Marine genomics meets ecology:

Diversity and divergence in South African sea stars of the genus *Parvulastra*

Katherine Dunbar

Thesis submitted for the degree of Doctor of Philosophy

Biodiversity and Ecological Processes Research Group

School of Biosciences

Cardiff University

December 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date 3rd October 2007.

STATEMENT 1

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references.

Signed ..... (candidate)

Date 3rd October 2007.

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations

Signed ..... (candidate)

Date 3rd October 2007.

Abstract

The coast of South Africa is situated between the warm Indian and the cold Atlantic Oceans, resulting in an extreme intertidal temperature gradient and potentially strong opposing selection pressures between the east and west coasts. Several intertidal biogeographic divides have been identified, including one at Cape Point between the cold west coast and the temperate south coast provinces. However, few studies have investigated the effects of these opposing environments on phylogeography or gene flow in intertidal organisms. A small intertidal sea star, *Parvulastra exigua*, was chosen as a model organism to investigate these issues. Taxonomic confusion in this species and its systematic relationship with a related South African *Parvulastra* species, *Parvulastra dyscrita*, was resolved using nuclear (Actin intron sequences and AFLP) and mtDNA molecular markers. At least one cryptic species was identified within *Parvulastra* in South Africa, which occupied an extremely restricted geographic distribution and therefore may be a candidate for conservation. Molecular and morphological evidence confirmed that *P. exigua* and *P. dyscrita* are separate species.

An ecological survey was conducted on *P. exigua* at 19 locations in South Africa covering a distance of 2500 km. *P. exigua* samples from each location were sequenced for mtDNA and screened for 421 AFLP loci. AFLP was also used to identify outlier loci that were potentially under selection. An 'unmottled' colour morph was distributed from the Namibian border to Cape Point and a 'mottled' morph was distributed from Cape Point to the Mozambique border, with an area of sympatry around Cape Point. The unmottled morphs were positively influenced by under boulder and bare rock habitats, but negatively affected by canopy, coralline algae and sand. Mottled morphs were positively influenced by under boulder, protected habitats, encrusting algae and bare rock, and negatively affected by algal tufts and sand. MtDNA revealed two divergent, reciprocally monophyletic clades, one comprising the east coast samples and the other encompassing the west coast samples. Both clades showed evidence for a recent, rapid population expansion. The genetic break-point was located on the south coast, but did not coincide with the divergence in colour morphs, being approximately 500 km to the east. AFLP indicated a strong isolation by distance pattern of genetic structure among sampling locations and did not recapitulate the mtDNA genetic divide. Such incongruence among data sets might be caused by a vicarance event if sea level changes separated the east and west coast populations, which expanded in isolation, followed by secondary contact, restoring present day gene flow between the coasts. Population genomic analysis revealed approximately 7% of the genome to potentially be under divergent selection, and the phenotype frequencies of the 'diverging outlier loci' revealed high directionality (spatial correlation). This suggests strong selection pressures between the east and west coasts may be acting on these loci, which could have arisen when the populations were in allopatry. The habitat and colour morph differences of *P. exigua* between the two coasts are potentially also influenced by selection. However, the isolation by distance pattern indicates that divergent selection pressures are not strong enough to cause reproductive isolation, or disrupt gene flow between the east and west coast populations.

Acknowledgements

This work was funded by the Natural Environment Research Council (NERC). I am extremely indebted to my supervisor Prof Mike Bruford for enabling me to undertake this research, and his support, patience, enthusiasm and encouragement throughout my Ph.D. My thanks also goes to Mike and Dr. David Lees for inspiring me to study molecular ecology and animal speciation. I would also like to sincerely thank my second supervisor Dr. Jo Cable for being extremely supportive and helpful throughout. She was both a great supervisor and friend. My thanks goes to my other supervisor Prof. Charlie Griffiths for providing me with lab space both times I visited South Africa, as well as drawing my attention to the little seastars that are the subject of this work. I would also like to very much thank Dr. Jo Lello for her invaluable statistical help and advice. My thanks goes to all the people who helped me in South Africa with my fieldwork, specifically Tom Peschak for making my first trip to South Africa so amazing, and inspiring me to go back, Ben Pizii, who kept me laughing all the way round the coast, and remains a good friend back in the UK, Grahame and Shane who kept me safe and provided me with their fantastic vehicles and experience, Bronwen Hobbs and her family for being kind to me and providing me with a place to live and the girls in Charlie's lab who made me laugh and kept me sane (just!). A massive thank you goes to all my friends and colleagues from G10, Jo's lab and Bill's lab who have given me advice and help with lab work and kept me laughing throughout good times and frustrating times. They are too numerous to name. My deepest thanks also go to my parents and my brother who have been supportive and encouraging throughout. My sincerest thanks also goes to my surrogate Cardiff family, Matt, Nicky, Jo, Justin, Penny, Stewart, Rebecca and Matt L, for their friendship and laughter for which I am truly grateful. Finally, but most importantly I would like to dedicate this PhD to Jon Russill, He has been supportive and patient every step of the way. Without his quiet steady love and encouragement I would not have been able to complete this work. He is my strength and my sunshine.

Chapter 1

General Introduction

Chapter 1: General Introduction

1.0. Introduction

Phylogeography is the study of the patterns and processes governing the geographic distributions of genealogical lineages, especially within and among closely related species. Phylogeography focuses explicitly on historical and phylogenetic components of population structure and how these are influenced by the processes of genetic drift, gene flow, natural selection and other evolutionary forces (Avise 2004). Throughout this study the phylogeography of *Parvulastra* species in South Africa will be examined using molecular approaches in the context of systematics, life history, global distribution, phenotype, past demography, ecology and adaptation. The intertidal coastline of South Africa is a dynamic arena within which to investigate these processes as it is the meeting point of two great ocean bodies, the Atlantic and the Indian, with their associated contrasting characteristics. In this Chapter, asterinid systematics will be explored, as well as the taxonomic confusion between the two study organisms *Parvulastra exigua* and *Parvulastra dyscrita*. This will be followed by a brief description of the two study species as well as an account of their habitat and geographical distribution in South Africa, and the life histories and what is currently known about the phylogeography of the asterinids. The molecular markers available to tackle gaps in our current knowledge of *Parvulastra* species will be described before the specific aims of the project are outlined.

1.1. Asterinid systematics

Within the Family Asterinidae (Grey 1840) there are 21 genera and 116 species worldwide (O'Loughlin and Waters 2004). The taxonomy of this family in the past has been confounded by morphological characters that are of dubious phylogenetic value, with some morphological characters being subject to strong selection, phenotypically plastic (O'Loughlin and Waters 2004) or

phylogenetically informative for some clades but homoplasious and unreliable for others (Mah 2000). Due to these problems with traditional taxonomy, researchers have turned to molecular data to reassess systematic relationships (Hart *et al.* 1997; Dartnall *et al.* 2003; O'Loughlin and Waters 2004; Waters *et al.* 2004a). Several morphological and molecular systematic revisions of this family have been reported (reviewed in O'Loughlin and Waters 2004) and the most recent of these phylogenies, compiled from mtDNA sequence (Waters *et al.* 2004a) and morphological data (O'Loughlin and Waters 2004) erected several new genera and re-assigned several species to new genera (Fig. 1.1), including the genus *Parvulastra*.

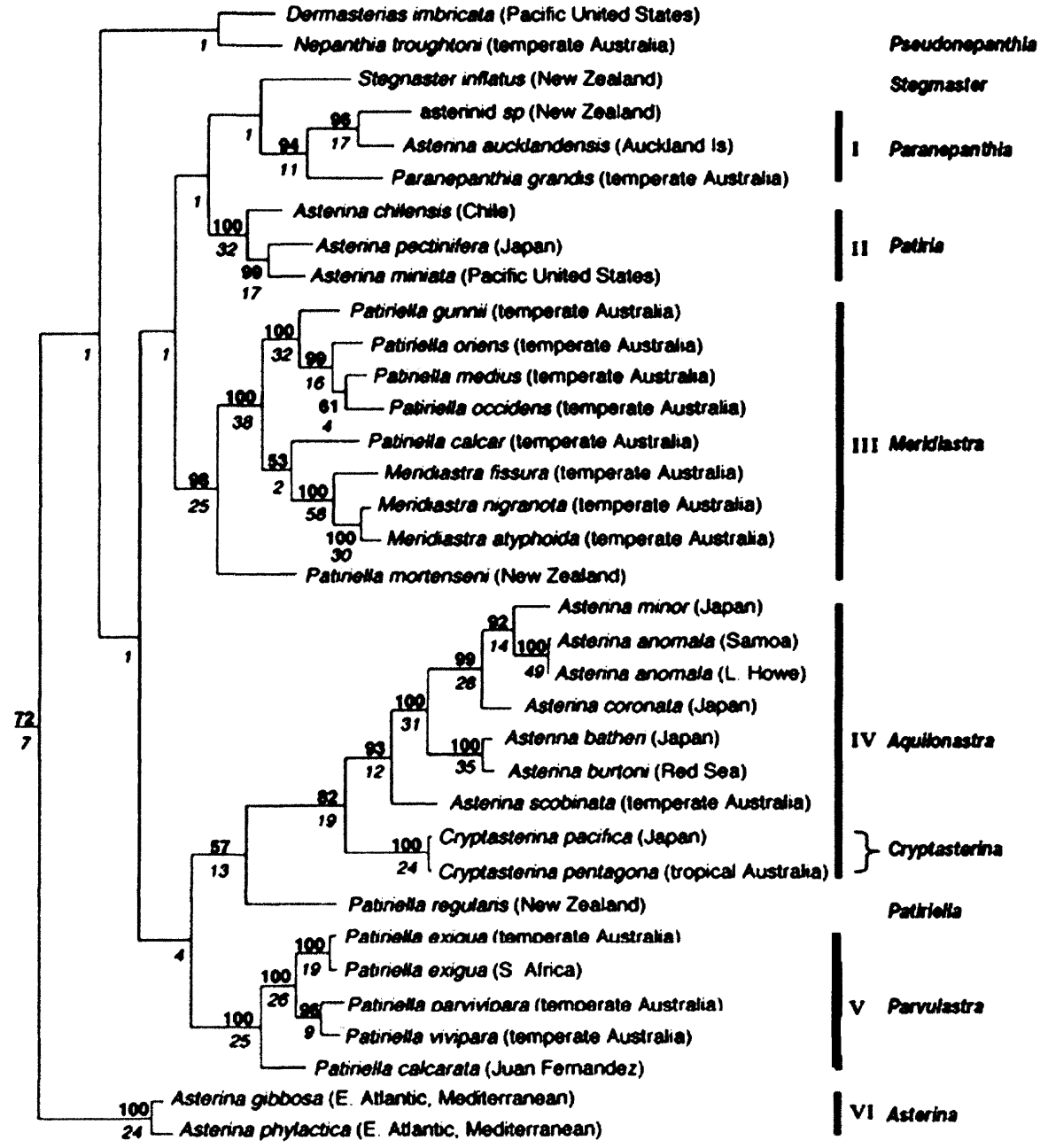

Figure 1.1. Phylogenetic relationships of asterinids based on mtDNA sequences (Waters et al. 2004a). Support for particular clades is indicated by bootstrap values, and by decay indices (in italics). MtDNA clades I – IV are identified on the basis of high bootstrap support and new generic assignments suggested by Waters *et al.* (2004a) appear on the right (from Waters *et al.* 2004a).

1.1.1 The genus *Parvulastra*

Following the taxonomic and systematic revision of O'Loughlin and Waters (2004), *Patiriella exigua* and *P. dyscrita* were re-assigned to a new genus *Parvulastra*, along with *P. parvivipara*, *P. vivipara* and *P. calcarata*, on the basis of their shared and consistent morphological and molecular characteristics (although no molecular work had been conducted on *P. dyscrita*). Some studies published since O'Loughlin and Waters' (2004) study have referred to the previous name *Patiriella exigua*, or *Parvulastra (=Patiriella) exigua*, when discussing this species (Waters and Roy 2004; Colgan *et al.* 2005; Hart *et al.* 2006). The *Parvulastra* genus is morphologically similar (but clearly differentiated) to the genera *Patiriella* and *Cryptasterina* (see Dartnall *et al.* 2003) but is not closely related on molecular grounds (O'Loughlin and Waters 2004). The latter authors suggest that the morphological characters have remained stable while divergence has occurred in characters not considered, there is strong morphological convergence among three unrelated genera, or that the molecular data are unreliable at levels more basal then inter-species.

Recently several new cryptic species within the Asterinidae have been identified (O'Loughlin 2002; O'Loughlin *et al.* 2002; Dartnall *et al.* 2003; Hart *et al.* 2003) with speculation that more species will be found as investigations proceed (Dartnall *et al.* 2003). As these genera contain the greatest diversity of larval types and life histories known among extant sea stars (see Section 1.4) (Hart *et al.* 1997), changes in larval development characters may also have driven speciation (Byrne *et al.* 1999).

1.1.2 Taxonomic confusion of *P. exigua* and *P. dyscrita*

In Southern Africa, two species of *Parvulastra* are currently recognised, *P. exigua* and *P. dyscrita*; however, it has been questioned several times whether these two taxa should be classed as the same species. Their taxonomic and systematic history was meticulously reviewed by Hart *et al.* (2006). To

summarize, *Asterias (Asterina) exigua* was first described by Lamarck in 1816. Verrill (1913) assigned *Asterina exigua* and several other species to a new genus, *Patiriella*. Mortensen (1921) (cited in Hart *et al.* 2006) emphasized that *P. exigua* had oral (underside of sea star, where mouth is located) gonopores (a key taxonomic characteristic) and suggested that *P. exigua* actually included several species.

The first description of *P. dyscrita* was by H. L. Clark (1923) who described *Asterina dyscrita*, which had aboral (upper surface of sea star) gonopores, but was otherwise similar to *P. exigua*, co-occurring with *P. exigua* in South Africa. In 1933, the two species were re-classified as synonyms (Mortensen 1933), but were then separated again into two species by Dartnall (1971) on the basis of gonopore position. Dartnall (1971) also raised a neotype for *P. dyscrita* as the original had been lost, from which A.M. Clark (1974) confirmed the aboral location of gonopores in H. L. Clarke's *P. dyscrita*. Hart *et al.* (2006) noted that the existence of *P. dyscrita* was not widely recognised until the 1990's. However, many South African workers, from the early 20th century differentiated *P. exigua* from *P. dyscrita* (C. Griffiths and K. Dunbar pers. observations), but within *P. exigua* striking phenotypic differences were noted (references cited in Hart *et al.* 2006).

Recently several authors have suggested that *P. exigua* might contain cryptic species. Walenkamp (1990) suggested that South African collections might include *P. dyscrita* as well as *P. exigua* and another species '*pentagona*', (which was subsequently placed in a new genus, *Cryptasterina* (see Dartnall *et al.* 2003) on the basis that the collections contained specimens with both oral and aboral gonopores). When the genus *Patiriella* was changed to *Parvulastra* (O'Loughlin and Waters 2004), several morphological (but not molecular) characteristics were noted that distinguished *P. exigua* from *P. dyscrita*. In 2004, Waters and Roy (2004) controversially published a phylogeographic account of the global distribution of *P. exigua* and proposed that South Africa is the ancestral origin of

P. exigua and that Australia was subsequently colonized via the west wind drift (see phylogeography Section 1.5). In 2005, the apparent mix of oral and aboral gonopores within *P. exigua* was again raised in the literature by Colgan *et al.* (2005) who confirmed that in South Africa and three Atlantic Ocean islands the presence of oral and aboral gonopores within the '*P. exigua*' taxon had been observed by Dartnall and Byrne (unpublished observations cited in Colgan *et al.* 2005). By re-opening the question of whether '*P. exigua*' contained cryptic species, Colgan *et al.* (2005) had to describe Waters and Roy's (2004) 'Out of Africa' hypothesis as 'uncertain'.

Hart *et al.* (2006) also questioned Waters and Roy's (2004) 'Out of Africa' hypothesis on the basis that samples used in the earlier study might have included cryptic species (based on gonopore and molecular divergence evidence) and therefore the assumptions behind the hypothesis would be wrong. Crucial to this argument, though, was the ability to link the '*P. exigua*' specimens with aboral gonopores to the highly divergent genotypes presented in Waters and Roy's (2004) study. This was not achieved by Hart *et al.* (2006) and so the debate remains unresolved. Hart *et al.* (2006) tentatively concluded that the Cape Town specimens from Waters and Roy's (2004) study were in fact *P. dyscrita* and that *P. exigua* contains further cryptic species. They also suggested that the centre for haplotype diversity (and implied ancestral origin) is Australia rather than South Africa, with the implication being that colonisation occurred from Australia to South Africa, in the opposite direction to that predicted by Waters and Roy's (2004). Hart *et al.* (2006) acknowledge, however, that sampling intensity was greater in Australia which may have biased the inference of the root haplotypes. Crucial to this ongoing debate is the fact that *P. dyscrita* has still not been genotyped or included in any phylogenies of asterinid taxa, despite much speculation over its taxonomic status.

1.2. *Parvulastra*: The study species

1.2.1. *Parvulastra exigua*

Parvulastra exigua (Lamarck 1816) is a small, abundant intertidal sea star found mainly in rocky shore habitats, but can occur in sandy lagoons and wave exposed cliffs (pers. obs). It is described as a scavenging omnivore and feeds by protruding its gut onto the substrate (Branch and Branch 1980). In South Africa *P. exigua* has a geographic distribution covering the whole coastal range, stretching from the Namibian to the Mozambique border (Fig 1.2). *Parvulastra exigua*'s range also extends across the southern hemisphere including Australia and several oceanic islands, but it does not occur in South America or New Zealand (Waters and Roy 2004). *Parvulastra exigua* is thought to have an entirely benthic life cycle, producing large eggs which hatch into benthic larvae which metamorphose into 'crawl-away' juveniles (Byrne 1995). *Parvulastra exigua* has a flattened pentagonal shape with short, stubby arms and its dorsal surface is made up of tile-like plates each with a cluster of knob-like spines (Branch *et al.* 1994). Most notably, in South Africa a strong phenotypic gradient exists with an unmottled, often greenish brown phenotype found in abundance on the cold water west coast and an irregular multicoloured mottled phenotype found in abundance on the warm east and south coasts (Fig. 1.3). There have been many studies on the ecology (Branch and Branch 1980; Arrontes and Underwood 1991; Byrne 1992; Stevenson 1992; Byrne and Anderson 1994; Byrne 1995; Anderson and Underwood 1997; Anderson 1999; Moreno and Hoegh-Guldberg 1999; Waters and Roy 2004) and genetics (Hunt 1993; Colgan *et al.* 2005; Hart *et al.* 2006) of *P. exigua* in Australia but no known ecological and genetic studies on this species in South Africa.

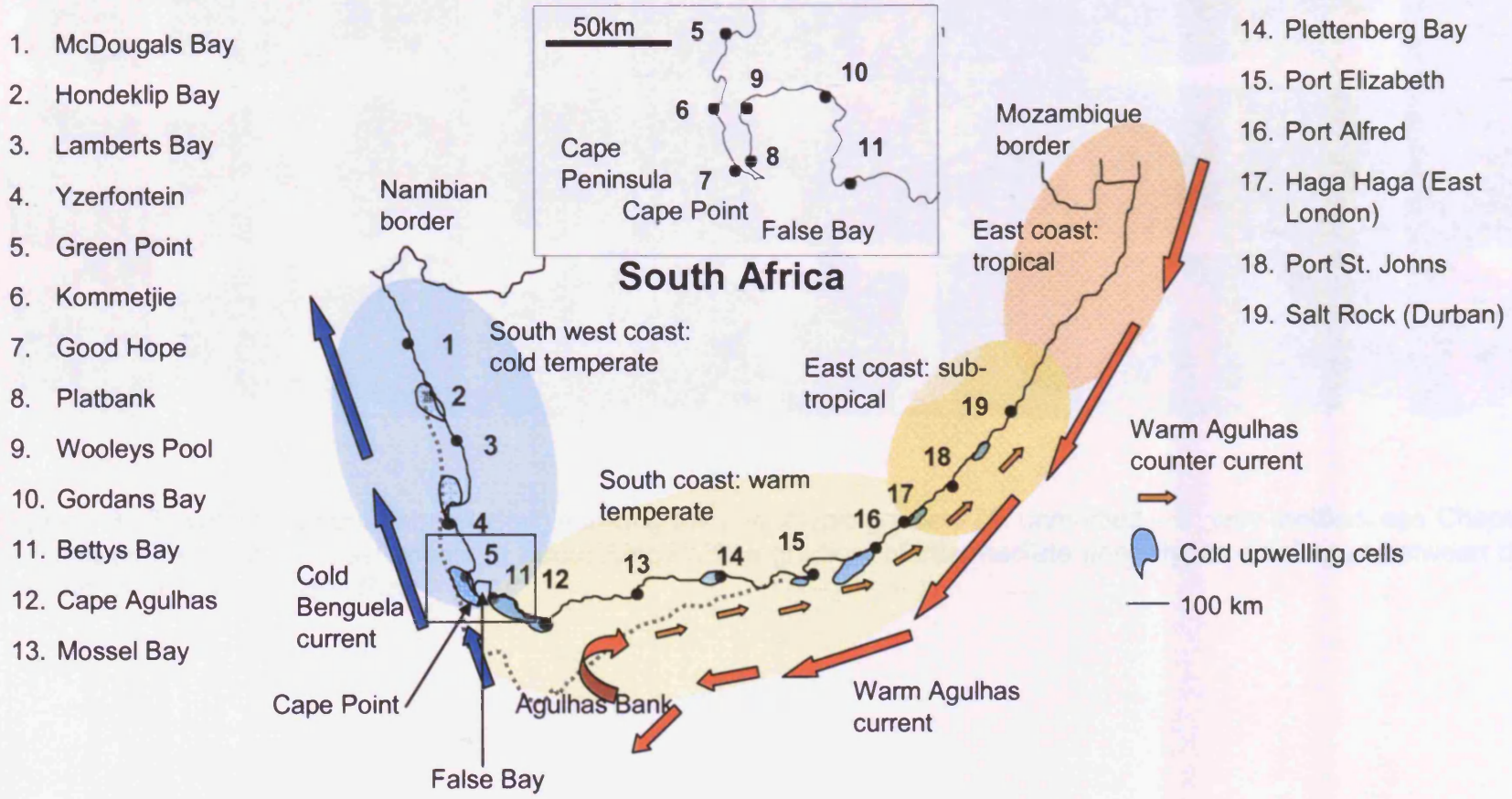

Figure 1.2. Oceanic currents and biogeographic zones (Section 1.2.2.) around the coast of South Africa.

Figure 1.3. *Parvulastra exigua* (aboral view) showing the phenotypic gradient (1: unmottled – 5: very mottled, see Chapter 3) found on the west and east coasts of South Africa with a gradient of intermediate phenotypes distributed between the two coasts (photograph Ben Pizii).

1.2.2 *Parvulastra dyscrita*

Parvulastra dyscrita (Clark 1923) is a common mainly subtidal to very low intertidal sea star endemic to South Africa, with a geographic range along the southern coast from Cape Point to approximately East London (Fig. 1.2). *Parvulastra exigua* and *P. dyscrita* differ in their life cycles since *P. dyscrita* is thought to have a planktonic phase (Branch and Griffiths 1994). *Parvulastra dyscrita* is larger than *P. exigua* and similar in shape, but can grow up to 5 cm diameter. *Parvulastra dyscrita* has an even granular surface texture rather than clusters of spines (Branch and Griffiths 1994). *Parvulastra dyscrita* has an irregular mottled pattern (Fig. 1.4) similar to the mottled phenotype of *P. exigua*. Nothing has been published on the life history, reproduction or ecology of *P. dyscrita* since the original species description by Clark (1923).

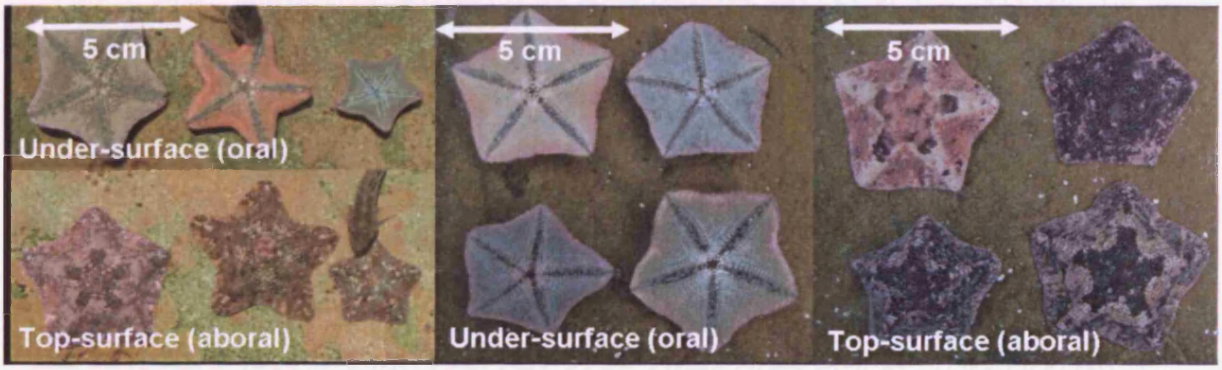

Figure 1.4. *Parvulastra dyscrita* phenotypes of the top-surface (aboral) and under-surface (oral), which shows the five lines of tube feet with the mouth location in the centre (photograph Ben Pizii).

1.3. South Africa: The study site

1.3.1 Intertidal environment and geological history

The intertidal environment is a unique habitat which is subject to extreme abiotic conditions and therefore the fauna and flora are dependent upon physiological and behavioural adaptations. The abiotic pressures include flooding with salt water twice a day, exposure to air, with possibly extreme temperatures which may cause

desiccation, freshwater precipitation, long or short term fluctuations in pH, oxygen, salinity and carbon dioxide and physical attrition in the form of wave action that may cause dislodgement (Branch and Branch 1981). The biotic pressures include both terrestrial and marine predators and competition for resources. The South African coastal environment has two great ocean bodies meeting at its tip, which have major influences on the marine biogeography of the South African coast. To the west of Africa is the cold Atlantic Ocean and to the east is the warm Indian Ocean (Fig. 1.5).

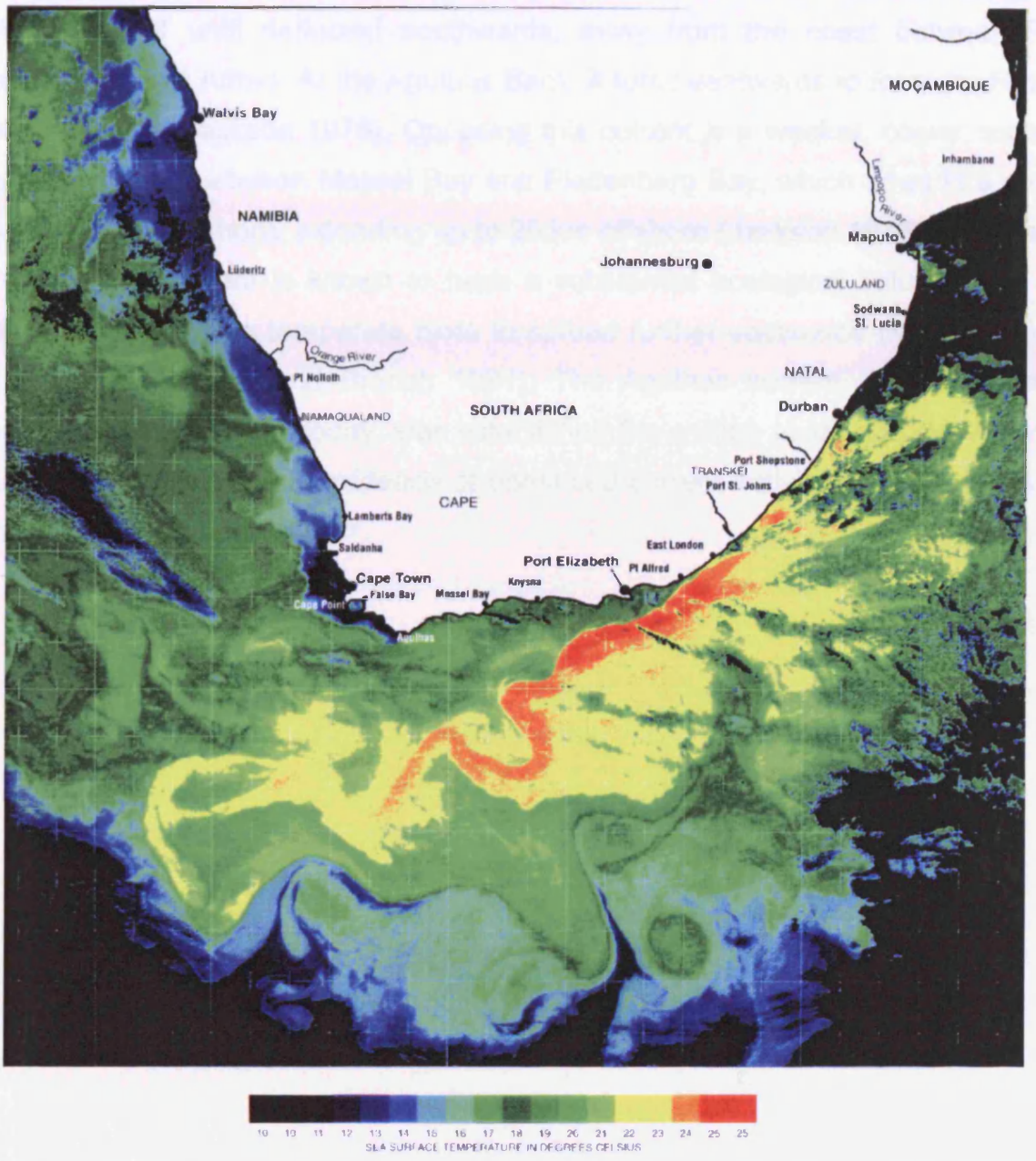

Figure 1.5. Satellite map of mean water temperatures around the coast of Southern Africa (taken from Branch and Griffiths 1994)

The temperature regime of the coastal zone and the associated flora and fauna are dependant on both the latitude and the two powerful prevailing ocean current systems (Fig. 1.2) (Brown and Jarman 1978; Emanuel *et al.* 1992). On the east coast of South Africa the fast, southward flowing, warm Agulhas current originates from the south west Indian Ocean sub gyre (Peschak 2005). At its northern end, this current can

reach 28°C and at its southern edge 21°C in summer and 16°C in winter (Peschak 2005), but it generally becomes cooler further south. The current flows along the continental shelf until deflected southwards, away from the coast between Port Elizabeth and Port Alfred. At the Agulhas Bank, it turns eastwards to form the Return Agulhas Current (Jackson 1976). Opposing this current is a weaker, cooler, surface current originating between Mossel Bay and Plettenberg Bay, which flows in a north-easterly direction inshore, extending up to 20 km offshore (Jackson 1976). Despite its low speed, this current is known to have a substantial ecological influence on the intertidal biota enabling temperate biota to spread further eastwards (Macnae 1961; Jackson 1976; Branch and Branch 1981). The Agulhas current, with flow paths broadly similar to those of today, was established five million years ago (MYA) (mid-Pliocene), although there is evidence of enhanced current activity dating back as far as the Oligocene (Hiller 1994).

The currents become complex on the south coast and the temperature is affected by a number of factors such as wind direction and force (Brown and Jarman 1978). Mid way along the south coast the Agulhas Bank, a triangular extension of the continental shelf, drops steeply from the coast to 50 m then gradually deepens to 200 m, at a distance offshore of 250 km, where it then drops steeply to more than 1000 m (Hutchings 1994). The eastern edge of the bank, between Mossel Bay and Plettenberg Bay is subject to occasional upwelling events and surface cooling, and has interjections of warmer water from the Agulhas current (Hutchings 1994). In the centre of the Agulhas Bank, south of Mossel Bay, cyclonic circulation around a cool water ridge, with an eastward flow on the inner margin and westward flow offshore, may prevent offshore loss of plankton into the Agulhas itself. From Cape Agulhas to East London, the mean annual temperatures are 17-18°C (Bolton and Anderson 1997). The western part of the Agulhas Bank, from Cape Agulhas to the Cape Peninsula is considered part of the upwelling regime of the Benguela system (Hutchings 1994) and has a westward convergence of currents towards Cape Point, with currents rounding the point and heading up the western side (Hutchings 1994). The area around the Cape Peninsula divides two bodies of water with very different temperatures. At the same latitude on either side of the peninsula the water temperature can differ by as much as 8°C on the same day. This is possibly the only ocean of this size in the world with such a drastic temperature gradient over such a small area (Brown and Jarman 1978).

In contrast, the west coast of Southern Africa is influenced by the slow northward flowing Benguela current which is reinforced by the Circum-Antarctic West Wind Drift current. This means that the west coast waters are considerably colder than those of the south and east coasts at the same latitude. The Benguela current is dominated by strong upwelling, bringing cold nutrient rich waters to the surface close inshore in sporadically distributed upwelling cells (Peschak 2005). Towards the end of the Pliocene, sea levels fluctuated and the seas gradually cooled, with a 10°C drop in the region (Marlow *et al.* 2000). During this time the Benguela system was established approximately 2-3 MYA (Shannon 1985; Bolton and Anderson 1997). The upwelling is wind-driven and so the sea surface temperature can fluctuate greatly from day to day.

The average temperature is 8-17°C but changes from 17°C to 8°C can occur over a 7 hour period (Branch and Branch 1981). False Bay can reach 23°C in the summer (Brown and Jarman 1978) but this is unusually high for a western location.

Over the past 300,000 years there have been four glacial periods (Branch and Branch 1978), the most recent of which, 20,000 YA, caused the sea level to drop by approximately 120 m, exposing much of the Agulhas Bank and extending the coastline between 50 and 150 km out to sea. During inter-glacial periods, the sea level may have risen to 50 m above present, causing the Cape Peninsula to become an island (Branch and Branch 1978).

1.3.2 Biogeography

Southern Africa has an extreme intertidal water temperature gradient ranging from 10-16°C on the west coast to 10-24°C for the east coast. Many authors have investigated the implications this has on the biogeography of the intertidal zone (Jackson 1976; Brown and Jarman 1978; Emanuel *et al.* 1992; Hiller 1994; Bustamante *et. al.* 1997; Neraudeau and Mathey 2000), but relatively few have investigated phylogeographic patterns and compared them to these well established biogeographic patterns (reviewed in Chapter 3). Emanuel *et al.* (1992) used presence / absence distribution data from the literature for 2000 marine invertebrate species and concluded that the South African coastal region should be divided into four biogeographic intertidal zones (Fig. 1.2): (i) The cool temperate north west coast (Namibia – not shown in Figure 1.2); (ii) the cool temperate south west coast; (iii) the warm temperate south coast; and (iv) the sub-tropical east coast which has a sub division at Durban, where north of this is the tropical east coast. Detailed studies by Stephenson (1944) indicate that the boundary between the cool temperate south west coast and the warm temperate south coast zones is actually at Kommetjie on the western side of the Cape Peninsula and not at Cape Point itself as other studies have concluded (Emanuel *et al.* 1992). The intertidal geology at Kommetjie also shows discontinuity (Brown and Jarman 1978). Although there has been much debate in the past over the number of biogeographic intertidal zones and their exact boundaries,

most studies have since adopted the view that the area between Cape Agulhas and Kommetjie is an overlap zone between the warm east and cold west biogeographic zones.

The intertidal biota on the west coast tends to be less diverse, but more abundant than that of the south and east coast provinces (Brown and Jarman 1978). Mid way up the west coast an almost landlocked lagoon (Langabaan Lagoon) has uncharacteristically warm water and contains biota which is more diverse and productive than the rest of the west coast. Many species occur here which have not been recorded in the rest of the west coast province and are characteristic of other provinces, notably the south coast warm temperate region (Brown and Jarman 1978). The algae and some invertebrate species (including *P. exigua*) on the west coast also occur in the cold water regions of Australia and New Zealand (Brown and Jarman 1978). The topography and level of exposure may also cause some species to occur in areas outside their normal biogeographic boundaries, but overall the data show distinctly different species sets in each zone.

1.4. Asterinid life histories

Life history strategies of asterinids are among the most diverse in any marine invertebrate group (Byrne 1992), and it has been suggested that many parallel changes in larval form, habitat and dispersal potential have occurred in the evolution of this family (Byrne 1992, 1995, 1996; Chen and Chen 1992; Hunt 1993; Hart *et al.* 1997; Hart 2000; Jeffery *et al.* 2003*).* Small genetic distances between lineages with different developmental modes suggest that some of these changes have been recent or rapid (Hart 2000). Table 1.1 shows the developmental patterns and larval types in the Family Asterinidae (Byrne 2006) and Figure 1.6 shows a maximum parsimony phylogenetic tree showing relationships of asterinid mtDNA sequences in relation to their developmental modes (Byrne 2006).

Patiriella regularis shows what is considered to be the ancestral form of development for asteroids, developing through planktotrophic bipinnaria and brachiolaria larvae

(Byrne 1995). Benthic lecithotrophic development, as in *Parvulastra exigua,* and viviparity in *P. vivipara* and *P. parvivipa*ra are considered to be derived features (Byrne 1995). However, complex feeding planktonic larvae appear to have been lost at least four times during the evolution of these species. There does not appear to be an ordered transformation series from feeding planktonic development to viviparous brooding, instead life history traits of these sea stars appear to have evolved freely under no obvious constraints, contrary to the widely assumed evolutionary conservatism of early development (Hart *et al.* 1997).

Hart *et al.* (1997) discuss the evolution of hermaphroditism in *Parvulastra* species and suggest that it has evolved several times from dioecy. Byrne (1996) suggests that species which have life history traits conducive to inbreeding, such as hermaphrodites, are generally derived from outbreeding taxa, however due to the potential for inbreeding depression in species with limited dispersal, it is hypothesised that self fertilization is less common than sexual reproduction. Lawrence and Herrera (2000) propose that hermaphroditism in echinoderms is a derived, adaptive, reproductive characteristic and that it may have evolved in response to stress in environmental conditions to increase the probability of successful reproduction. Inter-specific cross-fertilisation to produce viable hybrids is not unusual in asteroids and hybrid zones are evident between some species (Byrne and Anderson 1994). Multi-species spawnings occur in several sympatric echinoderms, which may facilitate gamete fusion and therefore hybridization and gene flow (Byrne and Anderson 1994).

Genus/species	Egg diameter (μm)	Spawning/fertilization	Dev. type	Larval type(s)
Asterina				
A. gibbosa	500	Benth. eggs	BL	Bilobed Brach.
A. phylactica	500	Broods benth. eggs	BL	Bilobed Brach.
A. stellifera	150	Broadcasts	Pt	Bip. and Brach.
Stegnaster				
S. inflatus	1000	—	L	—
Paranepanthia				
P. aucklandensis	400	—	L	—
P. grandis	800	—	L	—
Patiria				
P. miniata	169	Broadcasts	Pt	Bip. and Brach.
P. chilensis	160	—	Pt	—
P. pectinifera	170	Broadcasts	Pt	Bip. and Brach.
Meridiastra				
M. mortenseni	240	Broadcasts	Pt	Brach.
M. calcar	413	Broadcasts	PL	Brach.
M. oriens	400	Broadcasts	PL	Brach.
M. occidens	400	Broadcasts	PL	Brach.
M. gunnii	430	Broadcasts	PL	Brach.
M. atyphoida	400	—	L	—
Parvulastra				
P. exigua	390	Benth. eggs	BL	Tripod brach.
P. vivipara	150	Ig	IgL	Reduced
P. parvivipara	235	Ig	IgL	Reduced
Patiriella				
P. regularis	150	Broadcasts	Pt	Bip. and Brach.
Aquilonastra				
A. batheri	433	Broadcasts	PL	Brach.
A. burtoni	550	Broadcasts	PL	Brach.
A. coronata japonica	422	Broadcasts	PL	Brach.
A. minor	437	Benth. eggs	BL	Tripod brach.
A. scobinata	450	—	L	—
Aquilonastra new sp (Qld)	420[b]	—	L	—
Cryptasterina				
C. pacifica	400	Ig	IgL	Brach.
C. pentagona	413	Broadcasts	PL	Brach.
C. hystera	440	Ig	IgL	Brach.
Cryptasterina sp (Taiwan)	320	Broadcasts	PL	Brach.
Cryptasterina new sp #1 (Qld)	440[b]	Ig	IgL	Brach.
Cryptasterina new sp#2 (Qld)	380[b]	—	PL	—

[a]Data from MacBride (1896); James (1972); Komatsu (1975); Kano and Komatsu (1978); Komatsu and colleagues (1979); Marthy (1980); Emlet and colleagues (1987); Chen and Chen (1992); Chia and colleagues (1993); Byrne and Cerra (1996); and Byrne and colleagues (2003); personal communications from M. Barker, M. Fernandez, D. McClary, and R. Ventura.
[b]Data from unspawned eggs in gonad. BL, Benthic lecithotroph; Benth, benthic; Bip, bipinnaria; Brach, brachiolaria; Dev., development; Ig, intragonadal; IgL, intragonadal lecithotroph; L, lecithotroph; Pt, planktotroph; PL, planktonic lecithotroph; Dashed line, no data.

Table 1.1. Developmental patterns and larval types in the Family *Asterinidae* (from Byrne 2006). Definition of terms: Benthic: bottom dwelling; Lecithotroph: larvae that do not feed, but rather derive nutrition from the yolk; Bipinnaria: the complex, bilaterally symmetrical free swimming larval stage of most echinoderms; Brachiolaria: an early larval stage having bilateral structure and swimming by means of bands of vibrating cilia; Intragonadal: within the gonads; Planktotroph: larvae that feeds during the planktonic phase.

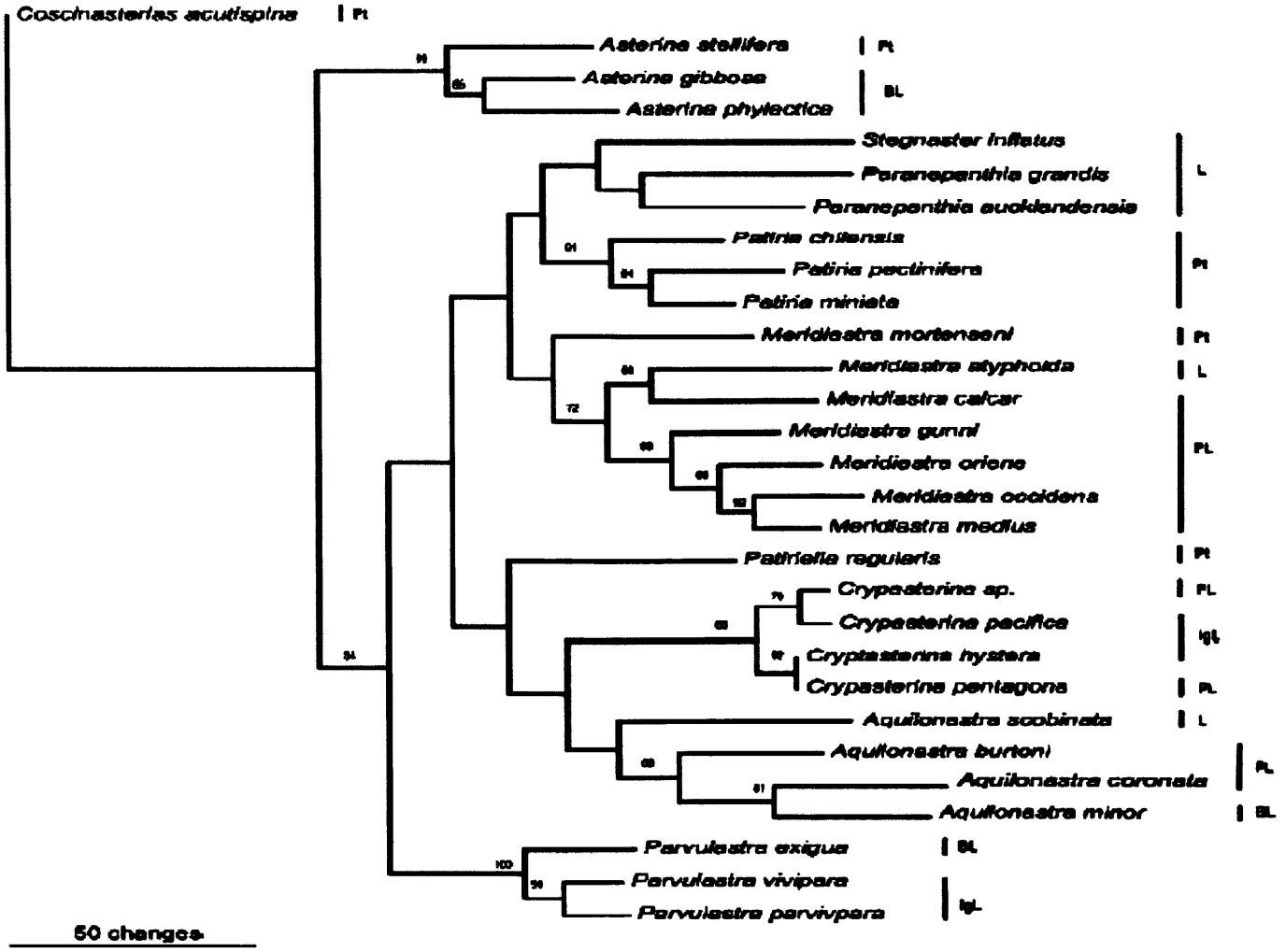

Figure 1.6. Maximum parsimony phylogenetic tree showing relationships of asterinid mtDNA sequences and developmental modes (see Table 1.1 above). Bootstrap values > 50% are indicated. Abbreviations for developmental mode as in Table. 1.1 above (from Byrne 2006).

The reproductive biology of Australian populations of *P. exigua* has been studied extensively (Byrne 1995; Hart *et al.* 1997; Waters and Roy 2004; Byrne 2006), but there are no comparable studies of South African populations, therefore the following description of the life history of *P. exigua* is based largely on the Australian populations. It is assumed that as the two populations are presently classed as a single species (but see Section 1.3.2) they have the same reproductive biology (Branch and Griffiths 1994; Waters and Roy 2004) despite differences in phenotype (A. Hewitt pers. comm. 2002) and mtDNA genotype (see phylogeography Section 1.5) (Waters and Roy 2004). The only three lines of evidence on South African *P. exigua* reproductive mode come from (a) photographic evidence of an unmottled phenotype in South Africa lying close to an egg mass (Fig. 1.7) (C. Griffiths pers. comm. 2001), (b) gonopore position included in the original species description Lamarck (1916) (see discussion on taxonomic confusion over gonopore location in Section 1.3.2) and (c) a sea-shore identification book (Branch and Griffiths 1994), which states that *P. exigua* hatch directly into tiny juvenile starfish. The latter two comments may indicate that the South African populations have a different mode of development to the Australian population, but this will be investigated further during the course of this study. It is assumed that the mottled variety in South Africa has the same reproductive biology as the unmottled phenotype in Australia (and South Africa), however this has not been confirmed.

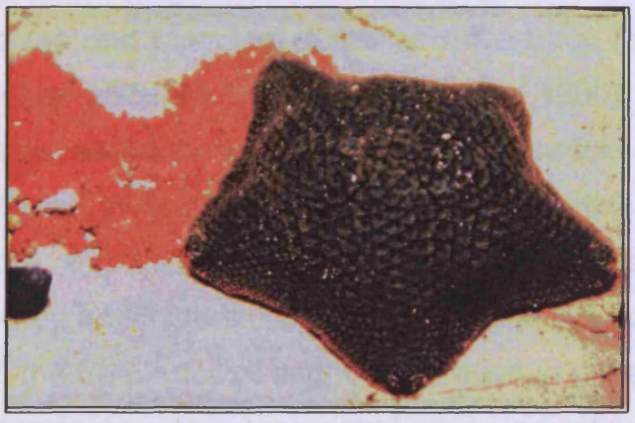

Figure 1.7. *Parvulastra exigua* (unmottled phenotype) possibly ovipositing eggs onto the substratum (photograph C. Griffiths)

Parvulastra exigua releases large yolky (Byrne 1992) eggs (390 μm dia.) through orally directed gonopores. Gravid ovaries are large and yellow whereas spent ovaries are small and yellow or brown. The sperm ducts also open on the oral surface and, where fertilization occurs externally, males appear to fertilize the eggs soon after deposition (Byrne 1992). The eggs have a sticky jelly coat, which enables them to adhere to the substratum. Clutch sizes are generally small (10^2 -10^3) (Hart *et al.* 1997). Such large eggs and lack of an exogenous food source are considered to be derived features (Byrne 1995), and development proceeds without parental care (Byrne 1995).

When first laid the fertilization membranes are transparent but after one day take on a red hue. Groups of *P. exigua* have been observed gathering round newly laid egg masses in what appear to be spawning groups. Egg masses at different stages of development have been observed together along with newly metamorphosed juveniles in microhabitats. The size of some egg masses also suggests that they were deposited by more than one female (Byrne 1992). *Parvulastra exigua's* large red benthic egg masses are very conspicuous on the rocky shore, which is a consumer rich habitat where predation pressure is great. Moreover, the eggs appear to be very palatable because of their large size, yolky consistency and bright red colour and therefore high visibility, but they appear to lack structural or morphological defenses against predators. These features of *P. exigua's* eggs may indicate that they may be unpalatable due to chemical defenses and therefore be protected from predation. Secondary metabolites and chemical defenses have been reported in asteroid eggs previously (Lucas *et al.* 1979; McClintock and Vernon 1990), as well as other marine invertebrates such as ascidians and nudibranchs (Pawlick *et al.* 1988; Young and Bingham 1997). Marine invertebrate larvae that are benthic or brooded have been shown to be more unpalatable than larvae of broadcast spawners as it is hypothesised that they need greater protection against predators (Lindquist and Hay 1996). Additionally, the frequency of bright colouration in unpalatable larvae and eggs is high in comparison to palatable larvae and eggs, indicating potential aposematism (warning coluration), (Lindquist and Hay 1996). The Spanish Dancer nudibranch has

bright pink egg masses which have been shown to be unpalatable and chemically defended against fish predators (Pawlick *et al.* 1988), as have the bright orange egg masses of *Ectyoplasia,* both of which derive their chemical defenses from their diets (Pawlick *et al.* 1988). Brightly coloured red, (as seen in *P. exigua*) orange or yellow pigments have been shown to be common in unpalatable marine larvae and eggs (Lindquist and Hay 1996).

The larvae of *P. exigua* develop into lecithotrophic (without feeding) modified brachiolaria larvae (similar to planktonic larvae) and remain benthic until metamorphosing into juveniles (Fig. 1.8). The larvae have a muscular stem with a central adhesive disk, which enables them to remain attached to the substratum. The transformation of the larval form to the juvenile takes six to seven days, about half the duration of larval life. This is much more gradual that in other *Parvulastra* species with planktonic development (Byrne 1995). The dominant stage of development in *P. exigua* is the brachiolaria stage which persists for about 15 days irrespective of temperature (Byrne 1995). There is no planktonic dispersal phase, but instead the juveniles are negatively geotactic and float to the surface attaching themselves to the surface waters. This behaviour may act as a dispersal mechanism by rafting. The length of time that *P. exigua* can raft is unknown, but the juveniles metamorphose after two - three weeks into adults (Byrne 1995; Waters and Roy 2004).

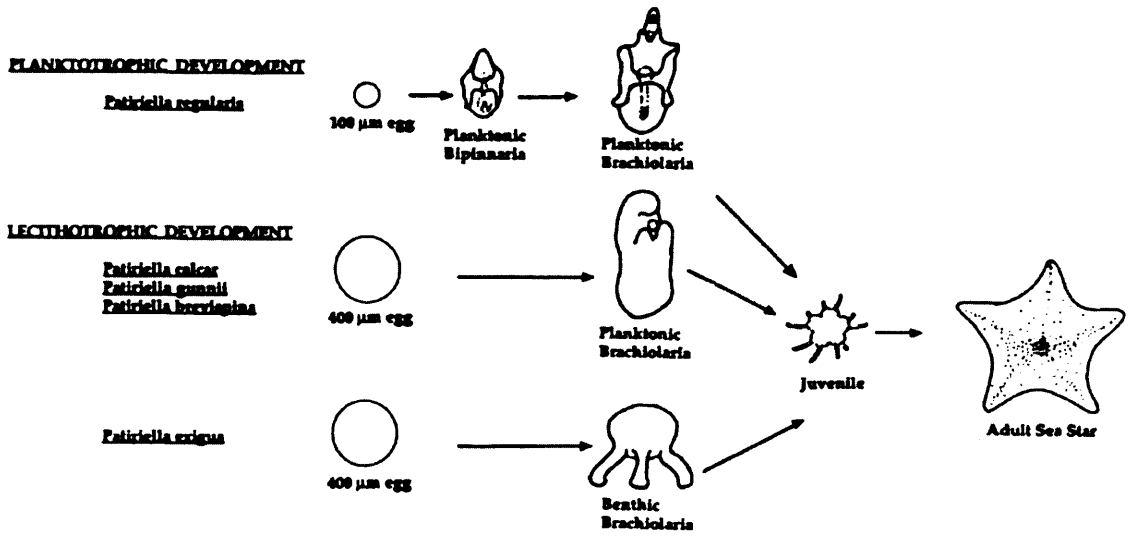

Figure 1.8. The planktonic and lecithotrophic stages of some *Parvulastra* species life-cycles (*Patiriella calcar, P. gunnii and P. bervispina* now in the family *Meridiastra*), (from Byrne 1995).

