

Genomic Resource Development for the South African
scallop, *Pecten sulcicostatus*

by

Natasha Kitchin

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



Supervisor: Clint Rhode, Ph.D., Pr.Sci.Nat.
Co-supervisor: Rouvay Roodt-Wilding, Ph.D.

Department of Genetics

March 2016

Declaration

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

March 2016

Summary

Although South Africa boasts a rich biodiversity, many South African species, especially marine species, remain uncharacterised. Many of these species show great potential for commercial use, in particular for the aquaculture of endemic species for high-value, specialised markets. Only once these species have been identified and genetically characterised can the feasibility of aquaculture be evaluated.

‘Scallop’ refers to species of marine bivalve molluscs in the family Pectinidae, although it may also refer to species in other closely related families within the superfamily Pectinoidea. There are 29 scallop species in the waters surrounding South Africa, and of these, *Pecten sulcicostatus* has been identified as a candidate species for aquaculture. Non-destructive sampling is necessary for studies on genetic fitness and population structure to not be hindered by the death of the study individuals. DNA extraction methods using non-destructive sample tissue have not been developed for scallops, therefore this study compared the use of various tissue types in DNA extraction, allowing for the development of an effective non-destructive DNA extraction method for *P. sulcicostatus* using tentacle tissue and mucus swabs. This study also allowed for the development of an effective DNA extraction method for use on dried or degraded tissue, which will, in turn, allow for the use of opportunistic and historic samples in future studies on *P. sulcicostatus*.

Despite the potential commercial value of *P. sulcicostatus*, no genetic resources are available for this endemic species. The development of genetic markers will assist in future studies on the genetic composition of this species as well as the genetic constitution of *P. sulcicostatus* populations along the South African coast – factors which are important for the formulation of effective genetic management strategies. This study therefore aimed to develop genetic markers for *P. sulcicostatus* and to conduct preliminary analyses using these markers to demonstrate their usefulness for future studies, which will assist in the establishment of a sustainable aquaculture industry. This study allowed for the optimisation of a fragment of the 16S rRNA gene which was successfully used to determine intra- and interspecific genetic diversity and shed light on the evolutionary relationship of five *Pecten* species. A set of 10 microsatellite markers was developed for *P. sulcicostatus* using cross-species

amplification from *Pecten maximus*, a sister species to *P. sulcicostatus*, with a success rate of 50%. The set of microsatellite markers was successfully applied to generate genetic diversity data, which, in future studies, could be used to evaluate the extent of intra- and interpopulation genetic partitioning and variation. This study also provided the first reduced genome sequences for *P. sulcicostatus*, with over 7.3 million reads. The use of two bioinformatic approaches aided in the identification of 55 putative microsatellite markers as well as 2 530 putative SNPs. Although currently limited, this study marks the first step towards providing genetic information to assist in the development of genetic management strategies within the context of establishing a sustainable aquaculture industry for this endemic species. The genetic resources developed in this study could be used in various downstream applications such as genetic diversity assessment, population structure inference, linkage studies as well as marker assisted selection. In future, the microsatellite markers and SNPs developed in this study could also be continuously used to monitor genetic diversity as the species is subjected to aquaculture.

Opsomming

Ten spyte van 'n ryk biodiversiteit, is baie Suid-Afrikaanse spesies, veral mariene spesies, steeds ongekarakteriseerd. Baie van hierdie spesies het groot potensiaal vir kommersiële gebruik, veral die akwakultuur van endemiese spesies vir hoë-waarde, gespesialiseerde markte. Slegs nadat hierdie spesies geïdentifiseer is en geneties gekarakteriseerd is, kan die volhoubaarheid van akwakultuur geëvalueer word.

'Kammossel' verwys na spesies van marine tweekleppigediere in die familie Pectinidae, alhoewel dit ook na spesies in ander naby verwante families binne die superfamilie Pectinoidea kan verwys. Daar is 29 kammossel spesies in die waters rondom Suid-Afrika. Van die 29 Suid-Afrikaanse kammossel spesies, is *Pecten sulcicostatus* geïdentifiseer as 'n kandidaat vir akwakultuur. Nie-vernietigende monsterneming is nodig vir studies van genetiese fiksheid en populasie struktuur sodat die dood van individue nie dié studies verhinder nie. Die gebruik van nie-vernietigende monster weefsel in DNS ekstraksie is nog nie vir kammossels ontwikkel nie, dus vergelyk hierdie studie die gebruik van verskillende tipes weefsel in DNS ekstraksie, om 'n nie-vernietigende DNS ekstraksie metode vir *P. sulcicostatus* te ontwikkel met behulp van tentakel weefsel en slym deppers. Hierdie studie het ook 'n DNS ekstraksie metode ontwikkel vir gebruik op droë of afgebreekte weefsel, wat gebruik kan word vir opportunistiese en historiese monsters in toekomstige studies op *P. sulcicostatus*.

Ten spyte van die feit dat *P. sulcicostatus* geïdentifiseer is as 'n potensiële akwakultuur spesies, is daar geen genetiese hulpbronne beskikbaar vir hierdie spesie nie. Die ontwikkeling van genetiese merkers vir hierdie endemiese spesies sal bydra tot toekomstige studies op die genetiese samestelling van hierdie spesie, asook die genetiese samestelling van *P. sulcicostatus* populasies rondom die Suid-Afrikaanse kus – belangrike faktore vir die formulering van effektiewe genetiese bestuur strategieë. Die doelwit van hierdie studie was dus om genetiese merkers vir *P. sulcicostatus* te ontwikkel en om voorlopige bevestiging van hierdie merkers uit te voer om hul nut te bewys vir toekomstige studies wat sal help met die vestiging van 'n volhoubare akwakultuurbedryf. Hierdie studie het 'n fragment van die 16S rRNA geen geoptimeer wat gebruik is om intra- en interspesifieke genetiese diversiteit te bepaal en die evolusionêre verwantskappe van vyf *Pecten* spesies te bestudeer. 'n

Stel van 10 mikrosatelliet merkers is ontwikkel vir *P. sulcicostatus* deur die gebruik van kruis-spesies amplifisering vanaf *Pecten maximus*, 'n suster spesie van *P. sulcicostatus*, met 'n suksesyfer van 50%. Die mikrosatelliet merker stel is gebruik om genetiese diversiteit data daar te stel, wat in toekomstige studies gebruik kan word om intra- en interpopulasie genetiese verskille/differensiasie en variasie te evalueer. Hierdie studie het ook die eerste verkleinde genoom volgorde bepaling vir *P. sulcicostatus* gedoen, met meer as 7.3 miljoen lesings. Twee bioinformatiese tegnieke het 'n totaal van 55 vermeende mikrosatelliet merkers asook 2 530 vermeende enkel nukleotied polimorfisme (ENP) in hierdie studie geïdentifiseer. Alhoewel tans beperk, is hierdie studie die eerste stap in die verskaffing van genetiese inligting vir die ontwikkeling van genetiese bestuur strategieë binne die konteks van die totstandkoming van 'n volhoubare akwakultuurbedryf vir hierdie endemiese spesie. Die genetiese hulpbronne wat in hierdie studie ontwikkel is, kan in verskeie toekomstige studies gebruik word vir genetiese diversiteit assessering, populasiestruktuur inferensie, koppeling studies asook merkerbemiddelde seleksie. Die mikrosatelliet merkers en ENP kan ook voortdurend gebruik word om genetiese diversiteit te monitor as die spesie onderworpe is aan akwakultuur.

Acknowledgements

I would like to thank the following persons for technical assistance, guidance, advice and support during the course of this project:

Jessica Vervalle

Dr. Clint Rhode

Prof. Rouvay Roodt-Wilding

The following people and institutions are thanked for biological samples, financial support and use of facilities:

Brett M. Macey and Dale Arends, Fisheries branch of the Department of Agriculture, Forestry and Fisheries (DAFF), South Africa

Catarina N. S. Silva, Victoria University of Wellington, New Zealand

Dr. Dai Herbert, KwaZulu-Natal Museum, South Africa

Dr. Gustav Paulay and Amanda Bemis, Florida Museum of Natural History, USA

Dr. Jennifer R. Ovenden, University of Queensland, Australia

Dr. Marianna Pauletto, University of Padova, Italy

Mary Cole, East London Museum, South Africa

Prof. Paolo Mariottini, Roma Tre University, Italy

Central Analytical Facilities (CAF), Stellenbosch University

National Research Foundation (NRF)

Stellenbosch University

Table of Contents

Declaration	i
Summary	ii
Opsomming	iv
Acknowledgements	vi
Table of Contents	vii
List of Figures	x
List of Tables	xiii
List of Abbreviations	xvi
Chapter I: Literature Review	
1. Introduction	1
2. Scallops: An Overview	1
2.1 Scallop Evolution, Phylogeny and Distribution	2
2.2 Scallop Biology and Ecology	11
3. Fisheries and Aquaculture	16
3.1 Potential South African Scallop Industry	19
4. Molecular Markers	20
4.1 Nuclear Molecular Markers	21
4.1.1 Microsatellite Markers: Characteristics and Development	21
4.1.2 Single Nucleotide Polymorphisms (SNPs): Characteristics and Development	23
4.1.3 Next Generation Sequencing (NGS)	24
4.2 Mitochondrial DNA (mtDNA)	26
4.3 Genetics and Molecular Markers in Aquaculture	27

5. Rationale, Aims and Objectives	29
6. References	31

**Chapter II: DNA Extraction in the South African Scallop, *Pecten sulcicostatus*:
Evaluation of Optimal Tissue Type**

1. Abstract	40
2. Introduction	41
3. Materials and Methods	43
3.1 Biological Specimens	43
3.2 Tissue Collection and DNA Extraction	43
3.3 Testing of Genomic DNA Amplification	45
4. Results and Discussion	46
5. Conclusion	50
6. References	52

**Chapter III: Cross-species Transfer of Microsatellite Markers and Optimisation
of Mitochondrial Markers for Applications in *Pecten sulcicostatus***

1. Abstract	55
2. Introduction	56
3. Materials and Methods	58
3.1 Biological Specimens and DNA Extraction	58
3.2 Cross-species Amplification of Microsatellite Markers	60
3.2.1 Microsatellite Marker Genotyping	64
3.2.2 Data Analysis	65
3.3 Assessment of Mitochondrial Genes for <i>Pecten sulcicostatus</i>	65
3.3.1 Sanger Sequencing	67

3.3.2 Data Analysis	68
4. Results and Discussion	68
4.1 Microsatellite Markers	68
4.2 Mitochondrial Genes	78
5. Conclusion	98
6. References	100
Chapter IV: Marker Discovery through Next Generation Sequencing	
1. Abstract	106
2. Introduction	107
3. Materials and Methods	109
3.1 Biological Specimens, DNA Extraction and Next Generation Sequencing (NGS)	109
3.2 Processing of Next-Generation Sequencing (NGS) Data	109
3.3 Identification of Putative Microsatellite Markers and Single Nucleotide Polymorphisms (SNPs)	110
4. Results and Discussion	111
4.1 Approach 1	111
4.2 Approach 2	116
4.3 Comparison of Bioinformatic Approaches	122
5. Conclusion	124
6. References	126
Chapter V: Summary of Findings and Conclusion	
1. Introduction	130
2. Synopsis of Main Results and Discussion	130
3. Future Work and Study Limitations	134
4. Final Remarks	136
5. References	137
Appendix	142

List of Figures

- Figure 1.1: A phylogeny of Pectinidae species plotted against time. Phyletic lines ending in arrows contain extant representatives while those ending in cross-bars are extinct. Taken from Shumway and Parsons (2006). 3
- Figure 1.2: Putative natural distribution range of *P. sulcicostatus* along the inner continental shelf from False Bay to East London, South Africa. 9
- Figure 1.3: Shell morphology of (a) *Pecten sulcicostatus*, (b) *Pecten maximus* and (c) *Pecten jacobaeus*, showing the similar shell shape and size of all three *Pecten* species. Taken from Pecten Site (www.pectensite.com/). 10
- Figure 1.4: Shell morphological characteristics of (a) *Pecten sulcicostatus*, (b) *Pecten maximus* and (c) *Pecten jacobaeus*, demonstrating how *Pecten sulcicostatus* (a) can be distinguished from *Pecten maximus* (b) and *Pecten jacobaeus* (c) by the rough surface of the valve due to secondary radial riblets. Shell morphological characteristics also show the more pronounced and square radial costae of *Pecten jacobaeus* (c) in comparison to the flatter radial costae of *Pecten maximus* (b). Taken from Pecten Site (www.pectensite.com/). 10
- Figure 1.5: Upper valve of the *Pecten sulcicostatus* shell with key morphological characteristics. The upper valve is sculptured with 12 to 15 radial costae with secondary radial riblets. The base of the shell has a pair of auricles which are equal in size and the hinge is located between the two auricles. 13
- Figure 1.6: Internal anatomy of *Pecten sulcicostatus*. Mantle musculature tissue encloses the internal organs (gonad and gills) while the adductor muscle holds the two valves together. Tentacles are located within the folds of the free edge of the mantle musculature. 14
- Figure 1.7: Depiction of the developmental stages in the life cycle of scallops. Broadcast spawning occurs primarily in summer, after which larvae remain in the water column for two to four weeks before dissipating to the ocean floor where they attach themselves to a substratum through byssus threads. Rapid growth within the first years of a scallop's life allows most scallops to reach commercial size at four to five years of age. Taken from Leavitt *et al.* (2010). 15

Figure 1.8: (A) The P1 adapter is ligated to digested gDNA. The P1 adapter contains a barcode, forward primer set and primer site. (B) Fragments are pooled and randomly sheared. (C) P2 adapters, containing a reverse complement of the reverse amplification primer site, are ligated onto the fragments. (D) Enrichment of RAD tag containing both the P1 and P2 adapters. Taken from Baird *et al.* (2008). 25

Figure 2.1: Various tissue types used in tissue sampling of *Pecten sulcicostatus*. Internal tissue types (adductor muscle, gills, gonad tissue and mantle musculature) represent destructive tissue samples, whilst tentacle tissue represent non-destructive tissue samples. 43

Figure 2.2: DNA quantification (ng/μl) and quality (absorbance ratio: 260/280 and 260/230) for gDNA extracted from various tissue types in *Pecten sulcicostatus*. 46

Figure 2.3: 2% agarose gel image of PCR products of a fragment of the 16S rRNA gene using gDNA extracted from various tissue types where “W” indicates tissue sampled from whole, dead individuals while “L” indicates tentacle tissue sampled from live individuals. 47

Figure 2.4: 2% agarose gel image of PCR products of a fragment of the 16S rRNA gene using gDNA extracted from mucus swabs. 48

Figure 2.5: 2% agarose gel image of PCR products of a fragment of the 16S rRNA gene using gDNA extracted from dried mantle musculature tissue. 48

Figure 3.1: 12% polyacrylamide gel image of PCR products of the PmNH11 (lanes 1-4) and PmGC01 (lanes 5-8) loci. 68

Figure 3.2: 12% polyacrylamide gel image of PCR products of the PmGC05 (lanes 1-4) and PmGC10 (lanes 5-8) loci. 68

Figure 3.3 Principle coordinates analysis (PCoA) of the False Bay *Pecten sulcicostatus* individuals based on 10 microsatellite markers visualising the level of similarity of individuals. 76

Figure 3.4: Shell morphology of (a) *Pecten sulcicostatus* and (b) *Pecten afribenedictus*, showing the similar shell shape and size as well as the more

convex lower valve of *Pecten aribenedictus* (b) and the almost flat upper valve of *Pecten sulcicostatus* (a) compared to the concave upper valve of *Pecten aribenedictus*. Taken from Pecten Site (www.pectensite.com/). 80

Figure 3.5: Haplotype network representing the relationships between the 18 *Pecten sulcicostatus* haplotypes identified using a 582 bp fragment of the 16S rRNA gene. Node size reflects the frequency of the haplotypes. The line length is proportional to the number of mutations/mutational steps between each haplotype. Intermediate missing haplotypes are indicated as red dots. 81

Figure 3.6: 2% agarose gel image of PCR products of the 16S primer pair using gDNA from two replications of the three *Pecten keppelianus* individuals. 82

Figure 3.7: Haplotype network representing the relationships between the 32 haplotypes identified using a 504 bp fragment of the 16S rRNA gene for various *Pecten* species. Node size reflects the frequency of the haplotypes. The number of mutations/mutational steps between each haplotype is one unless otherwise stated. Intermediate missing haplotypes are indicated as red dots. 94

Figure 3.8: Maximum Likelihood tree of a 504 bp fragment of the 16S rRNA gene of haplotype sequences of five *Pecten* species with *Euvola perulus* and *Euvola ziczac* as outgroups. The “*P. jacobeus*/*P. maximus*” haplotype indicates a shared haplotype. The numbers above branches are bootstrap support values in %.

95

List of Tables

Table 1.1: Fifteen extant <i>Pecten</i> species as described by Dijkstra (1998) including their natural distribution ranges around the world.	4
Table 1.2: Scallop species found in the waters surrounding South Africa as described by Dijkstra and Kilburn (2001).	6
Table 1.3: Descriptions of life habit classes in the family Pectinidae.	12
Table 1.4: The top five scallop species utilised in aquaculture according to production in tons (Astorga 2014).	18
Table 2.1: 16S mitochondrial primer pair used to test gDNA amplification.	45
Table 3.1: Sample details for various <i>Pecten</i> species included in this study.	58
Table 3.2: Microsatellite marker primers developed for <i>Pecten maximus</i> grouped according to annealing temperature (T_a).	60
Table 3.3: Mitochondrial primer pairs for both the <i>CO1</i> and 16S rRNA genes tested in the present study.	65
Table 3.4: PCR cycling conditions for the various mitochondrial primer pairs tested in the present study.	66
Table 3.5: Microsatellite markers grouped into three multiplexes with florescent label, motif, predicted and actual size range as well as annealing temperature (T_a).	69
Table 3.6: Number of alleles (A_n), effective number of alleles (A_e), information (Shannon-Weaver) index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), Hardy-Weinberg Equilibrium (HWE) p value, frequency of null alleles (fr_{null}), inbreeding coefficient within individuals (F_{is}), polymorphic information content (PIC), probability of identity (PI) and probability of exclusion (PE) for the False Bay <i>Pecten sulcicostatus</i> population.	73

Table 3.7: Comparison of number of alleles (A_n) and observed heterozygosity (H_o) for *Pecten maximus*, as taken from Charrier *et al.* (2012), Hold *et al.* (2013) and Morvezen *et al.* (2013), and *Pecten sulcicostatus* (current study). 74

Table 3.8: The number and composition of polymorphic sites and haplotype and nucleotide diversity for a 582 bp fragment of the 16S rRNA gene of *Pecten sulcicostatus*. 79

Table 3.9: BLAST results of the consensus sequences of the various *Pecten* - and outgroup species, indicating the best match as well as the percentage identity. 83

Table 3.10: Summary of the putative polymorphisms detected in a 504 bp fragment of the 16S rRNA gene. 84

Table 3.11: Polymorphic sites for sites 31 - 281 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species. 85

Table 3.12: Polymorphic sites for sites 284 - 319 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species. 86

Table 3.13: Polymorphic sites for sites 323 - 471 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species. 87

Table 3.14: The number and composition of polymorphic sites, haplotype and nucleotide diversity for a 504 bp fragment of the 16S rRNA gene of various *Pecten* species. 89

Table 3.15: Haplotype and nucleotide diversity for a fragment of the 16S rRNA gene of various *Pecten* species studied in this study as well as a study by Saavedra and Peña (2004). 91

Table 3.16: Haplotype and nucleotide diversity for a fragment of the 16S rRNA gene of *Pecten maximus* and *Pecten jacobaeus* in the current study (A), a study by Saavedra and Peña (2004) (B) and a study by Saavedra and Peña (2005) (C).

92

Table 4.1: Microsatellite markers identified from the contigs constructed using approach one. The repeat is shown in red.

111

Table 4.2: Microsatellite markers identified from the contigs constructed using approach two. The repeat is shown in red.

116

Table 4.3: Summary of the variants detected following *de novo* assembly of the 18 958 contigs generated using approach two.

120

Table 4.4: Summary of the putative single nucleotide polymorphisms (SNPs) detected.

120

Table 4.5: Comparison of the two bioinformatic approaches and their success in microsatellite marker and SNP discovery.

122

Table A1: Nucleotide variation in a 582 bp fragment of the 16S rRNA gene of 18 *Pecten sulcicostatus* haplotypes. Only variable sites are shown. Dots indicate identity with the Haplotype 1 sequence.

141

Table A2: Nucleotide variation in a 504 bp fragment of the 16S rRNA gene of 32 haplotypes. Only variable sites (site 32 to site 305) are shown. Dots indicate identity with the Haplotype 1 sequence.

142

Table A3: Nucleotide variation in a 504 bp fragment of the 16S rRNA gene of 32 haplotypes. Only variable sites (site 306 to site 471) are shown. Dots indicate identity with the Haplotype 1 sequence.

144

Table A4: Haplotype frequencies of the five *Pecten* species.

146

List of Abbreviations

%	Percentage
°C	Degrees Celsius
µl	Microlitre
µM	Micromolar
3'	Three prime
5'	Five prime
A	Adenine
A _e	Effective number of alleles
A _n	Number of alleles
BLAST	Basic Local Alignment Search Tool
bp	Base pair
C	Cytosine
CAF	Central Analytical Facility
CE	Capillary Electrophoresis
CO1	Cytochrome c oxidase subunit I
CTAB	Cetyl trimethylammonium bromide ((C ₁₆ H ₃₃)N(CH ₃) ₃ Br)
DAFF	Department of Agriculture, Forestry and Fisheries
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediaminetetraacetic acid (C ₁₀ H ₁₆ N ₂ O ₈)
EST	Expressed Sequence Tags
EtBr	Ethidium Bromide
F	Forward primer
F _{is}	Inbreeding coefficient
G	Guanine
GB	GigaByte

gDNA	Genomic DNA
GDP	Gross Domestic Product
H3	Histone 3
H _e	Expected heterozygosity
H _o	Observed heterozygosity
HPC	High-Performance Computing
HW	Hardy-Weinberg
HWE	Hardy-Weinberg Equilibrium
I	Information (Shannon-Weaver) index
MAF	Minor Allele Frequency
MAS	Marker Assisted Selection
MgCl ₂	Magnesium Chloride
mM	Millimolar
MNV	Multiple Nucleotide Variants
MSG	Multiplexed Shotgun Genotyping
mtDNA	Mitochondrial DNA
MY	Million years
MYA	Million years ago
ND1	NADH dehydrogenase subunit 1
ng	Nanogram
ng/μl	Nanogram per microlitre
NGS	Next Generation Sequencing
PAGE	Polyacrylamide gel electrophoresis
PCoA	Principle Coordinates Analysis
PCR	Polymerase Chain Reaction
PE	Probability of Exclusion

PI	Probability of Identity
PIC	Polymorphic Information Content
QTL	Quantitative Trait Loci
R	Reverse primer
RAD	Restriction-site Associated DNA sequencing
RAPD	Random Amplified Polymorphic DNA
RE	Restriction enzyme
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RRS	Reduced Representation Sequencing
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variants
SSCP	Single-strand Conformation Polymorphism
SSR	Simple Sequence Repeats
T	Thymine
T _a	Annealing temperature
Taq	<i>Thermus aquaticus</i> DNA polymerase
TB	TeraByte
ts:tv	transition to transversion
U	Units
uH _e	Unbiased expected heterozygosity
UV	Ultraviolet
VNTR	Variable Number Tandem Repeat
v:v	volume:volume
ZAR	South African Rand

Chapter I

Literature Review

1. Introduction

South Africa has a rich biodiversity, yet many species, especially within marine ecosystems, remain uncharacterised. Many of these marine species are underutilised and could be of commercial value within the bio-economy. One such sector is the aquaculture of endemic species for high-value, specialised markets. South Africa aims to promote an economically sustainable and globally competitive marine aquaculture industry (Chief Directorate: DAFF 2012), therefore aquaculture of endemic species, which promises a cost-effective means of providing a sustainable protein source, is of great importance (Directorate: Marketing of DAFF 2012). It is therefore necessary to identify and genetically characterise such species in order to evaluate the feasibility of aquaculture activities and formulate management strategies for long-term economic and environmental sustainability.

2. Scallops: An Overview

The term scallop is commonly applied to species of saltwater clams or marine bivalve molluscs in the family Pectinidae, although it may also refer to species in other closely related families within the superfamily Pectinoidea. Some scallop species have a very narrow distribution range, whilst most species are opportunistic, living under a wide variety of conditions over large distances. Most scallop species live in shallow waters from the low tide line to approximately 100 meters (Shumway and Parsons 2006), however a few species live as deep as 7 000 meters below sea level (Barucca *et al.* 2004).

Scallops are regarded as premium seafood, which has led to the farming of a number of species worldwide. Scallops are one of the most colourful and variable mollusc families, varying greatly in colour, pattern and shell morphology. The brightly coloured, fan-shaped shells of scallops are valued by shell collectors and used in art, architecture and design. Scallops produce pearls, however these pearls do not have the build-up of layers and therefore may not have lustre or iridescence. As such,

scallops pearls are often small, dull and of varying colour, but a few exceptions are appreciated for their aesthetic qualities (Shumway and Parsons 2006).

2.1 Scallop Evolution, Phylogeny and Distribution

The earliest known records of true scallops date to the Triassic period, over 245 million years ago. These archetypal species were classified into two groups based on shell morphology - *Pleuromerites*, with a nearly smooth exterior, and *Praechlamys*, with radial ribs and auricles (Shumway and Parsons 2006) (Figure 1.1). Fossil records suggest that the abundance of Pectinidae species varied over time. At the beginning of the Mesozoic era (252 to 65 million years ago (MYA)) the Pectinidae was the most diverse bivalve family, but it almost disappeared by the end of the Cretaceous period (145.6 to 65.0 MYA). The surviving species radiated during the Tertiary period (65.0 to 2.6 MYA) and currently the Pectinidae consists of approximately 350 extant species in more than 56 genera within the subfamilies Chlamydinae, Palliolinae and Pectininae, and the tribe Aequipectini (Puslednik and Serb 2008). Current species of the family Pectinidae live in a wide range of habitats, from shallow subtidal waters to depths of 7 000 meters and from the tropics to the polar regions (Barucca *et al.* 2004). *Pecten* is a genus of large scallops within the family Pectinidae in the *Praechlamys* group (Figure 1.1).

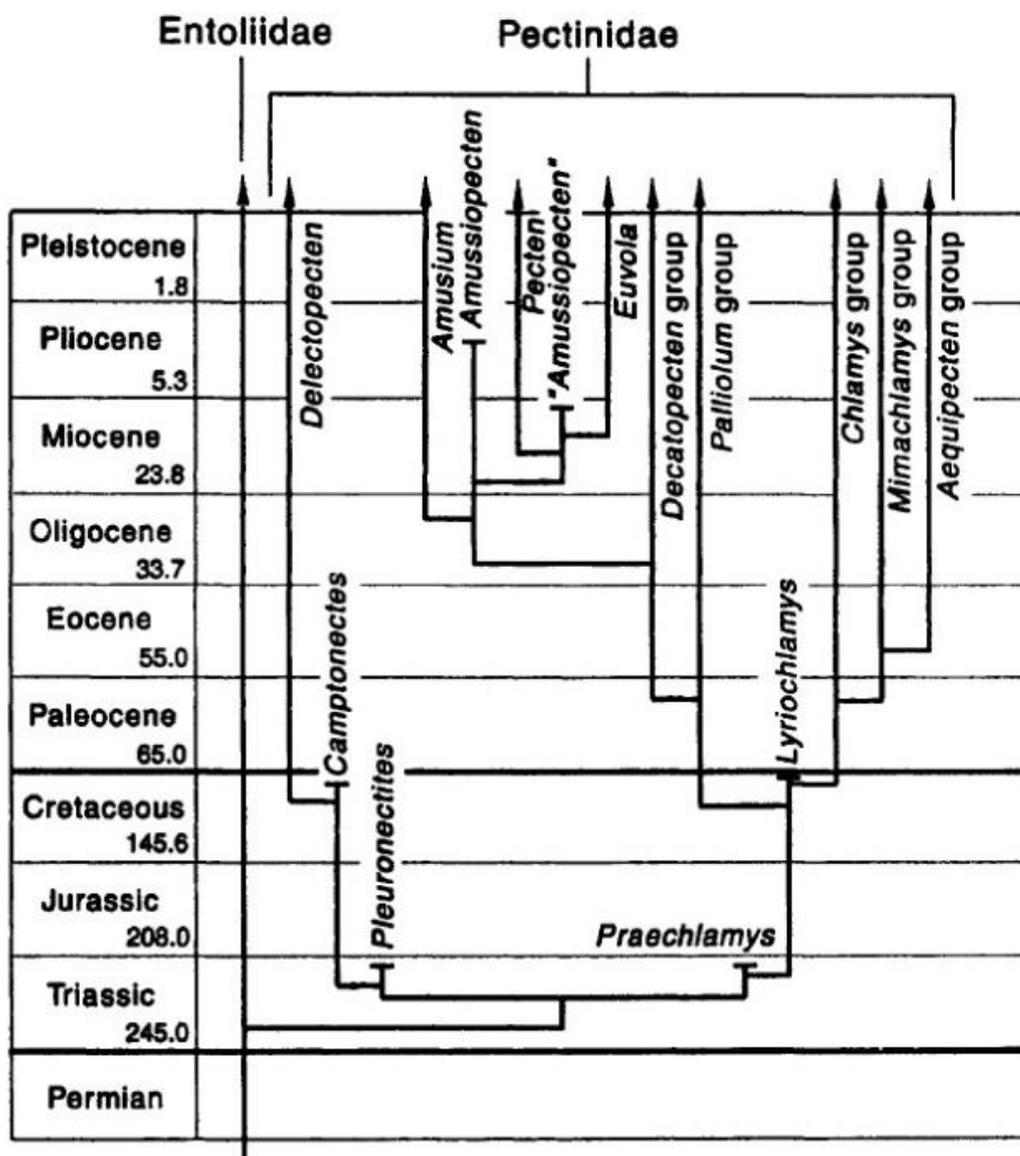


Figure 1.1: A phylogeny of Pectinidae species plotted against time. Phyletic lines ending in arrows contain extant representatives while those ending in cross-bars are extinct. Taken from Shumway and Parsons (2006).

Species within the genus *Pecten* are distributed throughout temperate and sub-tropical oceans in distant geographic areas such as Europe, Africa, Asia and Australia (Raines and Poppe 2006). There are a total of 15 extant *Pecten* species (Table 1.1) (Dijkstra 1998).

Table 1.1: Fifteen extant *Pecten* species as described by Dijkstra (1998) including their natural distribution ranges around the world.

Species	Authority	Distribution
<i>P. aribenedictus</i>	Kilburn and Dijkstra, 1995	East Coast of South Africa
<i>P. albicans</i>	Schröter, 1802	Japanese and South China Seas
<i>P. dijkstrai</i>	Duncan and G. Wilson, 2012	Western Australia
<i>P. diomedeus</i>	Dall, Bartsch and Rehder, 1938	North Pacific Ocean
<i>P. dorotheae</i>	Melvill and Standen, 1907	North West Indian Ocean
<i>P. erythraeensis</i>	G. B. Sowerby II, 1842	East Africa and the Red Sea
<i>P. excavates</i>	Anton, 1838	South China Sea
<i>P. fumatus</i>	Reeve, 1852	Australia
<i>P. jacobaeus</i>	Linnaeus, 1758	Eastern Basin of the Mediterranean Sea, North Atlantic Ocean and the North Coast of Tunisia
<i>P. keppelianus</i>	Sowerby III, 1905	North Atlantic Ocean, Luanda and Senegal
<i>P. maximus</i>	Linnaeus, 1758	Belgium, France, Ireland, Morocco, Norway, Sweden, United Kingdom, North Atlantic Ocean and the North Sea
<i>P. novaezelandiae</i>	Reeve, 1852	New Zealand
<i>P. raoulensis</i>	Powell, 1958	New Zealand
<i>P. sulcicostatus</i>	Sowerby II, 1842	South Coast of South Africa
<i>P. waikikius</i>	Dall, Bartsch and Rehder, 1938	North Pacific Ocean

Prior to the last decade of the 19th century only two scallop species, namely *P. sulcicostatus* and *Talochlamys multistriata* (Poli, 1795) had been recorded in the waters surrounding South Africa. Dredging surveys conducted between 1897 and 1901 added three benthic species, namely *Pseudamussium gilchristi* (Sowerby III, 1904), *Talochlamys humilis* (Sowerby III, 1904) and *Volachlamys fultoni* (Sowerby III, 1904), bringing the total to five. Over the following few decades a further two species, *Laevichlamys weberi* (Bavay, 1904) and *Semipallium coruscans coruscans* (Hinds, 1845), were added. During this time the first records of Indo-Pacific taxa were also recorded in South African waters. In 1964 it was believed that South Africa had only eight species of Pectinidae, but a few subsequent records brought the total number of Pectinidae species to 13. Over the last 30 years, two new species and one subspecies have been added from benthic samples collected during dredging, scuba and littoral collecting. Currently, scallops are represented by 29 species in South Africa (Table 1.2). These species have diverse origins with radiations from both the east (Indo-Pacific) and west (Mediterranean/West Africa) (Dijkstra and Kilburn 2001).

Table 1.2: Scallop species found in the waters surrounding South Africa as described by Dijkstra and Kilburn (2001).

Subfamily	Tribe	Genus	Species
Camptonectinae (Habe, 1977)		<i>Delectopecten</i> (Stewart, 1930)	<i>Delectopecten musorstomi</i> (Poutiers, 1981)
			<i>Delectopecten vitreus</i> (Gmelin, 1791)
	Palliolini (Waller, 1991)	<i>Pseudamussium</i> (Mörch, 1853)	<i>Pseudamussium gilchristi</i> (Sowerby III, 1904)
		<i>Anguipecten</i> (Dijkstra, 1995)	<i>Anguipecten picturatus</i> (Dijkstra, 1995)
		<i>Bractechlamys</i> (Iredale, 1939)	<i>Bractechlamys nodulifera</i> (Sowerby II, 1842)
Pectininae (Wilkes, 1810)	Decatopectinini (Waller, 1986)	<i>Decatopecten</i> (G.B Sowerby II, 1839)	<i>Decatopecten amiculum</i> (Philippi, 1851)
			<i>Decatopecten plica</i> (Linnaeus, 1758)
		<i>Glorichlamys</i> (Dijkstra, 1991)	<i>Glorichlamys elegantissima</i> (Deshayes in Maillard, 1863)
		<i>Gloripallium</i> (Iredale, 1939)	<i>Gloripallium pallium</i> (Linnaeus, 1758)
		<i>Juxtamusium</i>	<i>Juxtamusium maldivense</i>

		(Iredale, 1939)	(Smith, 1903)
		<i>Mirapecten</i>	<i>Mirapecten tuberosus</i>
		(Dall, Bartsch and Rehder, 1938)	(Dijkstra and Kilburn, 2001)
			<i>Pecten aribenedictus</i>
Pectinini	<i>Pecten</i>		(Dijkstra and Kilburn, 1995)
(Wilkes, 1810)	(Müller, 1776)		<i>Pecten sulcicostatus</i>
			(Sowerby II, 1842)
			<i>Laevichlamys deliciosa</i>
			(Iredale, 1939)
	<i>Laevichlamys</i>		<i>Laevichlamys lemniscata</i>
	(Waller, 1993)		(Reeve, 1853)
			<i>Laevichlamys weberi</i>
			(Bavay, 1904)
Chlamydinae	Chlamydini	<i>Pedum</i>	<i>Pedum spondyloideum</i>
(von Teppnes,	(von Teppnes, 1922)	(Lamarck, 1799)	(Gmelin, 1791)
1922)			<i>Semipallium coruscans coruscans</i>
			(Hinds, 1845)
		<i>Semipallium</i>	<i>Semipallium crouchi</i>
		(Lamy, 1928)	(E.A. Smith, 1892)
			<i>Semipallium flavicans</i>
			(Linnaeus, 1758)

		<i>Talochlamys humilis</i> (Sowerby III, 1904)
	<i>Talochlamys</i> (Iredale, 1929)	<i>Talochlamys multistriata</i> (Poli, 1795)
	<i>Veprichlamys</i> (Iredale, 1929)	<i>Veprichlamys africana</i> (Dijkstra and Kilburn, 2001)
Mimachlamydini (Waller, 1993)	<i>Mimachlamys</i> (Iredale, 1939)	<i>Mimachlamys sanguinea</i> (Linnaeus, 1758)
	<i>Aequipecten</i> (Fischer, 1886)	<i>Aequipecten commutatus peripheralis</i> (Dijkstra & Kilburn, 2001)
		<i>Cryptopecten bullatus</i> (Dautzenberg and Bavay, 1912)
Aequipectinini (Waller, 1993)	<i>Cryptopecten</i> (Dall, Bartsch and Rehder, 1938)	<i>Cryptopecten nux</i> (Reeve, 1853)
	<i>Haumea</i> (Dall, Bartsch and Rehder, 1938)	<i>Haumea minuta</i> (Linnaeus, 1758)
	<i>Volachlamys</i> (Iredale, 1939)	<i>Volachlamys fultoni</i> (Sowerby III, 1904)

Of the 29 South African scallop species, *Pecten aribenedictus* and *P. sulcicostatus* are the only *Pecten* species found in the waters around South Africa (Dijkstra and Kilburn 2001). *Pecten sulcicostatus* originated from the Mediterranean/West Africa (Dijkstra and Kilburn 2001) and, at present, has a natural distribution range along the inner continental shelf from False Bay to East London, South Africa (Figure 1.2). These scallops can be found at sub-littoral depths between 22 and 70 meters (Arendse *et al.* 2008), although the highest catches are found at depths of approximately 40 meters (Arendse and Pitcher 2012). These scallops are free-living on either sand or mud and interestingly, individuals inhabiting shallower water are smaller and often more brightly patterned than those inhabiting the continental shelf (Dijkstra and Kilburn 2001).



