alleles causes stutter bands which makes correct allele calling difficult. Null alleles and allele drop-out also introduces errors into genotyping data. This causes problems with segregation analysis and leads to segregation distortion. These problems can only be identified in pedigree analysis and so this should be done before microsatellites are used in further studies. This makes SNPs more useful when looking at linkage analysis as the high-throughput methods used to genotype these markers are very robust and genotyping errors are much lower (Ball *et al.* 2010).

Genomic microsatellite markers are more polymorphic than genic microsatellite markers, but the functional value of genic microsatellites greatly exceeds those of genomic microsatellite markers. Type 1 markers can be developed from transcriptome or EST sequences. Using more mapping families when type 1 markers

Chapter four- Conclusions and future applications

are used for map construction will solve the problems associated the lower levels of polymorphism. This increases the amount of informative marker data for linkage map construction. It will also facilitate the merging and linking of linkage groups and maps to each other. The preliminary linkage map of *H. midae* (Hepple 2010) used two unrelated mapping families, which was an improvement to the one family used for the construction of the AFLP-based linkage map (Badenhorst 2008). Further increasing the number of mapping families will lead to even further improvement of map density and map merging.

The linkage map created in this study provides an integrated, denser linkage map with increased average marker spacing than the previous linkage map (Hepple 2010), which will help facilitate QTL analysis for traits that the farms deem important to production. This is the first linkage map for *H. midae* to contain SNPs. Much more SNPs should be included in future studies for *H. midae* as SNPs have a very low PIC, which hinders linkage mapping when conducted with a small number of SNPs. In future a greater number of SNPs should be mapped in more mapping families. The SNPs should also be mapped first and then the microsatellites should be added so that as few as possible errors are introduced into the linkage map. A linkage map with sufficient marker density will enable accurate QTL analysis to be conducted in *H. midae*, which can be used in subsequent marker-assisted selection for this species.

References

References

- 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
- Allsopp, M., Lafarga-De la Cruz, F., Flores-Aquilar, F., and Watts, E. (2011) Abalone Culture In: *Recent advances and new species in aquaculture*. R.K. Fortedar and B.F. Phillips (eds) Blackwell Publishing Ltd., Oxford. pp. 231-236
- An, H.S., and Han, S.J. (2006) Isolation and characterization of microsatellite DNA markers in the Pacific abalone, *Haliotis discus hannai*. *Molecular Ecology Resources* **6**: 11-13
- Andersen, J.P., and Lübberstedt, T. (2003) Functional markers in plants. *Trends in Plant Science* **8**: 554-560
- 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
- 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., Rourke, 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., Loughnan, 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
- Bester, A.E., Roodt-Wilding, R., and Whitaker, H.A. (2008) Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae*: a targeted EST approach. *Animal Genetics* **39**: 321-324
- 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
- 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
- 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
- Britz, P.J., Lee, B. and Botes, L. 2009. Aquaculture Institute of South Africa (AISA) Aquaculture Benchmarking Survey: Primary Production and Markets. AISA report produced by Enviro-Fish Africa (Pty) Ltd. [Available at <http://www.soundinteraxions.co.za/2009AISABenchmarkingSurveyFINAL.pdf>, accessed June 2011] pp 12-20
- Bürgener, M. (2010) Evaluation of CITES appendix III listing and delisting of South African abalone *Haliotis midae*. *TRAFFIC Bulletin* **23**: 42-48
- Buschiazzo, E., and Gemmell, N.J. (2006) The rise, fall and renaissance of microsatellites in eukaryote genomes. *BioEssays* **28**: 1040-1050

- 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
- 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
- Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **100**: 713-722
- Ciobanu, 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*. *Aquaculture* **220**: 203-218
- Conner, S.J., and Hughes, D.C. (2003). Analysis of fish ZP1/ZPB homologous genes -evidence for both genome duplication and species-specific amplification models of evolution. *Reproduction* **126**: 347-352.
- 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*
- Cruz, P., Imbarra, M., Fiore-Amaral, G., Galindo-Sanchez, C.E., and Mendoza-Carrion, G. (2005) Isolation of microsatellite loci in green abalone (*Haliotis fulgens*) and cross-species amplification in two other North American red (*Haliotis rufescens*) and pink (*Haliotis corrugata*) abalones. *Molecular Ecology Resources* **5**: 857-859
- Dahle, G., Johansen, T., Vasemägi, A., Carvalho, G., Florin, A., Was, A., Prödohl, P., and Khrustaleva, A.M. (2008) Review the potential for application of SNP's (single nucleotide polymorphisms) in fisheries genetics and aquaculture. *Report of the Working Group on the Application of Genetics in Fisheries and Mariculture (WGAGFM)* pp. 4-16
- 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*. Zhanjiang (John) Liu (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

DEAT (2007) Press Release by the Minister of DEAT on 25 March 2007. Suspension of Wild Abalone Commercial Fishing. Department of Environmental Affairs and Tourism, Pretoria, South Africa.

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* **10**: 1048-1058

Dere, R., and Wells, R.D. (2006) DM2 CCTG•CAGG repeats are hotspots that are more prone to expansions than the DM1 CTG•CAG repeats in *Escherichia coli*. *Journal of Molecular Biology* **360**: 21-36

Diaz-Viloria, N., Perez-Enriques, R., Fiore-Amaral, G., Burton, R.S., and Cruz, P. (2008) Isolation and cross-amplification of microsatellites in pink abalone (*Haliotis corrugata*). *Molecular Ecology Notes* **8**: 701-703

Dichmont, C.M., Buterworth, D.S., and Cochrane, K.L. (2000) Towards adaptive approaches to management of the South African Abalone *Haliotis midae* Fishery. *South African Journal of Marine Science* **22**: 33-42

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

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

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

Evans, B., White, R.W.G., and Elliot, N.G. (2000) Characterization of microsatellite loci in the Australian Blacklip abalone (*Haliotis rubra*, Leach). *Molecular Ecology Resources* **9**: 1183-1184

- Fallu, R. (1991) *Abalone farming*. Oxford: Fishing News Books. A division of Blackwell Scientific Publications Ltd. pp. 1-120
- FAO (2010) The State of World Fisheries and Aquaculture 2008 (SOFIA). FAO Fisheries and Aquaculture Department. Food and Agriculture Organisation of the United Nations. [Available at <http://www.fao.org/docrep/013/i1820e/i1820e00.htm>, accessed June 2011] pp. 3-87
- 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., Paternello, T., Power, D.M., Katoulas, G., and Bargolloni, L. (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics* **174**: 851–861
- Franchini, P., Van der Merwe, M., and Roodt-Wilding, R. (2011) Transcriptome characterization of the South African abalone *Haliotis midae* using sequence-by-synthesis. *BMC Research Notes* **4**: 59
- Fraser, L.G., McNeillage, M.A., Tsang, G.K., Harvey, C.F., and De Silva, H.N. (2005) Cross-species amplification of microsatellite loci within the dioecious, polyploid genus *Actinidia* (Actinidiaceae). *Theoretical and Applied Genetics* **112**: 149-157
- 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
- Geiger, D.L. (2000) Distribution and biogeography of the Haliotidae (Gastropoda: Vestigastropoda) world-wide. *Bolletino Malacologia* **35**: 57-120
- 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

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

Groenen, M.A.M., Wahlberg, P., Foglio, M., Cheng, H.H., Megans, H., Crooijmans, R.P.M.A., Besnier, F., Lathrop, M., Muir, W.M., Wong, G.K., Gut, I., and Andersson, L. (2009) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research* **19**: 510-519

Guo, X., Li, Q., Wang, Q.Z., and Kong, L.F. (2011) Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Marine Biotechnology* DOI 10.1007/s10126-011-9405-4

Gupta, P.K, Roy, J.K., and Prasad, M. (2001) Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science* **80**: 524-535

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H.S. (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* **270**: 315-323