Parvulastra exigua is a protandrous hermaphrodite (male gametes mature and are shed before female gametes mature). Some *P. exigua* are simultaneous hermaphrodites from the outset of gonad formation and become increasingly female with growth (Byrne 1992) with large females often having a minute amount of sperm in their gonads (Byrne and Anderson 1994). They therefore have the potential for self-fertilization and have been recorded laying self fertilised eggs (Byrne 1995). It has been suggested that the decreased male investment in *P. exigua* is associated with its non-planktonic development (Byrne 1992). Hunt (1993) suggested that there is no evidence indicating that asexual reproduction occurs in *P. exigua* (but see Byrne and Anderson 1994) as genotype frequencies from Australian populations are in Hardy-Weinberg equilibrium. Gametogenesis in *P. exigua* continues throughout the year, but with enhanced oviposition during the colder months of winter and spring in Australia. An abundance of mature females has also been observed in June, July and September, indicating that there are definite periods of increased spawning. Fertilizable gametes may be obtained from females for at least nine months of the year. Juveniles were located in the intertidal zone from August to October (Byrne 1992). It has been suggested that enhanced oviposition in winter coincides with cooler sea temperatures, shorter photoperiods and a higher oxygen level in the water. This breeding cycle may be in response to optimal conditions for egg development (Byrne 1992).

The prolonged breeding period of *P. exigua* disguises potential environmental factors that may stimulate reproduction (Byrne 1995). In another *Parvulastra* species, *P. vivipara*, Byrne and Cerra (1996) suggested that temperature does play a key role in spawning time and duration. Temperature effects on the temporal breeding pattern in *P. exigua* may be important as the South African population occurs along a steep coastal temperature gradient and so individuals at the peripheries of the range will be exposed to very different environmental temperatures and potentially populations at either end of the temperature range could have different breeding cycles. Other

environmental factors, such as food availability, energy budgets (Moreno and Hoegh-Guldberg 1999; Lawrence and Herrera 2000) and habitat complexity for larval and juvenile settlement (Chen and Chen 1992) may also effect the reproductive cycles of *Parvulastra* species. In Australia, *P. exigua* occurs sympatrically with *Meridiastra gunnii* and *M. calcar* where they are found in close proximity in pools and under boulders (Byrne and Anderson 1994). *Parvulastra exigua* has a different spawning time and a different ontogeny to the other two species, which have the same spawning time and very similar ontogeny to each other. Hybridization between *M. gunnii* and *M. calcar* resulted in phenotypically intermediate viable hybrid juveniles *in vitro*. However, heterospecific crosses involving *P. exigua* resulted in low fertilisation rates and no viable hybrids *in vitro* (Byrne and Anderson 1994), but no information is available on the ability of *P. exigua* to hybridise with other species in the wild.

1.5. Parvulastra phylogeography

Parvulastra exigua has an entirely benthic life history except when juveniles float to the surface when rafting, which may be a dispersal mechanism. Yet it occurs throughout southern Africa, on two oceanic islands in the Southern Indian Ocean, on St. Helena in the South Atlantic, south eastern Australia and Lord Howe Island (Fig. 1.9; Waters and Roy 2004) and in the Andaman Sea (Bussarawit and Hansen 1987, cited in Putchakarn and Sonchaeng 2004), but is absent from New Zealand and South America. Australian *P. exigua* populations occupy a cold water biogeographic zone, and are also an unmottled greenish brown phenotype, and the mottled *P. exigua* (as seen in South Africa) phenotype does not occur in Australia.

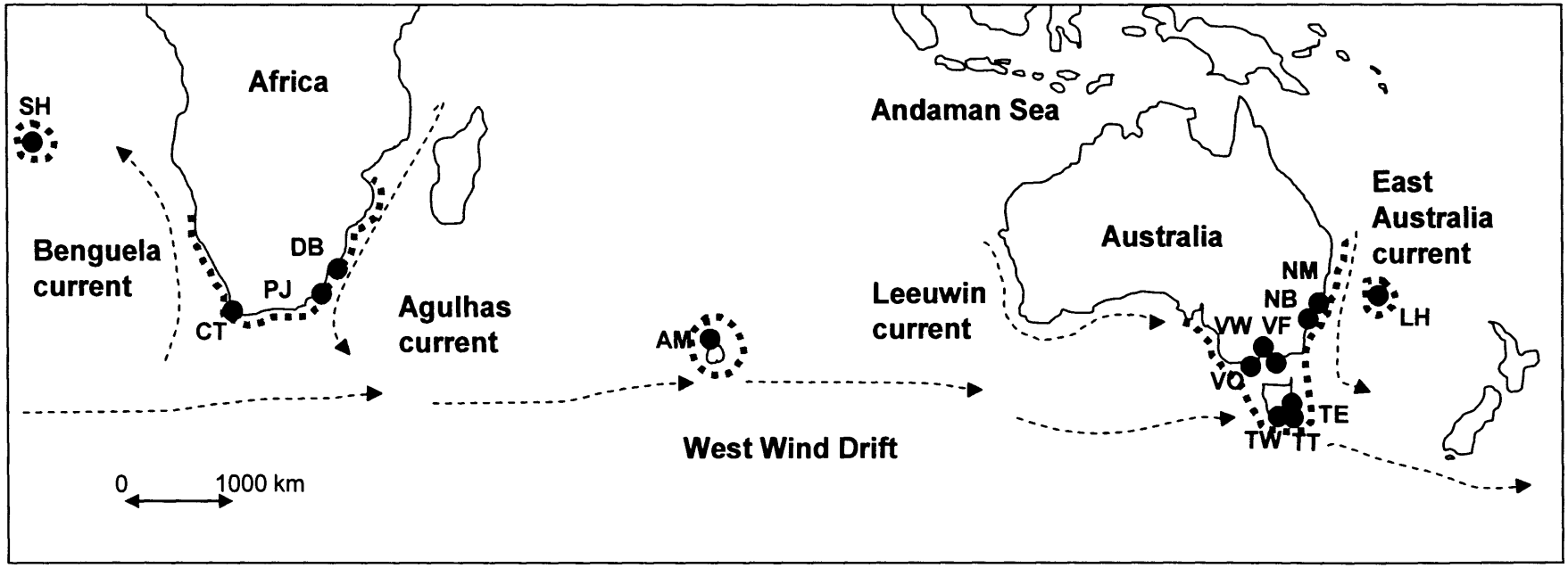

Figure 1.9: Geographic distribution of *Parvulastra exigua* in the Southern Hemisphere with known records indicated by grey dotted lines and Waters and Roy's (2004) sampling sites marked by black dots (SH: St. Helena Island; CT: Cape Town; PJ: Port St. Johns; DB: Durban; AM: Amsterdam Island; VO: Victoria Cape Otway; VW: Victoria Williamstown; VF Victoria Flinders; TW: Tasmania Woodbridge; TT: Tasmania Taranna; TE: Tasmania Eaglehawk Neck; NB: New South Wales Bondi; NM: New South Wales Mona Vale; LH: Lord Howe Island). Also illustrated on the map are major ocean currents (adapted from Waters and Roy 2004).

This distribution is unusually large for a species with limited or sporadic dispersal, which are generally hypothesised to have high inter-population genetic variation (Riginos and Victor 2001; Sponer and Roy 2002). *Parvulastra exigua's* sister species, *P. dyscrita* is suspected to have a planktonic stage to its life cycle (Branch and Griffiths 1994), and is therefore hypothesised to have a larger geographic range and little genetic structure (Booth and Ovenden 2000). Yet this species has a smaller sympatric range within *P. exigua*'s South African distribution and is endemic to South Africa. It is not known how long the planktonic larval stage of *P. dyscrita* lasts, but the duration and distance potential this species has in the plankton may be limited by its metamorphosis into adulthood and the need to settle on substratum, sparse habitat suitability or current systems around the south coast, preventing *P. dyscrita* achieving the geographic range of *P. exigua* in Southern Africa.

Parvulastra exigua's seemingly paradoxical southern hemisphere distribution has been investigated previously (Byrne 1995; Hart *et al.* 1997), as well as the distribution of *P. exigua* on a local (continental) scale (Hunt 1993; Colgan *et al.* 2005) to determine the genetic structure and limits to gene flow of this species. Adult and juvenile rafting on wood or macro algae is the main mechanism proposed for dispersal (Mortensen 1933; Fell 1962; Clark and Downey 1992; Hart *et al.* 1997; Waters and Roy 2004). However, the geographic distribution of a species is not defined by dispersal alone and Dartnall (1971) suggested that the distribution of *P. exigua* was "defined by thermal tolerance" (see Section 1.2.2 and Fig. 1.3). The distance covered by rafting *P. exigua* adults is unknown but it is thought that passive dispersal by rafting of both adults and juveniles can occur for many months in the open ocean (Waters and Roy 2004).

Waters and Roy (2004) investigated the possibility of long distance rafting in *P. exigua* by examining mitochondrial DNA Cytochrome Oxidase I and control region sequence variation. They found that the South African population had a paraphyletic assemblage, with South African haplotypes occurring in both the South African haplo-group and the Australian haplo-group, whereas Australian samples displayed a monophyletic group with Australian haplotypes occurring only in the Australian haplo-

group (Fig. 1.10). Waters and Roy (2004) concluded that *P. exigua* originated in South Africa and colonisation of Australia was achieved by a single colonisation event via the west wind drift current, subsequent population expansion was then facilitated by self fertilisation. Colonisation of Amsterdam, Lord Howe and St. Helena Islands might have been achieved in the same way during the late Pleistocene (<0.5 MYA). However, Waters and Roy (2004) based their investigation on just 11 South African samples and 16 Australian samples and they found no shared haplotypes between the South African and Australian populations. More extensive sampling might reveal additional mtDNA lineages and could reduce the divergence estimates. The 'Out of Africa' hypothesis is considered controversial by some authors (Colgan *et al.* 2005; Hart *et al.* 2006) on grounds that it may be based on a cryptic species complex (see taxonomy section), or that the *P. exigua* distribution may be caused by vicariant events, e.g. the break up of Gondwanaland (Heads 2005a, b) and not dispersal (Waters and Roy 2004) (see Chapter 2).

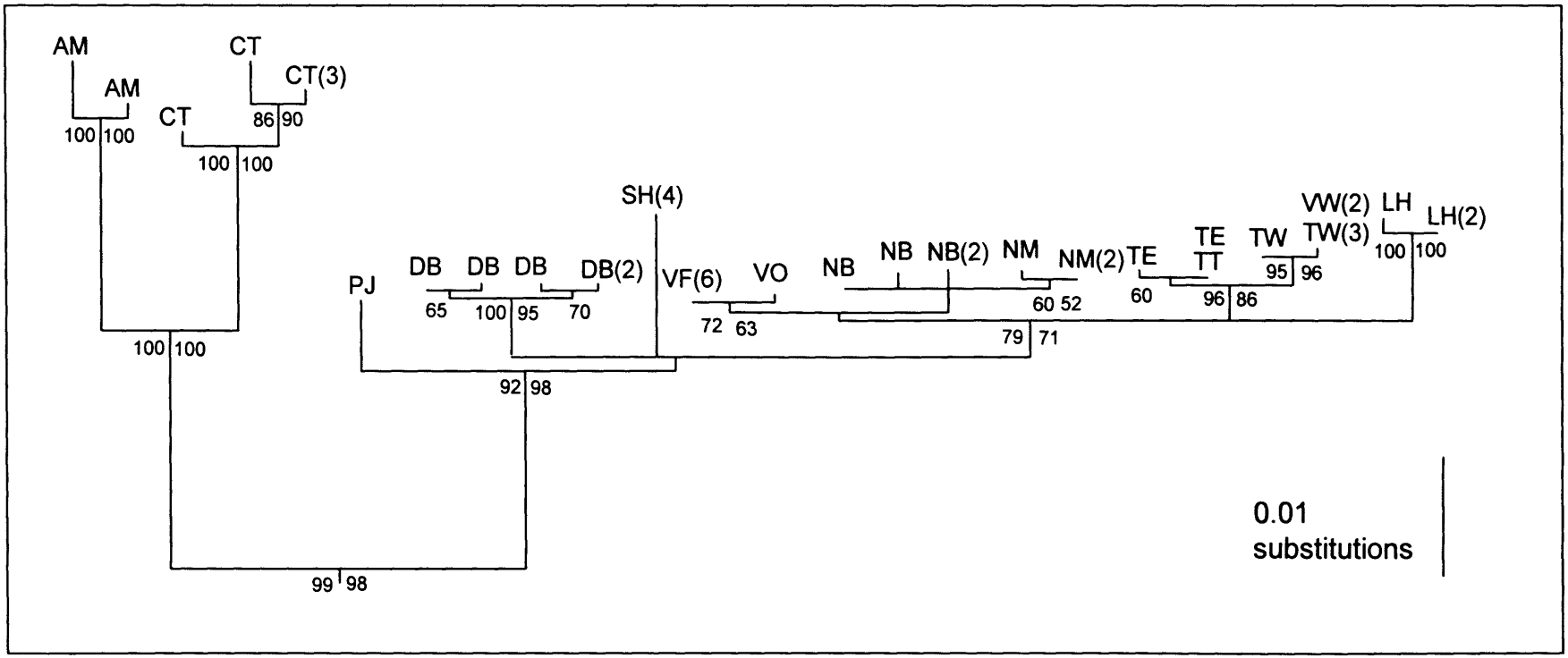

Figure 1.10: Phylogenetic relationships of *Parvulastra exigua* haplotypes indicated by location code on map (Fig. 1.9; SH: St. Helena Island; CT: Cape Town; PJ: Port St. Johns; DB: Durban; AM: Amsterdam Island; VO: Victoria Cape Otway; VW: Victoria Williamstown; VF Victoria Flinders; TW: Tasmania Woodbridge; TT: Tasmania Taranna; TE: Tasmania Eaglehawk Neck; NB: New South Wales Bondi; NM: New South Wales Mona Vale; LH: Lord Howe Island). Numbers in parentheses indicate the number of times a haplotype was recorded at a site. Bootstrap estimated to the left of nodes are derived from maximum likelihood analysis, and values to the right are from minimum evolution analysis (from Waters and Roy 2004).

The phylogeography, population structure and dispersal of *P. exigua* has also been investigated at a more local (continental) scale. Hunt (1993) used allozymes to investigate the population structure of *P. exigua* with its entirely benthic life cycle in comparison to a sympatric species *Patiriella calcar* (now *Meridiastra calcar*, see Fig. 1.1), which has planktonic larvae, which both occur in south east Australia. Hunt (1993) showed that over 230 km distance, *P. exigua* had a finer scale genetic structure than *M. calcar*, indicating that *P. exigua* has low inter-population immigration, as expected of a species with low dispersal ability. This pattern of genetic structure in the Australian populations of *P. exigua* was supported by Colgan *et al.* (2005) using mtDNA who suggested that there was very low migration between populations around the south eastern Australian and Tasmanian coasts, and that complete lineage sorting to regions had occurred. However, they also suggested that the evolutionary divergence between the regions was low, indicating that there has been a recent range expansion of the ancestor of the haplotypes currently observed in Australia. Subsequent gene flow was inferred to have been so restricted that there has been no sharing between regions of new haplotypes produced by mutation. Colgan *et al.* (2005) conclude that the range expansion could either be due to colonisation of Australia by a low number of animals with identical haplotypes (i.e. a founder event) as suggested by Waters and Roy (2004) or to a selective sweep through a pre-existing population by a favourable mutation. The population genetic structure and genetic demographic history of the South African *P. exigua* populations has never been investigated.

1.6. Molecular approaches

By measuring molecular genetic variation and by applying population genetic models, evolutionary ecologists can make inferences about the biology, phylogeography and demographic history of organisms (Sunnucks 2000). Previously single locus approaches have been widely used to investigate population genetics and locus specific effects (e.g. selection, mutation). Single

locus techniques investigate specific genes or non-coding regions in either the mitochondrial (e.g. control region, Cytochrome Oxidase I gene) or the nuclear genome (e.g. ribosomal DNA internal transcribed spacer regions), or investigate allele frequencies of highly polymorphic loci (e.g. microsatellites). The single locus approach has several benefits: Low quantities of DNA can be used, data are comparable across taxa and it is relatively cheap and easy (Sunnucks 2000). Sequence data can detect both silent (i.e. synonymous changes that do not alter the amino acid) and non-silent changes (i.e. non-synonymous changes that result in a change in the amino acid) (Singh 2003). Markers giving both allele and haplotype frequency and sequence data can be informative over a range of timescales. However, analysing only a few loci can provide an incomplete or biased view of the genome and the population history. Recently, approaches looking at the whole genome have become more accessible. The genomic approach is an emerging discipline in non-model organisms and it combines genomic concepts and technologies with the population genetics objective of understanding evolution (Luikart *et al.* 2003).

1.6.1 Single Gene approaches

1.6.1.1 Mitochondrial sequence markers

Animal mitochondrial (mt) DNA is a small closed circular double stranded DNA molecule approximately 15 - 17 kb, but this varies among taxa (Ballard and Whitlock 2004; but see Burger *et al.* 2003). In most species it encodes 37 genes, 24 of which encode the translation machinery of the mitochondrial DNA itself (Ballard and Whitlock 2004; Ballard and Rand 2005; but see Burger *et al.* 2003). The MtDNA gene content is considered conserved, and it contains no introns, intergenic sequences or interrupted genes (Moritz 1987) and its structure and organization have been well studied. Although in most mitochondrial genomes the gene content remains the same (but see Moritz 1987), there are several examples of large length variations in many phyla (reviewed by Burger *et al.* 2003), resulting from insertions in the control region, duplication or deletion of sequences or replication slippage (Moritz 1987). Gene arrangement also varies

in some taxa. There has been a major inversion in the mtDNA of sea stars compared to sea urchins of a 4.6 kb portion of the tRNA cluster, the 16S rRNA, and the ND1 and ND2 genes (Fig. 1.11). These two orders have been separated for at least 225 million years and so this rearrangement occurred after the separation as it does not occur in both orders. The sea urchin mtDNA gene arrangement is similar to that of the vertebrates with respect to gene order (Smith 1993).

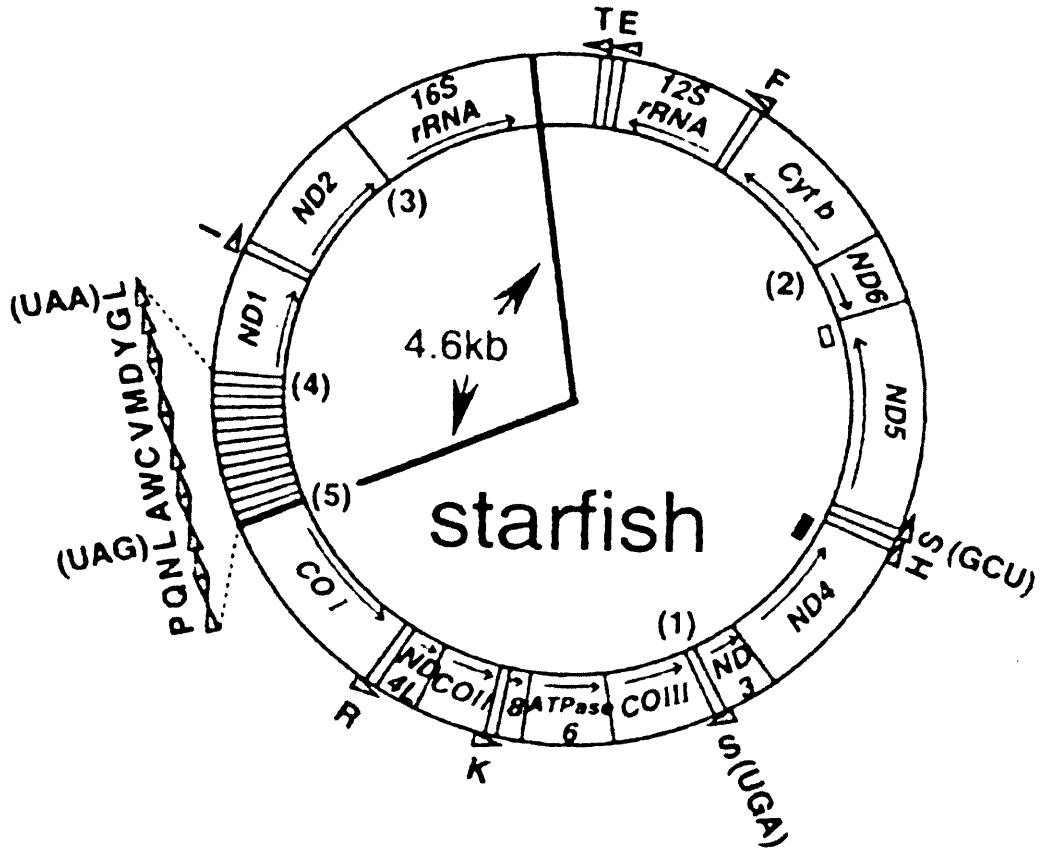

Figure 1.11. The sea star (starfish) mitochondrial genome showing the inversion of the gene arrangement between sea star and sea urchin (from Asakawa *et al.* 1995).

The mitochondrial DNA molecule is haploid meaning that there is only one copy unlike nuclear DNA which has two copies of the genome. The major implication of the mitochondrial molecule being haploid is that because it is normally passed

on only by half of the population through maternal inheritance (but see below) the effective population size when examining mitochondrial markers is only 0.25 the size of nuclear markers. This means that in most species, mitochondrial DNA would fix new alleles faster than nuclear DNA (summarized by Ballard and Whitlock 2004). Furthermore, most evolutionary rates for mtDNA are higher compared to nuclear genes (Birky *et al.* 1983), therefore making it possible to examine differentiation in populations or recently diverged species (Rokas *et al.* 2003). However, it is unsafe to assume that all evolutionary rates in the mtDNA genome are consistently higher than the nuclear genome. There are several factors that can affect the mtDNA evolutionary rate, including effective population size, mutational biases, availability of nucleotides in the cellular medium of the mitochondria and selection (Ballard and Whitlock 2004). Over or under-estimating the rate of evolution in mtDNA can lead to inaccurate phylogenetic inferences.

Within the mitochondrial genome itself, different rates of evolution of different mitochondrial genes and even parts of genes are observed, and this also varies across taxa. The ribosomal RNA genes (rRNA) within the mitochondrial genome are much simpler than the nuclear rRNA genes and have been used extensively because of their critical role in protein assembly, their universal occurrence and their sequences and secondary structure conservation. Their use in phylogenetic analysis can cover a wide range of divergence levels including the deepest levels of divergence. Due to the secondary structure of the ribosomal genes the rate of evolution of the rRNA genes varies considerably along the length of the molecule. Highly conserved nucleotide sites are associated with sites of ribosomal protein attachment, messenger RNA processing, tRNA attachment and core helices (Simon *et al.* 1994). Mitochondrial transfer RNA genes (tRNA) are also structurally and functionally constrained and therefore evolve more slowly than the mitochondrial protein coding genes. Again, like rRNA genes, different regions within the tRNA evolve at different rates, with the anticodon (AC) loop being most conserved and the ribothymidine pseudouridine cytosine (T)

loop and T stem being the most variable regions. The rate of evolution varies among as well as within tRNA's in different taxa (reviewed in Simon *et al.* 1994). These rate differences could result because tRNA's that are adjacent to protein coding genes may play a punctuating role in protein processing.

The mitochondrial protein coding genes, including the Cytochrome Oxidase I gene, the gene under investigation in this study, are also used extensively in phylogenetic studies and in general are useful for examining divergence at the species or recently diverged species level. The main difference between mitochondrial protein and RNA genes is that protein genes possess a triplet code for the assembly of proteins. This function places strong constraints on nucleotide changes at first and second codon positions, but because of the degenerate nature of the amino acid code, many third, and some first codon positions are less constrained and have been observed to evolve at a higher rate. Substitutions at these positions are termed synonymous substitutions as they do not cause amino acid changes. Most mitochondrial genomes have a non-coding region called the control region (although the size and number of control regions can vary across taxa) which surrounds the origin of replication of the molecule (Simon *et al.* 1994). As these regions are non-coding they are more variable than other mitochondrial regions as they are not constrained by function. Due to this hyper-variability, the control region has been used in many phylogenetic studies and is useful for looking at the population or below species level in vertebrates (reviewed in Simon *et al.* 1994). However, in invertebrates this region can also be extremely variable in length, probably caused by the presence or absence of repeated sequence blocks.

Mitochondrial DNA occurs in every cell of the organism in multiple copies making it abundant and easy to amplify, even from small amounts of starting DNA. Pioneered by Avise *et al.* (1987) for the past three decades mtDNA has been used as a marker in thousands of population and evolutionary biology studies (reviewed by Ballard and Rand 2005). The reasons for its success as a

phylogenetic marker are based on it being relatively easy to use in a wide variety of species, as well as the theoretical considerations outlined below. However, most, if not all of the theoretical assumptions about mitochondrial DNA have been questioned and in some cases disproved (outlined below).

1.6.1.1.1. Mitochondrial heteroplasmy

Although normally only one copy of the mitochondrial genome is present in all cells within an individual, heteroplasmy (where within a single cell there is a mixture of different mitochondrial haplotypes) (Bromham *et al.* 2003) has been detected for many animal and plant species (reviewed by Kmiec *et al.* 2006). Heteroplasmy can arise through paternal leakage, recombination or small scale mutations (Kmiec *et al.* 2006), and can lead to unreliable population or evolutionary inferences.

1.6.1.1.2. Paternal inheritance of the mitochondrial genome

In homogametic species, mtDNA normally is passed on only by the female, and as such it reflects matriarchal phylogenies in most species, providing an excellent marker for tracking patterns of colonization and founder events, which will not be influenced by male biased dispersal events (Harrison 1989). However, paternal inheritance (Bromham *et al.* 2003) has been shown to occur convincingly in mussels (Ladoukakis and Zouros 2001), some invertebrate species (Rokas *et al.* 2003), vertebrate cross species hybrids (Gyllensten *et al.* 1991; Kvist *et al.* 2002) and even humans (Kraytsberg *et al.* 2004). Mussels have 'doubly uniparental inheritance' (i.e. female offspring inheriting their mother's mtDNA and male offspring inheriting both parents mtDNA) leading to heteroplasmic offspring (Rokas *et al.* 2003). The extent of this phenomenon across species is not known however it is thought to be restricted to a few exceptions (Bromham *et al.* 2003). However, the above discussion focuses on species which have separate sexes.

In *P. exigua* the organism starts life as a male and turns into a female. Therefore the mitochondrial genome in the male and female part of the life cycle will be the same, negating the effects of maternal transmission in terms of population phylogenies.

1.6.1.1.3. Recombination in mitochondrial DNA

MtDNA lacks recombination and therefore is not affected by locus-specific effects (Luikart 2003; Rokas *et al.* 2003). However, in recent years recombination in animal mtDNA has become a hotly debated topic (Eyre-Walker *et al.* 1999; Macaulay *et al.* 1999, reviewed by Piganeau *et al.* 2004; Eyre-Walker 2000; Eyre-Walker and Awadalla 2001; Rokas *et al.* 2003). Some animal species do show evidence of mtDNA recombination which can cause serious problems when for example constructing phylogenies, inferring expansion events or dating most recent common ancestors (Rokas *et al.* 2003; Piganeau *et al.* 2004;). Furthermore, Rokas *et al.* (2003) suggest that the knowledge base of how common mtDNA recombination is in the animal kingdom is poor. The biochemical structures necessary for mtDNA recombination are present in the mitochondria (Ballard and Whitlock 2004). Since most approaches to phylogenetic analysis assume no recombination, mtDNA is an ideal phylogenetic marker. However, this lack of recombination means that when one part of the molecule is affected, it directly influences all other parts of the molecule, meaning that independent replication of data about a population history cannot be achieved using different sections of the mtDNA genome (Ballard and Whitlock 2004).

1.6.1.1.4. Mitochondrial DNA as a neutral marker

For many years mtDNA was considered a neutral marker and not affected by selection or mutation and therefore could be used to examine gene flow, biogeography, estimates of coalescence times and other stochastic effects (Ballard and Rand 2005). However, the assumption that mitochondrial DNA is neutral only reflects its 'apparent' neutrality. Non-synonymous base changes in

the 1st or 2nd codon are often deleterious mutations which are disfavoured by selection. This results in mainly detecting only 3rd base mutations when screening populations (which do not change the amino acid), giving the false impression that mutations in the mitochondrial DNA are neutral and not affected by selection. The assumption that mtDNA is a neutral marker has been addressed in the context of hybrid zones (Moritz 1987) and for direct and indirect selection acting on the mtDNA itself (reviewed in Ballard and Rand 2005). Ballard and Whitlock (2004) reviewed the direct impact of mitochondrial haplotypes on fitness in copepods, mice, *Drosophila* and very extensively in humans. In the latter two taxa, there is evidence that environmental temperature could impact on mtDNA fitness (Nagata and Matsuura 1991; Mishmar *et al.* 2003). These authors suggest that mtDNA variants (particularly in the ATPase 6 gene) that reduce the coupling efficiency of oxidative phosphorylation would reduce ATP production but increase heat production. Somero (2002) also noted that mtDNA may be under selection for thermal tolerance as the external temperature in poikilotherms is experienced by the mitochondria and the relative fitness of the different genotypes is likely to change as a result. Given the potential for temperature variation across species ranges, including the South African *Parvulastra* species, temperature may play a role in selection acting on the mtDNA. Additionally, human mtDNA point mutations are often non-neutral and there is growing evidence that many diseases and phenotypes such as differences in sperm motility are caused by mtDNA mutations (Ballard and Whitlock 2004).

1.6.1.1.5. Mitochondrial gene genealogies not species genealogies

In the past it has been assumed that phylogenetic inference using mtDNA gives a true representation of a species or population demographic history, i.e. a mtDNA tree represents a phylogenetic tree and not a gene tree (Ballard and Whitlock 2004). However, the lack of recombination in mtDNA means that the entire molecule has the same history (or gene geneology), but this can give rise to errors in interpreting demographic histories since gene genealogies often differ from species genealogies (reviewed in Nichols 2001; Hudson and Turelli 2003).

MtDNA may be unusual in the overall genealogy of the species as it may be affected by a number of processes including (i) sampling over possible coalescent processes, i.e. having incomplete lineage sorting due to the most recent common ancestor existing before the species split resulting in taxa being paraphyletic and some lineages of mtDNA occurring in more than one taxon (Ballard and Whitlock 2004); (ii) greater sensitivity to certain processes such as introgression from one species to another, which can cause signals from previous events to be erased and result in a significantly different genealogy than for most genes in the species (Ballard and Whitlock 2004); (iii) molecule-specific vicariant events such as selective sweeps (increase in the frequency of an allele caused by genomic selection, leading to a local excess of rare alleles) thus obscuring the typical mtDNA genealogy for that species (Ballard and Whitlock 2004). These problems can only be overcome by replicate sampling of independent gene trees (i.e. independent markers from the nuclear genome) which can represent different samples for the evolutionary process and therefore independent estimates of the species tree (Ballard and Whitlock 2004).

1.6.1.1.6. Nuclear copies of mitochondrial DNA (*NUMTS*)

Mitochondrial DNA is also known to insert into the nuclear genome (nuclear copies of mitochondrial DNA, *NUMTS*). Recently it has been hypothesized that many phylogenetic studies based on mitochondrial DNA may in fact be using *NUMTS* instead of true mitochondrial DNA (reviewed in Bensasson *et al.* 2001). If *NUMTS* are sequenced instead of true mitochondrial DNA then no accurate phylogenetic inferences can be made about the evolution of the organism because *NUMTS* have dissimilar constraints on substitution. As mtDNA primers may amplify both mtDNA and *NUMTS* in the same organism, sequences which contain stop codons in coding genes indicate that *NUMTS* are present and that sequences must be cloned prior to the data being used for phylogenetic inference.

Despite the potential problems outlined above, the mtDNA molecule does provide an invaluable tool in phylogenetic studies and provides insights for some phylogenetic processes. However, the above discussion illustrates that mtDNA markers alone could prove misleading about population history and emphasizes the need to use several independent nuclear markers in conjunction with mtDNA to provide more accurate and reliable inferences about population history.

1.6.1.2 Nuclear sequence markers

Mitochondrial markers alone can provide an incomplete or biased view of a species demographic history therefore for many molecular ecology studies, the use of nuclear genes is advocated to give a more accurate representation of the overall structure (Lucchini 2003). Nuclear genes are thought to evolve more slowly than mitochondrial DNA and are therefore not as useful for intra-specific or recently diverged, closely related species (Hewitt 2004). Several neutral nuclear sequence markers have been used for population genetics, speciation or phylogeographic studies in marine invertebrates (Waters and Roy 2003; Waters *et al.* 2004b), including the cytoplasmic actin gene family introns, which have already been successfully used in a phylogenetic study of a *Patiriella* species complex (now the *Meridiastra* species complex, see Fig 1.1, Chapter 1), (Waters *et al.* 2004b). The cytoplasmic actin gene family play a role in cell structure and motility. As an intron, however, the actin intron is non-coding and is therefore assumed to be a neutral marker.

1.6.2 Genomic approaches

Population genomics exploit genome wide sampling to identify and to separate locus specific effects from genome wide effects such as drift, bottlenecks, gene-flow and inbreeding, to improve our understanding of microevolution (Luikart *et al.* 2003). Only genome wide effects inform us reliably about population demography and phylogenetic history, whereas locus specific effects help identify regions of the genome that are linked to traits involving fitness and

adaptation. The two main principles of genomics are that neutral loci across the genome will be similarly affected by demography and the evolutionary history of populations, and that loci under selection will often behave differently and therefore reveal outlier patterns of variation (Luikart *et al.* 2003; Chapter 4 and references therein). The ideal molecular approach for population genomics would be to score hundreds of polymorphic markers that cover the entire genome in a single simple reliable experiment (Luikart *et al.* 2003). The technique that comes closest to matching these criteria, and is financially feasible within the timescale of the present project is Amplified Fragment Length Polymorphism (AFLP; Vos *et al.* 1995).

1.6.2.1 Amplified Fragment Length Polymorphisms (AFLP)

Amplified Fragment Length Polymorphism (AFLP) was developed in 1995 (Vos *et al.* 1995) and has since has become widely used in plant, fungi and bacterial genetics and is becoming increasingly used in population and evolution studies in the animal kingdom (Bensch and Akesson 2005 and references therein) including one phylogeographic study of sea stars (Baus *et al.* 2005). AFLP provides a cheap and relatively low cost method for the screening of a large number of markers across the genome to approximate genome wide variation (Lucchini 2003; Bensch and Akesson 2005). It requires comparatively short start up times and as such is a viable method for screening large numbers of loci. The technique can be used on DNA of any origin and complexity and because it relies on PCR very little starting DNA is required. No prior sequence knowledge is needed and the technique is robust, reliable and reproducible (Bensch and Akesson 2005). It can also be used to compare genetic diversity across taxa, which can prove difficult with microsatellites.

The major disadvantage with the AFLP technique is that the per-locus type of genetic information is relatively poor and cannot provide complete genotypic information for diploid organisms (Bensch and Akesson 2005). AFLP produces

dominant data, meaning that presence or absence of a DNA fragment of a certain length can be detected at a given locus, but it is very difficult to separate between dominant homozygous (1/1, i.e. the allele being present in two copies) or dominant heterozygous (1/0, i.e. only one copy of the allele being present). In theory, dominant homozygous bands should be twice as strong as heterozygous bands, but in practice this is an unreliable method for detecting homozygosity or heterozygosity and therefore only presence or absence of a band can be recorded. A band can be absent from the data if there was a base substitution in the sequence corresponding to the restriction sites for the restriction enzyme, or in the sequence corresponding to the additional bases in the primers (see below) in the absent band but not the present band. However, several problems can arise in the interpretation of dominant data for population analyses. The assumption that the absent band really is absent from the data is used for most population analyses based on AFLP data, but this is not always the reason for the absence of the band. Other types of mutations e.g. indel variations or substitutions that create a new cut site for the restriction enzymes, may cause the DNA fragment to be a different length resulting in the bands being at a different position on the gel. If this occurs two alleles representing the same locus may be mistakenly scored as presence alleles at two different loci, thus violating the assumption in analyses of population structure and genetic diversity of independent replication (Bensch and Akesson 2005). This assumption can also be violated if size homplasy occurs i.e. bands of the same length are not homologous and therefore represent two or more different AFLP loci but appear on the gel to be one band.

The decision to use AFLP takes into account the trade off between poor per locus data and the integral biases when using dominant data (outlined above) with the ease with which numerous (>1000) loci can be generated. AFLP provides effectively the ability to screen the whole genome as opposed to the traditional and more common approaches of single locus techniques. AFLP can be used to screen hundreds of marker loci and address a variety of different

problems such as DNA fingerprinting and parentage analysis; population or species level genetic structure and genetic diversity estimates; conservation and management of species; assignment of migratory individuals to populations; identifying hybridization or introgression; gene mapping and linkage; species phylogenies; population or species level genetic structure and genetic diversity estimates; and identifying loci affecting phenotypes or under selection (reviewed in Bensch and Akesson 2005). The latter three of these applications will be applied to the South African *Parvulastra* species in this study.

The AFLP method is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos *et al.* 1995). The technique follows three steps. Firstly, DNA digestion generally using two restriction enzymes, a rare cutter and a frequent cutter, which results in the predominant amplification of the restriction fragments which have a rare cutter on one end and a frequent cutter on the other. Two restriction enzymes are used in this way because the frequent cutter will generate small DNA fragments that will amplify well and are in the optimal size range for separation on denaturing gels. The rare cutter will reduce the number of fragments to be amplified as only the rare cutter/frequent cutter fragments are amplified (Vos *et al.* 1995). These restriction fragments then have double stranded ligated ends, which serve as binding sites for the PCR amplification. Secondly, selective amplification of sets of restriction fragments is performed, and finally the amplified fragments are analysed on a denaturing gel (Vos *et al.* 1995). As this method involves several steps, the potential to introduce error or contamination is high. However, as there are several steps, there are also several opportunities to check the accuracy of the results at different stages in the procedure. To monitor quality control, a portion of the preamplification product should be run on an agarose gel to check that the DNA has digested and amplified consistently, which should result in an even 'smear' centred around the 200 bp intensity which is the same in all samples. Additionally both positive (using re-extracted DNA from the same individual) and negative (replacing the template DNA volume with sterile water) control samples

should be included in each batch of samples processed. Using high quality DNA is also strongly advisable because degraded DNA may result in spurious bands which are not cut by the restriction site in a regular pattern and represent bands of differing lengths which are not comparable between samples (Bensch and Akesson 2005).

Another virtue of the AFLP technique is the ease by which the method can be adapted to address more specific questions, simply by changing the restriction enzymes, adaptors or primers. For example, the basic AFLP method has recently been adapted to identify simple single locus markers e.g. SNPs (Single Nucleotide Polymorphisms) (Meksem *et al.* 2001; Nicod and Largiader 2003) or sequence tagged sites (STSs) (Brugmans *et al.* 2003) which are both proving to be valuable markers. Microsatellites can also be isolated by using a primer which anchors to a simple sequence repeat (SSR) at the selective amplification stage, although the complex banding pattern makes it difficult to identify alleles. Further modifications of the method can be applied to compare gene expression using cDNA-AFLP (reviewed in Bensch and Akesson 2005).

1.7. Ph.D. Aims

This study assesses the systematic, taxonomic, phylogenetic and ecological issues surrounding the South African *Parvulastra* sea stars providing fundamental knowledge of population genetics, life history and selection. Using extensive ecological and genetic sampling, and a combination of phenotypic, genotypic and ecological data the overall aims are to:

1. Establish the number of distinct lineages within *Parvulastra* in South Africa using both morphological analysis and a variety of molecular markers to resolve evolutionary questions, specifically are (i) the different phenotypes of *P. exigua* genetically distinct? and (ii) *P. exigua* and *P. dyscrita* separate taxonomic units? The ecology, distribution and phenotype of *P. exigua* will be investigated and compared to gonopore morphology and

genotype. South African *Parvulastra* species will be discussed in the context of the global populations of *P. exigua*, the 'Out of Africa' hypothesis (Waters and Roy 2004), and the recent evidence of mixed gonopore position (Hart *et al.* 2006).

2. Estimate neutral phylogeographic structure within South African *P. exigua,* specifically addressing the question - Is geographic variation in colour polymorphism due to neutral factors such as genetic drift, founder effect or migration (as opposed to being the result of selective forces)? Neutral phylogeographic structure will be compared to the biogeographic provinces, phenotypic distribution and ecological differences among habitat types, and discussed with reference to past processes and phenotypic adaptation and plasticity.

3. Identify AFLP loci within *P. exigua* which may be linked to genomic regions subject to diversifying or stabilizing selection. The allele frequencies of such loci will be examined across populations and compared to the neutral phylogeographic and phenotypic structure to elucidate potential selective forces. The influence and strength of the effects of these loci on the population structure will be assessed.

Chapters two, three and four are written as stand-alone entities and will be condensed and reformatted with the intention of submission for publication.

1.8. References

Anderson MJ (1999) Distinguishing direct from indirect effects of grazers in intertidal estuarine assemblages. *Journal of Experimental Marine Biology and Ecology* **234**, 199-218.

Anderson MJ, Underwood AJ (1997) Effects of gastropod grazers on recruitment and succession of an estuarine assemblage: A multivariate and univariate approach. *Oecologia* **109**, 442-453.

Arrontes J, Underwood AJ (1991) Experimental studies on some aspects of the feeding ecology of the intertidal starfish *Patiriella exigua*. *Journal of Experimental Marine Biology and Ecology* **148**, 255-269.

Asakawa S, Himeno H, Miura K, Watanabe K (1995) Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. *Genetics* **140**, 1047-1060.

Avise J (2004) *Molecular markers, Natural history and Evolution* Chapman and Hall, New York.

Avise JC, Arnold J, Ball RM, *et al.* (1987) Intraspecific phylogeography - the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology Evolution and Systematics* **36**, 621-642.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Baus E, Darrock DJ, Bruford MW (2005) Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star *Asterina gibbosa*. *Molecular Ecology* **14**, 3373-3382.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899-2914.

Bensch S, Helbig AJ, Salomon M, Siebold I (2002) Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Molecular Ecology* **11**, 473-481.

Bingham B, Braithwaite L (1986) Defensive adaptations of the dendrochirote holothurian *Psolus chitonoides*. *Journal of Experimental Marine Biology and Ecology* **98**, 311-322.

Birky CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513-527.

Bolton JJ, Anderson RJ (1990) Correlation between intertidal seaweed community composition and sea water temperature patterns on a geographical scale. *Botanica Marina* **33**, 447-457.

Bolton JJ, J. AR (1997) Marine Vegetation. In: *Vegetation of Southern Africa* (eds. Cowling RM, Richardson DM, Pierce SM), pp. 348-370. Cambridge University Press, Cambridge.

Booth JD, Ovenden JR (2000) Distribution of *Jasus* spp. (*Decapoda: palinuridae*) *phyllosomas* in southern waters: implications for larval recruitment. *Marine Ecology-Progress Series* **200**, 241-255.

Branch G, Branch M (1980) Competition between *Cellana tramoserica* (*Gastrapoda*) and *Patiriella exigua* (*Asteroidea*), and their influence on algal standing stocks. *Journal of Experimental Marine Biology and Ecology* **48**, 35-49.

Branch G, Branch M, Bannister A (1981) *The living shores of Southern Africa*, 5th edn. Struik, Cape Town.

Branch GM, Griffiths CL, Branch ML, Beckley LE (1994) *Two Oceans, A guide to the marine life of southern Africa* David Philip Publishers (Pty) Ltd., Cape Town.

Bromham L, Eyre-Walker A, Smith NH, Smith JM (2003) Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology & Evolution* **18**, 2-4.

Brown AC, Jarman N (1978) Coastal Marine Habitats. In: *Biogeography and Ecology of Southern Africa* (eds. Werger MJA, Bruggen acV), pp. 1239-1227. The Hague, Cape Town.

Brugmans B, van der Hulst RGM, Visser RGF, Lindhout P, van Eck HJ (2003) A new and versatile method for the successful conversion of AFLP (TM) markers into simple single locus markers. *Nucleic Acids Research* **31**, art. no.-e55.

Bussarawit S, Hansen B (1897) Results of Echinoderms from the firts PMBC/DANIDA Training course and workshop on taxonomy, biology and ecology of echinoderms. Phuket Marine Biological Centre, Phuket.

Bustamante RH, Branch GM, Eekhout S (1997) The influences of physical factors on the distribution and zonation patterns of South African rocky-shore communities. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **18**, 119-136.

Byrne M (1992) Reproduction of sympatric populations of *Patiriella gunnii, P. calcar* and *P. exigua* in New-South-Wales, Asterinid sea stars with direct development. *Marine Biology* **114**, 297-316.

Byrne M (1995) Changes in larval morphology in the evolution of benthic development by *Patiriella exigua* (*Asteroidea, Asterinidae*), a comparison with the larvae of *Patiriella* species with planktonic development. *Biological Bulletin* **188**, 293-305.

Byrne M (1996) Viviparity and intragonadal cannibalism in the diminutive sea stars *Patiriella vivipara* and *P. parvivipara* (family *Asterinidae*). *Marine Biology* **125**, 551-567.

Byrne M (2006) Life history diversity and evolution in the *Asterinidae*. *Integrative and Comparative Biology* **46**, 243-254.

Byrne M, Anderson MJ (1994) Hybridization of sympatric *Patiriella* species (*Echinodermata, Asteroidea*) in New-South-Wales. *Evolution* **48**, 564-576.

Byrne M, Cerra A (1996) Evolution of intragonadal development in the diminutive Asterinid sea stars *Patiriella vivipara* and *P. parvivipara* with an overview of development in the *Asterinidae*. *Biological Bulletin* **191**, 17-26.

Byrne M, Cerra A, Villinski JT (1999) Oogenic strategies in the evolution of development in *Patiriella* (*Echinodermata: Asteroidea*). *Invertebrate Reproduction & Development* **36**, 195-202.

Chen BY, Chen CP (1992) Reproductive cycle, larval development, juvenile growth and population dynamics of *Patiriella pseudoexigua* (*Echinodermata, Asteroidea*) in Taiwan. *Marine Biology* **113**, 271-280.

Clark AM (1974) Notes on some echinoderms of southern Africa. *Bulletin of the British Museum of Natural History (Zoology)* **26**, 421-487.

Clark AM, Downey ME (1992) *Starfishes of the Atlantic* Chapman and Hall, London.

Clark HL (1923) The Echinoderm Fauna of South Africa. *Annals of the South African Museum* **XIII**.

Colgan DJ, Byrne M, Rickard E, Castro LR (2005) Limited nucleotide divergence over large spatial scales in the Asterinid sea star *Patiriella exigua*. *Marine Biology* **146**, 263-270.

Dartnall AJ (1971) Australian sea stars of the genus *Patiriella* (*Asteroidea, Asterinidae*). *Proceedings of the Linnean Society of New South Wales* **96**.

Dartnall AJ, Byrne M, Collins J, Hart MW (2003) A new viviparous species of asterinid (*Echinodermata, Asteroidea, Asterinidae*) and a new genus to accommodate the species of pan-tropical exiguoid sea stars. *Zootaxa* **359**, 1-14.

Emanuel BP, Bustamante RH, Branch GM, Eekhout S, Odendaal FJ (1992) A zoogeographic and functional approach to the selection of marine reserves on

the west coast of South Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **12**, 341-354.

Eyre-Walker A (2000) Do mitochondria recombine in humans? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **355**, 1573-1580.

Eyre-Walker A, Awadalla P (2001) Does human mtDNA recombine? *Journal of Molecular Evolution* **53**, 430-435.

Eyre-Walker A, Smith NH, Smith JM (1999) Reply to Macaulay *et al.* (1999): mitochondrial DNA recombination - reasons to panic. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 2041-2042.

Fell HB (1962) West-wind drift dispersal of echinoderms in the southern hemisphere. *Nature* **193**, 759-761.

Gray JE (1840) A synopsis of the genera and species of the class *Hypostoma* (*Asterias Linnaeus*). *The Annals and Magazine of Natural History* **1**, 175-184; 275 - 290.

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of Mitochondrial DNA in mice. *Nature* **352**, 255-257.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution* **4**, 6-11.

Hart M (2000) Phylogenetic analyses of mode of larval development. *Seminars in Cell & Developmental Biology* **11**, 411-418.

Hart MW, Byrne M, Johnson SL (2003) *Patiriella pseudoexigua* (*Asteroidea*: *Asterinidae*): a cryptic species complex revealed by molecular and embryological analyses. *Journal of the Marine Biological Association of the United Kingdom* **83**, 1109-1116.

Hart MW, Byrne M, Smith MJ (1997) Molecular phylogenetic analysis of life-history evolution in *Asterinid* starfish. *Evolution* **51**, 1848-1861.

Hart MW, Keever CC, Dartnall AJ, Byrne M (2006) Morphological and genetic variation indicate cryptic species within Lamarck's little sea star, *Parvulastra (=Patiriella) exigua*. *Biological Bulletin* **210**, 158-167.

Heads M (2005) Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* **21**, 62-78.

Heads M (2005) Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society* **84**, 675-723.

Hewitt A. (2004) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology* **1**, 1-16.

Hiller N (1994) The environment, biogeography, and origin of the southern African recent brachiopod fauna. *Journal of Paleontology* **68**, 776-786.

Hudson RR, Turelli M (2003) Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**, 182-190.

Hunt A (1993) Effects of contrasting patterns of larval dispersal on the genetic connectedness of local populations of two intertidal starfish, *Patiriella calcar* and *P. exigua*. *Marine Ecology-Progress Series* **92**, 179-186.

Hutchings L (1994) The Agulhas Bank - a synthesis of available information and a brief comparison with other east-coast shelf regions. *South African Journal of Science* **90**, 179-185.

Jackson L (1976) Aspects of the intertidal ecology of the east coast of South Africa, pp. 1 - 72. Oceanographic Research Institute, Durban.

Jeffery CH, Emlet RB, Littlewood DTJ (2003) Phylogeny and evolution of developmental mode in temnopleurid echinoids. *Molecular Phylogenetics and Evolution* **28**, 99-118.

Kai Y, Nakayama K, Nakabo T (2002) Genetic differences among three colour morphotypes of the black rockfish, *Sebastes inermis*, inferred from mtDNA and AFLP analyses. *Molecular Ecology* **11**, 2591-2598.

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics* **50**, 149-159.

Kraytsberg Y, Schwartz M, Brown TA, *et al.* (2004) Recombination of human mitochondrial DNA. *Science* **304**, 981-981.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution* **20**, 243-247.

Ladoukakis ED, Zouros E (2001) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution* **18**, 1168-1175.

Lamarck JBPA (1816) *Histoire Naturelle des Animaux Sans Vertebres* Verdiere, Paris.

Lawrence JM, Herrera J (2000) Stress and deviant reproduction in echinoderms. *Zoological Studies* **39**, 151-171.

Lindquist N, Hay M (1996) Palatability and chemical defense of marine invertebrate Larvae. *Ecological Monographs* **66,** 431-450.

Lucas J, Hart R, Howden M, Salathe R (1979) Saponins in eggs and larvae of *Acanthaster planci* (L.) (Asteroidea) as chemical defenses against planktivorous fish. *Journal of experimental Marine Biology and Ecology* **40,** 155-165.

Lucchini V (2003) AFLP: a useful tool for biodiversity conservation and management. *Biologies* **326**, S43-S48.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.

Macaulay V, Richards M, Sykes B (1999) Mitochondrial DNA recombination - no need to panic. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 2037-2039.