Figure 1.2: Putative natural distribution range of *P. sulcicostatus* along the inner continental shelf from False Bay to East London, South Africa.

Pecten sulcicostatus is considered to be closely related to the Atlantic *Pecten maximus* and Mediterranean *Pecten jacobaeus* on the basis of morphology (Saavedra and Peña 2004). All three *Pecten* species have similar shell shape and size (Figure 1.3), however, *P. jacobaeus* has more pronounced and square radial costae while *P. maximus* has flatter radial costae (Figure 1.4). *Pecten sulcicostatus* can be distinguished from the European *Pecten* species (*P. maximus* and *P. jacobaeus*) by the rough surface of both valves caused by secondary radial riblets.

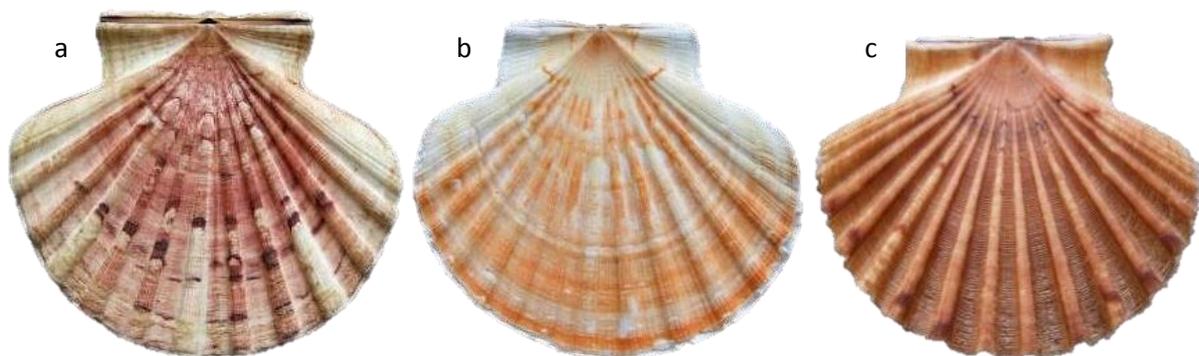


Figure 1.3: Shell morphology of (a) *Pecten sulcicostatus*, (b) *Pecten maximus* and (c) *Pecten jacobaeus*, showing the similar shell shape and size of all three *Pecten* species. Taken from Pecten Site (www.pectensite.com/).



Figure 1.4: Shell morphological characteristics of (a) *Pecten sulcicostatus*, (b) *Pecten maximus* and (c) *Pecten jacobaeus*, demonstrating how *Pecten sulcicostatus* (a) can be distinguished from *Pecten maximus* (b) and *Pecten jacobaeus* (c) by the rough surface of the valve due to secondary radial riblets. Shell morphological characteristics also show the more pronounced and square radial costae of *Pecten jacobaeus* (c) in comparison to the flatter radial costae of *Pecten maximus* (b). Taken from Pecten Site (www.pectensite.com/).

A study by Barucca *et al.* (2004), based on the mitochondrial 12S and 16S rRNA genes, found the genetic distance between *P. maximus* and *P. jacobaeus* to be

marginal. An earlier study by Wilding *et al.* (1999), based on two PCR-RFLP mitochondrial DNA (mtDNA) markers, determined the nucleotide divergence to be 0.045% between these two species. Mitochondrial DNA divergence is estimated at approximately 2% per million years (MY), therefore after five MY of separation, which is thought to have occurred during the Messinian salinity crisis (Ríos *et al.* 2002), a nucleotide divergence of approximately 10% would be expected, contradicting the 0.045% obtained in this study. Therefore the small genetic distance observed is regarded as being more consistent with two populations of the same species rather than with distinct species. According to Barucca *et al.* (2004), the two *Pecten* species may represent two varieties of the same species, which following adaptation to different environmental conditions, developed different morphological characteristics. This is further supported by the ability of *P. maximus* and *P. jacobaeus* to fully interbreed in captivity (Saavedra and Peña 2005).

Although *P. sulcicostatus* is presumed to be of Mediterranean/West African origin it is morphologically similar to the Eastern Atlantic *P. maximus*. The wide disjunction between the modern distribution ranges of these two species (over 9 000 km) suggests that vicariance occurred before the Pleistocene (2.6 MYA to 11 700 years ago) (Dijkstra and Kilburn 2001).

2.2 Scallop Biology and Ecology

Scallops exhibit a diverse set of lifestyles which are organised into six categories (Table 1.3) based on the methods and permanence of attachment to a substrate, locomotive ability and spatial relationship to a substrate (Alejandrino *et al.* 2011). Most scallop species are free-living, living in soft sand or mud where they swim short distances of more than five meters by rhythmically and rapidly opening and closing their valves. Other species are byssate, allowing for release and reorientation while a few species live permanently attached to rocky substrates as adults (Barucca *et al.* 2004).

Table 1.3: Descriptions of life habit classes in the family Pectinidae.

Life habit	Description
Nestle	Scallops settle and byssally attach to living Porites corals after which the coral grows around and permanently contains the scallop.
Cement	Scallops permanently attach to hard or heavy substratum.
Byssate	Scallops temporarily attach to a substratum through byssus threads, but are able to release and reorient.
Recess	Scallops excavate a cavity in soft sediment, which results in full or partial concealment.
Free-living	Scallops rest above soft sediment or hard substratum.
Gliding	Scallops are able to swim more than 5 meters per effort, which includes a level swimming phase with a glide component.

Scallop shells vary in colour, shape, texture and size depending on the species, however, all scallops have two valves, which are equal in length and width. Shell colouration is highly variable both within and between species, with the colourful upper valve providing camouflage while the bottom, more convex valve is usually much lighter (white or yellow) (Branch *et al.* 2005).

Pecten sulcicostatus has a strongly ribbed, pale coloured shell, with the lower valve curving outwards (convex) and the upper valve being flat. Both valves are sculptured with 12 to 15 radial costae with secondary radial riblets (Figure 1.5). The base of the shell has a pair of protrusions, known as auricles, which are equal in size (Branch *et al.* 2005). *Pecten sulcicostatus* shell length ranges from 60 to 100 mm, with an average shell length of 94 mm, although some animals exceed 150 mm. Interestingly, shallow water individuals are often smaller and more brightly patterned than those from the continental shelf (Dijkstra and Kilburn 2001), most likely due to differences in the intake of phytoplankton and organic detritus.

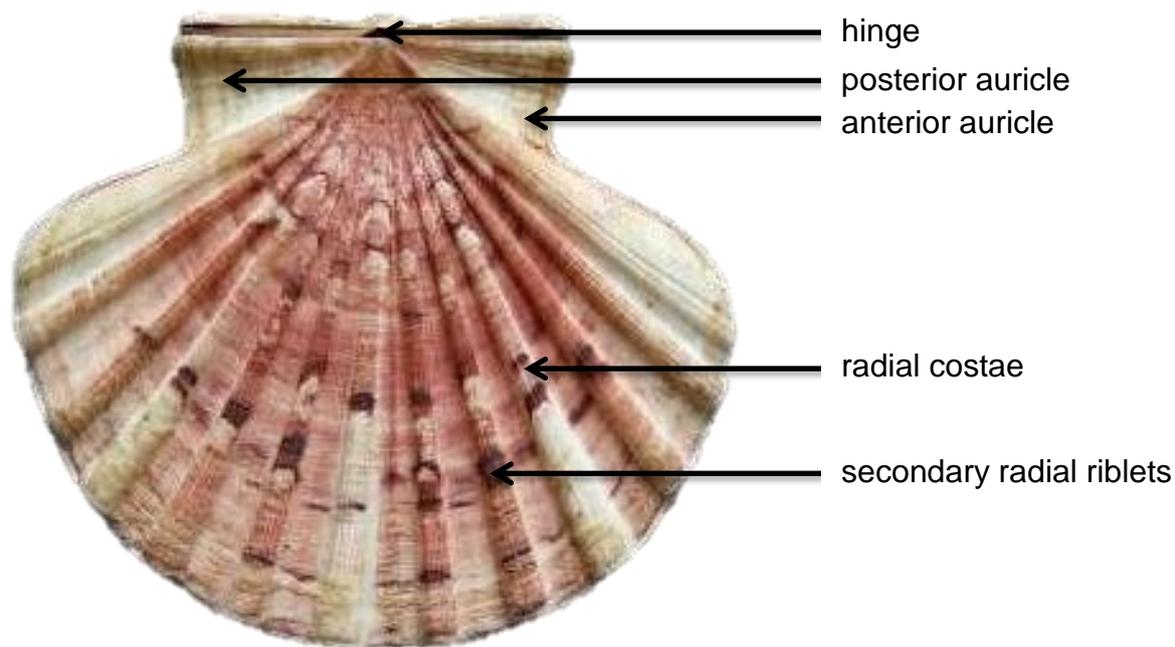


Figure 1.5: Upper valve of the *Pecten sulcicostatus* shell with key morphological characteristics. The upper valve is sculptured with 12 to 15 radial costae with secondary radial riblets. The base of the shell has a pair of auricles, which are equal in size and the hinge is located between the two auricles.

Pectinid shells are highly conserved in shape and offer few diagnostic characteristics, confounding phylogenetic inference (Alejandrino *et al.* 2011). To address the phylogenetic relationships among pectinids, Waller (1991) proposed a system based on microsculptural shell features and the morphological characteristics of juveniles, as, at this stage, scallops show few traces of the changes that provide adaptation to different habitats in adult life. Dijkstra and Kilburn's (2001) attempts at developing a dichotomous key for the Pectinidae occurring in the waters surrounding Southern Africa was unsuccessful due to the variability of many characters within most genera, as well as the difficulty of defining degree of development of these characters. Instead, Dijkstra and Kilburn made use of diagnoses of generic characters using shell characters.

Mantle musculature tissue (Figure 1.6) encloses the internal organs while the adductor muscle holds the two valves together, aiding the scallop to swim by rapidly opening and closing the valves and thrusting itself through the water. Scallop

tentacles are located within the folds of the free edge of the mantle musculature and are covered with epithelial sensory cells, which provide an early warning by detecting chemicals associated with approaching predators. Scallops also have approximately 100 eyes surrounding the mantle musculature that can detect light, dark and motion (Shumway and Parsons 2006).

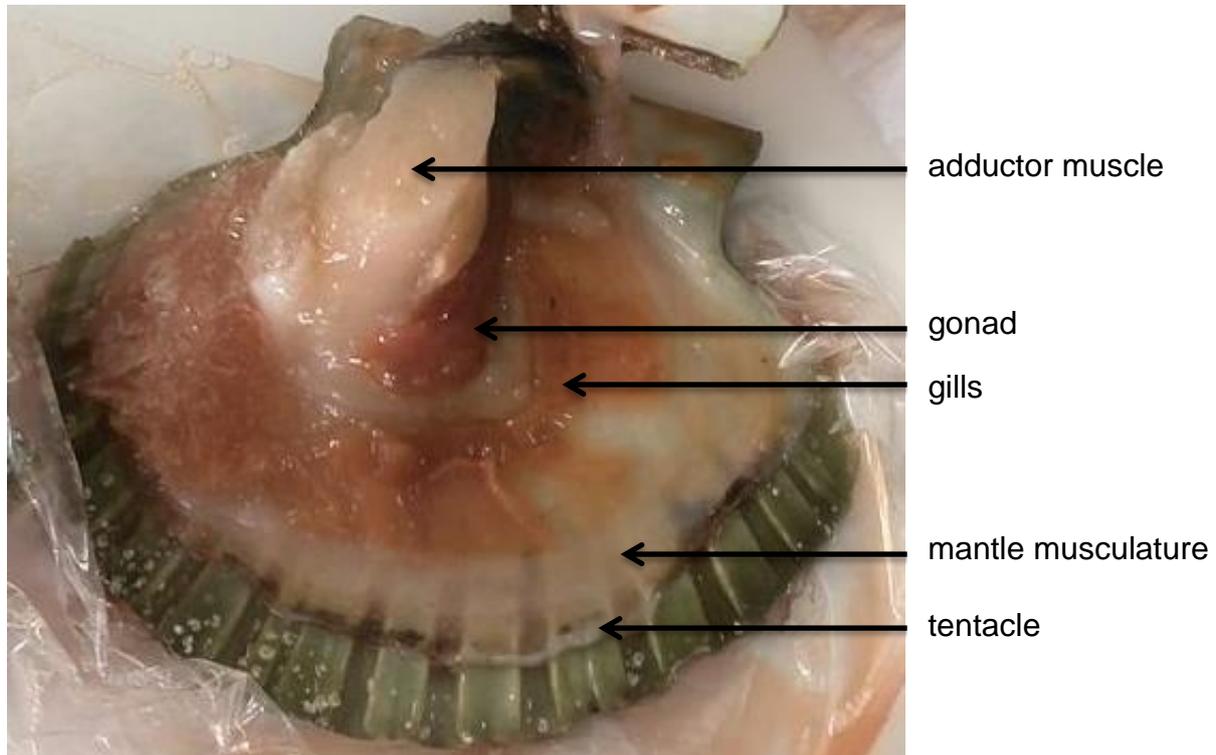


Figure 1.6: Internal anatomy of *Pecten sulcicostatus*. Mantle musculature tissue encloses the internal organs (gonad and gills) while the adductor muscle holds the two valves together. Tentacles are located within the folds of the free edge of the mantle musculature.

Many scallops are hermaphrodites, while others are dioecious, having definite sexes. A few scallop species are protandrous hermaphrodites, being male when young and then switching to female later in life. Most scallop species become sexually mature at two years of age, but do not contribute significantly to egg and sperm production until the age of four (Hart and Chute 2004). Reproduction takes place externally through broadcast spawning in which eggs and sperm (Figure 1.7) are released into the water. Simultaneously functional hermaphrodites are capable of self-fertilisation, although this results in low larval survival, likely a result of inbreeding depression. In

order to minimise self-fertilisation, scallops release eggs and sperm at different times within a single spawning event, usually starting with sperm and then switching to eggs. It is, however, possible for scallops to switch back and forth between sexual products during a single spawning event (Arendse *et al.* 2008).

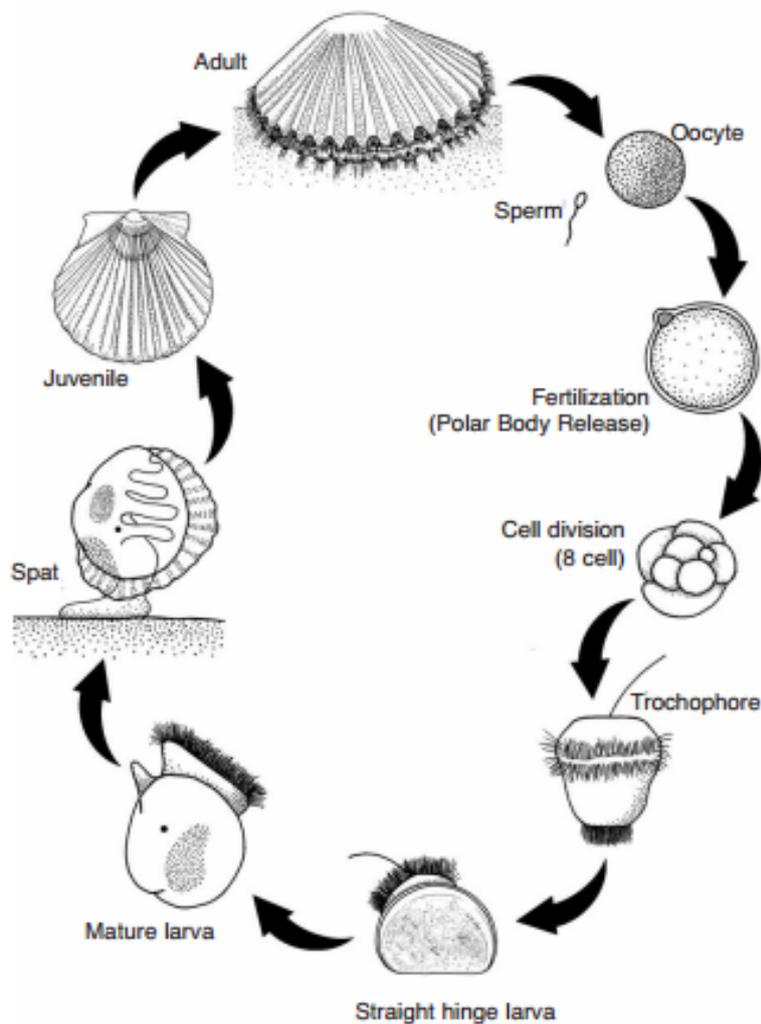


Figure 1.7: Depiction of the developmental stages in the life cycle of scallops. Broadcast spawning occurs primarily in summer, after which larvae remain in the water column for two to four weeks before dissipating to the ocean floor where they attach themselves to a substratum through byssus threads. Rapid growth within the first years of a scallop's life allows most scallops to reach commercial size at four to five years of age. Taken from Leavitt *et al.* (2010).

Most scallop species spawn primarily in summer, after which larvae remain in the water column for two to four weeks before dissipating to the ocean floor where they

attach themselves to a substratum through byssus threads (Figure 1.7). Byssus threads are eventually lost with adulthood, transitioning almost all scallop species into free swimmers. There is rapid growth within the first years of a scallop's life, with an increase of 50 to 80% in shell height and the quadrupling of meat weight (Leavitt *et al.* 2010). Most scallops reach commercial size at approximately four to five years of age, however, some scallops have been known to live more than 20 years (Arendse *et al.* 2008).

Pecten sulcicostatus seems to be a functional hermaphrodite (Arendse *et al.* 2008). Unlike most scallop species, which spawn primarily in summer, *P. sulcicostatus* spawns in winter and early spring (between June and September, peaking in August and September), similar to the Australian scallop, *Pecten fumatus*. Developing *P. sulcicostatus* oocytes are present throughout the year, likely due to the abundant availability of phytoplankton throughout the year (Arendse *et al.* 2008).

Scallops are filter feeders capable of ingesting phytoplankton and organic detritus from sea water. The microalgae *Isochrysis galbana* and marine planktonic diatom *Chaetoceros neogracile* are commonly used in scallop aquaculture, although the dietary requirements of scallops differ depending on species and life stage. For example, increased protein content of the phytoplankton diet of broodstock has been shown to reduce time to spawning maturity and increase fecundity. Similar positive results for growth and survival have been observed in larvae fed with high protein diets (Shumway and Parsons 2006).

3. Fisheries and Aquaculture

Aquaculture is defined as the farming of aquatic organisms including fish, molluscs, crustaceans and plants in controlled aquatic (both fresh- and saltwater) environments with some form of intervention in the rearing process. Intervention in the rearing process includes regular stocking, feeding and protection from predators, all of which enhances production (Chief Directorate: DAFF 2012).

South Africa has suitable environmental conditions for aquaculture development and opportunities for commercial production of various cultured species. Aquaculture is relatively new in South Africa and, despite the fact that South African aquaculture

was historically focused on high value species such as abalone, mussels and oysters, it has been identified as an area for expansion (McCord and Zweig 2011) as the South African aquaculture sector has performed below its potential, remaining a minor contributor to national fishery products and the country's Gross Domestic Product (GDP). Currently, the South African aquaculture sector is divided into two sub-sectors: freshwater aquaculture, which consists of species such as carp, trout, tilapia and catfish; and marine aquaculture, which consists of species such as abalone, oysters, mussels, kob, yellowtail and seaweed (Chief Directorate: DAFF 2012).

Currently, the South African marine aquaculture sector is estimated at approximately ZAR379 million, yet there is still room for growth. Marine aquaculture is a major contribution to the economy of countries such as China, Japan and the United States of America, therefore South Africa needs to ensure that this sector grows to its full potential. The growth of the South African aquaculture industry has the potential to contribute to the economy, poverty reduction, empowerment, employment and the sustainable use of coastal and inland resources to the benefit of local communities (Chief Directorate: DAFF 2012).

In 2012, global mollusc production in aquaculture reached a volume of 15.17 million tons, representing 23% of total aquaculture production, making molluscs the second highest category of aquaculture products after finfish (Astorga 2014). Clams and oysters have the highest production levels, followed in by mussels, scallops and abalone. Scallops are regarded as premium seafood world-wide with growing markets in Europe, North America and Asia (Arendse and Pitcher 2012). The first attempts to cultivate scallops were recorded in the 1950s and 1960s. In the past, fishing for wild scallops was the preferred practice, as intensive farming operations were costly. Recently, however, worldwide declines in wild scallop populations have resulted in the growth of scallop aquaculture. Scallops have an aquaculture production volume of 1.52 million tons, representing 10% of the total global mollusc production. There are ten scallop species that support commercially well-established aquaculture operations, as well as a few scallop species that support aquaculture operations that are under development. China and Japan are major contributors to the global scallop aquaculture industry (Table 1.4), cultivating four main scallop

species, namely *Argopecten irradians*, *Patinopecten yessoensis*, *Chlamys farreri* and *Chlamys nobilis*. Peru's aquaculture production of *Argopecten purpuratis* is also a significant contributor to the global scallop aquaculture industry (Table 1.4) (Astorga 2014). At present, Africa makes no contribution to the global scallop aquaculture industry, therefore the establishment of a scallop aquaculture industry in South Africa has great economic potential.

Table 1.4: The top five scallop species utilised in aquaculture according to production in tons (Astorga 2014).

Species	Country	Production (tons)
<i>Argopecten irradians</i>	China	850 000
<i>Patinopecten yessoensis</i>	Japan	546 749
<i>Chlamys farreri</i>	China	300 000
<i>Argopecten purpuratis</i>	Peru	58 101
<i>Chlamys nobilis</i>	China	50 000

There is a variety of aquaculture methods currently utilised in scallop aquaculture, however, the effectiveness of each method depends on the species of scallop as well as the local environment. Once scallops have been grown, harvested and processed the meat end product consists of the adductor muscle, either fresh or frozen. Top quality scallop adductor muscle can demand a high market price, which fluctuates depending on production, success of wild scallop fisheries and other global factors (Shumway and Parsons 2006).

Around the world, scallop aquaculture is considered to be a sustainable practice, with potentially positive effects on the ecosystem, due to the fact that scallops, as filter-feeding bivalves, are able to remove suspended solids, silt, unwanted nutrients, bacteria and viruses from the water thereby increasing water clarity which, in turn, improves pelagic and benthic ecosystems. Not only does aquaculture provide the resources to help feed a growing global population, it also reduces fishing pressure on wild stocks if the species can be produced through aquaculture. By decreasing the dependency on wild stocks, aquaculture allows natural populations to recover from overexploitation. Care should, however, be taken as positive impacts can be

area specific and scallop aquaculture could result in the eutrophication of water. Furthermore, aquaculture farms are a source of visual pollution and so often attract public opposition in highly populated coastal areas. Aquaculture may also compromise wild population gene pools if farmed animals and wild animals interbreed (Shumway and Parsons 2006).

3.1 Potential South African Scallop Industry

Large-scale aquaculture production of *P. maximus*, a sister species to *P. sulcicostatus*, was initiated in Europe during the 1980's. In more recent years the success of scallop culture, particularly the success of *C. nobilis* aquaculture in Japan and China (Table 1.4), has made *P. sulcicostatus* a prime candidate species for aquaculture in South Africa (Arendse and Pitcher 2012).

In 1972, exploratory fishing for *P. sulcicostatus* in False Bay revealed one to 106 scallops per 10 minute trawl with the highest densities near the centre of the bay (Arendse *et al.* 2008). The population size, however, was determined to be insufficient and the exploitable area was determined to be too small to support a viable fishery. Exploratory fishing in Mossel Bay revealed that the population size was also inadequate (Dijkstra and Kilburn 2001), hence any commercial utility of *P. sulcicostatus* would have to come from aquaculture. In this regard, initial studies on *P. sulcicostatus* have already focused on reproduction, life history characteristics (Arendse *et al.* 2008) and growth rates (Arendse and Pitcher 2012).

To date, *P. sulcicostatus* have been successfully spawned, in captivity, at the Sea Point Research Aquarium (Cape Town, South Africa) for pilot trials, however large scale commercial inductions has not yet been performed. A major determinant of the economic feasibility of scallop culture is the growth rate of the species, as the cost of production is largely determined by the length of the grow-out period. In 2012, a study was conducted on the growth and survival of hatchery-produced juveniles of *P. sulcicostatus* in suspended culture in Saldanha Bay. The influence of environmental conditions on the growth and survival of *P. sulcicostatus* was examined in relation to measures of temperature and phytoplankton biomass (Arendse and Pitcher 2012). The study found that the mean growth rate of 0.10 mm day⁻¹ compared favourably with other commercially cultured species such as *C. farreri* (mean growth rate of 0.09 mm day⁻¹) (Guo and Luo 2006), *P. yessoensis*

(mean growth rate of 0.10 mm day⁻¹) (Ventilla 1982), *P. maximus* (growth rate of 0.05 - 0.23 mm day⁻¹) (Louro *et al.* 2005), *Argopecten purpuratus* (Lamarck, 1819) (mean growth rate of 0.15 mm day⁻¹) (Von Brand *et al.* 2006) and *Placopecten magellanicus* (Gmelin 1791) (growth rate of 0.04 - 0.12 mm day⁻¹) (Kleinman *et al.* 1996). The mean growth rate of 0.10 mm day⁻¹ exceeded previous estimates of growth of naturally occurring populations of *P. sulcicostatus*. Scallop growth was poorly correlated with temperature; however the lowest growth rates coincided with maximum temperatures as well as highly variable temperatures over short periods of time. These periods of high temperatures or highly variable temperatures also correlated with high scallop mortalities, therefore the high and variable surface temperatures of Saldanha Bay during summer was concluded as an unsuitable environment for the grow-out of *P. sulcicostatus*. It has been proposed, however, that the deeper waters of Saldanha Bay may provide a more suitable environment due to lower temperatures as well as less temperature fluctuation over short periods of time (Arendse and Pitcher 2012).

4. Molecular Markers

Molecular markers have become increasingly significant due to the importance placed on genetic indicators for species conservation, the sustainability of natural populations as well as the management of these populations. Use of the genetic information determined using molecular markers allows the species in question to be developed sustainably, maintaining it for future generations. The information from molecular markers also forms the basis for genetic improvement programmes in species in both the early or advanced stages of aquaculture (Astorga 2014).

Genetic variation is measured using molecular markers, which are defined as any sequence variation or polymorphism between individuals that is inherited in a Mendelian fashion. Molecular markers include insertions and deletions, segment inversions and rearrangements, nucleotide base pair substitutions as well as variable number of tandem repeats (VNTRs). These markers provide a means to evaluate genome-wide genetic variation within and between individuals, populations and species. Molecular markers can be used in various applications in fisheries management and aquaculture including linkage mapping, population studies,

individual identification, parentage assignment and strain/species identification (Liu and Cordes 2004).

4.1 Nuclear Molecular Markers

Allozymes, variant forms of an enzyme that are coded by different alleles at the same locus, were the first molecular marker used in animal genetics, including fisheries and aquaculture (May *et al.* 1980; Seeb and Seeb 1987). Allozymes were used in aquaculture for assessing inbreeding, stock identification and parentage analysis; however the limited number of allozyme loci precludes their use in large-scale genome mapping. Other disadvantages of allozymes include heterozygote deficiencies, due to null alleles as well as the amount and quality of tissue samples required. Also some changes in DNA sequence are masked at the protein level, reducing the level of detectable variation. Allozyme studies revealed low levels of genetic variation, prompting a search for markers with greater genetic resolution (Liu and Cordes 2004).

In 1980 minisatellites, repetitive DNA in which motifs are repeated 5 to 50 times, were discovered (Nakamura *et al.* 1987). Minisatellites enabled genetic identification, but could not be used for applications such as population genetics due to the complexity of the banding patterns they produced (Wan *et al.* 2004). The mid 1990's saw the rise of a second wave of VNTRs, namely microsatellite markers, which quickly became the marker of choice in population genetic studies (O'Connell and Wright 1997). It was, however, the development of the polymerase chain reaction (PCR), by Kary Mullis that paved the way for wide scale use of DNA-based molecular markers. The development of PCR meant that any genomic region could now be studied following PCR amplification (Schlötterer 2004). Microsatellite markers were the first markers to be used in conjunction with PCR amplification (Barbará *et al.* 2007).

4.1.1 Microsatellite Markers: Characteristics and Development

Molecular markers, such as microsatellite markers and single nucleotide polymorphisms (SNPs), are based on the detection of the particular sequence variation. Microsatellite markers are a group of repetitive DNA elements that comprise multiple copies of tandemly organised simple sequence repeats (SSRs)

which range in size from two to six base pairs and are therefore classified as di-, tri-, tetra-, penta- and hexanucleotide repeats based on the number of nucleotides per repeat unit. Dinucleotide repeats, in particular $(CA)_n/(TG)_n$ and $(AT)_n/(TA)_n$ repeat are most abundant (Ellegren 2004). Microsatellite markers are uniformly distributed throughout the genomes of all known organisms at densities proportional to genome size and these markers can be found in introns, gene coding regions, as well as non-gene sequences (Liu and Cordes 2004; Hoffman and Nichols 2011).

High levels of polymorphism, co-dominant mode of inheritance, multiplexing potential, semi-automation and fluorescent dye capillary electrophoresis (CE) systems coupled to computer imaging programmes for easy allele scoring has made microsatellite markers one of the most popular molecular markers to date. The large number of alleles per locus results in the highest polymorphic information content (PIC) values of any DNA markers. Despite this, microsatellite markers suffer from technical difficulties, such as null alleles (failure of allele amplification) and stuttering (*in vitro* slippage of *Taq* polymerase causing multiple bandings of a single allele), leading to genotyping errors (Hoffman and Nichols 2011). Microsatellite markers have been used extensively in fisheries and aquaculture research including studies of genome mapping, parentage, kinship and genetic stock structure (Liu and Cordes 2004).

Traditionally, microsatellite markers are isolated by following three steps: construction of a partial genomic library, screening for positive clones and marker-specific primer design and optimisation. Initially, partial genomic libraries were constructed by selecting genomic fragments based on size. Clones were then screened via colony hybridisation using repeat probes, positive clones were sequenced and from the sequences, repeat flanking primers were designed for PCR optimisation. Despite being a relatively simple process, positive clone yields averaged between two and three percent, making this method inefficient (Hoffman and Nichols 2011). Later methods made use of microsatellite-enriched genomic DNA libraries which were used to develop a much higher number of microsatellite markers (Liu and Cordes 2004). *De novo* microsatellite marker development can, however, be both time-consuming and costly, therefore once a microsatellite marker set has been developed for a focal species, time and effort can be saved on marker

development in closely related species by using cross-species transfer of microsatellite markers. For microsatellite markers, closely related taxa, such as species belonging to the same genus or recently diverged genera, have a relatively high success rate with cross-species amplification (Zane *et al.* 2002). It should, however, be noted that this success rate decreases with genetic distance between species (Arif *et al.* 2010). Furthermore, even though cross-species amplification may be successful, high levels of polymorphism cannot be guaranteed (Chambers and MacAvoy 2002; Zane *et al.* 2002) and genetic diversity may be underestimated by the use of non-species specific microsatellite markers (Arif *et al.* 2010).

4.1.2 Single Nucleotide Polymorphisms (SNPs): Characteristics and Development

Single nucleotide polymorphisms (SNPs) are single base changes in a DNA sequence, with two possible nucleotides at a given position. These bi-allelic markers are inherited co-dominantly and represent the most abundant type of genetic marker in any organism's genome (Chauhan and Rajiv 2010), occurring every 0.3 to 1 kb (Lui and Cordes 2004). This high density makes SNPs ideal for studying the inheritance of genomic regions (Baird *et al.* 2008). Although DNA sequencing allowed for the characterisation of SNPs, it was not until the development of gene chip technology in the late 1990s that genotyping large numbers of SNPs became possible. When a purine to purine or pyrimidine to pyrimidine substitution occurs the SNP is classified as a transition. On the other hand if a purine to pyrimidine or pyrimidine to purine substitution occurs the SNP is classified as a transversion. In theory the transition to transversion (ts:tv) ratio should be one to one, however transitions are more common, possibly due to high rates of spontaneous deamination of cytosine which leads to the overrepresentation of C to T or T to C transitions (Vignal 2002). SNPs have become one of the most promising markers as they possess high information content, can be automated and can be used as a powerful analytical tool for various genetic applications. SNPs may be deemed superior to microsatellite markers as they are mutationally more stable, therefore their inheritance conforms more strictly to Mendelian expectations, increasing their resolving power by being less prone to homoplasia (Lui and Cordes 2004). Despite this, due to their bi-allelic nature which results in low PIC values, 30 to 50 SNPs

would be needed to provide equal information content to 10 to 15 microsatellite markers (Aitken *et al.* 2004).

Single-strand conformation polymorphism (SSCP) analysis, heteroduplex analysis and DNA sequencing have been traditionally used for SNP discovery, although DNA sequencing is the most accurate and most used approach. Random shotgun sequencing, amplicon sequencing using PCR and comparative expressed sequence tag (EST) analysis are among the most popular sequencing methods for SNP discovery. The most basic strategy for SNP discovery is to screen ESTs generated from sequencing of cDNA clones in order to identify polymorphic sites. Previously, an alternative method involved the pooling of individuals' DNA and subsequent sequencing of this DNA using shotgun genome sequencing. This method, however, produces large amounts of data which can be time-consuming and difficult to analyse (Lui and Cordes 2004).

4.1.3 Next Generation Sequencing (NGS)

With the advancement of Next Generation Sequencing (NGS) technologies it is now possible to identify thousands of molecular markers through the generation of high genome coverage with increased throughput while reducing the cost of sequencing (Etter *et al.* 2011). Although DNA sequencing costs continue to drop, whole genome Next Generation Sequencing remains costly, therefore Reduced Representation Sequencing (RRS), which makes use of multiplexed shotgun genotyping (MSG) or Restriction-site Associated DNA (RAD) sequencing, provides a way of increasing sample number while maintaining practical DNA sequencing costs. These NGS methods depend on restriction enzymes (REs) to produce a reduced representation of a genome for use in genome-wide marker discovery (Davey *et al.* 2011). Reduced Representation Sequencing is an alternative method to whole genome resequencing, as it focuses on compact panels of genomic markers throughout the genome. More broadly, RRS facilitates rapid, inexpensive microsatellite marker and SNP discovery and these markers allow for large-scale genotyping (Altshuler *et al.* 2000). Genome subsampling can be attained by using Restriction-site Associated DNA (RAD) sequencing, which creates short DNA fragments which are adjacent to recognition sites of specific REs. Restriction-site Associated DNA sequencing is

useful when studying organisms lacking a reference genome, as it accelerates marker discovery (Etter *et al.* 2011).

Restriction-site Associated DNA sequencing entails the digestion of genomic DNA (gDNA) using REs followed by ligation of adapters (P1) to the compatible ends. The ligated adapter consists of a barcode differing by three nucleotides (for identification purposes), forward amplification as well as primer sites. Fragments containing the adapters are pooled and sheared at random, followed by an electrophoresis step in which fragments of a specific size are selected. A second adapter (P2) is then ligated onto the correctly sized fragments (Figure 1.8). The reverse primer can only bind to the fragment when both the adapters (P1 and P2) are present due to the Y shaped nature of the P2 adapter. The two strands are only complementary for part of the sequence due to the adapters' divergent ends; therefore only fragments ligated with both of the adapters will be amplified during the polymerase chain reaction step (Baird *et al* 2008; Etter *et al.* 2011).