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

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
- Hauser, L., Adcock, G.J., Smith, P.J., Bernal Ramírez, J.H., and Carvalho, G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences, USA* **99**: 11742-11747
- 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., 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
- He, C., Chen, L., Simmons, M., Li, P., Kim, S., and Liu, Z.J. (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics* **34**: 445-448
- He, F., Wen, H.S., Dong, S.L., Shi, B., Chen, C.F., Wang, L.S., Yao, J., Mu, X.J., and Zhou, Y.G. (2008) Identification of single nucleotide polymorphism cytochrome P450-c19a and its relation to reproductive traits in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **279**: 177-181
- Hepple, J. (2010) An integrated linkage map of Perlemoen (*Haliotis midae*). [Unpublished Master of Science thesis] Stellenbosch University, South Africa.
- Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, G.I., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Steward, 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
- 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
- Huchette, S., Soulard, J.P., Koh, C.S., and Day, R.W. (2004). Maternal variability in the blacklip abalone, *Haliotis rubra* leach (Mollusca: Gastropoda): effect of egg size on fertilisation success. *Aquaculture* **231**: 181-195
- Huehn, M. (2010) Random variability of map distances based on Kosambi's and Haldane's mapping functions. *Journal of Applied Genetics* **51**: 27-31
- 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 soy bean whole genome sequence. *BMC Genomics* **11**: 38
- Hyten, D.L., Song, Q., Choi, I-K., Yoon, M-S., Specht, J.E., Matukumalli, L.K., Nelson, R.L., Shoemaker, R.C., Young, N.D., Gregan, P.B. (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theoretical Applied Genetics* **116**: 945-952
- Jansen, J., De Jong, A.G., and Van Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theoretical and Applied Genetics* **102**: 1113-1122
- Jarne, P., David, P., and Viard, F. (1998) Microsatellites, transposable elements and the X chromosome. *Molecular and Biology Evolution* **15**: 28-34
- Jerry, D.R., Preston, N.P., Crocos, P.J., Keys, S., Meadows, J.R.S., and Li, Y. (2004) Parentage determination of Kuruma shrimp *Penaeus (Marsupenaeus) japonicus* using microsatellite markers (Bate). *Aquaculture* **235**: 237-247
- Joshi, H.C., and Cleveland, D.W. (1990) Diversity among tubulin subunits: toward what functional end? *Cell Motility and the Cytoskeleton* **16**: 159-163

- Kang, J.H., Appleyard, S.A., Elliot, N.G., Jee, Y.J., Lee, J.B., Kang, S.W., Baek, M.K., Han, Y.S., Choi, T.J., and Lee, Y.S. (2010) Development of genetic markers in abalone through construction of a SNP database. *Animal Genetics* **42**: 309-315
- 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
- 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
- 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
- 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 Wiley and Sons, Inc. pp. 181-186
- 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., Streeleman, 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
- Lee, H., Cha, S., Lee, M-S., Cho, G.J., Choi, W.S., and Suk, K. (2003) Role of antiproliferative B cell translocation gene-1 as an apoptotic sensitizer in activation-induced cell death of brain microglia. *Journal of Immunology* **171**: 5802-5811
- 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
- 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, H., Liu, W., Gao, X., Zhu, D., Wang, J., Li, Y., and He, C. (2010) Identification of host-defense genes and development of microsatellite markers from ESTs of hard clam *Meretrix meretrix*. *Molecular Biology Reports* **38**: 769-775
- Li, L., and Guo, X. (2004) AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology* **6**: 26-36
- Li, L., Guo, X., and Zhang, G. (2009) Inheritance of 15 microsatellites in the Pacific oyster *Crassostrea gigas*: segregation and null allele identification for linkage analysis. *Chinese Journal of Oceanology and Limnology* **27**: 74-79
- Li, Q., Shu, J., Zhao, C., Liu S., Kong, L., and Zheng, X. (2010) Characterisation of genic microsatellite markers derived from expressed sequence tags in Pacific abalone (*Haliotis discus hannai*). *Chinese Journal of Oceanology and Limnology* **28**: 46-54
- Li, Q., Park, C., Endo, T., and Kijima, A. (2004) Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (*Haliotis discus hannai*). *Aquaculture* **235**: 207-222

- Li, W-H., and Sadler L.A. (1991) Low nucleotide diversity in man. *Genetics* **129**: 513-523
- 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
- Liou, H.C., Boothby, M.C., Finn, P.W., Davidon, R., Nabavi, N., Zeleznik-Le, N.J., Ting, J.P., and Glimcher, L.J. (1990) A new member of the leucine zipper class of proteins that binds to the HLA DR alpha promoter. *Science* **247**: 1581-1584
- Litt, M., and Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**: 397-401.
- Liu, X, Liu, X., and Zhang, G. (2007) Identification of quantitative trait loci for growth-related traits in the Pacific abalone *Haliotis discus hannai* Ino. *Aquaculture Research* **38**: 789-797
- 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
- Lo Presti, R., Lisa, C., and Di Stasio, L. (2009) Molecular genetics in aquaculture. *Italian Journal of Animal Science* **8**: 299-313
- 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

- Malosetti, M., van Eeuwijk, F.A., Boer, M.P., Casas, A.M., Elía, M., Moralejo, M., Bhat, P.R., Ramsay, L., and Molina-Cano, J. (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. *Theoretical and Applied Genetics* **122**: 1605-1616
- Maneeruttanarungroj, C., Pongsomboon, S., Wuthisuthimethavee, S., Klinbunga, S., Wilson, K.J., Swan, J., Li, J., Wahn, V., Chu, K-H., Li, C.P., Tong, J., Glenn, K., Rothchild, M., Jerry, D., and Tassanakajon, A. (2006) Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). *Animal Genetics* **37**: 363-368
- 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
- Maretto, F., Reffo, E., Dalvit, C., Barcaccia, G, and Mantovani, R. (2010) Finding 16S rRNA gene-based SNPs for the genetic traceability of commercial species belonging to Gadiformes. *Italian Journal of Animal Science* **6**: 161-163
- Massault, C., Bovenhuis, H., Haley, C., and De Koning, D. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285**: 23-29
- McAndrew, B., and Napier, J. (2011) Application of genetics and genomics to aquaculture development: current and future directions. *Journal of Agricultural Science* **149**: 143-151
- 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
- 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
- Ning, Y., Liu, X., Wang, Z.Y., Guo, W., Li, Y., and Xie, F. (2007) A genetic map of large yellow croaker *Pseudosciaena crocea*. *Aquaculture* **264**: 16-26
- 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
- O'Connell, M., and Wright, J.M (1997) Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* **7**: 331-363
- 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
- Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. 2002. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques Suppl* :56-61
- Palti, Y., Genit, C., Luo, M., Charlet, A., Gao, G., Hu, Y., Castaño-Sánchez, C., Tabet-Canale, K., Krieg, F., Yao, J., Vallejou, R.L., and Rexroad, C.E. (2011) A first generation integrated map of the rainbow trout genome. *BMC Genomics* **12**: 180
- Parchman, T.L., Geist, K.S., Grahnen, J.A., Benkman, C.W., and Buerkle, C.A. (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* **11**:180
- Pavy, N., Pelgas, B., Beauseigle, S., Blais, S., Gagnon, F., Gosselin, I., Lamothe, M., Isabel, N., and Bousquet, J. (2008) Enhancing genetic mapping of complex

genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* **9**: 21

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

Poompuang, S., and Na-Nakorn, U. (2004) A preliminary genetic map of walking catfish (*Clarias macrocephalus*). *Aquaculture* **232**: 195-203

Primmer, C.R., and Ellegren, H. (1998) Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* **15**: 997-1008

Qi, H., Liu, X., and Zhang, G. (2008) Characterization of 12 single nucleotide polymorphisms (SNPs) in Pacific abalone, *Haliotis discus hannai*. *Molecular Ecology Resources* **8**: 974-976

Qi, H., Liu, X., Wu, F., and Zhang, G. (2010) Development of gene-targeted SNP markers for genomic mapping in Pacific abalone *Haliotis discus hannai* Ino. *Molecular Biology Reports* **37**: 3779-3784

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

Rampling L.R., Harker N., Shariflou M.R., Morell M.K. (2001) Detection and analysis systems for microsatellite markers in wheat. *Australian Journal of Agricultural Research* **52**: 1131–1141.

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

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.