Macnae W (1961) The fauna and flora of the eastern coasts of southern Africa in relation to ocean currents. *South African Journal of Science* **58**, 208-212.

Mah CL (2000) Preliminary phylogeny of the *Forcipulatacean Asteroidea*. *American Zoologist* **40**, 375-381.

Marlow JR, Lange CB, Wefer G, Rosell-Mele A (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* **290**, 2288-2291.

McClintock J, Vernon J (1990) Chemical defense in the eggs and embryos of Antarctic sea stars (Echinodermata). *Marine Biology* **105**, 491-495.

Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot DA (2001) Conversion of AFLP bands into high-throughput DNA markers. *Molecular Genetics and Genomics* **265**, 207-214.

Mishmar D, Ruiz-Pesini E, Golik P, *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 171-176.

Moreno G, Hoegh-Guldberg O (1999) The energetics of development of three congeneric sea stars (*Patiriella Verrill*, 1913) with different types of development. *Journal of Experimental Marine Biology and Ecology* **235**, 1-20.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology systematics. *Annual Review of Ecology and Systematics* **18**, 269-292.

Mortensen (1933) Echinoderms of southern Africa (*Asteroidea* and *Ophiuroidea*). *Vidensk. Meddr dansk naturh. Foren.* **93**, 215-400.

Mortensen T (1921) *Studies of the Development and larval forms of Echinoderms* G.E.C. GAD, Copenhagen.

Nagata Y, Matsuura ET (1991) Temperature dependency of electron transport activity in mitochondria with exogenous mitochondrial DNA in *Drosophila*. *Japanese Journal of Genetics* **66**, 255-261.

Neraudeau D, Mathey B (2000) Biogeography and diversity of South Atlantic Cretaceous echinoids: implications for circulation patterns. *Palaeogeography Palaeoclimatology Palaeoecology* **156**, 71-88.

Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16**, 358-364.

Nicod JC, Largiader CR (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for non- model organisms. *Nucleic Acids Research* **31**, art. no.-e19.

O'Loughlin PM (2002) New genus and species of southern Australian and Pacific *Asterinidae* (*Echinodermata, Asteroidea*). *Memoirs of Museum Victoria* **59**, 277-296.

O'Loughlin PM, Waters JM (2004) A molecular and morphological **revision** of the genera of *Asterinidae* (*Echinodermata: Asteroidea*). *Memoirs of Museum Victoria* **61**, 1-40.

O'Loughlin PM, Waters JM, Roy MS (2002) Description of a new species of *Patiriella* from New Zealand, and review of *Patiriella regularis* (*Echinodermata, Asteroidea*) based on morphological and molecular data. *Journal of the Royal Society of New Zealand* **32**, 697-711.

Pawlick J, Kernan M, Molinski T, Harper M, Faulkner D (1988) Defensive chemicals of the Spanish Dancer Nudibranch, *Hexabranchus sanguineus*, and its egg ribbons: macrolides derived from a sponge diet. *Journal of Experimental Marine Biology and Ecology* **119**, 99-109.

Peschak T (2005) *Currents of Contrast - Life in Southern Africa's Two Oceans* Struik Publishers (Pty) Ltd, Cape Town.

Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution* **21**, 2319-2325.

Putchakarn S, Sonchaeng P (2004) Echinoderm Fauna of Thailand: History and Inventory Reviews. *ScienceAsia* **30**, 417-428.

Riginos C, Victor BC (2001) Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 1931-1936.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology & Evolution* **18**, 411-417.

Salvato P, Battisti A, Concato S, *et al.* (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology* **11**, 2435-2444.

Shannon LV (1985) The Benguela Ecosystem.1. Evolution of the Benguela, Physical Features and Processes. *Oceanography and Marine Biology* **23**, 105-182.

Singh RS (2003) Darwin to DNA, molecules to morphology: the end of classical population genetics and the road ahead. *Genome* **46**, 938-942.

Smith MJ, Arndt A, Gorski S, Fajber E (1993) The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *Journal of Molecular Evolution* **36**, 545-554.

Somero GN (2002) Thermal physiology and vertical zonation of intertidal animals: Optima, limits, and costs of living. *Integrative and Comparative Biology* **42**, 780-789.

Sponer R, Roy MS (2002) Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution* **56**, 1954-1967.

Stephenson (1944) The constitution of the intertidal fauna and flora of South Africa. *Annals of the natal museum* **10,** 261-358.

Stevenson JP (1992) A possible modification of the distribution of the intertidal sea star *Patiriella exigua* (Lamarck) (*Echinodermata, Asteroidea*) by *Patiriella calcar* (Lamarck). *Journal of Experimental Marine Biology and Ecology* **155**, 41-54.

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

Verrill AE (1913) Revision of the genera of starfishes of the subfamily *Asterinidae. American Journal of Science* **35**, 177-485.

Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Walenkamp JH (1990) Systematics and zoogeography of *Asteroidea* (*Echinodermata*) from Inhaca Island, Mozambique. *Zoologische Verhandelingen* **261**, 1-86.

Waters JM, O'Loughlin PM, Roy MS (2004a) Molecular systematics of some Indo-Pacific *asterinids* (*Echinodermata, Asteroidea*): Does taxonomy reflect phylogeny? *Molecular Phylogenetics and Evolution* **30**, 872-878.

Waters JM, O'Loughlin PM, Roy MS (2004b) Cladogenesis in a starfish species complex from southern Australia: evidence for vicariant speciation? *Molecular Phylogenetics and Evolution* **32**, 236-245.

Waters JM, Roy MS (2003) Marine biogeography of southern Australia: Phylogeographical structure in a temperate sea-star. *Journal of Biogeography* **30**, 1787-1796.

Waters JM, Roy MS (2004) Out of Africa: The slow train to Australasia. *Systematic Biology* **53**, 18-24.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP. *Journal of Evolutionary Biology* **14**, 611-619.

Young C, Bingham B (1987) Chemical defense and aposematic coloration in larvae of the ascidian *Ecteinascidia turbinate. Marine Biology* **96**, 539-544.

Chapter 2

Phylogenetic complexity and cryptic species in the African *Parvulastra* sea stars

Chapter 2: Phylogenetic complexity and cryptic species in the African *Parvulastra* sea stars

K. S. Dunbar[1], J. Cable[1], C. Griffiths[2], M.W. Bruford[1].

[1]School of Biosciences, Cardiff University, Cardiff CF10 3TL, U.K.

[2]Marine Biology Research Institute, University of Cape Town, Rondebosch, South Africa.

2.1. Abstract

The Family Asterinidae (Grey 1840) is one of the most speciose and reproductively diverse families of marine invertebrates. Two species of the genus *Parvulastra* are recognised in South Africa, *Parvulastra exigua* and *Parvulastra dyscrita*. However, taxonomic confusion, paradoxical geographic distributions, intra-specific colour variation, morphological diversity and variation in gonopore position, used as an indicator of reproductive mode, suggest that either *P. exigua* and *P. dyscrita* are synonymous, or there are cryptic *Parvulastra* species within South Africa. Using morphological observations of gonopore position, and both mitochondrial and nuclear sequence data, this study suggests that South African *P. dyscrita* and *P. exigua* are separate species; however both species contain several distinct reciprocally monophyletic groups. Genetic evidence indicates that *P. exigua* in South Africa encompasses at least one additional 'cryptic' species which has an extremely limited geographic distribution. Further morphological and genetic investigations into *P. exigua* and *P. dyscrita* are needed to ascertain whether there are further cryptic lineages within the genus *Parvulastra* in South Africa.

2.2. Introduction

Largely as a result of the use of molecular markers, it has become evident that many marine taxa contain cryptic species (Knowlton 2000). Resolution of such species is particularly common amongst taxa showing phenotypic diversity (O'Loughlin *et al.* 2003), with extensive distributions (Spooner and Roy 2002; Hart *et al.* 2003), occurring across biogeographic divides (reviewed in Briggs 2006), inhabiting heterogeneous environments where the potential for local adaptation is high (reviewed in Schluter 2001), or having poorly studied ecology or life histories. Establishing the phylogenetic structure of marine species and identifying cryptic species may provide insights into barriers to gene flow, reproductive isolating mechanisms, marine genetic biodiversity and may as a result aid conservation (Palumbi 2003, 2004). Moreover, taxonomic identity is an essential prerequisite prior to further population, ecological or genetic investigations (reviewed in Knowlton 1993; Palumbi 1994).

The Asterinidae is one of the most speciose and diverse families of marine invertebrates containing 21 genera and 116 species worldwide (O'Loughlin and Waters 2004). This family contains the greatest diversity of larval types and life histories known among extant sea stars (Hart *et al.* 1997). It has been suggested that parallel changes in life history, habitat and dispersal potential have occurred during the evolution of this family (Byrne 1992; Chen and Chen 1992; Hunt 1993; Byrne 1995; Byrne 1996; Hart *et al.* 1997; Hart 2000; Jeffery *et al.* 2003) and that changes in larval developmental strategies may have driven speciation (Byrne *et al.* 1999). Small genetic distances between lineages with different developmental modes suggest some of these changes have been recent or rapid (Hart 2000; Byrne 2006).

Historically, the systematics of the Asterinidae has been confounded by the use of unreliable morphological characteristics that may be subject to selection, phenotypic plasticity or homoplasy (O'Loughlin and Waters 2004), prompting

researchers to develop molecular markers (Hart *et al.* 1997; Dartnall *et al.* 2003; Waters and Roy 2003, 2004a, b; O'Loughlin and Waters 2004; Waters *et al.* 2004a, b). Morphological and molecular systematic revisions of this family (reviewed in O'Loughlin and Waters 2004) have resulted in several new genera, the re-assignment of several species to new genera and the identification of several cryptic species complexes (Campbell and Rowe 1997 cited in O'Loughlin *et al.* 2002; O'Loughlin 2002; Dartnall *et al.* 2003; Hart *et al.* 2003; Waters *et al.* 2004a).

2.2.1. Parvulastra diversity

The newly described genus *Parvulastra* (O'Loughlin and Waters 2004) contains five species, and displays three different life history modes, including the most derived and rare form in sea stars, viviparity (intra-gonadal development and production of live young). *Parvulastra vivipara* and *P. parvivipara* have 'crawl away' juveniles, poor dispersal capabilities and very limited geographic distributions, being endemic to Australia (Dartnall 1969 cited in Dartnall *et al.* 2003; Keough and Dartnall 1978). *Parvulastra calcarata* is thought to have planktonic larvae and is endemic to Juan Fernandez Island (O'Loughlin and Waters 2004). *Parvulastra dyscrita* (Clark 1923) is also considered to have planktonic larvae but is endemic to South Africa. *Parvulastra exigua* (Lamarck 1816) has 'crawl away' juveniles and poor dispersal capabilities, but paradoxically has an extensive southern hemisphere distribution including Australia, South Africa and several oceanic islands. The distribution of the South African *Parvulastra* (*P. exigua*, distributed between Namibia and Mozambique, covering 3,000 km of coastline and *P. dyscrita*, distributed between Cape Point and Port Alfred in South Africa, covering 800 km of coastline) is surprising because marine invertebrate species with benthic life histories (such as *P. exigua*) are thought to have restricted geographic distributions, poor dispersal ability and high inter-population genetic differentiation (Riginos and Victor 2001; Sponer and Roy 2002). Species with planktonic dispersal abilities (such as *P. dyscrita*) are, in contrast, thought to be panmictic over large geographic areas

(Booth and Ovenden 2000). Therefore, it is unknown how *P. exigua* has obtained such an extensive distribution, or why *P. dyscrita* has a more limited distribution than *P. exigua* when it has the potential to disperse widely.

The paradoxical global distribution of *P. exigua* led Waters and Roy (2004a) to investigate the phylogeography of this species and to propose their "Out of Africa" hypothesis. This hypothesis proposes that *P. exigua* has an African ancestral origin and subsequent founder events, followed by population expansion, gave rise to the Australian and oceanic island populations via the west wind drift. Despite large genetic distances between the *P. exigua* populations on different land masses, Waters and Roy (2004a) do not suggest that the populations of *P. exigua* are divergent enough to warrant separate species status. However, Colgan *et al.* (2005) and Hart *et al.* (2006) have seriously questioned the 'Out of Africa' hypothesis suggesting that cryptic species could be present within *P. exigua*, obscuring the true phylogeography and dispersal of *P. exigua*.

2.2.2. *Parvulastra* in South Africa

With two out of the five *Parvulastra* species occurring in South Africa, the postulated ancestral origin of *P. exigua* being South Africa and speculation that more species will be found within the asterinids as investigations proceed (Dartnall *et al.* 2003), there is a clear gap in our knowledge regarding the systematics, phylogeny and ecology of South African *Parvulastra* species. All current knowledge of *P. exigua* population genetics (Hunt 1993; Colgan *et al.* 2005), life history traits (Hunt 1993; Byrne 1995; Hart *et al.* 1997) ecology (Branch and Branch 1980) and phylogeography (Waters and Roy 2004a; Colgan *et al.* 2005) is based on Australian populations.

In South Africa, *P. exigua* has an intertidal distribution stretching from the Namibian to the Mozambique border (Fig. 2.1) and displays striking phenotypic divergence. On the west coast this species is largely unmottled and greenish or

dark in colour, whereas on the south and east coasts, populations exhibit a wide diversity of mottled patterns varying in colour (Branch *et al.* 1994; pers. obs. see Chapter 1). Only the unmottled morph occurs in Australia and the mottled phenotype is endemic to South Africa. *Parvulastra dyscrita* occurs sympatrically within *P. exigua*'s geographic distribution from Cape Point to Port Elizabeth (Branch *et al.* 1994) (Fig. 2.1) and is mainly subtidal, but in some locations can be found sympatrically in the low intertidal with *P. exigua. Parvulastra dyscrita* is about twice the size of *P. exigua* but otherwise similar in appearance to the mottled morph of *P. exigua* (see Branch *et al.* 1994). The phenotypic and ecological similarities between the two species have led to taxonomic confusion, which has never been resolved (reviewed by Dartnall 1971; Colgan *et al.* 2005; Hart *et al.* 2006), despite many South African workers and textbooks identifying them as separate species (Branch and Branch 1981; Branch *et al.* 1994; C. Griffiths pers. comm.)

Recently this taxonomic confusion has been confounded by the discovery of museum specimens of *P. dyscrita* and *P. exigua* with different gonopore morphology, a key species diagnostic character. *Parvulastra exigua* (in Australia) possesses oral gonopores and produces eggs which give rise to benthic larvae (Byrne 1995; Hart *et al.* 1997; Waters and Roy 2004a; Colgan *et al.* 2005). *Parvulastra dyscrita* possesses aboral gonopores and is thought to produce planktonic larvae (Dartnall 1971). The gonopore location and reproductive mode of the African *P. exigua* populations, although not examined, were assumed to be the same as the Australian populations. However, Hart *et al.* (2006) identified both oral and aboral gonopores from museum specimens of *P. exigua* collected in South African and Island populations, but not Australian populations. Unfortunately, no genetic data were obtained from the museum specimens of *P. exigua* with aboral gonopores, and it is unknown if some specimens were misclassified (and are possibly *P. dyscrita* or represent cryptic species).

South Africa has a highly heterogeneous intertidal environment characterized by a steep temperature gradient. This, coupled with the continuous distribution and low dispersal ability relative to the geographic range of *P. exigua*, provides conditions in which local adaptation and speciation are likely (Schluter 2001). Preliminary observations of *P. exigua* indicate the colour morph divergence coincides with the Cape Peninsula biogeographic divide which separates the cold west coast province from the warm south and east coasts provinces (Jackson *et al.* 1976; Brown and Jarman 1978; Emanuel *et al.* 1992; Branch *et al.* 1994; Hiller 1994; Bustamante and Branch 1996; Neraudeau and Mathey 2000). This phenotypic divergence may be the result of ecological adaptation, phenotypic plasticity or may be an indicator of cryptic species separated by a barrier to gene flow around Cape Point. Recent studies (Hunt 1993; Colgan *et al.* 2005) indicate that the *P. exigua* populations in Australia and Tasmania have marked inter-population variation, accompanied by high frequencies of patchily distributed alleles, which suggests an increase in the effects of genetic drift and a higher potential for localized selection (Hunt 1993). There have, however, been few studies addressing phylogeographic barriers to gene flow and cryptic speciation in intertidal marine invertebrates in South Africa, despite examples of phenotypic differentiation occurring across biogeographic divides (Ridgway *et al.* 1998; Ridgway *et al.* 2000; Laudien *et al.* 2003; Tolley *et al.* 2005).

2.3. Aims

Given the phenotypic (colour and gonopore) diversity and taxonomic confusion within and between *P. exigua* and *P. dyscrita,* further investigation into the phenotypic and genetic structure of the African *Parvulastra* is necessary in order to phylogenetically place them in the context of the global populations of *P. exigua*. Using molecular (partial mitochondrial Cytochrome Oxidase I sequences and partial intron 3 of the nuclear actin gene sequences) and gonopore data, this study addresses two questions.

1. Are *P. exigua* and *P. dyscrita* really different taxa, and is there any genetic or phenotypic evidence to indicate cryptic molecular diversity and possibly additional *Parvulastra* taxa within South Africa?
2. Explore the phylogeography and systematic relationship between the African *Parvulastra* and the global populations of *P. exigua* in order to discuss the validity of the 'Out of Africa' hypothesis (Waters and Roy 2004a) and the ambiguous '*P. exigua*' gonopore data observed by Hart et al. (2006).

The region around the Cape Peninsula in the Western Cape of South Africa was chosen as the primary sampling location for this study because (i) the colour morph divergence within *P. exigua* is found in this region, occurring across this well characterized biogeographic divide and (ii) *P. dyscrita* occurs on the eastern side of the Peninsula but not the west.

2.4. Materials and Methods

2.4.1. Sampling

Three colour morphs (unmottled, intermediate and mottled) for *Parvulastra exigua* (Fig. 1.3, Chapter 1) were identified *apriori*, and preliminary field observations revealed a colour morph geographic transition zone roughly coinciding with Cape Point. Therefore sampling locations either side of Cape Point were selected to collect samples for genetic analyses (Fig. 2.1, Table 2.1). The sampling was conducted on low spring tides between July and August 2001.

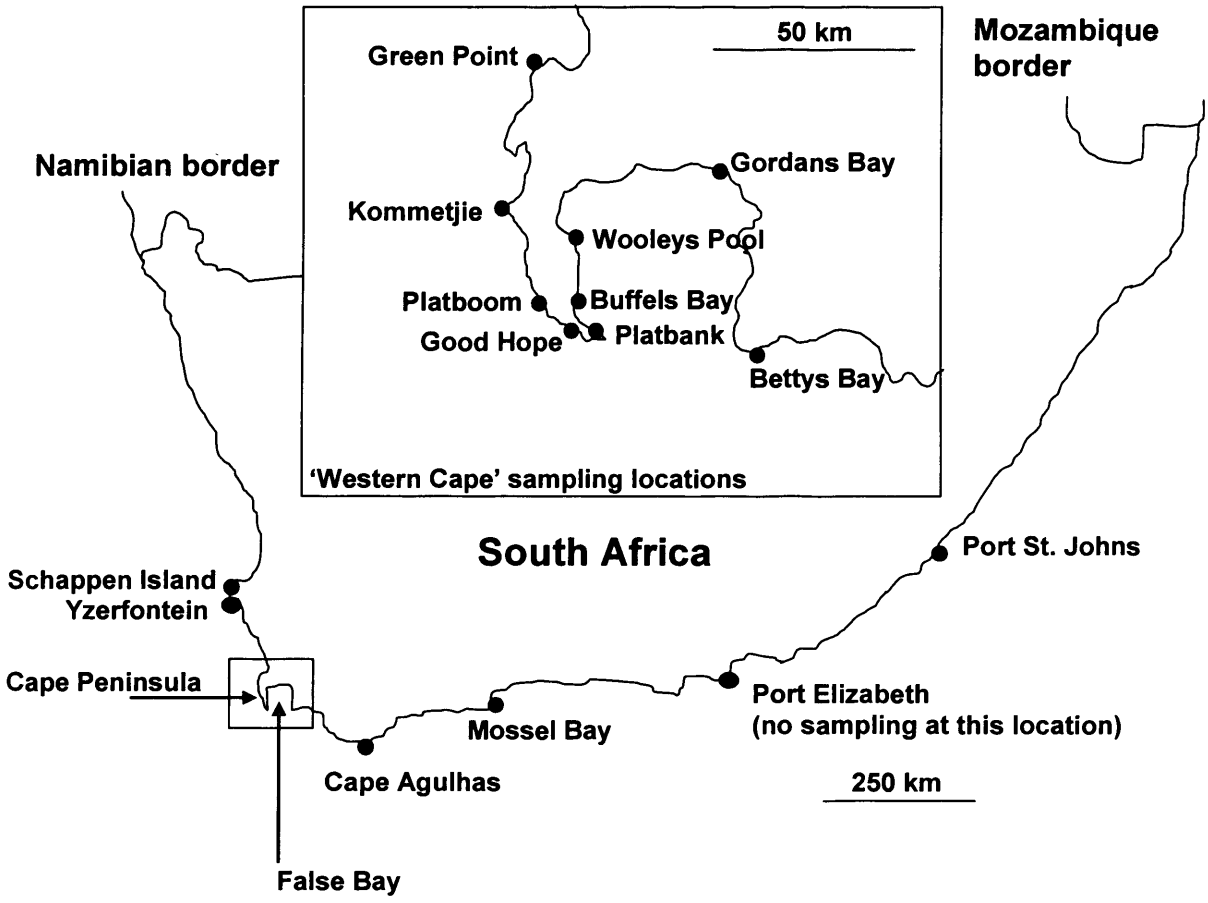

Figure 2.1. Map of the Western Cape, showing sampling locations for *Parvulastra exigua* and *P. dyscrita*.

Sampling location	Samples collected for genetic analysis	
	P. exigua	*P. dyscrita*
Schappen Island	Yes	No
Yzerfontein	Yes	No
Green Point	Yes	No
Kommetjie	Yes	No
Plat boom	Yes	No
Good Hope	Yes	No
Platbank	Yes	No
Buffels Bay	Yes	Yes
Wooleys Pool	Yes	Yes
Gordans Bay	Yes	Yes
Bettys Bay	Yes	No
Cape Agulhas	No	Yes
Mossel Bay	No	Yes
Port St. Johns	Yes	No

Table 2.1. Sampling locations for *Parvulastra exigua* and *P. dyscrita.*

Parvulastra exigua samples of each phenotype (unmottled, intermediate and mottled) were collected for genetic and morphological analyses from tide pools at all 12 sampling locations. Photographs were taken of all *P. exigua* samples collected for genetic analyses. *Parvulastra dyscrita* samples were collected from the subtidal and intertidal zones by snorkel and SCUBA from Buffels Bay, Mossel Bay, Cape Agulhas, Gordans Bay and Wooleys Pool. All samples were saturated in 100% ethanol and stored at-80°C prior to DNA extraction.

2.4.2. Gonopore Analyses

A subset of *P. exigua* samples (7 individuals from Western Cape sampling locations; nine individuals from east coast Port St. Johns, seven individuals from Kommetjie) and *P. dyscrita* samples (8 individuals) were examined for gonopore location using a stereo-microscope. The number of gonopores visible on the oral surfaces of the sea stars was recorded by two independent observers. The aboral surface of the animals was not examined as the gonopores on this surface are difficult to identify reliably without dissection (A. Dartnall pers. comm. 2005).

2.4.3. DNA extraction, Cytochrome Oxidase I (COI) and Actin Intron amplification and sequencing

DNA extractions were performed using either a Qiagen DNeasy tissue purification kit (Qiagen) according to the manufacturer's instructions, or by phenol-chloroform and CTAB purification (Arndt *et al.* 1996). The DNA was re-suspended in 100 µl TE buffer and treated with RNAase (20µg/µl) at 55°C for 1 h. Novel primers (P.ex.COI.29F (5' CCA AAC ACA AGG ACA TAG GAA 3') and P.ex.COI.575B (GCG GTA ACG AAT ACG GAT CA) were designed to amplify approximately the first 500 bases of the COI gene sequence from an Australian specimen of *P. exigua* (Accession number U50053: Hart *et al.* 1997) using PRIMER VS. 3 (Rozen and Skaletsky 2000) and the web software Oligonucleotide Properties Calculator (www.basic.northwestern .edu/biotools/oligocalc.html). Introns of the cytoplasmic actin gene have previously been used to study a *Parvulastra* species complex in southern Australia (Waters *et al.* 2004b), and therefore this nuclear marker was selected for the present study. A 427 bp region of intron 3 of the gene was sequenced using primers Actin Intron 3 2FB (5'-CTTTCACCACCACYGGTGAGA -3') and Actin 3R (5'– TTGSWGATCCACATCTG -3') (Waters *et al.* 2004b).

All PCRs were performed in a PE 9700 thermal cycler. COI amplification was performed using the following program: 94°C for 5 min; followed by 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 min for 35 cycles with a final extension of 72°C for 10 min. The reaction conditions were 1x PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs; 1 µM of each primer; 0.1U Taq (Invitrogen); and 1 µl DNA template (diluted between 1/10 and 1/1000) in a final volume of 15 µl. Actin Intron amplification was obtained using the following program: 94°C for 5 min; followed by 94°C for 1 min, 53°C for 30 seconds, 72°C for 30 seconds for 35 cycles with a final extension of 72°C for 10 min. The reaction conditions used were 1x PCR buffer, 1.5 mM $MgCl_2$; 0.2 mM dNTPs; 0.5 µM of each primer; 1 U Taq; 1 µl DNA template (diluted between 0 and 1/100) in a final volume of 15 µl.

PCR products were purified using a BIO 101 Geneclean *Turbo* For PCR Kit (Q-BIOgene) or using the enzymes Exonuclease I and shrimp alkaline phosphatase (Amersham Biosciences) according to the manufacturer's instructions. Sequencing PCRs were performed in 7.5 μl reactions. The PCR program involved an initial denaturation at 96°C for 1 min 30 seconds for 1 cycle, followed by denaturation at 96°C for 10 seconds; annealing at 50°C for 5 seconds; extension at 60°C for 4 minutes for 25 cycles. Both forward primer and reverse primer reactions were performed. The reaction conditions were as follows: PCR H_2O and template combined 3 μl; Better Buffer (Web Scientific Ltd) 2.5 μl; Big Dye 0.5 μl (ABI PRISM® Big Dye ™ Terminator dye vs. 3.1); 1.5 μl of either the forward or reverse primer (1.6 μM). An ABI Prism 3100 Genetic Analyzer (Applied Biosystems) was used for the sequencing according to the manufacturer's instructions. Both the COI and Actin Intron forward and reverse sequences were separately aligned in SEQUENCHER vs. 3.1.2 (GeneCode Corp.) and verified by eye.

2.4.4. COI haplotype identification

The COI alignment chromatograms were examined for ambiguous nucleotide positions. Where ambiguous positions were evident, products were cloned using the TA Cloning kit for PCR (Invitrogen), with One Shot TOP 10 chemically competent cells (Invitrogen) following manufacturer's instructions and the forward sequence of the colony insert was sequenced for six colonies within each sample. Identification of COI haplotypes, polymorphic loci and amino acid reading frame was limited to confidently aligned fragments that contained no ambiguous nucleotide positions (except for the cloned samples, see results). This meant that the COI fragment length used in the phylogenetic analyses was 358 bp. The amino acid reading frame was identified by aligning the full COI gene (Accession no. U50053) to the COI sequences identified in the present study in SEQUENCHER vs. 4.12 (GeneCode Corp.) and checking the echinoderm mitochondrial genetic code for stop codons and amino acid changes.

2.4.5. Actin intron allele identification

The actin intron sequences were produced in genotype format displaying double peaks on the chromatograms where single nucleotide polymorphisms (SNP) were present between the alleles. To infer the allelelic phase of the genotype data, the web based software HAP (Halperin and Eskin 2004) was used. Indels were excluded from the allele inference as the program was unable to process indels, instead they were treated as informative presence/absence markers.

2.4.6. Phylogenetic Analyses

Three datasets were used for the phylogenetic analyses. The first dataset used the 358 bp COI sequences obtained in the present study. The second dataset consists of a 244 bp COI alignment of African *Parvulastra* sequences (obtained in this study) with previously published sequences from the global *P. exigua* distribution (Waters and Roy 2004a: Accession No's. AY396051 - AY396074 and AY397622 - AY397643). The third dataset consisted of an alignment of 427 bp of the actin intron sequences. The three datasets will be referred to hereafter as the 358 bp COI, 244 bp COI and AI datasets.

To assess the genetic structure of *P. exigua* and *P. dyscrita* and identify haplogroups, minimum spanning networks and median joining networks were constructed. The former were produced for all three datasets using the software TCS vs. 1.13 (Clement *et al.* 2000). This method identifies haplotypes that cannot be reliably connected to each other at the 95% plausibility level, a measure often interpreted as evidence of separate species (Tarjuelo *et al.* 2004; Uthicke *et al.* 2004; Hart *et al.* 2006) and also assigns an 'out-group weighting' to each haplotype, which is an indicator of divergence. Four other *Parvulastra* species COI sequences (*P. vivipara*, Accession no. U50054, Hart *et al.* 1997; *P. parvivipara,* Accession no. U50055, Hart *et al.* 1997; *P. regularis,* Accession no. U50045, Hart *et al.* 1997; and *Meridiastra mortenseni,* Accession no. AY370750, Waters *et al.* 2004a) were included as outgroups in the minimum spanning network. Secondly, median joining networks (Bandelt *et al.* 1999)

were constructed using the two COI datasets using the software NETWORK vs. 4.1.1.2. (www.fluxus-engineering.com). This method shows the relationship between the haplotypes and the number of individuals possessing each haplotype.

To more rigorously explore the phylogenetic relationships among the haplotypes detected, a maximum likelihood phylogeny was constructed to assess both genetic structure and the level of bootstrap support for the haplogroups identified by the networks. The Akaike Information Criterion (AIC) in Model test 3.06 (Posada and Crandall 1998) was used to select the GTR+I+G model of sequence evolution with base frequencies: A=0.28; G=0.18; T=0.28; c=0.26; Proportion of invariable sites (I) = 0.5691; and Gamma shape distribution parameter = 2.3385. Six outgroups were included in the model selection and Maximum Likelihood analysis. *Parvulastra parvivipara* and *P. vivipara* were included to represent the most closely related species to *P. exigua* and *P. dyscrita* (see O'Loughlin and Waters 2004). *Asterina gibbosa* (Accession No. U50053, Hart *et al.* 1997) was included as a distant outgroup to root the tree and *P. regularis* (Accession No. U50045, Hart *et al.* 1997), *M. mortenseni* (Accession No. AY370750, Waters *et al.* 2004a) and *Stegmaster inflatus* (Accession No. AY370743, Waters *et al.* 2004a) were included as species with an intermediate level of phylogenetic similarity, to give the tree structure on three levels. The tree was constructed for the 244 bp COI dataset only, using the heuristic option with 10 replicates of random sequence addition in PAUP*4.0b10 (Swofford 1998). Phylogenetic confidence in the topology was assessed by bootstrapping (Felsenstein 1985) using the software PAUP DOS (Swofford 1998) with heuristic analysis of 100 maximum likelihood replicate datasets.

An analysis of molecular variance (AMOVA: Excoffier *et al.* 1992) was carried out to assess the partitioning of variation within and among apriori defined groups and was calculated using ARLEQUIN vs. 3. (Excoffier *et al.* 2005: http://cmpg.unibe.ch/software/arlequin3) using 1000 permutations. To assess indicative inter-haplogroup and inter-specific genetic divergence, Tamura-Nei pairwise distances (Tamura and Nei 1993) between haplogroups were calculated in

MEGA3 (Kumar *et al.* 2004). The Tamura-Nei model corrects for multiple hits, taking into account the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies. It distinguishes between transitional substitution rates between purines and transversional substitution rates between pyrimidines. It also assumes equality of substitution rates among sites (Kumar *et al.* 2004). The inter haplogroup distances were calculated for both COI datasets (358 and 244 bp). Three outgroups were included for inter-specific comparisons, *Parvulastra vivipara*, *P. parvivipara* and *Patiriella regularis*.

2.5. Results

2.5.1. Gonopore Analyses

All 23 *Parvulastra exigua* specimens examined had visible gonopores on their oral surface, but there was considerable variation in the number of gonopores detected per specimen (ranging from 2-10). Four of the nine specimens from the east coast of South Africa; four out of the seven Kommetjie *P. exigua* specimens and three out of the seven Western Cape *P. exigua* specimens had ≥8 gonopores visible. Of the eight *P. dyscrita* samples examined, seven had no visible gonopores on their oral surfaces, and one had gonopore like structures in the same location that the gonopores were found in the *P. exigua* specimens. This *P. dyscrita* specimen was much smaller than the other *P. dyscrita* specimens, but possessed a *P. dyscrita* COI genotype.

2.5.2. Ambiguous nucleotide positions within the COI chromatograms

Ambiguous nucleotides were evident from the COI chromatograms of 12 individuals from the sampling location Kommetjie only, and were not observed in individuals from any other location. A subset consisting of six of the individuals with ambiguous nucleotide positions was cloned and the forward sequence of the colony insert was sequenced for six colonies within each individual, yielding 27 sequences of readable quality from the 36 colonies sequenced. Only one of the six individuals cloned yielded all colonies with an identical sequence. The other five individuals yielded between two and five different sequences from the colonies. Within the six colonies sequenced per individual, one sequence was common to all individuals and appeared between one and four times in the colonies sequenced (Table 2.2).

The ambiguous peaks could indicate either (i), mitochondrial DNA that has been transposed into the nuclear genome (NUMTS), (ii) paternal leakage, or (iii) heteroplasmy.

	No. colonies sequenced	No. of 'readable' sequences obtained	No. of unique sequences within each individual	No. of colonies yielding most common sequence
Individual 1	6	5	3	3
Individual 2	6	5	5	1
Individual 3	6	6	4	3
Individual 4	6	4	1	4
Individual 5	6	5	2	4
Individual 6	6	2	2	1
Total	36	27	18	16

Table 2.2. Sequence results from the cloning of a subset of six *Parvulastra exigua* Kommetjie individuals which displayed ambiguous nucleotide positions on the COI chromatograms.

To test for the presence of NUMTS, the colony sequences were aligned with the amino acid reading frame and checked for stop codons, non-synonymous substitutions and unusual amino acid substitutions. The most common sequence was aligned with the most common haplotype present in the remaining *P. exigua* samples. These two sequences differed by 34 out of 363 bp. However, all the substitutions were synonymous (and at the 3rd codon position) and the amino acid sequence was conserved, despite the large sequence divergence. Eight of the remaining 11 colony sequences contained between one and 14 non-synonymous substitutions, and one of these contained a stop codon. Three remaining colony sequences contained only synonymous substitutions, appeared only once and were present in separate individuals. On the basis of this sequencing evidence, it was concluded that the most common haplotype found in 16 out of the 27 colony sequences obtained was most likely the 'correct' mitochondrial haplotype, and this was then assigned to the 12 individuals from Kommetjie for the remainder of the analyses. The additional sequences obtained from the colonies were considered to be NUMTS or heteroplasmic sequences and were discarded from further analysis.

2.5.3. Cytochrome Oxidase I (COI) Sequence variation

One hundred and fifty seven *Parvulastra exigua* individuals were sequenced for a 358 bp region of Cytochrome Oxidase I yielding 28 haplotypes (Table 2.3), and 35 *P. dyscrita* individuals yielded eight haplotypes (Table 2.4). All polymorphic sites for both species were synonymous substitutions. An alignment of the sequence ambiguities between the 358 bp dataset haplotypes is shown in 2.9. Appendices 2 and 3.

		Sampling locations												
		Yzerfontein	Green Point	Kommetjie	Platboom	Good Hope	Platbank	Buffels Bay	Wooleys Pool	Gordans Bay	Bettys Bay	Schappen Island	Port St. Johns	**Total no. in each haplotype**
Haplotypes	**1**	6	6	13	3	4	9	10	7	26	2	4		**90**
	2								3	11				**14**
	3						3	2	1		1			**7**
	4										1	3		**4**
	5								3					**3**
	6			12										**12**
	7					2								**2**
	8	2												**2**
	9										1			**1**
	10							1						**1**
	11							1						**1**
	12							1						**1**
	13		1											**1**
	14		1											**1**
	15			1										**1**
	16						1							**1**
	17						1							**1**
	18						1							**1**
	19											1		**1**
	20								1					**1**
	21								1					**1**
	22								1					**1**
	23	1												**1**
	24	1												**1**
	25												3	**3**
	26												2	**2**
	27												1	**1**
	28												1	**1**
Total (n)		**10**	**8**	**26**	**3**	**6**	**15**	**15**	**17**	**37**	**5**	**8**	**7**	**157**

Table 2.3. Haplotype identification and sampling location for *Parvulastra exigua* partial Cytochrome Oxidase I subunit sequences.

	Sampling locations						
Haplotypes		Buffels Bay	Wooleys Pool	Gordans Bay	Cape Agulhas	Mossel Bay	**Total no. in each haplotype**
	29	9	3	6	6		**24**
	30	2	1				**3**
	31	1					**1**
	32		1				**1**
	33			1			**1**
	34		1				**1**
	35				1	2	**3**
	36					1	**1**
Total (n)		**12**	**6**	**7**	**7**	**3**	**35**

Table 2.4. Haplotype identification and sampling location for *Parvulastra dyscrita* partial Cytochrome Oxidase I subunit sequences.

A further 11 *P. exigua* sequences (and six additional haplotypes) from Waters and Roy's (2004a) study were aligned with the sequences from this study and the overlapping portion of the two sequence sets comprises the 244 bp COI dataset. The shorter length of the 244 bp dataset resulted in the re-grouping of ten of the 358 bp haplotypes into just two 244 bp haplotypes because the polymorphic nucleotides were in the portion of the fragment that was trimmed (Table 2.5). Eight haplotypes identified from the 358 bp dataset (which differed from haplotype 1 by only one substitution) were re-grouped into a new haplotype, named SA Main. Haplotypes 25 and 26, which differed from each other by only one substitution were also re-grouped, along with two haplotypes (VICVO, SAFPJ) from Waters and Roy (2004a). An alignment of the sequence ambiguities between the 358 bp dataset haplotypes and the 244 bp dataset is shown in 2.9. Appendix 2 and Appendix 3.

244 bp haplotype names	**Samples represented in the new and re-grouped *Parvulastra exigua* haplotypes**	**n**	**Geographic locations represented in the haplotypes**
SA Main	Haplotypes 1, 7, 8, 11, 15, 16, 21, 24 (all from the 358 bp dataset)	98	Western Cape South Africa
25/26	Haplotypes 25 and 26 (from the 358 bp dataset) VICVO, SAFPJ (from Waters and Roy 2004)	7	Port St. Johns (east coast South Africa), Victoria (Australia)
TASTW	TASTW1, TASTW3 (from Waters and Roy 2004)	2	Tasmania (Australia)
LH	LH1, LH3 (from Waters and Roy 2004)	2	Lord Howe Island
SAFDB	SAFDB1, SAFDB2 (from Waters and Roy 2004)	2	Upper east coast South Africa
SH1	SH1 (from Waters and Roy 2004)	1	St Helena Island
AM1	AM1 (from Waters and Roy 2004)	1	Amsterdam Island
AM2	AM2 (from Waters and Roy 2004)	1	Amsterdam Island

Table 2.5. *Parvulastra exigua* haplotype identities using 244 bp of COI sequences from the present study and from Waters and Roy (2004a).

2.5.4. Actin Intron (AI) sequence variation

A 427 bp region of the AI was sequenced from 57 *P. exigua* individuals and 29 genotypes were identified. Twenty-five individuals were homozygotes for six different alleles. Thirty-two individuals were heterozygotes and the HAP software identified a further 16 alleles within the heterozygotes, making a total of 22 *P. exigua* alleles. The heterozygotes contained between one and five polymorphic loci. A four bp indel was present in six of the individuals which shared one very divergent homozygote AI genotype. These six individuals were also found to contain the COI NUMTS or heteroplasmy (see above) and were all recorded as intermediate phenotypes from the sampling location Kommetjie.

Within the five *P. dyscrita* individuals sequenced, five genotypes were present, all individuals were heterozygotes and from these the HAP software identified seven alleles. All further analyses for the AI marker for both *P. exigua* and *P. dyscrita* were conducted using the allelic data.

2.5.5. Phylogenetic Analysis

2.5.5.1. COI Networks

The haplogroups identified by the minimum spanning and median joining networks (using the 244 bp dataset were labelled (Table 2.6) and will be referred to by these labels hereafter).

Haplogroup name	Abbrv	n	sampling locations represented in haplogroup	Source of data
P. exigua western Cape	PeWC	138	Yzerfontein, Green Point, Kommetjie, Platboom, Good Hope Platbank, Buffels Bay, Wooleys Pool, Gordans Bay, Bettys Bay, Schappen Island	358 bp COI data
P. exigua Global	PeG	16	Port St. Johns, South Africa Durban, South Africa All samples from Australia, All samples from Tasmania, St Helena Island, Lord Howe Island	358 bp COI data Waters and Roy 2004a " " " "
P. exigua Kommetjie	PeK	12	Haplotype 6 identified in 12 / 26 samples from Kommetjie	358 bp COI data
P. exigua Amsterdam Island	PeAI	2	All samples from Amsterdam Island	Waters and Roy 2004a
P. dyscrita West	PdW	31	Buffels Bay, Gordans Bay, Cape Agulhas	358 bp COI data
P. dyscrita East	PdE	4	Cape Agulhas, Mossel Bay	358 bp COI data

Table 2.6. *Parvulastra* haplogroups identified from the Minimum Spanning Network and the Median Joining Network for the 244 bp COI dataset, with the sample size (n) and group name abbreviation (Abbrv).

The Minimum spanning network (MSN) and Median Joining Network (MJN) using both the COI datasets (244 bp dataset, MJN shown in Fig. 2.2; 244 bp dataset MSN shown in Fig. 2.5 Appendix 2.9;1; 358 bp dataset networks not shown) showed a similar overall structure. The MSN based on the 244 bp dataset indicated five networks at the 95% plausibility level (>6 substitutions). One of these networks was subdivided into two groups, making a total of six haplogroups. The network containing the two sub-groups, contained both *P. exigua* individuals from geographically widespread locations and *P. dyscrita* individuals. Although connected at the 95% level in the MSN, these sub-groups will be treated as separate haplogroups (PeG and PdW) because of the difference in gonopore position (see Discussion). The MJN for the 244 bp dataset showed the relationship between the six haplogroups.

The MSN using the 358 bp dataset indicated the same haplogroups identified using the 244 bp dataset (except the Amsterdam Island samples, from Waters and Roy 2004a, which were not included in this dataset). The haplogroup pattern for the 358 bp dataset was slightly different from that of the 244 bp dataset (the number of substitutions separating sub-groups PeG and PdW increased from three to four; 95% plausibility level increased from >6 to >8 substitutions; only haplotypes from the east coast of South Africa were represented in PeG) as a result of the longer haplotype sequences which formed the 358 bp dataset.

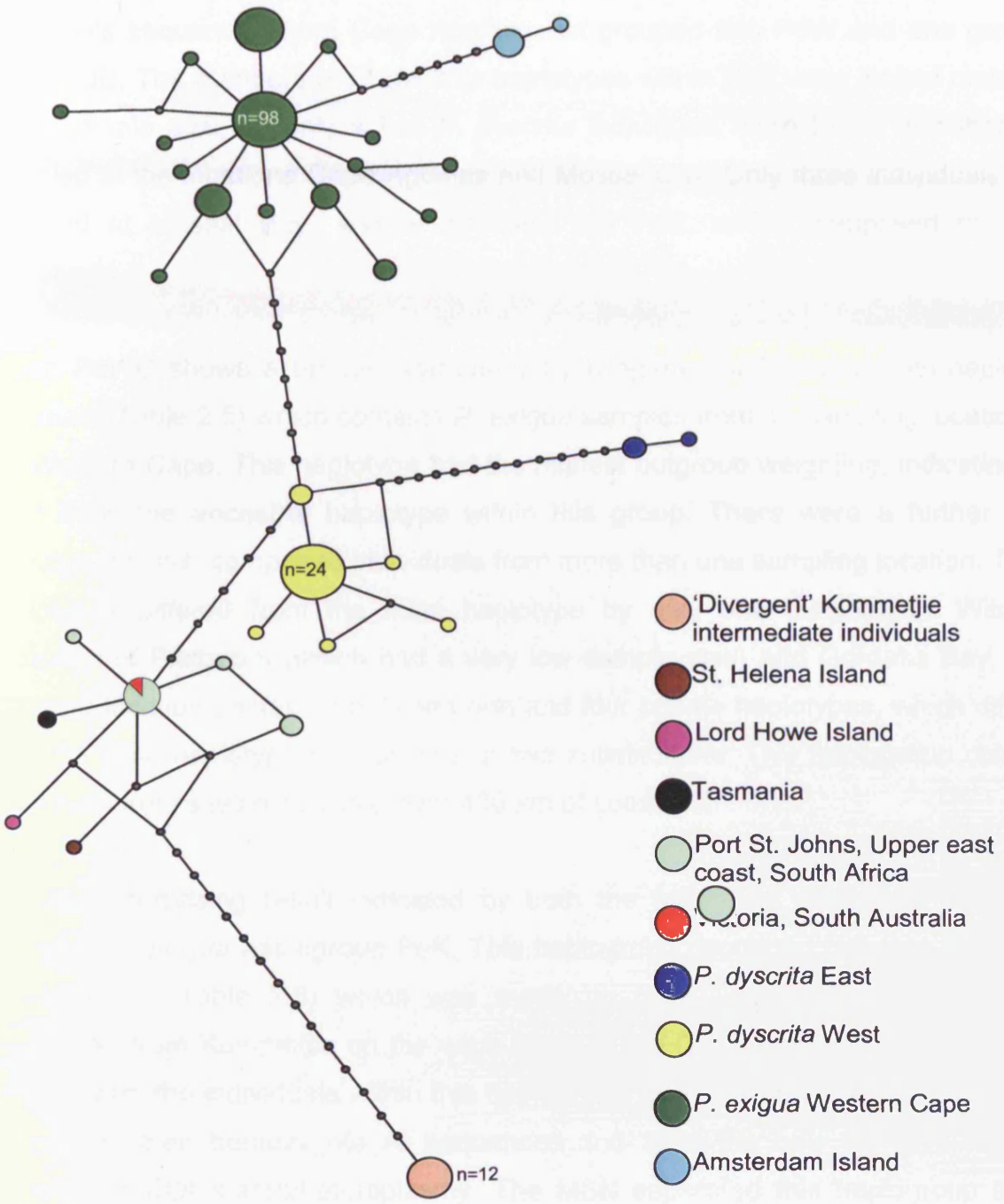

Figure 2.2. Median Joining Network for *Parvulastra exigua* and *P. dyscrita* 244 bp dataset COI haplotypes. Circles represent haplotypes, circle area represents number of individuals within that haplotype (except where 'n' is stated in or next to the circle). Small grey circles indicate number of nucleotide differences between haplotypes, lines without small grey circles indicate one nucleotide difference between haplotypes. No outgroups included.

The MJN indicated that the *P. dyscrita* samples within the haplogroups PdE and PdW were separated by 11 nucleotide substitutions. Out of the seven *P. dyscrita* individuals sequenced from Cape Agulhas, six grouped into PdW and one grouped within PdE. The numbers of *P. dyscrita* haplotypes within PdE were limited mainly by small sample size, as only a few *P. dyscrita* individuals were found and therefore sampled at the locations Cape Agulhas and Mossel Bay. Only three individuals were sampled at Mossel Bay, and all grouped into PdE, which comprised only two haplotypes.

Group PeWC shows a broadly star-shaped phylogeny, having one main haplotype (SA main, Table 2.5) which contains *P. exigua* samples from 11 sampling locations in the Western Cape. This haplotype had the highest outgroup weighting, indicating it is most likely the ancestral haplotype within this group. There were a further three haplotypes which comprised individuals from more than one sampling location. These haplotypes differed from the main haplotype by only one substitution. With the exception of Platboom (which had a very low sample size) and Gordans Bay, each sampling location contained between one and four private haplotypes, which differed from the main haplotype by only one or two substitutions. This haplogroup contains individuals from sites distributed over 400 km of coastline.

The most surprising result indicated by both the MSN and MJN was the highly divergent *P. exigua* haplogroup PeK. This haplogroup contained only one haplotype (Haplotype 6, Table 2.6) which was made up of 12 out of the 26 individuals sequenced from Kommetjie on the west coast of the Cape Peninsula, South Africa. Furthermore, the individuals within this haplogroup were the only ones to show the 4 bp indel in their homozygote AI sequences and were the only samples showing evidence of COI NUMTS/heteroplasmy. The MSN separated this haplogroup at the 95% plausibility level (<6 substitutions, 244 bp dataset) and the MJN indicated that this haplogroup was connected to the *P. exigua* individuals within PeG by 17 or 18 substitutions (depending on the alternative connections within PeG). On the discovery of this divergent Kommetjie haplogroup, photographs of these individuals (which were

initially recorded as 'intermediate' phenotypes when sampled) were re-examined revealing subtle phenotypic differences, with these individuals displaying a reddish green colouration, which was not detected in any other individuals of the intermediate phenotype, or from any other location.

Group PeAI contained only two individuals, both *P. exigua* with unique haplotypes and both from Amsterdam Island, which is located almost mid-way between South Africa and Australia, in the South Indian Ocean. This group was joined by eight nucleotide substitutions to the *P. exigua* individuals within the main haplotype in group PeWC.

2.5.5.2. COI Maximum Likelihood tree (244 bp dataset) and average COI Tamura-Nei genetic distances

Support for the COI groupings was also provided by the Maximum Likelihood phylogenetic analysis (Fig. 2.3), AMOVA and average Tamura-Nei pairwise divergences between the groups (Table 2.7). Generally, there was high comparability between all the analyses, however the divergence estimates were dramatically different between some groups when using the 244 bp dataset as opposed to the 358 bp dataset. Excluding outgroups, five distinct phylogenetic groups were evident in the maximum likelihood tree. Groups PeWC and PeAI were in the same clade (bootstrap support 65%), but PeAI formed a separate internal clade branching from group PeWC which had higher support (87%). The divergence estimates between these two groups was 3.3% (Table 2.7).

The *P. dyscrita* individuals within PdE formed a single well supported clade (91%) and the genetic distance between PdE and PdW, the most closely related group, was 7.7% (4.7% for the 358 bp dataset). Group PdW was separated from group PeG, but with low bootstrap support (59%), which also reflected the MSN and MJN structures for these two groups. Haplogroups PdW and PeG were the least divergent from each other with a distance estimate of 2.49% (2.07%, for the 358 bp dataset). Group PeK had the highest pairwise divergence estimates overall ranging from 11.66% (PeAI) to 7.07% (PeG). This haplogroup formed a loosely supported clade with the two

outgroups *P. parvivipara* and *P. vivipara*, (bootstrap support 60%). This grouping was reflected in the pairwise divergences only 7.91% (8.42 for the 358 bp dataset) between PeK and *P. parvivipara*, and the smallest divergence estimate between PeK and any other *P. exigua* group was 7.07 (PeG).