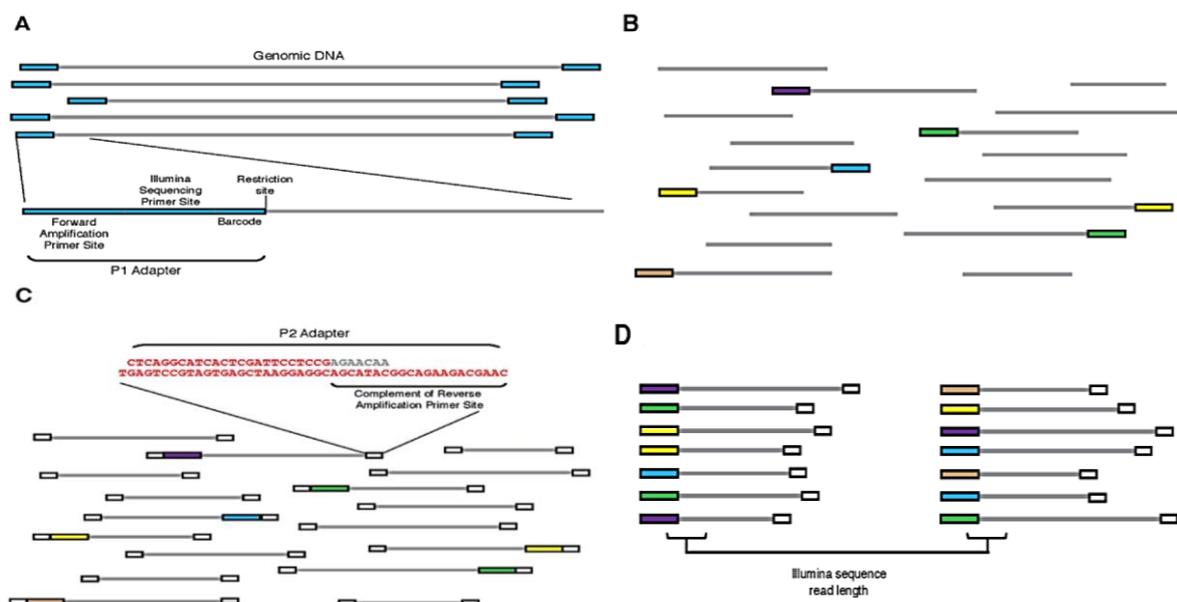


Figure 1.8: (A) The P1 adapter is ligated to digested gDNA. The P1 adapter contains a barcode, forward primer set and primer site. (B) Fragments are pooled and randomly sheared. (C) P2 adapters, containing a reverse complement of the reverse amplification primer site, are ligated onto the sheared fragments. (D) Enrichment of RAD tag containing both the P1 and P2 adapters. Taken from Baird *et al.* (2008).

With RAD sequencing, SNPs are detected by creating a reduced version of the genome, which allows the nucleotides next to the RE sites to be over-sequenced (Etter *et al.* 2011). Instead of using reads from the entire genome, which can be computationally straining, RAD sequencing produces short sheared-end sequencing which can then be compiled into larger contigs. Multiplexing with RAD tag libraries and parallel sequencing can lead to the generation of vast amounts of polymorphism data (Etter *et al.* 2011; Rowe *et al.* 2011; Davey *et al.* 2013).

Within the aquaculture environment, microsatellite markers and SNPs have been successfully applied over the last few decades in various applications including the characterisation of genetic stocks, parentage analysis, broodstock selection, linkage mapping, studies of quantitative trait loci, the evaluation of population-genetic differentiation as well as for monitoring genetic diversity (Guo *et al.* 2008; Kohlmann *et al.* 2010; Davey *et al.* 2011).

4.2 Mitochondrial DNA (mtDNA)

Studies have shown that sequence divergence accumulates more rapidly in mtDNA than in nuclear DNA. This rapid sequence divergence is due to the faster mutation rate, which is a result of a lack of repair mechanisms during replication and smaller effective population size due to the maternal inheritance of the haploid mitochondrial genome. The mtDNA molecule contains 12 protein coding genes and almost the entire mtDNA molecule is transcribed except for the 1 kb control region (D-loop), where replication and transcription is initiated. Non-coding sequences such as the D-loop exhibit more variation relative to coding sequences due to reduced functional constraints and relaxed selection pressure (Lui and Cordes 2004; Marín *et al.* 2015). Phylogenetic studies often make use of the Cytochrome c oxidase subunit I (CO1) and 16S rRNA genes (Barucca *et al.* 2004; Saavedra and Peña 2005; Yuan *et al.* 2012; Marín *et al.* 2013; Marín *et al.* 2015). Mitochondrial genes, such as the CO1 and 16S rRNA genes, are also commonly used in population genetic studies (Nagashima *et al.* 2005; Gaffney 2010; Yuan *et al.* 2012; Marín *et al.* 2013). Despite the fact that the 16S rRNA gene is one of the slowest evolving genes, it is often used in phylogenetic studies (Saavedra and Peña 2004; Saavedra and Peña 2005; Marín *et al.* 2013). As mtDNA is maternally inherited, the phylogenetic inferences and population genetic structuring derived from mtDNA data may not be completely

accurate if gender-biased migration, selection or introgression exists (Chauhan and Rajiv 2010). Mitochondrial DNA markers have also been used to investigate stock structure and identify broodstocks. Unfortunately mtDNA markers are subject to the same challenges that exist for other DNA markers: back mutation, where sites that have already undergone substitution are returned to their original state; parallel substitution, where mutations occur at the same site in independent lineages; as well as rate heterogeneity or mutational hot spots, where there are large differences in the rate at which some sites undergo mutation when compared to other sites in the same region (Lui and Cordes 2004).

4.3 Genetics and Molecular Markers in Aquaculture

The collapse of the global fisheries industry has sparked aquaculture endeavours to supply food for human nourishment. Aquaculture involves the establishment of strains of economically important animals that will outperform their wild counterparts in terms of production traits (Hulata 2001). The management and manipulation of genetic variation is, however, of the utmost importance in animal domestication and breeding (Bourdon 2000), as genetic factors are responsible for phenotypic variation. Effective animal breeding aims to increase the accuracy of breeding outcome predictions, secure animal welfare by preventing the introduction of deleterious genetic effects and effectively manage genetic diversity. It is therefore important to establish breeding objectives, models for the extent of genetic variation for important traits and determine the correlation between genotype and phenotype through the understanding of the molecular genetic processes (Flint and Woolliams 2008).

In order to create a genetically enhanced aquaculture strain a founder population representing the maximum genetic diversity of the wild population is needed. This founder population forms the baseline population for downstream artificial selection. A thorough understanding of the genetic history and structure of natural populations is therefore essential. A loss of genetic diversity in the founder population may lead to a loss of genetic variation that could have been capitalised on during selection programmes for advantageous traits (Flint and Woolliams 2008). Molecular markers can be used to maximise genetic diversity for aquaculture selective breeding schemes, by minimising kinship amongst broodstock individuals and selecting broodstock to maximise population heterozygosity (Hayes *et al.* 2006). The

establishment of the broodstock population simulates a bottleneck effect, as some of the natural genetic variation of the species is lost when individuals are chosen as broodstock, as they represent only a small fraction of the population (Gosling 2003). Genetic drift in small populations may lead to loss of potentially advantageous alleles or the fixation of detrimental alleles (Roodt-Wilding 2007), therefore it is important to maximise broodstock population genetic diversity. A further consequence of reduced population size is that over the following generations the offspring of these parents will be bred together, increasing the probability of inbreeding, resulting in a further loss of genetic variation (Gosling 2003). The number of individuals with advantageous alleles may be few therefore, when individuals are retained for breeding, the probability of these individuals being related is higher than that expected under natural conditions (Flint and Woolliams 2008). Management of hatcheries should therefore ensure long term retention of genetic diversity that will ensure sustainable breeding and long term genetic gains through artificial selection (Cardellino and Boyazoglu 2009).

Molecular markers have been used to investigate population structure and gene flow, as well as discriminate between taxa and investigate evolutionary relationships (Sweijd *et al.* 2000; Chistiakov *et al.* 2006; Wenne *et al.* 2007; Hauser and Seeb 2008). The use of microsatellite markers has assisted population genetic studies on a number of scallop species including *A. purpuratus* (Marín *et al.* 2013), *C. farreri* (Zhan *et al.* 2009; Yuan *et al.* 2012), *C. nobilis* (Wang *et al.* 2013), *Patinopecten caurinus* (Gaffney 2010), *P. magellanicus* (Kenchington *et al.* 2006), *Pecten novaezelandiae* (Silva and Gardner 2015) and *P. yessoensis* (An *et al.* 2009). A number of these population genetic studies used microsatellite markers in conjunction with a mitochondrial gene such as CO1 or the 16S rRNA gene (Gaffney 2010; Yuan *et al.* 2012; Marín *et al.* 2013), although older population genetic studies only made use of mitochondrial genes (Nagashima *et al.* 2005).

Mitochondrial genes, such as the 16S rRNA gene, have been used to study the evolutionary relationships of scallop species (Saavedra and Peña 2004; Saavedra and Peña 2005; Mahidol *et al.* 2007) as well as to assess the genetic differentiation of scallop populations in order to assess the genetic consequences of scallop transfers and introductions (Beaumont 2000). A study by Yuan *et al.* (2009) made

use of the 16S rRNA and *CO1* genes to investigate intraspecific genetic variation in domestic and wild populations of *C. nobilis*.

5. Rationale, Aims and Objectives

Despite the fact that *P. sulcicostatus* has been identified as having potential commercial value as an aquaculture species, no molecular markers are available for this species, as this species has, until now, not been studied genetically. No molecular studies have been conducted on the genetic composition of this species or the genetic constitution of these molluscs along the South African coast. Although the chromosome number of *P. sulcicostatus* has not been determined, the haploid chromosome number of three *Pecten* species (*P. maximus*, *P. jacobeus* and *P. albicans*) was determined to be 19 (Shumway and Parsons 2006), suggesting that this might be similar for *P. sulcicostatus* if haploid chromosome number is conserved in the *Pecten* genus.

Genetic resources, such as microsatellite markers, SNPs and mitochondrial genes, are vital for studying the genetic constitution of scallop populations along the South African coast. Developing genetic resources for this species will assist in the assessment of the genetic constitution of this species, which is important for formulating effective strategies for genetic management and sustainable utilisation of the species. Therefore the overall aim of this study is to develop molecular markers for *P. sulcicostatus* and to conduct a preliminary assessment of the genetic constitution of this species in order to aid the establishment of a sustainable aquaculture industry.

The first experimental chapter, chapter II, aimed to test the effectiveness of various tissue samples in DNA extraction as well as determine whether DNA extraction using dried tissue is feasible. Ideally the most effective tissue and DNA extraction protocol will entail an effective non-destructive sampling method for *P. sulcicostatus*, which can be implemented for use on broodstock individuals once an aquaculture industry is established for this endemic species. Chapter II entailed the extraction of gDNA from various tissue types as well as dried mantle musculature tissue followed by

amplification of a mitochondrial gene in order to test amplification of the extracted gDNA.

Chapter III aimed to optimise the amplification of mitochondrial genes and to establish microsatellite markers using cross-species transfer for phylogeographic and population genetic studies in *P. sulcicostatus*. Scallop phylogenetic classification systems differ greatly, therefore the evolutionary relationship among members of the Pectinidae family is poorly understood. In order to validate the optimised mitochondrial gene and prove its usefulness, Chapter III made use of the optimised mitochondrial gene in order to shed light on the evolutionary relationship of various scallop species in relation to *P. sulcicostatus*. Furthermore, in order to validate the set of microsatellite markers and prove their usefulness, the set of microsatellite markers was used to quantify the genetic diversity of a natural *P. sulcicostatus* population. These initial studies demonstrated the utility of both the mitochondrial marker as well as the set of microsatellite markers, allowing future studies to make use of these genetic resources in order to provide genetic information. Genetic information derived from these genetic resources can be used to develop strategies for genetic management within the context of establishing a sustainable aquaculture industry for this species. Chapter III entailed the extraction of gDNA from several *P. sulcicostatus* individuals as well as individuals representing various *Pecten* species. Mitochondrial markers were tested for amplification and optimised in order to optimise a mitochondrial marker which can be used in future studies. Microsatellite markers previously developed for *P. maximus* were then selected for cross-species transfer to the target species, *P. sulcicostatus*. A standardised microsatellite marker panel was subsequently constructed and used to quantify the genetic diversity of the False Bay *P. sulcicostatus* population.

Chapter IV aimed to make use of NGS technology in order to assess the utility of RRS and two bioinformatic approaches in the development of genetic resources for *P. sulcicostatus*. Development of species-specific microsatellite markers and SNPs will provide genetic resources for this species, which, in future studies, could be used to study the underlying population genetic structure of wild and cultured *P. sulcicostatus*. Chapter IV entailed extracting gDNA from *P. sulcicostatus* individuals followed by RRS through the use of NGS technology. This was followed

by bioinformatic analysis of sequencing data for microsatellite marker and SNP discovery and preliminary characterisation of this species.

6. References

- Aitken, N., S. Smith, C. Shwartz, and P. A. Morin, 2001 Single nucleotide polymorphisms (SNP) discovery: a targeted-gene approach. *Molecular Ecology* 13: 1423-1431.
- Alejandrino, A., L. Puslednik, and J. M. Serb, 2011 Convergent and parallel evolution in life habit of the scallops (Bivalvia: Pectinidae). *BMC Evolutionary Biology* 11: 164.
- Altshuler, D. V., J. Pollara, C. R. Cowles, W. J. Van Etten, J. Baldwin, L. Linton, and E. S. Lander, 2000 An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513-516.
- An, H. S., Y. G. Lee, J. Y. Park, and C. Lee, 2009 Genetic characterization of four East Asian giant scallop (*Mizuhopecten yessoensis*) populations using microsatellite markers. *Aquaculture Resources* 40: 619-624.
- Arendse, D. C. Z., and G. C Pitcher, 2012 Growth and survival of the South African scallop *Pecten sulcicostatus* in suspended culture. *African Journal of Marine Science* 34: 181-186.
- Arendse, D. C. Z., S. Andersen, N. J. Blake, and G. C. Pitcher, 2008 The reproductive cycle of the scallop *Pecten sulcicostatus* from the Southern Benguela upwelling system. *Journal of Shellfish Research* 27: 281-287.
- Arif, I. A., M. A. Bakir, H. A. Khan, A. H. Al Farhan, A. A. Al Homaidan, A. H. Bahkali, M. Al Sadoon, and M. Shobrak, 2010 A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences* 11: 2079-2096.
- Astorga, M. P., 2014 Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Frontiers in Genetics* 5: 1-6.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: 1-7.
- Barbará, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer, 2007 Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759-3767.

Barucca, M., E. Olmo, S. Schiaparelli, and A. Canapa, 2004 Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes. *Molecular Phylogenetics and Evolution* 31: 89-95.

Beaumont, A., 2000 Genetic considerations in transfers and introductions of scallops. *Aquaculture International* 8: 493-512.

Bourdon, R. M., 2000 *Understanding Animal Breeding* 2nd ed. Prentice-Hall, United States. Pages 198-226.

Branch, G.M., M.L. Branch, C.L. Griffiths, and L.E Beckley, 2008 *Two Oceans: a guide to the marine life of southern Africa*. Struik, South Africa. Pages 1-360.

Cardellino R. A., and J. Boyazoglu, 2009 Research opportunities in the field of animal genetic resources. *Livestock Science* 120: 166-173.

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

Chauhan, T., and K. Rajiv, 2010 Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology* 1: 281-291.

Chief Directorate: Aquaculture and Economic Development Fisheries Branch, Department of Agriculture, Forestry and Fisheries, 2012 *Aquaculture Yearbook 2012 South Africa*.

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

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews* 12: 499-510.

Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter, 2013 Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* 22: 3151-3164.

Dijkstra H. H., 1998 Pectinoidea (Mollusca: Bivalvia: Pectinidae: Propeamussiide) from Hansa Bay, Papua New Guinea. *Molluscan Research* 19: 11-52.

Dijkstra, H. H., and R. N. Kilburn, 2001 The family Pectinidae in South Africa and Mozambique (Mollusca: Bivalvia: Pectinoidea). *African Invertebrates* 42: 263-321.

Directorate: Marketing of the Department of Agriculture, Forestry and Fisheries, 2012 A profile of the South African aquaculture market value chain.

Ellegren, H., 2004 Microsatellite: simple sequences with complex evolution. *Nature Review Genetics* 5: 435-445.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson, 2011 Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE* 6: 1-10.

Flint A. P. F., and J. A. Woolliams, 2008 Precision animal breeding. *Philosophical Transactions of the Royal Society B* 363: 573-590.

Gaffney, P. M., C. M. Pascal, J. Barnhart, W. S. Grant, and J. E. Seeb, 2010 Genetic homogeneity of weathervane scallops (*Patinopecten caurinus*) in the northeastern Pacific. *Canadian Journal of Fisheries and Aquatic Sciences* 67: 1827-1839.

Gosling, E., 2003 *Bivalve Molluscs: Biology, Ecology and Culture*. Fishing News Books. Oxford, Blackwell Publishing. Page 337.

Guo X., and Luo Y., 2006 Scallop culture in China. In: Shumway, S. E., and G. J. Parsons (eds.), *Scallops: Biology, Ecology and Aquaculture*. Developments in Aquaculture and Fisheries Science. Amsterdam, Elsevier: 35: 1143-1161.

Guo, J. J., J. Zeng, S. L. Zhao, and Z. G. Zhao, 2008 Isolation and characterization of 19 microsatellite markers in a tropical and warm subtropical birch, *Betula alnoides*. *Molecular Ecology Notes* 8: 895-897.

Hart, D. R., and A. S. Chute, 2004 Essential fish habitat source document: sea scallop, *Placopecten magellanicus*, life history and habitat characteristics. NOAA Technical Memorandum NMFS-NE-189.

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

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

Hoffman, J. I., and H. J. Nichols, 2011 A novel approach for mining polymorphic microsatellite markers *in silico*. *PLoS ONE* 6: 1-9.

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

Kenchington, E. L., M. U. Patwary, E. Zouros, and C. J. Bird, 2006 Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology* 15: 1781-1796.

Kleinman S., B. G. Hatcher, R. E. Scheibling, L. H. Taylor, and A. W. Hennigar, 1996 Shell and tissue growth of juvenile sea scallops (*Placopecten magellanicus*) in suspended and bottom culture in Lunenburg Bay, Nova Scotia. *Aquaculture* 142: 75-97.

Kohlmann, K., P. Kersten, R. Panicz, D. Memis, and M. Flajs̃hans, 2010 Genetic variability and differentiation of wild and cultured tench populations inferred from microsatellite loci. *Reviews in Fish Biology and Fisheries*. 20: 279-288.

Leavitt, D. R. Karney, and A. Surier, 2010 Biology of the bay scallop. Northeastern Region Aquaculture Center Publication 213: 1-8.

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

Louro A., G. Christophersen, T. Magnesen, and G. Román, 2005 Suspension culture of the great scallop *Pecten maximus* in Galicia, NW Spain – intermediate primary culture of hatchery produced spat. *Journal of Shellfish Research* 24: 61-68.

Mahidol, C., U. Na-Nakorn, S. Sukmanomon, W. Yoosuk, N. Taniguchi, and T. T. T. Nguyen, 2007 Phylogenetic relationships among nine scallop species (Bivalvia: Pectinidae) inferred from nucleotide sequencing of one mitochondrial and three nuclear gene regions. *Journal of Shellfish Research* 26: 25-32.

Marín, A., T. Fujimoto, and K. Arai, 2013 Genetic structure of the Peruvian scallop *Argopecten purpuratus* inferred from mitochondrial and nuclear DNA variation. *Marine Genomics* 9: 1-8.

Marín, A., T. Fujimoto, and K. Arai, 2015 The mitochondrial genomes of *Pecten albicans* and *Pecten maximus* (Bivalvia: Pectinidae) reveal a novel gene arrangement with low genetic differentiation. *Biochemical Systematics and Ecology* 61: 208-217.

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

McCord, M., and T. Zweig, 2011 Fisheries: Facts and Trends South Africa. WWF Report South Africa 2011. WWF-SA, South Africa. Pages 3-32.

Nagashima, K., M. Sato, K. Kawamata, A. Nakamura, and T. Ohta, 2005 Genetic structure of Japanese scallop population in Hokkaido, analyzed by mitochondrial haplotype distribution. *Marine Biotechnology* 7: 1-10.

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

O'Connell, M., and J. M. Wright, 1997. Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* 7: 331-363.

Puslednik, L., and J. M. Serb, 2008 Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. *Molecular Phylogenetics and Evolution* 48: 1178-1188.

Raines, B. K., and Poppe, G. T. 2006 A conchological iconography. The family Pectinidae. ConchBooks, Hackenheim. Page 402.

Ríos, C., S. Sanz, C. Saavedra, and J.B. Peña, 2002 Allozyme variation in populations of scallops, *Pecten jacobaeus* (L.) and *P. maximus* (L.) (Bivalvia:Pectinidae), across the Almeria-Oran front. *Journal of Experimental Marine Biology and Ecology* 267: 223-244.

- Roodt-Wilding, R., 2007 Abalone ranching: a review on genetic considerations. *Aquaculture Resources* 38: 1229-1241.
- Rowe, H. C., S. Renaut, and A. Guggisberg, 2011 RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 1-4.
- Saavedra, C., and J. B. Peña, 2004 Phylogenetic relationships of commercial European and Australasian king scallops (*Pecten* spp.) based on partial 16S ribosomal RNA gene sequences. *Aquaculture* 235: 153-166.
- Saavedra, C., and J. B. Peña, 2005 Nucleotide diversity and Pleistocene population expansion in Atlantic and Mediterranean scallops (*Pecten maximus* and *P. jacobaeus*) as revealed by the mitochondrial 16S ribosomal RNA gene. *Journal of Experimental Marine Biology and Ecology* 323: 138-150.
- Schlötterer, C., 2004 The evolution of molecular markers - just a matter of fashion? *Genetics* 5: 63-69.
- Seeb J. E., and L. W. Seeb, 1986 Gene mapping of isozyme loci in chum salmon. *Journal of Heredity* 77: 399-402.
- Shumway, S. E., and G. J. Parsons (eds.), 2006 *Scallops: Biology, Ecology and Aquaculture*. Developments in Aquaculture and Fisheries Science. Amsterdam, Elsevier: 35: 1-594.
- Silva, C. N. S., and J. P. A. Gardner, 2015 Emerging patterns of genetic variation in the New Zealand endemic scallop *Pecten novaezelandiae* *Molecular Ecology* 24: 5379-5393.
- Sweijd N. A., R. C. K. Bowie, B. S. Evans, and A. L. Lopata, 2000 Molecular genetics and the management and conservation of marine organisms. *Hydrobiologia* 420: 153-164.
- Ventilla R. F., 1982 The scallop industry in Japan. *Advances in Marine Biology* 20: 309-382.

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

Von Brand E., G.E. Merino, A. Abarca, and W. Stotz, 2006 Scallop fishery and aquaculture in Chile. In: Shumway S. E. and G. J. Parsons (eds.), 2006 *Scallops: Biology, Ecology and Aquaculture. Developments in Aquaculture and Fisheries Science*. Amsterdam, Elsevier: 35: 1293-1314.

Waller, T. R., 1991 Evolutionary relationships among commercial scallops (Mollusca: Bivalvia: Pectinidae). In: Shumway, S. E., and G. J. Parsons (eds.), *Scallops: Biology, Ecology and Aquaculture. Developments in Aquaculture and Fisheries Science*. Amsterdam, Elsevier: 35: 1-73.

Wan, Q., H. Wu, T. Fujihara, and S. Fang, 2004 Which genetic marker for which conservation genetics issue? *Electrophoresis* 25: 2165-2176.

Wang, Y., D. Fu, and J. Xia, 2013 The genetic diversity of the noble scallop (*Chlamys nobilis*, Reeve 1852) in China assessed using five microsatellite markers. *Marine Genomics* 9: 63-67.

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

Wilding, C. S., A. R. Beaumont, and J. W. Latchford, 1999 Are *Pecten maximus* and *Pecten jacobaeus* different species? *Journal of the Marine Biological Association of the United Kingdom* 79: 949-952.

Yuan, T., M. He, and L. Huang, 2009 Intraspecific genetic variation in mitochondrial 16S rRNA and COI genes in domestic and wild populations of Huaguizhikong scallop *Chlamys nobilis* Reeve. *Aquaculture* 289: 19-25.

Yuan, Y., M. Kanno, and A. Kijima, 2012 Genetic diversity of wild populations of *Chlamys farreri* in Japan and their genetic relationship with cultured stocks in China. *Aquaculture* 370-371: 109-122.

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

Zhan, A., J. Hu, X. Hu, Z. Zhou, M. Hui, S. Wang, W. Peng, M. Wang, and Z. Bao, 2009 Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): Do local marine currents drive geographical differentiation? *Marine Biotechnology* 11: 223-235.

Chapter II

DNA Extraction in the South African Scallop, *Pecten sulcicostatus*: Evaluation of Optimal Tissue Type

1. Abstract

The quality of DNA obtained through DNA extraction depends both on the extraction method as well as the biological material used, and is vital for downstream applications which form an integral part of studies on population structure, gene flow and phylogeny of marine invertebrates. Destructive tissue sampling, such as the sampling of internal tissue types, is unsuitable for use in aquaculture, as individuals need to be sampled, but remain viable as broodstock. Non-destructive sampling generally involves the sampling of tentacles or the use of mucus as a sampling medium, and is necessary for studies on genetic fitness and population structure to not be hindered by requiring the death of the study individuals. DNA extraction methods using non-destructive sample tissue have not been developed for scallops, therefore this chapter compared the use of various tissue types (fresh tissue and dried tissue as well as tissue sampled by means of destructive tissue sampling and tissue sampled by means of non-destructive tissue sampling) in DNA extraction, allowing for the development of an effective non-destructive DNA extraction method for *Pecten sulcicostatus* using tentacle tissue and mucus swabs. Mucus swabbing of live individuals is a useful alternative to tentacle tissue sampling, as there is no risk to the study individual, making this method the least harmful and time-consuming way of collecting samples for DNA extraction. This study also allowed for the development of an effective DNA extraction method for use on dried or degraded tissue, which will, in turn, allow for the use of opportunistic and historic samples in future studies on *Pecten sulcicostatus* or similar bivalve species.

2. Introduction

DNA-based molecular markers provide a means to evaluate genome-wide genetic variation within and between individuals, populations and species (Hoffman and Nichols 2011). These molecular markers can be used in various applications in fisheries and aquaculture including individual identification, strain/species identification, parentage assignment and population studies. Studies using these genetic markers also aid in linkage mapping as well as the identification of genes involved in quantitative trait loci (QTL) which can be used to aid in the establishment of marker assisted selection (MAS) (Liu and Cordes 2004).

The first step in molecular genetic studies is DNA extraction. The quality of the DNA obtained is dependent both on the extraction method as well as the biological material used. DNA quality is important for downstream applications such as polymerase chain reaction (PCR) and DNA sequencing (Karlsson *et al.* 2013). These downstream applications are invaluable in studies on the evolution, systematics and population genetics of marine invertebrates (Dawson *et al.* 1998). The selection of a DNA extraction method is critical when downstream analyses depend on DNA quality. DNA extraction methods should be straightforward, quick, efficient and high throughput while minimising the potential for cross-contamination. DNA quality is of the utmost importance as DNA amplification is influenced by the presence of co-purifying inhibitors, which can reduce PCR efficiency (Berg *et al.* 1995). DNA damage may also occur during the extraction procedure due to oxidation and enzymatic hydrolysis problems associated with extraction buffer formulation and excessive mechanical shearing (Pereira *et al.* 2011). Unlike mammals, for which numerous DNA extraction methods have been developed, for molluscs there are few optimised DNA extraction protocols. There are several difficulties in DNA extraction from molluscs. For example, molluscs secrete mucopolysaccharides and polyphenolic proteins which co-purify with DNA, interfering with enzymatic processing of nucleic acids (Pereira *et al.* 2011).

Increasing importance has been placed on genetic information of aquaculture species, as this genetic information can assist with genetic maintenance and thus the sustainability of production (Astorga 2014). Sustainable aquaculture practices require information on the genetic structure of populations as well as the genetic

fitness of individuals. Traditionally, techniques used to collect tissue samples for genetic analyses and biochemical measures of fitness (glycogen and lipid concentrations, enzyme activity, *etc.*) required the death of the individuals. This prevented the use of such techniques for studies in which animals should be kept alive, such as those involving repeated sampling from study individuals or continued use as broodstock (Berg *et al.* 1995).

In scallops, internal tissue types, such as the adductor muscle, mantle musculature, gonad and gill, are representative of destructive tissue sampling and is therefore unsuitable for use in the aquaculture environment where individuals need to be sampled, but remain viable as broodstock. Previous studies on various scallop species made use of adductor muscle (Saavedra and Peña 2004; Saavedra and Peña 2005; Watts *et al.* 2005; Ibarra *et al.* 2006; Kenchington *et al.* 2006; Yuan *et al.* 2009; Zhan *et al.* 2009; Gaffney *et al.* 2010; Marín *et al.* 2013; Morvezen *et al.* 2013; Silva and Gardner 2015) and to a lesser extent, mantle musculature (An *et al.* 2009). No previous studies have made use of other internal tissue types, such as gonad and gill tissue, most likely due to the fact that both the adductor muscle and mantle musculature are easily distinguishable and easy to dissect in order to sample a piece of tissue.

Fresh scallop individuals are difficult to obtain due to low population densities as well as the depth at which these individuals can be found. Often scallop records along the South African coast are only due to shells washed up on beaches. The use of dried tissue in DNA extraction will not only allow for the use of tissue attached to shells that have washed up on beaches, but will also allow for the use of tissue attached to shells in museums, which may represent historic specimens. DNA has successfully been extracted from dry remains of soft tissues that are over 100 000 years old (Svante 1989). Although DNA obtained from ancient specimens, known as ancient DNA, may be damaged by oxidative processes, downstream applications can still be used to generate DNA sequences that are of anthropological and evolutionary significance. The use of ancient DNA opens up the possibility of performing diachronical studies of molecular evolutionary genetics and allows for the use of museum specimens as well as paleontological finds to address questions of historic, evolutionary and taxonomic significance (Svante 1989).

Non-destructive sampling has become fundamental for genetic monitoring and conservation of vulnerable species, providing a useful alternative to blood or tissue sampling (Taberlet *et al.* 1997; Schwartz *et al.* 2007). Similarly, non-destructive tissue sampling has become increasingly important in aquaculture, as non-destructive tissue sampling is needed in order for studies of genetic structure and fitness to not be hindered by requiring the death of the study individuals (Berg *et al.* 1995). Non-destructive sampling in bivalves generally involves the sampling of tentacles or the use of mucus as a sampling medium (Berg *et al.* 1995) and has been used in studies on other bivalves such as abalone (Slabbert and Roodt-Wilding 2006) and freshwater mussels (Karlsson *et al.* 2013). DNA extraction methods using non-destructive sample tissue have not been developed for scallops, therefore the aim of this chapter is to compare the use of various tissue types (fresh tissue and dried tissue as well as tissue sampled by means of destructive tissue sampling and tissue sampled by means of non-destructive tissue sampling) in DNA extraction in order to develop optimal tissue sampling methods for *Pecten sulcicostatus*.

3. Materials and Methods

3.1 Biological Specimens

Two fresh, whole *P. sulcicostatus* individuals as well as tentacle tissue from an additional two *P. sulcicostatus* individuals were collected from the Department of Agriculture, Forestry and Fisheries (DAFF). The *P. sulcicostatus* individuals were originally sampled from False Bay in 2013 and transferred to the DAFF Sea Point Research Aquarium for use in studies on reproduction and feeding. Dried *P. sulcicostatus* mantle musculature tissue was also obtained from the KwaZulu-Natal Museum. The dried museum sample was originally collected in False Bay, off Buffels Bay in 1991.

3.2 Tissue Collection and DNA Extraction

The valves of the two fresh, whole *P. sulcicostatus* individuals were opened to expose the internal organs. The mantle musculature was moved to clearly expose the internal organs. Duplicate tissue samples were collected from the following tissue types: tentacle, adductor muscle, gills, gonad and mantle musculature (Figure 2.1).

Duplicate swabbing of the mucus using sterile cotton balls was also performed on both *P. sulcicostatus* individuals. Tissue as well as mucus swab cotton balls were stored in 99.9% ethanol until DNA extractions were performed. Internal tissue types (adductor muscle, gills, gonad tissue and mantle musculature) represent destructive tissue samples, whilst tentacle tissue and mucus swabs represent non-destructive samples.

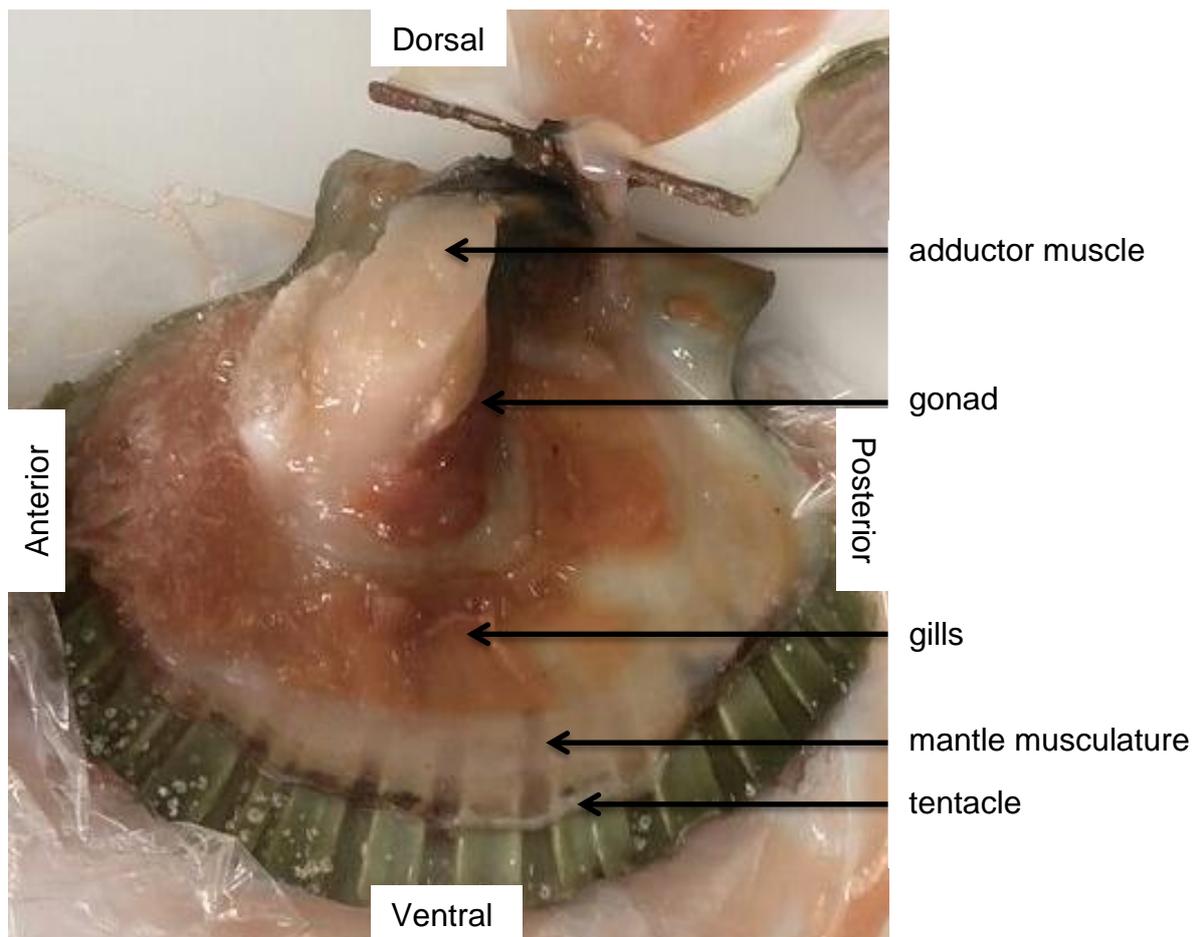


Figure 2.1: Various tissue types used in tissue sampling of *Pecten sulcicostatus*. Internal tissue types (adductor muscle, gills, gonad tissue and mantle musculature) represent destructive tissue samples, whilst tentacle tissue represent non-destructive tissue samples.

Tentacle tissue was also sampled from two live *P. sulcicostatus* individuals by cutting a single tentacle using sterile dissection scissors and storing the tissue in 99.9% ethanol until DNA extractions were performed. The individuals were monitored for

two months after tissue collection in order to ensure that the tentacle sampling had no negative effect on the individuals.

Total genomic DNA (gDNA) was extracted from 10 mg of each fresh tissue type as well as the entire mucus swab cotton balls from both individuals using a modified CTAB extraction method (Saghai-Marooif *et al.* 1984). Genomic DNA (gDNA) was also extracted from 10 mg of tentacle samples from the additional two individuals as well as the dried mantle musculature tissue using the same method. Modification of the CTAB extraction method described by Saghai-Marooif *et al.* (1984) included the doubling of the concentration of the buffer reagents as well as the extension of the incubation time from 60 minutes to an overnight incubation. Instead of chloroform:octanol (24:1 (v:v)), chloroform:isoamylalcohol (24:1 (v:v)) was used to separate DNA from the cell organelles. Two volumes of 100% ethanol were added to precipitate DNA and the precipitated DNA was rehydrated using 20 μ l MilliQ water instead of 1.5 mL NH_4OAc and 0.25 mM EDTA.