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

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

Rose, O., and Falush, D. (1998) A threshold size for microsatellite expansion. *Molecular and Biology Evolution* **15**: 613-615

Roux, A. (2011) Reproduction of the South African abalone, *Haliotis midae*. [Unpublished Doctor of Philosophy dissertation] Stellenbosch University, South Africa

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

Sainudiin, R., Durrett, R.T., Aquadro, C.F., and Nielsen, R. (2004) Microsatellite mutation models: Insights from a comparison of humans and chimpanzees. *Genetics* **168**: 383-395

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

Sekar, M., Suresh, E., Kumar, N.S., Nayak, S.K. and Balakrishna, C. (2009) Microsatellite DNA markers, a fisheries perspective Part 1: The nature of microsatellites. *Aquaculture Asia Magazine* **14**: 27-29

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

- 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
- Sekino, M., Saido, T., Fujita, T., Kobayashi, T., and Takami, T. (2005) Microsatellite DNA markers of Ezo abalone (*Haliotis discus hannai*): a preliminary assessment of natural populations sampled from heavily stocked areas. *Aquaculture* **243**: 33-47
- Serapion, J., Kucuktas, H., Feng, J., and Liu, Z.J. (2004) Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Marine Biotechnology* **6**: 364-377
- Sha, Z., Wang, S., Zhuang, Z., Wang, Q., Wang, Q., Li, P., Ding, H., Wang, N., Liu, Z., and Chen, S. (2010) Generation and analysis of 10 000 ESTs from the half-smooth tongue sole *Cynoglossus semilaevis* and identification of microsatellite and SNP markers. *Molecular Ecology Resources* **76**: 1190-1204
- Shendure, J., and Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**: 1135-1145
- Shi, Y., Guo, X., Gu, Z., Wang, A., and Wang, Y. (2010) Preliminary genetic linkage map of the abalone *Haliotis diversicolor* Reeve. *Chinese Journal of Oceanology and Limnology* **28**: 549-557
- Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B., and Postlethwait, J. (2002) Sex-specific recombination rates in Zebrafish (*Danio rerio*). *Genetics* **160**: 649-657
- 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 dissertation] 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
- Smit, A.F.A., Hubley, R., and Green, P. RepeatMasker Open-3.0. 1996-2010 <<http://www.repeatmasker.org>>.
- Smith, J.A., Turner, M.J., DeLay, M.L., Klenk, E.I., Sowders, D.P., and Colbert, R.A. (2008) Endoplasmic reticulum stress and the unfolded protein response are linked to synergistic IFN- α induction via XBP-1. *European Journal of Immunology* **38**: 1194-1203
- Stam, P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* **3**: 739-744
- Steinberg, J. (2005) The illicit abalone trade in South Africa. *Institute for Security Studies Occasional Papers* **105**: 1-16
- 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
- Tarr, R.J.Q. (1989) Abalone. In: *Oceans of life off southern Africa*. A.I.L. Payne, and R.J.M. Crawford (eds). Vlaeberg Publishers, Cape Town. pp. 62-69

- Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471
- Thomson, J.D., Higgins, D.G., and Gibson, T.J. (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequences weighting, position-specific gap penalties and weight matrix choice *Nucleic Acids Research* **22**: 4673-4680
- Troell, M., Robertson-Andersson, D., Anderson, R.J., Bolton, J.J., Maneveldt, G., Halling, C., and Probyn, T. (2006) Abalone farming in South Africa: An overview with perspectives on kelp resources, abalone feed, potential for on-farm seaweed production and socio-economic importance. *Aquaculture* **257**: 266-281
- 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., and Roodt-Wilding, R. (2008) Chromosome number of the South African abalone *Haliotis midae*. *African Journal of Marine Science* **30**: 195-198
- Van der Merwe, M., Franchini, P., and Roodt-Wilding, R. (2011) Differential growth-related gene expression in Abalone (*Haliotis midae*). *Marine Biotechnology* **13**: 1125-1139
- 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 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
- Varshney, R.K., Graner, A., and Sorrells, M.E. (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48-55

- Vasemägi, A., Nilsson, J., and Primmer, C.R. (2005) Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species. *Molecular Ecology Notes* **5**: 282-288
- Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**: 275-305
- Vogiatzi, E., Lagnel, J., Pakaki, V., Louro, B., Canario, A.V.M., Reinhardt, R., Kotoulas, G., Magoulas, A., and Tsigenopoulos, C.S. (2011) *In silico* mining and characterization of simple sequence repeats from gilthead sea bream (*Sparus aurata*) expressed sequence tags (EST-SSRs); PCR amplification, polymorphism evaluation and multiplexing and cross-species assays. *Marine Genomics* **4**: 83-91
- Voorrips, R.E. (2002) MapChart: software for the graphical presentation of linkage maps and QTL. *The Journal of Heredity* **93**: 77-78
- 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
- Wang, H., Li, F., and Xiang, J. (2005) Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. *Aquaculture* **249**: 107-114
- Wang, S., Sha, Z., Sonstegard, T.S., Liu, H., Xu, P., Somridhivej, B., Peatman, E., Kucuktas, H., and Liu, Z. (2008) Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics* **9**: 40
- 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
- Weber, J.L., and Wong, K. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2**: 1123-1128

- 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
- Whittaker, J.C., Harbord, R.M., Boxall, N., Mackay, I., Dawson, G., and Sibly, R.M. (2003) Likelihood-based estimation of microsatellite mutation rates. *Genetics* **164**: 781-787
- Wood, A., and Buxton, C. (1996) Aspects of the biology of the abalone *Haliotis midae* on the east coast of South Africa. *South African Journal of Marine Science* **17**: 69-78
- 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
- Wu, R., Ma, C-X., and Casella, G. (2007) *Statistical Genetics of Quantitative Traits*. Springer, U.S.A. pp. 25-71
- Wu, S-B., Wirthensohn, M.G., Hunt, P., Gibson, J.P., and Sedgley, M. (2008) High resolution melting analysis of almond SNPs derived from ESTs. *Theoretical and Applied Genetics* **118**: 1-14
- 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
- Xu, Z., Dhar, A.K., Wyrzykowski, J., and Alcivar-Warren, A. (1999) Identification of abundant and informative microsatellites from shrimp (*Penaeus monodon*) genome. *Animal Genetics* **30**: 150-156
- 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

- Yu, H., and Li, Q. (2007) EST-SSR markers from the Pacific oyster, *Crassostrea gigas*. *Molecular Ecology Notes* **7**: 860-862
- Yu, H., and Li, Q. (2008) Exploiting EST databases for the development and characterization of EST-SSRs in the Pacific oyster (*Crassostrea gigas*). *Journal of Heredity* **99**: 208-214
- Yue, G.H., Ho, M.J., Orban, L., and Komen, J. (2004) Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* **234**: 85-98
- Yue, H., Yuan, H., and Zhang, X. (2009) Fifteen novel polymorphic microsatellites in rock carp, *Procypris rabaudi* (Tchang), an endemic fish species in the upper reaches of the Yangtze River drainage. *Conservation Genetics* **10**: 539-542
- Zane, L., Bargelloni, L., and Patarnello, T. (2002) Strategies for microsatellite isolation: A review. *Molecular Ecology* **11**: 1-16
- Zerbino, D.R., and Birney, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*: **18**: 821-829
- Zhan, A., Bao, Z., Hu, X., Wang, S., Peng, W., Wang, M., Hu, J., Linag, C., and Yue, Z. (2008a) Characterisation of 95 novel microsatellite markers for Zhikong scallop *Chlamys farreri* using FIASCO-colony hybridisation and EST database mining. *Fisheries Science* **74**: 516-526
- Zhan, A., Bao, Z., Wang, M., Chang, D., Yuan, J., Wang, X., Hu, X., Liang, C., and Hu, J. (2008b) Development and characterisation of microsatellite markers for the Pacific abalone (*Haliotis discus*) via EST database mining. *Journal of Ocean University of China* **7**: 219-222
- Zhan, X., Fan, F., You, W., Yu, and Ke, C. (2011) Construction of an integrated map of *Haliotis diversicolor* using microsatellite markers. *Marine Biotechnology* DOI 10.1007/s10126-011-9390-7

References

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

Zhang, Z., Zhan, A., and Liu, X. (2010) A panel of genic single nucleotide polymorphism (SNP) markers for the Pacific abalone, *Haliotis discus hannai*. *Conservation Genetics Resources* **2**: 133-135

Appendices

Appendices

Appendix 1: Chemical reagents and electrophoresis gels

A: 5X TBE:

- 54g Tris
- 27.5g boric acid
- 3.725g EDTA
- Fill 1L with d.H₂O

B: 1X TBE:

- 100ml 10X TBE
- Fill to 1L with d.H₂O

C: 2% Agarose gel:

- 2g of agarose for every 100ml 1XTBE
- 0.05ng/μl of ethidium bromide

D: 12% Polyacrylamide gel:

- Polyacrylamide (12% m/v) 3ml
- 5X TBE 2ml
- ddH₂O 4.67ml
- APS (10% m/v) 300μl
- Temed 30μl
- 49 Acrylamide: 1 Bisacrylamide cross-linkage.
- Run at 150V for 1.5 hours.
- Staining: 0.05ng/μl ethidium bromide in 1X TBE solution.

Appendix 2: Marker information

Primer information

Table S1: Primer information on all microsatellite markers developed from Illumina sequence results.