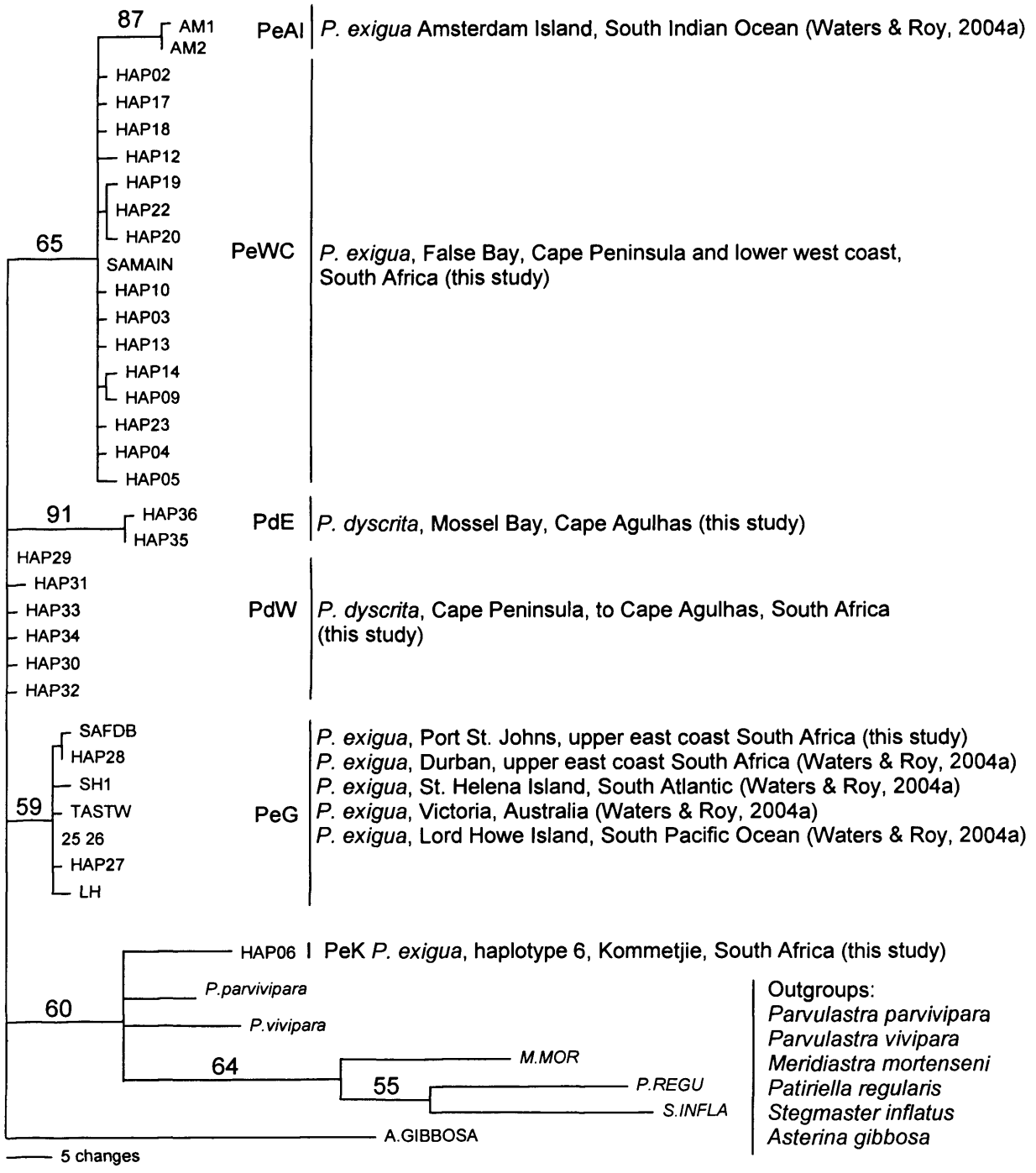

Figure 2.3. Maximum Likelihood tree of all South African *Parvulastra* sequences (and outgroups) sequenced for a portion of the COI gene with 100 bootstraps. Numbers on branches indicate bootstrap probability.

	PeWC	PeG	PdW	PdE	PeK	PeAl	*P. vivipara*	*P. parvivipara*
PeWC	-							
PeG	6.03 (3.97)	-						
PdW	4.31 (3.48)	2.49 (2.07)	-					
PdE	7.13 (6.33)	6.57 (5.75)	7.70 (4.76)	-				
PeK	11.17 (10.07)	7.07 (7.27)	7.48 (7.21)	9.57 (8.26)	-			
PeAl	3.30	6.64	4.12	8.19	11.66	-		
P. vivipara	12.72 (10.63)	11.49 (9.80)	10.01 (8.19)	13.90 (10.39)	10.36 (9.05)	14.61	-	
P. parvivipara	10.14 (9.97)	9.95 (9.87)	8.38 (8.16)	11.05 (9.56)	7.91 (8.42)	12.88	9.40 (7.17)	-
P. regularis	23.45 (21.44)	23.83 (21.87)	22.00 (19.81)	25.50 (23.37)	21.66 (20.76)	21.85	25.86 (22.61)	22.07 (19.74)

Table 2.7. Average Tamura-Nei genetic distances of all pairwise distances between groups, calculated in MEGA3, for partial (244 bp) Cytochrome Oxidase I in *Parvulastra exigua* and *P. dyscrita* samples (current study; Waters and Roy 2004a). The TrN inter-group distances for all haplogroups (except Amsterdam Island) using the 358 bp COI dataset were also calculated and are shown in parentheses. Outgroups were included for comparison.

2.5.5.3. Analysis of molecular variance (AMOVA)

The AMOVA results (Table 2.8) support the network with >95% of the genetic variation accounted for among groups, <1% accounted for among populations within groups and <4% accounted for within populations. Although groups PeG and PdW are connected at the 95% plausibility level (as indicated by the MSN, Fig. 2.5 Appendix 2.9;1) the gonopore results indicate that they are likely to be two separate taxa and therefore should be classed as separate. The AMOVA results support this separation showing 84-90% (depending on dataset) of the genetic variation is accounted for between the groups. These two groups however, also showed a higher level of intra-population genetic variation, probably a reflection of the large geographic area that the samples represent. AMOVA does not support the hypotheses that the different phenotypes are divergent (0.44% among groups) or that the Cape Peninsula represents a barrier to gene flow between *P. exigua* individuals from east and west of the peninsula (1.78% among groups).

Source of variation (group partitioning)	Reason for choosing groups	Genetic variation accounted for:	% variation	Fixation Indices (2 d.p.)	P values (3 d.p.)
244 bp dataset [PeWC][PeK] [PeAl]PeG] [PdE][PeW]	To test the validity of the groups defined by the MJN using both datasets (Fig. 2.2) and (358 bp dataset MJN not shown)	AG APWG WP	95.26 0.97 3.77	FSC : 0.21 FST : 0.96 FCT : 0.95	0.000 0.000 0.000
358 bp dataset [PeWC][PeK] [Port St. Johns] [PdE][PeW]		AG APWG WP	95.90 0.38 3.72	FSC : 0.09 FST : 0.96 FCT : 0.96	0.000 0.001+-0.001 0.000
244 bp dataset [PeG][PdW]	To test the rationale for classing PdW and PeG as separate taxa, despite the MSN joining them at the 95 % plausibility level (Fig. 2.5. Appendix 2.9;1)	AG APWG WP	84.51 6.97 8.52	FSC : 0.45 FST : 0.92 FCT : 0.85	0.000 0.000 0.011+-0.003
358 bp dataset [Port St. Johns] [PdW]		AG APWG WP	89.79 -0.33 10.54	FSC : -0.03 FST : 0.90 FCT : 0.90	0.000 0.556+-0.015 0.202+-0.013
358 bp dataset [Green] [Intermediate] [Unmottled]	To test the hypothesis that the different phenotypes are divergent	AG APWG WP	-0.44 10.34 90.10	FSC : 0.10 FST : 0.10 FCT : -0.00	0.000 0.000 0.634+-0.014
358 bp dataset [*P. exigua* east of Cape Peninsula] [*P. exigua* west of Cape Peninsula (excluding PeK)]	To test the hypothesis that the Cape Peninsula represents a barrier to gene flow between *P. exigua* individuals from east and west of the peninsula	AG APWG WP	1.78 9.09 89.13	FSC : 0.09 FST : 0.11 FCT : 0.018	0.000 0.000 0.183+-0.014

Table 2.8. Analysis of Molecular Variance (AMOVA) results for all *Parvulastra* samples for the cytochrome oxidase I partial gene sequence for the 244 and 358 bp datasets indicating the apriori groups tested and the reason for testing them. AG: Among Groups; APWG: Among Populations Within Groups; WP: Within Populations.

2.5.5.4. Actin Intron network

The actin intron (AI) minimum spanning network, calculated using a sub-sample of *P. exigua* and *P. dyscrita* individuals from South Africa (Fig. 2.4) supports PeK as a separate lineage which separates into an independent minimum spanning network at the 95% plausibility level (8 substitutions). The AI minimum spanning network did not divide the other haplogroups at the 95% level. However, all the *P. exigua* alleles clustered together but were separated from the *P. dyscrita* alleles by only one inferred allele. The *P. dyscrita* alleles were much more divergent, separating from each other by between three and six inferred alleles. No individuals from groups PeG, PdE, PeAI were sequenced for AI due to sample degradation.

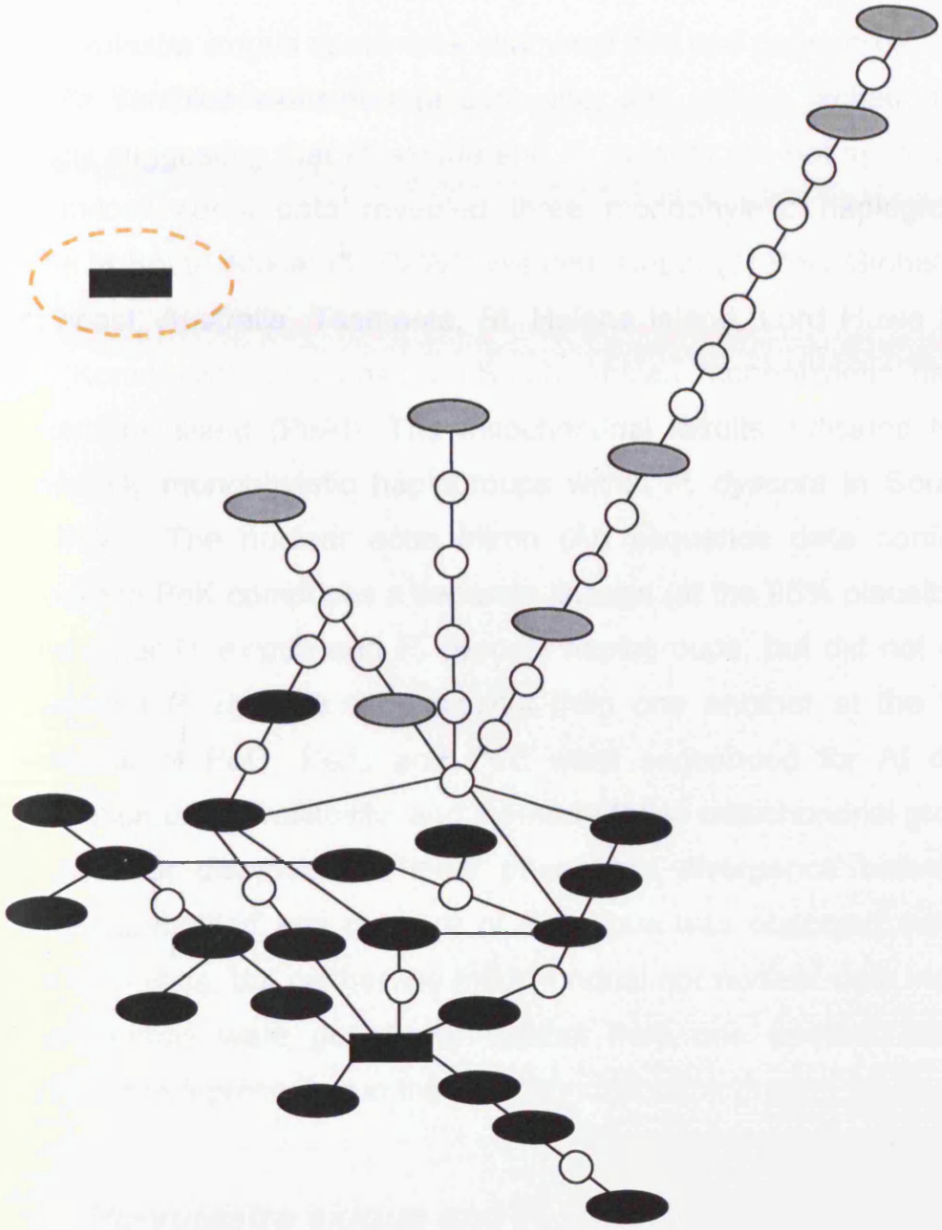

Figure 2.4. Minimum spanning network for the South African *Parvulastra* alleles for partial sequence of actin intron 3. Black ovals and rectangles represent *P. exigua* alleles. Rectangles represent alleles with a high outgroup weighting (indicating probable root haplotypes or outgroups). Grey ovals represent *P. dyscrita* alleles. Small white circles represent nucleotide differences between alleles. There are two separate allele networks (95% plausible connection limit; < 8 substitutions). The large network contains all the alleles from *P. exigua* and *P. dyscrita*, except the six divergent *P. exigua* alleles from the location Kommetjie, which are represented in the small network (outlined by a dashed orange line) which contains only one homozygote allele.

2.6. Discussion

All *Parvulastra exigua* specimens examined had oral gonopores, and all of the *P. dyscrita* samples examined (except one, see below) lacked oral gonopores, strongly suggesting that *P. exigua* and *P. dyscrita* are not synonymous species. The mitochondrial data revealed three monophyletic haplogroups within *P. exigua* in South Africa: (1) PeWC Western Cape, (2) PeG Global: South African east coast, Australia, Tasmania, St. Helena Island, Lord Howe Island, and (3) PeK (Kommetjie), and one non-South African monophyletic haplogroup from Amsterdam Island (PeAI). The mitochondrial results indicated there were two reciprocally monophyletic haplogroups within *P. dyscrita* in South Africa (PdE and PdW). The nuclear actin intron (AI) sequence data confirmed that the haplogroup PeK comprises a separate lineage (at the 95% plausibility level) from all the other *P. exigua* and *P. dyscrita* haplogroups, but did not confirm that *P. exigua* and *P. dyscrita* were distinct from one another at the 95% level. No specimens of PeG, PeAI and PdE were sequenced for AI due to sample degradation or unavailability, and therefore these mitochondrial groups cannot be supported or disproved. A clear phenotypic divergence between the colour morphs (unmottled and mottled) of *P. exigua* was observed either side of the Cape Peninsula, but neither the mitochondrial nor nuclear data indicated that the colour morphs were genetically distinct from one another, and both colour morphs were represented in the mitochondrial haplogroup PeWC.

2.6.1. *Parvulastra exigua* and *P. dyscrita* are distinct lineages

The relationship between these two lineages is complicated by the presence of more than one haplogroup within both *P. dyscrita* and *P. exigua.* However, evidence that the *P. exigua* and *P. dyscrita* lineages represent separate taxa comes from the gonopore location, as all *P. exigua* haplogroups had oral gonopores and all *P. dyscrita* haplogroups (except one individual) lacked them. One small specimen genetically identified as *P. dyscrita,* and therefore assumed

to be a juvenile, had oral gonopores. It was concluded that it was either mislabelled during collection or represents a hybrid. Hybridization does occur in asterinids (Byrne and Anderson 1994), but has not been investigated in South African *Parvulastra* species.

Although the mitochondrial DNA indicates that *P. exigua* and *P. dyscrita* are genetically distinct from one another, the nuclear sequence data did not confirm this, probably because the actin intron is not sensitive enough to detect divergence between closely related taxa. Waters *et al.* (2004b) used actin intron sequences to investigate another Australian six rayed asterinid species complex of the genus *Patiriella* (now *Meridastra,* see Fig. 1.1 Chapter 1). These authors found that the actin intron sequences only partially confirmed the COI structure between closely related species (Waters *et al.* 2004b).

Hart *et al.* (2006) found aboral gonopores in *P. exigua* museum specimens from the Cape of Good Hope, Table Bay (Green Point) and Port Elizabeth in South Africa. Although no samples were collected from Port Elizabeth during the current study, no *P. exigua* samples were detected with aboral gonopores. Due to the extensive sampling regime, it is unlikely that any groups with aboral gonopores (which would also presumably be genetically divergent) were not sampled unless they were (i) niche specialists and overlooked due to sampling bias, (ii) have become extinct or rare since the time of the museum collections, or (iii) are cryptic taxa with restricted distributions, such as at Kommetjie.

Parvulastra dyscrita is not known to occur west of the Cape Peninsula and therefore the suggestion that the museum '*P. exigua*' specimens with aboral gonopores found at Table Bay (Green Point) and Good Hope (both west of the Cape Peninsula) were misidentified *P. dyscrita* samples (Hart *et al.* 2006) seems unlikely unless the geographic distribution of *P. dyscrita* extended to the west coast in the past. The museum '*P. exigua*' with aboral gonopores found at Port Elizabeth (Hart *et al.* 2006) could have been misidentified *P. dyscrita* as *P.*

dyscrita is known to occur at Port Elizabeth (Branch *et al.* 1994). Until genetic materiel can be obtained from the '*P. exigua*' museum specimens with aboral gonopores, or live specimens of '*P. exigua*' with aboral gonopores can be found and genotyped, the identity of these anomalous museum specimens remains unknown.

2.6.2. Cryptic species within South African *Parvulastra*

Having established that the *P. exigua* and the *P. dyscrita* lineages are probably distinct from one another, this study also provides compelling evidence for a separate South African Parvulastra taxon (found only at Kommetjie to date) which is distinct from the other *P. exigua* and *P. dyscrita* haplogroups.

2.6.2.1. The divergent Kommetjie haplogroup

Twenty-six samples were sequenced for COI from Kommetjie. Thirteen of these samples were represented in the PeWC haplogroup and the remaining 12 samples formed a separate monophyletic and divergent haplogroup (PeK). These latter samples were also the only samples from South Africa which formed a separate minimum spanning network (at the 95 % plausibility level) using the AI nuclear marker. The divergence estimates between PeK and other South African *Parvulastra* lineages ranged between 7.0 and 11.7%, values much higher than other asterinid intra-specific divergence estimates (typically < 1% Waters *et al.* 2004b; reviewed by Hart *et al.* 2006; but see Waters and Roy 2004a) and as high as inter-specific estimates (6 - 26%, Hart *et al.* 1997; Hart *et al.* 2003; 2.4% - >10% (mean 26.2%); 7.5 - 12.8%; 1.1 - 4.3% Waters *et al.* 2004a; Waters *et al.* 2004b;). Genetic evidence of reproductive isolation in this lineage further comes from the presence of NUMTS or heteroplasmic variants in the COI sequences and a four base pair indel in the AI sequences, only evident in the individuals in the PeK haplogroup. This evidence implies little or no genetic exchange between this group and the other *Parvulastra* lineages as the NUMTS or heteroplasmic variants

and the indel have not been integrated into the other *Parvulastra* lineages, and they would be evident if genetic exchange was occurring.

Assigning species status can be problematic, especially as the underlying concepts are hotly debated (Turelli *et al.* 2001; Orr 2001). Defining species based on the 'Biological Species Concept' (Mayr 1942) which states that taxa must be reproductively isolated, is difficult to confirm without breeding experiments. Furthermore, other than a subjective difference in colour morph, no other diagnostic morphological characteristics were identified between PeK and the other *P. exigua* haplogroups. Therefore other methods were employed to investigate if the haplogroup PeK deserved separate taxon status. Monophyly at mitochondrial sequences and significant differences at nuclear loci are taken as practical guidelines for defining evolutionary significant units (ESUs; Moritz 1994). Moreover, the Phylogenetic Species Concept (PSC; Rosen 1978; Nixon and Wheeler 1990; Cracraft 1992) indicates that species can be defined on the basis of reciprocal monophyly of mitochondrial DNA alone. Hart *et al.* (2006) state that in the absence of observed phenotypic differences, mitochondrial sequence divergence of 5-7% among echinoderms is often interpreted as evidence for cryptic speciation (Lessios *et al.* 2001; O'Loughlin *et al.* 2003). Hart *et al.* (2006) also states that marine phylogeographers have interpreted haplotype samples that break down into separate 95% plausible networks (as seen in both the AI and COI data in this study) as evidence of multiple biological species (Tarjuelo *et al.* 2004; Uthicke *et al.* 2004). Therefore on the basis of the genetic evidence presented here, the PeK lineage should be classified (at least) as an ESU, and is probably worthy of classification as a new *Parvulastra* species. However, until further genetic, morphological or reproductive investigation occurs, this remains to be confirmed.

The ML analysis clustered PeK with the outgroup taxa, *P. parvivipara* and *P. vivipara*, in a sister clade to the *P. exigua* / *P. dyscrita* complex. The PeK haplogroup appears to share several similarities with these species. Firstly, like

PeK, both *P. parvivipara* and *P. vivipara* occur sympatrically within *P. exigua*'s range in Australia (Byrne and Cerra 1996). Secondly, a characteristic of *P. parvivipara* and *P. vivipara* is an extremely restricted geographic distribution. *Parvulastra parvivipara* is found only on west side of the Eyre Peninsula where it is recorded from five locations (Keough and Dartnall 1978) and *P. vivipara* is endemic to south east Tasmania where it is recorded from only four locations on the island (Dartnall 1969 cited in Dartnall *et al.* 2003; Byrne 1996). Both species are intertidal, inhabit about 100 m of the shoreline at each location where they occur and have a geographic distributions spanning only 150 - 300 km. These are the most restricted distributions known in echinoderms (Byrne and Cerra 1996). Although rocky shore locations directly either side of Kommetjie were not searched, no members of the PeK haplogroup were detected at the sampling locations either side of Kommetjie indicating this group also has a very restricted distribution. Thirdly, both *P. parvivipara* and *P. vivipara* species are morphologically very similar to *P. exigua* and were originally classified as *P. exigua* and have only recently been recognized as separate species (Dartnall 1969 cited in Dartnall *et al.* 2003; Dartnall 1971; Keough and Dartnall 1978; Byrne 1996; Byrne and Cerra 1996). In museum *P. parvivipara* and *P. vivipara* collections, examination of the gonads and the presence of juveniles is often necessary for species identification (Byrne 1996). However, living *P. parvivipara* and *P. vivipara* specimens are easily distinguished from *P. exigua* by their distinct yellow orange colour (Byrne and Cerra 1996). All of the PeK group were recorded as the intermediate colour morph, and on subsequent examination of photographs taken at the time of sampling it was noted that all of these animals appeared to have a blotchy red-orange colour in comparison to the other *P. exigua* intermediate colour morphs. However, other than this subjective difference, PeK appears very similar to *P. exigua.*

Speculation over how PeK might fit into the evolutionary sequence of reproductive modes in this family is intriguing. It has been suggested that *P. parvivipara* and *P. vivipara* were derived from an 'exigua' like ancestor,

supported by molecular data, approximately three million years ago (Dartnall 1971; Keough and Dartnall 1978; Byrne and Cerra 1996; Byrne 1996; Hart *et al.* 2003; Byrne 2006). *Parvulastra parvivipara* and *P. vivipara* are both hermaphroditic and viviparous species (Byrne and Cerra 1996), releasing newborn juveniles from aboral gonopores. The PeK lineage has oral gonopores and is therefore unlikely to be viviparous, however it is unknown if PeK is hermaphroditic. Other asterinid species with oral gonopores where the method of reproduction has been investigated are *P. exigua*, *Asterina gibossa*, *Asterina phylactica, Aquilonastra minor* and *Aquilonastra scobinata* (see Byrne 2006). All of these species, except *A. phylactica,* produce large yolky eggs from oral gonopores which develop into benthic larvae without parental care (Byrne 2006). However, *A. phylactica* produces large yolky eggs from oral gonopores but broods its eggs until they develop into benthic larvae (Emson and Crump 1984). Furthermore, Strathmann *et al.* (1984) observed occasional development of embryos and larvae in the gonad of *A. phylactica* indicating that internal brooding is possible, and that brooding may be subject to plasticity (Hart *et al.* 2003). *A. phylactica* is also characterized by a somewhat restricted geographic distribution. Morphological evidence suggests that the benthic life history of *P. exigua*, which abandons its eggs, gave rise to the evolution of external brooding of eggs produced from oral gonopores, with occasional retention of eggs in the gonads, as seen in *A. phylactica* (see Strathmann *et al.* 1984) and ultimately to intragonadal brooding, as seen in *P. parvivipara* and *P. vivipara* (see Byrne 1996; Byrne and Cerra 1996; Byrne 2006). This sequence of reproductive evolution, as well as the other similarities that PeK has to *A. phylactica* (oral gonopores, restricted dispersal) and *P. parvivipara* and *P. vivipara* suggests that PeK may represent the 'reproductive transition step' (with a life cycle similar to *A. phylactica*) between the benthic life cycle of *P. exigua* and the viviparous life cycle of *P. parvivipara* and *P. vivipara.*

The distribution of *P. parvivipara* in two small populations on either side of a 60 km long peninsula suggests that this species is a remnant of a previously wider

range and its population size may be diminishing. Therefore it has been suggested that this species may be a good candidate for conservation (Byrne and Cerra 1996). In light of the similarities between *P. parvivipara* and the PeK lineage, conservation status for the PeK haplogroup may also be a consideration. Clearly PeK warrants further investigation to clarify the taxonomic status, distribution, reproductive mode and evolution of this lineage.

2.6.2.2. *Parvulastra exigua* haplogroups

The South African *P. exigua* mitochondrial haplogroups (PeWC and PeG) are divergent from one another by approximately 6% (244 bp COI data) and 4% (358 bp COI data), and comprise independent networks at the 95% plausibility level. Waters and Roy (2004a) reported large intraspecific divergences between *P. exigua* populations Cape Town and Australia (6 - 7.4%) and Cape Town and eastern South Africa (5.9 - 7.2%), whereas divergences within sites here were always < 1%. Using the same reasoning applied to the haplogroup PeK above (separate networks at the 95% plausibility level; large divergence estimates between haplogroups; reciprocal monophyly of mitochondrial DNA indicating separate species under the PSC), haplogroups PeWC and PeG could be classed as separate species. However, without either morphological or nuclear (no samples from the haplogroup PeG were sequenced for AI) evidence of distinction between them, this classification may be premature. Mitochondrial DNA can define the pattern of reciprocal monophyly more clearly than nuclear markers because of its smaller effective population size: divided populations will drift to reciprocal monophyly more quickly (Hudson 1992; Hellberg *et al.* 2002; Waters *et al.* 2004b). Therefore reciprocal monophyly in the haplogroups PeWC and PeG may merely represent a period of allopatry in the past and not represent lack of present day gene flow or reproductive isolation.

2.6.2.3. *Parvulastra dyscrita* haplogroups

The two *P. dyscrita* haplotypes were divergent from one another by 7.7% (244 bp COI data) and 4.7% (358 bp COI data), and separated into different networks at the 95% plausibility level. Therefore the same line of reasoning applied to the *P. exigua* haplogroups PeWC and PeG applies to the *P. dyscrita* haplogroups.

2.7. Conclusions

Parvulastra exigua, *P. dyscrita* and the divergent group found at Kommetjie (PeK) seem to fall within currently accepted parameters for distinct species. Within both *P. exigua* and *P. dyscrita* in South Africa two lineages were identified. These are not thought to represent further cryptic species, but additional genetic and morphological investigation is required to confirm their taxonomic status. This study cannot confirm the existence of '*P. exigua*' samples with aboral gonopores from South Africa, nor confirm that the anomalous '*P. exigua*' museum specimens with aboral gonopores were misclassified as *P. dyscrita*. As this study provides no evidence to indicate that the wide distribution of *P. exigua* can be explained by the presence of cryptic lineages (see Hart *et al.* 2006), the dispersal mode by which benthic *P. exigua* in the haplogroup PeG achieved such a widespread distribution remains a paradox. The suggestion that Africa is the ancestral origin of *P. exigua* (see Waters and Roy 2004a) is premature. Waters *et al.* hypothesize that if identical haplotypes on different landmasses are observed (as seen in this study) that the most likely explanation for the distributions would be human translocation associated with early shipping activities. However, as none of the seastar species are commercially important, and neither *P. exigua* nor the new Kommetjie lineage has planktonic larvae that may be caught in ballast water, this explanation seems unlikely. The method of dispersal responsible for the observed haplotypic distribution remains unclear. The results of this study concur with the conclusions of Hart *et al.* (2006) in that further population genetic analyses within identified *Parvulastra* species are required before the origin and direction of gene flow between the global populations of *P. exigua* can be ascertained with certainty.

2.8. References

Arndt A, Marquez C, Lambert P, Smith MJ (1996) Molecular phylogeny of Eastern Pacific sea cucumbers (*Echinodermata: Holothuriodea*) based on mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution* **6**, 425-437.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Booth JD, Ovenden JR (2000) Distribution of *Jasus spp.* (*Decapoda: Palinuridae*) *phyllosomas* in southern waters: implications for larval recruitment. *Marine Ecology-Progress Series* **200**, 241-255.

Branch GM, Branch ML (1980) Competition between *Cellana tramoserica* (Sowerby) (*Gastropoda*) and *Patiriella* exigua (Lamarck) (*Asteroidea*), and their influence on algal standing stocks. *Journal of Experimental Marine Biology and Ecology* **48**, 35-49.

Branch GM, Griffiths CL, Branch ML, Beckley LE (1994) *Two Oceans, A guide to the marine life of southern Africa* David Philip Publishers (Pty) Ltd. Cape Town.

Briggs JC (2006) Proximate sources of marine biodiversity. *Journal of Biogeography* **33**, 1-10.

Brown AC, Jarman N (1978) Coastal Marine Habitats. In: *Biogeography and Ecology of Southern Africa* (eds. Werger MJA, Bruggen ACV), pp. 1239 - 1227. The Hague, Cape Town.

Bustamante RH, Branch GM (1996) Large scale patterns and trophic structure of southern African rocky shores: The roles of geographic variation and wave exposure. *Journal of Biogeography* **23**, 339-351.

Byrne M (1992) Reproduction of sympatric populations of *Patiriella gunnii, P. calcar* and *P. exigua* in New-South-Wales, Asterinid seastars with direct development. *Marine Biology* **114**, 297-316.

Byrne M (1995) Changes in larval morphology in the evolution of benthic development by *Patiriella exigua* (*Asteroidea, Asterinidae*), a comparison with the larvae of *Patiriella* species with planktonic development. *Biological Bulletin* **188**, 293-305.

Byrne M (1996) Viviparity and intragonadal cannibalism in the diminutive sea stars *Patiriella vivipara* and *P. parvivipara* (Family Asterinidae). *Marine Biology* **125**, 551-567.

Byrne M (2006) Life history diversity and evolution in the Asterinidae. *Integrative and Comparative Biology* **46**, 243-254.

Byrne M, Anderson MJ (1994) Hybridization of sympatric *Patiriella* Species (*Echinodermata, Asteroidea*) in New-South-Wales. *Evolution* **48**, 564-576.

Byrne M, Cerra A (1996) Evolution of intragonadal development in the diminutive asterinid sea stars *Patiriella vivipara* and *P. parvivipara* with an overview of development in the *Asterinidae*. *Biological Bulletin* **191**, 17-26.

Byrne M, Cerra A, Villinski JT (1999) Oogenic strategies in the evolution of development in *Patiriella* (*Echinodermata: Asteroidea*). *Invertebrate Reproduction & Development* **36**, 195-202.

Campbell AC, Rowe FWE (1997) A new species in the *asterinid* genus *Patiriella* (*Echinodermata, Asteroidea*) from Dhofar, Southern Oman: a temperate taxon in a tropical locality. *Bulletin of the British museum, Natural History (Zoology)* **63**.

Chen BY, Chen CP (1992) Reproductive-cycle, larval development, juvenile growth and population-dynamics of *Patiriella pseudoexigua* (*Echinodermata, Asteroidea*) in Taiwan. *Marine Biology* **113**, 271-280.

Clark HL (1923) The Echinoderm Fauna of South Africa. *Annals of the South African Museum XIII.*

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Colgan DJ, Byrne M, Rickard E, Castro LR (2005) Limited nucleotide divergence over large spatial scales in the asterinid sea star *Patiriella exigua. Marine Biology* **146**, 263-270.

Cracraft J (1992) The species of the Birds-of-Paradise (Paradisaeidae) - Applying the Phylogenetic Species Concept to a complex pattern of diversification. *Cladistics - the International Journal of the Willi Hennig Society* **8**, 1-43.

Dartnall AJ (1969) *The taxonomy and biogeography of the sea star genus Patiriella in Tasmania* Unpublished MSc thesis, University of Tasmania.

Dartnall AJ (1971) Australian Seastars of the genus *Patiriella* (*Asteroidea, Asterinidae*). *Proceedings of the Linnean Society of New South Wales* **96**.

Dartnall AJ, Byrne M, Collins J, Hart MW (2003) A new viviparous species of asterinid (*Echinodermata, Asteroidea, Asterinidae*) and a new genus to accommodate the species of pan-tropical exiguoid sea stars. *Zootaxa* **359**, 1-14.

Emanuel BP, Bustamante RH, Branch GM, Eekhout S, Odendaal FJ (1992) A zoogeographic and functional approach to the selection of marine reserves on the west coast of South Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **12**, 341-354.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online, Application Note* **2005**, 47-50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Felsenstein J (1985) Confidence limits on phylogenies - an approach using the bootstrap. *Evolution* **39**, 783-791.

Gray JE (1840) A synopsis of the genera and species of the class *Hypostoma* (*Asterias Linnaeus*). *The Annals and Magazine of Natural History* **1**, 175-184; 275-290.

Halperin E, Eskin E (2004) Haplotype reconstruction from genotype data using Imperfect Phylogeny. *Bioinformatics* **20**, 1842-1849.

Hart M (2000) Phylogenetic analyses of mode of larval development. *Seminars in Cell & Developmental Biology* **11**, 411-418.

Hart MW, Byrne M, Johnson SL (2003) *Patiriella pseudoexigua* (*Asteroidea: Asterinidae*): a cryptic species complex revealed by molecular and embryological analyses. *Journal of the Marine Biological Association of the United Kingdom* **83**, 1109-1116.

Hart MW, Byrne M, Smith MJ (1997) Molecular phylogenetic analysis of life-history evolution in Asterinid starfish. *Evolution* **51**, 1848-1861.

Hart MW, Keever CC, Dartnall AJ, Byrne M (2006) Morphological and genetic variation indicate cryptic species within Lamarck's little sea star, *Parvulastra* (=*Patiriella*) *exigua. Biological Bulletin* **210**, 158-167.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.

Hiller N (1994) The environment, biogeography, and origin of the southern African recent brachiopod fauna. *Journal of Paleontology* **68**, 776-786.

Hudson RR (1992) Gene trees, species trees and the segregation of ancestral alleles. *Genetics* **131**, 509-512.

Hunt A (1993) Effects of contrasting patterns of larval dispersal on the genetic connectedness of local populations of 2 intertidal starfish, *Patiriella calcar* and *P. exigua. Marine Ecology-Progress Series* **92**, 179-186.

Jackson L (1976) Aspects of the intertidal ecology of the east coast of South Africa, pp. 1-72. Oceanographic Research Institute, Durban.

Jeffery CH, Emlet RB, Littlewood DTJ (2003) Phylogeny and evolution of developmental mode in temnopleurid echinoids. *Molecular Phylogenetics and Evolution* **28**, 99-118.

Keough MJ, Dartnall AJ (1978) A new species of viviparous asterinid asteroid from Eyre Peninsula, South Australia. *Records of the South Australian Museum* **17**, 407-416.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics* **24**, 189-216.

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* **420**, 73-90.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

Lamarck JBPA (1816) *Histoire Naturelle des Animaux Sans Vertebres* Verdiere, Paris.

Laudien J, Flint NS, van der Bank FH, Brey T (2003) Genetic and morphological variation in four populations of the surf clam *Donax serra* (Roding) from southern African sandy beaches. *Biochemical Systematics and Ecology* **31**, 751-772.

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution* **55**, 955-975.

Mayr E (1942) *Systematics and the origin of species* Columbia University Press, New York.

Moritz C (1994) Defining Evolutionarily-Significant-Units for conservation. *Trends in Ecology & Evolution* **9**, 373-375.

Neraudeau D, Mathey B (2000) Biogeography and diversity of South Atlantic Cretaceous echinoids: implications for circulation patterns. *Palaeogeography Palaeoclimatology Palaeoecology* **156**, 71-88.

Nixon KC, Wheeler QD (1990) An amplification of the Phylogenetic Species Concept. *Cladistics-the International Journal of the Willi Hennig Society* **6**, 211-223.

O'Loughlin PM (2002) New genus and species of southern Australian and Pacific *Asterinidae* (*Echinodermata, Asteroidea*). *Memoirs of Museum Victoria* **59**, 277-296.

O'Loughlin PM, Waters JM (2004) A molecular and morphological revision of the genera of *Asterinidae* (*Echinodermata: Asteroidea*). *Memoirs of Museum Victoria* **61**, 1-40.

O'Loughlin PM, Waters JM, Roy MS (2002) Description of a new species of *Patiriella* from New Zealand, and review of *Patiriella regularis* (*Echinodermata, Asteroidea*) based on morphological and molecular data. *Journal of the Royal Society of New Zealand* **32**, 697-711.

O'Loughlin PM, Waters JM, Roy MS (2003) A molecular and morphological review of the *asterinid, Patiriella gunnii* (Gray) (*Echinodermata: Asteroidea*). *Memoirs of Museum Victoria* **60**, 181-195.

Orr HA (2001) The genetics of species differences. *Trends in Ecology & Evolution* **16**, 343-350.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, S146-S158.

Palumbi SR (2004) Marine reserves and ocean neighborhoods: The spatial scale of marine populations and their management. *Annual Review of Environment and Resources* **29**, 31-68.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Ridgway TM, Branch GM, Stewart BA, Hodgson AN (2000) Taxonomic status of the '*Patella miniata*' species complex (*Mollusca: Gastropoda*) in southern Africa. *Hydrobiologia* **420**, 103-118.

Ridgway TM, Stewart BA, Branch GM, Hodgson AN (1998) Morphological and genetic differentiation of *Patella granularis* (*Gastropoda: Patellidae*): recognition of two sibling species along the coast of southern Africa. *Journal of Zoology* **245**, 317-333.

Riginos C, Victor BC (2001) Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 1931-1936.

Rosen DE (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology* **27**, 159-188.

Rozen S, Skaletsky H, (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods*

and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.

Sponer R, Roy MS (2002) Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (*Echinodermata*) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution* **56**, 1954-1967.

Strathmann RR, Strathmann MF, Emson RH (1984) Does limited brood capacity link adult size, brooding, and simultaneous hermaphroditism - a test with the starfish *Asterina phylactica*. *American Naturalist* **123**, 796-818.

Swofford DL (1998) *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10* Sinauer, Sunderland, Massachusetts.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tarjuelo I, Posada D, Crandall KA, Pascual M, Turon X (2004) Phylogeography and speciation of colour morphs in the colonial ascidian *Pseudodistoma crucigaster*. *Molecular Ecology* **13**, 3125-3136.

Tolley KA, Groeneveld JC, Gopal K, Matthee CA (2005) Mitochondrial DNA panmixia in spiny lobster *Palinurus gilchristi* suggests a population expansion. *Marine Ecology-Progress Series* **297**, 225-231.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.

Uthicke S, O'Hara TD, Byrne M (2004) Species composition and molecular phylogeny of the Indo-Pacific teatfish (*Echinodermata : Holothuroidea*) beche-de-mer fishery. *Marine and Freshwater Research* **55**, 837-848.

Waters JM, O'Loughlin PM, Roy MS (2004a) Molecular systematics of some Indo-Pacific asterinids (*Echinodermata, Asteroidea*): does taxonomy reflect phylogeny? *Molecular Phylogenetics and Evolution* **30**, 872-878.

Waters JM, O'Loughlin PM, Roy MS (2004b) Cladogenesis in a starfish species complex from southern Australia: evidence for vicariant speciation? *Molecular Phylogenetics and Evolution* **32**, 236-245.

Waters JM, Roy MS (2003) Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography* **30**, 1787-1796.

Waters JM, Roy MS (2004a) Out of Africa: The slow train to Australasia. *Systematic Biology* **53**, 18-24.

Waters JM, Roy MS (2004b) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology* **13**, 2797-2806.

2.9. Appendix 1.

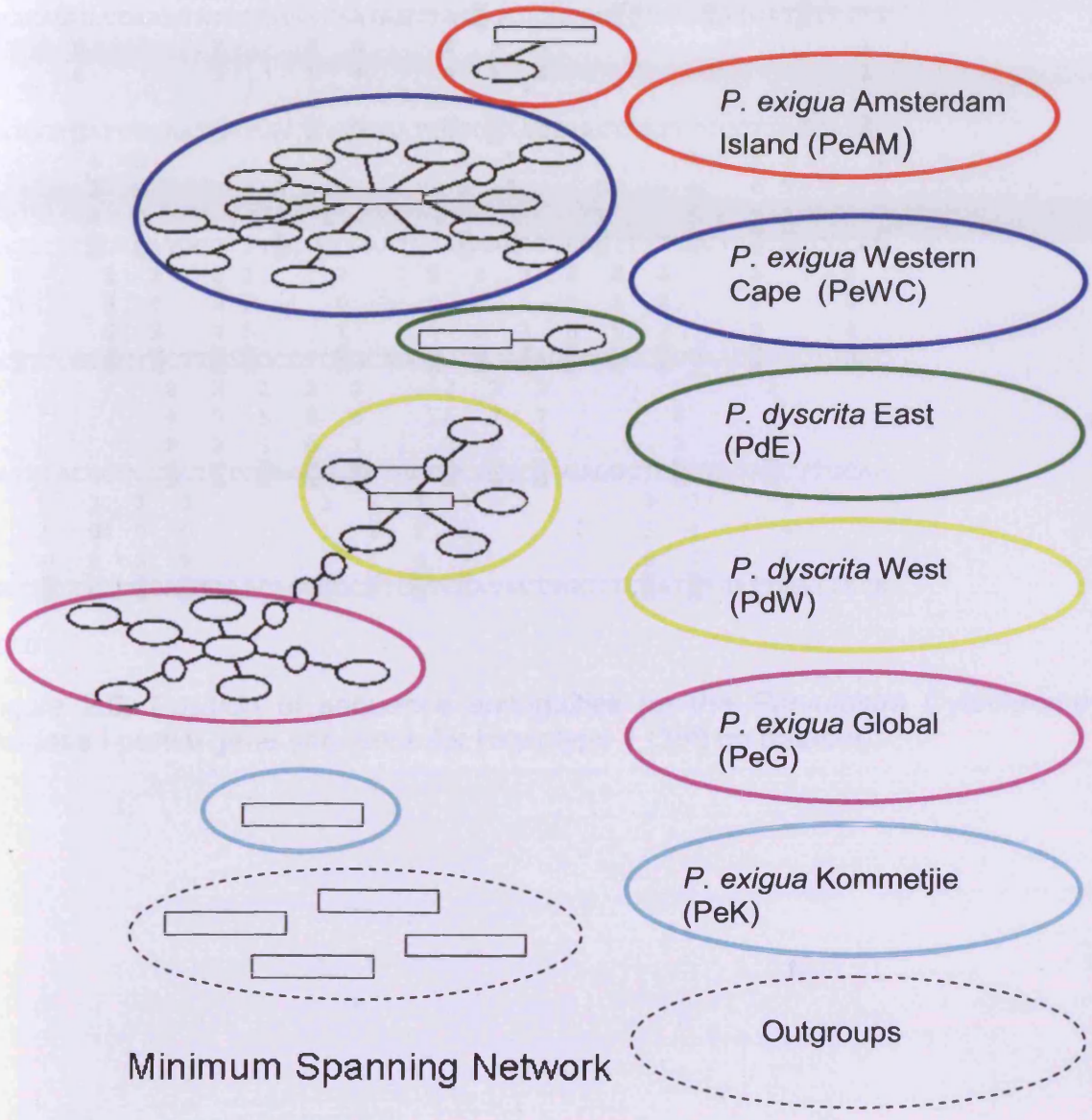

Figure 2.5. Minimum spanning network: Ovals represent haplotypes, rectangles represent haplotypes with a high outgroup weighting (indicating probable root haplotypes or outgroups). Smaller circles represent nucleotide differences between haplotypes. There are four separate haplotype networks (95% plausible connection limit; < 6 substitutions), with one network separated into two sub-groups. Outgroups included: *Parvulastra vivipara, P. parvivipara, Patiriella regularis, M. mortenseni.*

2.9. Appendix 2.

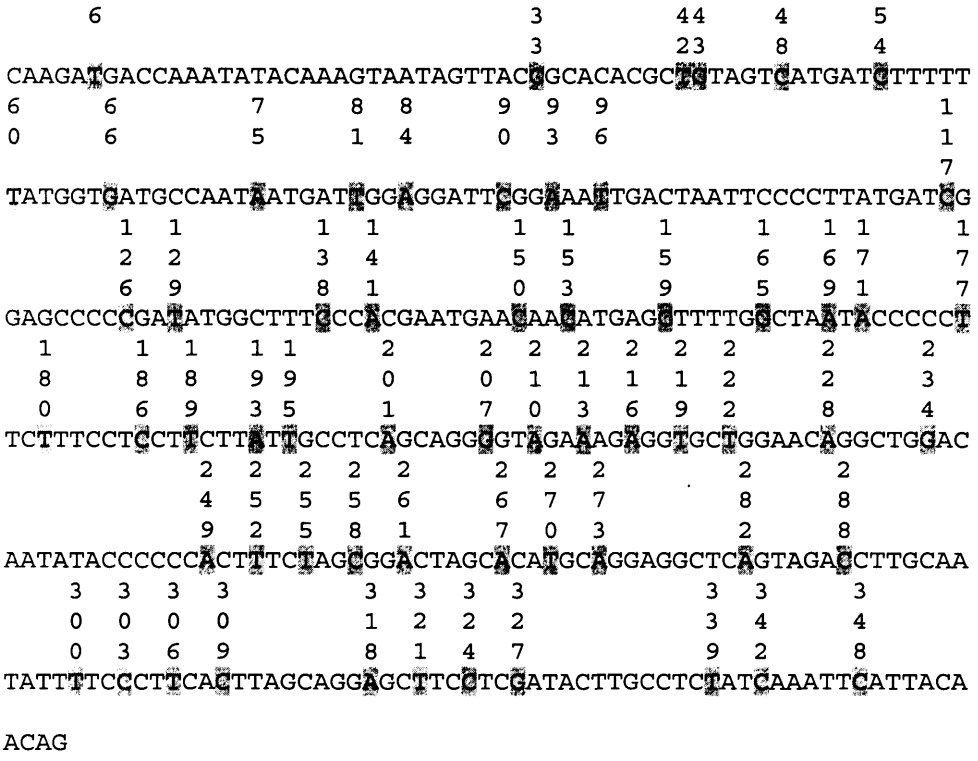

Figure 2.6. Position of sequence ambiguities for the *Parvulastra* Cytochrome Oxidase I partial gene sequence for Haplotype 1 (358 bp dataset).

2.9. Appendix 3.

```
            6344456678 8999111111 1111111111 1222222222 2222222222 3333333333 3
             323840651 4036122345 5566778889 9001111223 4555667788 0000122234 4
                   769810 3959170693 5170369284 9258170328 0369814792 8

Hap 1       TCTCCCTGAT ACATCCTCAC CCGAATTCTA TAGAAATTAG ATTCAATAAC TCTCATCGTC C
Hap 2       .......... .......... .......... .......... .C........ .......... .
Hap 3       .......... .......... ..A....... .......... .......... .......... .
Hap 4       .......... ....T..... .......... .......... .......... .......... .
Hap 5       .......... ....T..... .......... ....G..... .......... .......... .
Hap 6       CTC.T..T.C GT..TA.T.T T.A..C.TA. A.A..T.C.A ...TGT.CCT .A.T...AA. T
Hap 7       .....T.... .......... .......... .......... .......... .......... .
Hap 8       .......... ...C...... .......... .......... .......... .......... .
Hap 9       .......... .......... .......... .......... ...A...... .........T .
Hap 10      .......... .......... .......... ...G...... .......... .......... .
Hap 11      .......... ..G....... .......... .......... .......... .......... .
Hap 12      .......... .........T ..A....... .......... .......... .......... .
Hap 13      .......... .......... .......... .......... ..G....... .......... .
Hap 14      .......... .......... .......... .......... ...T...... .......... .
Hap 15      ......C... .......... .......... .......... .......... .......... .
Hap 16      .......A.. .......... .......... .......... .......... .......... .
Hap 17      .......... .......... .......... ........G. .......... .......... .
Hap 18      .......... .......... .......... .......... .......... .......C.. .
Hap 19      .......... .......... .......... .......... ......C... ..C....... .
Hap 20      .......... .......... .......... .......C.. .......... ..C....... .
Hap 21      .........C .......... .......... .......... .......... .......... .
Hap 22      .......... .......... .......... .......... .......... ..C....... .
Hap 23      .......... .......... .......... .......... .......C.. .......... .
Hap 24      ...T...... .......... .......... .......... .......... .......... .
Hap 25      .......... ....TA.... T.A...CT.. ..A....C.A ........CT .A...C.... .
Hap 26      .......C.. ....TA.... T.A...CT.. ..A....C.A ........CT .A...C.... .
Hap 27      .......... ....TAC... T.A...CT.. ..A....C.A ........CT .A...C.... .
Hap 28      .......... ....TA.... T.A...CT.. ..A....C.A G.......CT .A...C.... .
Hap 29      ....T..A.. ....TA.... ..A....T.. ..A....... ........CT .A...C.A.. .
Hap 30      ....T..... ....TA.... ..A....T.. ..A....... ........CT .A...C.... .
Hap 31      ....T..A.. ....TA.... ..A.C..T.. ..A..G.... ........CT .A...C.A.. .
Hap 32      ....T..A.. ....TA...T ..A....T.. ..A....... ........CT .A...C.A.. .
Hap 33      ....T..A.. ....TA.... ..A....T.. ..A..C.... ........CT .A...C.A.. .
Hap 34      ....T..A.. ....TA.... ..A....T.. ..A.G..... ........CT .A...C.A.. .
Hap 35      C.C.T..AT. ....TA.T.T TTA....... ..T...A... .......G.T CA..GC..C. .
Hap 36      C.C.T..AT. ....TA.T.T TTA....... ..T...A... ....G..G.T CA..GC..C. .
```

Figure 2.7. Alignment of the sequence ambiguities for the *Parvulastra* Cytochrome Oxidase I partial gene for all haplotypes from chapter 2 indicating the 358 bp (all haplotypes up to and including Hap36) and 244 bp datasets (next page: all haplotypes including and below 'SAMAIN', with first 14 ambiguous bases marked as a dash, as they are not included in the 244 bp dataset).

```
          6344456678 8999111111 1111111111 1222222222 2222222222 3333333333 3
           323840651 4036122345 5566778889 9001111223 4555667788 0000122234 4
                     769810 3959170693 5170369284 9258170328 0369814792 8
SAMAIN    ---------- ----..... .......... .......... .......... .......... .
Hap 02    ---------- ----..... .......... .......... .C........ .......... .
Hap 03    ---------- ----..... ..A....... .......... .......... .......... .
Hap 04    ---------- ----T..... .......... .......... .......... .......... .
Hap 05    ---------- ----T..... .......... ....G..... .......... .......... .
Hap 06    ---------- ----TA.T.T T.A..C.TA. A.A..T.C.A ...TGT.CCT .A.T...AA. T
Hap 09    ---------- ----..... .......... .......... ...A...... .........T .
Hap 10    ---------- ----..... .......... ...G...... .......... .......... .
Hap 12    ---------- ----.....T ..A....... .......... .......... .......... .
Hap 13    ---------- ----..... .......... .......... ..G....... .......... .
Hap 14    ---------- ----..... .......... .......... ...T...... .......... .
Hap 17    ---------- ----..... .......... ........G. .......... .......... .
Hap 18    ---------- ----..... .......... .......... .......... .......C.. .
Hap 19    ---------- ----..... .......... .......... ......C... ..C....... .
Hap 20    ---------- ----..... .......... .......C.. .......... ..C....... .
Hap 22    ---------- ----..... .......... .......... .......... ..C....... .
Hap 23    ---------- ----..... .......... .......... .......C.. .......... .
25 26     ---------- ----TA.... T.A...CT.. ..A....C.A ........CT .A...C.... .
Hap 27    ---------- ----TAC... T.A...CT.. ..A....C.A ........CT .A...C.... .
Hap 28    ---------- ----TA.... T.A...CT.. ..A....C.A G.......CT .A...C.... .
Hap 29    ---------- ----TA.... ..A....T.. ..A....... ........CT .A...C.A.. .
Hap 30    ---------- ----TA.... ..A....T.. ..A....... ........CT .A...C.... .
Hap 31    ---------- ----TA.... ..A.C..T.. ..A..G.... ........CT .A...C.A.. .
Hap 32    ---------- ----TA...T ..A....T.. ..A....... ........CT .A...C.A.. .
Hap 33    ---------- ----TA.... ..A....T.. ..A..C.... ........CT .A...C.A.. .
Hap 34    ---------- ----TA.... ..A....T.. ..A.G..... ........CT .A...C.A.. .
Hap 35    ---------- ----TA.T.T TTA....... ..T...A... .......G.T CA..GC..C. .
Hap 36    ---------- ----TA.T.T TTA....... ..T...A... ....G..G.T CA..GC..C. .
LH        ---------- ----TA..G. T.A...CT.. ..A....C.A ........CT .A...C.... T
SAFBD     ---------- ----TA.... T.A...CT.. ..A....C.A G..T....CT .A...C.... .
SH1       ---------- ----TA.... T.AG..CT.. ..A....C.A ...A....CT .A...C.... .
TASTW     ---------- ----TA.... T.A...CT.. ..A.G..C.A ........CT .A...C.... .
AM1       ---------- ----..... .........G .CA....... .....G...T .....CTA.. .
AM2       ---------- ----..... .........G .CA....... .........T .....CTA.. .
```

Chapter 3

Lack of phenotypic and molecular congruence in an intertidal sea star: implications for ecological adaptation and phylogeography

Chapter 3: Lack of phenotypic and molecular congruence in an intertidal sea star: implications for ecological adaptation and phylogeography

K. S. Dunbar[1], J. Cable[1], J. Lello[1], C. Griffiths[2], M.W. Bruford[1].

[1]School of Biosciences, Cardiff University, Cardiff CF10 3TL, U.K.