Following DNA extraction, DNA quantification (ng/ μ l) and quality (absorbance ratio: 260/280 and 260/230) was determined using a Nanodrop spectrophotometer. Nucleic acids and proteins have absorbance maxima at 260 and 280 nm, respectively. The ratio of absorbance at these wavelengths is a measure of the purity of the DNA solution after extractions. A 260/280 ratio of approximately 1.8 is accepted as pure for DNA. Similarly, absorbance at 230 nm is as a result of other contamination, therefore the 260/230 ratio is often also calculated. The 260/230 ratio values for pure DNA are often higher than the respective 260/280 ratio values with expected 260/230 ratio values ranging between 2.0 and 2.2 (Pereira *et al.* 2011).

3.3 Testing of Genomic DNA Amplification

Amplification of the extracted gDNA was tested *via* PCR using the 16S rRNA gene (Table 2.1). Extracted gDNA was standardised to 20 ng/ μ l and used as a template to amplify the target 16S rRNA gene fragment *via* PCR. Individual reactions were carried out in 10 μ l total volume, containing 20 ng of template DNA, 2.0 mM of MgCl_2 , 200 μ M of deoxynucleotide triphosphates (dNTPs), 0.5 U of *Taq* DNA polymerase and 0.2 μ M of each primer. Amplification was carried out in the presence of negative (no template) controls.

Table 2.1: 16S mitochondrial primer pair used to test gDNA amplification.

Primer name	Region	Primer sequence (5' – 3')	Reference
Pec16S-F1	16S	GTTTTAAGGTCGGGGAAAG	Saavedra and Peña 2005
16Sbr		CCGGTCTGAACTCAGATCACGT	

The PCR cycling conditions for the 16S primer pair included an initial activation step at 94°C for 4 minutes followed by 35 cycles of a denaturing step at 95°C for 1 minute, an annealing step at 55°C for 30 seconds and an extension step at 72°C for 30 seconds. A final extension was conducted at 72°C for 10 minutes.

PCR products were separated by electrophoresis in an ethidium bromide (EtBr) stained 2% (w/v) agarose gel and viewed under shortwave ultraviolet (UV) in order to verify amplification of a DNA fragment of the expected size (650 bp), as determined by comparison to a 1 kb HyperLadder size standard (Bioline).

4. Results and Discussion

The non-destructive tissue sampling of a single tentacle was successful as both *P. sulcicostatus* individuals survived the removal of these tentacles. All fresh tissue types, dried mantle musculature as well as mucus swabs were successful sources of DNA in DNA extraction. Of the fresh tissue types, gonad tissue yielded the highest concentration (1482.54 ng/μl), good quality (260/280: 1.98; 260/230: 2.16) DNA (Figure 2.2), whereas adductor muscle yielded the lowest concentration (118.40 ng/μl), average quality (260/280: 1.91 260/230: 0.96) DNA (Figure 2.2). The fact that adductor muscle tissue yielded the lowest DNA concentration and quality is surprising, as the majority of studies on scallops have made use of the adductor muscle (Saavedra and Peña 2004; Saavedra and Peña 2005; Watts *et al.* 2005; Ibarra *et al.* 2006; Kenchington *et al.* 2006; Yuan *et al.* 2009; Zhan *et al.* 2009; Gaffney *et al.* 2010; Marín *et al.* 2013; Morvezen *et al.* 2013; Silva and Gardner 2015). Although gonad tissue yielded the highest concentration (1482.54 ng/μl), good quality (260/280: 1.98; 260/230: 2.16) DNA, it represents destructive sampling

and is therefore only an option in recently deceased individuals or individuals that are not needed for future studies or use as broodstock. Tentacle tissue yielded an average DNA concentration of 217.33 ng/ μ l with a 260/280 ratio of 2.00 and a 260/230 ratio of 1.55 (Figure 2.2), making it suitable for use in DNA extraction without harming the scallop. The other form of non-destructive sampling, namely mucus swabbing, yielded low concentration (10.91 ng/ μ l), poor quality (260/280: 1.70; 260/230: 1.02) DNA (Figure 2.2), making it less suitable for DNA extraction. The very low 260/230 ratio obtained from the mucus swabs (260/230: 1.02) indicates co-isolation of impurities such as proteins, which may inhibit downstream applications such as PCR (Wilfinger *et al.* 1997).

The dried mantle musculature tissue yielded a very high concentration of DNA (2853.20 ng/ μ l), although the DNA was of poor quality (260/280: 1.58; 260/230: 0.71). The poor DNA quality indicates contamination with proteins and other contaminants, most likely due to degradation of the tissue before drying. This contamination may inhibit downstream analyses such as PCR, resulting in failed amplification (Pereira *et al.* 2011).

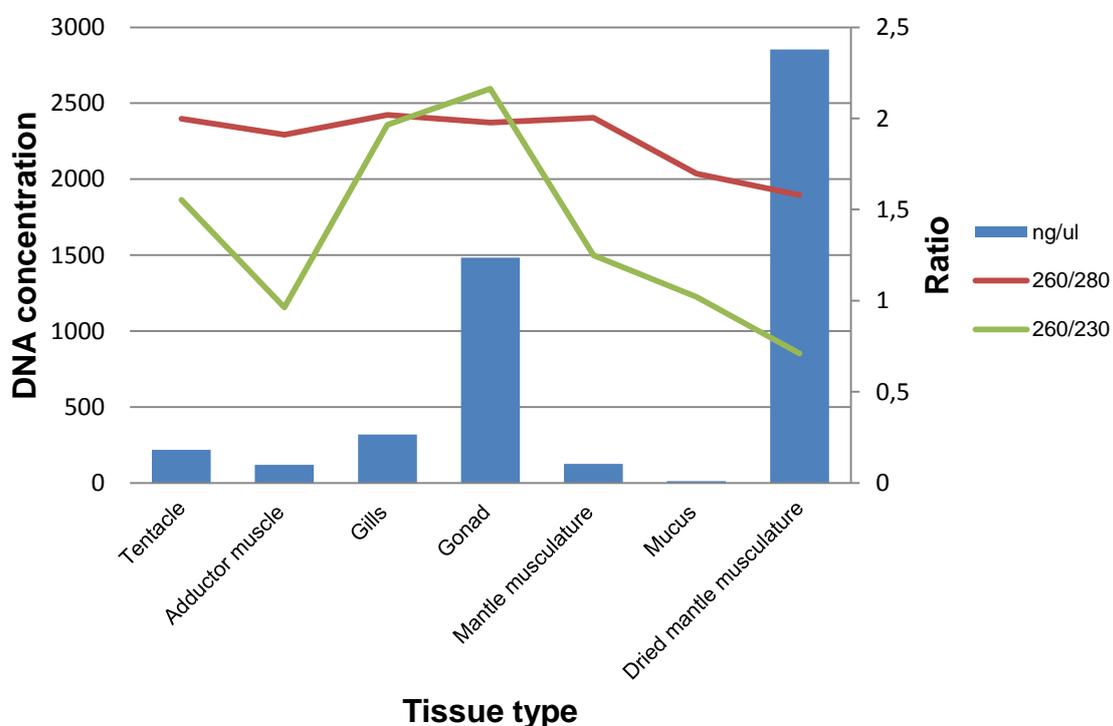


Figure 2.2: DNA quantification (ng/ μ l) and quality (absorbance ratio: 260/280 and 260/230) for gDNA extracted from various tissue types in *Pecten sulcicostatus*.

Although DNA quantification (ng/ μ l) and quality (absorbance ratio: 260/280 and 260/230) is a good measure of the usefulness of DNA for downstream analyses, the only way to confirm successful DNA extraction is through the amplification of the extracted gDNA *via* PCR. Genomic DNA (gDNA) from all tissue samples, both internal (destructive tissue sampling) and external (non-destructive tissue sampling), amplified an approximately 650 bp fragment of the 16S rRNA gene with equal efficacy (Figure 2.3), making all tissue types suitable for use in DNA extraction in *P. sulcicostatus*.

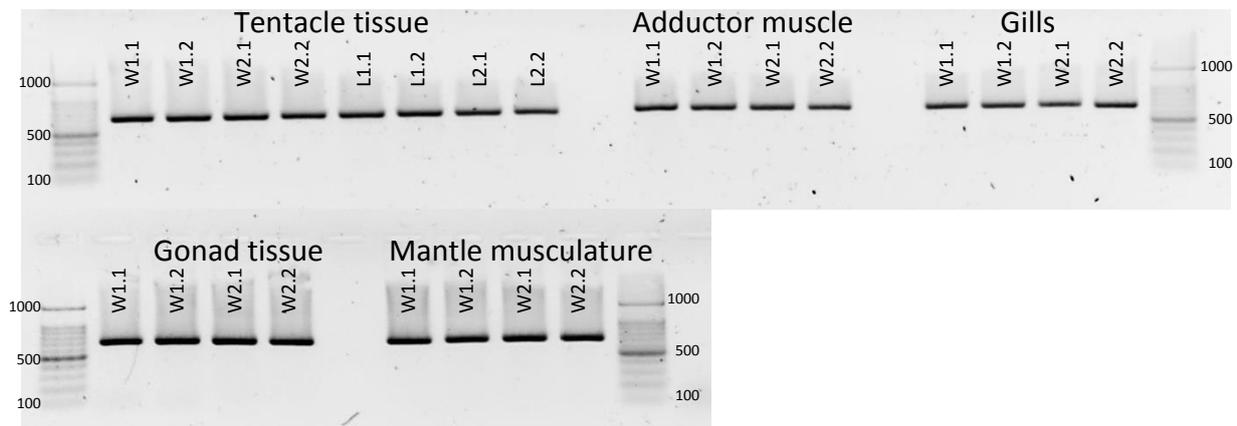


Figure 2.3: 2% agarose gel image of PCR products of 16S rRNA gene using gDNA extracted from various tissue types where “W” indicates tissue sampled from whole, dead individuals while “L” indicates tentacle tissue sampled from live individuals.

Although the non-destructive tissue sampling of a single tentacle was successful, resulting in the survival of both *P. sulcicostatus* individuals, yielding good quality DNA as well as amplifying a fragment of the 16S rRNA gene, tentacle sampling causes disturbance and stress of the study individual and may therefore affect breeding behaviour and reproductive success (Pidancier *et al.* 2003). It is therefore necessary to investigate alternative non-destructive sampling methods such as mucus swabbing.

Despite the fact that the gDNA extracted from the mucus swabs had a very low concentration (10.91 ng/ μ l) and was of poor quality (260/280: 1.70; 260/230: 1.02), the gDNA successfully amplified an approximately 650 bp fragment of the 16S rRNA gene (Figure 2.4). Mucus swabbing of live individuals is therefore a useful alternative

to tentacle tissue sampling, as there is no risk to the study individual. Mucus swabbing is the least harmful and time-consuming way of collecting samples for DNA extraction, allowing for the swabbing of a large number of individuals in a short period of time (Pidancier *et al.* 2003).

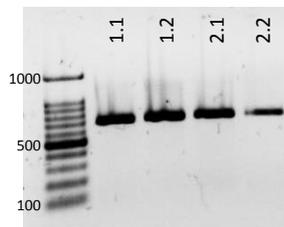


Figure 2.4: 2% agarose gel image of PCR products of a fragment of the 16S rRNA gene using gDNA extracted from mucus swabs.

The dried mantle musculature tissue also successfully amplified an approximately 650 bp fragment of the 16S rRNA gene (Figure 2.5), although the band produced was fainter than that produced using gDNA extracted from fresh tissue, most likely due to protein contaminants which inhibit PCR, resulting in suboptimal amplification (Pereira *et al.* 2011). The fact that DNA extraction was successful using tissue that has been dried for over 20 years is of great significance for future studies on this endemic species, as this DNA extraction method can be used on tissue sampled from shells found washed up on beaches, as this tissue may be decayed or dried. It is also possible for this DNA extraction method to be used on museum samples as well as paleontological finds as historic samples. The use of historic samples can aid in addressing questions of historic, evolutionary and taxonomic significance as well as aid in diachronical studies of molecular evolutionary genetics (Svante 1989).

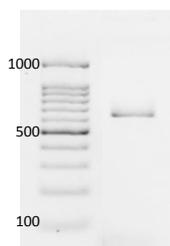


Figure 2.5: 2% agarose gel image of PCR products of a fragment of the 16S rRNA gene using gDNA extracted from dried mantle musculature tissue.

The testing of various tissue types for DNA extraction allowed for the development of an effective non-destructive DNA extraction method using tentacle tissue and mucus swabs, thus ensuring the survival of sampled individuals. Similar non-destructive sampling methods have been developed for other invertebrates such as abalone (Slabbert and Roodt-Wilding 2006) and freshwater mussels (Karlsson *et al.* 2013). Conservation requires information on the genetic structure of populations and fitness of individuals (Sandlund *et al.* 1992), however traditional tissue sampling methods require the death of study individuals. Such techniques can therefore not be used on individuals that must be kept alive, such as those involving threatened and endangered species or otherwise sensitive populations (Berg 1995). The non-destructive sampling method developed in this study can therefore be applied to similar bivalve species where individuals must be kept alive for conservation purposes.

Similarly, in order to formulate effective strategies for genetic management and sustainable utilisation of aquaculture species, information on the genetic structure of populations is necessary (Pereira *et al.* 2011). Non-destructive sampling methods are vital in aquaculture as they allow for repeated sampling from study individuals. In light of the fact that *P. sulcicostatus* has been identified as having potential commercial value as an aquaculture species, the non-destructive sampling method/s developed in this study can be used on broodstock without causing stress to the study individual (Pidancier *et al.* 2003). The development of a non-destructive sampling method for *P. sulcicostatus* will allow individuals in future breeding programmes to be used for genetic studies while remaining viable (Slabbert and Roodt-Wilding 2006; Nowland *et al.* 2015). Further, this non-destructive sampling method could be used to identify genetically superior *P. sulcicostatus* individuals at an early age, after the development of technologies to enable marker assisted selection (MAS) (Slabbert and Roodt-Wilding 2006).

5. Conclusion

This study compared the use of different tissue types (fresh tissue and dried tissue as well as tissue sampled by means of destructive and non-destructive tissue sampling) in DNA extraction. This study allowed for the development of an effective

non-destructive DNA extraction method for *P. sulcicostatus* using tentacle tissue and mucus swabs, ensuring the survival of sampled individuals. In future this non-destructive sampling method will allow individuals in future breeding programmes to be used for genetic studies while remaining viable. This non-destructive sampling method can also be applied to similar bivalve species where individuals must be kept alive for aquaculture as well as conservation purposes. This study also allowed for the development of an effective DNA extraction method for use on dried or degraded tissue such as the tissue attached to shells washed up on beaches or shells in museum collections, which will, in turn, allow for the use of opportunistic and historic samples in future studies on *P. sulcicostatus* or similar species.

6. References

- An, H. S., Y. G. Lee, J. Y. Park, and C. Lee, 2009 Genetic characterization of four East Asian giant scallop (*Mizuhopecten yessoensis*) populations using microsatellite markers. *Aquaculture Resources* 40: 619-624.
- Astorga, M. P., 2014 Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Frontiers in Genetics* 5: 1-6.
- Berg, D. J., W. R. Haag, S. I. Guttman, and J. B. Sickel, 1995 Mantle biopsy: A technique for nondestructive tissue-sampling of freshwater mussels. *Journal of the North American Benthological Society* 14: 577-581.
- Dawson, M. N., K. A. Raskoff, and D. K. Jacobs, 1998 Field preservation of marine invertebrate tissue for DNA analyses. *Molecular Marine Biology and Biotechnology* 7: 145-152.
- Gaffney, P. M., C. M. Pascal, J. Barnhart, W. S. Grant, and J. E. Seeb, 2010 Genetic homogeneity of weathervane scallops (*Patinopecten caurinus*) in the northeastern Pacific. *Canadian Journal of Fisheries and Aquatic Sciences* 67: 1827-1839.
- Hoffman, J. I., and H. J. Nichols, 2011 A novel approach for mining polymorphic microsatellite markers *in silico*. *PLoS ONE* 6: 1-9.
- Ibarra, A. M., J. L. Petersen, T. R. Famula, and B. May, 2006 Characterization of 35 microsatellite loci in the Pacific lion-paw scallop (*Nodipecten subnodosus*) and their cross-species amplification in four other scallops of the Pectinidae family. *Molecular Ecology Notes* 6: 153-156.
- Karlsson, S., B. M. Larsen, L. Eriksen, and M. Hagen, 2013 Four methods of nondestructive DNA sampling from freshwater pearl mussels *Margaritifera margaritifera* L. (Bivalvia:Unionoida). *Freshwater Science* 32: 525-530.
- Kenchington, E. L., M. U. Patwary, E. Zouros, and C. J. Bird, 2006 Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology* 15: 1781-1796.
- Liu, Z. J., and J. F. Cordes, 2004 DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238: 1-37.

Marín, A., T. Fujimoto, and K. Arai, 2013 Genetic structure of the Peruvian scallop *Argopecten purpuratus* inferred from mitochondrial and nuclear DNA variation. *Marine Genomics* 9: 1- 8.

Morvezen, R., F. Cornette, G. Charrier, B. Guinand, S. Lapègue, B. Pierre, and L. Jean, 2013 Multiplex PCR sets of novel microsatellite loci for the great scallop *Pecten maximus* and their application in parentage assignment. *Aquatic Living Resources* 26: 207-213.

Nowland, S. J., D. R. Jerry, and P. C. Southgate, 2015 A non-destructive tissue sampling technique for holothurians to facilitate extraction of DNA for genetic analysis. *Invertebrate Biology* 134: 252-259.

Pereira, J. C., R. Chaves, E. Bastos, A. Leitão, and H. Guedes-Pinto, 2011 An efficient method for genomic DNA extraction from different molluscs species. *International Journal of Molecular Sciences* 12: 8086-8095.

Pidancier, N., C. Miquel, and C. Miaud, 2003 Buccal swabs as a non-destructive tissue sampling method for DNA analysis in amphibians. *Herpetological Journal* 13: 175-178.

Saavedra, C., and J. B. Peña, 2004 Phylogenetic relationships of commercial European and Australasian king scallops (*Pecten* spp.) based on partial 16S ribosomal RNA gene sequences. *Aquaculture* 235: 153-166.

Saavedra, C., and J. B. Peña, 2005 Nucleotide diversity and Pleistocene population expansion in Atlantic and Mediterranean scallops (*Pecten maximus* and *P. jacobaeus*) as revealed by the mitochondrial 16S ribosomal RNA gene. *Journal of Experimental Marine Biology and Ecology* 323: 138-150.

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

Sandlund, O. T., K. Hindar, and A. H. D. Brown (eds.), 1992 Conservation of biodiversity for sustainable development. Scandinavian University Press, Oslo, Norway. Pages 1-324.

Schwartz M. K., G. Luikart, and R. S. Waples, 2007 Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution* 22: 25-33.

Silva, C. N. S., and J. P. A. Gardner, 2015 Emerging patterns of genetic variation in the New Zealand endemic scallop *Pecten novaezelandiae* *Molecular Ecology* 24: 5379-5393.

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

Svante, P., 1989 Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences USA* 86: 1939-1943.

Taberlet P., J. J. Camarra, S. Griffin, E. Uhrès, O. Hanotte, L. P. Waits, C. Dubois-Paganon, T. Burke, and J. Bouvet, 1997 Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* 6: 869-876.

Watts, P. C., P. J. Mallanaphy, C. McCarthy, B. D. Beukers-Stewart, M. W. J. Mosley, A. R. Brand, and I. J. Saccheri, 2005 Polymorphic microsatellite loci isolated from the great scallop, *Pecten maximus* (Bivalvia: Pectinidae). *Molecular Ecology Notes* 5: 902-904.

Wilfinger, W. W., K. Mackey, and P. Chomczynski, 1997 Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques* 22: 474-481.

Yuan, T., M. He, and L. Huang, 2009 Intraspecific genetic variation in mitochondrial 16S rRNA and COI genes in domestic and wild populations of Huaguizhikong scallop *Chlamys nobilis* Reeve. *Aquaculture* 289: 19-25.

Zhan, A., J. Hu, X. Hu, Z. Zhou, M. Hui, S. Wang, W. Peng, M. Wang, and Z. Bao, 2009 Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): Do local marine currents drive geographical differentiation? *Marine Biotechnology* 11: 223-235.

Chapter III

Cross-species Transfer of Microsatellite Markers and Optimisation of Mitochondrial Markers for Applications in *Pecten sulcicostatus*

1. Abstract

Although *Pecten sulcicostatus* has been identified as having potential commercial value as an aquaculture species, no molecular markers are available for this species. No molecular studies have been conducted on the genetic composition of this species or the genetic constitution of these molluscs along the South African coast, therefore genetic resources are vital for use in future studies addressing these gaps. This chapter therefore attempted to develop a set of microsatellite markers for the South African scallop, *Pecten sulcicostatus*, as well as to optimise a mitochondrial marker, which was applied to shed light on the evolutionary relationship of various *Pecten* species. This study allowed for the development of a set of 10 microsatellite markers for *Pecten sulcicostatus*, using cross-species transfer from *Pecten maximus*, a sister species to *Pecten sulcicostatus*, with a success rate of 50%. The set of microsatellite markers was successfully applied to generate genetic diversity data, which, in future studies, could be used to evaluate the extent of intra- and interpopulation genetic partitioning and variation. This study also allowed for the optimisation a fragment of the 16S rRNA gene, which was successfully used to determine intra- and interspecific genetic diversity. The genetic resources developed in this study will be used in future studies to assess the genetic constitution of this species, which is important for formulating effective strategies for genetic management and sustainable utilisation of this species. In future the microsatellite markers developed for *Pecten sulcicostatus* could also be continuously used to monitor genetic diversity as the species is subjected to aquaculture.

2. Introduction

Microsatellite markers, which are based on the detection of a particular sequence variation, are currently one of the most popular molecular markers (Renon *et al.* 2001; Hoffman and Nicols 2011). Microsatellite markers are classed as variable number of tandem repeats (VNTRs) which are co-dominantly inherited, comprise of multiple copies of tandemly organised simple sequence repeats (SSRs) and are uniformly distributed throughout the genome at densities proportional to genome size. These markers can be found in introns, gene coding regions and non-gene sequences (Hoffman and Nichols 2011). Microsatellite markers have widespread use in studying the genetic differentiation within and among populations and have been used in the characterisation of genetic stocks, parentage analysis, broodstock selection, linkage mapping and studies of quantitative trait loci (QTL) (Chistiakov *et al.* 2006). These markers have also been used in aquaculture to monitor genetic diversity (Guo *et al.* 2008; Kohlmann *et al.* 2010).

De novo microsatellite marker development can be both time-consuming and expensive, therefore once a set of microsatellite markers has been developed for a focal species, time and effort can be saved on marker development in closely related species by using cross-species transfer of microsatellite markers. Closely related taxa have a high success rate with microsatellite marker cross-species amplification (Zane *et al.* 2002), however, this success rate decreases with genetic distance between species (Arif *et al.* 2010). As the genetic distance between species increases, primers fail to bind to the target region due to the large number of mutations that distinguish species separated by a large genetic distance. Cross-species transfer of microsatellite markers has a success rate of approximately $72 \pm 7\%$ in invertebrates (Zane *et al.* 2002). Even though cross-species amplification may be successful, high levels of polymorphism cannot be guaranteed (Chambers and MacAvoy 2002; Zane *et al.* 2002). Furthermore, genetic diversity may be underestimated by the use of non-species specific microsatellite markers (Arif *et al.* 2010).

Despite abundant fossil records extending from the early Triassic, the evolutionary relationship among members of the Pectinidae family is poorly understood. Scallop shell morphological features are useful for classification at the species level, but

provide limited phylogenetic information (Barucca *et al.* 2004). With current morphology-based systems, the attribution of species at both the genus and the subfamily level is far from uniform, as different authors give different degrees of importance to morphological features. Scallop shell characteristics show great variability within a single species, while the shells of different species can display very similar characteristics.

In order to obtain further insights into the complex taxonomy and evolutionary relationships among *Pecten* species, more studies on the mitochondrial genomes of these scallop species are needed. The mitochondrial genome is a double stranded, circular DNA molecule which encodes for 37 genes - 12 protein coding genes, two ribosomal RNA genes and 22 transfer RNAs. Almost the entire mitochondrial DNA (mtDNA) molecule is transcribed except for the 1 kb control region (D-loop), where replication and transcription is initiated. Mitochondrial genes are important for evolutionary and phylogenetic studies due to their high mutation rates and maternal inheritance. Studies have shown that sequence divergence accumulates more rapidly in mtDNA than in nuclear DNA due to a faster mutation rate in mtDNA. The faster mutation rate can be attributed to a lack of repair mechanisms during replication as well as smaller effective population size due to maternal inheritance of the haploid mitochondrial genome. The relatively fast mutation rate of mtDNA results in the generation of genetic diversity within and between populations over a relatively short time (thousands of generations) (Liu and Cordes 2004). It was previously believed that the 16S rRNA gene is one of the slowest evolving mitochondrial genes, however recent studies show that this gene contains both rapidly and slowly evolving regions. The rapidly evolving regions are useful for determining closely related species, whereas the slowly evolving regions are useful for determining distant relationships. The 16S rRNA gene has been extensively used in phylogenetic studies (Barucca *et al.* 2004; Saavedra and Peña 2004; Saavedra and Peña 2005; Marín *et al.* 2013; Marín *et al.* 2015).

This chapter will therefore attempt to develop a set of microsatellite markers for the South African scallop, *Pecten sulcicostatus*, using cross-species amplification from *Pecten maximus*, a sister species to *P. sulcicostatus*. In order to validate the set of microsatellite markers and prove their usefulness, the set of microsatellite markers

was used to quantify the intrapopulation genetic diversity of a natural *P. sulcicostatus* population. This chapter will also attempt to optimise a mitochondrial marker which will be applied to shed light on the evolutionary relationship of various *Pecten* species. In future the microsatellite markers developed for *P. sulcicostatus* could be continuously used to monitor genetic diversity as the species is subject to aquaculture.

3. Materials and Methods

3.1 Biological Specimens and DNA Extraction

Four *P. sulcicostatus* individuals were sampled from False Bay in 2013 and transferred live to the DAFF Sea Point Research Aquarium for use in studies on reproduction and feeding. During 2014 an additional 33 *P. sulcicostatus* individuals were sampled from False Bay. A historic *P. sulcicostatus* individual, sampled from False Bay, off Buffels Bay in 1991, was obtained from the Natal Museum. False Bay represents the western geographical distribution extreme, west of the Agulhas upwelling, a major bio-geographical barrier to many marine species around the South African coast. Finally a total of 39 individuals representing various *Pecten* species (Table 3.1) were obtained from various institutions around the world.

DNA extraction was performed from tentacle tissue samples from the 38 *P. sulcicostatus* individuals, dried mantle musculature tissue from the historic *P. sulcicostatus* individual and adductor muscle tissue from the 39 individuals representing various *Pecten* species using a modified CTAB extraction method (Saghai-Marooif *et al.* 1984). For the initial testing of microsatellite marker transferability as well as mitochondrial gene fragment amplification, the four *P. sulcicostatus* individuals sampled in 2013 were used. The remaining 33 *P. sulcicostatus* individuals were included for final microsatellite marker genotyping. All 38 *P. sulcicostatus* individuals as well as the 39 individuals representing various *Pecten* species were included for final mitochondrial gene fragment amplification.

Table 3.1: Sample details for various *Pecten* species included in this study.

Sample provider	Species	ID	Locality	
University of Padova, Italy	<i>P. jacobeus</i>	1_a	Caorle, Italy	
		2_a		
		3_a		
		4_a		
			1_b	Chioggia, Italy
			2_b	
			3_b	
			4_b	
		<i>P. maximus</i>	Manto_1	Galicia, Spain
			Manto_2	
			Manto_3	
			PM_MN 48	
PM_MN 50				
Victoria University of Wellington, New Zealand	<i>P. novaezelandiae</i>	PM_MN 53	Norway	
		PM_MN 55		
		MER 1		Mercury Islands, northern New Zealand
		MER 2		
		MER 3		
		TAS 1	Tasman Bay, central New Zealand	
		TAS 2		
		TAS 3		
		FIO 1	Fiordland, southern New Zealand	
		FIO 2		
FIO 3				
CHA 2	Chatham Islands, eastern Archipelago			
CHA 3				
CHA 4				
Florida Museum of Natural History, USA	<i>P. fumatus</i>	296997	Shark Bay, Australia	
	<i>Euvola raveneli</i> *	327784	Florida Straits	
	<i>Euvola perulus</i> *	351616	Gulf of Panama,	

		371284	Panama
		PeFu 1	
University of Queensland, Australia	<i>P. fumatus</i>	PeFu 2 PeFu 3 PeFu 4 PeFu 5	Eddystone, eastern coast of Tasmania
Roma Tre University, Italy	<i>P. keppelianus</i>	Pke 1 Pke 2 Pke 3	Atlantic Ocean, Senegal

***Outgroup species.**

3.2 Cross-species Amplification of Microsatellite Markers

A total of 20 microsatellite markers previously developed for *P. maximus* (Watts *et al.* 2005; Charrier *et al.* 2012; Hold *et al.* 2013; Morvezen *et al.* 2013) (Table 3.2) were selected to test for transferability to the target species, *P. sulcicostatus*, via PCR cross-amplification in a GeneAmp PCR System 2700 (Applied Biosystems). These microsatellite markers were selected as most of them displayed perfect repeats and have approximately the same annealing temperature (T_a).

Table 3.2: Microsatellite marker primers developed for *Pecten maximus* grouped according to annealing temperature (T_a).

Locus	Primers	Primer sequence (5' – 3')	T _a (°C)	Motif	Reference
PmNH11	PmNH11F	GCCATGGTCGGAATCACC	51	(ATTT) ₄	Hold <i>et al.</i> 2013
	PmNH11R	CAAACGCGCCAAGTCTACG			
PmGC01	PmGC01Fb	CAAAAATGCAAAGCAGAAACAAC		(ATG) ₈	
	PmGC01R	ATATCATAGAAGAGCTTTATTAAGG			
PmGC05	PmGC05Fb	AATTGTACTTTCAATCATAAACTGAG		(TA) ₈	Charrier <i>et al.</i> 2012
	PmGC05Rb	ACAGTAATCTAGGAAACACAATG			
PmGC10	PmGC10Fb	TGTGACACTATATGACGTCAC		(AT) ₇	
	PmGC10Rb	AACCTTCTTTCATCAGTGTAAC			
PmGC18	PmGC18F	TCCACTCCTTTTCTGTGACG	55	(AT) ₃ T(AT) ₃	
	PmGC18Rb	TGAGAAATATATGAGTTGACTATAG			
	LIST15-009	TGGATAAGACAACAGGCTGAGC		(TACAC) ₃₄	Watts <i>et al.</i> 2005
	BV676522	GGTCTGTACATGCATTATTGGC			
PmRM014	F	CGCCATTCAACAGACAAAT		(TGT) ₁₄	Morvezen <i>et al.</i> 2013
	R	CAGCACATCTGTCACGACCT			
PmRM041	F	CAGAATTGCCGATGGATGT		(CAA) ₁₀	
	R	GACACCAAAGCGTTACCACA			

PmGC03	PmGC03Fb	TGTAAGACTTTCCCTTCACTC		(AT) ₇	Charrier <i>et al.</i> 2012
	PmGC03Rb	GTTCACATAGTCCTCAACCTC			
PmGC06	PmGC06F	ACAAAGGTTTCTTGGCAGAAG	56	(TG) ₆	
	PmGC06Rb	AAGTCAGGATGTATCTACACC			
PmGC20	PmGC20Fb	GTAATGTTTTTTACTTGTCTCTTG		(AT) ₄ T(AT)	
	PmGC20Rb	GGTATACATGTGTAAAGCGTG			
PmNH09	PmNH09F	CTGATTTACCCTCCGCAACG		(CTGT) ₄	
	PmNH09R	TCAGTAATTCTACTCGTGACCC			
PmNH62	PmNH62F	GGGACCACTGTAAACAATGTG	58	(AAC) ₉	Hold <i>et al.</i> 2013
	PmNH62R	GCGTGACAGTCGACCATTTC			
PmNH70	PmNH70F	AGTTGTGCTATTGAATGGGAAC		(AAT) ₈	
	PmNH70R	ATGCACTGCTTGTCCACTTC			
PmNH73	PmNH73F	CATAGCGATGCAGGACAAGG		(GGT) ₈	
	PmNH73R	ATTCCAATGTCTGCCGTCTG			
PmRM002	F	GTGACAATGTGTCCACCTGC		(ATGT) ₇	
	R	CGTCGAGGGAAAAGTGAAGT			
PmRM007	F	TGTCTGCAGGTCAGGAGAT	59	(AGG) ₁₀	Morvezen <i>et al.</i> 2013
	R	TTACCTCCCCTTCCACATTG			
PmRM016	F	TGTGATCGGACGAGTTTAC		(ACG) ₈	
	R	ATGCAATGAGCCTCAGACCT			
PmRM027	F	GAGGTAAAGTTGATGTGATGAAGAA		(AGG) ₁₁	

PmRM028	R	ATCCCACCGTCCTGTCATAC	
	F	ATCCCTTTCCAAGATTTTCG	
	R	CCTTGTTGTGTAACGATGGA	(ATCC) ₇

Optimisation of microsatellite marker amplification was performed using the selected four *P. sulcicostatus* individuals. Individual reactions were carried out in 10 µl total volume, containing 20 ng of template DNA, 2.0 mM of MgCl₂, 200 µM of deoxynucleotide triphosphates (dNTPs), 0.5 U of *Taq* DNA polymerase and 0.2 µM of each primer. Amplification of the microsatellite markers was carried out in the presence of negative (no template) controls. The PCR cycling conditions included an initial activation step at 94°C for 5 minutes followed by 30 cycles of a denaturing step at 94°C for 45 seconds, an annealing step at T_a°C for 45 seconds and an extension step at 72°C for 45 seconds. A final extension was conducted at 72°C for 10 minutes. Subsequent to the completion of individual PCR reactions, the amplicons were separated by electrophoresis in an ethidium bromide (EtBr) stained 2% (w/v) agarose gel and viewed under shortwave ultraviolet (UV) in order to verify amplification of a DNA fragment of the expected size, as determined by comparison to a 1 kb HyperLadder size standard (Bioline). Amplicons of successfully amplifying microsatellite markers were electrophoresed in a 12% polyacrylamide gel and viewed under shortwave UV in order to detect polymorphism for each microsatellite marker. A microsatellite marker was deemed polymorphic when two bands were distinguishable in a single individual and/or there were clear size differences between bands of different individuals.

3.2.1 Microsatellite Marker Genotyping

Successfully amplifying, polymorphic microsatellite markers were fluorescently labelled with FAM, VIC, PET or NED dyes (Applied Biosystems) and optimised into multiplex genotyping systems in order to genotype the individuals for each of the microsatellite markers. Amplification of the microsatellite marker multiplexes was performed using a KAPA2G Fast Multiplex kit according to manufacturer's instructions. Fragment analysis was conducted via capillary electrophoresis on an ABI 3730XL DNA Analyzer (Applied Biosystems) at Stellenbosch University Central Analytical Facilities (CAF) with the GeneScan-600 (LIZ) internal size standard (Applied Biosystems). Peaks were scored based on fragment sizes using Genemapper software version 4.1 (Applied Biosystems).

3.2.2 Data Analysis

Deviation from Hardy-Weinberg (HW) equilibrium was tested in Genepop version 4.0 (Rousset 2008) using the Markov chain method and Fisher's exact test. Genepop (Rousset 2008) was also used to test for the presence of null alleles using default parameters. For marker characterisation, GenAlEx version 6.501 (Peakall and Smouse 2006, 2012) was used to calculate number of alleles (A_n), effective number of alleles (A_e), information (Shannon-Weaver) index (I), observed heterozygosity (H_o), expected heterozygosity (H_e) and unbiased expected heterozygosity (uH_e). GenAlEx (Peakall and Smouse 2006, 2012) was also used to calculate probability of identity (PI), probability of exclusion (PE), inbreeding coefficient within individuals (F_{is}) and polymorphic information content (PIC), which are used to measure the usefulness of molecular markers. GenAlEx was used to calculate the pairwise relatedness based on the Queller and Goodnight (1989) estimator. In order to visualise the level of similarity between individuals, a Principal Coordinates Analysis plot (PCoA) was constructed using GenAlEx (Peakall and Smouse 2006, 2012).

3.3 Assessment of Mitochondrial Genes for *Pecten sulcicostatus*

Two candidate mtDNA genes were originally targeted, namely the Cytochrome c oxidase subunit I (CO1) gene and the 16S rRNA gene (Table 3.3), as these mtDNA genes are often used in phylogenetic studies (Barucca *et al.* 2004; Saavedra and Peña 2005; Yuan *et al.* 2012; Marín *et al.* 2013; Marín *et al.* 2015). The primer pairs tested in this study were chosen as they have been used in a number of species (Saavedra and Peña 2005; Ward *et al.* 2005; Ivey and Santos 2007; von Ritelen *et al.* 2007) and are thus considered universal primers.

Table 3.3: Mitochondrial primer pairs for both the CO1 and 16S rRNA genes tested in the present study.