Sequence Id	Orientation	Length	Ta	GC%	Sequence (5'-3')	Prod Size	Repeat Motif
NODE_112233	FORWARD	20	59.69	45	TCAAAAGGTGTCATGGCTTG	152	TGAG
NODE_112233	REVERSE	20	58.8	50	GTTGCCGCTCACAGGTATAA		
NODE_140858	FORWARD	21	59.71	47.62	TTCTAGATGTCGCCATGACCT	269	GTGA
NODE_140858	REVERSE	20	59.07	55	GGTGTCTTTGGCCAGGTAGT		
NODE_161284	FORWARD	20	60.51	55	CCAAGGGCATTACTGAGTGG	120	GATG
NODE_161284	REVERSE	20	59.91	45	TTTCGGTTGATTTGGAGGTC		
NODE_71359	FORWARD	20	60.06	50	GGTTTGCAATGACCCCATAC	117	GCAT
NODE_71359	REVERSE	20	60.35	50	TGTGCATCTATGCGTGTGTG		
NODE_87955	FORWARD	20	60.23	50	ATTCTCAGCTGCATGGTTCC	274	GTGA
NODE_87955	REVERSE	20	60.56	55	CGAGCGTGCCTCTATTCTGT		
NODE_128551	FORWARD	20	60.08	60	CTGAGCTCTGCAGCTGTGAC	118	TGGA
NODE_128551	REVERSE	21	60.84	52.38	GCTCTGTTCCGGTTGTATGCTG		
NODE_98293a	FORWARD	20	59.19	45	TGCAGTGTGTTTGTGG	109	TTTG
NODE_98293a	REVERSE	18	59.78	50	AAACGGGTTGGGTTGTCA		
NODE_68575	FORWARD	20	59.52	50	CTAAGGCGCTGCAAAGGTAT	110	TCAC
NODE_68575	REVERSE	20	59.5	45	TGTTGTGAGCATCATTTCC		
NODE_104411	FORWARD	25	61.03	40	TTGTCTGTGTTGCACATATATACCG	200	CA
NODE_104411	REVERSE	25	62.4	40	AGGGATGGATACAAATGTCAGAATG		
NODE_118779	FORWARD	25	60.21	40	AAGATTCCTCCCTGAATAGAATGTC	242	AT
NODE_118779	REVERSE	25	60.23	40	GTGTATCTCAATAATGAGCGAATCC		
NODE_126949	FORWARD	25	61.45	48	GGATACACACCTACACACACTCGAT	320	CA

NODE_126949	REVERSE	24	60.39	41.67	GTATGTGTTTCGTACGTGTTTGCTT		
NODE_31121	FORWARD	25	60.18	40	CTGCTAATCATCCTAAGTGCAGAAT	208	AT
NODE_31121	REVERSE	25	60.34	40	GTATGCATAGTGTGTTGTTCTTCCA		
NODE_47613	FORWARD	25	59.53	40	GTGTTTACAAGGCGTCATATCAGTA	216	AC
NODE_47613	REVERSE	23	60.05	43.48	CCTGCGCCTACTTACAATAAATG		
NODE_50505	FORWARD	25	60.13	40	CCATCATAGTTTCAGTCCCATAAAG	214	TA
NODE_50505	REVERSE	24	60.62	41.67	CAGCCCAACTGTACCATAAATTCT		
NODE_50652	FORWARD	25	60.92	44	CTTCTTTGTGCGTAAGTACAGGTGA	283	AC
NODE_50652	REVERSE	24	60.72	45.83	GTATGTGTGTGTGCGTGTATGTGT		
NODE_52703	FORWARD	25	59.94	40	ATCAAAGTTCTCTCTCTTTGTCCA	190	AT
NODE_52703	REVERSE	25	61.04	40	GTCTCTCCGTTGGAATATTCTTGTT		
NODE_70036	FORWARD	26	60.22	42.31	TCTACATCTAAACTGGCATCTACAGC	215	TG
NODE_70036	REVERSE	25	60.32	40	CATAAACTGGTTTAACTGCTGGAC		
NODE_7020	FORWARD	25	59.5	40	ACCCACAATACGGCAGTATAATATC	205	GT
NODE_7020	REVERSE	25	60.11	44	CGAGAATCCGTTACATAGCTACACT		
NODE_76149	FORWARD	25	59.2	40	GGAATGGGCTAGTAAGTAGTTCAA	216	AT
NODE_76149	REVERSE	25	59.38	40	ACTAAATGTTTCAGATGTGTGGTGAC		
NODE_8276	FORWARD	25	59.74	40	TATGGAGTCTCTGACAAAACAGTTG	196	AT
NODE_8276	REVERSE	25	59.73	40	GCATCTTACGGTATGAGGTGTATTT		
NODE_86892	FORWARD	25	60.69	44	GTCCAGATCCTCTGTAACACACATT	200	CA
NODE_86892	REVERSE	25	60.66	40	ACAGATACATTCCAAAGGACAAGTG		
NODE_98293b	FORWARD	22	60.17	45.45	AATATCTGCGCATCACTCACAC	270	CA
NODE_98293b	REVERSE	25	60.41	44	GGTCGTAGTAGTCGATTTCTGAATG		
NODE_29450	FORWARD	25	60.04	40	ATCTGCAGGAGTATTTCAAGTATTC	227	TG
NODE_29450	REVERSE	24	59.81	45.83	TATCACATAGAAGCGTCTCTGACC		

Appendices

NODE_46349	FORWARD	25	60.2	40	TCCATCAGTCTAACAGCTATCACAA	233	GT
NODE_46349	REVERSE	25	59.66	40	ATGGTCGAGCACTGTAAATCTACTT		
NODE_124227	FORWARD	25	59.71	44	GGTGTGATGGACACAGTGATATTAG	193	TA
NODE_124227	REVERSE	25	60.57	48	GCAGGTATTCTGTTCCACCAGTAGAG		
NODE_105204	FORWARD	25	60.07	44	GTACAGAGAATTAAGGACCGAGACA	173	AAG
NODE_105204	REVERSE	25	59.49	40	TCTCCTTCTTCCTTCTTTTCTACCT		
NODE_112066	FORWARD	24	60.16	41.67	AAGATTGACCTTCCTCACTTTCTG	275	GAA
NODE_112066	REVERSE	25	59.82	40	CAATTTCTTCTTCAGTCCAACCTCTC		
NODE_11646	FORWARD	24	60.05	45.83	GGTACAGAGAGATATTTGCGTCCT	208	CTC
NODE_11646	REVERSE	25	60.71	40	CACAGAAGAATCAGTCATCAAGTCA		
NODE_116595	FORWARD	25	60.27	40	CTGATCCTGTATTTCTTCAACAAC	205	TTC
NODE_116595	REVERSE	25	60.17	40	GAAACGAATTGAGACAGAGGTAGAA		
NODE_118613	FORWARD	23	60.98	47.83	CTAAACCATGGCTAAGAGGCAGT	212	GGT
NODE_118613	REVERSE	24	60.54	41.67	ACACTCATAGCATCGGTCTTCTTT		
NODE_122684	FORWARD	24	60.42	41.67	ATGTCAGCCAGGAAGTATTGAGTT	199	AAG
NODE_122684	REVERSE	25	59.64	40	ACACTGCATAAAGACACGATATGAG		
NODE_125907	FORWARD	24	60.17	45.83	GATCTGAGACATGTCATAGCGAAC	230	TGC
NODE_125907	REVERSE	25	59.35	44	CTGATCTTGGTCTGGACTCTCTAAT		
NODE_128274	FORWARD	24	60.05	50	GAGAAGAGAACTGAGGAAGTGAGG	260	GAA
NODE_128274	REVERSE	25	60.97	40	ATAAGTTTAGCTGTCTGCGTTTCTT		
NODE_128607	FORWARD	24	61.46	41.67	ATTCACATTATTGACCAGCCAGAC	235	GCA
NODE_128607	REVERSE	25	59.97	40	CATCACAGATGTGTAGCTAAATTGG		
NODE_129966	FORWARD	24	60.61	45.83	GTCAAATGGCAGAGAGAATAGGAG	197	AAG
NODE_129966	REVERSE	25	60.59	44	GAGCACAGTCAAAAAGTCCAGTTAC		
NODE_130590	FORWARD	22	60.8	40.91	ATGATGCAGCAGCAATCAAGTA	230	GTT