[2]Marine Biology Research Institute, University of Cape Town, Rondebosch, South Africa.

3.1. Abstract

The processes which have led to present day patterns of phenotypic and phylogeographic structure in the intertidal fauna of South Africa are poorly understood. Furthermore, the processes maintaining assemblages of intertidal fauna in biogeographic provinces have never been investigated in terms of gene flow or phylogeographic structure. South Africa is an ideal location to investigate such processes as the intertidal zone is divided into biogeographic provinces defined by temperature. Using a large scale sampling regime, habitat survey, phenotypic and neutral genetic diversity analyses, the impact of these processes was investigated in a continuously distributed intertidal sea star, *Parvulastra exigua*. This species displays phenotypic structure with an 'unmottled' morph inhabiting the cold west coast and a 'mottled' morph inhabiting the warm east coast. It also consists of two deeply divergent monophyletic clades corresponding to populations on the east and west coasts, with the west showing a rapid population expansion. However, the phenotypic and genotypic divides do not occur in the same geographic location. Furthermore, the different phenotypes showed different habitat preferences indicating that particular habitats might incur a selective advantage. These combined results suggest that colour polymorphism in this species is not a product of neutral genetic structure, but is more likely caused by selective forces or phenotypic plasticity corresponding to the different temperature regimes on the east and west coasts of South Africa.

3.2. Introduction

Genetic variation has long been supposed to enable the evolution, adaptability and survival of organisms living in heterogeneous habitats, but with the general prediction that the evolution of such populations and species in heterogeneous habitats is often tightly linked to their ecology (Hedrick 1986). Rocky shores are one of the most extreme environments with respect to physical heterogeneity, with variable substratum and sharp gradients in abiotic variables, such as temperature, salinity and wave action (Johannesson 2003). Population survival depends on phenotypic (e.g. colour morphs or shell shape), behavioural (e.g. habitat choice or diurnal / nocturnal feeding patterns) and physiological adaptations (e.g. salinity, humidity or temperature tolerance) which may lead to different ecotypes evolving in response to local selection pressures, differing in one or several adaptive traits. Spatial and temporal variation overlying the physical heterogeneity and selection pressures combine to mould the genetic structure of organisms. Neutral molecular markers can be employed to determine the demographic structure of species. Demographic patterns and processes can then be compared to phenotypic structure and habitat preferences to provide baseline ecological knowledge with which inferences can be made about how historical and present day environmental conditions affect the genetic, phenotypic and phylogeographic structure of species (Brisson *et al.* 2005).

3.2.1. Phenotypic diversity

Colour variation is common in marine invertebrates (Haylor 1984; Williams and Benzie 1998; Merilaita 2001; Sokolova and Berger 2000; Sponer *et al.* 2001; Johannesson and Ekendahl 2002; Mackenzie *et al.* 2004; Tarjuelo *et al.* 2004; Le Gac *et al.* 2004; Stoletzki and Schierwater 2005) but its significance is usually unknown (reviewed in Gray and McKinnon in press). Colour polymorphism could be the result of three major processes. Firstly, neutral processes or population structure may be responsible for colour morph frequencies (Oxford 2005; Hoffmann *et al.* 2006), and colour polymorphisms which coincide with neutral

genetic variation may represent cryptic species (Haylor 1984; Tarjuelo *et al.* 2004) or founder events (Le Gac *et al.* 2004). Secondly, colour variation could be controlled by phenotypic plasticity. This is defined as the production of multiple phenotypes from a single genotype, depending on environmental conditions, and can be expressed either within the lifespan of an individual or across generations (Miner *et al.* 2005). For example, Hull *et al.* (2001) reported that the base colour frequency of species in the genus *Idotea* was derived from the algal pigments in the diet, and if reared on different substrates, base colour could change within weeks (Lee 1966, cited in Hull *et al.* 2001). Thirdly, colour variation could be an adaptation in response to selective forces (i.e. reflecting a genetically based response to selection). In marine invertebrates the selective pressure(s) responsible for colour variation can be difficult to identify. The evolution of cryptic colouration is thought to be closely linked with the evolution of habitat selection (de Meeus *et al.* 1993) and may reduce the risk of detection by predators (Hull *et al.* 2001). A clear link is often shown between colour variants and selective advantage (Haylor 1984; Merilaita 2001; Johannesson and Ekendhal 2002; Stoletzki and Schierwater 2005). Distinguishing between these three processes when trying to identify the cause of colour variation can be challenging and requires knowledge of the species ecology and genetics. For example, Gillespie and Oxford (1998) reported that colour morph frequencies in one of the most intensively studied animals (*Cepea* land snails) appears to be influenced by a number of factors, including predation (Cain and Sheppard 1954 cited in Gillespie and Oxford 1998) climate (Jones *et al.* 1977) and neutral processes, such as sampling drift (Wright 1978, cited in Gillespie and Oxford 1998) or founder / bottleneck effects (Cameron and Dillon 1984).

3.2.2. South African intertidal environment

Within South Africa, temperature variation and the presence of distinct biogeographic provinces provides an ideal arena for investigating the effects of environmental conditions on phenotypic and phylogeographic structure. The intertidal environment in South Africa has an extreme temperature gradient

between the east and west coasts, mediated by two powerful ocean currents. On the east coast, the fast warm Agulhas current (16-28°C), established in the mid-Pleistocene, flows southwards, and a significant but weaker, colder, surface current, flows northerly inshore up the east coast (Jackson 1976). This influences the intertidal biota enabling temperate biota to spread further eastwards (Macnae 1962; Jackson 1976; Branch and Branch 1981). On the west coast, the slow, cold (8-17°C) northerly, Benguela current, formed around the Neocene / Holocene transition (Shannon 1985; Bolton and Anderson 1997; Marlow *et al.* 2000), is dominated by strong wind driven upwelling which brings cold nutrient rich waters to the surface inshore (Peschak 2006) and can cause the sea surface temperature to fluctuate greatly (Brown and Jarman 1978; Branch and Branch 1981). The south coast has a shallow triangular extension of the continental shelf called the Agulhas Bank (Hutchings 1994). The eastern edge of the bank has upwelling events and interjections of warmer water from the Agulhas current and the western edge is considered part of the Benguela system (Hutchings 1994). In the centre of the Agulhas Bank, cyclonic circulation prevents offshore loss of plankton into the Agulhas current itself. The most recent glacial period 20,000 years ago, caused a drop in sea level exposing much of the Agulhas Bank and extending the coastline up to 150 km out to sea. During inter-glacial periods, the sea level may have risen to 50 m above present, causing the Cape Peninsula to become an island (Branch and Branch 1981).

3.2.3. South African biogeography and phylogeography

Based on studies from diverse taxonomic groups, extant intertidal biota in South Africa is divided into three biogeographic regions (Stephenson and Stephenson 1972; Day 1974; Brown and Jarman 1978; Emanuel *et al.* 1992; Stegenga and Bolton 1992; Bustamante *et al.* 1996; Bolton and Anderson 1997; Turpie *et al.* 2000), the boundaries of which are disputed depending on the taxa considered (Brown and Jarman 1978; Hockey *et al.* 1983; Bolton 1986; Stegenga and Bolton 1992; Turpie *et al.* 2000; Harrison 2002; Sink *et al.* 2004). However, most authors agree that a cool temperate west coast region stretches from Namibia to

a region between Kommetjie and Cape Agulhas (Turpie *et al.* 2000); a warm temperate south coast region stretches from Cape Point to between Port Alfred and Durban (Turpie *et al.* 2000; Peschak 2006) and the sub-tropical east coast extends from around Port St. Johns to Mozambique (Emanuel *et al.* 1992). Subtidally, to a depth < 30 m, the temperature around the coast from Namibia to Port Elizabeth is fairly uniform throughout the year (12-14°C), which can mask and shift the biogeographic boundaries depending on the depth of an organism's distribution (Brown and Jarman 1978; Turpie *et al.* 2000).

To our knowledge only 12 studies have assessed the genetic structure of intertidal marine organisms in South Africa, none of which investigated continuously distributed species occurring around the entire coastline, and therefore previous studies could make only isolated phylogeographic predictions. Of these studies most only investigated phenotypic structure, but one (Ridgway *et al.* 1998 studying *Patella granularis,* the patellid limpet) of these concurrently examined the species ecology. Seven of these studies used allozymes (Lombard and Grant 1986 using *Choromytilus meridionalis* (black mussel); Grant *et al.* 1992 using *Perna perna* (brown mussel); Grant and daSilva-Tatley 1997 using *Bullia digitalis* (gastropod, whelk); Ridgway *et al.* 1998; Ridgway *et al.* 2000 using *Patella miniata* (gastropod); Laudien *et al.* 2003 using *Donax serra* (surf clam)) and five collected mtDNA sequence data (Teske *et al.* 2003 using Hippocampus capensis (Knysna seahorse); Evans *et al.* 2004 using *Haliotis midae* (abalone); Oosthuizen *et al.* 2004 using *Octopus vulgaris* (octopus); Tolley *et al.* 2005 using *Palinurus gilchristi* (spiny lobster); Teske *et al.* 2006 using *Upogebia Africana* (mudprawn), Exosphaeroma hylecoetes (isopod), *Iphinoe truncate* (cumacean)). For four species, phenotypic differences were identified between biogeographic provinces (Ridgway *et al.* 1998; Ridgway *et al.* 2000; Laudien *et al.* 2003, Tolley *et al.* 2005), with only one indicating phenotypic divergence either side of the Cape Peninsula (Ridgway *et al.* 2000). Only Ridgway *et al.* (2000) found genetic differences corresponding to phenotype, whereas the other studies concluded that phenotypic plasticity resulting from different selection pressures in different

biogeographic provinces was responsible for phenotypic differences. Several of these studies found genetically divergent lineages within species (Grant *et al.* 1992; Ridgway *et al.* 1998; Evans *et al.* 2004; Teske *et al.* 2006 within all three species) which may indicate barriers to gene flow. Phylogeographic breaks have been identified near Port St. Johns (Coffee Bay), on the east coast (Ridgway *et al.* 1998; Ridgway *et al.* 2000); around False Bay and the Cape Peninsula (Grant *et al.* 1992; Teske *et al.* 2006) and around Cape Agulhas (Evans *et al.* 2004; Teske *et al.* 2006). Teske *et al.* (2006) concluded that phylogeographic breaks in estuarine crustaceans roughly coincide with the biogeographic boundaries. Mitochondrial DNA provides an ideal tool for establishing phylogeographic breaks (Reeb and Avise 1990; Waters and Roy 2003).

3.2.4. *Parvulastra exigua,* the study organism

In South Africa, the small seastar, *Parvulastra exigua* provides a potentially interesting model to study how phylogeographic and phenotypic structure is affected by environmental conditions. In South Africa *P. exigua* is continuously distributed over approximately 2500 km of coastline from Mozambique to Namibia (Branch and Griffiths 1994). It reportedly occupies the mid to low tidal zones (although appears sub-tidal in some areas; C. Griffiths pers. obs.) in both South Africa and Australia (Byrne 1992; Stevenson 1992). It can reach densities of 150 m^2 (Branch and Branch 1980) and is found in a wide range of microhabitats including exposed rocky outcrops, sheltered still pools and cryptic habitats e.g. crevices or under boulders (Arrontes and Underwood 1991). Although the exact diet of *P. exigua* is unknown (Arrontes and Underwood 1991), it is reportedly a scavenging omnivore (Branch and Branch 1980) with the ability to exploit a variety of resources (such as macro algae), but shows a preference for bare rock surfaces (Arrontes and Underwood 1991). Its distribution may be affected by other grazing species (Branch and Branch 1980; Stevenson 1992), but it has no known predators in South Africa (Griffiths unpublished).

Previous work has identified phenotypic divergence in *P. exigua*, with an unmottled morph occurring in the cold biogeographic province west of the Cape Peninsula and a mottled morph occurring in the warm and subtropical provinces east of the Cape Peninsula, with a gradient of intermediate morphs occurring between these (see Chapter 2). The evolutionary significance of the colour morph differences, and the extent and distribution of this phenotypic divergence in South Africa is unknown. Moreover, colour morph frequencies and associations between habitat heterogeneity, biogeography and neutral genetic variation have not previously been investigated. MtDNA sequencing (Chapter 2) revealed that *P. exigua* phenotypes are not reciprocally monophyletic, indicating that they are not different species, and there is no apparent geographical structuring of *P. exigua* haplotypes in a 400 km stretch of coastline from west of the Cape Peninsula to east of False Bay. This suggests that there is no impediment to gene flow around Cape Point. However, *P. exigua* samples from Port St. Johns on the east coast are genetically divergent from the Cape Peninsula samples (Chapter 2) indicating that *P. exigua* contains unsampled genetic structure somewhere on the south or east coasts.

3.3. Aims

Ecological and phenotypic surveys of *P. exigua* populations across approximately 2500 km of South African coastline and neutral genetic analyses were conducted to:

1. Compare the colour morph frequency distributions and neutral phylogeographic structure of *P. exigua* (Hypothesis: sea star phenotype does not correspond to neutral phylogeographic structure).
2. Examine the neutral phylogeographic structure within *P. exigua* and compare this to biogeographic provinces and current systems (Hypothesis: Biogeographic divides represent barriers to gene flow in *P. exigua*.)

3. Determine whether colour morph frequency distributions are related to ecological differences between habitat types. Hypotheses: a) some or all of the measured environmental variables are associated with changes in the abundance of each sea star phenotype; b) the predicted effects of the environmental variables will differ between the phenotypes.

The findings from this study are discussed in the context of general predictions about intertidal phylogeography in South Africa.

3.4. Materials and Methods

3.4.1. Ecological survey and sampling

This study was conducted in South Africa between February and April 2005. The sampling locations (Fig. 3.1) were approximately equally spaced (approx 200 km between sampling locations) along each coast covering the entire coastal temperature gradient, with a more intensive sampling regime around the Cape Peninsula and False Bay where the phenotypic overlap occurs (Fig. 3.1). At each of the 19 sampling locations, suitable intertidal rocky shore habitat for *Parvulastra exigua* was selected for habitat survey and sampling. The sampling location Kommetjie was included in the ecological survey and samples from here were collected for genetic analysis. However, the divergent lineage identified in Chapter 2 from Kommetjie (PeK) was excluded from both the ecological survey and genetic analysis, and only the *P. exigua* lineage was included. Sampling was conducted during low spring tides. At each sampling location the rocky shore was vertically divided into low, mid and high tidal zones. Within each tidal zone, 10 tide pools were identified as suitable habitats for *P. exigua*, and sampled using a 32 x 32 cm quadrat. Tidal pools deeper than 40 cm were excluded as observations at this depth were inaccurate.

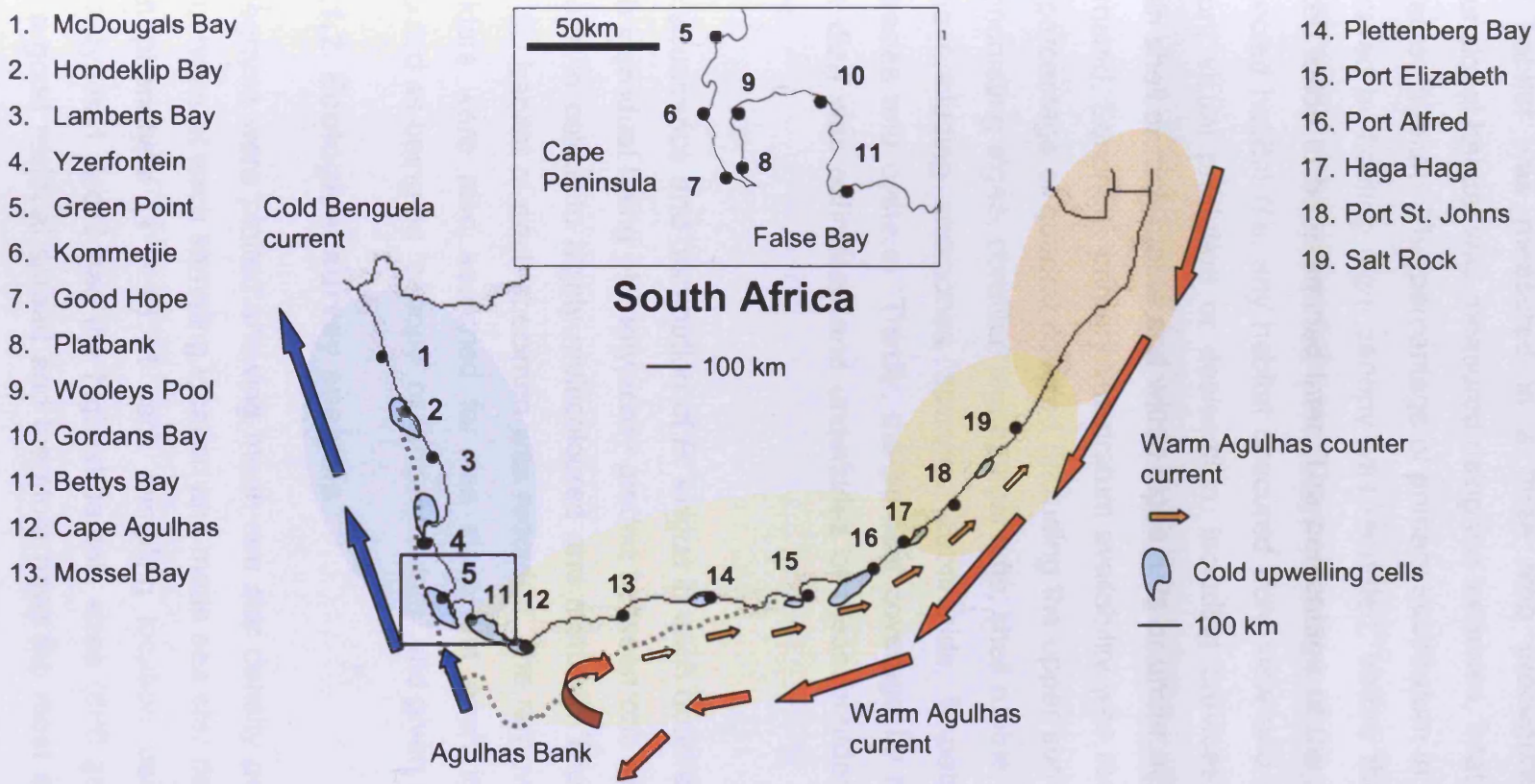

Figure 3.1. Map of South Africa showing the sampling locations, biogeographic provinces, major currents and upwelling cells.

3.4.1.1. Ecological survey

The habitat was measured in a three step procedure. Firstly, the three dimensional habitat was measured using the variables, floating algal canopy and protected habitat. The percentage of primary substratum in the quadrat that was obscured by floating algal canopy was recorded, before this was removed and placed aside to be examined later. The percentage of the quadrat taken up by protected habitat (*i.e.* any habitat obscured from view and protected from wave action, visual predation or desiccation, including crevices, sides of boulders, within shell or rock rubble and within algae tufts or under algal canopy) was also estimated. Secondly, primary substratum availability was recorded by estimating the percentage of quadrat covered (including the upper surface of any boulders) by encrusting algae, coralline algae, algal tufts, shell rubble, sand, bare rock and fauna (including anemones, sponges, zooanthids, limpets, whelks, mussels, barnacles and oysters). Thirdly, the percent coverage by movable boulders (> 7cm dia.) was estimated and undersides of these boulders examined for sea stars.

The abundance and distribution of *P. exigua* in each quadrat was recorded, with each individual being phenotypically graded between one and five (ranging from uniform in colour to highly multicoloured and mottled). The primary substratum and 3D habitat of each specimen was recorded. The removed algal canopy and boulders were also searched for sea stars and each individual found was recorded as being on "canopy" or "under boulder" and given a phenotypic grade.

3.4.1.2. Ecological survey analysis

Histograms were plotted showing mean sea star density per quadrat across all phenotypes at each sampling location and mean sea star density per quadrat for each phenotype (1 – 5) at each sampling location using Microsoft Excel. Phenotypes 1 and 5 had the highest sample sizes (886 and 946, respectively) and a good residual spread and therefore have the most statistical power. The

results for the other phenotypes, particularly phenotype 4, may be less reliable due to smaller samples sizes (of 456, 361 and 177 for phenotypes 2, 3 and 4, respectively).

Sea star abundance was analysed using Generalized Linear Mixed Modelling (GLMM). The data was divided into five data sets corresponding to sea star phenotypes 1 – 5, which represented the dependant variable in the models (Fig 3.2). The dataset for each phenotype included all sampling sites where more than two individuals of that phenotype were present. However, sites with zero or less than 2 specimens of a particular phenotype were still included if adjacent sites contained two or more sea stars of that phenotype. This method of defining a phenotypic dataset removed zero counts which were simply a result of the phenotype being outside its natural geographical range as opposed to being a zero count for a valid biological reason. Each dataset also contained all the ecological variables: coast, tidal zone (low, mid and high), canopy, under boulder, protected habitat, encrusting algae, coralline algae, sessile fauna, mobile fauna, shell rubble, sand, bare rock. Therefore, while certain sites were excluded from an individual phenotype's analysis, all sites contained each of the measured environmental variables and so a comparison, of the effects of each environmental variable upon sea star abundance, was still possible between phenotypes.

	Shaded sampling sites that were included in each phenotype dependant variable dataset.					Coastal zone
	Pheno1	Pheno2	Pheno3	Pheno4	Pheno5	
McDoug	■	■	■			cold west
Hondek	■	■	■			
Lam	■	■	■			
Yzer	■	■	■			
Green	■	■	■			
Komm	■	■	■			
Good	■	■	■			
Plat	■	■	■	■	■	False Bay
Wool	■	■	■	■	■	
Gordans		■	■	■	■	
Bettys			■	■	■	cold west
Agulhas			■	■	■	
Mossel			■	■	■	temperate south
Plett			■	■	■	
Eliz			■	■	■	
Alfred			■	■	■	
Haga			■	■	■	sub-tropical east
Johns				■	■	
Salt					■	

Figure 3.2. Sites included in each phenotype dataset (shaded grey) with the corresponding coastal zones in which they occur in South Africa.

All analyses were undertaken using ASReml2, a dedicated mixed modelling statistical package. In all models sample site was included as a random term to control for any variation due to location that could not explained by the measured environmental variables. The maximal model included all habitat types as fixed terms (independent variables) (Table 3.1).Total sea star abundance was initially considered as a dependent variable. However, in order for this analysis to be meaningful, given the changing environment and varying phenotypic distribution at each sampling site, phenotypic proportions were included as independent variables. These preliminary models had very poor residual distributions showing heteroskedacity, which could not be overcome by transformation. Sea star abundance was then considered by phenotype, with each of the five phenotypes used as the dependent variable, resulting in five final models, one for each phenotype. Initially total sea star abundance was included in the phenotype

models as an independent variable, however inclusion of this term produced poorly fitting models with heteroskedastic residuals and so it was removed from the models for subsequent analysis.

Dependant Variables	Random term (independent variable)	Fixed terms (Independent Variables)
Sea star counts for each phenotype (1 – 5).	Sampling site	tidal zone (low, mid and high) canopy under boulder protected habitat encrusting algae coralline algae sessile fauna mobile fauna shell rubble sand bare rock coast

Table 3.1. Terms initially included in the maximal Generalised Linear Mixed Models for each phenotype (1 – 5).

The histograms of the raw data for all the phenotypes (dependant variables) and all the independent variables showed an overdispersed (aggregated) distribution. A natural log link function was used in the model to transform the dependent variables. All independent variables were normalised by base 10 log transformation (log x + 1) prior to analysis. Non-significant terms were removed from the model in a stepwise manner until a minimal model was reached. The estimates for each fixed and random term were used to plot graphical predictions (Appendix 3.9) of the effects of each of the environmental variables upon the abundance of each phenotype. All graphical predictions from the models were presented as back transformed values of the dependent variable and the proportional effect of each habitat type on the predicted effect of that habitat type between the phenotypes was estimated by examining the comparative changes in predicted effect over the same range of values. Interactions between the independent variables were not calculated because of the large number of

habitats recorded resulting in too many potential interactions between these variables for all of them to be considered. However, the current analyses should highlight those habitat features which have a dominant effect upon the sea star phenotype numbers.

3.4.2. Molecular methods

3.4.2.1. DNA extraction, Cytochrome Oxidase I (COI) amplification and sequencing

DNA extractions were performed using either a Qiagen DNeasy tissue purification kit (Qiagen) with RNAase A treatment according to the manufacturer's instructions, or by phenol-chloroform and CTAB purification (Arndt *et al.* 1996) where the DNA was re-suspended in 100 µl TE buffer and treated with RNAase A (20µg/µl) at 55°C for 1 hour. The samples were subsequently stored at -20°C. Mitochondrial cytochrome oxidase I (COI) primers (P.ex.COII.29F (5' CCA AAC ACA AGG ACA TAG GAA 3') and P.ex.COI.575B (GCG GTA ACG AAT ACG GAT CA) were designed from a COI sequence from an Australian specimen of *P. exigua* (Accession number U50053: Hart *et al.* 1997) using PRIMER 3 (Rozen and Skaletsky 2000) and the free web based software: OLIGONUCLEOTIDE PROPERTIES CALCULATOR: (http://www.basic.northwestern.edu/biotools/oligocalc.html).

All PCRs were performed in a PE 9700 thermal cycler. COI amplification was performed using the following program: 94°C for 5 minutes; followed by 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 minutes for 35 cycles with a final extension of 72°C for 10 minutes. The reaction conditions used were 1x PCR buffer, 1.5 mM $MgCl_2$; 0.2 mM dNTPs; 1 µM of each primer; 0.1U Taq (Invitrogen); and 1 µl DNA template (diluted between 1/10 and 1/1000) in a final volume of 15 µl.

PCR products were purified using a BIO 101 Geneclean Turbo PCR Kit (Q-BIOgene) or ExoSap (Amersham Biosciences) according to the manufacturer's

instructions. Sequencing PCR reactions were performed in 7.5 μl reactions. The PCR program involved an initial denaturation at 96°C for 1 minute 30 seconds for 1 cycle; denaturation at 96°C for 10 seconds; annealing at 50°C for 5 seconds; extension at 60°C for 4 minutes for 25 cycles. Both forward primer and reverse primer reactions were performed. The reaction conditions were as follows: PCR H_2O and template combined 3 μl; Better Buffer (Web Scientific Ltd) 2.5 μl; Big Dye 0.5 μl (ABI PRISM® Big Dye TM Terminator dye vs. 3.1); either forward or reverse primer (1.6 μM) 1.5 μl. An ABI Prism 3100 semi-automated genetic analyser (Applied Biosystems) was used for the sequencing according to the manufacturer's instructions.

3.4.2.2. Genetic structure analyses

The COI forward and reverse sequences were aligned in SEQUENCHER vs. 4.12 (GeneCode Corp.) and verified by eye. The amino acid reading frame was identified by aligning the full COI gene (Accession no. U500053, Hart *et al.* 1997) with the haplotypes identified and checking the echinoderm mitochondrial genetic code for stop codons and amino acid changes. To visualize the relationship between the haplotypes and the sample sizes in each haplotype a median joining network (Bandelt *et al.* 1999) was constructed in NETWORK vs. 4.1.1.2. (www.fluxus-engineering.com). Pairwise divergences between populations (sampling locations) and within and between haplogroups (east and west) were calculated in MEGA 3 (Kumar *et al.* 2004) using the Tamura Nei model of evolution (Tamura and Nei 1993). The Tamura-Nei model corrects for multiple hits, taking into account the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies. It distinguishes between transitional substitution rates between purines and transversional substitution rates between pyrimidines. It also assumes equality of substitution rates among sites (Kumar *et al.* 2004).

Divergence time between haplogroups was estimated based on molecular calibration using a 'clock-like' evolution approach. However, this method of

estimating divergence times relies on a constant rate of evolution for both the gene and the lineage and therefore must be treated with caution (Waters and Roy 2004; Heads 2005). As no definitive geological events are known that can provide a calibration for *P. exigua*, published Kimura two-parameter (K2P; Kimura 1980) calibrations for a variety of echinoderm COI sequences of 3.1 - 3.5 % / million years ago (MYA) were applied (Lessios *et al.* 1999; McCartney *et al.* 2003; Waters and Roy 2004). The % sequence divergence estimate was re-calculated in MEGA 3 (Kumar *et al.* 2004) using the K2P model to compare with the Tamura Nei to estimate the level of accuracy when using a K2P calibration estimate for the time since divergence. The Kimura's two parameter model corrects for multiple hits, taking into account transitional and transversional substitution rates, while assuming that the four nucleotide frequencies are the same and that rates of substitution do not vary among sites (Kumar *et al.* 2004).

3.4.2.3. Nucleotide variation, tests for neutrality and population expansion

Nucleotide diversity, haplotype diversity and mean number of pairwise differences were calculated in ARLEQUIN vs. 3. (Excoffier *et al.* 2005: http://cmpg.unibe.ch/software/arlequin3). Comparing haplotype diversity and nucleotide diversity can reveal information about patterns of historical demography. High haplotype diversity in conjunction with low nucleotide diversity can suggest recent population growth while high haplotype diversity with high nucleotide diversity is indicative of a stable population (Mila *et al.* 2000).

Historical demography was inferred by testing for population expansion and neutrality. Mismatch distributions calculated for the sudden expansion model in ARLEQUIN vs. 2 were plotted using EXCEL. However, one of the assumptions of the mismatch distribution is that there must be random mating within the population being tested. Therefore, to establish if there was any population structure, which reflects the any degree of random mating, and Analysis of Molecular Variance (AMOVA) was conducted both using both a distance based analysis and ΦST, as

well as a frequency based analysis and Fst. The P values were examined and if significant population structure was observed, the assumption of random mating required by the mismatch distribution was broken. Therefore a mismatch distribution analysis was not performed.

If a population has undergone rapid expansion, a unimodal mismatch distribution approximating a Poisson curve is expected, and the P value is expected to be not significantly different from the model of sudden expansion (i.e. P>0.05) (Rogers and Harpending 1992), whereas populations approaching mutation drift equilibrium are expected to produce a multimodal or 'ragged' mismatch distribution and a significantly different P value. The raggedness statistic quantifies the smoothness of the observed pairwise differences distribution. However, as this statistic has low power for detecting population expansion, more powerful combined statistics, namely Tajima's D (Tajima 1989), Fu's Fs (Fu 1997) and Fu and Li's F* and D* (Fu and Li 1997), were conducted.

Tajima's D and Fu's Fs (10000 simulated samples) were calculated in ARLEQUIN and Fu and Li's F* and D* statistics were calculated in DNASP vs. 4.0 (Rozaz *et al.* 2003). A negative Tajima's D statistic and significance value indicates either population expansion or background selection. P is the probability of obtaining the observed D value under the neutral mutation hypothesis (P< 0.05 = the data differs significantly from neutral mutation, P> 0.05 the data does not differ significantly from zero).

The effects of background selection can be distinguished from population growth or range expansion by examining the pattern of significance between Fu's Fs and Fu and Li's F*, and D*. Fu's Fs can be used to test specifically for population growth while Fu and Li's F* and D* are sensitive to background selection and test the hypothesis that all mutations are selectively neutral (Kimura 1983). Expansion is indicated if Fu's Fs is significant but Fu and Li's F* and D* are not,

whereas background selection is indicated if Fu and Li's F* and D* are significant and Fu's Fs is not (Russell *et al.* 2005)

An approximate estimation of time since expansion was calculated using $\tau = 2ut$ (Rogers and Harpending 1992), where τ (tau) is the mode of the mismatch distribution, i.e. the estimation of the age of expansion (obtained from ARLEQUIN vs. 3), u = the mutation rate per sequence per generation and t = time in generations since expansion. The value of t was calculated using the equation $t = \tau / 2u$, and the value of u was calculated using the equation $u = 2\mu k$, where 2μ = the nucleotide divergence rate (twice the mutation rate per nucleotide) per million years (3.1 - 3.5%, Lessios *et al.* 1999; McCartney *et al.* 2003; Waters and Roy 2004) and k = sequence length. Time in generations since expansion (t) was then multiplied by the generation time (estimated at four years, based on the generation time of *Asterina gibossa*, another asterinid of similar size and with a similar life history, Emson and Crump 1978, cited in Emson and Crump 1979) which gave an estimated time in years since expansion.

3.5. Results

3.5.1. Ecology

Sea star mean density across all phenotypes was generally high on the upper west coast but declines southwards (Figure 3.3). However, at the tip of the Cape Peninsula at Good Hope, the density increases dramatically again. False Bay (Platbank, Wooleys Pool and Gordans Bay) has high densities, which decline eastwards along the south coast (with the exception of high numbers at Cape Agulhas). The east coast (from Port Alfred to Salt Rock) has the lowest numbers of sea stars.

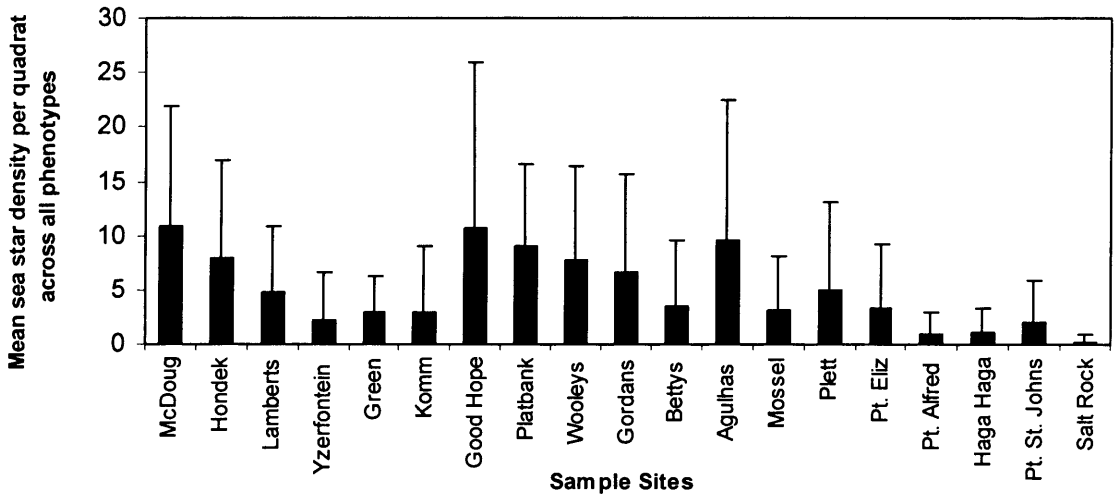

Figure 3.3. Mean sea star density per quadrat across all phenotypes at each sampling location (from west to east). Error bars show standard deviations from the mean.

Figure 3.4, showing mean sea star density / quadrat (32 cm^2) for each phenotype at each sampling location, reveals a clear gradient of phenotypic density around the coast. Unmottled phenotype 1 is present only on the west coast with lowest densities in False Bay. Phenotype 2 is present sporadically along the entire coast, but concentrated on the west coast and in False Bay. Intermediate phenotype 3 has a more consistent distribution around the coast, but is found in high densities in False Bay and the upper west coast only. Phenotype 4 is found

only in False Bay and on the south and east coasts, with high densities in False Bay. Mottled phenotype 5 is again found only in False Bay and on the south and east coasts, but its densities peak around Cape Agulhas, the central sampling location on the south coast.

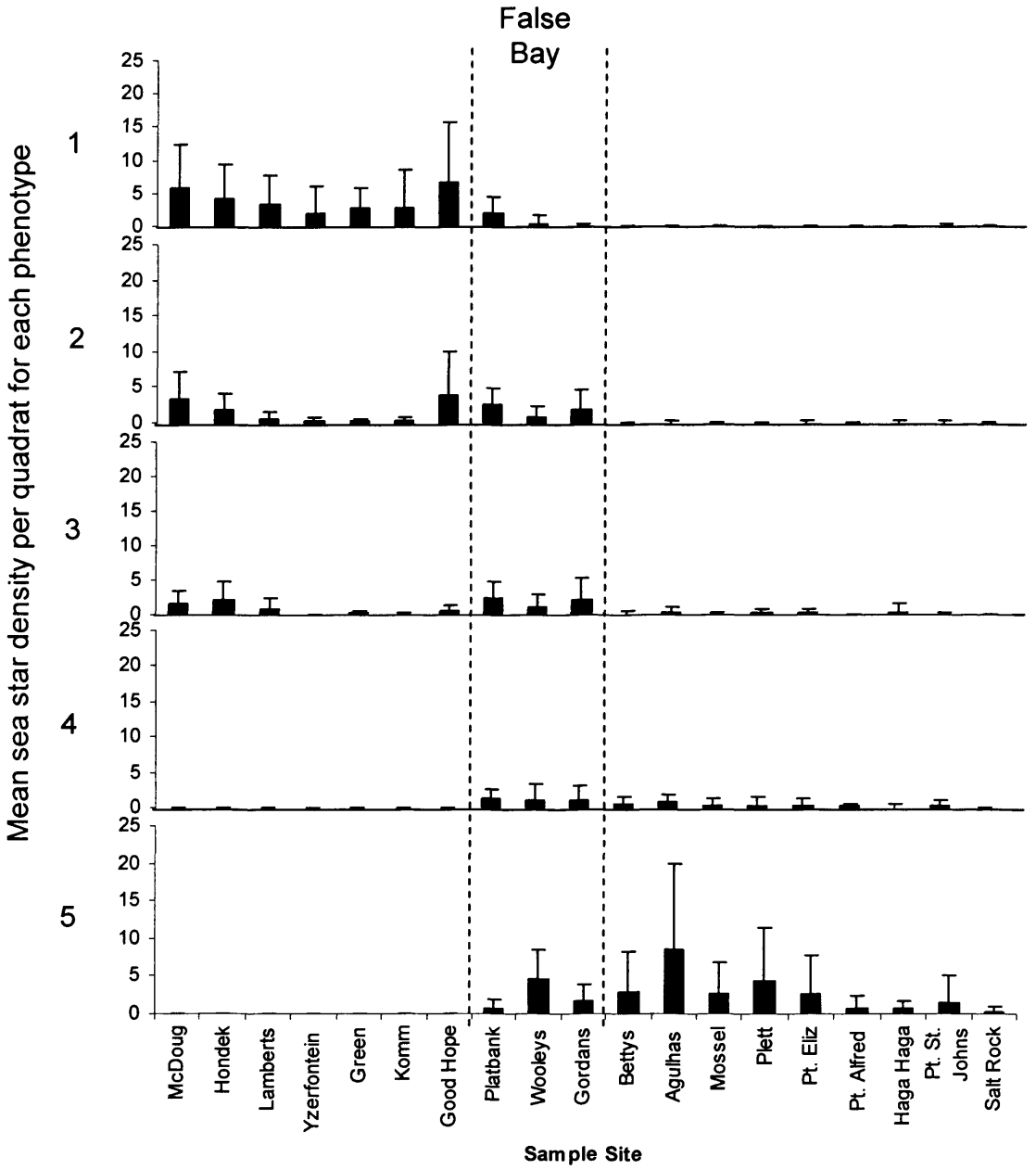

Figure 3.4. Mean sea star density per quadrat (32 cm^2) for each phenotype at each sampling location (from west to east). Error bars show standard deviations from the mean.

3.5.1.2. General Linear Mixed Models

The general linear mixed model (GLMM) revealed that 'coastal zone' (which corresponded to biogeographic provinces Fig. 3.2), and the habitat types 'sessile fauna' and 'shell rubble' had no effect upon sea star abundance or distribution and these terms were therefore not included in any of the minimum models. All other habitat variables had a positive or negative significant effect on at least one of the phenotypes (comparative summary between phenotypes shown in Table 3.2, with the proportional strength of the effects displayed in the plots of the predicted effects (Figs 3.9 – 3.14 Appendix 3.9; 3.). The variables included in the minimal models for each phenotype are shown in Table 3.3.

	Phenotype models				
Habitat type	**1**	**2**	**3**	**4**	**5**
Tide	Mid = most Low = intermediate High = least	As for Phenotype 1	Mid = most Low and High at similar levels	High = most Mid = intermediate Low = least	Mid =most but not significantly higher than High so similar to Pheno4 Low = least
Under Boulder	+++	+	NIIM	++	+
Protected	NIIM	NIIM	NIIM	NIIM	+
Sand	- -	- - -	-	- - - -	- - - - -
Bare rock	+++++	++	+	NIIM	+
Canopy cover	-	- - -	- -	- - - -	NIIM
Algal tufts	NIIM	NIIM	- -	- - -	-
Coralline algae	- -	- -	-	NIIM	NIIM
Encrusting algae	NIIM	NIIM	NIIM	NIIM	+
Mobile fauna	NIIM	NIIM	-	NIIM	NIIM

Table 3.2. General Linear Mixed Model summary table showing comparative predicted effects of each habitat type on each phenotype. Dashes and crosses indicate the comparative strength of the negative and positive effects, respectively. NIIM indicates habitat variables that had no effect on the phenotype and therefore were not included in the model.

The models indicate statistically significant differences in the response of mottled and unmottled phenotypes to tidal zone. The unmottled phenotypes (1 and 2) had higher abundances in the mid zone, with intermediate abundances in the low zone and the lowest abundances in the high zone. Conversely, the mottled

phenotypes (4 and 5) had the highest abundances in the high zone and the lowest abundances in the low zone, with the mid zone being intermediate. Phenotype 3 was more similar to the unmottled phenotypes, with the highest abundances in the mid zone and the low and high zones at similar levels.

As well as tide zone differences, the ecological survey revealed some habitat differences between the phenotypes. The presence of boulders unsurprisingly had a positive effect on all phenotypes (except phenotype 3 where small sample sizes may have caused an unrepresentative significance value) with the largest proportional effect on phenotype 1. Under boulder had a weak proportional positive effect on phenotypes 2 (but again this may be an artefact of small sample size) and 5. This habitat type also had a quite large proportional positive effect on phenotype 4, however, again this may be affected by small sample sizes. The 'protected' habitat type was, surprisingly, only significant for phenotype 5, where it had a reasonably strong positive effect.

The habitat type sand had a significant negative effect on all phenotypes, with the largest proportional negative effect on phenotype 5, followed by phenotypes 4, 2 and 1, with a very small proportional negative (but very large confidence intervals) effect on phenotype 3 (Fig 3.9 Appendix 3.9;3). Bare rock had a significant positive effect on all phenotypes except phenotype 4. Bare rock had by far the largest proportional positive effect on phenotype 1, the unmottled phenotype, and had intermediate proportional positive effect on phenotype 2. It had a weak proportional positive effect on phenotypes 3 and 5.

Canopy had a significant negative effect on the unmottled phenotypes 1-4, but no effect on phenotype 5, with the largest proportional negative effect on phenotype 4, followed in order by phenotypes 2, 3 and 1 (Fig 3.10 Appendix 3.9;3). However, algal tufts also had a negative effect but on the mottled phenotypes instead, with the largest proportional negative effect on phenotype 4, followed by 3 and then 5 (Figs 3.11 Appendix 3.9;3), and no effect on phenotypes 1 and 2.

Coralline algae had a negative effect on the unmottled phenotypes 1-3 (Fig 3.12 Appendix 3.9;3), with proportionally similar negative effects on phenotypes 1 and 2 compared to on phenotype 3 where it had a slightly smaller proportional negative effect, but no effect on phenotypes 4 and 5. Conversely, encrusting algae only had a significant (highly positive) effect on phenotype 5.

The presence of mobile fauna had no effect on any phenotypes except phenotype 3, where it had a weak negative effect.

Significant Terms	d.f.	F value	P value	Estimate	Estimate standard error
Phenotype 1					
Mu	1	38.80	<0.001	0.03423	0.27075
Tide	2	12.75	<0.001	Low 0.29659 Mid 0.58950 High 0.0000	Low 0.15685 Mid 0.13101 High 0.0000
Canopy	1	8.96	0.003	-0.05802	0.08170
Under boulder	1	31.44	<0.001	0.43637	0.09910
Coralline algae	1	7.54	0.007	-0.51730	0.19357
Sand	1	4.49	0.036	-0.09519	0.09570
Bare Rock	1	15.47	<0.001	0.46133	0.11730
Phenotype 2					
Mu	1	24.60	<0.001	0.28630	0.19547
Tide	2	16.73	<0.001	Low 0.09650 Mid 0.34770 High 0.0000	Low 0.10725 Mid 0.09021 High 0.0000
Canopy	1	23.15	<0.001	-0.16392	0.05735
Under boulder	1	5.96	0.016	0.09689	0.06548
Coralline algae	1	6.79	0.010	-0.34623	0.13457
Sand	1	6.12	0.015	-0.10440	0.06639
Bare Rock	1	8.45	0.004	0.24151	0.08308
Phenotype 3					
Mu	1	17.98	<0.001	0.43410	0.1222
Tide	2	13.00	<0.001	Low 0.05410 Mid 0.16780 High 0.0000	Low 0.06435 Mid 0.05539 High 0.0000
Canopy	1	8.94	0.003	-0.05701	0.03736
Coralline algae	1	13.14	<0.001	-0.15650	0.05090
Algal tufts	1	19.99	<0.001	-0.16000	0.04149
Mobile fauna	1	4.95	0.027	-0.18460	0.05744
Sand	1	10.82	0.001	-0.09829	0.03843
Bare Rock	1	5.67	0.018	0.11580	0.04862
Phenotype 4					
Mu	1	15.45	0.003	0.48210	0.0842586
Tide	2	14.70	<0.001	Low -0.19906 Mid - 0.14824 High 0.0000	Low 0.06992 Mid 0.06035 High 0.0000
Canopy	1	5.73	0.018	-0.09205	0.0525446
Under boulder	1	21.54	<0.001	0.15310	0.0430823
Algal tufts	1	12.74	<0.001	-0.14440	0.0477346
Sand	1	8.08	0.005	-0.13155	0.0462783
Phenotype 5					
Mu	1	30.40	<0.001	0.2308	0.25390
Tide	2	11.61	<0.001	Low -0.1536 Mid 0.09551 High 0.000	Low 0.1188 Mid 0.1074 High 0.000
Under boulder	1	29.44	<0.001	0.1547	0.08431
Protected	1	6.15	0.014	0.2070	0.10470
Encrusting algae	1	12.71	<0.001	0.1882	0.09156
Algal tufts	1	15.21	<0.001	-0.2693	0.08114
Sand	1	8.58	0.004	-0.1867	0.08532
Bare Rock	1	6.03	0.015	0.2261	0.09209

Table 3.3. Analysis of Variance table for the minimal generalised linear mixed models for each phenotype of *Parvulastra exigua* showing d.f. = degrees of freedom, F value, P value, estimates for the logged dependent variables and the standard error of the estimate.

3.5.2. Molecular data

A 358 bp sequence of the mitochondrial cytochrome oxidase I (COI) gene was sequenced from 170 individuals, comprising between 5 and 17 individuals from 19 sampling locations. This resulted in 36 haplotypes (Table 3.4) and 39 polymorphic sites, of which 29 were parsimony informative. All polymorphic sites were synonymous substitutions (resulting in no amino acid changes).

Haplotypes	*Parvulastra exigua* sampling locations																			
	Mcdougals Bay	Hondeklip Bay	Lamberts Bay	Yzerfontein	Green Point	Kommetjie	Good Hope	Platbank	Wooleys Pool	Gordans Bay	Bettys Bay	Cap Agulhas	Mossel Bay	Plettenberg Bay	Port Elizabeth	Port Alfred	Haga Haga	Port St. Johns	Salt Rock	**Total in each haplotype**
1	2	4	7	6	6	14	4	9	7	12	2	3								**76**
2									3	5										**8**
3								3	1		1									**5**
4											1									**1**
5									3											**3**
6							2													**2**
7				2																**2**
8											1									**1**
9					1															**1**
10					1															**1**
11						1														**1**
12								1												**1**
13								1												**1**
14								1												**1**
15									1											**1**
16									1											**1**
17									1											**1**
18				1																**1**
19				1																**1**
20																	2	3		**5**
21																		2		**2**
22																		1		**1**
23															2			1		**3**
24																7	5			**12**
25												1	6							**7**
26	5	1																		**6**
27														6						**6**
28																			6	**6**
29															4					**4**
30												2								**2**
31		2																		**2**
32													1							**1**
33												1								**1**
34														1						**1**
35															1					**1**
36																			1	**1**
Total	**7**	**7**	**7**	**10**	**8**	**15**	**6**	**15**	**17**	**17**	**5**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**170**

Table 3.4. Cytochrome oxidase I haplotypes of *Parvulastra exigua* in South Africa. Data comprise: total number of haplotypes identified in each sampling location, number of individuals found with each haplotype in each sampling location and total number of individuals sampled from each sampling location.

3.5.2.1. Genetic structure

The median joining network (Fig. 3.5) indicated two deeply divergent reciprocally monophyletic clades (haplogroups) to the west and east coasts separated by 11 nucleotide substitutions, with complete lineage sorting between them and an average Tamura Nei pairwise divergence of 4.38% (Table 3.5), (4.38% Kimura 2 Parameter). The time since divergence between the east and west haplogroups was estimated at between 1.4 - 1.25 MYA. The west haplogroup contained all individuals from all 13 sites located from Mcdougals Bay to Mossel Bay, a total of 128 individuals within 25 haplotypes. The east haplogroup contained all individuals from all six sites from Plettenberg Bay to Salt Rock. This radical genetic split between these two haplogroups lay somewhere between Mossel Bay and Plettenberg Bay (Fig. 3.5), a distance of only 140 km, but the exact location was not identified, and it is unknown whether there is an area of sympatry. This genetic divide lies approximately 460 - 680 km west of False Bay, the site of the main area of phenotypic sympatry where the unmottled and the mottled phenotypes appear to diverge.