Primer name	Gene	Primer sequence (5' – 3')	T _a (°C)	Reference
COI-F-Car	CO1	GCTGCTAATTTTATATCTACAG	56	von Ritelen <i>et al.</i> 2007
COI-R-Car		TGTGTAGGCATCTGGGTAATC		
LCO1490		GGTCAACAAATCATAAAGATATTG	49	Ivey and Santos 2007
HCO2198		TAAACTTCAGGGTGACCAAAAAATCA		
FishF1		TCAACCAACCACAAAGACATTGGCAC		
FishR1		TAGACTTCTGGGTGGCCAAAGAATCA		
FishF2		TCGACTAATCATAAAGATATCGGCAC	50	Ward <i>et al.</i> 2005
FishR2		ACTTCAGGGTGACCGAAGAATCAGAA		
CRUST16SF	16S	TAATTCAACATCGAGGTCGCAA	53	Ivey and Santos 2007
CRUST16SR		TTTGTACCTTKTGTATTAGG		
Pec16S-F1		GTTTTAAGGTCGGGGAAAG		
16Sar		CGCCTGTTTATCAAAAACAT	55	Saavedra and Peña 2005
16Sbr		CCGGTCTGAACTCAGATCACGT		

Extracted genomic DNA (gDNA) was standardised to 20 ng/ μ l and used as a template to amplify the target DNA regions *via* PCR. Individual reactions were carried out in 10 μ l total volume, containing 20 ng of template DNA, 2.0 mM of MgCl₂, 200 μ M of dNTPs, 0.5 U of *Taq* DNA polymerase and 0.2 μ M of each primer. Amplification of the mitochondrial markers was carried out in the presence of negative (no template) controls. The PCR cycling conditions for all primer pairs are listed in Table 3.4.

Table 3.4: PCR cycling conditions for the various mitochondrial primer pairs tested in the present study.

Steps	Primer pairs		
	COI-Car, CRUST16S, LCO1490 and HCO2198	Fish	16S
Initial activation	94°C - 5 minutes	94°C - 2 minutes	94°C - 4 minutes
Core denaturing, annealing and extension	30 cycles	35 cycles	35 cycles
	94°C - 45 seconds	94°C - 30 seconds	95°C - 1 minute
	T _a °C - 45 seconds	T _a °C - 30 seconds	T _a °C - 30 seconds
	72°C - 45 seconds	72°C - 1 minute	72°C - 30 seconds
Final extension	72°C - 10 minutes	72°C - 10 minutes	72°C - 10 minutes

PCR products were separated by electrophoresis in an EtBr stained 2% (w/v) agarose gel and viewed under shortwave UV in order to verify amplification of a DNA fragment of the expected size, as determined by comparison to a 1 kb HyperLadder size standard (Bioline). In order to remove unused primers, primer dimers and non-target genomic DNA, PCR products were purified using the GeneJET PCR Purification kit (Thermoscientific) according to manufacturer's instructions.

3.3.1 Sanger Sequencing

Sanger sequencing of purified PCR products was performed with the forward primer using the BigDye terminator kit version 3.0 (Applied Biosystems). Individual reactions were carried out in 10 μ l total volume, containing 10 to 20 ng of template DNA, 1x

BigDye Sequencing Buffer and 1 μ M of forward primer. The sequencing reaction products were resolved using a 3100 ABI automated sequencer (Applied Biosystems) at Stellenbosch University CAF.

3.3.2 Data Analysis

Chromatograms were viewed in FinchTV version 1.4.0 (Geospiza, <http://www.geospiza.com/finchTV>) in order to verify good quality sequence data. Sequences were manually edited and aligned using ClustalX (Thompson *et al.* 1997) as implemented in the computer software MEGA version 6.06 (Tamura *et al.* 2013). Following multiple sequence alignment, polymorphic sites were identified by eye and inspected to further ensure accuracy of raw data reads. In order to confirm that the target genes were sequenced as well as to confirm species identification, a Basic Local Alignment Search Tool (BLAST) search was performed on the publicly available databases GenBank, EMBL and DDBJ (Altschul *et al.* 1990) using consensus sequences generated in BioEdit version 7.2.5 (Hall 1999).

Haplotype inference, the number and composition of polymorphic sites, nucleotide diversity and haplotype diversity were calculated using DNA Sequence Polymorphism (DnaSP) version 5.10 (Librado and Rozas 2009). A haplotype network was drawn using the Median Joining method (Bandelt *et al.* 1999) as implemented in the computer software NETWORK version 4.6.1.3 (available at <http://www.fluxus-engineering.com/sharenet.htm>). A Maximum Likelihood tree was constructed in MEGA with significance testing of the bifurcations using the Bootstrap method (1 000 replications).

4. Results and Discussion

4.1 Microsatellite Markers

Of the 20 microsatellite markers designed for the great scallop, *P. maximus*, 14 were successfully amplified in *P. sulcicostatus* (Table 3.5). A microsatellite marker was deemed polymorphic when two bands were distinguishable in a single individual and/or there were clear size differences between bands of different individuals

(Figure 3.1 and 3.2). All 14 successfully amplifying microsatellite markers also displayed polymorphism following polyacrylamide gel electrophoresis (PAGE).

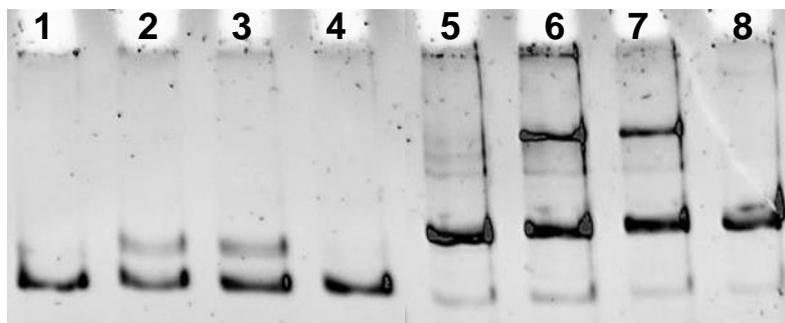


Figure 3.1: 12% polyacrylamide gel image of PCR products of the PmNH11 (lanes 1-4) and PmGC01 (lanes 5-8) loci.

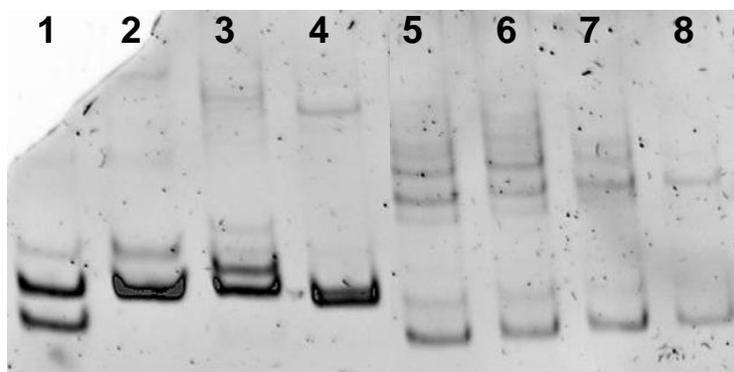


Figure 3.2: 12% polyacrylamide gel image of PCR products of the PmGC05 (lanes 1-4) and PmGC10 (lanes 5-8) loci.

The 14 successfully amplifying polymorphic microsatellite markers were fluorescently labelled with FAM, VIC, PET or NED dyes (Applied Biosystems) and optimised into three multiplex genotyping systems (Table 3.5). Of the 14 successfully amplifying polymorphic microsatellite markers only 11 microsatellite markers successfully amplified in the multiplex PCR and displayed clear, scorable peaks (Table 3.5). Of the 11 microsatellite markers, one microsatellite marker (PmGC20) was found to be monomorphic (Table 3.5) and was therefore removed from the set of microsatellite markers.

Table 3.5: Microsatellite markers grouped into three multiplexes with florescent label, motif, predicted and actual size range as well as annealing temperature (T_a).

Multiplex	Locus	Label	Motif	Predicted	Amplification	Actual	T _a (°C)
				size range (bp)		size range (bp)	
1	PmNH11	FAM	(ATTT) ₄	304 - 308	+	289 - 305	53
	PmGC01	PET	(ATG) ₈	125 - 134	+	102 - 111	
	PmGC05	VIC	(TA) ₈	175 - 201	+	164 - 178	
	PmGC10	NED	(AT) ₇	187 - 193	+	159 - 194	
2	PmRM041	FAM	(CAA) ₁₀	268 - 299	+	258 - 312	55
	PmGC18	NED	(AT) ₃ T(AT) ₃	219 - 223	-	-	
	PmGC03	VIC	(AT) ₇	242 - 246	+	218 - 225	
	PmGC06	FAM	(TG) ₆	90 - 94	+	64 - 71	
	PmGC20	PET	(AT) ₄ T(AT)	105 - 107	+	78	
3	PmNH62	FAM	(AAC) ₉	254 - 266	-	-	58
	PmNH70	PET	(AAT) ₈	140 - 146	+	116 - 125	
	PmRM002	FAM	(ATGT) ₇	97 - 155	+	90 - 208	
	PmRM007	VIC	(AGG) ₁₀	212 - 257	-	-	
	PmRM027	NED	(AGG) ₁₁	164 - 206	+	143 - 149	

Of the original 20 microsatellite markers designed for the great scallop, *P. maximus*, 10 microsatellite markers successfully amplified and displayed polymorphism in *P. sulcicostatus*, relating to a success rate of 50%. Typically cross-species transfer of microsatellite markers has a success rate of approximately $72 \pm 7\%$ in invertebrates (Barbará *et al.* 2007). Although high cross-species amplification success has been observed within genera, a number of factors may affect microsatellite marker transferability, including the phylogenetic distance between the source and target species (Barbará *et al.* 2007). Furthermore, even though cross-species amplification may be successful, high levels of polymorphism cannot be guaranteed (Chambers and MacAvoy 2002; Zane *et al.* 2002), evident by the fact that, although 14 of the 20 microsatellite markers (70%) designed for *P. maximus* successfully amplified in *P. sulcicostatus*, only 10 of these microsatellite markers successfully amplified in the multiplex PCR and displayed polymorphism. It is important to note that, although microsatellite marker PmGC20 was monomorphic in the False Bay population, it may be polymorphic in other *P. sulcicostatus* populations.

In a study by Marín *et al.* (2012), 12 microsatellite loci were isolated for *Argopecten purpuratus* and cross-amplified into four other scallop species, namely *A. irradians*, *P. maximus*, *P. yessoensis* and *P. albicans*. The success rate of the cross-species amplification from *A. purpuratus* to the other four scallop species ranged from 0% to 41.67% (Marín *et al.* 2012). Similarly, in a study by Ibarra *et al.* (2006), 35 microsatellite loci were isolated for *Nodipecten subnodosus* and cross-amplified into four other scallop species, namely *A. ventricosus*, *Nodipecten nodosus*, *A. irradians* and *A. purpuratus*. The success rate of the cross-amplification from *N. subnodosus* to the other four scallop species ranged from 0% to 77.14% (Ibarra *et al.* 2006). The 50% success rate observed in this study is therefore comparable to that of similar studies on scallop species.

The final set of microsatellite markers developed for *P. sulcicostatus* included 10 microsatellite markers previously developed for *P. maximus* (Charrier *et al.* 2012; Hold *et al.* 2013; Morvezen *et al.* 2013). The set of 10 microsatellite markers was used to determine population genetic statistics for the False Bay population in order to evaluate the extent of intrapopulation genetic diversity. The number of alleles

ranged from 2 (PmNH70) to 36 (PmRM002), with observed heterozygosity ranging from 0.000 to 0.946 and unbiased expected heterozygosity ranging from 0.053 to 0.977 (Table 3.6).

With a null allele frequency above 5% (0.050), microsatellite markers PmGC06, PmNH11, PmNH70 and PmRM002 (Table 3.6) showed the presence of null alleles. Null alleles are often due to a point mutation within the primer binding site and are therefore common when using non-species specific microsatellite markers such as those developed *via* cross-species transfer (O' Connell and Wright 1997). Although microsatellite markers PmGC06 and PmNH11 displayed a null allele frequency of 0.077 and 0.073, respectively (Table 3.6), these low null allele frequencies will most likely not influence the results. The high null allele frequency for PmNH70 (0.986) is most likely responsible for the low PIC value found for this microsatellite marker (0.051) (Table 3.6), as the presence of null alleles may cause a significant deficit of heterozygotes (O' Connell and Wright 1997). Interestingly, despite a null allele frequency of 0.229, microsatellite marker PmRM002 has the highest PIC value (0.963), although the high PIC value is most likely due to the high number of alleles (36) (Table 3.6). Null alleles are commonly found in bivalves (Marín *et al.* 2012; Silva and Gardner 2014), resulting in the underestimation of heterozygotes and consequently deviations from Hardy-Weinberg Equilibrium (HWE) (Silva and Gardner 2014). For that reason these markers should be used with caution (Shaw *et al.* 1999). Microsatellite markers PmGC05, PmNH70 and PmRM002 deviated from Hardy-Weinberg expectations (Table 3.6), likely due to the presence of null alleles.

The inbreeding coefficient, F_{IS} , ranged from -0.148 (PmGC01) to 1.000 (PmNH70) with an average of 0.189. Although an F_{IS} value of 1.000 is extremely high, this is due to the fact that microsatellite marker PmNH70 only had 2 alleles (Table 3.6). Four of the ten microsatellite markers (PmGC01, PmGC05, PmRM027 and PmRM041) showed negative F_{IS} values (Table 3.6) indicating an excess of heterozygotes. Polymorphic Information Content (PIC) ranged from 0.051 (PmNH70) to 0.963 (PmRM002) with an average PIC of 0.505 (Table 3.6). The mean PIC value of 0.505 suggests that these microsatellite markers could be useful in future studies on both natural populations as well as during aquaculture endeavours. As PIC values greater than 0.5 indicate highly informative microsatellite markers (Lafarga-de

la Cruz *et al.* 2015), microsatellite markers PmGC05 (0.554), PmGC10 (0.869), PmNH11 (0.752), PmRM002 (0.963) and PmRM041 (0.912) are deemed highly informative (Table 3.6). Another way to evaluate the usefulness of microsatellite markers is by determining their probability of inclusion (PI) and probability of exclusion (PE), which are useful in parentage analyses. The PI is the power of a microsatellite marker that allows the inclusion of a candidate parent when neither parent is known (Lafarga-de la Cruz *et al.* 2015). The PI for the tested microsatellite markers ranged from 0.003 (PmRM002) to 0.899 (PmNH70). The PI for all 10 loci combined is therefore 1.059×10^{-9} . On the other hand, the PE is the power of a microsatellite marker to exclude a candidate parent when neither parent is known (Lafarga-de la Cruz *et al.* 2015) and is therefore inversely proportional to the PI. In this study the PE for the tested microsatellite markers ranged from 0.049 (PmNH70) to 0.990 (PmRM002). The total exclusionary power, the combined power of the set of microsatellite markers to exclude a randomly selected unrelated candidate parent from parentage of an arbitrary offspring, was 0.999. These high probability values indicate that these microsatellite markers would be highly effective in parentage assignment, which will aid in the assessment of broodstock contribution to the next generation. Unequal broodstock contribution is a common occurrence in aquaculture and may result in a lower effective population size than the actual broodstock population size (Frankham 1995).

Table 3.6: Number of alleles (A_n), effective number of alleles (A_e), information (Shannon-Weaver) index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), Hardy-Weinberg Equilibrium (HWE) p value, frequency of null alleles (fr_{null}), inbreeding coefficient within individuals (F_{is}), polymorphic information content (PIC), probability of identity (PI) and probability of exclusion (PE) for the False Bay *Pecten sulcicostatus* population.

Locus	A_n	A_e	I	H_o	H_e	uH_e	p	fr_{null}	F_{is}	PIC	PI	PE
PmGC01	4	1,316	0,482	0,216	0,240	0,243	0,504	0,039	-0,148	0,222	0,596	0,208
PmGC03	6	1,827	0,921	0,432	0,453	0,459	0,833	0,012	0,228	0,417	0,335	0,411
PmGC05	9	2,456	1,291	0,595	0,593	0,601	0,003*	0,028	-0,076	0,554	0,204	0,566
PmGC06	4	1,364	0,574	0,216	0,267	0,271	0,079	0,077**	0,415	0,255	0,549	0,256
PmGC10	17	8,297	2,425	0,865	0,879	0,892	0,160	0,029	0,154	0,869	0,025	0,918
PmNH11	8	4,602	1,707	0,703	0,783	0,793	0,223	0,073**	0,036	0,752	0,078	0,771
PmNH70	2	1,056	0,124	0,000	0,053	0,053	0,014*	0,986**	1,000	0,051	0,899	0,049
PmRM002	36	27,657	3,441	0,541	0,964	0,977	0,000*	0,229**	0,410	0,963	0,003	0,990
PmRM027	3	1,056	0,143	0,054	0,053	0,054	1,000	0,000	-0,023	0,052	0,898	0,052
PmRM041	22	12,009	2,786	0,946	0,917	0,929	0,804	0,000	-0,103	0,912	0,012	0,958

*Microsatellite markers which deviate from Hardy-Weinberg expectations.

**Microsatellite markers which showed the presence of null alleles.

In comparison with the studies in which the microsatellite markers were originally developed (Charrier *et al.* 2012; Hold *et al.* 2013; Morvezen *et al.* 2013), the number of alleles was nearly similar for five of the ten microsatellite markers (PmGC01, PmGC03, PmGC05, PmGC06 and PmNH70) (Table 3.7), however, the number of alleles for *P. sulcicostatus* was significantly higher than that for *P. maximus* for four of the ten microsatellite markers (PmGC10, PmNH11, PmRM002 and PmRM041) (Table 3.7). This higher number of alleles in *P. sulcicostatus* is surprising as the sample sizes for the two studies from which these microsatellite markers were developed (Hold *et al.* 2013 - 125 individuals; Morvezen *et al.* 2013 - 48 individuals) were much higher than that of the current study (37 individuals). Although the use of non-species specific microsatellite markers may result in the underestimation of genetic diversity (Arif *et al.* 2010), wild populations of *P. maximus* are over-exploited and may therefore exhibit low genetic diversity (Gosling 2003), especially in comparison to *P. sulcicostatus*, which has never been exploited (Arendse and Pitcher 2012). Interestingly, for microsatellite marker PmRM027, the number of alleles was much lower for *P. sulcicostatus* (3 alleles) than for *P. maximus* (15 alleles) (Table 3.7).

Table 3.7: Comparison of number of alleles (A_n) and observed heterozygosity (H_o) for *Pecten maximus*, as taken from Charrier *et al.* (2012), Hold *et al.* (2013) and Morvezen *et al.* (2013), and *Pecten sulcicostatus* (current study).

Locus	<i>P. maximus</i>		<i>P. sulcicostatus</i>	
	A_n	H_o	A_n	H_o
PmGC01	3	0.227	4	0.216
PmGC03	3	0.256	6	0.432
PmGC05	11	0.703	9	0.595
PmGC06	3	0.039	4	0.216
PmGC10	4	0.426	17	0.865
PmNH11	2	0.180	8	0.703
PmNH70	3	0.050	2	0.000
PmRM002	12	0.810	36	0.541
PmRM027	15	0.850	3	0.054
PmRM041	12	0.850	22	0.946

The pairwise relatedness for the False Bay *P. sulcicostatus* population ranged from -0.613 to 0.566 with an average of -0.028, indicating that individuals in this population show low relatedness. Similarity between individuals was visualised using a PCoA, in which the majority of *P. sulcicostatus* individuals clustered together, with the exception of a few individuals (Scallop 3, Scallop 4, False Bay 04, False Bay 08, False Bay 11, False Bay 26 and False Bay 30). The genetic similarity amongst the majority of the *P. sulcicostatus* individuals may be as a result of relatedness due to the fact that the individuals were sampled from one location (False Bay), with a maximum distance of five kilometers between sampling, although the pairwise relatedness indicated low relatedness. Although the life cycle of *P. sulcicostatus* is not well studied (Arendse *et al.* 2008), in most scallop species, larvae remain in the water column for two to four weeks before dissipating to the ocean floor (Leavitt *et al.* 2010), therefore it is possible for larvae to travel large distances, aided by ocean currents.

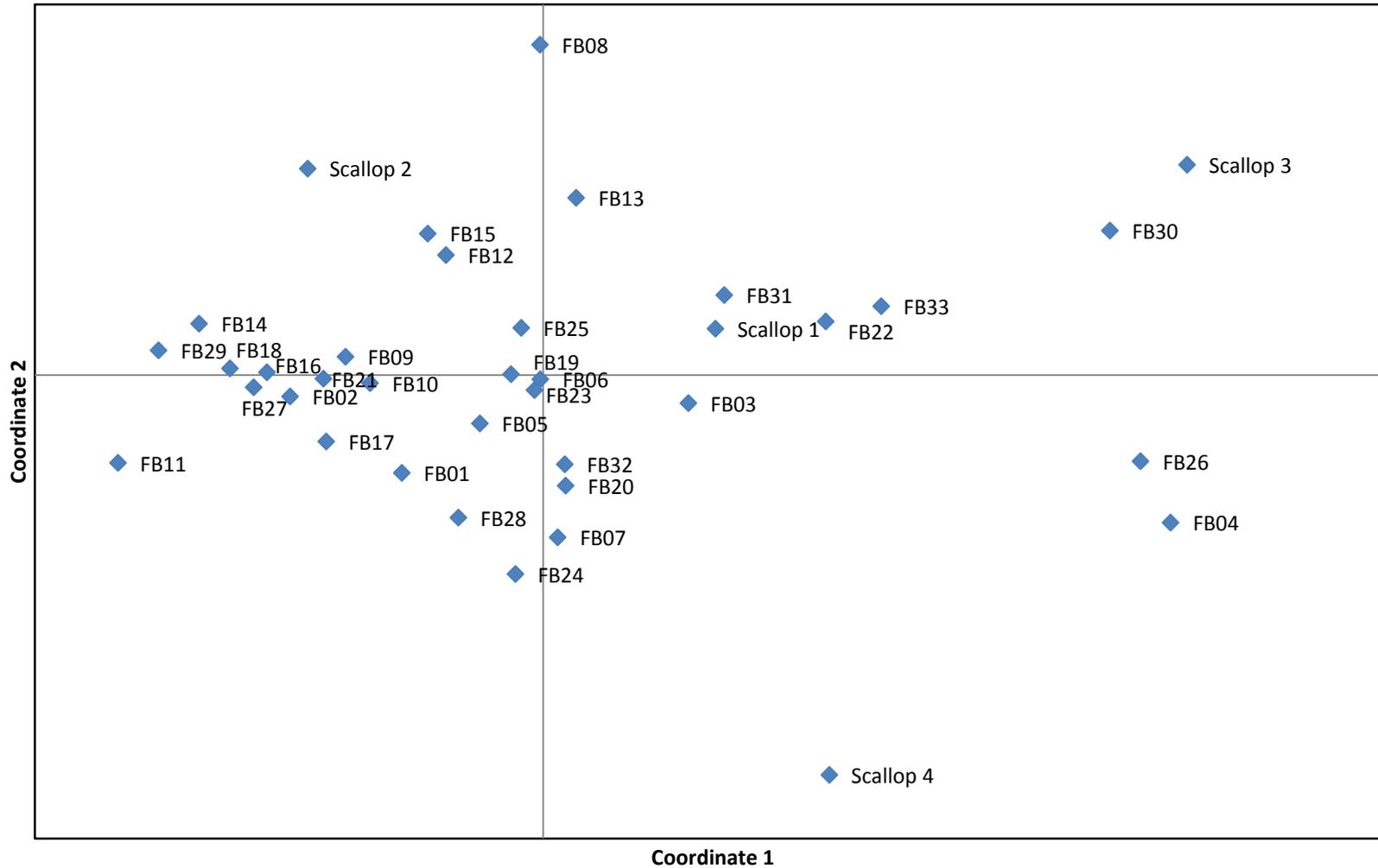


Figure 3.3 Principle coordinates analysis (PCoA) of the False Bay *Pecten sulcicostatus* individuals based on 10 microsatellite markers visualising the level of similarity of individuals.

4.2 Mitochondrial Genes

The COI-F-Car and COI-R-Car (von Ritzel *et al.* 2007), LCO1490 and HCO2198 (Ivey and Santos 2007) and Fish2 (FishF2 and FishR2) primer pairs (Ward *et al.* 2005) failed to amplify a fragment of the *CO1* gene. The CRUST16S (CRUST16SF and CRUST16SR) primer pair (Ivey and Santos 2007) also failed to amplify a fragment of the 16S rRNA gene. Successful amplification of a 438 base pair (bp) fragment of the *CO1* gene was obtained using the Fish1 (FishF1 and FishR1) primer pair (Ward *et al.* 2005), with an optimal T_a of 50°C, however only two of the four individuals were successfully sequenced. Following a BLAST search of the *P. sulcicostatus* *CO1* sequence, the best match, with 79% identity, was the *CO1* gene of *Patelloida striata* (sea snail) from Japan (GenBank accession number: AB161589.1). Following the construction of a Maximum Likelihood tree it was clear that the *CO1* sequences generated for *P. sulcicostatus* did not cluster with that of any other scallop species; therefore it is likely that the primer pair failed to amplify the target region, instead amplifying a non-target region.

Sequencing of the 16S amplicons generated using the Pec16S-F1 and 16Sbr primer pair (Saavedra and Peña 2005), with an optimal T_a of 55°C, yielded a sequence of approximately 650 bp in length. After manual trimming, sequences of 582 bp in length were generated for all *P. sulcicostatus* individuals. The best match for the BLAST searches of the *P. sulcicostatus* 16S sequences, with 97 to 98% identity, was the 16S rRNA gene of *P. maximus* (GenBank accession number: KF982791.1). Although the 16Sar and 16Sbr primer pair (Saavedra and Peña 2005) also successfully amplified a fragment of the 16S rRNA gene, the Pec16S-F1 forward primer (Saavedra and Peña 2005) was selected for sequencing, as it amplified a longer fragment of the 16S rRNA gene.

Comparison of the 38 aligned sequences allowed the identification of 23 polymorphic sites (Table 3.8). The 23 polymorphic sites consisted of 13 singleton variable sites with two variants and 10 parsimony informative sites with two variants. The 23 polymorphic sites allowed for the identification of 18 distinct haplotypes (Appendix: Table A1). A large number of the polymorphic sites were located in the 5' end of the 16S rRNA gene. A study by Marín *et al.* (2015) found that the intraspecific variation

at the 5' end of the 16S rRNA gene is twice as high as that at the 3' end, likely due to the fact that different regions of the 16S rRNA gene evolve at different rates.

The 582 bp fragment of the 16S rRNA gene analysed in this study yielded high levels of intraspecific variation, with overall haplotype diversity of 0.9132 and overall nucleotide diversity of 0.0094 (Table 3.8), making the 582 bp fragment of the 16S rRNA gene suitable for studying phylogeographic patterns. The high haplotype diversity (0.9132) observed for *P. sulcicostatus* in this study (Table 3.8) is higher than that observed in other *Pecten* species for the 16S rRNA gene - 0.857 for *P. maximus*, 0.667 for *P. jacobeus*, 0.800 for *P. novaezelandiae* and 0.000 for *P. fumatus* (Saavedra and Peña 2004). Although no haplotype diversity was observed for *P. fumatus* (0.000), it is important to note that only six individuals were included in the study, all of which shared a common haplotype (Saavedra and Peña 2004). The high haplotype diversity (0.9132) observed for *P. sulcicostatus* in the current study indicates that this group of individuals has maintained a relatively large population size. The fact that the *P. sulcicostatus* haplotype diversity was higher than that of other *Pecten* species (0.000 to 0.857) further corroborates that this group of individuals has maintained a large population size, as *P. maximus*, *P. jacobeus*, *P. novaezelandiae* and *P. fumatus* are all commercially exploited (Saavedra and Peña 2004), while *P. sulcicostatus* remains unexploited (Arendse *et al.* 2012). Although there is little information available on the population size of *P. sulcicostatus* in False Bay, the high haplotype diversity supports the observed high densities near the centre of the bay (Arendse *et al.* 2008).

The low nucleotide diversity (0.0094) exhibited in the current study (Table 3.8) indicates a low level of genetic variation. Despite evidence of low genetic variation, the nucleotide diversity observed in the current study is still much higher than that observed for other *Pecten* species - 0.0039 for *P. maximus*, 0.0015 for *P. jacobeus*, 0.0020 for *P. novaezelandiae* and 0.000 for *P. fumatus* (Saavedra and Peña 2004). As mentioned before, commercial exploitation of these four species may have resulted in the observed low nucleotide diversity (Saavedra and Peña 2004). High haplotype diversity coupled with low nucleotide diversity, as seen in both this study as well as the study by Saavedra and Peña (2004), is usually characteristic of the

occurrence of a genetic bottleneck followed by rapid population expansion (Avisé 2000; D'Arcy *et al.* 2013).

Table 3.8: The number and composition of polymorphic sites and haplotype and nucleotide diversity for a 582 bp fragment of the 16S rRNA gene of *Pecten sulcicostatus*.

<i>P. sulcicostatus</i>	
Number of individuals	38
Number of haplotypes	18
Monomorphic sites	559
Polymorphic sites	23
Singleton variable sites	13
Parsimony informative sites	10
Haplotype diversity	0.9132
Nucleotide diversity	0.0094

A Median Joining network of haplotypes was drawn to infer evolutionary relationships among the *P. sulcicostatus* haplotypes. The network depicted a complex branching pattern that was dominated by two highly divergent clusters (clades) of haplotypes (Figure 3.5), suggesting that two highly divergent maternal lineages were present when *P. sulcicostatus* colonised False Bay. Both clades are characterised by many unique, low-frequency haplotypes. Furthermore, both clades were characterised by a star-shaped structure, where many low frequency haplotypes differ by a single mutational step from a main common haplotype. Such patterns are often associated with populations that originated from a small number of individuals leading to reduced genetic diversity, in which enough time has elapsed for single-step mutations to arise from common haplotypes (D'Arcy *et al.* 2013). Interestingly, haplotype 1 (Figure 3.5), which represents the historic *P. sulcicostatus* individual sampled in 1991, clusters closely with six other *P. sulcicostatus* haplotypes sampled in 2014, despite the 23 years between sampling.

Another possible explanation for the two divergent haplotype groups could be that one clade represents the other *Pecten* species that can be found in South Africa, *Pecten afribenedictus*. Although the natural distribution range of *P. afribenedictus* is

believed to be along the continental shelf of the east coast of South Africa, from East London to northern Zululand (Kosi Bay), most scallop “sightings” are of shells washed up on beaches and therefore the distribution range is most likely inaccurate. Morphologically, *P. sulcicostatus* and *P. aribenedictus* have similar shell shape and size, however the lower valve of *P. aribenedictus* is more convex than that of *P. sulcicostatus* and the ribs on the lower valve are less defined, irregular and have a smooth surface (Figure 3.4). Also, the upper valve of *P. sulcicostatus* is almost flat compared to the concave upper valve of *P. aribenedictus* (Figure 3.4). Although *P. sulcicostatus* and *P. aribenedictus* exhibit morphological differences, species identification based on morphological characteristics was not performed, therefore, although it is unlikely, there is a small possibility that *P. aribenedictus* individuals inhabit False Bay and that a few of these individuals were sampled in this study.



Figure 3.4: Shell morphology of (a) *Pecten sulcicostatus* and (b) *Pecten aribenedictus*, showing the similar shell shape and size as well as the more convex lower valve of *Pecten aribenedictus* (b) and the almost flat upper valve of *Pecten sulcicostatus* (a) compared to the concave upper valve of *Pecten aribenedictus*. Taken from Pecten Site (www.pectensite.com/).

It is important to note that, although the 16S rRNA gene shows two highly divergent clades (Figure 3.5), the set of microsatellite markers did not show two clusters of individuals, therefore the two divergent clades are most likely due to two highly divergent maternal lineages, as opposed to the sampling of *P. aribenedictus* individuals in this study.

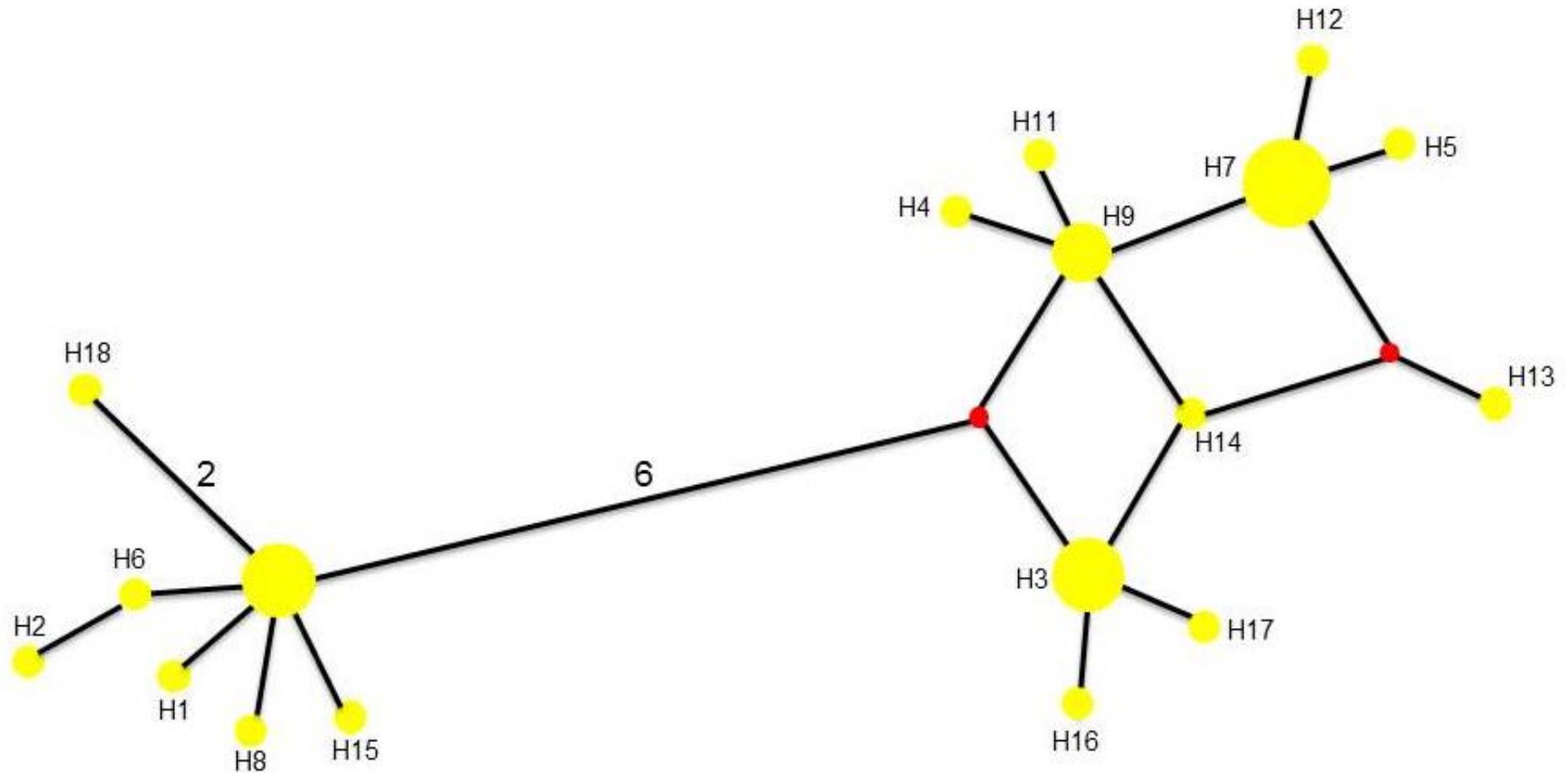


Figure 3.5: Haplotype network representing the relationships between the 18 *Pecten sulcicostatus* haplotypes identified using a 582 bp fragment of the 16S rRNA gene. Node size reflects the frequency of the haplotypes. The number of mutations/mutational steps between each haplotype is one unless otherwise stated. Intermediate missing haplotypes are indicated as red dots.

Sequencing of the 16S amplicons yielded a sequence of approximately 650 bp in length for the various *Pecten* species. After manual trimming, sequences of 504 bp in length were generated for all the 38 *P. sulcicostatus*, 33 individuals of various *Pecten* species and two *Euvola* species. Amplification of the fragment of the 16S rRNA gene produced a single thick band on the 2% (w/v) agarose gel for all of the *Pecten* species, except *P. keppelianus*. Interestingly, amplification of the fragment of the 16S rRNA gene in all *P. keppelianus* samples produced two faint bands (Figure 3.6). Despite changing various PCR conditions, the two faint bands remained and sequencing of the PCR products produced poor quality sequence data which was not usable. The occurrence of two bands may indicate that the primer pair amplified two mitochondrial genome regions of similar lengths or it may indicate nuclear copies of mitochondrial DNA (NUMTs), which are as a result of recombination or crossing-over that transposed a portion of the 16S rRNA gene onto the nuclear genome (Mirimin *et al.* 2015). At present no NUMTs are known from bivalves (Bensasson *et al.* 2001; Zbawicka *et al.* 2007), although this possibility cannot be excluded.