NODE_130590	REVERSE	25	59.98	40	ACAGATACACAGACAAATCCAGACA		
NODE_130822	FORWARD	23	60.33	43.48	GAAACATTTTCCCTGGACTTCTC	179	AAG
NODE_130822	REVERSE	25	60.17	40	AGCTCATCATATGTGTAATCCCTGT		
NODE_131826	FORWARD	25	60.66	40	GTCCCCAATAAACTATGTGGAGAAT	191	GAA
NODE_131826	REVERSE	25	59.49	40	CTTTCTGAATGTTCTCAGACTTCCT		
NODE_134845	FORWARD	25	61.45	40	GATAGATTTTGGAAACGAAGACAGG	257	AGG
NODE_134845	REVERSE	24	62.23	54.17	GTAGTGCAGGGGGTACTTTACCAG		
NODE_134916	FORWARD	25	61.15	40	ATAAATTCTGTGGAAAGTGCTCCTC	135	GAT
NODE_134916	REVERSE	24	60.17	41.67	CTCTTCAACGTCAGAAATGTCATC		
NODE_139279	FORWARD	25	59.37	40	GGCACAACATGTGATAATCTAGGTA	200	AAT
NODE_139279	REVERSE	25	60.21	40	TATTAGTGCCACTGAGCATCTGTAA		
NODE_154893	FORWARD	25	60.05	40	TAAACATCTGAAGAACCAGTCTCC	204	GCA
NODE_154893	REVERSE	24	60.46	41.67	CAGATGATGAAGACACAAGGTGTT		
NODE_20396	FORWARD	24	60.83	41.67	AAGAATGGTTCTCTCTTCCAGGTT	191	CCT
NODE_20396	REVERSE	23	59.96	43.48	CCAAACTCAACTTTACCGTCAAC		
NODE_20908	FORWARD	25	61.74	48	CTAAAGGAGAAGCTGGTATGACAGG	197	AGG
NODE_20908	REVERSE	23	61.03	43.48	GACATCATTCTGGATGTCCATGT		
NODE_20970	FORWARD	25	61.96	48	ATTGTCTCCCTCTACAGGATGTGAC	212	TCG
NODE_20970	REVERSE	25	61.24	40	GCTGTAGCGTAAACAAAACAACATC		
NODE_2192	FORWARD	25	60.02	40	TATCAACAGCTCTAGCAACACAAAG	228	CAT
NODE_2192	REVERSE	25	60.21	40	TCACTCAGCCTTGACATACACTTA		
NODE_23452	FORWARD	25	60.67	40	ATCTAATGACAAGGTTTCGTGTGAC	199	AAT
NODE_23452	REVERSE	25	59.88	40	CTGGCTGTAACGATTATTTACCAGT		
NODE_24872	FORWARD	25	59.5	48	GCTGCAGTAGTTGTAGTTGTAGTGG	208	GTG
NODE_24872	REVERSE	25	60.16	40	GCAGAAACGATAGGAGATACAGAAA		

Appendices

NODE_27253	FORWARD	25	60.95	40	AGCATCGAGTCCTGCTTTAGTTATT	204	ATT
NODE_27253	REVERSE	24	60.26	41.67	AGTATCTGGGATGACTTCGACATT		
NODE_36066	FORWARD	25	59.65	44	ATGTGGAGACCTTAGTTACAGCCTA	269	AGC
NODE_36066	REVERSE	25	59.77	40	TGCTGTGAGGTTTATACAGAAACAG		
NODE_36257	FORWARD	25	61.61	40	CCAATGGCATCATATTCTCATAAC	198	TCT
NODE_36257	REVERSE	24	60.45	41.67	AGACGTTACAGGCGAAGTTGTATT		
NODE_38396	FORWARD	25	60.01	40	ATATCCCAATGTGACGGAGATACTA	208	ATT
NODE_38396	REVERSE	25	59.69	44	GGAGAAACCTATACCGATAAGGAAC		
NODE_39658	FORWARD	25	61.11	40	GAGCCTTTTGTAAAGAGTGGGATTTA	191	CCA
NODE_39658	REVERSE	23	59.78	47.83	ATCTGGCTATGGAGACAGCTATG		
NODE_39227	FORWARD	23	60.13	52.17	CCCTACCCCTTATCTAGTTGGTG	209	CCT
NODE_39227	REVERSE	25	59.64	44	AGACTACAAAGTGACAGGTGGATCT		
NODE_39923	FORWARD	25	60.65	40	ATCTTGTCTAGCAGAGGTTGCAATA	176	TGT
NODE_39923	REVERSE	25	60.71	44	CAAGACCAACTGTAAGAAGTTCTGC		
NODE_44063	FORWARD	25	60.63	40	ATTATTCCACTGTTGGTGCTGATAG	219	CTC
NODE_44063	REVERSE	25	60.17	44	TCTTCTCTATCCTGAGTTTGGACAC		
NODE_47161	FORWARD	24	60.46	41.67	TAACTACATCGGGATTTCGTTTAC	189	AAT
NODE_47161	REVERSE	25	60.29	44	CACTGTCACTTTATACAGCTTGTGC		
NODE_50824	FORWARD	25	60.11	40	AAAAGCAGTTATGTTGACAGTCCTC	194	CAG
NODE_50824	REVERSE	24	61.18	50	CGACAGAACTGACACGTGAGTAAC		
NODE_51766	FORWARD	25	61.05	44	GGGAGAGCATATAACCTTCACTTGT	224	TCG
NODE_51766	REVERSE	25	59.37	40	TACCTTATGACACAGTGGATATTGC		
NODE_57236	FORWARD	24	61.48	41.67	ATTAATGCGAGCTGCTGTGTAAGT	162	ATC
NODE_57236	REVERSE	25	59.54	40	ACATTGGAACATATCCTCTACATCC		
NODE_60025	FORWARD	24	59.78	41.67	GGTCAGAAGGAGTTTATTGGTGAT	245	AGG

NODE_60025	REVERSE	25	59.71	40	ACCAGACTGATCTCAGATTTCAACT		
NODE_60863	FORWARD	25	60.59	40	TGTCAACATCAGCTAGTTTGTTCATC	209	TCA
NODE_60863	REVERSE	25	60.24	44	TATGATCAAGGAGAGCATAGAGGAG		
NODE_62696	FORWARD	25	59.15	40	ACATACAGAAACAGTGA CTCCGGATA	209	CTT
NODE_62696	REVERSE	25	60.35	44	CTGATTCATCAAGACAGTCAGTCAG		
NODE_6458	FORWARD	25	60.37	40	TCATTCTTCCAGTCAGATATCAGGT	189	CAT
NODE_6458	REVERSE	25	60.32	40	TGATGTTACTAGTCTTGCTCGATGA		
NODE_64699	FORWARD	25	60.81	40	TTGGCCTCAATAATGCTATGAGTAG	195	TGA
NODE_64699	REVERSE	25	60.28	40	CGTACAGCTGGTATGATGATGAATA		
NODE_66010a	FORWARD	25	60.48	44	GAGTACCAAAGGGAGATAACCAAAC	201	ATT
NODE_66010a	REVERSE	25	59.98	40	AAGAATGCCTACGATACTTCAACAC		
NODE_69910b	FORWARD	25	59.92	44	CCTACCCAACATCCATGTTATAGTC	325	ACA
NODE_69910b	REVERSE	23	59.97	47.83	GTCTTGTCTCAGCAGAACTTGGT		
NODE_71363	FORWARD	25	60.6	52	GAGTACTACAGACAGCAGGGTATGC	213	GCA
NODE_71363	REVERSE	24	60.96	45.83	GAGTGAATGTGTAGCCATCAAGTG		
NODE_79387	FORWARD	25	59.68	40	TGATGAAGAGAGAAAGAAGAAGGAG	197	AAG
NODE_79387	REVERSE	25	59.18	44	AGTACTGTCCCCATCTGAGTTAAG		
NODE_80333	FORWARD	25	60.29	44	CCTGAATATCAGAGTCGTTTACACC	178	CCT
NODE_80333	REVERSE	25	60.64	40	AGGACAAAAGAGGTACCAGATGTTT		
NODE_82292	FORWARD	22	60.15	50	GACCTGACCAAGAGGTTTGAAG	267	AAG
NODE_82292	REVERSE	25	58.93	40	GTCTTCTGTATTTACCAGTGGGATT		
NODE_89045	FORWARD	25	60.7	40	CGGAGGTCAATTCAAGTATATCAGA	229	CAT
NODE_89045	REVERSE	25	61.85	40	CCCAAATACCATTGTTAATCCTCAG		
NODE_94542	FORWARD	25	60	44	TGCTGTATCAGTCTCTCAA ACTCAC	230	GAG
NODE_94542	REVERSE	24	59.2	41.67	GACAACAACCACACACATGTTATC		