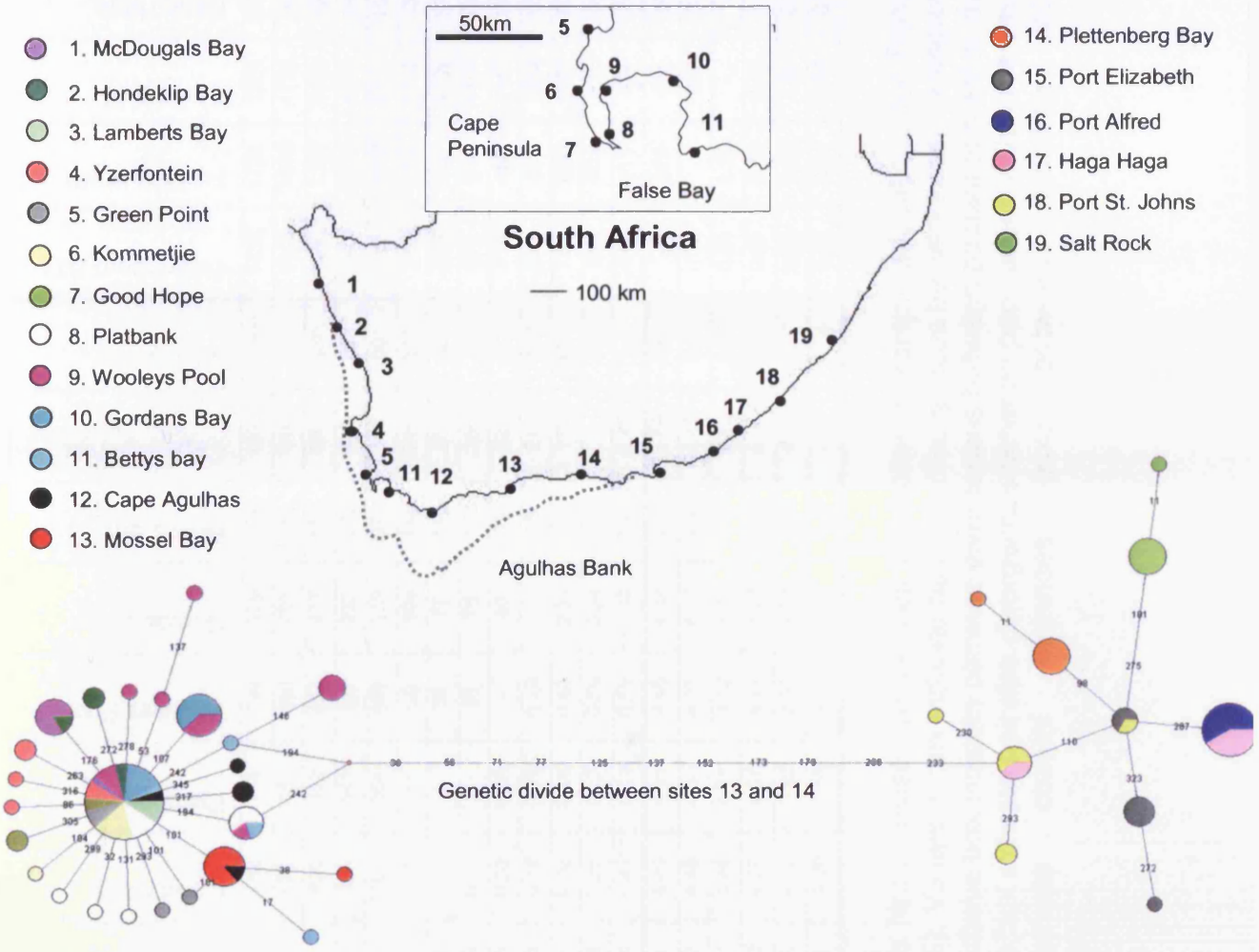

Figure 3.5. Median Joining Network (MJN) for *Parvulastra exigua* COI haplotypes, indicating the haplotype relationships. Circles represent haplotypes, circle area represents number of individuals within that haplotype. Lines with numbers indicate number of nucleotide differences, and position of nucleotide differences in the sequence between the haplotypes, lines without numbers indicate one nucleotide difference between haplotypes. Coloured circles represent proportion of individuals within that haplotype from each sampling location. Each sampling location represented by an individual colour, indicated by coloured circle next to the sample location name. Sampling location names arranged so that the west coast haplogroup sampling locations are on the left, and the east coast haplogroup sampling locations are on the right. This figure indicates the genetic divide is between sampling locations 13 and 14.Inset enlarged map of False Bay area between sampling locations 5 and 11.

	McDougals Bay	Hondeklip Bay	Lamberts Bay	Yzerfontein	Green Point	Kommetjie	Good Hope	Platbank	Wooleys Pool	Gordans Bay	Bettys Bay	Cape Agulhas	Mossel Bay	Plettenberg Bay	Port Elizabeth	Port Alfred	Haga Haga	Port St. Johns	Salt Rock
McDougals Bay	-	122	352	592	694	719	752	754	789	829	864	1039	1291	1432	1702	1861	2034	2238	2551
Hondeklip Bay	0.26	-	229	469	571	597	629	631	666	706	741	916	1168	1310	1580	1739	1912	2116	2429
Lamberts Bay	0.20	0.12	-	240	342	367	400	402	437	477	512	687	939	1080	1350	1509	1682	1886	2199
Yzerfontein	0.31	0.23	0.11	-	102	127	160	162	197	237	272	447	699	840	1110	1269	1442	1646	1959
Green Point	0.27	0.19	0.07	0.18	-	25	262	364	669	339	374	549	801	942	1212	1371	1544	1748	2061
Kommetjie	0.22	0.14	0.02	0.13	0.09	-	32.5	34	69	109	144	319	571	713	983	1142	1315	1519	1832
Good Hope	0.29	0.21	0.09	0.21	0.16	0.11	-	2	35	77	112	287	539	680	950	1109	1282	1486	1799
Platbank	0.31	0.23	0.11	0.22	0.18	0.13	0.21	-	35	75	110	285	537	678	948	1107	1280	1484	1797
Wooleys Pool	0.43	0.35	0.23	0.34	0.30	0.25	0.33	0.34	-	40	72	325	577	718	988	1147	1320	1524	1837
Gordans Bay	0.28	0.20	0.08	0.20	0.15	0.10	0.18	0.19	0.29	-	35	210	462	603	873	1032	1205	1409	1722
Bettys Bay	0.43	0.35	0.22	0.34	0.29	0.24	0.32	0.31	0.43	0.31	-	175	637	778	1048	1207	1380	1584	1897
Cape Agulhas	0.36	0.28	0.16	0.27	0.23	0.18	0.25	0.27	0.39	0.24	0.37	-	252	393	663	822	995	1199	1512
Mossel Bay	0.52	0.44	0.32	0.43	0.36	0.34	0.41	0.43	0.55	0.40	0.43	0.40	-	141	411	570	743	947	1260
Plettenberg Bay	4.61	4.52	4.39	4.51	4.47	4.41	4.49	4.39	4.46	4.48	4.39	4.56	4.64	-	270	429	602	806	1119
Port Elizabeth	4.52	4.40	4.30	4.42	4.37	4.32	4.40	4.30	4.37	4.39	4.30	4.47	4.55	0.56	-	159	332	363	676
Port Alfred	4.56	4.48	4.35	4.47	4.42	4.37	4.45	4.38	4.42	4.44	4.35	4.52	4.60	0.61	0.52	-	173	377	690
Haga Haga	4.40	4.31	4.18	4.30	4.25	4.20	4.28	4.18	4.25	4.27	4.18	4.35	4.43	0.61	0.52	0.16	-	204	517
Port St. Johns	4.14	4.05	3.92	4.04	4.00	3.94	4.02	3.92	3.99	4.01	3.92	4.10	4.17	0.69	0.60	0.65	0.51	-	313
Salt Rock	4.91	4.82	4.69	4.81	4.69	4.71	4.80	4.69	4.76	4.78	4.63	4.82	4.64	0.91	0.84	0.88	0.88	0.97	-

Table 3.5. Below diagonal, average Tamura Nei pairwise divergences (%) between sampling locations for *Parvulastra exigua* COI sequences (calculated in mega 3). Values in box indicate pairwise divergences between east haplogroup and west haplogroup sampling locations, values above box indicate pairwise divergences between populations within the west coast haplogroup and values to the left of the box indicate pairwise divergences between populations within the east coast haplogroup. Above diagonal, approximate coastal distances (km) between sampling locations.

An analysis of molecular variance (AMOVA) was performed to test for population structure between the east and the west coast haplogroups, as well as to test for population structure within the east and west halpogroups separately (Table 3.6).

Source of variation (group partitioning)	Reason for choosing groups	Genetic variation accounted for:		% variation	Fixation Indices (2 d.p.)	P values (3 d.p.)
West populations: [McDougals Bay, Hondeklip Bay, Lamberts Bay, Yzerfontein, Green Point, Kommetjie, Good Hope, Platbank, Wooleys Pool, Gordans Bay, Bettys Bay, Cape Agulhas, Mossel Bay] **East populations**: [Plettenberg Bay, Port Elizabeth, Port Alfred, Haga Haga, Port St. Johns, Salt Rock]	To test for population structure in order to determine if there is random mating between the populations and therefore if the datasets comply with the assumption of random mating necessary for analysis using a mismatch distribution.	Distance based ΦST	AG APWG WP	91.69 3.93 4.39	ΦSC: 0.47 ΦST: 0.96 ΦCT: 0.92	0.000+-0.000 0.000+-0.000 0.000+-0.000
		Frequency based FST	AG APWG WP	23.79 23.80 52.41	FSC: 0.31 FST: 0.48 FCT: 0.24	0.000+-0.000 0.000+-0.000 0.000+-0.000
East populations: [Plettenberg Bay, Port Elizabeth, Port Alfred, Haga Haga, Port St. Johns, Salt Rock]		Distance based ΦST	APWG WP	76.31 23.69	ΦST: 0.76315	0.000+-0.000
		Frequency based FST	APWG WP	55.32 44.68	FST: 0.55323	0.000+-0.000
West populations: [McDougals Bay, Hondeklip Bay, Lamberts Bay, Yzerfontein, Green Point, Kommetjie, Good Hope, Platbank, Wooleys Pool, Gordans Bay, Bettys Bay, Cape Agulhas, Mossel Bay]		Distance based ΦST	APWG WP	24.37 75.63	ΦST: 0.24374	0.000+-0.000
		Frequency based FST	APWG WP	20.77 79.23	FST: 0.20773	0.000+-0.000

Table 3.6. Analysis of Molecular Variance (AMOVA) results for all *Parvulastra exigua* samples indicating the apriori groups tested and the reason for testing them, and showing the percentage of genetic variation accounted for when the data is divided up into different groups. AG: Among Groups; APWG: Among Populations Within Groups; WP: Within Populations

AMOVA indicated a significant genetic structure was detected between the east and the west groups ($P<0.001$), and therefore these were subsequently treated as separate populations. The AMOVA also indicated significant population structure in both the east and west coast haplogroups, suggesting that the assumption of random mating required for the Mismatch distribution (to test for

population expansion) is violated, and that the mismatch distribution should not be performed on either of the groups (Rogers and Harpending 1992). However, Rogers (1995, 1997) explored if, and if so, how much, structure within a population affects the statistical inferences of the mismatch distribution in predicting population expansions. He concluded that population structure tends to make mismatch distributions ragged and so acts against the inference of expansion, but that his results provided no support for the view that population structure reduces the value of mismatch distributions for statistical inference (Rogers 1995, 1997). Therefore the topology of the MJN's and the associated neutrality statistics were examined to see if it was justified to perform a mismatch distribution on either the east or the west group despite the violated assumption of random mating.

The MJN's for both the east and the west haplogroups was star-shaped (central haplotype with divergent haplotypes radiating from it) and both haplogroups also had high haplotype diversity and low nucleotide diversity (Table 3.7), possibly suggesting either a population expansion event or a selective sweep has occurred in both haplogroups.

	West coast haplogroup	**East coast haplogroup**
Sample Size (N)	128	48
No. of haplotypes	25	11
Nucleotide diversity	0.002458 +/- 0.001920	0.005820 +/- 0.003687
Haplotype (gene) diversity	0.6394 +/- 0.0489	0.8653 +/- 0.0301
Mean no. pairwise differences	0.879798 +/- 0.620954	2.083624 +/- 1.188381
Fu's Fs statistic	-30.44523 (P< 0.01)	-3.01766 (P= 0.09)
Fu and Li's F*	-3.66786 (P<0.02)	0.06313 (P>0.10)
Fu and Li's D*	-3.54154 (P<0.02)	0.22413 (P>0.10)
Tajimas D	-2.30904 (P<0.01)	0.30555 (P>0.10)

Table 3.7. Neutrality statistics, and either standard deviations or P values, for *Parvulastra exigua* COI sequences showing the whole dataset and east and west haplogroups.

3.5.2.1.1. East coast haplogroup

The structure of the east coast haplogroup shows a central (ancestral) haplotype which contains far fewer individuals than the west coast haplogroup. This haplogroup also has a star-shape but the haplotypes diverging from the central haplotype are typically more than one base divergent and comprise more samples than the divergent haplotypes in the west coast haplogroup. This shape suggests that a population expansion may have occurred earlier in the east coast haplogroup than the shape of the MJN in the west coast haplogroup suggests. The intra-haplogroup average pairwise divergences estimate is 0.58%. The sampling sites at both ends of the east haplogroup geographic range (Plettenberg Bay and Salt Rock) have no individuals that share the central haplotype. The sampling locations in the middle of the haplogroups geographic distribution share haplotypes, but not always with adjacent sampling sites. These combined results suggest some degree of population structure in the east coast haplogroup.

The Tajima D statistic was positive and did not differ significantly from the neutrality model, indicating neither population expansion or background selection. Fu's Fs was negative but not significant whereas Fu and Li's F* and D* were both positive but not significant. These results suggest that the east coast haplogroup is not expanding and therefore no mismatch distribution was performed on this haplogroup. However, the network structure, high haplotype diversity and low nucleotide diversity may indicate that at some time in the past, this haplogroup did experience a population expansion or mitochondrial selective sweep event.

3.5.2.1.2. West Coast haplogroup

The west coast haplogroup had high haplotype diversity and low nucleotide diversity (Table 3.7), possibly suggesting either a population expansion event or selective sweep event. The star-shaped MJN pattern of the west coast haplogroup shows one dominant (probably ancestral) central haplotype, with several haplotypes containing only one or two individuals diverging from this, suggesting recent or ongoing population expansion. This pattern is reflected in the low average intra-haplogroup pairwise divergence estimate (0.24%) (Table 3.5). The 'divergent' haplotypes occur at most locations. Notably, no individuals from Mossel Bay are represented in the central (ancestral) haplotype circle, but this site does share haplotypes with Cape Agulhas, and has a haplotype from Betty's Bay diverging from it.

The negative Tajima's D test statistic for the west haplogroup was highly significant for deviations from the model of neutrality, indicating either range expansion or background selection. Fu's Fs for the west haplogroup was highly negative and highly significant and Fu and Li's F* and D* were both negative and significant, often indicating background selection. This suggests that the west coast haplogroup is either expanding or has undergone a mitochondrial selective sweep (Rogers 1995). Therefore a mismatch distribution was performed on this haplogroup (Fig. 3.6), but the inferences made were treated with caution because the random mating assumption is violated.

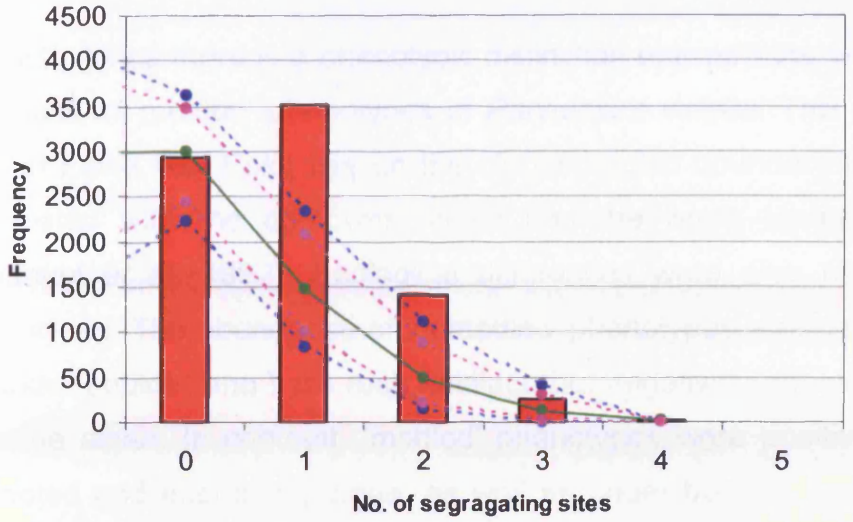

Figure 3.6. Pairwise mismatch distribution plot for the west coast haplogroup for the *Parvulastra exigua* aligned COI sequences; the green line represents the expected mismatch distribution of a population under the sudden expansion model. The red bars represent the observed mismatch distribution for the data. The pink dashed lines represent the confidence intervals at 0.05%, and the blue dashed lines represent the confidence intervals at 0.01%. The mismatch observed mean: 0.88, mismatch observed varience: 0.685, Sum of Squared deviation: 0.005494, P(Sim. Ssd >= Obs. Ssd): 0.0867, Harpending's Raggedness index: 0.093821, P(Sim. Rag. >= Obs. Rag.): 0.0298.

The Mismatch analysis indicated that the west coast population has undergone a sudden expansion (P>0.08) or there has been a mitochondrial selective sweep, as the same shape of the mismatch graph is produced by both processes (Harpending et al. 1998). However, considering the associated statistics and the shape of the MJN, a sudden expansion seems more likely. The mismatch distribution (Fig. 3.6) shows a uni-modal distribution, and the close proximity of the curve to the left axis also suggests a very recent or ongoing expansion, thus supporting the results of the network structure. The time since expansion was estimated at approximately 39,000 years ago (Tau = 0.971 confidence intervals at 0.05%: up-bound tau = 0.705 (28,000 years ago), low bound tau = 1.449 (58,671 years ago). Due to the shape of the mismatch distribution, negative Tajima's D and Fu's Fs statistics, high haplotype and low nucleotide diversity and star shape of the network, population expansion is a more plausible explanation than background selection, although the latter cannot be ruled out.

3.6. Discussion

In South Africa there is a phenotypic distinction between the western 'unmottled' and eastern 'mottled' phenotypes of *Parvulastra exigua*. This phenotypic divide around False Bay Point lies on the biogeographic boundary in fauna and flora associated with the cool west coast and the warm temperate south coast (Emanuel *et al.* 1992). Ecological differences were also noted between the phenotypes. The abundance of 'unmottled' phenotypes was positively influenced by under boulder and bare rock habitats, but negatively affected by canopy and coralline algae. In contrast, 'mottled' phenotypes were positively influenced by protected and encrusting algae, as well as under boulder and bare rocks, and negatively affected by algal tufts. Both the mottled and unmottled phenotypes were negatively affected by sand. Genetic analysis revealed a deep divide between reciprocally monophyletic 'east' and 'west' haplogroups, near the Agulhas Bank on the south coast. However, the genetic divide does not coincide with this phenotypic divide, but instead is situated approximately 500 km to the east.

3.6.1. Phenotypic diversity

If colour variation was the result of neutral demographic processes, it would be expected that the colour morphs would be distributed randomly, or that divergence in colour morphs would coincide with neutral genetic divergence (Hoffmann *et al.* 2006). As neither of these predictions were realized, neutral processes or population structure are unlikely to explain the colour morph frequencies. Instead, colour variation coincided more closely with environmental variables and therefore either phenotypic plasticity or adaptations in response to ecological selective forces are more likely explanations (Growns and Ritz 1994; Hoekstra *et al.* 2004; Brisson *et al.* 2005; reviewed in Gray and McKinnon 2007).

Distinguishing between phenotypic plasticity and natural selection as the causal mechanism for phenotypic variation can be difficult and usually requires breeding (reviewed in Scheiner 2002), physiological (Sokolova and Berger 2000) and/or translocation experiments (Trussell 2000; Johannesson and Ekendahl 2002). In the case of *P. exigua* it is unknown whether individuals change colour if transferred to different environments or if phenotype is inherited. For example, colour may be the result of diet (Hull *et al.* 2001; Jordaens *et al.* 2001), or the association between morphotypes and habitat may be a behavioural response to increased crypsis, indicating that both colour and behaviour are plastic traits, and that selection is acting on behaviour and is not the causative agent responsible for the colour morphs. However, phenotypic plasticity does not account for the distinct phenotypic divide around False Bay as broadly similar habitats exist on the east and west coasts. Therefore although phenotypic plasticity cannot be ruled out, it is more likely that selection is responsible for the colour morph divergence of *P. exigua.*

A mottled morph is also seen in the sister species of *P. exigua*, *P. dyscrita* (H. L. Clark 1923), indicating that phenotype may be under such strong selection pressure that it transcends species, indicating convergent evolution. Polymorphisms that transcend species boundaries are most likely retained from a common ancestor and are unlikely to be selectively neutral (Golding 1992; Oxford 2005). If the mottled phenotype is ancestral, the rapid population expansion seen in the west coast haplogroup of *P. exigua* may indicate that the 'unmottled' phenotypes have colonized the west coast relatively recently. In this case, the unmottled phenotype may have evolved very rapidly in response to strong environmental selection. If selection (as opposed to phenotypic plasticity) is assumed to be the causative mechanism responsible for colour, then habitat differences of the colour morphs might identify the selective forces underlying the colour variation.

3.6.2. Ecology

Phenotypic variation as a selective response to predation is a common cause for colour variation in marine invertebrates (Palma and Steneck 2001; Johannesson and Ekendahl 2002). The unmottled (phenotypes 1 - 3) and the mottled morphs (phenotypes 3 - 5) of *P. exigua* may increase crypsis from different types of visual predators in different habitats.

Tidal zone significantly influenced the abundance of mottled and unmottled sea stars. Phenotypes 1 - 3 were most abundant in the mid tide level, and least abundant in the high zone (phenotype 3 had similar levels in the low and high zones), whereas phenotypes 4 and 5 were more common at the high tide level and least abundant in the low zone. This may indicate that the mottled and unmottled morphs are susceptible to different types of predation, with the unmottled sea stars being more at risk from aerial predators and the mottled sea stars being more at risk from marine predators. This may be a result of the colouring of the morphs, with the reds and light colours of the mottled sea stars being more visible than darker colours of the unmottled morphs underwater. However, this distribution could also be the result of other abiotic factors (such as dissolved oxygen, nitrogen), and other biotic differences (competition) between the tidal zones or the level of susceptible to desiccation of the different morphs.

The abundance of all phenotypes was negatively affected by the presence of sand. This may reflect the fact that all phenotypes are conspicuous against the uniform colour of sand, which possibly results in a high level of predation on this substrate. Alternatively, this correlation could reflect the lack of food sources present in this habitat. The 'encrusting' habitat only had a positive effect on the most mottled phenotype 5. Mottled sea stars may appear more cryptic on heterogeneous backgrounds (e.g. bare rock / encrusting algae mixtures) due to background matching or breaking up the outline of the sea star, (Merilaita 2002;

Rosenblum 2006; Todd *et al.* 2006). The lack of effect between encrusting algae and phenotype 4, the next most mottled phenotype, which should also benefit from crypsis, may be an artefact of a smaller sample size.

Algal tufts had a negative effect on phenotypes 3 - 5 but no effect on phenotypes 1 - 2; conversely, coralline algae had a significant negative effect on phenotypes 1, 2 and 3 and no effect on phenotypes 4 and 5. As Algal tufts are mostly green or brown and coralline algae is mainly pink or white, these effects could be explained by the unmottled sea stars being more cryptic on algal tuft backgrounds and the mottled sea stars being more conspicuous on this background, and vice versa for the coralline algae.

All phenotypes, except phenotype 4, were positively associated with bare rock with by far the strongest effect exhibited by the unmottled sea stars. *Parvulastra* sea stars feed on microalgal film on the surfaces of the bare rock (Branch and Griffiths 1994), but this habitat could also incur a cryptic advantage as the rock surfaces tend to be fairly homogenous in colour (although this was not quantified) providing a uniform background for the unmottled phenotypes. Alternatively, this positive association may reflect lack of competition from other grazers.

With the exception of phenotype 5, algal canopy had a significant negative effect on all phenotypes possibly as the canopy attracted other predators, thereby increasing the risk of predation for the sea stars. The mottled appearance of phenotype 5 may have provided this phenotype with some level of protection from predation due to breaking up the sea star outline, thus causing a lack of negative effect. Alternatively, the algal canopy could be changing the physical environment in the pools by either decreasing the amount of oxygen and nitrogen in the water, or providing protection from desiccation. The unmottled morphs may be more negatively susceptible to the effects of these changes that the mottled morphs.

With the exception of phenotype 3, all other phenotypes showed a weak positive association with under boulder habitats, potentially indicating that cover is important for all sea stars and may a play a role in preventing terrestrial and / or aerial predation (Johannesson and Ekendahl 2002), desiccation (Etter 1988) or even physical damage from wave action. Surprisingly however, the habitat type protection was only positively associated with phenotype 5. This could be due to the effect of protective cover being masked by the stronger effects of 'under boulder'.

Only two of the habitat types recorded, sessile fauna and shell rubble, had no effect on the abundances of any of the phenotypes. Whilst some of the sessile organisms may provide protective cover in the form of crevices for small sea stars (e.g. mussels, oysters or barnacles) this protective effect may be masked by stronger associations between sea stars and other habitat types (such as under boulder and protection). The level of shell rubble may be linked to the degree of wave action or abrasion (although this was not tested), however, again, any influence that this habitat category may have on the sea star numbers in each phenotype may be obscured by other stronger effects.

As highlighted by the ecology survey, the interactions between phenotype and habitat type are complex and there may be many selective forces or secondary selective forces that cannot be identified from these data. These results indicate only general trends and should be treated with caution because the generic terms used in this survey encompass many different species within them (e.g. many different species of algae, and lots of different groups of mobile and sessile fauna, including mussels, gastropods, zooanthids, oysters etc.) and all of these species may have different effects on the phenotypes and their interactions in the environment. However, the potential selective forces acting upon the phenotypes could be influenced by predation and the level of crypsis each habitat type provides. The different colour morphs of *P. exigua* may have evolved in response to different levels and forms of predation in different areas. If some morphs are

more conspicuous in different habitats, this may indicate that they have different levels of chemical toxins and the colour morphs may have evolved as warning colouration, advertising the colour morphs unpalatability to protect them against physical attacks. The red colouration of some adult morphs may indicate chemical defences. Red colouration in other organisms often indicates antifeedant carotinoids (reviewed in Bandaranayake 2005). Some asteroids and holothurians are known to produce saponins as chemical defences in both the larval and adult stages (Lucas et al. 1979; Bingham and Braithwaite 1986; McClintock and Vernon 1990).

Additional to the ecological survey data, two lines of evidence indicate that phenotype in *P. exigua* may also be linked with water temperature, either directly or indirectly. Firstly, there was a pronounced phenotypic overlap zone in False Bay which corresponds to the biogeographic divide between the cold west coast and the cool temperate south coast. Secondly, *P. exigua* in Australia occurs in the cold water provinces but only the 'unmottled' phenotype has been observed, and *P. exigua* does not occur in the warm water provinces. In South Africa, the unmottled phenotype also occurs in the cold water biogeographic province. Additionally, *P. dyscrita* also has a mottled phenotype and is distributed on the south coast where the water is warmer, but is absent from the cold west coast. This distributional evidence may indicate a selective advantage for the 'mottled' phenotype in warmer waters (Etter 1998). Colour in other asterinids may be affected by light intensity or water temperature. For example, Strong (1975 cited in Williams 1999) showed that brown morphs of the starfish genus *Linckia* occur at greater depths than the royal-blue morph and are less tolerant of higher water temperatures. On the other hand, royal-blue morphs are more common in very shallow waters, where water temperatures and exposure to direct sunlight are beyond the tolerance limits of many species. Thermal tolerance in other species has been suggested as a driving selective force in maintaining colour polymorphisms (Forsman 1999). The role of selection in maintaining colour polymorphisms in *P. exigua* remains unclear, and will almost certainly be the

result of several selective forces acting simultaneously on different aspects of survival in the heterogeneous intertidal environment.

3.6.3. Biogeography and phylogeography

Two divergent haplogroups of *P. exigua* were identified corresponding to the west and east coasts, separated between Mossel Bay and Plettenberg Bay on the south coast. The star-shaped pattern of the west coast haplogroup indicated that the population expansion is recent or ongoing since the number of individuals sharing the central haplotype appears to have increased more rapidly than the evolution of mutations resulting in new divergent haplotypes. The divergent haplotypes occur at most locations indicating some incomplete lineage assortment and/or that gene flow between the sampling locations may be restricted. Lineage assortment has occurred towards the eastern end of this haplogroup's geographic distribution, indicated by no individuals from Mossel Bay represented in the ancestral haplotype, but shared haplotypes between Mossel Bay, Cape Agulhas and Betty's Bay, suggesting some gene flow between these three locations, but low or absent gene flow between these locations and the rest of the geographic range. The structure of the east coast haplogroup suggests that gene flow has been low at the extreme ends of the geographic range (i.e. between Plettenberg Bay and Salt Rock and the rest of the sampling locations), and that complete lineage sorting of haplotypes has occurred at these locations. The sampling locations in the middle of the haplogroup's geographic distribution share haplotypes indicating that there is potentially more gene flow between sampling sites on this stretch of coast, possibly due to the northerly inshore current flowing closer to the coast between these locations. The west coast population expansion prediction was supported by the mismatch distribution, associated statistics and time since expansion estimates which placed the west coast expansion at approximately 40,000 years ago. The west coast expansion appears to coincide with the period of strongest upwelling in the Benguela system which occurred between 42 and 20,000 years ago (Lindesay

1998; Partridge 2001), suggesting that the colder water temperature, or the associated increase in phytoplankton caused by upwelling, resulted in favourable conditions for the unmottled morph of *P. exigua*.

It is unknown whether the east and west haplogroups are reproductively isolated. The genetic structure and demographic history of *P. exigua* observed in South Africa has some similarities with that observed in Australia for the same species (Hunt 1993; Colgan *et al.* 2005). Colgan *et al.* (2005) found no shared haplotypes between populations of *P. exigua* in New South Wales, Tasmania and South Australia (approximate distances of 1100, 2250 and 1650 km, respectively), implying complete lineage assortment between these regions. Haplotypes were shared between the two populations within New South Wales implying some gene flow between these populations. However, the evolutionary distance between the haplotypes and therefore regions of their study was surprisingly low, at most only 0.006%. Colgan *et al.* (2005) also suggested that range expansion has occurred in the Australian populations of *P. exigua*, with subsequent low gene flow between regions. Similarly, Hunt (1993) showed that over a 230 km distance, *P. exigua* populations were strongly genetically partitioned, and gene flow was rare between widely separated populations, but some gene flow occurred between nearby populations.

Colgan *et al.* (2005) suggested that geographically limited haplotype distributions are characteristic of *P. exigua*, but Waters and Roy (2004) recorded one haplotype shared between Tasmania and Victoria (Australia). This pattern of limited haplotype distribution is not seen in either of the haplogroups in South Africa, with haplotypes shared between populations separated by 360 km (Table 3.5) on the east coast and haplotypes shared between almost all populations over a distance of 1040 km on the west coast. Additionally, the divergences estimated between populations are much higher (Table 3.5) than those estimated between Australian populations (Colgan *et al.* 2005). This discrepancy in genetic structure between South Africa and Australia could indicate either (i) more recent

or ongoing gene flow between the South African populations in comparison to low gene flow between the Australian populations, or (ii) more recent population expansion in both haplogroups in South Africa, indicated by the ancestral haplotype still being evident in most populations. An earlier investigation (Chapter 2) revealed that both the east and west South African haplogroups had oral gonopores, therefore it is assumed that the South African and Australian populations have the same reproductive mode and dispersal ability. As species with entirely benthic life cycles are assumed to have high genetic structure and restricted dispersal capability (Hunt 1993; Arndt and Smith 1998; Gaylord and Gaines 2000; Colgan *et al.* 2005), the latter explanation of the genetic structure in South Africa is more likely.

The history of population expansion in the west coast haplogroup does not explain the reciprocal monophyly and lack of introgression between the east and west haplogroups or how this genetic structure was formed. Previous studies have suggested upwelling can cause a barrier to gene flow (Bowen *et al.* 2001; Lessios *et al.* 2001; Waters and Roy 2004). Although there is some upwelling on the Agulhas Bank which may have contributed to the marked phylogeographic divide, the strong upwelling (Fig. 3.1) on the west coast did not appear to disrupt gene flow. The most plausible past demographic process that may have given rise to the reciprocally monophyletic haplogroup structure between the east and west coasts is allopatric divergence caused by a vicariance event. The time since divergence between the east and west haplogroups was estimated at between 1.4 - 1.25 MYA, near the beginning of the Pleistocene epoch. However, there have been no major geological changes along the South African coastline since the formation of the present continental shelf during the early Cretaceous (Heydorn *et al.* 1978), and there are no geological features that could have acted as an absolute barrier to the dispersal of benthic intertidal species during this time (Teske *et al.* 2006). Sea level or temperature changes during the Pleistocene (Branch and Branch 1981) may have exposed the continental shelf, creating a physical barrier to gene flow over the Agulhas Bank, separating the

east and west populations. Local adaptation, neutral genetic divergence and range expansion may have ensued in allopatry on both sides of the barrier, followed by further sea level or temperature changes causing the phylogeographic divide to break down and secondary contact to occur. The present inshore currents diverge northwards up either sides of the coast which may reduce or prevent present day gene flow between the two coasts, accounting for the lack of introgression between the haplogroups. As phylogeographic divides are seen in other species around the Agulhas Bank region (Teske *et al.* 2006), this is a plausible explanation. Teske *et al.* (2006) also found that the strength of the Agulhas Bank barrier to gene flow increases as the dispersal ability of the organism decreases. Vicariance events have been invoked as an explanation to account for other major genetic divergences in continuously distributed intertidal species (Reeb and Avise 1990).

The findings of this study have four general phylogeographic implications for the South Africa intertidal zone. Firstly, this study supports the presence of a phylogeographic divide on the south coast by the Agulhas Bank (Evans *et al.* 2004; Teske *et al.* 2006). Secondly, phylogeographic structure in *P. exigua* does not correspond to the four intertidal biogeographic provinces identified by Emanuel *et al.* (1992), indicating that biogeographic divides do not necessarily represent genetic divides, at least in this species. Thirdly, previous biogeographic investigations have suggested that the temperature regimes of the Benguela and Agulhas currents are the primary factors responsible for the distributions of intertidal organisms in South Africa (Stephenson and Stephenson 1972; Brown and Jarman 1978; Emanuel *et al.* 1992). However, the temperature gradient around the coast appears to have little influence on the genetic structure of *P. exigua*. This study indicates that past geography and the weaker inshore counter currents, especially on the east coast, may be more important for maintaining genetic structure than previously recognized. The east coast northerly inshore current may be reducing or preventing present day gene flow between the east and the west coast across the Agulhas Bank. Finally, genetic structure and

phylogeographic divides are not considered in the formation of marine reserves at present in South Africa, despite an awareness of the need to protect genetic as well as species diversity (Branch and Odendaal 2003; Palumbi 2003; Palumbi 2004). As the existence of a phylogeographic boundary across species in intertidal fauna around the Agulhas Bank will have implications for reproductive stocks, larval and juvenile dispersal, it should be considered when designating biogeographic provinces and marine reserves (Branch and Odendaal 2003).

3.7. Conclusions

The discontinuity between the neutral genetic structure and the colour morph frequency distribution of *P. exigua* leads us to conclude that phenotypic variation is not the result of population structure and neutral demographic processes. The concordance between the colour morph frequency distribution and the environmental biogeographic temperature provinces and habitat preferences suggest that colour is either under ecological selection directly, or linked to part of the genome that is under selection. These results suggest that the selective force maintaining the colour variation is related to predation and / or water temperature. However, the causative selective mechanisms maintaining colour polymorphisms in *P. exigua* remain unclear, and will almost certainly be the result of several selective forces acting simultaneously on different aspects of survival in the heterogeneous intertidal environment.

This study of neutral genetic variation indicates no correlation between the phylogeographic structure of *P. exigua* in South Africa and the presently recognized biogeographic provinces, but there is a major genetic divide in these organisms around the Agulhas Bank, which corresponds to genetic divides seen in other intertidal species, possibly created by a past vicariance event mediated by sea level or temperature changes. Furthermore, it appears that inshore currents on the east coast and upwelling events on the west coast may play a more important role in maintaining the genetic structure of near shore

invertebrates than the temperature regimes created by the large powerful offshore currents of the Benguela and Agulhas.

The next chapter will use a population genomics approach to identify potential loci under selection, and compare the distribution of these within populations to environmental variables to infer the potential selective forces shaping phenotype. Additionally, this approach will enable us to assess levels of present day gene flow in *P. exigua* and test whether the two genetic haplogroups are reproductively isolated.

3.8. References

Arndt A, Marquez C, Lambert P, Smith MJ (1996) Molecular phylogeny of Eastern Pacific sea cucumbers (*Echinodermata: Holothuriodea*) based on mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution* **6**, 425-437.

Arndt A, Smith MJ (1998) Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. *Molecular Ecology* **7**, 1053-1064.

Arrontes J, Underwood AJ (1991) Experimental studies on some aspects of the feeding ecology of the intertidal starfish *Patiriella exigua*. *Journal of Experimental Marine Biology and Ecology* **148**, 255-269.

Bandaranayake WM (2005) The nature and role of pigments of marine invertebrates. *Natural Product Reports* **23**, 223 - 255

Bandelt HJ, Forster P, Rohl A (1999) Median joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bingham B, Braithwaite L (1986) Defensive adaptations of the dendrochirote holothurian *Psolus chitonoides*. *Journal of Experimental Marine Biology and Ecology* **98**, 311 – 322.

Bolton JJ (1986) Marine phytogeography of the Benguela upwelling region on the west coast of southern Africa - a temperature dependent approach. *Botanica Marina* **29**, 251-256.

Bolton JJ, Anderson RJ (1997) Marine Vegetation. In: *Vegetation of southern Africa* (eds. Cowling RM, Richardson DM, Pierce SM), pp. 348 - 370. Cambridge University Press, Cambridge.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution* **55**, 1029-1039.

Branch G, Branch M (1980) Competition between *Cellana tramoserica* (*Gastropoda*) and *Patiriella exigua* (*Asteroidea*), and their influence on algal standing stocks. *Journal of Experimental Marine Biology and Ecology* **48**, 35 - 49.

Branch G, Branch M, Bannister A (1981) *The living shores of Southern Africa*, 5th edn. Struik, Cape Town.

Branch GM, Griffiths CL, Branch ML, Beckley LE (1999) *Two Oceans, A guide to the marine life of southern Africa*, 4th edn. David Philip Publishers (Pty) Ltd., Cape Town.

Branch GM, Odendaal F (2003) The effects of marine protected areas on the population dynamics of a South African limpet, *Cymbula oculus*, relative to the influence of wave action. *Biological Conservation* **114**, 255-269.

Brisson JA, De Toni DC, Duncan I, Templeton AR (2005) Abdominal pigmentation variation in *Drosophila polymorpha*: Geographic in the trait, and underlying phylogeography. *Evolution* **59**, 1046-1059.

Brown AC, Jarman N (1978) Coastal Marine Habitats. In: *Biogeography and Ecology of Southern Africa* (Eds. Werger MJA, Bruggen ACV), pp. 1239-1227. The Hague, Cape Town.

Bustamante RH, Branch GM (1996) Large scale patterns and trophic structure of southern African rocky shores: The roles of geographic variation and wave exposure. *Journal of Biogeography* **23**, 339-351.

Byrne M (1992) Reproduction of sympatric populations of *Patiriella gunnii, P. calcar* and *P. exigua* in New South Wales, *Asterinid* seastars with direct development. *Marine Biology* **114**, 297-316.

CAIN AJ, SHEPPERD PM (1954) Natural selection in *Cepaea*. *Genetics* **39**, 89-116.

Cameron RAD, Dillon PJ (1984) Habitat stability, population histories and patterns of variation in *Cepaea Nemoralis*. *Malacologia* **25**, 271-290.

Clark (1923) The Echinoderm Fauna of South Africa. *Annals of the South African Museum* **XIII**.

Colgan DJ, Byrne M, Rickard E, Castro LR (2005) Limited nucleotide divergence over large spatial scales in the asterinid sea star *Patiriella exigua*. *Marine Biology* **146**, 263-270.

Day JH (1974) *A guide to marine life on South African shores*, 2nd edn. A. A. Balkema, Cape Town.

Demeeus T, Michalakis Y, Renaud F, Olivieri I (1993) Polymorphism in heterogeneous environments, evolution of habitat selection and sympatric speciation - soft and hard selection models. *Evolutionary Ecology* **7**, 175-198.

Emanuel BP, Bustamante RH, Branch GM, Eekhout S, Odendaal FJ (1992) A zoogeographic and functional approach to the selection of marine reserves on

the west coast of South Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **12**, 341-354.

Emson RH, Crump RG, (1978) Brooding in *Asterina gibossa* Pennant. *Thalassia jugoslavica* **12**, 99-108.

Emson RH, Crump RG, (1979) Description of a new species of *Asterina* (*Asteroidea*), with an account of its ecology. *Journal of the Marine Biological association of the United Kingdom* **59**, 77-94.

Etter RJ (1988) Physiological stress and color polymorphism in the Intertidal Snail *Nucella lapillus*. *Evolution* **42**, 660-680.

Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliott NG (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology-Progress Series* **270**, 163-172.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN (vs. 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online, Application Note* **2005**, 47 - 50.

Forsman A (1999) Variation in thermal sensitivity of performance among colour morphs of a pygmy grasshopper. *Journal of Evolutionary Biology* **12**, 869-878.

Forsman A, Merilaita S (2003) Fearful symmetry? Intra-individual comparisons of asymmetry in cryptic vs. signalling colour patterns in butterflies. *Evolutionary Ecology* **17**, 491-507.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Fu YX, Li WH (1997) Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution* **14**, 195-199.

Gaylord B, Gaines SD (2000) Temperature or transport? Range limits in marine species mediated solely by flow. *American Naturalist* **155**, 769-789.

Gillespie RG, Oxford GS (1998) Selection on the color polymorphism in Hawaiian happy-face spiders: Evidence from genetic structure and temporal fluctuations. *Evolution* **52**, 775-783.

Golding B (1992) The prospects for polymorphisms shared between species. *Heredity* **68**, 263-276.

Grant WS, daSilvaTatley FM (1997) Lack of genetically subdivided population structure in *Bullia digitalis*, a southern African marine gastropod with lecithotrophic development. *Marine Biology* **129**, 123-137.

Grant WS, Schneider AC, Leslie RW, Cherry MI (1992) Population genetics of the brown mussel *Perna perna* in southern Africa. *Journal of Experimental Marine Biology and Ecology* **165**, 45-58.

Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution* **22**, 71-9.

Growns JE, Ritz DA (1994) Color variation in southern Tasmania populations of *Heliocidaris erythrogramma* (*Echinometridae, Echinoidea*). *Australian Journal of Marine and Freshwater Research* **45**, 233-242.

Harrison TD (2002) Preliminary assessment of the biogeography of fishes in South African estuaries. *Marine and Freshwater Research* **53**, 479-490.

Hart MW, Byrne M, Smith MJ (1997) Molecular phylogenetic analysis of life-history evolution in *Asterinid* starfish. *Evolution* **51**, 1848-1861.

Haylor GS, Thorpe JP, Carter MA (1984) Genetic and ecological differentiation between sympatric color morphs of the common intertidal sea anemone *Actinia equina*. *Marine Ecology-Progress Series* **16**, 281-289.

Heads M (2005) Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* **21**, 62-78.

Hedrick PW (1986) Genetic polymorphism in heterogeneous environments - a decade later. *Annual Review of Ecology and Systematics* **17**, 535-566.

Heydorn AEF, Bang ND, Pearce AF, *et al.* (1978) Ecology of Agulhas current region - assessment of biological responses to environmental parameters in southwest Indian Ocean. *Transactions of the Royal Society of South Africa* **43**, 151-190.

Hockey PAR, Cooper J, Duffy DC (1983) The roles of coastal birds in the functioning of marine ecosystems in southern Africa. *South African Journal of Science* **79**, 130-134.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution* **58**, 1329-1341.

Hoffman EA, Schueler FW, Jones AG, Blouin MS (2006) An analysis of selection on a colour polymorphism in the northern leopard frog. *Molecular Ecology* **15**, 2627-2641.

Hull SL, Winter LJ, Scott GW (2001) Habitat heterogeneity, body size and phenotypic diversity in *Idotea granulosa* (*Isopoda*) on the north-east coast of England. *Journal of the Marine Biological Association of the United Kingdom* **81**, 949-954.

Hunt A (1993) Effects of contrasting patterns of larval dispersal on the genetic connectedness of local populations of two intertidal starfish, *Patiriella calcar* and *P. exigua*. *Marine Ecology Progress Series* **92**, 179-186.

Hutchings L (1994) The Agulhas Bank - a synthesis of available information and a brief comparison with other east coast shelf regions. *South African Journal of Science* **90**, 179-185.

Jackson L (1976) Aspects of the intertidal ecology of the east coast of South Africa, pp. 1 - 72. Oceanographic Research Institute, Durban.

Johannesson K (2003) Evolution in *Littorina*: ecology matters. *Journal of Sea Research* **49**, 107-117.

Johannesson K, Ekendahl A (2002) Selective predation favouring cryptic individuals of marine snails (*Littorina*). *Biological Journal of the Linnean Society* **76**, 137-144.

Jones JS, Leith BH, Rawlings P (1977) Polymorphism in *Cepaea* - problem with too many solutions. *Annual Review of Ecology and Systematics* **8**, 109-143.

Jordaens K, Van Riel P, Geenen S, Verhagen R, Backeljau T (2001) Food-induced body pigmentation questions the taxonomic value of colour in the self-fertilizing slug *Carinarion* spp. *Journal of Molluscan Studies* **67**, 161-167.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Kimura M (1983) Rare variant alleles in the light of the Neutral Theory. *Molecular Biology and Evolution* **1**, 84-93.

Kumar S, Tamura K, Nei M (2004) MEGA 3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

Laudien J, Flint NS, van der Bank FH, Brey T (2003) Genetic and morphological variation in four populations of the surf clam *Donax serra* (Roding) from southern African sandy beaches. *Biochemical Systematics and Ecology* **31**, 751-772.

Le Gac M, Feral JP, Poulin E, Veyret M, Chenuil A (2004) Identification of allopatric clades in the cosmopolitan ophiuroid species complex *Amphipholis squamata* (*Echinodermata*). The end of a paradox? *Marine Ecology Progress Series* **278**, 171-178.

Lee WL (1966) Pigmentation of the marine isopod *Idothea granulosa* (Rathke). *Comparative Biochemistry and Physiology* **19**, 13 - 27.

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution* **55**, 955 - 975.

Lessios HA, Kessing BD, Robertson DR, Paulay G (1999) Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution* **53**, 806-817.

Lindesay J (1998) Past climates of southern Africa. In: *Climates of Southern Continents, Present, Past and Future* (eds. Hobbs JE, Lindesay JA, Bridgman HA). John Wiley & Sons, Chichester.

Lombard AT, Grant WS (1986) Biochemical population genetics of the black mussel *Choromytilus meridionalis*. *South African Journal of Zoology* **21**, 131-135.

Lucas J, Hart R, Howden M, Salathe R (1979) Saponins in eggs and larvae of *Acanthaster planci* (L.) (Asteroidea) as chemical defenses against planktivorous fish. *Journal of experimental Marine Biology and Ecology* **40,** 155 – 165.

Mackenzie JB, Munday PL, Willis BL, Miller DJ, van Oppen MJH (2004) Unexpected patterns of genetic structuring among locations but not colour morphs in *Acropora nasuta* (*Cnidaria; Scleractinia*). *Molecular Ecology* **13**, 9-20.

Macnae W (1961) The fauna and flora of the eastern coasts of southern Africa in relation to ocean currents. *South African Journal of Science* **58**, 208 - 212.

Marlow JR, Lange CB, Wefer G, Rosell-Mele A (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* **290**, 2288-2291.

McCartney MA, Keller G, Lessios HA (2000) Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. *Molecular Ecology* **9**, 1391-1400.

McClintock J, Vernon J (1990) Chemical defense in the eggs and embryos of Antarctic sea stars (Echinodermata). *Marine Biology* **105**, 491 – 495.

Merilaita S (2001) Habitat heterogeneity, predation and gene flow: colour polymorphism in the isopod, *Idotea baltica*. *Evolutionary Ecology* **15**, 103-116.

Merilaita S, Kaitala V (2002) Community structure and the evolution of aposematic coloration. *Ecology Letters* **5**, 495-501.

Mila B, Girman DJ, Kimura M, Smith TB (2000) Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 1033-1040.

Miner BG, Sultan SE, Morgan SG, Padilla DK, Relyea RA (2005) Ecological consequences of phenotypic plasticity. *Trends in Ecology & Evolution* **20**, 685-692.

Oosthuizen A, Jiwaji M, Shaw P (2004) Genetic analysis of the octopus *vulgaris* population on the coast of South Africa. *South African Journal of Science* **100**, 603-607.

Oxford GS (2005) Genetic drift within a protected polymorphism: Enigmatic variation in color morph frequencies in the candy-stripe spider, *Enoplognatha ovata*. *Evolution* **59**, 2170-2184.

Palma AT, Steneck RS (2001) Does variable coloration in juvenile marine crabs reduce risk of visual predation? *Ecology* **82**, 2961-2967.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, S146-S158.

Palumbi SR (2004) Marine reserves and ocean neighborhoods: The spatial scale of marine populations and their management. *Annual Review of Environment and Resources* **29**, 31-68.

Partridge TC, Scott L (2001) Between Agulhas and Benguela: responses of southern African climates of the late Pleistocene to current fluxes, orbital precession and the extent of the circum-Antarctic vortex. *Past Climate Variability Through Europe and Africa*. Centre des Congrès, Aix-en-Provence, France

Peschak T (2006) *Currents of Contrast: Life in Southern Africa's Two Oceans* Struik Publishers (Pty) Ltd.

Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species – mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**, 397-406.

Ridgway TM, Branch GM, Stewart BA, Hodgson AN (2000) Taxonomic status of the *'Patella miniata'* species complex (*Mollusca : Gastropoda*) in southern Africa. *Hydrobiologia* **420**, 103-118.

Ridgway TM, Stewart BA, Branch GM, Hodgson AN (1998) Morphological and genetic differentiation of *Patella guanularis* (*Gastropoda : Patellidae*): recognition of two sibling species along the coast of southern Africa. *Journal of Zoology* **245**, 317-333.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.

Rosenblum EB (2006) Convergent evolution and divergent selection: Lizards at the White Sands ecotone. *American Naturalist* **167**, 1-15.

Rozen S, Skaletsky H, (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods*

and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Russell AL, Medellin RA, McCracken GF (2005) Genetic variation and migration in the Mexican free-tailed bat (*Tadarida brasiliensis mexicana*). *Molecular Ecology* **14**, 2207-2222.

Scheiner SM (2002) Selection experiments and the study of phenotypic plasticity. *Journal of Evolutionary Biology* **15**, 889-898.

Shannon LV (1985) The Benguela Ecosystem. 1. Evolution of the Benguela, physical features and processes. *Oceanography and Marine Biology* **23**, 105-182.

Sink K, Harris J, Lombard A, (2004) Appendix 1. South African marine bioregions. *South African National Spatial Biodiversity Assessment 2004: Technical Report, Marine Component* **4.**

Sokolova IM, Berger VJ (2000) Physiological variation related to shell colour polymorphism in White Sea *Littorina saxatilis*. *Journal of Experimental Marine Biology and Ecology* **245**, 1-23.

Sponer R, Deheyn D, Roy MS (2001) Large genetic distances within a population of *Amphipholis squamata* (*Echinodermata; Ophiuroidea*) do not support colour varieties as sibling species. *Marine Ecology Progress Series* **219**, 169-175.

Stegenga H, Bolton JJ (1992) *Ceramiaceae* (*Rhodophyta)* of the Cape Province, South Africa: Distribution in relation to concepts of marine provinces. *Botanica Marina* **35**, 99 - 107.

Stephenson TA, Stephenson A (1972) *Life between tidemarks on rocky shores.* Freeman, San Francisco.

Stevenson JP (1992) A possible modification of the distribution of the intertidal seastar *Patiriella exigua* (Lamarck) (*Echinodermata, Asteroidea*) by *Patiriella calcar* (Lamarck). *Journal of Experimental Marine Biology and Ecology* **155**, 41-54.

Stoletzki N, Schierwater B (2005) Genetic and color morph differentiation in the Caribbean sea anemone *Condylactis gigantea*. *Marine Biology* **147**, 747-754.

Tajima F (1989) Statistical method for testing the Neutral Mutation Hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tarjuelo I, Posada D, Crandall KA, Pascual M, Turon X (2004) Phylogeography and speciation of colour morphs in the colonial ascidian *Pseudodistoma crucigaster*. *Molecular Ecology* **13**, 3125-3136.

Teske PR, Cherry MI, Matthee CA (2003) Population genetics of the endangered Knysna seahorse, *Hippocampus capensis*. *Molecular Ecology* **12**, 1703-1715.

Teske PR, McQuaid CD, Froneman PW, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology-Progress Series* **314**, 283-293.

Todd PA, Briers RA, Ladle RJ, Middleton F (2006) Phenotype-environment matching in the shore crab (*Carcinus maenas*). *Marine Biology* **148**, 1357-1367.

Tolley KA, Groeneveld JC, Gopal K, Matthee CA (2005) Mitochondrial DNA panmixia in spiny lobster *Palinurus gilchristi* suggests a population expansion. *Marine Ecology Progress Series* **297**, 225-231.

Trussell GC (2000) Phenotypic clines, plasticity, and morphological trade-offs in an intertidal snail. *Evolution* **54**, 151-166.

Turpie JK, Beckley LE, Katua SM (2000) Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation* **92**, 59-72.

Waters JM, Roy MS (2003) Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography* **30**, 1787-1796.

Waters JM, Roy MS (2004) Out of Africa: The slow train to Australasia. *Systematic Biology* **53**, 18-24.

Williams ST (2000) Species boundaries in the starfish genus *Linckia*. *Marine Biology* **136**, 137-148.