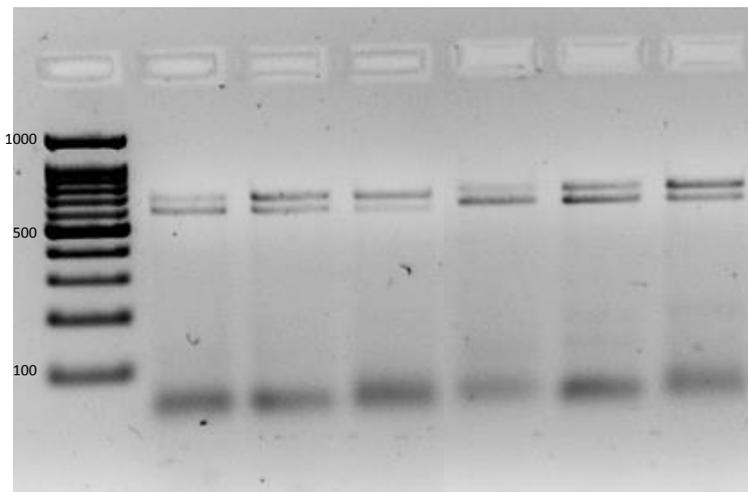


Figure 3.6: 2% agarose gel image of PCR products of the 16S primer pair using gDNA from two replications of the three *Pecten keppelianus* individuals.

In order to confirm species identification, a BLAST search was performed on publicly available databases (Altschul *et al.* 1990) using the consensus sequences generated for each species. All species were confirmed as the correct species, except for *P. jacobus*, which was matched to *P. maximus* with 99% identity (Table 3.9), and *P. fumatus*, which was matched to *P. novaezelandiae* with 98% identity (Table 3.9).

The *P. fumatus* consensus sequence also matched to the *P. fumatus* 16S rRNA gene (GenBank accession number: JF339110.1) with 98% identity, however the coverage was only 90%. The *P. sulcicostatus* consensus sequence matched to *P. maximus* with 94% identity (Table 3.9) as there are no *P. sulcicostatus* sequences available on the publicly available databases.

Table 3.9: BLAST results of the consensus sequences of the various *Pecten* - and outgroup species, indicating the best match as well as the percentage identity.

Species	Match	GenBank accession number	Identity (%)
<i>P. fumatus</i>	<i>P. novaezelandiae</i> 16S ribosomal RNA gene	AY650055.1	98
<i>P. novaezelandiae</i>			99
<i>P. jacobeus</i>	<i>P. maximus</i> mitochondrial 5.5kb fragment encoding genes flanking putative origin of replication	X92688.1	99
<i>P. maximus</i>			
<i>P. sulcicostatus</i>	<i>P. maximus</i> partial mitochondrion genome	KP900975.1	94
<i>E. perulus</i>	<i>Euvola perulus</i> isolate UF371263 16S ribosomal RNA gene	HM630517.1	99
<i>E. ziczac</i>	<i>Euvola ziczac</i> isolate 6 partial 16S rRNA gene	AJ972434.1	99

Comparison of the 71 aligned *Pecten* sequences allowed the identification of 40 polymorphic sites - 18 singleton variable sites with two variants, 19 parsimony informative sites with two variants and three parsimony informative sites with three variants (Table 3.11 - 3.13). A total of 33 (82.50%) transitions and seven (17.5%) transversions were observed, giving an observed transition to transversion (ts:tv) ratio of 4.71 (Table 3.10). Transitions (33) outnumbered transversions (7), as is typical of animal mitochondrial genomes (Saavedra and Peña 2004). The ts:tv ratio

observed in this study (4.71) is similar to that observed in a study by Saavedra and Peña (2004) for *P. maximus*, *P. jacobaeus*, *P. fumatus* and *P. novaezelandiae*, which found ts:tv ratios varying from 1.0 to 9.2 with an average of 5.5. At position 275 a single base deletion occurred in all of the *P. maximus* and *P. jacobaeus* sequences.

Table 3.10: Summary of the putative polymorphisms detected in a 504 bp fragment of the 16S rRNA gene.

Polymorphism	Total	Percentage (%)	
Transition	T/C	7	17.5
	A/G	4	10.0
	C/T	8	20.0
	G/A	14	35.0
Transversion	T/A	2	5.0
	T/G	2	5.0
	A/T	0	0.0
	A/C	0	0.0
	C/A	0	0.0
	C/G	0	0.0
	G/T	3	7.5
	G/C	0	0.0
Total	40	100.0	

Table 3.11: Polymorphic sites for sites 31 - 281 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species.

Site	32	33	39	67	154	258	264	265	266	268	270	275	279	281
Polymorphism	G/T	T/C	G/A	C/T	C/T	G/A	T/C	T/C	C/T	A/G	G/A	A/-	G/T	G/A
Minor allele	T	C	A	T	T	A	C	C	T	G	A	-	T	A
MAF (%)	1.41	23.94	1.41	16.90	1.41	36.62	2.82	1.41	36.62	1.41	1.41	21.13	1.41	9.86
<i>P. fumatus</i>	G	T	G	C	C	G	T	T	C	A/G	G	A	G	G
<i>P. jacobeus</i>	G	C	G	C	C	G	T	T	C	A	G	-	G/T	G/A
<i>P. maximus</i>	G	C	G	C	C	G	T/C	T	C	A	G	-	G	G/A
<i>P. novaezelandiae</i>	G	T/C	G	T	C	G	T	T	C	A	G	A	G	G
<i>P. sulcicostatus</i>	G/T	T	G/A	C	C/T	G/A	T	T/C	C/T	A	G/A	A	G	G

Table 3.12: Polymorphic sites for sites 284 - 319 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species.

Site	284	285	296	298	301	304	305	306	307	308	309	314	317	319
Polymorphism	C/T	G/A	G/A	G/A	C/T	T/G/C	G/A	G/A	T/C	T/C	T/G	T/C	C/T/A	G/T
Minor allele	T	A	A	A	T	G C	A	A	C	C	G	C	T A	T
MAF (%)	2.82	36.62	1.41	1.41	25.35	21.13 1.41	46.48	46.48	19.72	21.13	16.90	1.41	25.35 1.41	1.41
<i>P. fumatus</i>	C	G	G/A	G/A	C	C/T	A	A	T	T/C	T	T	T	G
<i>P. jacobeus</i>	C/T	G	G	G	C	G	A	A	C	C	T	T	C	G
<i>P. maximus</i>	C	G	G	G	C	G	A	A	T/C	T/C	T	T	C	G
<i>P. novaezelandiae</i>	C/T	G	G	G	C	T	A	A	T	T/C	T	T	T	G
<i>P. sulcicostatus</i>	C	G/A	G	G	C/T	T	G	G	T/C	T	T/G	T/C	C/A	G/T

Table 3.13: Polymorphic sites for sites 323 - 471 of a 504 bp fragment of the 16S rRNA gene of various *Pecten* species showing the type of polymorphism, minor allele, minor allele frequency (MAF) as well as the allele occurring in the various *Pecten* species.

Site	323	326	327	367	384	385	389	391	395	411	412	418	471
Polymorphism	G/A	T/A	G/A	C/T	G/A	T/A/C	A/G	A/G	G/A	A/G	C/T	G/A	T/C
Minor allele	A	A	A	T	A	A C	G	G	A	G	T	A	C
MAF (%)	36.62	1.41	14.08	1.41	25.35	16.90 1.41	1.41	2.82	1.41	1.41	1.41	14.08	1.41
<i>P. fumatus</i>	G	T	G	C	G/A	T/C	A/G	A	G	A	C/T	G	T
<i>P. jacobeus</i>	G	T	G	C	G/A	T	A	A	G	A	C	G	T/C
<i>P. maximus</i>	G	T	G	C	G	T	A	A	G	A	C	G	T
<i>P. novaezelandiae</i>	G	T	G	C	A	T	A	A	G	A	C	G	T
<i>P. sulcicostatus</i>	G/A	T/A	G/A	C/T	G	T/A	A	A/G	G/A	A/G	C	G/A	T

Despite the fact that all polymorphic sites were identified by eye and inspected to further ensure accuracy of raw data reads, 21 polymorphic sites had an allele frequency below 5%. Although it is possible that these polymorphic sites are as a result of sequencing error (Bensasson 2001), they cannot be excluded, as the sample size in the present study is very small. Small sample size may have resulted in unique, low frequency haplotypes being underrepresented. For the three parsimony informative sites with three variants, one of the three variants at all three sites had an allele frequency below 5%, therefore it is likely that they are in fact parsimony informative sites with only two variants (Bensasson 2001). Interestingly, 22 of the polymorphisms are only found in *P. sulcicostatus*, however 12 (54.5%) of the polymorphisms have an allele frequency below 5% and may therefore be as a result of sequencing error. In order to confirm polymorphism at these positions, sequencing in both the forward and reverse direction would have to be performed.

The deletion at position 275 only occurs in *P. maximus* and *P. jacobeus* and not in the other three *Pecten* species studied in the current study. Four *P. maximus* 16S rRNA sequences (GenBank accession numbers: AY650084.1, AY650057.1, FN667668.1, and EU379454.1) and four *P. jacobeus* 16S rRNA sequences (GenBank accession numbers: AJ245394.1, JQ611455.1, FN667671.1 and FN667670.1) were downloaded from NCBI and included in the haplotype alignment. All of the *P. maximus* and *P. jacobeus* 16S rRNA sequences contained this deletion, confirming that the deletion at position 275 is specific to *P. maximus* and *P. jacobeus*, which may indicate an adaptation to the North Atlantic and Mediterranean environment.

All 40 polymorphic sites were substitutions (Appendix: Table A2 and Table A3), which allowed the identification of 32 distinct haplotypes (Appendix: Table A4) across the 71 16S sequences. The overall haplotype diversity was determined to be 0.951 while the overall nucleotide diversity was determined to be 0.0153 (Table 3.14). The species-specific haplotype diversity ranged from 0.455 to 0.913 while the species-specific nucleotide diversity ranged from 0.0013 to 0.0094 (Table 3.14), with *P. novaezelandiae* having the lowest haplotype and nucleotide diversity (0.455 and 0.0013, respectively) and *P. sulcicostatus* having the highest haplotype and nucleotide diversity (0.913 and 0.0094, respectively) (Table 3.14).

Table 3.14: The number and composition of polymorphic sites, haplotype and nucleotide diversity for a 504 bp fragment of the 16S rRNA gene of various *Pecten* species.

	<i>P. maximus</i>	<i>P. jacobeus</i>	<i>P. novaezelandiae</i>	<i>P. fumatus</i>	<i>P. sulcicostatus</i>	All haplotypes
Number of individuals	7	8	12	6	38	71
Number of haplotypes	3	5	4	4	18	32
Polymorphic sites	3	5	3	9	23	40
Singleton variable sites	3	4	2	8	13	18
Parsimony informative sites	0	1	1	1	10	22
Haplotype diversity	0.762	0.857	0.455	0.800	0.913	0.951
Nucleotide diversity	0.0030	0.0031	0.0013	0.0064	0.0094	0.0153

A study by Saavedra and Peña (2004) also made use of the 16S rRNA gene to determine haplotype and nucleotide diversity in *P. maximus*, *P. jacobeus*, *P. novaezelandiae* and *P. fumatus* (Table 3.15). The number of individuals used in this study was comparable to that of the study by Saavedra and Peña (2004), except in the case of *P. novaezelandiae*, as the current study made use of double the number of samples (12). The current study observed lower haplotype and nucleotide diversity for *P. maximus* and *P. novaezelandiae* and higher haplotype and nucleotide diversity for *P. jacobeus* and *P. fumatus* (Table 3.15). Saavedra and Peña (2004) identified only one *P. fumatus* haplotype, resulting in a haplotype and nucleotide diversity of 0.000. Therefore the current study was able to shed light on a more realistic haplotype and nucleotide diversity for *P. fumatus*, namely 0.800 and 0.0064, respectively (Table 3.15).

In the present study, the lower haplotype and nucleotide diversity observed for *P. maximus* and *P. novaezelandiae* and higher haplotype and nucleotide diversity observed for *P. jacobeus*, may be due to sampling variation between the two studies. Individuals obtained in this study were sampled from a number of populations (*P. maximus* - two locations, *P. jacobeus* - two locations and *P. novaezelandiae* - four locations) whereas the study by Saavedra and Peña (2004) only made use of one wild (*P. jacobeus* and *P. novaezelandiae*) or cultured (*P. maximus* and *P. fumatus*) sampling location.

Table 3.15: Haplotype and nucleotide diversity for a fragment of the 16S rRNA gene of various *Pecten* species studied in this study (A) as well as a study by Saavedra and Peña (2004) (B).

Study	<i>P. maximus</i>		<i>P. jacobeus</i>		<i>P. novaezelandiae</i>		<i>P. fumatus</i>		<i>P. sulcicostatus</i>	
	A	B	A	B	A	B	A	B	A	B
Number of individuals	7	8	8	7	12	6	6	6	38	-
Number of haplotypes	3	7	5	4	4	4	4	1	18	-
Haplotype diversity	0.762	0.857	0.857	0.667	0.455	0.800	0.800	0.000	0.913	-
Nucleotide diversity	0.0030	0.0039	0.0031	0.0015	0.0013	0.0020	0.0064	0.000	0.0094	-

A later study by Saavedra and Peña (2005) investigated multiple *P. maximus* and *P. jacobeus* populations with increased sample size and found similar haplotype and nucleotide diversities as was found in the present study as well as the study by Saavedra and Peña (2004). For *P. maximus*, haplotype diversity ranged from 0.804 to 0.900 with an average of 0.817, while the nucleotide diversity ranged from 0.0028 to 0.0038 with an average of 0.0030 (Table 3.16). For *P. jacobeus*, haplotype diversity ranged from 0.650 to 0.857 with an average of 0.761, while the nucleotide diversity ranged from 0.0018 to 0.0034 with an average of 0.0027 (Table 3.16).

Table 3.16: Haplotype and nucleotide diversity for a fragment of the 16S rRNA gene of *Pecten maximus* and *Pecten jacobeus* in the current study (A), a study by Saavedra and Peña (2004) (B) and a study by Saavedra and Peña (2005) (C).

Study	<i>P. maximus</i>			<i>P. jacobeus</i>		
	A	B	C	A	B	C
Number of individuals	7	8	37	8	7	48
Number of haplotypes	3	7	15	5	4	15
Haplotype diversity	0.762	0.857	0.817	0.857	0.667	0.761
Nucleotide diversity	0.0030	0.0039	0.0030	0.0031	0.0015	0.0027

Most of the haplotypes identified in this study were species-specific, only being found in one species, except for two haplotypes (Haplotype 6 and Haplotype 8), which were shared by *P. jacobeus* and *P. maximus* (Appendix: Table A1). The fact that the fragment of the 16S rRNA gene used in this study was able to distinguish between all *Pecten* species except *P. jacobeus* and *P. maximus* supports a study by Barucca *et al.* (2004), based on the mitochondrial 16S and 12S rRNA genes, which found a small genetic distance between *P. maximus* and *P. jacobaeus*, suggesting that these species may in fact be two varieties of the same species. *Pecten maximus* is distributed along the Atlantic coast from Norway to Morocco, although it enters the Mediterranean along the Alboran Sea coast. In the remainder of the Mediterranean,

P. maximus is replaced by *P. jacobaeus* (Saavedra and Peña 2005). Although the Eastern Atlantic (*P. maximus*) and Mediterranean (*P. jacobaeus*) scallops have traditionally been considered distinct species, it is most likely that these two species are in fact the same species. The two *Pecten* species may represent two varieties of the same species, which following adaptation to different environmental conditions, developed different morphological characteristics (Barucca *et al.* 2004). Conspecificity is supported by the ability of *P. maximus* and *P. jacobaeus* to fully interbreed in captivity (Saavedra and Peña 2005).

A network of haplotypes was drawn to infer evolutionary relationships among the 32 16S rRNA gene haplotypes for various *Pecten* species (Figure 3.7). The network depicted a complex branching pattern characterised by many unique or low-frequency haplotypes (D'Arcy *et al.* 2013). As expected, *P. jacobaeus* and *P. maximus*, from the Eastern Atlantic- and Mediterranean Sea, cluster together while the Australian and New Zealand scallops, *P. novaezelandiae* and *P. fumatus* cluster together (Figure 3.7). The *P. sulcicostatus* haplotypes show two distinct clades, with the smaller clade clustering with the other *Pecten* species, whilst the larger clade is quite removed from the other scallop species (Figure 3.7). Despite the fact that *P. sulcicostatus* has previously been considered to be closely related to *P. maximus* and *P. jacobaeus* on the basis of morphology (Saavedra and Peña 2004), the haplotype network shows *P. sulcicostatus* to be quite removed from any of the other *Pecten* species. Also, the number of mutational steps between *P. sulcicostatus* and the other *Pecten* species is significant. Although the clear separation between *P. sulcicostatus* and the other *Pecten* species is significant, it is important to note that this separation is based solely on the 16S rRNA gene and therefore may not be completely accurate. In order to confirm these findings, more than one mitochondrial gene will have to be studied.

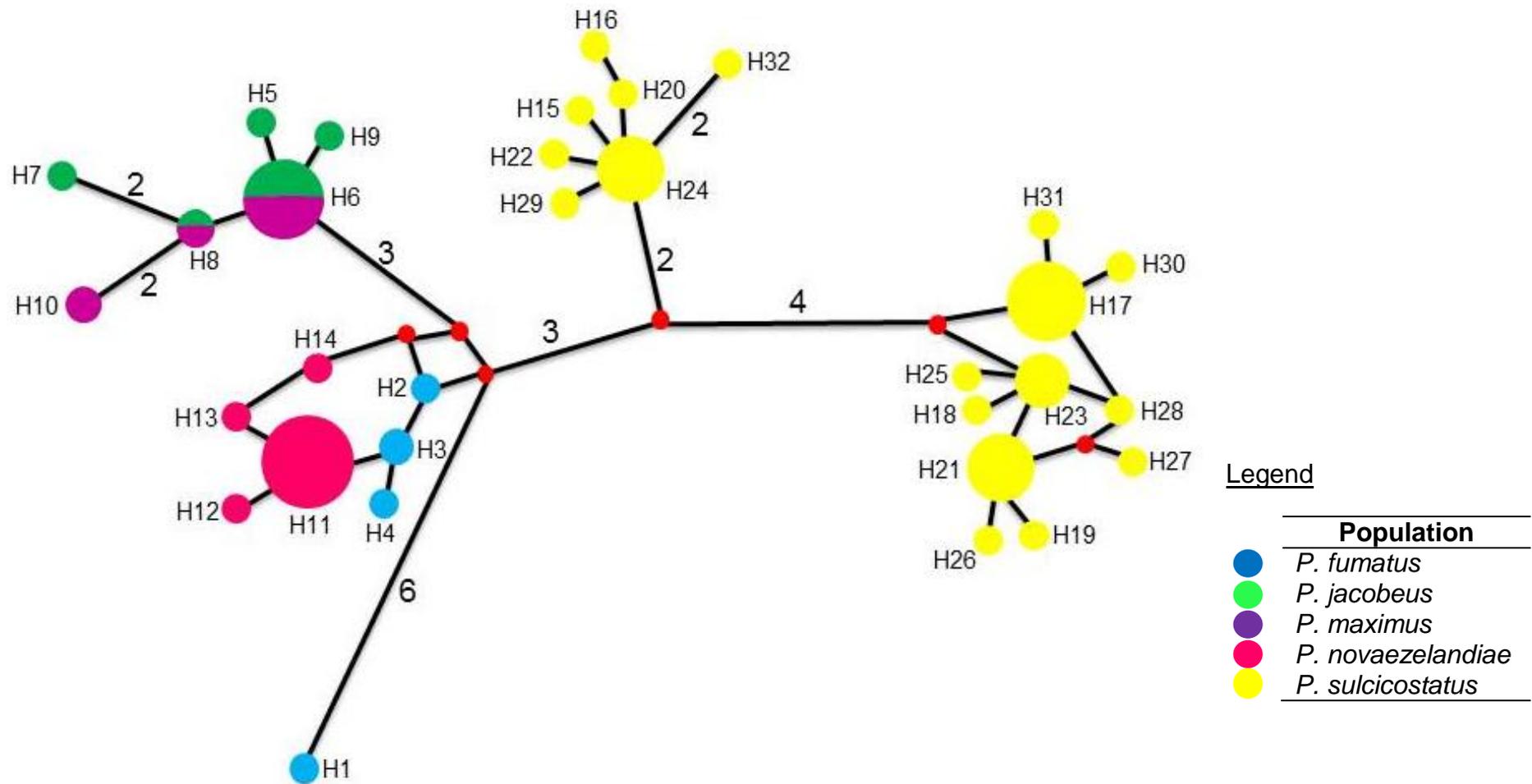


Figure 3.7: Haplotype network representing the relationships between the 32 haplotypes identified using a 504 bp fragment of the 16S rRNA gene for various *Pecten* species. Node size reflects the frequency of the haplotypes. The number of mutations/mutational steps between each haplotype is one unless otherwise stated. Intermediate missing haplotypes are indicated as red dots.

The maximum likelihood tree supported the network of haplotypes with *P. jacobeus* and *P. maximus* clustering together while *P. novaezelandiae* and *P. fumatus* clustered together (Figure 3.8). The *P. sulcicostatus* haplotypes show two distinct clades, with the larger clade clustering with the other *Pecten* species, whilst the smaller clade is quite removed from the other scallop species (Figure 3.8). As previously mentioned, the two divergent *P. sulcicostatus* clades are most likely due to two highly divergent maternal lineages, with a very slight possibility that the one clade represents *P. aribenedictus*, which is predominantly found along the east coast of South Africa.

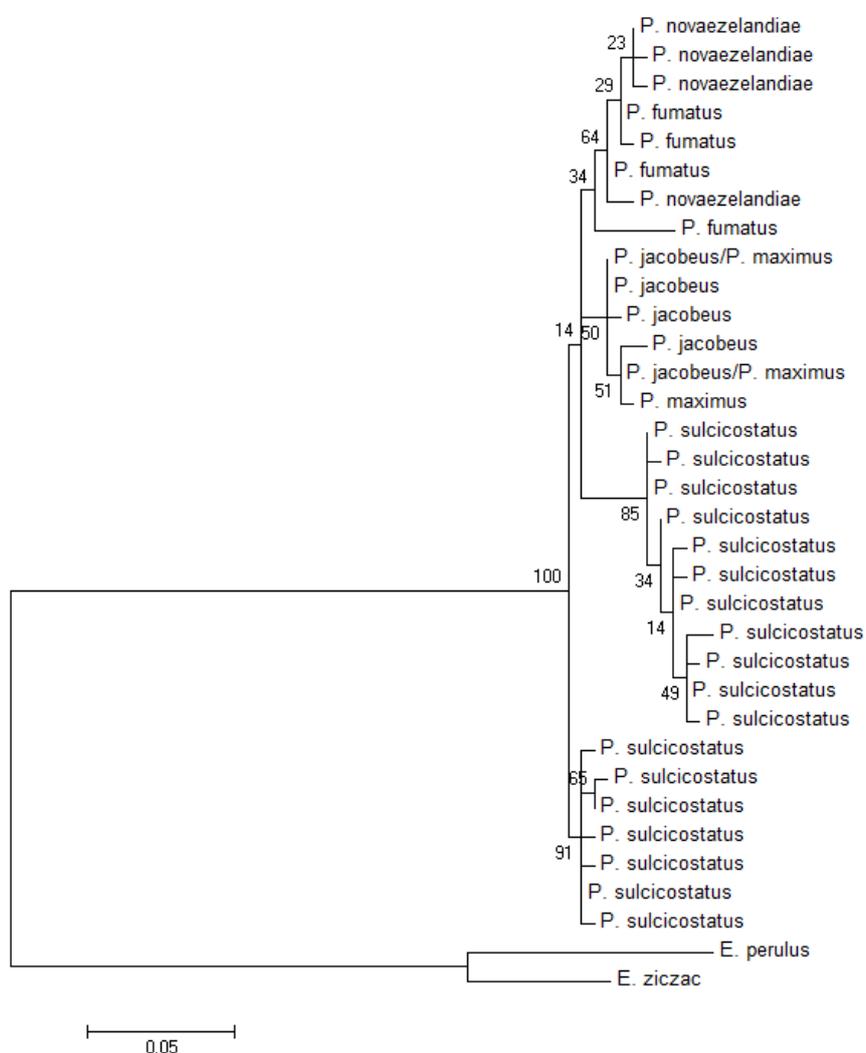


Figure 3.8: Maximum Likelihood tree of a 504 bp fragment of the 16S rRNA gene of haplotype sequences of five *Pecten* species with *Euvola perulus* and *Euvola ziczac* as outgroups. The “*P. jacobeus*/*P. maximus*” haplotype indicates a shared haplotype. The numbers above branches are bootstrap support values in %.

The Indo-Pacific area during the early Miocene (23.8 to 5.3 million years ago (MYA)) has been accepted as the *Pecten* diversification center (Waller 1991). From there, these scallops would have entered the Atlantic through the Tethys Sea, a water bridge formed by the separation of the Eurasian and African landmasses. The closure of this connection 20 to 14 MYA would have resulted in the Atlantic and Indo-Pacific *Pecten* lineages evolving in isolation ever since (Rögl 1998). *Pecten jacobaeus* fossils have been present in the Mediterranean region since the Pliocene (5.3 to 2.6 MYA) and *P. maximus* first appears in the fossil record in the Pleistocene (2.6 MYA to 11 700 years ago) (Waller 1991). *Pecten* first occurred in the Australian fossil record 1.8 MYA, and, in New Zealand, they appear “only a little more than 1 MYA” (Beu and Darragh 2001). Fleming (1957) proposed that *Pecten* scallops dispersed into Australia and New Zealand from the Red Sea and northern Indian Ocean. An alternative explanation, explaining the modest divergence observed between the Atlantic and Pacific *Pecten* lineages (Saavedra and Peña 2004), is that the two lineages separated later than commonly assumed. Beu and Darragh (2001) proposed that *Pecten* could have dispersed in the Australian region from the Atlantic and southern Africa, aided by the Antarctic Circum-Polar Current, and not from the northern Indian Ocean and the Red Sea, as suggested by Fleming (1957).

Despite abundant fossil records, the evolutionary relationship among members of the Pectinidae family is poorly understood. With current morphology-based systems, the attribution of species at both the genus and the subfamily level is not uniform, as different authors give different degrees of importance to morphological features. Scallop shell characteristics show great variability within a single species, while the shells of different species can display very similar characteristics. Therefore the use of molecular markers, such as a mitochondrial marker, is necessary to understand the evolutionary relationship among various scallop species. Although the deductions made in this study are based solely on the 16S rRNA gene and may therefore not be completely accurate, this study shows *P. sulcicostatus* to be quite removed from any of the other *Pecten* species.

In this study, the 16S rRNA gene was employed successfully in determining the relationship between various *Pecten* species. Although this study only made use of

the 16S rRNA gene to investigate the relationship between five *Pecten* species, this gene has been extensively used in phylogenetic studies (Barucca, *et al.* 2004; Saavedra and Peña 2004; Saavedra and Peña 2005; Marín *et al.* 2013; Marín *et al.* 2015). Future studies using more than one mitochondrial gene are necessary to understand the evolutionary relationship among these scallop species.

5. Conclusion

In conclusion, this study allowed for the development of a set of 10 microsatellite markers for the South African scallop, *P. sulcicostatus*, using cross-species transfer from *P. maximus*, a sister species to *P. sulcicostatus*, with a success rate of 50%. Microsatellite marker cross-species transfer provides insight into the genomic organisation of related species, allowing genomic information from a well characterised species (*P. maximus*) to be transferred to a genetically uncharacterised species such as *P. sulcicostatus*. The set of 10 microsatellite markers was applied to generate molecular marker data which was used to evaluate the extent of intrapopulation genetic diversity in the False Bay *P. sulcicostatus* population. A large range in the number of alleles was observed (2 to 36) for the set of microsatellite markers. Four of the 10 microsatellite markers showed the presence of null alleles, and three microsatellite markers deviated from Hardy-Weinberg expectations, likely due to the presence of null alleles. An average PIC of 0.505 indicates that the set of microsatellite markers could be useful in future studies. The high inclusion and exclusion probability values observed also strongly indicate that the set of microsatellite markers developed in this study would be effective in parentage assignment. The set of microsatellite markers will allow for further studies into the genetic composition of *P. sulcicostatus* populations along the South African coast and will also assist in genetic variability management and sustainable utilisation of the species. Selective breeding programmes implemented in aquaculture benefit from knowledge of the genetic composition of broodstock as well as their offspring. The set of microsatellite markers can be used to genotype future broodstock and their progeny in order to assess fertilisation success and maintain pedigree information for selective breeding (Selvamani *et al.* 2001). The set of

microsatellite markers can also be applied to the other South African scallop species, *P. aribenedictus*, which has no genetic resources available.

This study also allowed for the optimisation a mitochondrial marker, specifically a fragment of the 16S rRNA gene, which was successfully used to determine intra- and interspecific genetic diversity. *Pecten sulcicostatus* shows the highest haplotype diversity (0.9132) in comparison to that of other *Pecten* species, although this may be due to the current study's large sample size (38 individuals) in comparison with similar studies (Saavedra and Peña 2004; Saavedra and Peña 2005). The low nucleotide diversity (0.0094) observed for *P. sulcicostatus* indicates a low level of genetic variation, however it is still much higher than that observed for other species of *Pecten*. In the present study, the lower haplotype and nucleotide diversity observed for *P. maximus* and *P. novaezelandiae* and higher haplotype and nucleotide diversity observed for *P. jacobaeus*, may be due to differences in sampling between the two studies. Individuals obtained in this study were sampled from a number of populations (*P. maximus* - two locations, *P. jacobaeus* - two locations and *P. novaezelandiae* - four locations) whereas the study by Saavedra and Peña (2004) only made use of one wild (*P. jacobaeus* and *P. novaezelandiae*) or cultured (*P. maximus* and *P. fumatus*) sampling location. The 16S rRNA gene was also used to shed light on the evolutionary relationship between five *Pecten* species and corroborated previous studies which found that *P. maximus* and *Pecten jacobaeus*, which have traditionally been considered distinct species, most probably represent two variants of the same species.

6. References

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.
- Arif, I. A., M. A. Bakir, H. A. Khan, A. H. Al Farhan, A. A. Al Homaidan, A. H. Bahkali, M. Al Sadoon, and M. Shobrak, 2010 A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences* 11: 2079-2096.
- Arendse, D. C. Z., and G. C Pitcher, 2012 Growth and survival of the South African scallop *Pecten sulcicostatus* in suspended culture. *African Journal of Marine Science* 34: 181-186.
- Arendse, D. C. Z., S. Andersen, N. J. Blake, and G. C. Pitcher, 2008 The reproductive cycle of the scallop *Pecten sulcicostatus* from the Southern Benguela upwelling system. *Journal of Shellfish Research* 27: 281-287.
- Avise, J. C., 2000 *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA. Pages 3-441.
- Barbará, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer, 2007 Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759-3767.
- Barucca, M., E. Olmo, S. Schiaparelli, and A. Canapa, 2004 Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes. *Molecular Phylogenetics and Evolution* 31: 89-95.
- Bensasson, D., D. X. Zhang, D. L. Hartl, and G. M. Hewitt, 2001 Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* 16: 314-321.
- Beu, A., and T. A. Darragh, 2001 Revision of southern Australian Cenozoic fossil Pectinidae (Mollusca: Bivalvia). *Proceedings of the Royal Society of Victoria* 113: 1-205.
- Chambers G. K., and E. S. MacAvoy, 2000 Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology B* 126: 455-476.

Charrier, G., R. Morvezen, I. Calves, and J. Laroche, 2012 Development of new microsatellite markers derived from expressed sequence tags for the great scallop (*Pecten maximus*). Conservation Genetics Resources 4: 931-934.

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

D'Arcy, J., L. Mirimin, and R. FitzGerald, 2013 Phylogeographic structure of a protogynous hermaphrodite species, the ballan wrasse *Labrus bergylta*, in Ireland, Scotland, and Norway, using mitochondrial DNA sequence data. ICES Journal of Marine Science 70: 685-693.

Daniels, S. R., B. A. Stewart, and P. A. Cook, 2002 Congruent patterns of genetic variation in a burrowing freshwater crab revealed by allozymes and mtDNA sequence analysis. Hydrobiologia 468: 171-179.

Fleming, C. A., 1957 The genus *Pecten* in New Zealand. New Zealand Geological Survey Paleontological Bulletin 26: 1-69.

Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a review. Genetic Research 66: 95-107.

Gosling, E., 2003 Bivalve Molluscs: Biology, Ecology and Culture. Fishing News Books. Oxford, Blackwell Publishing: 249.

Guo, J. J., J. Zeng, S. L. Zhao, and Z. G. Zhao, 2008 Isolation and characterization of 19 microsatellite markers in a tropical and warm subtropical birch, *Betula alnoides*. Molecular Ecology Notes 8: 895-897.

Hebert, P. D. N., S. Ratnasingham, and J. R. de Waard, 2003 Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Biology Letters 270: 96-99.

Hoffman, J. I., and H. J. Nichols, 2011 A novel approach for mining polymorphic microsatellite markers *in silico*. PLoS ONE 6: 1-9.

Hold, N., L. Dawnay, and M. I. Taylor, 2013 Development of microsatellite markers from 454 transcriptome derived sequences for the scallop *Pecten maximus*. Conservation Genetics Resources 5: 663-666.

Ibarra, A. M., J. L. Petersen, T. R. Famula, and B. May, 2006 Characterization of 35 microsatellite loci in the Pacific lion-paw scallop (*Nodipecten subnodosus*) and their cross-species amplification in four other scallops of the Pectinidae family. Molecular Ecology Notes 6: 153-156.

Ivey, J. L., and S. R. Santos, 2007 The complete mitochondrial genome of the Hawaiian anchialine shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Decapoda: Atyidae). Gene 394: 35-44.

Kilburn, R. N., and H. H. Dijkstra, 1995 A new species of *Pecten* Muller, 1776, from South Africa, with a note on "*Pecten sulcicostatus* var. *casa*" van Bruggen, 1961. Annals of the Natal Museum 36: 271-279.

Kohlmann, K., P. Kersten, R. Panicz, D. Memis, and M. Flajs̃hans, 2010 Genetic variability and differentiation of wild and cultured tench populations inferred from microsatellite loci. Reviews in Fish Biology and Fisheries. 20: 279-288.

Lafarga-de la Cruz, F., A. Aguilar-Espinoza, and C. Gallardo-Escarate, 2015 Parentage assignment in hybrid abalones (*Haliotis rufescens* x *Haliotis discus hanna*) based on microsatellite DNA markers. Aquaculture Research 46: 216-225.

Leavitt, D., R. Karney, and A. Surier, 2010 Biology of the bay scallop. Northeastern Region Aquaculture Center Publication 213: 1-8.

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

Marín, A., T. Fujimoto, and K. Arai, 2012 Isolation and characterization of 12 microsatellite loci in the Peruvian scallop *Argopecten purpuratus* and cross-species amplification in other scallop species (family Pectinidae). Conservation Genetics Resources 4: 179-182.

Marín, A., T. Fujimoto, and K. Arai, 2013 Genetic structure of the Peruvian scallop *Argopecten purpuratus* inferred from mitochondrial and nuclear DNA variation. *Marine Genomics* 9: 1-8.

Marín, A., T. Fujimoto, and K. Arai, 2015 The mitochondrial genomes of *Pecten albicans* and *Pecten maximus* (Bivalvia: Pectinidae) reveal a novel gene arrangement with low genetic differentiation. *Biochemical Systematics and Ecology* 61: 208-217.

Mirimin L., N. Kitchin, D. N. Impson, P. F. Clark, J. Richard, S. R. Daniels, and R. Roodt-Wilding, 2015 Genetic and morphological characterization of freshwater shrimps (*Caridina africana* Kingsley, 1882) reveals the presence of alien shrimps in the Cape Floristic Region, South Africa. *Journal of Heredity* 106: 711-718.

Morvezen, R., F. Cornette, G. Charrier, B. Guinand, S. Lapègue, B. Pierre, and L. Jean, 2013 Multiplex PCR sets of novel microsatellite loci for the great scallop *Pecten maximus* and their application in parentage assignment. *Aquatic Living Resources* 26: 207-213.

Norris, A. T., D. G. Bradley, and E. P. Cunningham, 1999 Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture* 180: 247-264.

Peakall, R., and P. E. Smouse, 2006 GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Peakall, R., and P. E. Smouse, 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics*. 28: 2537-2539.

Renon, P., M. M. Colombo, F. Colombo, R. Malandra, and P. A. Biondi, 2001 Computer-assisted evaluation of isoelectric focusing patterns in electrophoretic gels: Identification of smooth hounds (*Mustelus mustelus*, *Mustelus asterias*) and comparison with lower value shark species. *Electrophoresis* 22: 1534-1538.

Rögl, F., 1998 Palaeogeographic considerations for Mediterranean and Paratethys seaways (Oligocene to Miocene). *Annalen des Naturhistorischen Museums in Wien* 99A: 279-310.