NODE_94556	FORWARD	25	59.66	44	CGTAGCCATCAGTACGTCTACAATA	221	CTC
NODE_94556	REVERSE	22	59.94	54.55	CAGAGGAAGCTGGAGTACTGGT		
NODE_95616	FORWARD	25	59.95	40	AAAGAACTGAAGAAGAGGGAGAAG	199	GAG
NODE_95616	REVERSE	25	59.86	48	GAGCCTCAGTTGTAGTATCCTTGAG		
NODE_71693	FORWARD	25	60.42	40	GATCACGTTTAAGAAACCTGTGAAG	167	AAG
NODE_71693	REVERSE	22	59.17	40.91	CCTCTTCTTCATCATCACCAA		
NODE_84315	FORWARD	24	60.5	45.83	ATATGTACGGACCGATCGTAGAAC	135	AGG
NODE_84315	REVERSE	25	60.25	40	AGGAAAGGCCATATGTCTTTAGAAC		
NODE_108706	FORWARD	23	60.34	43.48	AGGGCTTGATCTCAGATGGTAAT	128	ATC
NODE_108706	REVERSE	25	59.96	40	CAAGAAAATACCTGCAGAAGAAGAG		
NODE_125430	FORWARD	22	59.36	40.91	GCAGGAGTCTTCATTGAACAAA	207	ATC
NODE_125430	REVERSE	25	59.57	40	CTTAATTCTGCTTTAGTCTGCCTGT		
NODE_64121	FORWARD	25	60.48	40	GTGTTGGTTGTATTAATGTGTGTGC	299	CCT
NODE_64121	REVERSE	25	60.64	44	AAATAGAAAGTAGGAGGGAGGGAAG		
NODE_71352	FORWARD	25	60.71	40	GGATTTAGACCACAGCATTATCTA	196	ACT
NODE_71352	REVERSE	25	60.27	40	TGTCTGTCTGAAGCATGTTAGTGTT		
NODE_140027	FORWARD	25	59.68	40	TCCATTCCTTCATCTCTTACACTCT	202	TTC
NODE_140027	REVERSE	25	59.76	40	AACAAGAAGAACCCAGAACACTCTA		
NODE_72758	FORWARD	25	61.23	44	GAAGTACATGATGACGAAGATGAGG	262	CAT
NODE_72758	REVERSE	24	61.8	41.67	CGTGCTTCAACAACATGTCAATG		
NODE_93709a	FORWARD	25	61.04	40	AGAACAAGGATCCTATCAACGAAAC	266	AAG
NODE_93709a	REVERSE	25	60.31	40	ATTCTTCATCAGCTTGTTTCAGAGAC		
NODE_93709b	FORWARD	24	60.18	41.67	AATGCTAACAACCTTGAGAGGAAG	193	CCG
NODE_93709b	REVERSE	22	60.27	54.55	TCTCTCTTCTCAGAGCCTCCAC		
NODE_97931a	FORWARD	25	60.05	40	GAACAGAGACATTTGACCCTAAAGA	182	TGA

Appendices

NODE_97931a	REVERSE	25	59.19	40	TATAGACTTGCAAAGAGTTGTCTCG		
NODE_97931b	FORWARD	25	60.66	40	TAGTGGTCTGTGATGATTCATGGTA	201	GTT
NODE_97931b	REVERSE	25	60.25	48	ACACATCTGACCTCGCTAGTTACTC		
NODE_1000b	FORWARD	24	60.41	44	AAGACAATGAAGCTGAAGAGGAAG	410	GAT
NODE_1000b	REVERSE	25	60.46	40	GGCTAAAGTTCAAAGTTCAACACAC		
NODE_117200	FORWARD	25	60.6	40	AAGAGATTCCTCAAGAGAAAACAGG	406	GAT
NODE_117200	REVERSE	25	60.54	44	GCGTCTCGTTCATGAGTTAAGATAC		
NODE_24288a	FORWARD	25	59.55	44	AGAGTTATCAGAGCCACGTCTAAAG	193	CCT
NODE_24288a	REVERSE	23	60.15	52.17	CTGATGATAGTCAGGAGGACGAC		
NODE_24288b	FORWARD	25	60.97	40	CATCTTCATATGCCCTCCTAAATCT	213	CTT
NODE_24288b	REVERSE	25	60.73	40	AAGAACAGCCTCTGAAAGAAGAATC		
NODE_36269	FORWARD	22	60.53	50	GTCATCAGCCGTGATAGAACCT	204	(GAC)n(ACA)
NODE_36269	REVERSE	21	59.82	52.38	CGATGATTCAGAGGACGACTC		
NODE_36952	FORWARD	25	60.58	40	AGACCTATTCAGATCATTGGAACCT	307	(TGA)n(TGG)
NODE_36952	REVERSE	26	60.06	42.31	CTATAAAGATGGAAGGCTAGTTGAGG		
NODE_58101a	FORWARD	24	60.78	41.67	ACAAAGAAGGATGTTGGTGAAGAC	180	GAG
NODE_58101a	REVERSE	25	59.95	40	AGGACAATCTAGGTTTTCTTCTGGT		
NODE_58101b	FORWARD	25	59.95	40	AAGAAAACCTAGATTGTCCTGACCT	586	AAG
NODE_58101b	REVERSE	25	61.87	40	GGCCAACAGATCCATGTTAAAATAC		
NODE_64307	FORWARD	25	60.21	40	ATGAGGAAGAGAATGAGGATAAACC	202	AAG
NODE_64307	REVERSE	25	60.01	44	GATAAACACACGTCTCACATACAGC		
NODE_8738	FORWARD	25	60.78	40	ACATTCATAAGCTGGTCTTTACACG	204	(TGT)n(TGC)
NODE_8738	REVERSE	25	60.1	40	AAGTGACTTGCCTCATGTTATTAGC		
NODE_56762	FORWARD	25	60.49	40	AATGTCTGTCTTTTCCTCTTCTCC	366	(CCT)n(CTT)
NODE_56762	REVERSE	25	61.55	40	GAATACACTTTTGCAGTATCGGACA		

Optimised marker information:

Table S2: Marker information of all the microsatellite markers that could be optimised, the PCR cycle used, primer sequence, fluorescent label used for polymorphic loci and the motif.

Loci	Orientation	Ta	PCR cycle	Primer sequence (5'-3')	Motif
ILL2.140858	FORWARD	50°C	1	PET -TTCTAGATGTCGCCATGACCT	GTGA
	REVERSE			GGTGTCTTTGGCCAGGTAGT	
ILL2.161284	FORWARD	50°C	1	CCAAGGGCATTACTGAGTGG	GATG
	REVERSE			TTTCGGTTGATTTGGAGGTC	
ILL2.71359	FORWARD	50°C	1	GGTTTGCAATGACCCCATAC	GCAT
	REVERSE			NED -TGTGCATCTATGCGTGTGT	
ILL2.87955	FORWARD	50°C	1	FAM -ATTCTCAGCTGCATGGTTCC	GTGA
	REVERSE			CGAGCGTGCCTCTATTCTGT	
ILL2.104411	FORWARD	50°C	1	FAM -TTGTCTGTGTTGCACATATATACCG	CA
	REVERSE			TTGTCTGTGTTGCACATATATACCG	
ILL2.128551	FORWARD	50°C	1	VIC -CTGAGCTCTGCAGCTGTGAC	TGGA
	REVERSE			GCTCTGTTCCGGTTGTATGCTG	
ILL2.98293a	FORWARD	50°C	1	TGCAGTGTGTTTGTGGTGG	TTTG
	REVERSE			AAACGGGTTGGGTTGTCA	
ILL2.118779	FORWARD	50°C	1	VIC -AAGATTCCTCCCTGAATAGAATGTC	AT
	REVERSE			GTGTATCTCAATAATGAGCGAATCC	
ILL2.126949	FORWARD	50°C	1	GGATACACACCTACACACACTCGAT	CA
	REVERSE			VIC -GTATGTGTTTCGTACGTGTTTGCTT	
ILL2.47613	FORWARD	50°C	1	FAM -GTGTTTACAAGGCGTCATATCAGTA	AC
	REVERSE			CCTGCGCCTACTTACAATAAATG	
ILL2.50505	FORWARD	50°C	1	CCATCATAGTTTCAGTCCCATAAAG	TA