Williams ST, Benzie JAH (1998) Evidence of a biogeographic break between populations of a high dispersal starfish: Congruent regions within the Indo-West Pacific defined by color morphs, mtDNA, and allozyme data. *Evolution* **52**, 87-99.

Wright S (1978) *Variability within and among natural populations* Univ. of Chicago Press, Chicago.

3.9. Appendix 1.

```
     6                              3          44     4      5
                                    3          23     8      4
CAAGATGACCAAATATACAAAGTAATAGTTACGGCACACGCTGTAGTGATGATGTTTTT
6     6         7      8  8      9  9  9                     1
0     6         5      1  4      0  3  6                     1
                                                            7
TATGGTGATGCCAATAATGATTGGAGGATTGGGAAATTGACTAATTCCCCTTATGATCG
       1  1         1             1  1      1      1      1      1
       2  2         3             5  5      5      6      7      7
       6  9         8             0  3      9      5      1      7
GAGCCCCCGATATGGCTTTCCCACGAATGAAAAACATGAGCTTTTGGCTAATACCCCCT
  1      1  1      1            2  2  2  2  2  2      2      2
  8      8  8      9            0  1  1  1  1  2      2      3
  0      6  9      5            7  0  3  6  9  2      8      4
TCTTTCCTCCTTCTTATTGCCTCAGCAGGGGTAGAAAGAGGTGCTGGAACAGGCTGGAC
             2  2  2  2  2      2  2  2         2      2
             4  5  5  5  6      6  7  7         8      8
             9  2  5  8  1      7  0  3         2      8
AATATACCCCCCACTTTCTAGCGGACTAGCACATGCAGGAGGCTCAGTAGACCTTGCAA
    3  3  3  3          3  3      3             3  3      3
    0  0  0  0          1  2      2             3  4      4
    0  3  6  9          8  1      7             9  2      8
TATTTTCCCTTCACTTAGCAGGAGCTTCCTCGATACTTGCCTCTATCAAATTCATTACA

ACAG
```

Figure 3.7. Position of sequence ambiguities for the *Parvulastra exigua* Cytochrome Oxidase I partial gene sequence for Haplotype 1 from Chapter 3.

3.9. Appendix 2.

```
            6344456678 8999111111 1111111122 2222222222 2222223333 33333
             323840651 4036122355 5677888901 1112234555 6677880000 22344
                   769803 9517069570 3692849258 1703280369 17928
Hap_1       TCTCCCTGAT ACATCCTCCC CGATTCTTGA AATTAGATTC AATAACTCTC TGTCC
Hap_2       .......... .......... .......... .......C.. .......... .....
Hap_3       .......... .......... .A........ .......... .......... .....
Hap_4       .......... ....T..... .......... .......... .......... .....
Hap_5       .......... ....T..... .......... G......... .......... .....
Hap_6       CTC.T..T.C GT..TA.TTT .A.C.TAAA. .T.C.A...T GT.CCT.A.T .AA.T
Hap_7       .....T.... .......... .......... .......... .......... .....
Hap_8       .......... ...C...... .......... .......... .......... .....
Hap_9       .......... .......... .......... .........A .......... ...T.
Hap_10      .......... .......... .........G .......... .......... .....
Hap_11      .......... ..G....... .......... .......... .......... .....
Hap_12      .......... ........T. .A........ .......... .......... .....
Hap_13      .......... .......... .......... ........G. .......... .....
Hap_14      .......... .......... .......... .........T .......... .....
Hap_15      ......C... .......... .......... .......... .......... .....
Hap_16      .......A.. .......... .......... .......... .......... .....
Hap_17      .......... .......... .......... ....G..... .......... .....
Hap_18      .......... .......... .......... .......... .......... .C...
Hap_19      .......... .......... .......... .......... ..C.....C. .....
Hap_20      .......... .......... .......... ...C...... ........C. .....
Hap_21      .........C .......... .......... .......... .......... .....
Hap_22      .......... .......... .......... .......... ........C. .....
Hap_23      .......... .......... .......... .......... ...C...... .....
Hap_24      ...T...... .......... .......... .......... .......... .....
Hap_25      .......... ....TA...T .A..CT..A. ...C.A.... ....CT.A.. C....
Hap_26      .......C.. ....TA...T .A..CT..A. ...C.A.... ....CT.A.. C....
Hap_27      .......... ....TAC..T .A..CT..A. ...C.A.... ....CT.A.. C....
Hap_28      .......... ....TA...T .A..CT..A. ...C.AG... ....CT.A.. C....
Hap_29      ....T..A.. ....TA.... .A...T..A. .......... ....CT.A.. CA...
Hap_30      ....T..... ....TA.... .A...T..A. .......... ....CT.A.. C....
Hap_31      ....T..A.. ....TA.... .AC..T..A. .G........ ....CT.A.. CA...
Hap_32      ....T..A.. ....TA..T. .A...T..A. .......... ....CT.A.. CA...
Hap_33      ....T..A.. ....TA.... .A...T..A. .C........ ....CT.A.. CA...
Hap_34      ....T..A.. ....TA.... .A...T..A. G......... ....CT.A.. CA...
Hap_35      C.C.T..AT. ....TA.TTT TA......T. ..A....... ...G.TCA.. C.C..
Hap_36      C.C.T..AT. ....TA.TTT TA......T. ..A....... G..G.TCA.. C.C..
```

Figure 3.8. Alignment of the sequence ambiguities for the *Parvulastra exigua* Cytochrome Oxidase I partial gene for all haplotypes from Chapter 3.

3.9. Appendix 3.

Graphical predictions (with 95% confidence intervals) of the GLMM for the habitat variables of *Parvalustra exigua* presented as back transformed values of the dependent variable.

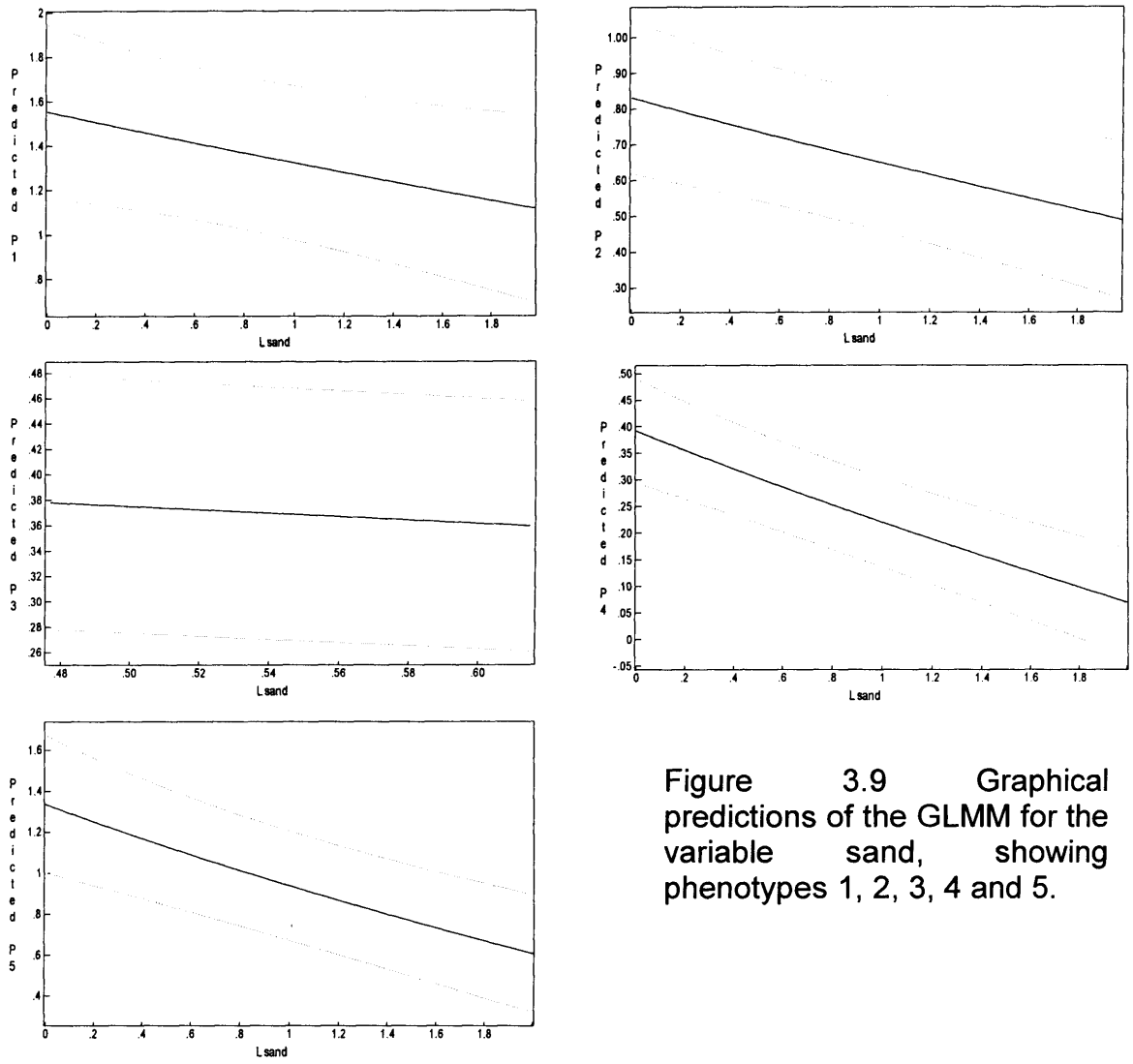

Figure 3.9 Graphical predictions of the GLMM for the variable sand, showing phenotypes 1, 2, 3, 4 and 5.

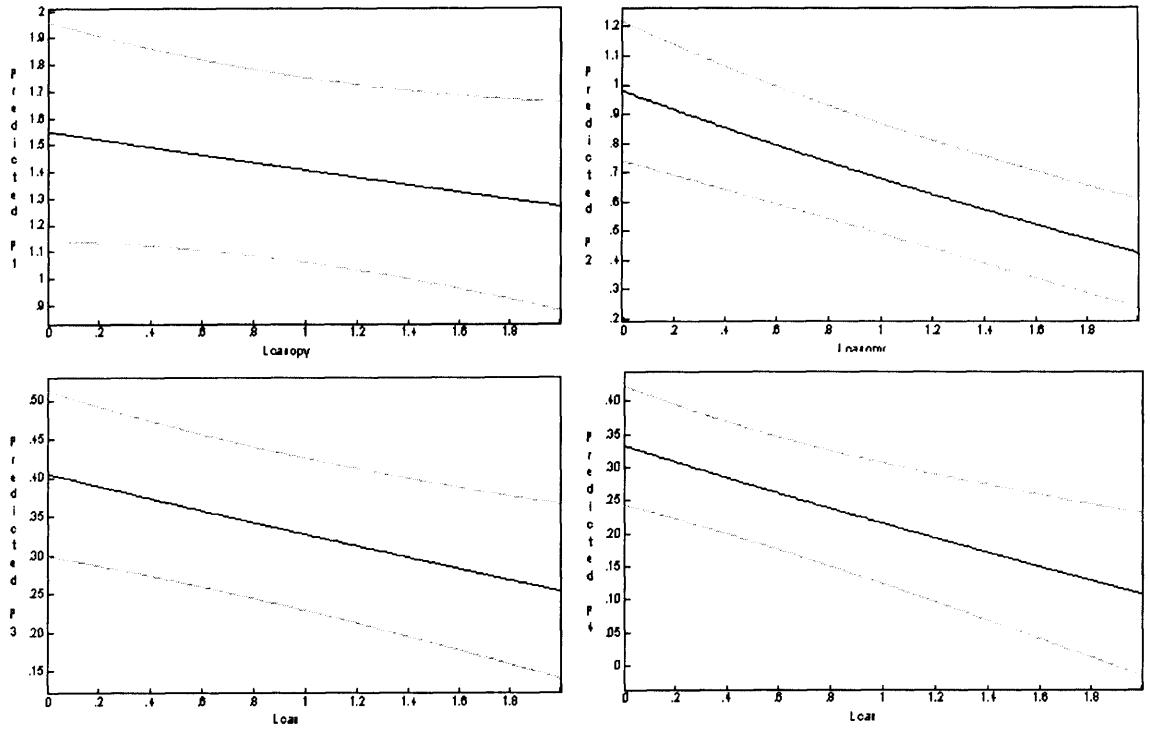

Figure 3.10. Graphical prediction of the GLMM for the variable Canopy, showing phenotypes 1, 2, 3 and 4.

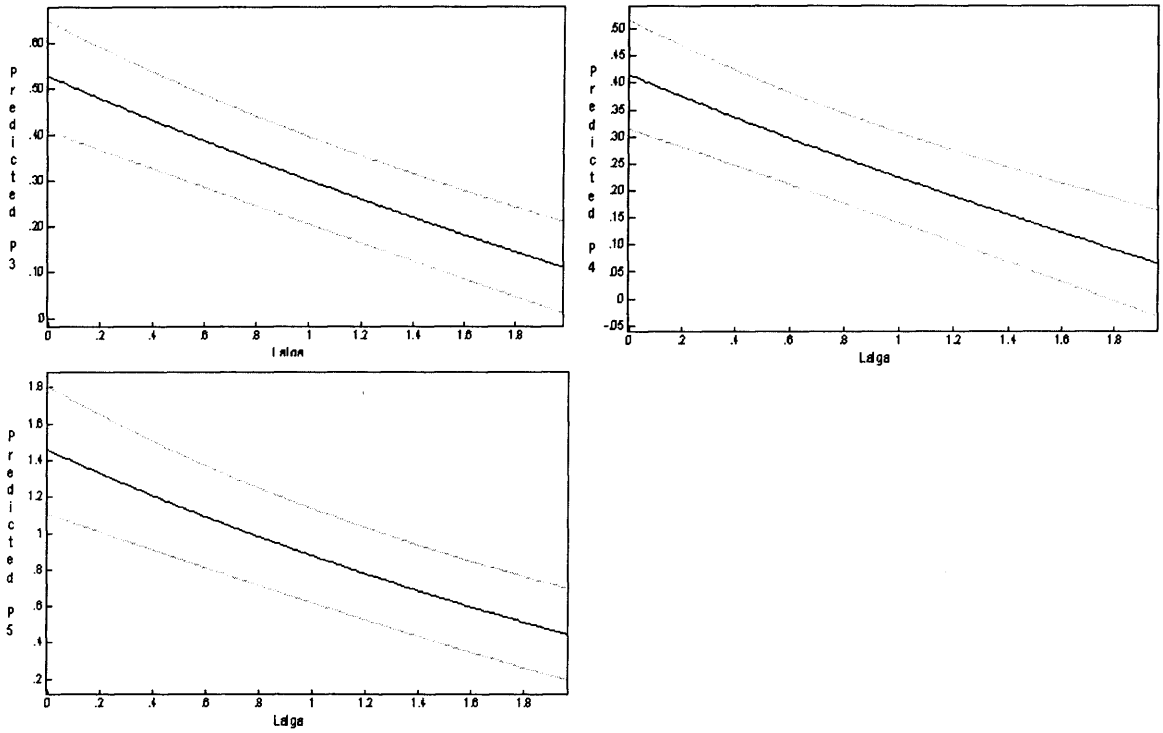

Figure 3.11. Graphical prediction of the GLMM for the variable algal tufts, showing phenotypes 3, 4, and 5.

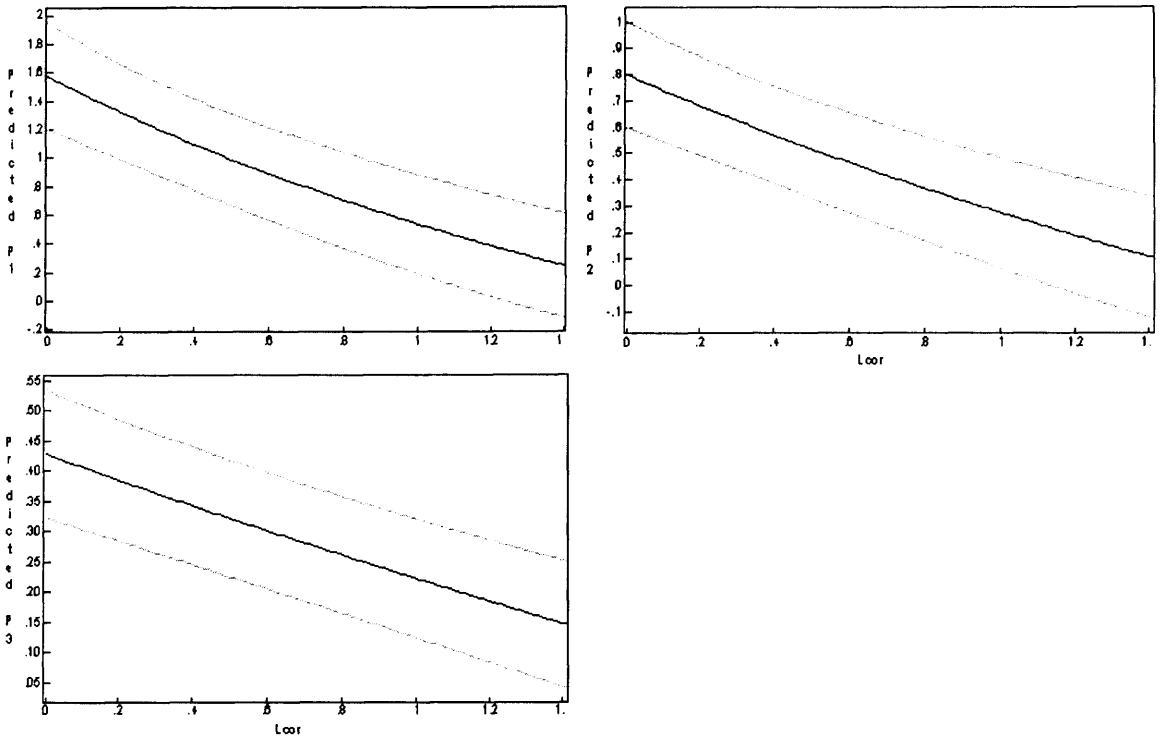

Figure 3.12. Graphical prediction of the GLMM for the variable Coralline algae, showing phenotypes 1, 2, 3 and 4.

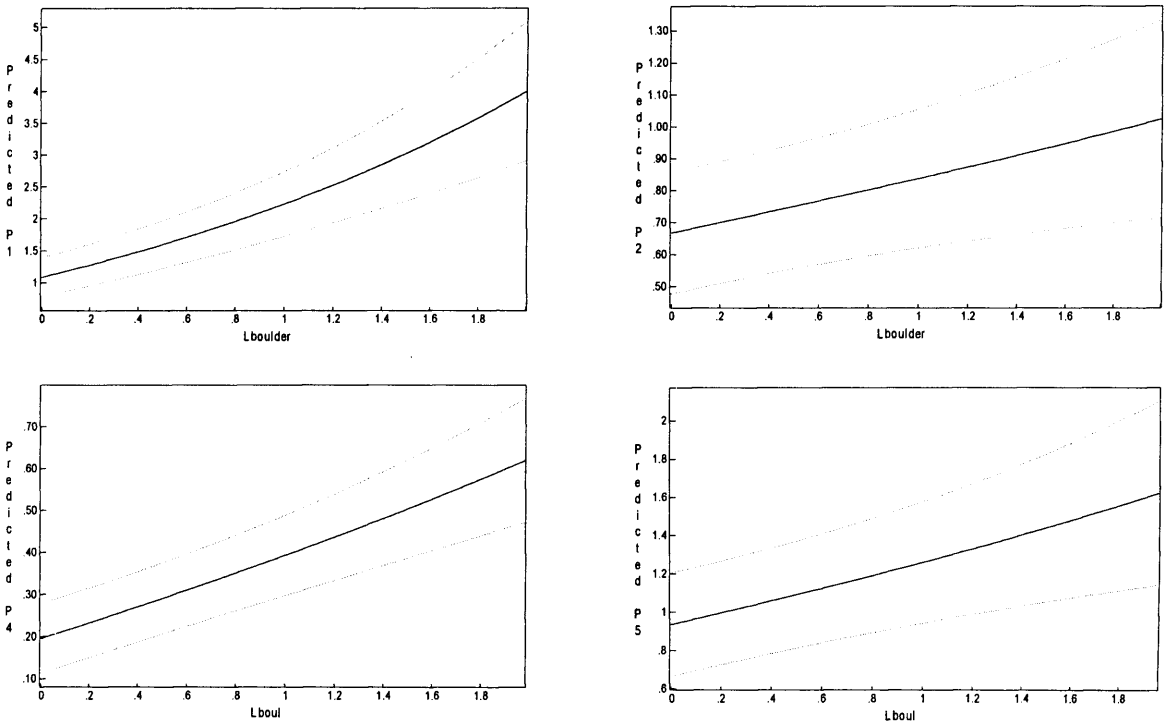

Figure 3.13. Graphical prediction of the GLMM for the variable under boulder, showing phenotypes 1, 2, 4, and 5.

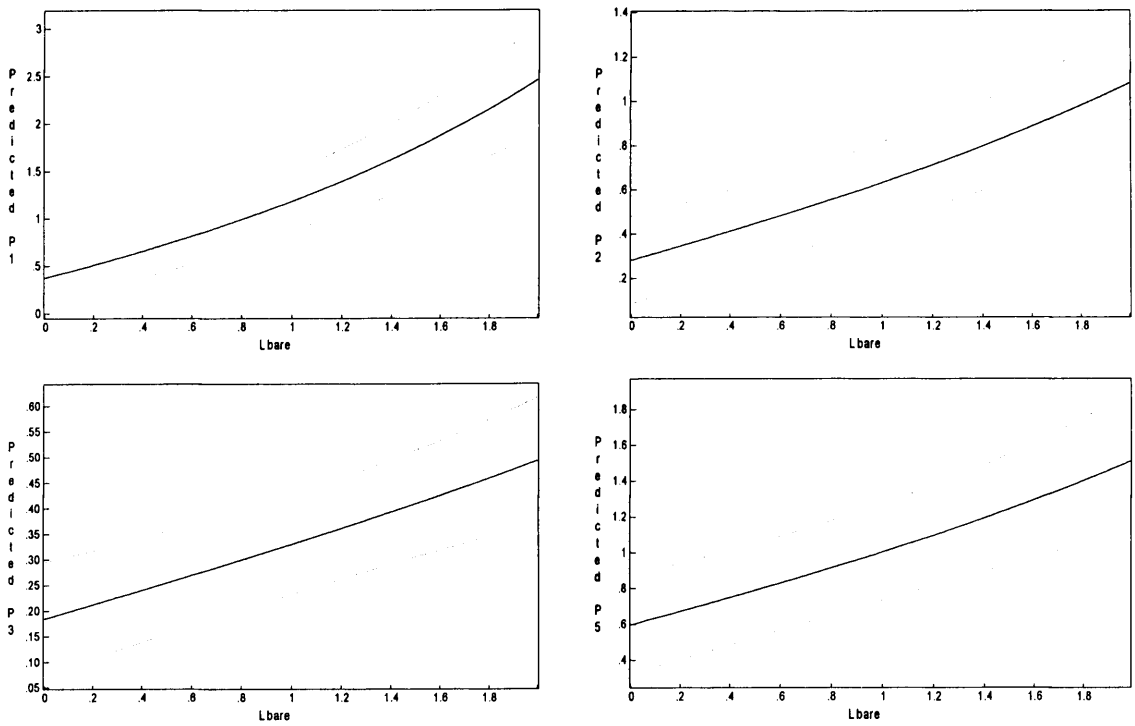

Figure 3.14. Graphical prediction of the GLMM for the variable bare rock, showing phenotypes 1, 2, 3, and 5.

Chapter 4

Spatial genetic structure and directional selection along an environmental gradient in an intertidal sea star

Chapter 4: Spatial genetic structure and directional selection along an environmental gradient in an intertidal sea star

K. S. Dunbar[1], J. Cable[1], C. Griffiths[2], M.W. Bruford[1].

[1]School of Biosciences, Cardiff University, Cardiff CF10 3TL, U.K.

[2]Marine Biology Research Institute, University of Cape Town, Rondebosch, South Africa.

4.1. Abstract

Diversifying selection and local adaptation can potentially occur in populations distributed across environmental gradients, or inhabiting geographic ranges where habitat heterogeneity exists over a larger spatial scale than the dispersal potential of the populations. To identify diversifying selection, neutral patterns of genetic structure need to be established in order to detect deviations from neutrality. A genome scan approach, using 307 AFLP loci was applied to establish the population genetic structure among 16 populations of a South African intertidal sea star species, *Parvulastra exigua*, which displays phenotypic differentiation across a temperature gradient between the Indian and Atlantic oceans. Coalescent-based simulation was used to establish neutral expectations of genetic structure and loci were identified and analysed where they deviated from neutral expectations. We found a strong pattern of isolation by distance around the coast, but approximately 6.8% of the genome displayed a higher than expected differentiation from this neutral model, potentially identifying genomic regions under diverging selection. Phenotype frequencies of divergent outlier loci within populations showed evidence for directional selection, broadly corresponding to temperature and phenotypic gradients between the east and west coasts of South Africa. This pattern could be the result of local adaptation linked to the temperature gradient or the phenotypic differentiation, although a

genetic signature of secondary contact and introgression following a period of allopatry cannot be ruled out.

4.2. Introduction

Four major evolutionary processes shape genetic diversity in wild populations; genetic drift and selection reduce genetic diversity, whereas mutation and gene flow increase it (Luikart *et al.* 2003). Selection and mutation tend to be locus specific, whereas genetic drift and gene flow tend to be genome-wide effects. Genetic drift involves random changes in allele frequencies that occur over generations due to chance events alone. Natural selection, the process highlighted by Darwin, relies on the fact that within every population there is random variation that is of different fitness value. Variations which aid fitness are 'selected' by being preferentially inherited in succeeding generations. Mutation involves changes in the genome which result in new or rearranged gene combinations. Gene flow is defined as the proportion of individuals within a population of each generation that successfully breed after migration (Hellberg *et al.* 2002). This chapter focuses on detecting spatial patterns of gene flow and selection, and investigating the interplay between these two forces in a continuously distributed intertidal population of sea stars occurring along a substantial environmental gradient. Using a genome wide sampling framework, a coalescent based simulation model was used to determine the genome-wide effects of population demography and phylogenetic history, in order to identify loci by contrast that may be under divergent selection (Luikart *et al.* 2003).

4.2.1. Determinants of gene flow in marine species

Examining what determines gene flow in the marine environment can provide insights into neutral genetic structure of populations and potential selective forces (Palumbi 1994; Hilbish 1996; Grosberg and Cunningham 2001; Hellberg *et al.* 2002). Gene flow can be shaped by both past and present abiotic conditions, such as plate tectonics, natural barriers or current systems, as well as the biology

of the species (Hilbish 1996; Grosberg and Cunningham 2001; Hellberg *et al.* 2002). The dispersal potential of different life history stages (e.g. gametes, larvae, adults) can greatly influence the level of gene flow (Holt and Gaines 1992). Species with high dispersal potential tend to show low genetic differentiation over large spatial scales (Booth and Ovenden 2000), whereas those with poor dispersal tend to be more genetically structured (Riginos and Victor 2001; Sponer and Roy 2002 but see Hilbish 1996), but often dispersal potential does not correlate with actual gene flow. For example, on the east coast of North America biogeographic boundaries are set by steep temperature gradients and near shore currents, but gene flow estimates based on larval biology indicate that genetic variation is structured over a much shorter geographic scale than predicted (Palumbi 1994 and references therein). Conversely, sea stars with low dispersal potential which have an entirely benthic life history and 'crawl away' juveniles have global geographic distributions that far exceed their predicted dispersal potential (Hart *et al.* 2006). Therefore gene flow alone cannot account for the different patterns of geographic and genetic structure in populations, and selection may be a contributing factor. There is potential for diversifying selection and local adaptation, e.g. phenotypic adaptations, or speciation to occur where habitat heterogeneity exists within a species' geographic range (Kreitman and Akashi 1995; Hilbish 1996; Schluter 2001; Veliz *et al.* 2006). Over large spatial scales habitat heterogeneity often exists in the marine environment as different circulation patterns and currents provide different ecological niches (Grosberg and Cunningham 2001). Before diversifying selection can be investigated, it is important to understand the patterns of gene flow among populations. If no gene flow is occurring between populations under different selective pressures, then there is no homogenising effect counteracting adaptation, and divergence can occur in allopatry to different environmental conditions. However, if there is gene flow, then diversifying selection may occur despite the homogenising effects of gene flow (Smith *et al.* 1997).

4.2.2. Selection

Investigating the genetic basis of adaptation and population divergence may help us to understand natural selection and adaptive evolution in changing environments, and ultimately protect genetic biodiversity. Ecological theories of adaptive radiation (e.g. Schluter 1996), suggest that phenotypic divergence and speciation are the ultimate consequences of divergent natural selection in contrasting environments (Storz 2005). Selection is the principal force shaping phenotypic variation (Rieseberg *et al.* 2002), the direction and strength of which may vary considerably in space (Steinger *et al.* 2002). Furthermore, selection may shape phenotypic differences between populations even when there is gene flow between differing phenotypes (Smith *et al.* 1997; Bensch and Akesson 2005). Several studies have suggested that selection is responsible for maintaining divergent phenotypes within sympatric or parapatric populations and that gene flow among the divergent phenotypes is evident (Wilding *et al.* 2001; Storz and Dubach 2004). Where phenotypic polymorphisms occur over steep environmental gradients, diversifying selection may be responsible for driving adaptive differentiation in the face of gene flow, which may lead to the acquisition of reproductive isolation and eventual speciation. Alternatively, phenotypic polymorphisms may be maintained by non-selective mechanisms and governed only by quasi-neutral processes such as mutation or drift (Hoffman *et al.* 2006). Neutral demographic processes are predicted to have relatively uniform effects across the entire genome, whereas natural selection will generally be more locus-specific, affecting only regions of the genome under selection or those linked to loci under selection (Beaumont and Nichols 1996; Luikart *et al.* 2003; Beaumont and Balding 2004).

An approach to investigate selective versus neutral genetic processes and to identify local adaptation or selection at genes with key functional roles was first proposed by Lewontin and Krakauer (1973). Since then several modified statistical methods have been developed (Bowcock *et al.* 1991; Beaumont and Nichols, 1996; Vitalis *et al.* 2001; Schlotterer 2002; Porter 2003; Beaumont and

Balding 2004) using a variety of different genetic markers (Andres *et al.* 2000; Akey *et al.* 2002; Kayser *et al.* 2003; Hoffman *et al.* 2006) to identify areas of the genome that are potentially under diversifying or stabilising selection. Similar approaches have also been used to investigate selection on specific candidate loci linked to colour pattern polymorphism (Gillespie and Oxford 1998; Andres *et al.* 2000, 2002; Hoffman *et al.* 2006), selection on divergent phenotypes (Campbell and Bernatchez 2004), speciation (Savolainen *et al.* 2006) and selection in populations distributed across environmental gradients (Wilding *et al.* 2001; Storz and Dubach 2004). Several authors have implemented these approaches at a genome-wide scale, termed 'genome scans' (Storz 2005) or a population genomics approach (reviewed in Luikart *et al.* 2003) using Amplified Fragment Length Polymorphism (AFLP; Wilding *et al.* 2001; Campbell and Bernatchez 2004; Achere *et al.* 2005; Mealor and Hild 2006; Murray and Hare 2006; Savolainen *et al.* 2006). AFLP analysis is ideal for rapidly screening many loci that are randomly distributed throughout the genome, allowing an accurate estimate of the genomic differentiation among populations and a thorough screen for regions of the genome under divergent selection (Murray and Hare 2006).

These methods for detecting selection use simulations to generate a null distribution of a particular summary statistic (such as Fst) under a neutral model of population structure. The majority of loci will conform to the neutral model but outlying loci may exhibit lower than expected levels of differentiation, suggesting some form of balancing selection, and some loci may exhibit higher than expected differentiation, suggesting diversifying selection (Storz 2005). These differences will occur at a small number of DNA sites but are potentially identifiable because linkage will lead to 'islands' of differentiation around the selected sites and any markers sampled within an 'island' should also show differentiation (Beaumont and Balding 2004). Phenotype frequencies of such 'outlier loci' can then be examined across populations. Large inter-population allele frequency differences or clines in allele frequencies may be evident at loci that control traits under differential adaptation or selection (Beaumont and

Balding 2004). The identified outlier loci phenotype frequencies may reflect diversifying or stabilising selection that correspond to a phenotypic gradient or to suspected selective agents within or among populations, such as environmental gradients.

In the absence of barriers to dispersal steep clines in allele frequencies may reflect a history of diversifying selection. However, they can also reflect the interplay between drift and spatially restricted gene flow (isolation by distance) or by admixture between previously isolated populations that have come into secondary contact (Wilding *et al.* 2001; Storz and Dubach 2004; Murray and Hare 2006). These two correlates of clinal variation in allele frequencies are very difficult to distinguish, since introgression of neutral markers and recombination are expected to break down the signal of past separation, unless the cline is stabilized by biotic or abiotic factors (Barton and Hewitt 1985; Wilding *et al.* 2001; Grahame *et al.* 2006; Murray and Hare 2006).

4.2.3. The intertidal sea star *Parvulastra exigua*

Due to its continuous distribution over a geographically heterogeneous habitat, the small South African intertidal sea star, *Parvulastra exigua* (see previous Chapters for intertidal environment of southern Africa, *Parvulastra* taxonomy, ecology, geographic range, life history, phenotypic variation of *P. exigua* and neutral *P. exigua* population structure) is a useful model system for exploring evidence of adaptive divergence. Several features of *P. exigua* are comparable to other biological systems in which diversifying selection may be driving adaptive differentiation (Savolainen *et al.* 2006). Firstly, there is phenotypic divergence in *P. exigua* coupled with low dispersal ability over a well characterised intertidal temperature gradient (Chapters 1, 2 and 3). In populations that are distributed across steep environmental gradients the potential for local adaptation is determined by the spatial scale of fitness variation relative to the dispersal ability of the species (Storz and Dubach 2004). The lack of a planktonic larval stage in *P. exigua* and its continuous distribution around the

South African coast suggests that its dispersal ability and gene flow may be restricted to adjacent populations, and therefore an isolation by distance pattern of genetic structure is predicted. Previous authors have found evidence of adaptive divergence within populations that show phenotypic differentiation over environmental gradients (Wilding *et al.* 2001; Storz and Dubach 2004). However, these studies tested gradients which span relatively small geographical distances, such as latitudinal clines (Storz and Dubach 2004) or vertical clines on a rocky shore (Wilding *et al.* 2001) and therefore stable maintenance of adaptive divergence would require strong selection to counterbalance the homogenizing effects of gene flow (Storz and Dubach 2004).

Secondly, the mottled polymorphism is evident in both *P. exigua* and its sister species *P. dyscrita*, over the same geographical locations. The presence of colour polymorphisms that transcend species boundaries suggests that they have been retained from a common ancestor and are therefore unlikely to be selectively neutral (Golding 1992; Richman 2000; Oxford 2005). Thirdly, the unmottled and mottled phenotypes of *P. exigua* show a different habitat usage (Chapter 3). Ecological shifts in habitat usage, accompanied by phenotypic differentiation represent a specific case of divergence where deterministic forces, rather than random processes, are likely to be implicated (Orr and Smith 1998). Other studies have tentatively suggested that some loci display divergent selection which may be operating on the phenotype and that phenotype is influenced by habitat use (Wilding *et al.* 2001; Campbell and Bernatchez 2004, but see Hoffman *et al.* 2006). Grahame *et al.* (2006) briefly discussed whether habitat choice by different *Littorina saxatilis* phenotypes acted to maintain shell polymorphism, or if habitat choice was evolving as a result of habitat related fitness differences of different phenotypes. They concluded that strong selection and habitat choice are responsible for maintaining a cline in allele frequency at some loci and that habitat choice would reduce the opportunities for mating between morphs and so increase the barrier to gene exchange. In the *P. exigua* system, the colour morphs are partially geographically separated (Chapter 3) as

well as displaying some habitat preference, and therefore the potential for restricted mating also exists.

However, prior to using *P. exigua* as a model for exploring evidence of adaptive divergence and selection, a taxonomic issue has to be resolved. *Parvulastra exigua* has two divergent monophyletic mtDNA clades geographically separated on the south coast, but this genetic divide does not coincide with the phenotypic divergence (Chapter 3). Thus, *P. exigua* might not represent a cohesive interbreeding taxonomic unit ('species'). Moreover, the taxonomic history of the genus *Parvulastra* in South Africa is complex, with recent speculation questioning the species status of both *P. exigua* and *P. dyscrita* (see Hart *et al.* 2006; Chapters 1 and 2). Recent mitochondrial data (Chapter 2) suggest that *P. exigua* and *P. dyscrita* are separate taxa, each containing two distinct parapatric lineages. Therefore, before inferences on loci under selection within *P. exigua* can be investigated, a genomic approach (Luikart *et al.* 2003) is needed to ascertain genetic structure within and between the *Parvulastra* species in South Africa.

Defining lineages using mitochondrial DNA alone can be misleading (Kai *et al.* 2002; Bensch *et al.* 2006; Gompert *et al.* 2006) as mtDNA may behave quite stochastically in terms of differentiation between sites, even if gene flow is constant and continuous (Irwin 2002). Mitochondrial genomes are not independent estimates of phylogenies but instead represent a gene tree which may not be congruent with a species tree because signals of a species' evolutionary history may be obscured by selective sweeps, demographic history or hybridization (Moore 1995; Barton 2001; Mishmar *et al.* 2003). The AFLP method has been used effectively to distinguish between species and subspecies, and between recently diverged taxa (Douek *et al.* 2002; Ogden and Thorpe 2002; Salvato *et al.* 2002; Carisio *et al.* 2004; but see Despres *et al.* 2003; Sullivan *et al.* 2004; Gompert *et al.* 2006), as well as in phylogeographic studies of marine invertebrates and other taxa (Darling *et al.* 2004; Baus *et al.*

2005; Timmermans *et al.* 2005), and in distinguishing between colour morphs within species (Kai *et al.* 2002). Once distribution of *Parvulastra* lineages is established, the neutral genetic structure of *P. exigua* can be elucidated without biasing the genetic signal.

4.3. Aims

There were three main aims of this study.

Firstly, we aimed to establish the number of distinct lineages within *Parvulastra* in South Africa using a genomic approach to compare to the mitochondrial structure evident in Chapter 3.

Secondly, we aimed to identify the AFLP (nuclear genomic) structure of *P. exigua* around the coast, and compare it to (i) the mitochondrial sequence structure, (ii) the phenotypic cline, or (iii) the expected pattern of isolation by distance for a benthic species with poor dispersal.

Thirdly, we aimed to identify outlier AFLP loci which may be subject to diversifying or stabilising selection within *P. exigua* using Beaumont and Nichols' (1996) approach. The outlier loci phenotype frequencies across populations will be compared to the phylogeographic and phenotypic structure and the influence and strength of these loci on the population structure will be assessed.

4.4. Methods

4.4.1. Sampling and DNA extraction

Parvulastra exigua and *P. dyscrita* samples were collected from 16 sites (Fig. 4.1) around the South African coastline between January and May 2005 as described in Chapter 3.

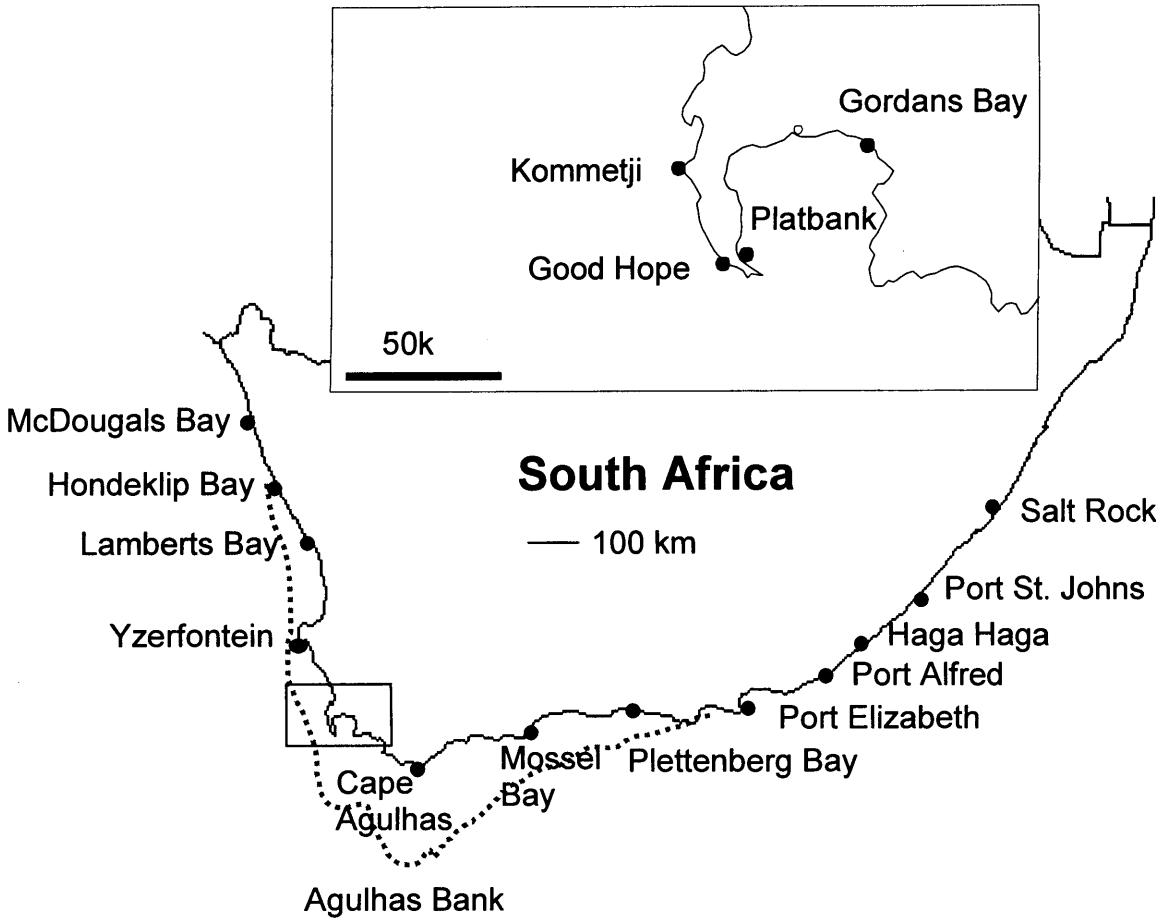

Figure 4.1. Map of *Parvulastra exigua* and *P. dyscrita* sampling sites within South Africa.

DNA samples were extracted between June and August 2005 using the Qiagen DNeasy tissue purification kit (Qiagen), according to the manufacturer's instructions. The extracted DNA was checked for quality and quantity before

starting the AFLP procedure using agarose gel electrophoresis and spectrophotometry. Only high quality DNA was used for the AFLP procedure as the method relies on digestion of the DNA at specific sites reliably producing fragments of consistent length. Degraded DNA was discarded as the fragment sizes would be unreliable.

4.4.2. AFLP

The AFLP procedure was based on Ajmone-Marsan *et al.* (1997) using one pre-selective primer combination and four selective primer combinations, but with the following adaptations to maximise reproducibility (recommended by Baus *et al.* 2005). Firstly, 200 ng of DNA was restricted to ensure complete digestion, instead of 400 ng (Ajmone-Marsan *et al.* 1997). Secondly, three independent pre-selective PCR reactions were performed on the same digested DNA. The products from one of these reactions were visualised on agarose gel to check that the digestion had proceeded efficiently and produced a consistent smear with no banding present. If bands were present in the smear, the sample was discarded. To maximise the probability of amplifying all restriction fragments produced by the digestion reaction, the PCR products of the three pre-selective PCRs were combined, diluted and used in the selective PCR reactions. To check and maintain consistent results, positive and negative control samples were incorporated into every procedure from the selective PCR stage onwards and comparisons between plates were made to assess repeatability.

4.4.2.1. Digestion of genomic DNA and ligation of adaptors

The digestion reaction contained 2.5 μl of One-Phor-All Buffer PLUS; 1.25 μl of DTT (100 mM); 3.1 μl of BSA (0.4 mg/ml); 0.5 μl of *Taq*I (10 U/μl); 200 ng of DNA and adjusted to 25 μl final volume with ddH2O. This was incubated for 1 h at 65°C before the following solution was added: 10.52 μl ddH_2O; 1.5 μl of One-Phor-All Buffer PLUS; 0.75 μl of DTT (100 mM); 1.88 μl of BSA (0.4 mg/ml); 0.34

μl of *Eco*R1 at 15 U/μl. This 40 μl reaction volume was incubated for a further 1 h at 37°C immediately prior to ligation of adaptors. The ligation reaction contained: 4.15 μl of ddH2O; 1 μl of *Eco*RI adaptors stock (5 pmol/μl; 2.5 pmol/μl of each EcoRI adaptor); 1 μl of *Taq*I adaptors stock (50 pmol/μl) 25 pmol/μl of each *Taq*I adaptor; 0.1 μl of ATP (100 mM); 1 μl of One-Phor-All Buffer PLUS; 0.5 μl of DTT (100 mM); 1.25 μl of BSA (0.4 mg/μl) and 1 μl of DNA ligase (1 /μl) total volume 10 μl. This mix was added to the 40 μl of digested DNA and incubated at 37°C for 3 h. This mix was then diluted 1:10 with a low TE buffer (1 ml Tris-HCL 1M pH 7.5, 20 ml 0.5 EDTA pH 8.0, up to 100 ml with ddH_2O) 1:100.

4.4.2.2. Pre-selective PCR

The reactions contained: 19.3 μl of ddH_2O; 5 μl of Invitrogen *Taq* polymerase buffer; 1.5 μl $MgCl_2$ (50 mM); 4 μl dNTPs (10 mM); 7.5 μl of E01 (10 ng/μl) (Pre-selective *Eco*RI primer E01 5'...GAC TGC GTA CCA ATT CA...3'); 7.5 μl of T01 or T02 (10 ng/μl) (Pre-selective *Taq*I primer T02 5'...GAT GAG TCC TGA CCG AC...3'); 0.2 μl of Invitrogen *Taq* polymerase 5 u/μl) and 5 μl of diluted template DNA in a total volume of 50 μl (Ajmone-Marsan *et al.* 1997). PCRs were performed in a Perkin Elmer thermal cycler at 72°C for 1 - 2 min for 1 cycle; 94°C for 30 s, 56°C for 1 min and 72°C for 1 min for 30 cycles; 72°C for 10 min for 1 cycle. The PCR products were then checked on a 1.25% agarose gel to see a smear between 100 and 1000 bp. The pre-amplified template was then diluted 20 fold with a low TE buffer.

4.4.2.3. Selective PCR and Primer combinations

Reactions contained: 7.3 μl of ddH_2O; 2 μl of Invitrogen *Taq* polymerase Buffer; 0.6 μl $MgCl_2$ (50 mM); 1.6 μl dNTPs (10 mM); 0.5 μl of EcoRI primer (labelled with 6 FAM), (10 ng/μl); 3 μl of unlabelled *Taq*I primer (10 ng/μl); 0.08 μl of

Invitrogen Taq 5 U/μl); 5 μl of the diluted pre-amplified template in a total volume of 20 μl (Ajmone-Marsan *et al.* 1997). The PCR conditions were as follows: Initial denaturation at 94°C for 2 min for 1 Cycle. The PCR cycle was repeated 36 times using the following conditions: denaturation at 94°C for 30 s; annealing at 65°C for 30 s (this annealing temperature was then reduced by 0.7°C each cycle to 56°C (13 cycles) and thereafter kept constant until the completion of the PCR run (a remaining 23 cycles)); extension at 72°C for 1 min. After the cycle completion, the final extension was at 72°C for 10 min for 1 cycle. Four primer combinations were used for the selective PCRs (Table 4.1, (Ajmone-Marsan *et al.* 1997).

Primer Pair	EcoR1 primers (labelled)	Taq1 primers (unlabelled)
1	E32 (5' GAC TGC GTA CCA ATT CAA C 3')	T51 (5' GAT GAG TCC TGA CCG ACC A 3')
2	E33 (5' GAC TGC GTA CCA ATT CAA G 3')	T51
3	E38 (5' GAC TGC GTA CCA ATT CAC T 3')	T51
4	E38	T48 (5' GAT GAC TCC TGA CCG ACA C 3')

Table 4.1. Primer combinations for the selective PCR amplifications in the AFLP procedure.

4.4.2.4. Detection of AFLP bands

One μl of the labelled amplification product was mixed with 0.5 μl of ROX 500 size standard (Applied Biosystems) and 10 μl of formamide (Applied Biosystems). The products were denatured at 94°C for 2 min and then the fragments were separated using an ABI prism 3100 genetic analyzer. The samples were analyzed using GENESCAN VS. 2.0 (Applied Biosystems) and the size standard peaks were checked for accuracy. The fragments were visualised with GENOTYPER VS. 3.6 (Applied Biosystems). The fragments for each sample were scored automatically by using a function in GENOTYPER which scores all fragments between specified size and intensity ranges. The size and intensity parameters were set according to the maximum accuracy and repeatability of the control samples, and were adjusted for each primer pair. The accuracy of the AFLP method and the repeatability was assessed using the positive and negative

controls on each plate. Primer pair three, which was inconsistent in its repeatability, was discarded.

4.4.3. Data analysis

4.4.3.1. Evolutionary lineage identification of *Parvulastra* in South Africa

Genetic diversity was calculated as the percentage of polymorphic loci in each primer pair, and in the whole dataset, for all lineages. Average gene diversity over loci (H_e) and the standard deviations were calculated in ARLEQUIN VS. 3.11. Principal Coordinates Analysis (PCA) was conducted on the whole dataset to visualise how the AFLP data clustered, and to ascertain if the mitochondrial lineages identified in Chapters 2 and 3 (*P. exigua* east and *P. exigua* west, *P. dyscrita* and Kommetjie) were also distinguishable using AFLP data. Firstly, a Jaccard's similarity matrix was computed using the software MVSP VS. 3.1 (Kovach Computing Services 1999, http://www.kovcomp.com/msvp/index.html). This was converted into a distance matrix in EXCEL using the equation: '1 - similarity indices', resulting in the distance matrix, which was used to compute the PCA using the EXCEL 'add in' software GENALEX (Peakall and Smouse 2006). A Jaccard's matrix was used since it is most suitable for analysing dominant markers because it takes into account shared presence of bands only, and not shared absence of bands between individuals (Lowe *et al.* 2000). Generally the only reason that two individuals share a band is because they both have the correct restriction sites to create the band, and therefore contain the same genetic information at that locus. There are however many potential reasons why individuals may share the absence of a band. For example, one individual may lack the *Taq*1 restriction site and the other individual may lack the *Eco*R1 restriction site due to an Indel, or recombination etc. This means that these individuals do not possess the same genetic information, and therefore accounting for shared presence of bands only makes the analyses more accurate (Lowe *et al.* 2000).

Analyses of molecular variance (Excoffier *et al.* 1992: AMOVA, conducted in ARLEQUIN vs. 3, Excoffier *et al.* 2005*:* http://cmpg.unibe.ch/software/arlequin3 using 1000 permutations) were used to infer how the genetic variation was apportioned using different combinations of pre-defined lineage groupings. Three groups were identified from the PCA and AMOVA analyses, *P. dyscrita*, *P. exigua* Kommetjie and *P. exigua*. The three separate group datasets were divided up and each lineage was treated as a separate entity. A further PCA was conducted on the *P. dyscrita* populations to assess if there was any genetic clustering within the *P. dyscrita* group. No further PCAs were performed using the Kommetjie lineage as this group contained a small sample size, had very low genetic diversity (as calculated from the percentage polymorphism and gene diversity) and was collected from only one sampling location. Once the groups had been defined within *Parvulastra* in South Africa, no further analyses was conducted using the *P. dyscrita* and Kommetjie samples.

4.4.3.2. Parvulastra exigua population genetic structure

Pairwise genetic differentiation was assessed between populations and across all populations by calculating Θ^B (a Bayesian analogue of Fst) using the software HICKORY vs. 1.0.4. (Holsinger *et al.* 2002). HICKORY uses a hierarchical Bayesian approach that does not assume any prior knowledge of the degree of inbreeding or other genetic structure within populations and is therefore useful for assessing genetic variation for dominant markers. The f free model option in HICKORY was selected for the calculation of Θ^B because it does not attempt to estimate F_{IS}, as this has been shown to be unreliable if using dominant marker data (Holsinger and Wallace 2004). To identify any genetic differentiation between groups of populations, the genetic variation accounted for when the data were divided up into different geographical groups was assessed using AMOVA. A Mantel test, conducted in ARLEQUIN vs. 3.01 and plotted in EXCEL, and a Partial Mantel test, conducted in ARLEQUIN VS. 3.11 (both with 1000 permutations), were performed to

determine whether the genetic structure among populations could be accounted for by geographic distance. The Θ^B pairwise genetic distance matrix, generated in HICKORY, and a matrix of geographic coastal distances were used to compute the Mantel correlation coefficient.