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

Saavedra, C., and J. B. Peña, 2004 Phylogenetic relationships of commercial European and Australasian king scallops (*Pecten* spp.) based on partial 16S ribosomal RNA gene sequences. *Aquaculture* 235: 153-166.

Saavedra, C., and J. B. Peña, 2005 Nucleotide diversity and Pleistocene population expansion in Atlantic and Mediterranean scallops (*Pecten maximus* and *P. jacobaeus*) as revealed by the mitochondrial 16S ribosomal RNA gene. *Journal of Experimental Marine Biology and Ecology* 323: 138-150.

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

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

Silva, C. N. S., and J. P. A. Gardner, 2014 Development and characterisation of 12 microsatellite markers for the New Zealand endemic scallop *Pecten novaezelandiae*. *Conservation Genetic Resources* 6: 327-328

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013 MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

von Rintelen, K., T. von Rintelen, M. Meixner, C. Lüter, Y. Cai, and M. Glaubrecht, 2007 Freshwater shrimp–sponge association from an ancient lake. *Biology Letters* 3: 262-264.

Waller, T. R., 1991 Evolutionary relationships among commercial scallops (Mollusca: Bivalvia: Pectinidae). In: Shumway, S. E., and G. J. Parsons (eds.), *Scallops: Biology, Ecology and Aquaculture*. Developments in Aquaculture and Fisheries Science. Amsterdam, Elsevier: 35: 1-73.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N Hebert, 2005 DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* 360: 1847-1857.

Watts, P. C., P. J. Mallanaphy, C. McCarthy, B. D. Beukers-Stewart, M. W. J. Mosley, A. R. Brand, and I. J. Saccheri, 2005 Polymorphic microsatellite loci isolated from the great scallop, *Pecten maximus* (Bivalvia: Pectinidae). *Molecular Ecology Notes* 5: 902-904.

Yuan, Y., M. Kanno, and A. Kijima, 2012 Genetic diversity of wild populations of *Chlamys farreri* in Japan and their genetic relationship with cultured stocks in China. *Aquaculture* 370-371: 109-122.

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

Chapter IV

Marker Discovery through Next Generation Sequencing

1. Abstract

To date there are no microsatellite markers or single nucleotide polymorphisms (SNPs) available for the South African scallop, *Pecten sulcicostatus*. As Next Generation Sequencing (NGS) technology allows for the identification of thousands of potential molecular markers, this study aimed to investigate the feasibility of using NGS technology for marker discovery in this endemic species. This study provided the first reduced genome sequences for *Pecten sulcicostatus*, with over 7.3 million reads. Quality based trimming and contig construction through *de novo* assembly using two bioinformatic approaches aided in the identification of 55 putative microsatellite markers as well as 2 530 putative SNPs. The two bioinformatic approaches were equally successful with regards to microsatellite marker discovery, however, approach two was successful in regard to SNP detection, identifying 2 530 putative SNPs, while approach one failed to detect any SNPs. Once validated, the microsatellite markers and SNPs identified in this study could be used in various downstream applications such as genetic diversity assessment, population structure inference, genetic linkage studies as well as marker assisted selection (MAS).

2. Introduction

In 2005 the development of Illumina's Genome Analyzer allowed for a short read, massively parallel sequencing technique which was a different approach to sequencing that revolutionised sequencing capabilities and allowed a single sequencing run to produce 1 GigaByte (GB) of data. By 2014, the rate climbed to 1.8 TeraByte (TB) of data in a single sequencing run. With the advancement and reduction in cost of Next Generation Sequencing (NGS) technologies it is now possible to identify thousands of molecular markers using NGS platforms (Etter *et al.* 2011) such as Roche 454, Illumina MiSeq and HiSeq as well as Ion Torrent and Proton platforms (Wei *et al.* 2014).

Traditionally, microsatellite marker development involved making use of enrichment protocols, however current use of NGS platforms has several advantages over enrichment methods when it comes to identifying microsatellite markers. Enrichment methods only target a few motifs, however these motifs can differ largely amongst different taxa. With NGS, no prior motif selection is necessary. Microsatellite marker development using NGS platforms is more cost effective, faster and yields more microsatellite markers than conventional enrichment protocols (Gardner *et al.* 2011).

Since the onset of DNA sequencing it has been possible to identify single nucleotide polymorphisms (SNPs), however only following the development of gene chip technology did genotyping of these markers become feasible (Lui and Cordes 2004). Single Nucleotide Polymorphism (SNP) identification makes use of methods such as expressed sequence tag (EST) analysis, random shotgun sequencing as well as polymerase chain reaction (PCR) amplicon sequencing. For large scale SNP discovery, methods such as pyro-sequencing technology were used (Ahmadian *et al.* 2000; Alderborn *et al.* 2000). For large, complex genomes however, pyro-sequencing proved to be too expensive and time consuming, therefore reducing the complexity of the genome before sequencing became a more viable and attractive option (Mascher *et al.* 2013).

Genome reduction can be performed using restriction enzymes (REs) which fragment DNA into smaller segments. These segments can then be pooled and sequenced in parallel in order to capture a targeted segment of the genome whilst still maintaining sufficient coverage (Mascher *et al.* 2013). In order to identify SNPs,

contigs must be constructed and mapped against a reference genome (Everett *et al.* 2011; Nielsen *et al.* 2011), however for non-model organisms lacking a reference genome, genome segments of multiple individuals can be screened in order to identify SNPs (Morin *et al.* 2004).

Reduced Representation Sequencing (RRS) can be utilised as an alternative method to whole genome sequencing by focusing on compact panels of genomic markers throughout the genome. Genome sub-sampling can be attained by using Restriction-site Associated DNA (RAD) sequencing, which creates short DNA fragments which are adjacent to recognition sites of specific REs. RAD sequencing is useful when studying organisms lacking a reference genome as it allows for simultaneous sequencing and genotyping, thereby accelerating SNP discovery while at the same time allowing for the genotyping of hundreds or thousands of markers throughout the genome. This is accomplished by focusing on and sequencing a subset of genomic regions from several individuals (Baird *et al.* 2008; Etter *et al.* 2011; Davey *et al.* 2013). SNPs are detected by creating a reduced version of the genome, which allows the nucleotides next to the RE sites to be over-sequenced. RAD sequencing can be used to genotype populations that have been pooled as well as a multiplex of individuals (Baird *et al.* 2008). RAD markers were initially used in conjunction with low cost microarray genotyping resources (Miller *et al.* 2007), but the advent of massively parallel NGS technologies and drop in sequencing costs lead to the integration of short-read sequencing with RAD genotyping (Baird *et al.* 2008). In the case of RAD sequencing, ligation of sequencing adapters to restriction digested DNA prior to shearing focuses sequencing effort on tagged restriction sites, rather than randomly sequencing throughout the genome (Rowe *et al.* 2011). This approach dramatically increases the coverage for a given sequenced site, increasing both the confidence in base identity and the likelihood that the same sites will be sequenced in multiple samples (Barchi *et al.* 2011).

To date no species-specific microsatellite markers or SNPs are available for the South African scallop, *Pecten sulcicostatus*. This study therefore aimed to investigate the feasibility of using NGS technology for marker identification, in particular microsatellites and SNPs. Once identified, these markers could form the base of genetic assessments for this species in future studies.

3. Materials and Methods

3.1 Biological Specimens, DNA Extraction and Next Generation Sequencing (NGS)

Genomic DNA (gDNA) was extracted from eight *P. sulciostatus* individuals, standardised to 500 ng/μl and pooled in equimolar concentrations. Reduced genomic library preparation and sequencing was performed by Inqaba Biotechnical Industries (Hatfield, Pretoria). The protocol, in brief, was as follows: 1 μg of DNA was digested using the *Msp1* restriction enzyme (RE). Digested DNA was then size selected for fragments ranging from 100 to 200 bp. This fragment pool then underwent NGS library preparation which involved end repair, adapter ligation and indexing. Sequencing was then performed using a MiSeq version 3 (150 cycle kit), producing 150 bp long single end reads on Illumina's MiSeq platform.

3.2 Processing of Next-Generation Sequencing (NGS) Data

The sequencing data was obtained in fastq file format (text-based files which store both nucleotide sequences and their quality scores). Data was analysed using two different bioinformatics approaches, as described below. Approach one involved uploading the fastq file to the Stellenbosch University High-Performance Computing (HPC) cluster for processing. The data file was then run through FastQC version 3.0 (Patel and Jain 2012) using default parameters in order to determine the quality of the sequence data. Sequencing data is annotated with quality scores, known as phred scores, which indicate the probability of base-calling errors. Low quality reads (reads containing a large number of bases with low phred scores) often lead to misassemblies or ambiguous contigs and are therefore trimmed or removed. Trimmomatic version 0.33 (Bolger *et al.* 2014) was used to filter out poor quality reads as well as remove sections of the reads that fell below a certain quality (phred score = 15). Base quality decreases towards the end of each read, therefore, in this study, 3 bps were trimmed from both the leading as well as the trailing reads and reads with a length below 35 bp were discarded. Data was then reanalysed using FastQC in order to determine the sequence quality after trimming. Velvet version 1.2.10 (Zerbino and Birney 2008) was used to construct contigs from the trimmed reads using default parameters and a 100 bp minimum contig length.

Approach two made use of CLC Genomics Workbench version 8.5 to determine sequence quality and trim the original data. Trimming of the original data involved filtering out poor quality reads (error probability > 0.05; phred score > 13) as well as removing sections of the reads that fell below a certain quality (error probability = 0.05; phred score = 13). As base quality decreases towards the end of each read, 3 bps were trimmed from both the 5' and 3' end of all of the reads. Reads with a length below 35 bp were also discarded. CLC Genomics Workbench was then used to perform a *de novo* assembly in order to generate contigs. *De novo* assembly was performed using default parameters with a minimum contig length of 100 bp.

3.3 Identification of Putative Microsatellite Markers and Single Nucleotide Polymorphisms (SNPs)

Following contig construction using the two approaches, perfect repeat microsatellite markers with more than five repeats were identified using Simple Sequence Repeat Identification Tool (SSRIT) (Temnykh *et al.* 2001). Once candidate microsatellite markers were identified, contigs with putative microsatellite markers towards the middle of the contig (more than 20 bp on either side of the repeat) were manually selected in Microsoft Excel (Microsoft 2010) and a Basic Local Alignment Search Tool (BLAST) search was performed on the publicly available databases GenBank, EMBL and DDBJ (Altschul *et al.* 1990).

The minimum coverage, minimum minor allele frequency (MAF) and minimum count greatly influences the number of SNPs detected as well as the confidence that the SNPs detected are not as a result of sequencing error. Studies use varying minimum parameters, ranging from 20 to 100 times coverage, MAFs ranging from 2% to 20% and minimum counts ranging from two to 100 (Franchini *et al.* 2011; Lal *et al.* 2015). Following contig construction using the two approaches, SNP discovery was performed using CLC Genomics Workbench's Fixed Ploidy Variant Detection tool with a required variant probability of 90%, minimum coverage of 100, minimum count of 100 and MAF larger than 10%. The Fixed Ploidy Variant Detection tool calculates the probabilities of each of the site types. One of the site types is the site that is homozygous for the reference - it stipulates that whatever differences are observed from the reference nucleotide, it is due to sequencing error. The remaining site-types are those which stipulate that at least one of the alleles in the sample is different

from the reference. The sum of the probabilities for these latter site types is the posterior probability that the sample contains at least one allele that differs from the reference at this site. This posterior probability is known as the variant probability, which defines how good the evidence has to be at a particular site for the Fixed Ploidy Variant Detection tool to report a variant at that location. If the site passes this threshold, then the variant with the highest probability at that site is reported. The Fixed Ploidy Variant Detection tool is therefore stringent, removing variants that likely originate from mapping or sequencing errors.

4. Results and Discussion

4.1 Approach 1

The original fastq file contained 7 387 243 reads with lengths between 35 and 151 bp. Following trimming using Trimmomatic, the fastq file contained 6 483 851 reads with lengths between 36 and 151 bp. Through *de novo* assembly using Velvet, 1 752 contigs were generated with an average length of 124.45 bp.

A total of 39 microsatellite markers were identified using this approach, of which 28 were useable (Table 4.1). Eleven microsatellite markers were unsuitable as the repeat was either at the beginning or the end of the contig, therefore primers could not be designed to amplify the repeat. For approach one, eight contigs returned BLAST results for marine species and one contig returned a BLAST result for a scallop species, namely *Chlamys farreri*. Four contigs showed no similarity to any sequences on the publically available databases.

Table 4.1: Microsatellite markers identified from the contigs constructed using approach one. The repeat is shown in red.

Microsatellite marker	Repeat	Sequence (5' → 3')	BLAST result
Psul1	(AT) ₈	TTTCCATTTCCATAGAATTATGGCTATGAAATTGATCATTACATA C ATATATATATATATAT TTAAAATGTCTTAACTTGAAATAAATCG TAAAAGCAAAGTCTGAAAAACAATCAAAAACAAAAAAA	<i>Apteryx australis mantelli</i> North Island brown kiwi
Psul2	(AC) ₅	ATGTATAGACACATAATGTGACCCCATTCCTTACCCACACTT TATACACTG ACACACACAC GGTAGCTTATAGGTCAAGAACGAA AGAGCAAACGTCTGAACTCCA	<i>Cyprinus carpio</i> Common carp
Psul3	(AT) ₁₀	CGATGGGCAGTATACACTGATACATGTAGGCTTAGTTTGTGTC CAATCCATTC ATATATATATATATATATAT AATATTTATGTAATTC TCTGTATGAAATATTTGAA	<i>Trichobilharzia regent</i> Bird schistosome
Psul4	(TA) ₅	ATAACTGTATGCAAATTTGTGAATAATTCCATATCTATATC TATAT ATATA CCGTTTCGTGTATATATTTACACATATCAAATTGGAATGA CACAGGAGATTTGCAAGTTTGATATGTGTAATATATACTG	<i>Toxocara canis</i> Dog roundworm
Psul5	(TA) ₆	AGTATATATATATAGCTCAGGGGAGTCTTCAACTCAAGAGATAA ACAATCCACAAAAAGTCTGAGCT TATATATATATA CTAGAACGGA AGAACACACGTCTGA	<i>Cyprinus carpio</i> Common carp
Psul6	(TA) ₅	GTCATGTCAAACCTAGATAGCT TATATATATA AACAGTCTTTGATT CAAATCATCATCACGGTCAGTGGCTAATCAAAAATAGAACTGTC GCCAGGATGGCTGA	<i>Trichobilharzia regent</i> Bird schistosome
Psul7	(TA) ₇	TAACATTTATACATAGAGCCCGGTATATGCTTTATCGTTACCCCC TCTTACTGATAACCGT TATATATATATATA CTATAACATTTATAC ATAGAGCCCGGT	<i>Cyprinus carpio</i> Common carp
Psul8	(AC) ₈	GGCCTATCGCCATGAGACATACCAAAAACATACCAAAGTCTCCC AAA ACACACACACACACAC TTACCACACGCATGCATATCGCAC GCACGGGAGGCCGTCCTT	<i>Chlamys farreri</i> Chinese scallop
Psul9	(CA) ₅	ACAGGGTGTACAGAATATAGT CACACACACA AATGGATCGGGTG TACAGAATATAATCACACAAACAATGGAAAGGGTGGAACAGAAA AGCAAACGTCTGAAATCCAG	<i>Ovis canadensis</i> Bighorn sheep

Psul10	(AC) ₆	GACACACTCACACAGAGATACATACACACAGAGACAACCTCACA CAGAGACACACTAACACAGAG ACACACACACAC AGAGACA ACTACACACAGAGACAACCACACACAAAGACAAACACACACAAAA AA	No similarity
Psul11	(ACC) ₅	GTGGCGCCCGCCACAGTTAAAGCAACCATCACCACGACCACGA CCACCACCACCACC ACTTGAACCTTTGGCCATCATTATTAGGGG AAGAGCACACGTCTGAACTC	<i>Spirometra erinaceieuropaei</i> Tapeworm
Psul12	(AT) ₅	ATACATATATCAACACAATTATTTCTGTTTTTAATTCAGATGTCTA TTTAATAATTATATATACTATCTATATTTAAATACAG ATATATAT AT CTAATAACATTAATACATCTATCTGTATTTCTAATTTAATGT	<i>Danio rerio</i> Zebrafish
Psul13	(TG) ₉	GTGTGTCCCCAGTGTTGTCTGTGTGTGTGTGTCCCCAGCGTTGTC TGTGTGTGTGTGTGTGTGT TCCCCAGTGTTGCCTGTGTGTAAGTC CCCAGTGTTGTGTAGGGAAAGAGTGTC	No similarity
Psul14	(TA) ₅	CATGTAGGGGTATATATGTAGTGAAATACCCGTCAAACACATGT ATGGG TATATATATA GTGATATTCACGTCACCCACATGTAGGGG TATATATATAGTGATATTC	<i>Fragaria vesca</i> Wild strawberry
Psul15	(AC) ₅	AACACACACACTCAATCTAGTCTACCAAACA ACACACACACA AAT CCAGTCTACCAAAACAACACCCACACACAATCCAATCTACCAAA CAACTCACACTCAAAAAAACTACCAAACAACACACAAAA AAAA	No similarity
Psul16	(TC) ₅	GCTCTTCCGATCCACTTGTTTAATATCTATTGGATGGATTTAGTT TTGAGCAGATTTGATACTTGTTATTAATGGGTTTTTAT TCTCTCT CTC TATCTTTTTTTGTTTTTTGTTTT	<i>Cyprinus carpio</i> Common carp
Psul17	(AT) ₅	CGCCACATTGGCTGTTTCACTGGCGC ATATATATAT CTTGTATA GGTTTTTTGAGCTGTTGTATATTTATTATATTAACATGCACACGT CTGAAATCCAGTCA	<i>Nippostrongylus brasiliensis</i> Roundworm
Psul18	(TA) ₆	ACGACATCTAACGACAAGAACAACATACGATATATGTTTGTTTTA CGAAAGGAGGAGTGCCATTTTTGTCTTCTGGTGGCTATCAATT GTACGTTTTATTATATCT TATATATATATA AAAGAGCACACGTCT GAACTCC	<i>Plasmodium falciparum</i> Protozoan parasite

Psul19	(TA) ₆	AACTAAATGTTAATGATTGATTGCCTATAATGTTTGAGAAGTTCA AAGTCCTTGTTTTCCCTTACAGATGAGTTACAACAAGGCTAT GG TATATATATATA AAAAATATAAAAAAAAAAACACAAA	<i>Solanum lycopersicum</i> Tomato
Psul20	(AG) ₇	GAGAGAGAGAGAGAGAGAGAGAT AGAGAGAGAGAGAG TGA TAGAGAGAAAGAGAGAAAGAGAAATAGAGAGAGAGAGAGAGAG AAAAAAAAAAATATAAAAAAAC	No similarity
Psul21	(AG) ₉	GAGAGAGAGAGAGAGAGAGAGATAGAGAGAGAGAGAGTGA TAGAGAGAAAGAGAGAAAGAGAAAT AGAGAGAGAGAGAGAGA G AAAAAAAAAAATATAAAAAAAC	
Psul22	(TA) ₇	CACATAATAGCCGAGAACCCAATATATGTTATCAGTGATAAGAA ACAGTACTTATAAGCATTGTTGGCAATTTATCCTGTAACAGCTAT AGCCTTCCATATTTGACATCCCACAGATTGTTTCTTT TATATATA TATATA GTTTTAAAGGTCAACTCGGAAGAGCGTTCGTGTAGGGA	<i>Cyprinus carpio</i> Common carp
Psul23	(TA) ₅	TTGCTTACAACCTTACCACCAGCTACCGCTTACG TATATATATA CTTGATACATATAAGCTGTGTATATCATAAGAGTAATAATTAATA ATACAGAACGGAAGAACACACGTC	<i>Cucumis melo</i> Muskmelon
Psul24	(AC) ₆	ATAGAGCCTTTTTCCCTTTGAAATTGCGTAGATCTACATTTTCAG CTACAGTTACA ACACACACACAC ATCTATAGATATCTTAATTA CTCATGTACAATGTATGAATTAAGATAGATAGGAAAAGCACAAG TCTAAACTCCA	<i>Cyprinus carpio</i> Common carp
Psul25	(TAA) ₅	AAATTGCACCAAGAACGACAACCTTGATATTGTCTATCATGTACTA AGATATGAACGTTCTGTCT TAATAATAATAATA AATAAAGTTATC GGAAGAGCACACGACTGAA	<i>Alocasia macrorrhizos</i> Elephant ear taro
Psul26	(TG) ₅	CGACCATTAACCTAACAGCCTGGGAGAATCAGGTAAAT TGTGTGT GTG TCTTAATACAAAGAACGGTTTGATCCTCGAAAAAGTTAAAC GGGATGAGCACAAATCGGAAGAGAACAAGTATGAAATCCAGTC ACAGT	<i>Cyprinus carpio</i> Common carp
Psul27	(CAA) ₅	AACAACAGAAATGTGTACATCACCGACAACACCAACAATAACGA CAGAAATAAGTACACCAC CAACAACAACAACA TAACAACAGA AATGTGTACATCACC	<i>Plasmodium reichenowi</i> Parasite

Psul28	(ATG) ₅	AGAACATGCGACGACACCCGATGTTTGACATGGACGAAGATGA CGACGAAATGCAT ATGATGATGATGATG AGGGACATGAGACCG GACAGTAAGTCGAATTTCCACGAGCACACGTCTGAACTCCAGT	<i>Echinostoma caproni</i> Intestinal fluke
--------	--------------------	--	---

Analysis of the 1 752 contigs generated through approach one detected no putative SNPs, most likely due to the low number of contigs generated using this approach in combination with the strict SNP detection parameters used in this study (minimum coverage of 100 and MAF larger than 10%).

4.2 Approach 2

The original fastq file contained 7 387 243 reads with an average length of 136.2 bp. Following trimming, the fastq file contained 6 503 842 reads with an average length of 141.5 bp. Through *de novo* assembly, 18 958 contigs were generated with a length ranging from 18 to 5 374 bp and an average length of 119 bp. The N50 for the *de novo* assembly was 117. CLC Genomics Workbench's *de novo* assembly makes use of the de Bruijn graph algorithm and therefore generates a large number of unique contigs (CLC Bio 2012) as seen in the current study in which CLC Genomics Workbench generated 18 958 contigs.

A total of 47 microsatellite markers were identified from the contigs generated through approach two, of which 28 were useable (Table 4.2). Nineteen microsatellite markers were unsuitable as the repeat was either at the beginning or the end of the contig, therefore primers could not be designed to amplify the repeat. For approach two, 13 contigs returned BLAST results for marine species and one contig returned a BLAST result for a scallop species, namely *Chlamys farreri*. Only one contig showed no similarity to any sequences on the publically available databases.

Table 4.2: Microsatellite markers identified from the contigs constructed using approach two. The repeat is shown in red.

Microsatellite marker	Repeat	Sequence (5' → 3')	BLAST result
Psul29	(AAC) ₅	GAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTA TGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAAAT AACAACAAC AACAAC AAAAAAAAACAAAATTAAACAAAACAAAAAAAAACAAAA AACAAACATAATA	<i>Conus episcopatus</i> Dignified cone
Psul30	(AC) ₅	ACTCCAGTCCCAGTCAACAATATCGTATGCCGTCTTCTGCCTGA AAAAAAAAAAAAAAAAACAAAAAAAAAAT ACACACACAC AAAAAAAA AAAAAAAAACACAAAAACAAACAAACACAAAAAAAAAAAA	<i>Mus musculus</i> House mouse
Psul31	(AG) ₅	CAGTCACAGCCAACAACACTCGTATGCCGTCTTCTGCTTGCAAAA AAAAAAAAAAGACAAGAAAACATACAAT AGAGAGAGAG ACAATA ATAAAAGATATTAACACTACAAAAAAATAAAAT	<i>Cyprinus carpio</i> Common carp
Psul32	(AC) ₅	TCCAGTGACAGTCAACAATCTCGTATGCCGTCTTCTGCTTGAAA AAAAAAAAAAAAAAAAAAAAAAAAACAAAACACAAGAACAAAAAAAA ACAAAAAA ACACACACAC AAAAACAAAAAAAAAAAAACAAA	<i>Pongo abelii</i> Sumatran orangutan
Psul33	(AC) ₅	CAACAACCTCGTATGCCGCCTTCTGCTTGAAAAAAAAAAAAAAAA AAAAAAAAACAAAAAAACAATACAACAAAAAAAAAAAA ACACACAC ACA ACAAAAACAAACAAAAACAAA	<i>Hordeum vulgare</i> Barley
Psul34	(AT) ₅	ACTCCAGTTACAGTCAACAATATCGTATGCCGTCTTCTGCTTGA AAAAAAAAAAAAAAAAAAAAAAAAAATCTAC ATATATATAT AGTATAAAA AAAAAAAAAAAAATAACAAACCCA	<i>Cyprinus carpio</i> Common carp
Psul35	(AC) ₅	GTCACAGTCAACAATTCGTATGCCGTCTTCCGCTTGAAAAAAA AAAAAAAAAAAAAAAA ACACACACAC AAAAAAAAACAAAAAAAC AAAAACAACAAA	<i>Zea mays</i> Maize
Psul36	(AC) ₅	GAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTA TGCCGTCTTATGCTTGAAAAAAAAAAAAAAAAACAAAAAA ACACACACA C ATAAAAAACAAAAACAAACACAAAAAAACAACAAAAAA AAAAAAAAAAAAACAAAATA	<i>Conus episcopatus</i> Dignified cone

Psul37	(TG) ₅	AACAACAGATACTAGAAAGTGTGTGATATTACAGTAAACAACAG ATACTT TGTGTGTGTG ATATTACAGGAAACAACAGATACTTGTGT GTGTGATATTACAGGA	<i>Ovis Canadensis</i> Bighorn sheep
Psul38	(CAA) ₅	ATAACAACAGAAATGTGTACATCACCGACAACACCAACAATAAC GACAGAAATAAGTACACCAC CAACAACAACAACA TAACAACA GAAATGTGTACATCACCAAAAACACCAACAATAACAAAAAATA A	<i>Cucumis melo</i> Muskmelon
Psul39	(AC) ₅	GAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTA TGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAATAAACATAAAAA ACAC ACACAC AACACGAAAAAAAAAAAAAAAAATCAAAGAAAACAAAACA	<i>Conus episcopatus</i> Dignified cone
Psul40	(ACA) ₅	GTATGCCGTATTCTGCTAGAAAAAAAAAAAAAAAA ACAACAACAA CAACAAT CACAAAAAAGACAACAGCAAAAAACAAAAAAAAAAAA ATAATAAAAAAA	<i>Haliotis midae</i> South African abalone
Psul41	(AAC) ₅	AGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATG CCGTATTCTGCTTGAAAAAACACAAAACAAAAAAAAAAAAAAAA TA AACAACAACAACAAC AAAAACAACCAAAAAAAAAAAAAAAAA AAACAACATAAAACAACAAAAGAACATAA	<i>Conus episcopatus</i> Dignified cone
Psul42	(AAC) ₅	TCAACAATTTTCGTATGCCGACTTCTGCTTGAAAAAAAAAAAAAAAA ACAAA AACAACAACAACAAC ATTAATAAACAAAAAAAAAAC ACAATTAACAATCAACAAAA	<i>Botrytis cinerea</i> Necrotrophic fungus
Psul43	(AAC) ₅	ACAACTCGTATGCAGTCTTCTGCTTGAAAAAAAAAAAAAAAACAAC AAAAAAAAAAAA AACAACAACAACAAC AAACAACAAAAAAAAAAAA AAAAGAAACAAAACAACAAGAAAA	<i>Oryza sativa</i> Asian rice
Psul44	(AC) ₅	AAAAAAAAAACAACAATATTACGAACATTAAC ACACACAC ACCAACAATAACACACTACATAATAAACAAAA	No similarity
Psul45	(AC) ₆	CTCGTATTCCGTCTTCTGCTTAAAAAAAAAAAAACATAAAATATACA AT ACACACACACAC AAAAATACTAAAAACAAAAAAAAAATAAAA CAATAAACAAAATAAAA	<i>Pongo abelii</i> Sumatran orangutan
Psul46	(ACA) ₆	CAACAACTCGTATGACGTCATCTGCTAGAAAAAAAAAAAAACAAAA AAAAAAAAAAAAATAAACAAAAAAG CAACAACAACAACAAC AA AGATAAAAAACAAACAAAACAA	<i>Haliotis midae</i> South African abalone

Psul47	(TA) ₅	TTATTTTTTTTTTATTGTGACTGGTTTTCTTTTTTTTTTTTTTTTT TTCTGTTCTAAAGAGGTTTTTGTAGTTGTAGAGTGTACACTCTT TCCCTACACGTGCTCTTCCGATCT TATATATATA GTATATTTAA GATCGGAAGAGCAAAAGTATGAACTCCAGTCACAGTCAACAAA CTAGTATGACGTATTCTGCTTTAAAAAAAAAAAAAAAAACAAA AAAAAAAAAAAAAAAAACAAAAAAAAATAAAATAAAAAAAAAATAAGAA	Danio rerio Zebrafish
Psul48	(AC) ₆	ACAAACTCGTATGCCGTATTCTGCTTGAAAAAAAAAAAAAAAAACA AAAAAGAGCATA ACACACACACAC AAATATAAACACAACAAAC AAAAAGCAAAAAAA	Cyprinus carpio Common carp
Psul49	(CA) ₅	AAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTA TGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAAAAAAAAA CA CACACACA CCAACCAAATAAACAAAAAAAAAAAAAAAAACAAA AAAAAAAAAACACACCAC	Conus episcopatus Dignified cone
Psul50	(TG) ₅	TTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATTCCC GGTTCGATTTCTGGTGGTTATTTTGGTAGTTTTGTTGTCCCTAG CCCGGCCCACTGTCC TGTGTGTG TTTACTCAATTATATAC AAAAATGCTCTTGTCTTAGGTTTTTTTTAAGATAAAGGCGAGAAA TAAGGGGGGTGGGAGATAGCATGGAGTTATCTGCCCTTGAT CTTCAGGGTATTTCTTATTTGTTTATTTTTTTCAGTAG	Conus episcopatus Dignified cone
Psul51	(CA) ₅	GTATTCTGATTGAAAAAAAAACAAAAAAAAAACAGAATAAAAA AAA CACACACACA AAAAAAAAAATACAAAAAATATAAAAA AAACAAAA	Zea mays Maize
Psul52	(AT) ₅	GTCACAGTCAACAAATTCGTATGCCGTCTTCTGCTTGAAAA AAAAAAAAAGGAA ATATATATAT AAACAACAAAAAGACAGCAC AAAACAAAAAATAAAAAAGAAGACA	Zea mays Maize
Psul53	(AC) ₅	ACAATCTCGAATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAA AAAAAAAAACCCCAACAAAAAAAAACAA ACACACACAC AACCA CACAACAAACAAAAAA	Angiostrongylus cantonensis Parasitic nematode
Psul54	(CA) ₆	AAAAAACAAAAAACAAAAAAC CACACACACACA AAAAAA AAGACCAAATAAACCAAAAAAAAAAAGAATACA	Cyprinus carpio Common carp
Psul55	(CGTC) ₈	CACATTTTTAGCCCACCATCATCAGATGGTGGGCTATTCAAT CGCCCTGCGTCCGTGGTCCGT CGTCCGTCCGTCCGTCCGTCC	Chlamys farreri Chinese scallop

		<p>GTCCGTCCGTCCGTAAACAATTCTTGTTATCGCTATTTCTCAGA AAGTACTGAAGGGATCTTTCTCAAATTTTCATATGTAGGTTCCCC TTGGTGCCTAGTTATGCATATTGCATTTTGAGACCAATCGGAAA ACAACATGGCCGACAGG</p>	
Psul56	(AC) ₅	<p>AACAATATCGTATGCCGTATTCTGCTTGAAAAAAAAAAAAAGACA CACACACAACATCAAACGAAACAGAAAAAAAAAAAAAAAAAAC AAAGAAAAAAAAAATAAACAGAAAAA</p>	<p><i>Pinellia ternate</i> Crow-dipper</p>

A total of 2 759 variants were detected from the contigs generated through approach two, of which 2 530 were putative Single Nucleotide Variants (SNVs) also known as putative SNPs (Table 4.3). A total of 606 (21.96%) transitions and 1 924 (69.74%) transversions were observed, giving an observed transition to transversion (ts:tv) ratio of 0.31 (Table 4.4). Interestingly, 28 Multiple Nucleotide Variants (MNVs) were also identified. Multiple Nucleotide Variants are neighboring SNVs where there is evidence that they occur together.

Table 4.3: Summary of the variants detected following *de novo* assembly of the 18 958 contigs generated using approach two.

Variant	Number
Single Nucleotide Variants (SNV)	2 530
Multiple Nucleotide Variants (MNV)	28
Deletion	192
Insertion	6
Replacement	3
Total	2 759

Table 4.4: Summary of the putative single nucleotide polymorphisms (SNPs) detected.

SNP type	Total	Percentage (%)	
Transition	T/C	5	0,20
	A/G	158	6,25
	C/T	4	0,16
	G/A	439	17,35
Transversion	T/A	978	38,66
	T/G	4	0,16
	A/T	113	4,47
	A/C	44	1,74
	C/A	749	29,60
	C/G	2	0,08
	G/T	11	0,43
	G/C	23	0,91
Total	2 530	100	

Although the transition to transversion (ts:tv) ratio of 0.31 is low, six of the eight fourfold degenerate groups have a G or C at the second position. As a result, neighbouring base composition will be highly biased toward one or more G or C (Etter *et al.* 2011). Therefore, the low (ts:tv) ratio could be a result of this bias. One

possibility is that misincorporation and/or proofreading by the DNA polymerase is influenced by the nucleotide 5' to the site of incorporation. Sites with a G and/or C both 5' and 3' would be biased to transitions during polymerisation using either strand as a template. The second possibility is that repair efficiency is influenced by neighbouring base composition. One consequence of the neighbouring base influence is that regions with a higher A+T content have a higher proportion of transversions than regions with lower A+T content (Baird *et al.* 2008).

4.3 Comparison of Bioinformatic Approaches

Approach two generated over 10 times more contigs than approach one, despite using the same basic parameters (trimming of 3 bps from both the leading and trailing reads and discarding of reads with a length below 35 bp). Approach one used a slightly higher phred score (15) than approach two (13) when trimming reads based on quality. In regards to the *de novo* assembly, Velvet and CLC Genomics Workbench make use of different algorithms, however both algorithms utilise de Bruijn graphs (CLC Bio 2012). Velvet removes errors before producing high quality contigs, making it sensitive and therefore stricter during contig construction, which explains the lower number of contigs constructed using approach one in comparison to approach two (1 752 and 18 958, respectively).

Interestingly, only one putative microsatellite marker was identified using both approaches (P_{sul27} and P_{sul38}). The two approaches were equally successful with regards to microsatellite marker identification, both identifying 28 useful putative microsatellite markers (Table 4.5), despite the fact that approach two generated over 10 times more contigs than approach one. Although the two approaches identified the same number of putative microsatellite markers, approach two identified a higher number of tri-nucleotides (7) in comparison with approach one (4 tri-nucleotides) (Table 4.5). Approach two also identified a single tetra-nucleotide microsatellite marker, whereas approach one failed to identify SSRs of more than three nucleotides (Table 4.5). Tri- and tetra-nucleotide repeats amplify more reliably and provide more easily interpretable results than dinucleotide repeats. Despite this, tri- and tetra-nucleotide repeats are less abundant than dinucleotide repeats, occurring every 300 to 500 kb, in contrast with dinucleotide repeats which occur approximately every 6 kb (Carracedo *et al.* 1996; Liu and Cordes 2004).

The MiSeq Illumina platform boasts great sequencing cost reduction while delivering high sequence throughput, although very short reads are produced (Illumina Incorporation 2013), which may hamper microsatellite marker identification. The low number of microsatellite markers identified by both approaches may be attributed to the short reads and therefore shorter contigs, as longer contigs provide an improved ability to detect polymorphism (Zalapa *et al.* 2012).

Approach two was successful in regard to SNP detection, identifying 2 530 putative SNPs, while approach one failed to detect any SNPs (Table 4.5). The large discrepancy between the two approaches in regard to SNP detection is most likely due to the fact that approach two generated over 10 times more contigs than approach one.

Table 4.5: Comparison of the two bioinformatic approaches and their success in microsatellite marker and SNP discovery.

	Approach 1	Approach 2
Total number of reads	7 387 243	7 387 243
Total number of trimmed reads	6 483 851	6 503 842
Total number of contigs	1 752	18 958
Average length of contigs (bp)	124	119
Number of putative microsatellite markers	39	47
Number of useable microsatellite markers	28	28
Microsatellite markers: Di-nucleotides	24	20
Microsatellite markers: Tri-nucleotides	4	7
Microsatellite markers: Tetra-nucleotides	-	1
Total number of putative SNPs	-	2 530
SNPs: Transitions	-	606
SNPs: Transversions	-	1 924

From these 55 microsatellite marker sequences (27 from approach one, 27 from approach two and one from both approaches), forward and reverse primers can be designed to amplify the repeat which can then be tested for amplification and

optimised to be used for future studies on *P. sulcicostatus*. The 2 530 putative SNPs identified will need to be validated through further sequencing and genotyping in a number of individuals from more than one population.