	REVERSE			CAGCCCAACTGTACCATAAATTCT	
<i>ILL2.70036</i>	FORWARD	50°C	1	TCTACATCTAAACTGGCATCTACAGC	TG
	REVERSE			NED -CATAAAACTGGTTTAACTGCTGGAC	
<i>ILL2.76149</i>	FORWARD	50°C	1	FAM -GGAATGGGCTAGTAAGTAGTTCAAA	AT
	REVERSE			ACTAAATGTTTCAGATGTGTGGTGAC	
<i>ILL2.8276</i>	FORWARD	50°C	1	TATGGAGTCTCTGACAAAACAGTTG	AT
	REVERSE			GCATCTTACGGTATGAGGTGTATTT	
<i>ILL2.98293b</i>	FORWARD	50°C	1	AATATCTGCGCATCACTCACAC	CA
	REVERSE			VIC -GGTCGTAGTAGTCGATTTCTGAATG	
<i>ILL2.86892</i>	FORWARD	50°C	1	GTCCAGATCCTCTGTAACACACATT	CA
	REVERSE			ACAGATACATTCCAAAGGACAAGTG	
<i>ILL2.29450</i>	FORWARD	50°C	1	ATCTGCAGGAGTATTTTCAGTGATTC	TG
	REVERSE			PET -TATCACATAGAAGCGTCTCTGACC	
<i>ILL2.46349</i>	FORWARD	50°C	1	TCCATCAGTCTAACAGCTATCACAA	GT
	REVERSE			ATGGTCGAGCACTGTAAATCTACTT	
<i>ILL2.124227</i>	FORWARD	50°C	1	GGTGTGATGGACACAGTGATATTAG	TA
	REVERSE			GCAGGTATTCTGTTCCACCAGTAGAG	
<i>ILL2.105204</i>	FORWARD	50°C	1	GTACAGAGAATTAAGGACCGAGACA	AAG
	REVERSE			TCTCCTTCTTCCTTCTTTTCTACCT	
<i>ILL2.112066</i>	FORWARD	50°C	1	AAGATTGACCTTCCTCACTTTCTG	GAA
	REVERSE			NED -CAATTTCTTCTTCAGTCCAACCTCTC	
<i>ILL2.11646</i>	FORWARD	50°C	1	GGTACAGAGAGATATTTGCGTCCT	CTC
	REVERSE			CACAGAAGAATCAGTCATCAAGTCA	
<i>ILL2.116595</i>	FORWARD	50°C	1	CTGATCCTGTATTTCTTCAACAAC	TTC
	REVERSE			GAAACGAATTGAGACAGAGGTAGAA	

<i>ILL2.125907</i>	FORWARD	50°C	1	GATCTGAGACATGTCATAGCGAAC	TGC
	REVERSE			CTGATCTTGGTCTGGACTCTCTAAT	
<i>ILL2.128607</i>	FORWARD	50°C	1	FAM -ATTCACATTATTGACCAGCCAGAC	GCA
	REVERSE			CATCACAGATGTGTAGCTAAATTGG	
<i>ILL2.128274</i>	FORWARD	50°C	1	GAGAAGAGAACTGAGGAAGTGAGG	GAA
	REVERSE			ATAAGTTTAGCTGTCTGCGTTTCCT	
<i>ILL2.129966</i>	FORWARD	50°C	1	GTCAAATGGCAGAGAGAATAGGAG	AAG
	REVERSE			GAGCACAGTCAAAAAGTCCAGTTAC	
<i>ILL2.131826</i>	FORWARD	50°C	1	GTCCCAATAAACTATGTGGAGAAT	GAA
	REVERSE			CTTTCTGAATGTTCTCAGACTTCCT	
<i>ILL2.134845</i>	FORWARD	50°C	1	VIC -GATAGATTTTGGAAACGAAGACAGG	AGG
	REVERSE			GTAGTGCAGGGGGTACTTTACCAG	
<i>ILL2.134916</i>	FORWARD	50°C	1	ATAAATTCTGTGGAAAGTGCTCCTC	GAT
	REVERSE			CTCTTCAACGTCAGAAATGTCATC	
<i>ILL2.139279</i>	FORWARD	50°C	1	GGCACAACATGTGATAATCTAGGTA	AAT
	REVERSE			TATTAGTGCCACTGAGCATCTGTAA	
<i>ILL2.154893</i>	FORWARD	50°C	1	TAAAACATCTGAAGAACCAGTCTCC	GCA
	REVERSE			CAGATGATGAAGACACAAGGTGTT	
<i>ILL2.20970</i>	FORWARD	50°C	1	ATTGTCTCCCTCTACAGGATGTGAC	TCG
	REVERSE			GCTGTAGCGTAAACAAAACAACATC	
<i>ILL2.38396</i>	FORWARD	50°C	1	PET -ATATCCCAATGTGACGGAGATACTA	ATT
	REVERSE			GGAGAAACCTATAACCGATAAGGAAC	
<i>ILL2.39227</i>	FORWARD	50°C	1	CCCTACCCCTTATCTAGTTGGTG	CCT
	REVERSE			FAM -AGACTACAAAGTGACAGGTGGATCT	
<i>ILL2.39658</i>	FORWARD	50°C	1	GAGCCTTTTGTAAAGAGTGGGATTTA	CCA

	REVERSE			ATCTGGCTATGGAGACAGCTATG	
<i>ILL2.47161</i>	FORWARD	50°C	1	TAACTACATCGGGGATTCGTTTAC	AAT
	REVERSE			CACTGTCACTTTATACAGCTTGTGC	
<i>ILL2.57236</i>	FORWARD	50°C	1	ATTAATGCGAGCTGCTGTGTAAGT	ATC
	REVERSE			ACATTGGAACATATCCTCTACATCC	
<i>ILL2.60863</i>	FORWARD	50°C	1	VIC -TGTC AACATCAGCTAGTTTGT CATC	TCA
	REVERSE			TATGATCAAGGAGAGCATAGAGGAG	
<i>ILL2.6458</i>	FORWARD	50°C	1	NED -TCATTCTTCCAGTCAGATATCAGGT	CAT
	REVERSE			TGATGTTACTAGTCTTGCTCGATGA	
<i>ILL2.64699</i>	FORWARD	50°C	1	TTGGCCTCAATAATGCTATGAGTAG	TGA
	REVERSE			CGTACAGCTGGTATGATGATGAATA	
<i>ILL2.66010a</i>	FORWARD	50°C	1	NED -GAGTACCAAAGGGAGATAACCAAAC	ATT
	REVERSE			AAGAATGCCTACGATACTTCAACAC	
<i>ILL2.71363</i>	FORWARD	50°C	1	GAGTACTACAGACAGCAGGGTATGC	GCA
	REVERSE			GAGTGAATGTGTAGCCATCAAGTG	
<i>ILL2.80333</i>	FORWARD	50°C	1	CCTGAATATCAGAGTCGTTTACACC	CCT
	REVERSE			AGGACAAAAGAGGTACCAGATGTTT	
<i>ILL2.82292</i>	FORWARD	50°C	1	GACCTGACCAAGAGGTTTGAAG	AAG
	REVERSE			GTCTTCTGTATTTACCAGTGGGATT	
<i>ILL2.89045</i>	FORWARD	50°C	1	CGGAGGTCAATTCAAGTATATCAGA	CAT
	REVERSE			CCCAAATACCATTGTTAATCCTCAG	
<i>ILL2.94556</i>	FORWARD	50°C	1	CGTAGCCATCAGTACGTCTACAATA	CTC
	REVERSE			CAGAGGAAGCTGGAGTACTGGT	
<i>ILL2.71693</i>	FORWARD	50°C	1	GATCACGTTTAAGAAACCTGTGAAG	AAG
	REVERSE			CCTCTTCTTCATCATCACCAA	

<i>ILL2.84315</i>	FORWARD	50°C	1	ATATGTACGGACCGATCGTAGAAC	AGG
	REVERSE			AGGAAAGGCCATATGTCTTTAGAAC	
<i>ILL2.108706</i>	FORWARD	50°C	1	AGGGCTTGATCTCAGATGGTAAT	ATC
	REVERSE			CAAGAAAATACCTGCAGAAGAAGAG	
<i>ILL2.64121</i>	FORWARD	50°C	1	GTGTTGGTTGTATTAATGTGTGTGC	CCT
	REVERSE			PET -AAATAGAAAGTAGGAGGGAGGGAAG	
<i>ILL2.71352</i>	FORWARD	50°C	1	GGATTTTCAGACCACAGCATTATCTA	ACT
	REVERSE			TGTCTGTCTGAAGCATGTTAGTGTT	
<i>ILL2.140027</i>	FORWARD	50°C	1	NED -TCCATTCCTTCATCTCTTACTCT	TTC
	REVERSE			AACAAGAAGAACCCAGAACACTCTA	
<i>ILL2.93709b</i>	FORWARD	50°C	1	VIC -AATGCTAACAACCTTGAGAGGAAG	CCG
	REVERSE			TCTCTCTTCTCAGAGCCTCCAC	
<i>ILL2.97931a</i>	FORWARD	50°C	1	GAACAGAGACATTTGACCCTAAAGA	TGA
	REVERSE			PET -TATAGACTTGCAAAGAGTTGTCTCG	
<i>ILL2.97931b</i>	FORWARD	50°C	1	TAGTGGTCTGTGATGATTCATGGTA	GTT
	REVERSE			ACACATCTGACCTCGCTAGTTACTC	
<i>ILL2.24288b</i>	FORWARD	50°C	1	CATCTTCATATGCCCTCCTAAATCT	CTT
	REVERSE			AAGAACAGCCTCTGAAAGAAGAATC	
<i>ILL2.36952</i>	FORWARD	50°C	1	AGACCTATTCAGATCATTGGAACCT	(TGA) _n (TGG)
	REVERSE			CTATAAAGATGGAAGGCTAGTTGAGG	
<i>ILL2.58101a</i>	FORWARD	50°C	1	ACAAAGAAGGATGTTGGTGAAGAC	GAG
	REVERSE			AGGACAATCTAGGTTTTCTTCTGGT	
<i>ILL2.64307</i>	FORWARD	50°C	1	ATGAGGAAGAGAATGAGGATAAACC	AAG
	REVERSE			FAM -GATAAACACACGTCTCACATACAGC	
<i>ILL2.8738</i>	FORWARD	50°C	1	NED -ACATTCATAAGCTGGTCTTTACACG	(TGT) _n (TGC)

REVERSE

AAGTGACTTGCCTCATGTTATTAGC

*EST and cross-species microsatellite marker information:***Table S3: PCR conditions and primer information for the cross-species microsatellite markers from *Haliotis rubra* and *Haliotis discus hannai* (Rhode 2010)**

Locus Name (Genbank Accession number)	T _a (°C)	[MgCl ₂] ¹ (mM)/ Buffer ² (X)	PCR cycle	K/P ³	Primer sequence (5'-3')
<i>Hrub11.A07</i> (DQ278009)	55	2.5/1	1	P	F: NED-AAAGAACTTCTCGCCGAACA R: CAGCATGACCAAACACCTG
<i>Hrub12.E10</i> (DQ278024)	55	2.5/1	1	P	F: VIC-TGCAGCATAACACTTGCTCA R: CGTAGCTGCCTTCATCCTTC
<i>Hrub15.A01</i> (DQ278045)	51	2.5/1	1	P	F:FAM-ACCGTTGGGATGACTGAAAG R: CCTGCGTAGGCGACATTTA
<i>Hrub17.D11</i> (DQ278053)	56	2.5/1	1	P	F: NED-GTGGCTGAAAGGTTCAAACG R: GCCGAACGTTGAGGAGTATG
<i>Hrub12.B10</i> (DQ278020)	50	2.5/1	1	P	F: VIC-GGCGAGGTATTGCTTCTTTG R: GCGTAAGATAAAACCGTTTGAGA
<i>Hrub13.F06</i> (DQ278037)	55	2.5/1	1	P	F: PET-GACAGGTGCTCCCCTATTCA R: CCAGGTGTCAACATGACCTG
<i>Hrub9.B05</i> (DQ278104)	50	2.5/1	1	P	F: AATCCGGAATACTGCACTGG R:NED-AGGTCATATTGTCCACCGGA
<i>Awb083</i> (AB177936)	50	2.0/1.5	2	K	F: NED- GCTTAGAAGGGACATAACTCGCAATA R: AATAGACATTCTACAAGCGAGGAAA
<i>Awb098</i> (AB177939)	60	1.5/1.5	3	K	F:VIC-ACATGGAAGTGCAGTCCTAGAAGC R: TGATTATTTTCAGATCGCCGTCATA
<i>Awb041</i> (AB177924)	45	2.0/1.5	2	K	F:FAM-CAGCAATATTCTAGCATGACGGTGG R: ACATACGCATCATGTTGGAAAGCAC

1 – Magnesium chloride concentration in millimolar; 2 – Buffer concentration; 3 – KAPATM or Promega PCR Kit. Primers taken from: Bananski et al. (2006a) (*Hrub.- primers*); Sekino et al. (2004) (*Awb-primers*)

Table S4: PCR conditions and primer information for the EST-microsatellite markers (Rhode 2010)

Locus Name (Genbank Accession number)	T _a (°C)	[MgCl ₂] ¹ (mM)/ Buffer ² (X)	PCR cycle	K/P ³	Primer sequence (5'-3')
<i>HdSSRex495</i>	55	1.5/1.5	2	K	F: CGCGGCATTAAGGAAATAAA

(ex534495)					R: NED-CAGTGTTAAACGTCGCATTGA
<i>HaSSRgd842</i>	55	1.5/1.5	2	K	F: GAAAAGCAGTCTCAGCGTCAG
(gd241842)					R: VIC-CGAGATCAGCTTGTAGACTTGG
<i>HaSSRdw239</i>	52	1.5/1.5	2	K	F: VIC-CACCAGGCAACAATCATCTG
(dw986239)					R: CCTGTTGTGTTTGACCGTTG
<i>HmSSRex489b</i>	55	1.5/1.5	2	K	F: PET-ATGAATCATGAAGCCAAGACG
(ex534489)					R: ACAATATGGCACCATGAGGAT
<i>HaSSRdw503</i>	55	2.0/1.5	2	K	F: FAM-TAAGAGCATGGGGGTGACTC
(dw986503)					R: TTCCCTGATGCAACCATAACA
<i>HmSSRex489a</i>	55	1.5/1.5	2	K	F: GGTTGCTCTGGTTCAAATCC
(ex534489)					R: PET-TCAAGGTATGCTCAGTGTGGA
<i>HdhSSRfe537</i>	50	2.0/1.5	2	K	F: CGCCCGGGCAGGTACAAA
(FE041537)					R: NED-TCTATGTTAGGGTCCCTCCT
<i>HdSSRcx732</i>	55	1.5/1.5	2	K	F: NED-GTTCTCTGACACGCCTCTCC
(cx726732)					R: CTGCGGTAGGCGATGTTCT
<i>HmSSRex446a</i>	56/	2.0/1.5	3	K	F: PET-GGTAGGGTGGGTTGGTTGAT
(ex534446)	(TD: 66-56)				R: AGACAAATCTCCTGAATCTCCA
<i>HaSSRc.571a</i>	55	1.5/1.5	2	K	F: CCAGTTGCCAAGGAGACACT
(GU263799)					R: VIC-ACAACCAACACGCACTGACAT
<i>HmSSRex446b</i>	55	1.5/1.5	2	K	F: VIC-AATGTGTGAAGCCATTTCTG
(ex534446)					R: TCGAAGCACGTGTAAAATCCT
<i>HdhSSRc.60b</i>	55	2.0/1.5	3	K	F: FAM-CTAGGTTTGTGAGCACCATAC
(GU263800)	(TD: 65-55)				R: ACCCTGTGGAATAAGGTTTT
<i>HaSSRgd475c</i>	55	1.5/1.5	2	K	F: TCCTGGACAAAATGGTGTAGC
(gd272475)					R: VIC-TGTGTATCTGCTCCCTCCTTG
<i>HaSSRdy903</i>	55	1.5/1.5	2	K	F: FAM-AGTGTGGAGGAACCAAGGTG
(dy402903)					R: GCAGTTTCCTTCCATTCTCC
<i>HdSSRcx009</i>	56/	2.0/1.5	3	K	F: AGACAAATGAGCTGCCTGAAG
(CX726009)	(TD: 66-56)				R: NED-GATACAAAGTCAGCAGGACAG:

1 – Magnesium chloride concentration in millimolar; 2 – Buffer concentration; 3 – KAPATM or Promega PCR Kit.

Appendix 3: PCR cycling programs*Cycle-program 1: General microsatellite*

Initial denaturing	10 min	95°C	
Denaturing	45 sec	94°C	} X 30
Annealing	45 sec	T _a °C	
Extension	45 sec	72°C	
Final extension	10 min	72°C	

Cycle-program 2: KAPA2G™ Fast HotStart Program

Initial denaturation	5 min	95°C	
Denaturing	15 sec	94°C	} X 30
Annealing	30 sec	T _a °C	
Extension	30 sec	72°C	
Final extension	7 min	72°C	

Cycle-program 3: KAPA2G™ Fast HotStart Touch down Program

Initial denaturing	5 min	95°C	
Denaturing	15 sec	94°C	} X 10
Annealing	30 sec	T _a °C	
Denaturing	15 sec	94°C	} X 30
Annealing	30 sec	T _a °C	

Cycle-program: BigDye sequencing

Denaturing	10 sec	96 °C	} X 25
Annealing	5 sec	50 °C	
Extension	4 min	60 °C	