The importance of the number of markers generated from NGS lies in downstream applications such as genetic diversity assessment, population structure inference, linkage studies as well as marker assisted selection (MAS) (Slate *et al.* 2009; Varshney *et al.* 2009; Ujino-Ihara *et al.* 2010). The inference of diversity and population structure is important in the conservation and management of both commercial and natural populations. Genetic markers, such as microsatellite markers and SNPs, are able to detect variations within and among populations and therefore play an integral role in studies on genetic diversity and population structure. Traditionally, linkage maps in aquaculture species have been predominantly constructed with microsatellite markers, however, the focus is shifting to the employment of SNPs in aquaculture research (Slate *et al.* 2009). The use of SNPs along with microsatellite markers for the development of linkage maps could be beneficial in non-model species such as *P. sulcicostatus*. Marker assisted selection (MAS) is commonly employed in aquaculture, as high performance individuals can be selected based on genotype at an early age, before the commercially beneficial phenotype is expressed. Molecular breeding is divided into four key phases: the development of molecular markers (DNA polymorphisms that segregate in a given population), linkage mapping, identification of QTL associated to commercially beneficial phenotypes and the implementation of MAS (Collard *et al.* 2005).

5. Conclusion

This study provided the first reduced genome sequences for *P. sulcicostatus* with a total of 7 387 243 reads using Illumina's MiSeq platform. Quality based trimming and contig construction through *de novo* assembly using two bioinformatic approaches aided in the identification of 55 putative microsatellite markers as well as 2 530 putative SNPs. The MiSeq Illumina platform produces very short reads (300 bp for paired-end and 150 bp for single-end reads) (Illumina Incorporation 2013), which

hampers microsatellite marker identification. The low number of microsatellite markers identified by both bioinformatic approaches may be attributed to these short reads and therefore short contigs. Approach two was successful in regard to SNP detection, identifying 2 530 putative SNPs, while approach one failed to detect any SNPs, most likely due to the fact that approach two generated over 10 times more contigs than approach one. The putative microsatellite markers and SNPs identified in this study, once validated, will form the base of molecular markers which can be used in future genetic studies on *P. sulcicostatus*.

6. References

- Ahmadian, A., B. Gharizadeh, A. C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen, and J. Lundeberg, 2000 Single nucleotide polymorphism analysis by pyrosequencing. *Analytical Biochemistry* 280: 103-110.
- Alderborn, A., A. Kristofferson, and U. Hammerling, 2000 Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Research* 10: 1249-1258.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.
- Astorga, M. P., 2014 Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Frontiers in Genetics* 5: 1-6.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: 1-7.
- Barchi, L., S. Lanteri, E. Portis, A. Acquadro, G. Valè, L. Toppino, and G. L. Rotino, 2011 Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics* 12: 304.
- Carracedo, A., B. Brinkmann, and W. Bär (eds.), 1996 16th Congress of the International Society for Forensic Haemogenetics. *Advances in Forensic Haemogenetics*. New York, Springer Science and Business Media 6: 1-701.
- Collard, B. C. Y., M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang, 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142: 169-196.
- CLC Bio, 2012 White paper on de novo assembly in CLC Assembly Cell 4.0. CLC Bio, Denmark. Pages 3-8.
- Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter, 2013 Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* 22: 3151-3164.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson, 2011 Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. PLoS ONE 6: 1-10.

Everett, M. V., E. D. Grau, and J. E. Seeb, 2011 Short reads and non-model species: exploring the complexities of Next-Generation Sequence assembly and SNP discovery in the absence of a reference genome. Molecular Ecology Resources 11: 93-108.

Franchini, P., M. van der Merwe, and R. Roodt-Wilding, 2011 Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. BMC Research Notes 4: 1-11.

Gardner, M. G., A. L. Fitch, T. Bertozzi, and A. J. Lowe, 2011 Rise of the machines- recommendations for ecologists when using next generation sequencing for microsatellite development. Molecular Ecology Resources 11: 1093-1101.

Illumina Incorporation, 2013 Illumina systems. <http://www.illumina.com/systems.ilmn>.

Lal, M. M., P. C. Southgate, D. R. Jerry, and K. R. Zenger, 2015 Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Marine Genomics (In Press).

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

Mascher, M., S. Wu, P. S. Armand, N. Stein, and J. Poland, 2013 Application of genotyping-by-sequencing on semiconductor sequencing platforms: A comparison of genetic and reference-based marker ordering in barley. PLoS ONE 8: e76925.

Microsoft, 2010 Microsoft Excel [computer software]. Redmond, Washington: Microsoft.

Milano, I., M. Babbucci, F. Panitz, R. Ogden, R. O. Nielsen, M. I., Taylor, S. J. Helyar, G. R. Carvalho, M. Espineira, M. Atanassova, F. Tinti, G. E. Maes, T. Patarnello, FishPopTrace Consortium, and L. Bargelloni, 2011 Novel tools for

conservation genomics: comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE* 6: e28008.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson, 2007 Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Resources* 17: 240-248.

Morin, P. A., G. Luikart, R. K. Wayne, and the SNP workshop group, 2004 SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19: 208-216.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song, 2011 Genotype and SNP calling from Next-Generation Sequencing data. *Nature Reviews Genetics* 12: 443-451.

O'Bryhim, J., J. Pin Chong, S. L. Lance, K. L. Jones, and K. J. Roe, 2012 Development and characterization of sixteen microsatellite markers for the federally endangered species: *Leptodea leptodon* (Bivalvia: Unionidae) using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources* 4; 787-789.

Rowe, H. C., S. Renaut, and A. Guggisberg, 2011 RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 1-4.

Silva, P. I. T., A. M. Martins, E. G. Gouvea, M. Pessoa-Filho, and M. E. Ferreira, 2013 Development and validation of microsatellite markers for *Brachiaria ruzizensis* obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics* 14: 1-9.

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

Ujino-Ihara, T., Y. Taguchi, Y. Moriguchi, and Y. Tsumura, 2010 An effective method for developing SNP markers based on EST data combined with high resolution melt (HRM) analysis. *BMC Research Notes* 3: 51.

Varshney, R. K., S. N. Nayak, G. D. May, and S. A. Jackson, 2009 Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522-530.

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

Wei, N., J. B. Bemmels, and C. W. Dick, 2014 The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Molecular Ecology Resources* 14: 953-965.

Zalapa, J. E., H. Cuevas, H. Zhu, S. Steffan, D. Senalik, E. Zeldin, B. McCown, R. Hardbut, and P. Simon, 2012 Using Next-Generation Sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant species. *American Journal of Botany* 99: 193-208.

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

Chapter V

Summary of Findings and Conclusion

1. Introduction

South Africa has a rich biodiversity which facilitates the utilisation of indigenous species, especially marine species, in the expansion of the bio-economy. Many of these species show great potential for commercial use, in particular the aquaculture of endemic species for high-value, specialised markets (Directorate: Marketing of DAFF 2012), yet they remain genetically uncharacterised. Identification and genetic characterisation of such species is necessary to evaluate the feasibility of aquaculture as well as formulate long-term sustainability strategies.

A number of scallop species support well established aquaculture industries worldwide. Of the 29 South African scallop species, *Pecten sulcicostatus*, a scallop species endemic to South Africa (Arendse and Pitcher 2008), has been identified as a candidate species for aquaculture (Arendse and Pitcher 2012). Despite having potential commercial value, until now, no genetic resources were available for this species. Consequently there are no molecular studies that could shed light on the genetic composition of this species or the genetic constitution of *P. sulcicostatus* populations along the South African coast. Development of genetic markers such as microsatellite markers, single nucleotide polymorphisms (SNPs) and mitochondrial markers will assist in the assessment of the genetic composition of *P. sulcicostatus*, which in turn, is important to maximise genetic diversity for future selective breeding programmes (Hayes *et al.* 2006) and for the formulation of effective genetic management strategies. The overall aim of this study was therefore to develop genetic markers for the South African scallop, *P. sulcicostatus*, and to conduct preliminary validation of these genetic markers in order to demonstrate their usefulness for future studies which aim to assess the genetic composition of this species in order to aid the future establishment of a sustainable aquaculture industry.

2. Synopsis of Main Results and Discussion

The first step in molecular genetic studies is DNA extraction (Berg *et al.* 1995). Traditionally, scallop tissue sampling methods involved sampling of the adductor

muscle (Watts *et al.* 2005; Ibarra *et al.* 2006; Kenchington *et al.* 2006; Yuan *et al.* 2009; Zhan *et al.* 2009; Gaffney *et al.* 2010; Marín *et al.* 2013; Morvezen *et al.* 2013; Silva and Gardner 2015) and the mantle musculature (An *et al.* 2009). Sampling of these internal tissue types results in the death of the study individuals, making it unsuitable for studies in which individuals should be kept alive, such as those involving repeated sampling or continued use as broodstock. This study found that, although adductor muscle tissue is the most commonly used tissue type in studies on scallops, it yields a low concentration of low quality DNA. Due to the fact that low quality DNA can negatively affect downstream applications (Karlsson *et al.* 2013), it may be necessary for future scallop studies to make use of alternate tissue types such as gonad tissue, which yields a high concentration of good quality DNA. Better yet, future studies on scallops could make use of the non-destructive DNA extraction method developed in this study. The use of tentacle tissue and mucus swabs in DNA extraction ensures no risk to the study individual, providing a non-harmful sampling method. Studies, not only on other scallop species, but also on similar bivalve species, can make use of this non-destructive sampling method when the study individuals must be kept alive for aquaculture and conservation purposes. Once aquaculture of *P. sulcicostatus* is established, the non-destructive sampling method developed in this study can be used on broodstock individuals, allowing these individuals to be used for genetic studies while remaining viable. As live *P. sulcicostatus* are difficult to obtain, this DNA extraction method can be used on dried or degraded tissue such as the tissue attached to shells washed up on beaches or shells in museum collections. This will, in turn, allow for the use of opportunistic and historic samples in future studies on *P. sulcicostatus* or similar bivalve species.

Although the use of additional mitochondrial genes will provide better accuracy, the fragment of the 16S rRNA gene used in this study revealed high haplotype diversity for *P. sulcicostatus*, much higher than that observed in other *Pecten* species, likely due to the fact that *P. sulcicostatus* is not commercially exploited (Arendse and Pitcher 2012). The low nucleotide diversity observed for *P. sulcicostatus* indicates a low level of genetic variation, however the nucleotide diversity is still much higher than that of other *Pecten* species (Saavedra and Peña 2004; Saavedra and Peña 2005). The four other *Pecten* species investigated in the present study are all

commercially exploited (Saavedra and Peña 2004) and show signs of genetic bottlenecks (D'Arcy *et al.* 2013), likely as a result of this commercial exploitation. It is therefore vital to avoid similar occurrences in the endemic *P. sulcicostatus* by preserving the natural *P. sulcicostatus* populations along the South African coast.

The 16S rRNA gene has been extensively used in phylogenetic studies (Barucca *et al.* 2004; Saavedra and Peña 2004; Saavedra and Peña 2005; Marín *et al.* 2013; Marín *et al.* 2015), and, in this study, the use of the 16S rRNA gene shed light on the evolutionary relationship of five *Pecten* species, corroborating previous studies (Saavedra and Peña 2004; Saavedra and Peña 2005), while providing the first ever insight into the relationship between *P. sulcicostatus* and the Eastern Atlantic- and Mediterranean Sea as well as the Australian and New Zealand *Pecten* scallops. The shared haplotypes observed in this study corroborate previous studies which found that *Pecten maximus* and *Pecten jacobaeus*, which have traditionally been considered distinct species, most probably represent two variants of the same species (Saavedra and Peña 2004; Saavedra and Peña 2005).

While mitochondrial genes are useful in phylogenetic studies, microsatellites and SNPs are currently the most popular molecular markers (Astorga 2014). In this study, a set of 10 microsatellite markers was developed for *P. sulcicostatus*, using cross-species transfer from *P. maximus*, with a success rate of 50%. Cross-species transfer of microsatellite markers typically has a success rate of approximately $72 \pm 7\%$ in invertebrates (Barbará *et al.* 2007), although similar studies on a number of scallop species reported success rates ranging from 0% to 77.14% (Ibarra *et al.* 2006; Marín *et al.* 2012), making the 50% success rate observed in this study comparable to that of similar studies. Even though cross-species amplification may be successful, high levels of polymorphism cannot be guaranteed (Chambers and MacAvoy 2002; Zane *et al.* 2002). Despite this, the average polymorphic information content (PIC) of 0.505 indicates that the set of microsatellite markers could be useful in future studies (Lafarga-de la Cruz *et al.* 2015). Also, the high inclusion and exclusion probability values observed in this study strongly indicate that the set of microsatellite markers will be effective in parentage assignment (Frankham 1995).

Although cross-species microsatellite marker transfer is a useful alternative to *de novo* microsatellite marker development, the use of non-species specific

microsatellite markers may result in the underestimation of genetic diversity (Zane *et al.* 2002; Arif *et al.* 2010). Newer genetic marker development methods make use of Next Generation Sequencing (NGS) technology, which allows for the identification of thousands of species-specific molecular markers (Silva *et al.* 2013; Wei *et al.* 2014). Reduced Representation Sequencing (RRS) such as Restriction-site Associated DNA (RAD) sequencing offers a reduction in sequencing cost by reducing the data obtained (Etter *et al.* 2011; Davey *et al.* 2013). This study made use of RRS in order to provide the first reduced genome sequences for *P. sulcicostatus*. Use of two bioinformatic approaches allowed for the identification of 55 putative microsatellite markers as well as 2 530 putative SNPs. Comparison of the utility of the two downstream analysis pipelines showed that the two bioinformatic approaches were equally successful with regards to microsatellite marker discovery, however approach two identified a higher number of tri-nucleotides and tetra-nucleotides in comparison with approach one. Tri- and tetra-nucleotide repeats amplify reliably and provide easily interpretable results, but are less abundant than di-nucleotide repeats, occurring every 300 to 500 kb (Liu and Cordes 2004). Approach one failed to detect any SNPs, most likely due to the low number of contigs generated using this approach, while approach two identified 2 530 putative SNPs.

Although currently limited, this pilot study allowed for the development of an effective non-destructive DNA extraction method for *P. sulcicostatus* and demonstrated the utility of the microsatellite- and mitochondrial markers developed in the current study, allowing future studies to provide sufficient genetic information in order to develop strategies for genetic management within the context of establishing a sustainable aquaculture industry for this species. In future the molecular markers developed for *P. sulcicostatus* could be used to assess the genetic diversity and underlying population structure of natural *P. sulcicostatus* populations. These molecular markers could also be continuously used to monitor genetic diversity as the species is subject to aquaculture. Once the microsatellite markers and SNPs identified *via* NGS have been validated and characterised they can add to the base of genetic markers available for *P. sulcicostatus*. The molecular markers (microsatellite markers and SNPs) developed in this study can also form a basis for linkage studies

as well as marker assisted selection (MAS) (Slate *et al.* 2009; Varshney *et al.* 2009; Ujino-Ihara *et al.* 2010).

3. Future Work and Study Limitations

This study is, to date, the only genetic study on the South African scallop, *P. sulcicostatus*. The sample size for this study was extremely limited, which may create sampling errors that could bias estimates of genetic diversity. Future studies should establish an accurate natural distribution range for this species, locating *P. sulcicostatus* populations along the South African coast for future sampling. A recent study by Foote *et al.* (2012) made use of water samples to detect harbour porpoises through environmental DNA (eDNA). As *P. sulcicostatus* is found at sublittoral depths between 22 and 70 meters (Arendse *et al.* 2008; Arendse and Pitcher 2012), they are only reachable by scuba. Therefore future studies could test the feasibility of using water samples to detect *P. sulcicostatus* eDNA, aiding in documenting *P. sulcicostatus*'s distribution range along the South African coast.

Once *P. sulcicostatus* populations are discovered, more samples from multiple locations along the South African coast need to be studied in order to evaluate the genetic variation displayed in these natural populations so as to create a broodstock population with maximum genetic diversity. A larger sample size containing individuals from multiple populations will be necessary to assess the usefulness of the set of microsatellite markers in other *P. sulcicostatus* populations. Although the set of microsatellite markers developed in this study *via* cross-species transfer proved useful, use of these non-species specific microsatellite markers may result in the underestimation of genetic diversity (Chambers and MacAvoy 2002; Zane *et al.* 2002; Arif *et al.* 2010). Following validation and optimisation of the species-specific microsatellite markers developed in this study using NGS, these microsatellite markers will allow for the more accurate estimation of genetic diversity (Chistiakov *et al.* 2006; Barbará 2007). These microsatellite markers can be added to the set of 10 microsatellite markers that have already been optimised and validated in the False Bay population, providing a larger number of microsatellite markers which can be used in future studies on *P. sulcicostatus*.

Although the 16S rRNA gene was useful in the evaluation of intra- and interspecific diversity and was able to shed light on the evolutionary relationship between the five *Pecten* species, future studies using more than one mitochondrial gene as well as nuclear markers will be able to further resolve the evolutionary relationship between *P. sulcicostatus* and other *Pecten* species. The use of additional mitochondrial genes, such as the 12S rRNA (Barucca *et al.* 2004; Malkowsky and Klusmann-Kolb 2012; Cadahía *et al.* 2014), NADH dehydrogenase subunit 1 (ND1) (Roe 2013) or cytochrome *b* (Springer 2001) genes, as well as the nuclear genes 18S (Malkowsky and Klusmann-Kolb 2012), 28S (Malkowsky and Klusmann-Kolb 2012; Oskars *et al.* 2015) and histone three (H3) (Cadahía *et al.* 2014; Oskars *et al.* 2015), will give confidence to the results obtained on the evolutionary relationship between these *Pecten* species. As the divergence time of the various *Pecten* lineages is still unclear, studies on the time since divergence will aid in the determination of the period in which divergence occurred, which will, in turn, provide insight into the events that may have facilitated divergence, such as glaciation and closure of seaways. Understanding the time since divergence will aid in understanding the evolutionary relationship between *P. sulcicostatus* and other *Pecten* species.

Although outside the scope of this study, the observed deletion specific to *P. maximus* and *P. jacobus* presents an interesting scenario of a mutation which might be beneficial to adaptation to the North Atlantic and Mediterranean environment. Future studies should, therefore, further characterise this deletion by determining whether this deletion is synonymous or non-synonymous, causing a change in gene or protein expression.

Although it is unlikely that *P. aribenedictus* individuals inhabit False Bay and that a few of these individuals were sampled in this study, this possibility cannot be ruled out. The set of microsatellite markers as well as the mitochondrial marker optimised in this study can be used on *P. aribenedictus* in order to generate the first molecular information on this species, which can then be compared to that of the *P. sulcicostatus*, to confirm or reject the possibility that *P. aribenedictus* individuals were sampled in this study as well as add to the growing molecular information on the South African *Pecten* species.

In future NGS studies, following DNA digestion, fragments larger than 200bp could be size selected in order to increase data diversity or a second restriction enzyme (RE) could be used to digest the sample pool. The microsatellite markers and SNPs developed in this study using NGS will need to be tested for amplification and then validated in a subset of individuals from more than one population. Any of the molecular markers that amplify and can be successfully genotyped can be used in future studies by applying these markers in genetic diversity and population genetic studies.

4. Final Remarks

In summary, this study provided an effective non-destructive DNA extraction method for *P. sulcicostatus* as well as an optimised and validated set of microsatellite markers and a mitochondrial marker. Although currently limited, this study provides the first step towards providing genetic information to assist in the development of genetic management strategies within the context of establishing a sustainable aquaculture industry for this endemic species. In addition, the microsatellite markers and SNPs identified in this study could be applied in assessments of regional populations on a molecular level in order to monitor natural populations over time and determine the genetic connectivity between natural *P. sulcicostatus* populations along the South Africa coast. Furthermore, the microsatellite markers and SNPs could assist in the establishment and implementation of effective aquaculture management programmes. In future the molecular markers developed for *P. sulcicostatus* could be used to monitor genetic diversity as the species is subject to aquaculture, form a basis for future linkage studies and, in the long run, be used in the implementation of MAS.

5. References

- An, H. S., Y. G. Lee, J. Y. Park, and C. Lee, 2009 Genetic characterization of four East Asian giant scallop (*Mizuhopecten yessoensis*) populations using microsatellite markers. *Aquaculture Resources* 40: 619-624.
- Arendse, D. C. Z., and G. C Pitcher, 2012 Growth and survival of the South African scallop *Pecten sulcicostatus* in suspended culture. *African Journal of Marine Science* 34: 181-186.
- Arendse, D. C. Z., S. Andersen, N. J. Blake, and G. C. Pitcher, 2008 The reproductive cycle of the scallop *Pecten sulcicostatus* from the Southern Benguela upwelling system. *Journal of Shellfish Research* 27: 281-287.
- Arif, I. A., M. A. Bakir, H. A. Khan, A. H. Al Farhan, A. A. Al Homaidan, A. H. Bahkali, M. Al Sadoon, and M. Shobrak, 2010 A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences* 11: 2079-2096.
- Astorga, M. P., 2014 Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Frontiers in Genetics* 5: 1-6.
- Barbará, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer, 2007 Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759-3767.
- Barucca, M., E. Olmo, S. Schiaparelli, and A. Canapa, 2004 Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes. *Molecular Phylogenetics and Evolution* 31: 89-95.
- Berg, D. J., W. R. Haag, S. I. Guttman, and J. B. Sickel, 1995 Mantle biopsy: A technique for nondestructive tissue-sampling of freshwater mussels. *Journal of the North American Benthological Society* 14: 577-581.
- Cadahía, L., J. Harl, M. Duda, H. Sattmann, L. Kruckenhauser, Z. Feher, L. Zopp, and E. Haring, 2014 New data on the phylogeny of Ariantinae (Pulmonata, Helicidae) and the systematic position of *Cylindrus obtusus* based on nuclear and mitochondrial DNA marker sequences. *Journal of Zoological Systematics and Evolutionary Research* 52: 163-169.

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

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

D'Arcy, J., L. Mirimin, and R. FitzGerald, 2013 Phylogeographic structure of a protogynous hermaphrodite species, the ballan wrasse *Labrus bergylta*, in Ireland, Scotland, and Norway, using mitochondrial DNA sequence data. *ICES Journal of Marine Science* 70: 685-693.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews* 12: 499-510.

Dijkstra, H. H. and R. N. Kilburn, 2001 The family Pectinidae in South Africa and Mozambique (Mollusca: Bivalvia: Pectinoidea). *African Invertebrates* 42: 263-321.

Directorate: Marketing of the Department of Agriculture, Forestry and Fisheries, 2012 A profile of the South African aquaculture market value chain.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson, 2011 Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE* 6: 1-10.

Foot A. D., P. F. Thomsen, S. Sveegaard, M. Wahlberg, J. Kielgast, L. A. Kyhn, A. B. Salling, A. Galatius, L. Orlando, M. Thomas, and P. Gilbert, 2012 Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS One* 7: e41781.

Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a review. *Genetic Research* 66: 95-107.

Gaffney, P. M., C. M. Pascal, J. Barnhart, W. S. Grant, and J. E. Seeb, 2010 Genetic homogeneity of weathervane scallops (*Patinopecten caurinus*) in the northeastern Pacific. *Canadian Journal of Fisheries and Aquatic Sciences* 67: 1827-1839.

Hoffman, J. I., and H. J. Nichols, 2011 A novel approach for mining polymorphic microsatellite markers *in silico*. PLoS ONE 6: 1-9.

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

Ibarra, A. M., J. L. Petersen, T. R. Famula, and B. May, 2006 Characterization of 35 microsatellite loci in the Pacific lion-paw scallop (*Nodipecten subnodosus*) and their cross-species amplification in four other scallops of the Pectinidae family. Molecular Ecology Notes 6: 153-156.

Karlsson, S., B. M. Larsen, L. Eriksen, and M. Hagen, 2013 Four methods of nondestructive DNA sampling from freshwater pearl mussels *Margaritifera margaritifera* L. (Bivalvia:Unionoida). Freshwater Science 32: 525-530.

Kenchington, E. L., M. U. Patwary, E. Zouros, and C. J. Bird, 2006 Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). Molecular Ecology 15: 1781-1796.

Lafarga-de la Cruz, F., A. Aguilar-Espinoza, and C. Gallardo-Escarate, 2015 Parentage assignment in hybrid abalones (*Haliotis rufescens* x *Haliotis discus hannai*) based on microsatellite DNA markers. Aquaculture Research 46: 216-225.

Malkowsky, Y., and A. Klusmann-Kolb, 2012 Phylogeny and spatio-temporal distribution of European Pectinidae (Mollusca: Bivalvia). Systematics and Biodiversity 10: 233-242.

Marín, A., T. Fujimoto, and K. Arai, 2012 Isolation and characterization of 12 microsatellite loci in the Peruvian scallop *Argopecten purpuratus* and cross-species amplification in other scallop species (family Pectinidae). Conservation Genetics Resources 4: 179-182.

Marín, A., T. Fujimoto, and K. Arai, 2013 Genetic structure of the Peruvian scallop *Argopecten purpuratus* inferred from mitochondrial and nuclear DNA variation. Marine Genomics 9: 1- 8.

- Marín, A., T. Fujimoto, and K. Arai, 2015 The mitochondrial genomes of *Pecten albicans* and *Pecten maximus* (Bivalvia: Pectinidae) reveal a novel gene arrangement with low genetic differentiation. *Biochemical Systematics and Ecology* 61: 208-217.
- Morvezen, R., F. Cornette, G. Charrier, B. Guinand, S. Lapègue, B. Pierre, and L. Jean, 2013 Multiplex PCR sets of novel microsatellite loci for the great scallop *Pecten maximus* and their application in parentage assignment. *Aquatic Living Resources* 26: 207-213.
- Oskars, T. R., P. Bouchet, and M. E. Malaquias, 2015 A new phylogeny of the Cephalaspidea (Gastropoda: Heterobranchia) based on expanded taxon sampling and gene markers. *Molecular Phylogenetics and Evolution* 89: 130-150.
- Roe, K. J., 2013 Phylogenetic analysis of the freshwater mussel genus *Ptychobranthus* (Bivalvia: Unionidae). *American Malacological Bulletin* 31: 257-265.
- Saavedra, C., and J. B. Peña, 2004 Phylogenetic relationships of commercial European and Australasian king scallops (*Pecten* spp.) based on partial 16S ribosomal RNA gene sequences. *Aquaculture* 235: 153-166.
- Saavedra, C., and J. B. Peña, 2005 Nucleotide diversity and Pleistocene population expansion in Atlantic and Mediterranean scallops (*Pecten maximus* and *P. jacobaeus*) as revealed by the mitochondrial 16S ribosomal RNA gene. *Journal of Experimental Marine Biology and Ecology* 323: 138-150.
- Sharma P. P., V. L. González, G. Y. Kawauchi, S. C. S. Andrade, A. Guzmán, T. M. Collins, E. A. Glover, E. M. Harper, J. M. Healy, P. M. Mikkelsen, J. D. Taylor, R. Bieler, and G. Giribet, 2012 Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). *Molecular Phylogenetics and Evolution* 65: 64-74.
- Silva, P. I. T., A. M. Martins, E. G. Gouvea, M. Pessoa-Filho, and M. E. Ferreira, 2013 Development and validation of microsatellite markers for *Brachiaria ruzizensis* obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics* 14: 1-9.

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

Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope, 2001 Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution* 18: 132-143.

Ujino-Ihara, T., Y. Taguchi, Y. Moriguchi, and Y. Tsumura, 2010 An effective method for developing SNP markers based on EST data combined with high resolution melt (HRM) analysis *BMC Research Notes* 3: 51.

Varshney, R. K., S. N. Nayak, G. D. May, and S. A. Jackson, 2009 Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522-530.

Watts, P. C., P. J. Mallanaphy, C. McCarthy, B. D. Beukers-Stewart, M. W. J. Mosley, A. R. Brand, and I. J. Saccheri, 2005 Polymorphic microsatellite loci isolated from the great scallop, *Pecten maximus* (Bivalvia: Pectinidae). *Molecular Ecology Notes* 5: 902-904.

Wei, N., J. B. Bemmels, and C. W. Dick, 2014 The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Molecular Ecology Resources* 14: 953-965.

Yuan, T., M. He, and L. Huang, 2009 Intraspecific genetic variation in mitochondrial 16S rRNA and COI genes in domestic and wild populations of Huaguizhikong scallop *Chlamys nobilis* Reeve. *Aquaculture* 289: 19-25.

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

Zhan, A., J. Hu, X. Hu, Z. Zhou, M. Hui, S. Wang, W. Peng, M. Wang, and Z. Bao, 2009 Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): Do local marine currents drive geographical differentiation? *Marine Biotechnology* 11: 223-235.

Appendix**Table A1: Nucleotide variation in a 582 bp fragment of the 16S rRNA gene of 18 *Pecten sulcicostatus* haplotypes. Only variable sites are shown. Dots indicate identity with the Haplotype 1 sequence.**

Site	32	39	154	258	265	266	270	285	301	307	309	314	317	319	323	326	327	367	385	391	395	411	418	
H1	G	G	C	G	C	C	G	G	C	T	G	T	C	G	G	T	G	C	A	A	G	A	G	
H2	•	•	•	•	T	•	•	•	•	•	•	•	•	T	•	•	•	•	•	G	•	•	•	
H3	•	•	•	A	T	T	•	A	•	•	T	•	•	•	A	•	A	•	T	•	•	•	•	
H4	•	A	•	A	T	T	•	A	T	•	T	•	•	•	A	•	•	•	T	•	•	•	•	
H5	•	•	•	A	T	T	•	A	T	•	T	C	•	•	A	•	•	•	T	•	•	•	A	
H6	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•	•
H7	•	•	•	A	T	T	•	A	T	•	T	•	•	•	A	•	•	•	T	•	•	•	A	
H8	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•
H9	•	•	•	A	T	T	•	A	T	•	T	•	•	•	A	•	•	•	T	•	•	•	•	
H10	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
H11	•	•	•	A	T	T	•	A	T	•	T	•	A	•	A	•	•	•	T	•	•	•	•	
H12	•	•	T	A	T	T	•	A	T	•	T	•	•	•	A	•	•	•	T	•	•	•	A	
H13	•	•	•	A	T	T	•	A	T	•	T	•	•	•	A	A	A	•	T	•	•	•	A	
H14	•	•	•	A	T	T	•	A	T	•	T	•	•	•	A	•	A	•	T	•	•	•	•	
H15	T	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
H16	•	•	•	A	T	T	•	A	•	C	T	•	•	•	A	•	A	•	T	•	•	•	•	
H17	•	•	•	A	T	T	A	A	•	•	T	•	•	•	A	•	A	•	T	•	•	•	•	
H18	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	T	•	•	A	•	•	

Table A2: Nucleotide variation in a 504 bp fragment of the 16S rRNA gene of 32 haplotypes. Only variable sites (site 32 to site 305) are shown. Dots indicate identity with the Haplotype 1 sequence.

	32	33	39	67	154	258	264	265	266	268	270	275	279	281	284	285	296	298	301	304	305
H1	G	T	G	C	C	G	T	T	C	G	G	A	G	G	C	G	A	A	C	C	A
H2	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H3	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H4	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H5	•	C	•	•	•	•	•	•	•	A	•	-	•	•	•	•	G	G	•	G	•
H6	•	C	•	•	•	•	•	•	•	A	•	-	•	•	•	•	G	G	•	G	•
H7	•	C	•	•	•	•	•	•	•	A	•	-	•	A	T	•	G	G	•	G	•
H8	•	C	•	•	•	•	•	•	•	A	•	-	•	A	•	•	G	G	•	G	•
H9	•	C	•	•	•	•	•	•	•	A	•	-	T	•	•	•	G	G	•	G	•
H10	•	C	•	•	•	•	C	•	•	A	•	-	•	A	•	•	G	G	•	G	•
H11	•	•	•	T	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H12	•	•	•	T	•	•	•	•	•	A	•	•	•	•	T	•	G	G	•	T	•
H13	•	C	•	T	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H14	•	C	•	T	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	•
H15	•	•	•	•	•	•	•	C	•	A	•	•	•	•	•	•	G	G	•	T	G
H16	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G
H17	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	•	T	G
H18	•	•	A	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H19	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G

H20	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G
H21	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H22	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G
H23	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H24	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G
H25	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H26	•	•	•	•	T	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H27	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H28	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	T	T	G
H29	T	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G
H30	•	•	•	•	•	A	•	•	T	A	•	•	•	•	•	A	G	G	•	T	G
H31	•	•	•	•	•	A	•	•	T	A	A	•	•	•	•	A	G	G	•	T	G
H32	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	G	G	•	T	G

Table A3: Nucleotide variation in a 504 bp fragment of the 16S rRNA gene of 32 haplotypes. Only variable sites (site 306 to site 471) are shown. Dots indicate identity with the Haplotype 1 sequence.

	306	307	308	309	314	317	319	323	326	327	367	384	385	389	391	395	411	412	418	471
H1	A	T	T	T	T	T	G	G	T	G	C	G	C	A	A	G	A	T	G	T
H2	•	•	•	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H3	•	•	C	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H4	•	•	C	•	•	•	•	•	•	•	•	A	T	G	•	•	•	C	•	•
H5	•	C	•	•	•	C	•	•	•	•	•	•	T	•	•	•	•	C	•	C
H6	•	C	•	•	•	C	•	•	•	•	•	•	T	•	•	•	•	C	•	•
H7	•	C	•	•	•	C	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H8	•	C	•	•	•	C	•	•	•	•	•	•	T	•	•	•	•	C	•	•
H9	•	C	•	•	•	C	•	•	•	•	•	•	T	•	•	•	•	C	•	•
H10	•	•	•	•	•	C	•	•	•	•	•	•	T	•	•	•	•	C	•	•
H11	•	•	C	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H12	•	•	C	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H13	•	•	C	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H14	•	•	•	•	•	•	•	•	•	•	•	A	T	•	•	•	•	C	•	•
H15	G	•	•	G	•	C	•	•	•	•	•	•	A	•	•	•	•	C	•	•
H16	G	•	•	G	•	C	T	•	•	•	•	•	A	•	G	•	•	C	•	•
H17	G	•	•	•	•	C	•	A	•	A	•	•	T	•	•	•	•	C	•	•
H18	G	•	•	•	•	C	•	A	•	•	•	•	T	•	•	•	•	C	•	•
H19	G	•	•	•	C	C	•	A	•	•	•	•	T	•	•	•	•	C	A	•

H20	G	•	•	G	•	C	•	•	•	•	•	•	A	•	G	•	•	C	•	•
H21	G	•	•	•	•	C	•	A	•	•	•	•	T	•	•	•	•	C	A	•
H22	G	•	•	G	•	C	•	•	•	•	•	•	A	•	•	•	G	C	•	•
H23	G	•	•	•	•	C	•	A	•	•	•	•	T	•	•	•	•	C	•	•
H24	G	•	•	G	•	C	•	•	•	•	•	•	A	•	•	•	•	C	•	•
H25	G	•	•	•	•	A	•	A	•	•	•	•	T	•	•	•	•	C	•	•
H26	G	•	•	•	•	C	•	A	•	•	•	•	T	•	•	•	•	C	A	•
H27	G	•	•	•	•	C	•	A	A	A	•	•	T	•	•	•	•	C	A	•
H28	G	•	•	•	•	C	•	A	•	A	•	•	T	•	•	•	•	C	•	•
H29	G	•	•	G	•	C	•	•	•	•	•	•	A	•	•	•	•	C	•	•
H30	G	C	•	•	•	C	•	A	•	A	•	•	T	•	•	•	•	C	•	•
H31	G	•	•	•	•	C	•	A	•	A	•	•	T	•	•	•	•	C	•	•
H32	G	•	•	G	•	C	•	•	•	•	T	•	A	•	•	A	•	C	•	•

Table A4: Haplotype frequencies of the five *Pecten* species.

	<i>P. fumatus</i>	<i>P. jacobeus</i>	<i>P. maximus</i>	<i>P. novaezelandiae</i>	<i>P. sulcicostatus</i>
Haplotype 1	1				
Haplotype 2	1				
Haplotype 3	3				
Haplotype 4	1				
Haplotype 5		1			
Haplotype 6		3	3		
Haplotype 7		1			
Haplotype 8		2	2		
Haplotype 9		1			
Haplotype 10			2		
Haplotype 11				9	
Haplotype 12				1	
Haplotype 13				1	
Haplotype 14				1	
Haplotype 15					1
Haplotype 16					1
Haplotype 17					6
Haplotype 18					1
Haplotype 19					1
Haplotype 20					1

Haplotype 21					7
Haplotype 22					1
Haplotype 23					5
Haplotype 24					6
Haplotype 25					1
Haplotype 26					1
Haplotype 27					1
Haplotype 28					1
Haplotype 29					1
Haplotype 30					1
Haplotype 31					1
Haplotype 32					1
Number of individuals	6	8	7	12	38