4.4.3.3. Identification of outlier loci, potentially under selection within *Parvulastra exigua*

To identify loci that may be under divergent (or stabilising) selection, the approach of Beaumont and Nichols (1996) was implemented using the program DFdist (© 2005 Beaumont; program distributed by the author http://www.rubic.rdg.ac.uk/cgi-bin/MarkBeaumont/dirlist1.cgi modified from Beaumont and Nichols 1996). Similar approaches to identifying loci under selection have previously been explored (Lewontin and Krakauer 1973; Nei and Maruyama 1975; Bowcock *et al.* 1991) but the approach taken by Beaumont and Nichols (1996) differs since it considers the relationship of Fst to heterozygosity, rather than gene frequency (Bowcock *et al.* 1991). This enables unlinked or loosely linked loci to be treated as independent units and therefore Beaumont and Nichols' (1996) approach can be applied to markers such as microsatellites, AFLP, RFLP or allozymes. The Beaumont and Nichols' (1996) approach also generates a null distribution of the relationship using the coalescent, which allows genealogical variance to be accounted for, and therefore reduces the likelihood of 'false positives' – the major criticism levelled at Lewontin and Krakauer's (1973) original method. The DFdist program was adapted for dominant markers from the programs FDist (Beaumont and Nichols 1996) and Fdist2 (Beaumont and Balding 2004) and uses Zhivotovsky's (1999) approach to calculate the simulated values for heterozygosity and Fst, and it uses the overall heterozygosity of the pooled sample (Mealor and Hild 2006).

The general approach works by identifying loci that show unusually low or high levels of genetic differentiation, which are often assumed to be under selection. Differentiation at these loci can be quantified using the statistic Fst, which is strongly related to the heterozygosity at a locus for a range of population structures and demographic histories (Beaumont and Nichols 1996). A null distribution close to the empirical distribution (Mealor and Hild 2006) for the dataset of Fst versus heterozygosity is generated by coalescent simulations using a simple model of a subdivided population (the symmetrical Island Model, Wright 1951, Beaumont and Nichols 1996), and then upper and lower quantiles of the distribution are estimated in which 95% of the data points are expected to lie (Beaumont and Nichols 1996). The Fst and heterozygosity for each locus in the actual dataset is then plotted onto the simulated distribution. Loci that fall above the 97.5% quantile, which have atypically high levels of differentiation (high Fst's) are interpreted as being under disruptive or diversifying selection. Loci that fall below the 2.5% quantile, which have atypically low levels of differentiation (low Fst) could arise from balancing or stabilising selection which keeps alleles at similar frequencies in different populations (Beaumont and Nichols 1996).

A range of model parameters were tested to explore the effect of the model parameters on the shape of the data distribution and the fit to the data of the simulation of the 95% quantiles. Theta values (the scaled mutation rate theta (Θ) = $N\mu$ where N is the size of each subpopulation and μ is the mutation rate, assuming an infinite allele mutation process (Beaumont and Nichols 1996)) 0.01, 0.1 and 1.0 were tested as well as deme sizes 16 (the same number as the number of populations), 30 and 100 demes (which accounted for many unsampled demes which are known to exist). The number of realizations (50,000 and 100,000) was also tested to explore the effect on the simulations and data distribution. The target average Weir and Cockerham's (1984) Fst required by the DFdist program was generated in two different ways to explore the effect of this value on the fit simulated distribution to the data. Firstly, this value was

calculated as the average theta B (0.2661) across all populations (calculated using HICKORY). Secondly, the trimmed weighted mean F 0.167910 based on the number of actual observations (loci in used in the simulations after loci with a frequency of above 0.98 and below 0.02 were removed), calculated by the Ddatacal program in the DFdist package was used as the target Fst for the DFdist program. However, an average theta B of 0.1 and 0.4 were also tested to observe the effect on the simulations and data distribution.

All bands which appeared only once were removed from the dataset, leaving 307 loci on which the DFdist analyses were performed. The maximum allele frequency allowed by the program was used (0.98), meaning that the DFdist program also excluded all loci with allele frequencies over 0.98 and under 0.02. The final parameters used for the simulations were as follows: 100 demes and 16 populations; theta 0.1; and as recommended by Beaumont and Nichols (1996) 50,000 realizations and '0.25 / 0.25' beta prior. A P-value of 0.95 was used and a smoothing value of 0.04 was used for plotting the graph. Several target average Weir and Cockerham's (1984) Fst, values were tested to explore the fit of the data to the simulated model.

Using Beaumont and Nichols' (1996) method, a few loci are expected to lie outside the 97.5% quantile and below the 2.5% quantile by chance (Mealor and Hild 2006), so identified outlier loci were removed and the phenotype frequency of each of these loci in the 16 populations was plotted in a histogram to detect the presence of any geographic patterns evident in the phenotype frequency of these loci. Mantel tests, conducted in ARLEQUIN vs. 3.01 with 100000 permutations, were performed to determine whether the phenotype frequencies of the outlier loci among populations could be accounted for by geographic structure.

In order to asses the strength of the outlier loci, the remaining loci in the dataset were plotted in a PCA to ascertain if the removal of the outlier loci affected the geographic population structure of *P. exigua* around the coast. PCA analyses were based on Jaccard's distance matrices. Additionally, several AMOVA analyses were conducted to test the apportioning of genetic variation in different geographical groupings with the outlier loci removed.

4.5. Results

4.5.1. Accuracy of AFLP methodology

Primer pair three was found to have consistent contamination in the negative controls and inconsistent band amplification, making its repeatability and accuracy low. Therefore, this primer pair was discarded from further analysis. The three remaining primer pairs differed in the number of bands they amplified, and the fragment size and intensity at which they reliably amplified bands (Table 4.2).

Primer pair	Fragment length scoring parameters	Intensity of bands scored	% Average Repeatability
PP1	> 40 - < 301	76 and over	88.5
PP2	> 60 - < 200	80 and over for one sample set of 96 75 and over for two sample sets of 96	90.5
PP4	> 49 - < 300	91 and over	91
		Total across all three primer pairs	**90**

Table 4.2. Assessment of the accuracy and repeatability of each primer pair, and the whole dataset using Amplified Fragment Length Polymorphism (AFLP).

The factors potentially affecting AFLP repeatability are: (1) Differences in peak intensity due to irregular PCR efficiencies. (2) Slight shifts between two homologous peaks occurring during migration on the gel. (3) The selective primers having more than two selective bases compared to the pre-selective primers because of possible non-specific annealing during amplification. (4) PCR inhibition. (5) Restriction anomalies. (6) Non-target or parasite peaks caused by contamination as the protocol is known to be sensitive to contamination from exogenous DNA (e.g. bacterial) as the amplification is non taxon specific. (7) Using a manual scoring method generates the most errors as it relies on experience and subjective decisions by the scorer (reviewed in Bonin *et al.* 2004). Overall, the average repeatability for the current dataset was 90%. Although lower than previously reported (reviewed in Bonin *et al.* 2004), this accuracy level was deemed an acceptable trade-off between accuracy and data generation for three reasons: (i) the automated scoring method used in

GENOTYPER is more time efficient than manual scoring, enabling a larger number of bands to be generated compared to previous studies (Dodd *et al.* 2002; Douek *et al.* 2002; Giannasi *et al.* 2001; Gompert *et al.* 2006; Parchman *et al.* 2006); (ii) the automated scoring method introduced consistency and removed human error caused by subjectivity when scoring, which previous studies have shown to be the main source of inaccuracies in AFLP, especially when marker specific approaches, such as outlier loci detection are used (reviewed in Bonin *et al.* 2004); and (iii) identification of outlier loci removes all bands with a frequency of more than 0.98 and less than 0.02 in the dataset, which will remove erroneous single bands and monomorphic bands which may have some bands missing. The lower repeatability percentage seen in this dataset was probably caused by (1) some bands being inconsistently amplified at the intensity threshold set due to irregular PCR efficiencies, but due to the automated scoring system, these were not identified. (2) Scoring all the markers instead of selecting markers to be scored. Choosing only bands which were consistently and reliably amplified during the repeatability tests, and removing from the dataset altogether, bands which were consistently unreliably amplified may have improved the repeatability percentage. However, to do this the bands would have to be scored manually, therefore negating the advantages gained by scoring automatically (reasons 1, 2 above). Due to the scoring method used, the low repeatability was unlikely to be caused by (i) slight shifts between homologous peaks occurring during migration, (ii) non-target or parasite peaks caused by contamination, as negative controls for contamination were also examined, (iii) or subjectivity in scoring methods. Bonin *et al.* (2004) report that even if the data is not perfectly repeatable, the extracted biological inferences can be unaffected, and the genotyping errors do not bias the results but just add some noise to the biological signal. As the results of this study are as expected from the apriori knowledge know of the system, it is suspected that the lower repeatability score seen in this study will not affect the biological inference.

4.5.2. Genetic diversity within *Parvulastra*

In total, 205 *P. exigua*, 39 *P. dyscrita* and 10 Kommetjie samples were analysed. For the purposes of this study, the term 'population' will refer to sampling location. Primer pair one produced on average the highest number of bands per population and primer pair two the lowest (Table 4.3). There were large differences in the average percentage polymorphism between the three groups which were also reflected in the average gene diversity results. *Parvulastra exigua* showed an average percent polymorphism across populations of 78%, ranging from 63% in Salt Rock to 84% in Yzerfontein. The degree of polymorphism did not appear to be related to any geographic trends, with the exception of Salt Rock. This population consistently had a lower percentage polymorphism in all three primer pairs, which may have been a reflection of a smaller population size at this location (Chapter 3, Fig. 3.3). *Parvulastra dyscrita* had a much higher level of polymorphism (average 97%) than the other two groups. Additionally, the number of single bands appearing only once in only one individual was much higher in *P. dyscrita*. The Kommetjie samples had a very low percentage polymorphism (average 33%) with only 72 bands present from all primer pairs combined.

Sampling location	N	Total No. scored loci / primer pair				% polymorphic loci / primer pair				Average Gene Diversity over loci (H_e) +/- standard deviation (4.d.p)
		1	2	4	All	1	2	4	All	
P. exigua										
McDougal's Bay	12	61	38	49	148	74	71	78	74	0.051745 +/- 0.0273
Hondeklip Bay	14	64	47	58	169	78	79	72	76	0.062042 +/- 0.0322
Lamberts Bay	13	55	42	52	149	82	76	79	79	0.058379 +/- 0.0305
Yzerfontein	13	56	47	66	169	84	79	88	84	0.066735 +/- 0.0348
Kommetjie	9	66	43	51	160	71	72	67	70	0.058862 +/- 0.0321
Good Hope	15	57	49	55	161	82	86	73	80	0.059892 +/- 0.0309
Platbank	11	62	46	54	162	81	78	78	79	0.064881 +/- 0.0344
Gordans Bay	15	60	42	55	157	80	83	76	80	0.056973 +/- 0.0294
Cape Agulhas	15	67	41	61	169	81	76	84	80	0.060346 +/- 0.0312
Mossel Bay	14	66	41	59	166	76	80	90	82	0.063023 +/- 0.0327
Plettenberg Bay	13	71	46	63	180	80	87	84	83	0.068834 +/- 0.0359
Port Elizabeth	13	66	47	59	172	80	87	81	83	0.069750 +/- 0.0364
Port Alfred	14	74	51	55	180	82	92	76	83	0.065525 +/- 0.0340
Haga Haga	13	60	40	51	151	78	75	82	79	0.055327 +/- 0.0290
Port St. Johns	12	57	33	45	135	81	73	69	75	0.049603 +/- 0.0262
Salt Rock	9	45	27	38	110	64	63	61	63	0.038277 +/- 0.0211
P. exigua all pops	205	172	98	151	421	99	98	99	99	
P. exigua mean pops						78	79	77	78	0.079427 +/- 0.0382
P. dyscrita										
Buffels Bay	12	136	74	116	326	96	97	96	96	0.141437 +/- 0.0737
Gordans Bay	13	131	76	101	308	99	95	98	98	0.133127 +/- 0.0689
Cape Agulhas	14	145	73	110	328	99	96	95	97	0.134059 +/- 0.0689
P. dyscrita All pops	39	204	106	170	480	100	99	99	99	
P. dyscrita mean pops						98	96	96	97	0.137511 +/- 0.0671
Kommetjie										
	10	18	20	34	72	22	30	41	33	0.012731 +/- 0.0073
Total all taxa	254	272	149	251	672	100	99	100	100	

Table 4.3. Amplified Fragment Length Polymorphism (AFLP) number of scored loci, percentage polymorphic loci and average gene diversity over loci (H_e) +/- standard deviation for each primer pair all sampling locations (population) and lineages of *Parvulastra* in South Africa.

4.5.3. AFLP differentiation among *Parvulastra* sea star groups in South Africa

The PCA plot for the whole dataset (Fig. 4.2) indicated the existence of three distinct clusters within *Parvulastra* species in South Africa, precisely corresponding to *P. exigua*, *P. dyscrita* and the Kommetjie samples. The PCA confirmed the low genetic diversity in the Kommetjie lineage as this group was very tightly clustered. Neither *P.*

exigua (Chapter 3) nor *P. dyscrita* (Chapter 2) consisted of two separate groups which corresponded to previously identified mitochondrial clades. On inspection, the *P. exigua* cluster appeared to show some evidence of two groups (coloured green and pink in Fig 4.2), but, in a simple AMOVA test, these were not supported (Table 4.4) However, the AMOVA test for the partitioning of the genetic variation within *P. exigua* into the previously identified mitochondrial clades may have been confounded by the inclusion of the other taxa, and therefore, the partition of the genetic data within *P. exigua* will be explored further in Section 4.5.4 below.

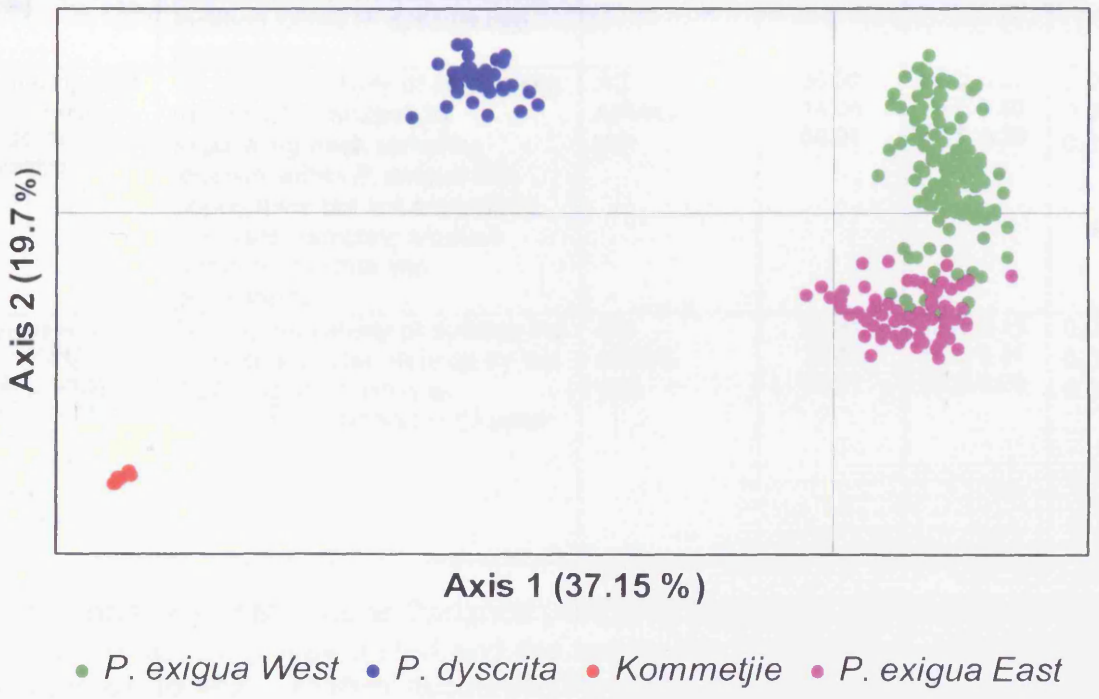

Figure 4.2. Principal Coordinate analysis (PCA) for the AFLP data from all samples of *Parvulastra* in South Africa, with the percentage of genetic variation accounted for in parenthesis after the axis label.

AMOVA indicated that the highest component of genetic variation was explained within populations, irrespective of whether regional groups were, or were not defined. The highest 'among group' variance component was indicated when the groups were defined as separate populations (sampling sites) within each *Parvulastra* group,

including Kommetjie, suggesting that this partitioning of the genetic variation was the most plausible for the data.

Source of variation (group partitioning, Whole dataset: 672 loci)	Reason for choosing groups	Genetic variation accounted for:	% variation	Fixation Indices (2 d.p.)	P values (3 d.p.)
No regional groups	To test the validity of clusters defined by the PCA (Fig. 4.2)	AG APWG	35.06 64.94	FST: 0.35	0.000
[All sampling sites 16 x *P. exigua]* [All sampling sites 3 x *P. dyscrita*] [Kommetjie sample]	To test the validity of partitioning the genetic variation by separating each sampling location within *P. exigua* into populations and each sampling location within *P. dyscrita* into populations.	AG APWG WP	39.02 12.74 48.24	FSC: 0.21 FST: 0.52 FCT: 0.39	0.000 0.000 0.000
[All sampling sites 16 x *P. exigua*] [*P. dyscrita*] [Kommetjie sample]	To test the validity of partitioning the genetic variation by separating each sampling location within *P. exigua* into populations but not separating and each sampling location within *P. dyscrita* into populations.	AG APWG WP	35.92 14.08 50.01	FSC: 0.22 FST: 0.50 FCT: 0.36	0.000 0.000 0.014+-0.004
[*P. exigua* west coast mtDNA sampling sites] [*P. exigua* east coast mtDNA sampling sites] [*P. dyscrita*] [Kommetjie sample]	To test the validity of splitting the *P. exigua* cluster, defined by the PCA (Fig. 4.2), into the haplogroups defined in Chapter 3 (Fig. 3.5)	AG APWG WP	28.31 12.48 59.21	FSC: 0.17 FST: 0.41 FCT: 0.28	0.000 0.000 0.000

Table 4.4. Analysis of Molecular Variance (AMOVA) results for all *Parvulastra* samples indicating the apriori groups tested and the reason for testing them, and showing the percentage of genetic variation accounted for when the *Parvulastra* species data is divided up into different regional groups. AG: Among Groups; APWG: Among Populations Within Groups; WP: Within Populations.

Despite the AMOVA results indicating that there was a higher percentage of genetic variation accounted for 'among groups' when the *P. dyscrita* samples were separated into the three separate sampling locations, the PCA conducted only on the *P. dyscrita* samples (Fig. 4.3) indicated that the three *P. dyscrita* populations were mixed, with no obvious geographic structure.

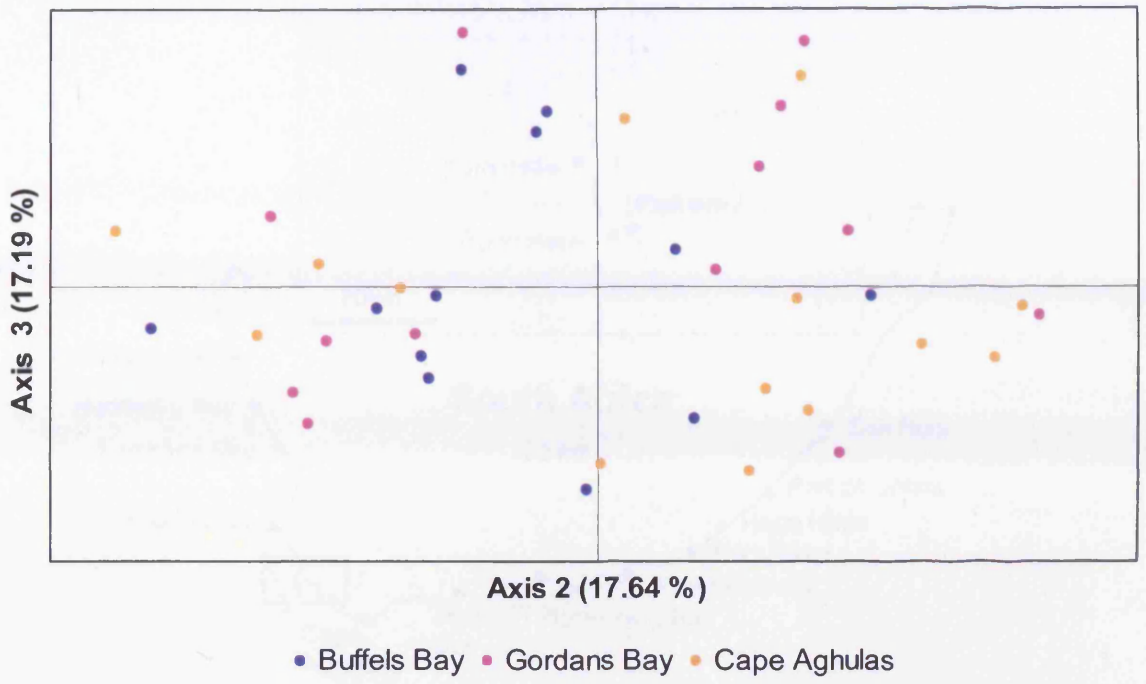

Figure 4.3. Principal Coordinate analysis (PCA) for the AFLP data from the *Parvulastra dyscrita* samples from three sampling locations in South Africa, with the percentage of genetic variation accounted for indicated in parenthesis after the axis label.

AMOVA and PCA strongly suggested that *P. exigua*, *P. dyscrita* and Kommetjie should be classed as separate taxa, therefore for the remainder of this study, no further analyses was performed on *P. dyscrita* and Kommetjie.

4.5.4. Parvulastra exigua population genetic structure

The PCA plot of all *P. exigua* samples (Fig. 4.4) indicated that all individuals within a sampling location clustered together, and that the sampling locations were positioned in relation to each other in accordance with their geographic distances, suggesting an isolation by distance pattern of genetic structure. The PCA plot, by coincidence reflected the shape of the coastline in South Africa, showing clearly that sampling location influenced genetic separation.

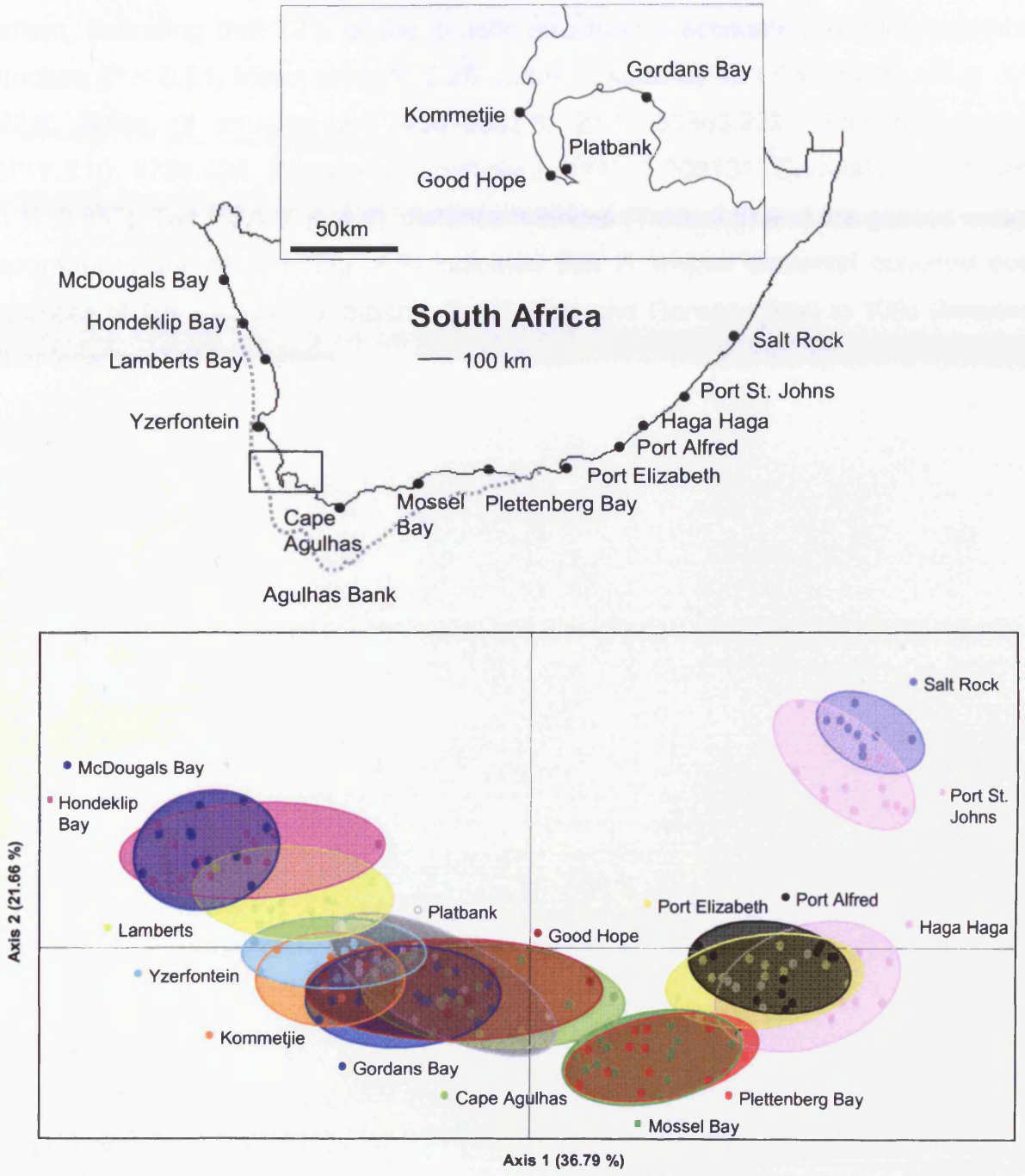

Figure 4.4. Principal Coordinate analysis (PCA) for the AFLP data from the *Parvulastra exigua* samples from 16 sampling locations in South Africa, with the percentage of genetic variation accounted for indicated in parenthesis after the axis label.

The isolation by distance pattern is reflected in the pairwise theta B comparisons (Table 4.5). The Mantel test plot between the geographic distance matrix and the genetic distance matrix (theta B) (Fig 4.5), also indicated an isolation by distance pattern, indicating that 72% of the genetic structure is accounted for by geographic structure ($P < 0.01$, Mean value Y: 0.26, Sums of squares Y: 1.033, Mean value X1: 895.6, Sums of squares X1: 43870532.8, ZY1: 33863.279, Sum of products (SP(Y,X1)): 5736.424, Regression coefficient (bY1): 0.000131, Correlation coefficient (rY1): 0.852). The PCA (Fig. 4.4), distance matrices (Table 4.5) and the genetic verses geographic distance plot (Fig. 4.5) indicated that *P. exigua* dispersal occurred over distances of 10s (between Platbank, Good Hope and Gordans Bay) to 100s (between other locations) of kilometers.

	Mcdoug	Honde	Lam	Yzer	Komm	GoodHo	PltBK	Gord	Agul	Mossel	Plett	PtEliz	Alfred	Haga	PtJohn	Salt
Mcdoug	-	122.5	352	592	719.5	752	754	829	1039	1291	1432.5	1702.5	1861.5	2034.5	2238.5	2551.5
Honde	0.06	-	229.5	469.5	597	629.5	631.5	706.5	916.5	1168.5	1310	1580	1739	1912	2116	2429
Lam	0.20	0.17	-	240	367.5	400	402	477	687	939	1080.5	1350.5	1509.5	1682.5	1886.5	2199.5
Yzer	0.24	0.21	0.16	-	127.5	160	162	237	447	699	840.5	1110.5	1269.5	1442.5	1646.5	1959.5
Komm	0.27	0.23	0.19	0.14	-	32.5	34.5	109.5	319.5	571.5	713	983	1142	1315	1519	1832
GoodHo	0.26	0.25	0.17	0.14	0.11	-	2	77	287	539	680.5	950.5	1109.5	1282.5	1486.5	1799.5
PltBK	0.25	0.22	0.17	0.09	0.11	0.08	-	75	285	537	678.5	948.5	1107.5	1280.5	1484.5	1797.5
Gord	0.31	0.29	0.24	0.20	0.22	0.17	0.13	-	210	462	603.5	873.5	1032.5	1205.5	1409.5	1722.5
Agul	0.27	0.25	0.20	0.15	0.18	0.11	0.09	0.14	-	252	393.5	663.5	822.5	995.5	1199.5	1512.5
Mossel	0.37	0.36	0.30	0.24	0.26	0.23	0.18	0.27	0.21	-	141.5	411.5	570.5	743.5	947.5	1260.5
Plett	0.36	0.35	0.29	0.24	0.26	0.21	0.20	0.27	0.20	0.19	-	270	429	602	806	1119
PtEliz	0.33	0.32	0.27	0.25	0.26	0.21	0.20	0.27	0.21	0.19	0.16	-	159	332	363	676
Alfred	0.36	0.34	0.30	0.28	0.30	0.23	0.23	0.28	0.21	0.22	0.17	0.07	-	173	377	690
Haga	0.37	0.36	0.33	0.30	0.31	0.24	0.25	0.29	0.24	0.20	0.19	0.11	0.14	-	204	517
PtJohn	0.40	0.41	0.37	0.37	0.39	0.35	0.34	0.38	0.36	0.36	0.35	0.27	0.29	0.31	-	313
Salt	0.46	0.44	0.41	0.41	0.44	0.41	0.38	0.43	0.40	0.41	0.40	0.32	0.34	0.32	0.25	-

Table 4.5. Below diagonal are the pairwise genetic differentiations (estimated as θB values in Hickory (Holsinger *et al.* 2002)) between 16 sampling locations of *Parvulastra exigua* in South Africa. Above the diagonal are the approximate geographic distances between sampling locations in kilometres.

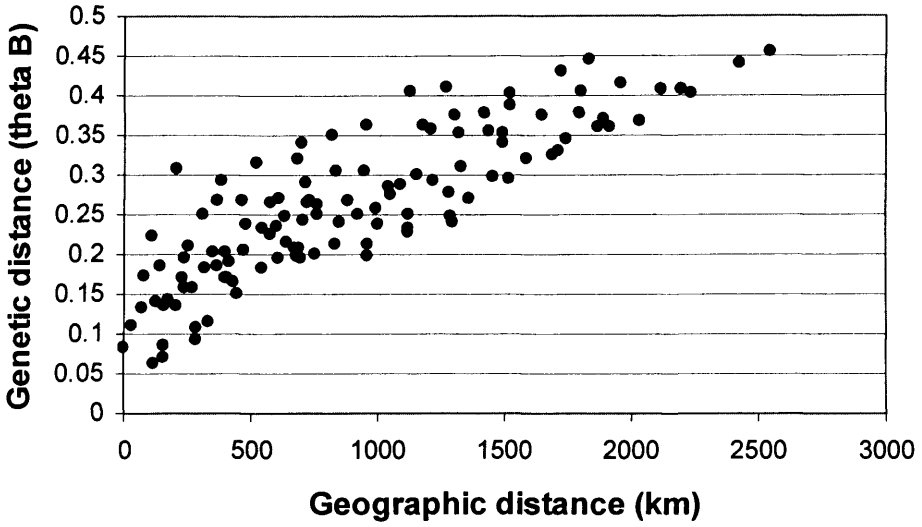

Figure 4.5. Mantel test plot for *Parvulastra exigua* AFLP data indicating an isolation by distance pattern of genetic structure.

To determine if the east and west mtDNA groups (see Chapter 3) differ at nuclear loci, independent of distance, a partial Mantel test was performed using three matrices. A matrix indicating whether samples came from the same or different mtDNA groups (with 0 and 1, respectively) was constructed and this matrix was tested to see if it significantly correlated with AFLP genetic distance, after controlling for geographic distance. This matrix was not significantly correlated with AFLP genetic distance, when accounting for geographic distance ($P = 0.956$) therefore it was concluded that the mtDNA groups did not differ at nuclear loci, suggesting mitochondrial divergence but not nuclear divergence. This result was supported by AMOVA, which indicated that only 11.81 % of the nuclear genetic variation was accounted for among groups when the mtDNA groups were defined (Table 4.6).

Several AMOVA analyses were performed to further examine hierarchical spatial elements of genetic variation within the overall isolation by distance structure, and detect any genetic breaks that may exist within *P. exigua* (Table 4.6).

Source of variation (group partitioning, Whole dataset: 672 loci)	Reason for choosing groups	Genetic variation accounted for:	% variation	Fixation Indices (2 d.p.)	P values (3 d.p.)
No regional groups:	To test validity of assuming any partitions in the genetic data	AG APWG	25.69 74.31	FST: 0.26	0.000
MtDNA groups: [West coast mtDNA sites] [East coast mtDNA sites]	To test validity of partitioning the genetic data according to mitochondrial groups identified in Chapter 3	AG APWG WP	11.81 18.33 69.86	FSC: 0.21 FST: 0.30 FCT: 0.12	0.000 0.000 0.001+-0.001
Phenotypic distribution: [West of Cape Point sites] [East of Cape Point sites]	To test validity of partitioning the genetic data according to east of False Bay (mottled) / west of False Bay (unmottled) phenotype divide	AG APWG WP	9.11 19.97 70.92	FSC: 0.22 FST: 0.29 FCT: 0.09	0.000 0.000 0.003+-0.002
Phenotypic distribution: [West of Cape Point sites] [False Bay sites] [East of False Bay sites]	To test validity of partitioning the genetic data according to east of False Bay (mottled) / False Bay (Intermediate) / west of False Bay (unmottled) phenotype divide	AG APWG WP	9.88 18.49 71.63	FSC: 0.21 FST: 0.28 FCT: 0.10	0.000 0.000 0.000
Biogeographic provinces: [West of Cape Point sites] [East of Cape Point incl. Port Alfred] [East coast sites: Haga Haga - Salt Rock]	To test validity or partitioning the genetic data according to biogeographic provinces (Fig 1.2)	AG APWG WP	10.89 17.53 71.58	FSC: 0.20 FST: 0.29 FCT :0.11	0.000 0.000 0.000
PCA groups: [McDougals Bay - Haga Haga] [Port St. johns - Salt Rock]	To test validity of partitioning the genetic data to identify any further geographic sub-structuring within the isolation by distance pattern identified by the PCA plot (Fig 4.4)	AG APWG WP	17.46 18.65 63.88	FSC: 0.23 FST: 0.36 FCT: 0.17	0.000 0.000 0.009+-0.003
[McDougal's Bay - Cape Agulhas] [Mossel Bay - Haga Haga] [Port St. johns - Salt Rock]		AG APWG WP	16.19 14.37 69.44	FSC: 0.17 FST: 0.31 FCT: 0.16	0.000 0.000 0.000
[McDougal's Bay - Cape Agulhas] [Mossel Bay - Plettenberg Bay] [Port Elizabeth - Haga Haga] [Port St. johns - Salt Rock]		AG APWG WP	16.27 13.66 70.08	FSC: 0.16 FST: 0.30 FCT: 0.16	0.000 0.000 0.000
[McDougal's Bay - Lamberts Bay [Witzerfontein - Cape Agulhas] [Mossel Bay - Plettenberg Bay] [Port Elizabeth - Haga Haga] [Port St. johns - Salt Rock]		AG APWG WP	17.17 11.01 71.82	FSC: 0.13 FST: 0.28 FCT: 0.17	0.000 0.000 0.000
[McDougal's Bay - Hondeklip Bay] [Lamberts Bay] [Witzerfontein - Cape Agulhas] [Mossel Bay - Plettenberg Bay] [Port Elizabeth - Haga Haga] [Port St. johns - Salt Rock]		AG APWG WP	17.62 10.41 71.98	FSC: 0.13 FST: 0.28 FCT: 0.18	0.000 0.000 0.000

Table 4.6. Analysis of molecular variance (AMOVA) results for *Parvulastra exigua* samples only, indicating the apriori groups tested and the reason for testing them, and showing the percentage of genetic variation accounted for when the *samples are* divided up into different regional groups. AG: Among Groups; APWG: Among Populations Within Groups; WP: Within Populations.

The majority of genetic variance was explained within populations regardless of grouping. Other than the clear pattern of isolation by distance among all populations (as indicated by 25.69% among populations when no regional groups were specified), there was only one genetic break-point present, between the most easterly samples (Port St. Johns and Salt Rock) and the remaining populations (17.46% among groups). This suggests that in general there are no strong barriers to gene flow around the coast.

4.5.5. Identification of AFLP outlier loci within *Parvulastra exigua*

DFdist performs coalescent simulations using an island model with 100 islands and assumes an infinite alleles mutational model. To generate the expected distribution, samples of the same size and number as the data are simulated, where each sample is taken from a different island (Beaumont and Nichols 1996). However, the Mantel test indicates that the *P. exigua* AFLP data displays a geographic isolation by distance distribution, and therefore violates the assumption of the Island model simulations. Beaumont and Nichols (1996) investigated the effects of this violation on the expected distributions by investigating the effects of Fst against heterozygosity for three stepping stone models which all represent different patterns of observed isolation by distance. They found that even when there is quite a high level of isolation by distance observed in the models, this does not have a marked effect on the expected distribution. They concluded that weak isolation by distance does not affect the method strongly, especially when the number of sampled populations is large. Additionally, they comment that larger numbers of samples reduce the variability in Fst among loci, and if taken over a wide geographic area, the effects of selection may be more easily detected (Beaumont and Nichols 1996).

No parameter changes greatly affected the distribution of the observed Fst versus heterozygosity data distribution. However, some parameters affected the simulated distribution and the shape and position of the 95% quantiles, which in turn affected the number and selection of data points which were identified as outlier loci. Initial simulations indicated that the results remained robust across the range of theta values

tested (0.01, 0.1 and 1.0 tested) as previously reported (Beaumont and Nichols 1996; Beaumont and Balding 2004; Savolainen *et al.* 2006). The simulation results also remained robust using deme sizes 16, 30 and 100 and using 50,000 and 100,000 realizations.

Using 100 demes and the average theta B of 0.2661 (used as the target average Weir and Cockerham's (1984) Fst), produced a simulated data distribution (Fig. 4.6) which showed a poor fit of the lower quantile of the simulated model to the data and many loci (approx 46 loci out of 220 observations = 21%) fell below the 2.5% quantile, with some falling below zero Fst. Eight loci (approx 3.6%) fell above the 97.5% quantile indicating a better fit to the data, and in total 24% of the loci fell outside the 95% quantiles.

Additionally the graph displayed an unusual shape of the lower quantile and mode (middle line), with a 'step' indicated on the simulated distribution at around a heterozygosity of 0.25. The simulated data reflects the distribution of the actual loci, and the loci responsible for the 'step' and the loci on both sides of the 'step' in the graph (Fig. 4.6) are spread throughout the dataset and are not clustered in any one primer pair. The raw data for these loci was checked and no errors in the data generation or simulation process were detected, indicating that this distribution is genuine. However, the shape of the simulated distribution is affected by the average target W&C Fst parameter of the model entered into DFdist (see below).

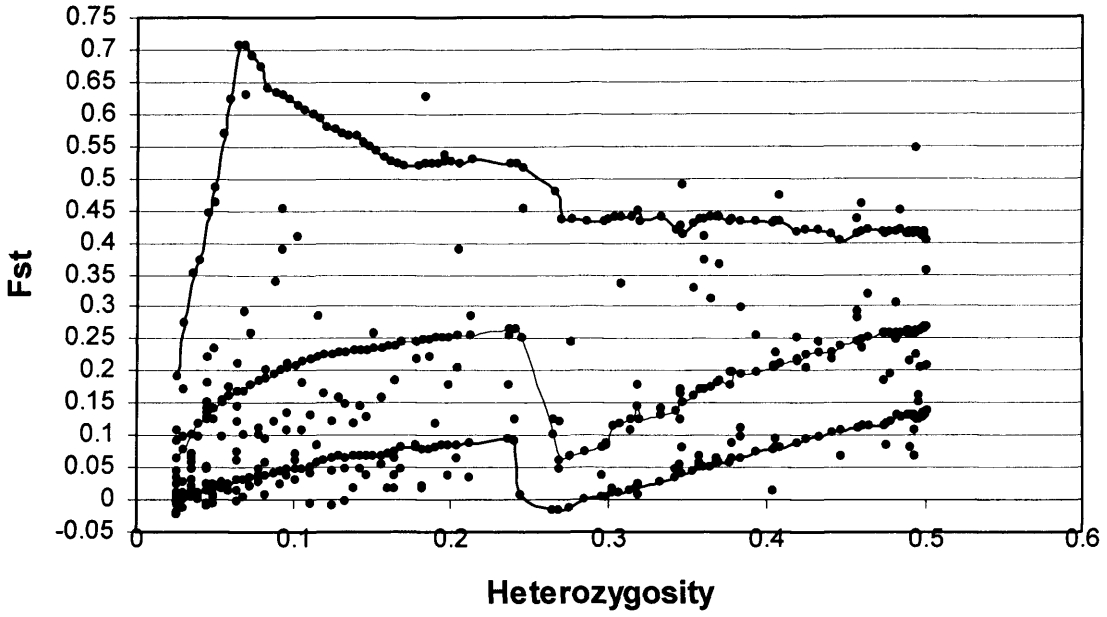

Figure 4.6. DFdist results of Fst for simulated AFLP markers as a function of heterozygosity using the recalculated theta B (0.2661), as calculated by Hickory, as the target average W&C Fst. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle dashed line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites. The scatter points above and below the 97.5% and 2.5% quantiles are outlier loci.

Using 100 demes and the trimmed weighted mean F of 0.167910 (based on the number of actual observations, generated by Ddatacal and following the recommendation of Beaumont in the DFdist manual) produced a simulated distribution (Fig. 4.7) which indicated that a total of approx 18% (20 loci above the 97.5% quantile out of 220 observations, 9%, and 19 loci below the 2.5% quantile out of 220 observations, 8%) of the loci fell outside of the 95% quantiles. This produced a better 'fit' of the lower quantile of the simulated distribution to the data, however this did not entirely remove the unusual 'step' in the shape of the simulated 2.5% quantile and the mode. The upper quantile is lower, indicating that there are more outlying loci above this quantile.

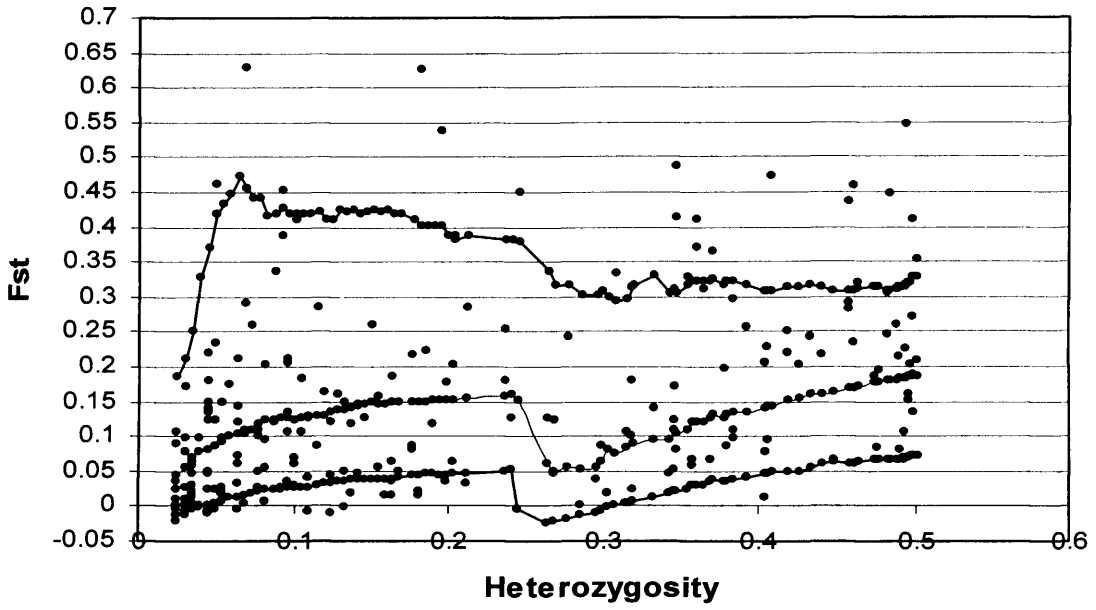

Figure 4.7. DFdist results of Fst for simulated AFLP markers as a function of heterozygosity using the recalculated trimmed weighted mean F (0.167910), as calculated by Ddatacal, as the target average W&C Fst. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle dashed line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites. The scatter points above and below the 97.5% and 2.5% quantiles are outlier loci.

The above simulations demonstrate that the average target W&C Fst dictates the shape of the 95% quantiles, and therefore which, and how many loci fall outside 95% limits. This has been previously noted by Beaumont and Balding (2004) and is further confirmed here when simulations were run with 0.1 as the average target W&C Fst (Fig 4.8) which shows 'flatter' quantiles with a bias towards more high outliers and a less pronounced 'step' as opposed to a target value of 0.4 (Fig 4.9) which shows the reverse.

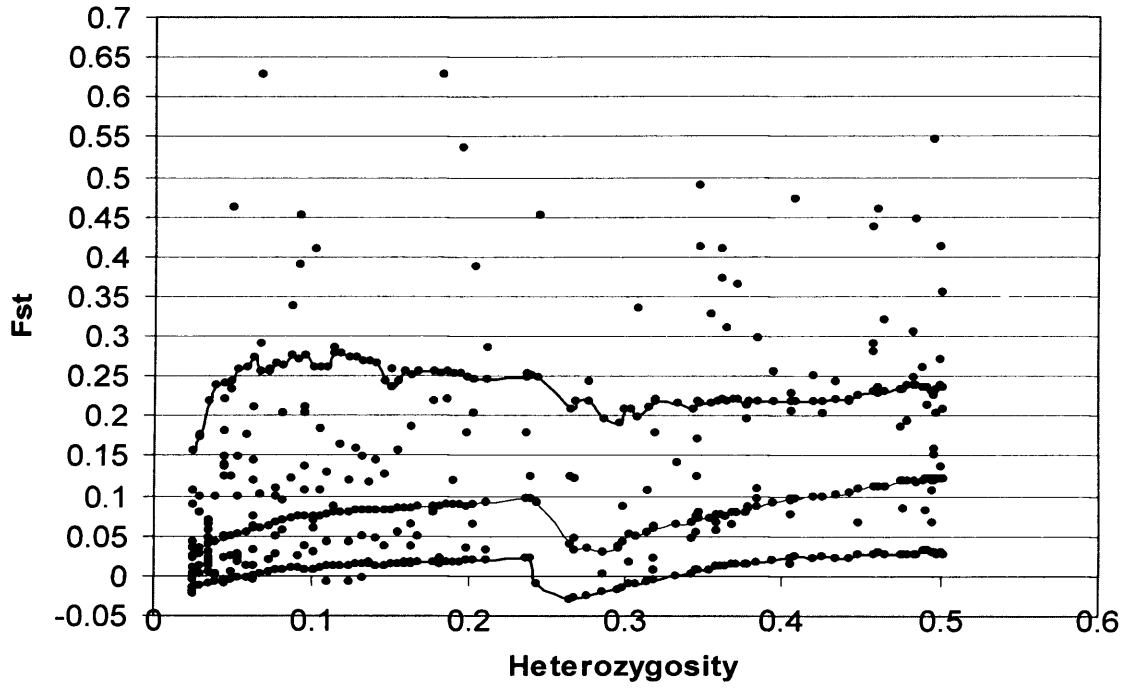

Figure 4.8. DFdist results of Fst for simulated AFLP markers as a function of heterozygosity using an average target W&C Fst of 0.1. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle dashed line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites. The scatter points above and below the 97.5% and 2.5% quantiles are outlier loci.

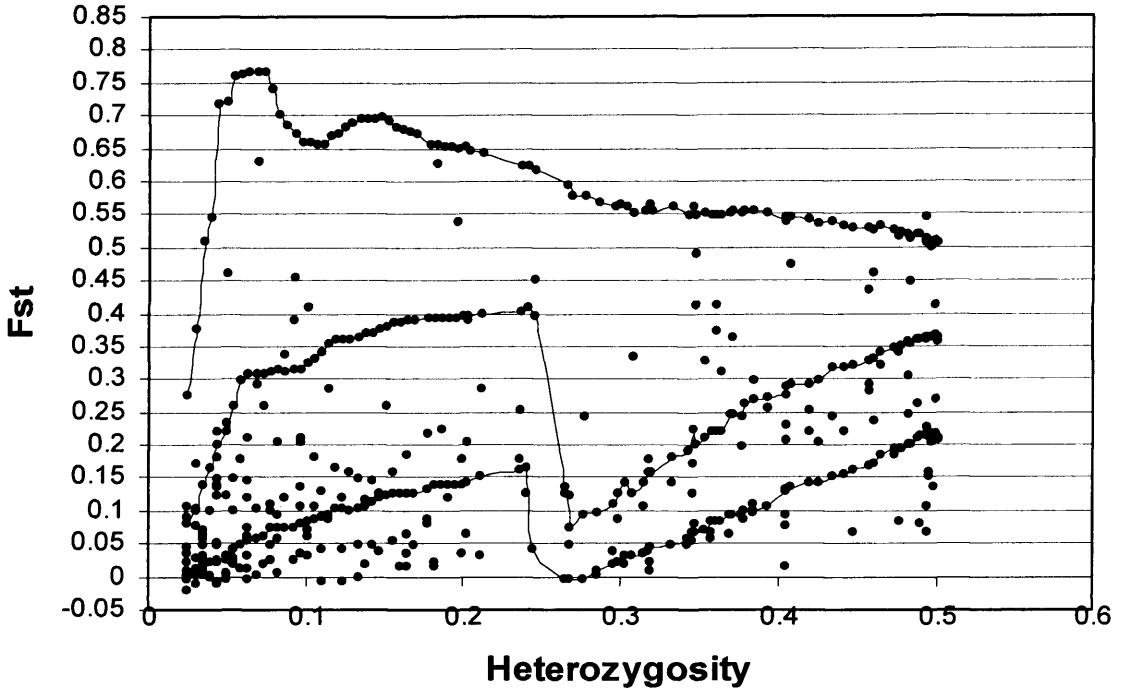

Figure 4.9. DFdist results of Fst for simulated AFLP markers as a function of heterozygosity using an average target W&C Fst of 0.4. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle dashed line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites. The scatter points above and below the 97.5% and 2.5% quantiles are outlier loci.

Beaumont and Nichols (1996) recommend that the DFdist analysis is better viewed as a guide to exploration of the data, with the aim of identifying outlier loci that may be subject to selection. This method has been reported to be reliable, under reasonably realistic conditions, at identifying the majority of loci under adaptive selection (Beaumont and Balding 2004). However, several authors have reported that this method cannot reliably identify loci that are potentially under balancing selection because the lower confidence limit (the 2.5% quantile) often falls close to or lower than zero (Beaumont and Nichols 1996; Beaumont and Balding 2004). However, Hoffmann *et al.* (2006) did report that loci under balancing selection can be detected (at least in one of the regions they investigated) because the lower 95% confidence interval lies

well above zero. The lower 95% confidence interval in this study also fell close to zero when using lower target average W&C Fst values, and did not fall far above zero even at very high target average W&C Fst values. The 'step' shape of the lower quantile, apparent in all simulated distributions to a greater or lesser extent regardless of target average W&C Fst, also caused a portion of the lower quantile to fall below zero. Most studies of this nature did not report proportions of the genome which fell below the 2.5% quantile, potentially representing loci under balancing selection as it is considered an unreliable method (Beaumont and Nichols 1996; Beaumont and Balding 2004). For this reason it was considered more conservative to not interpret the loci below the 2.5% quantile as being potentially under balancing selection, and they were left in the data set for further simulations.

The outlier loci that fell above the 97.5% quantile were removed and the DFdist analysis was repeated to investigate the effect of the removal of these loci on the simulated distribution. This was done using both the recalculated trimmed weighted mean F (0.134198), as calculated by Ddatacal as the target average W&C Fst (Fig. 10), and using the recalculated theta B (0.2183), as calculated by Hickory, as the target average W&C Fst (Fig. 11) with 100 demes in both analyses.

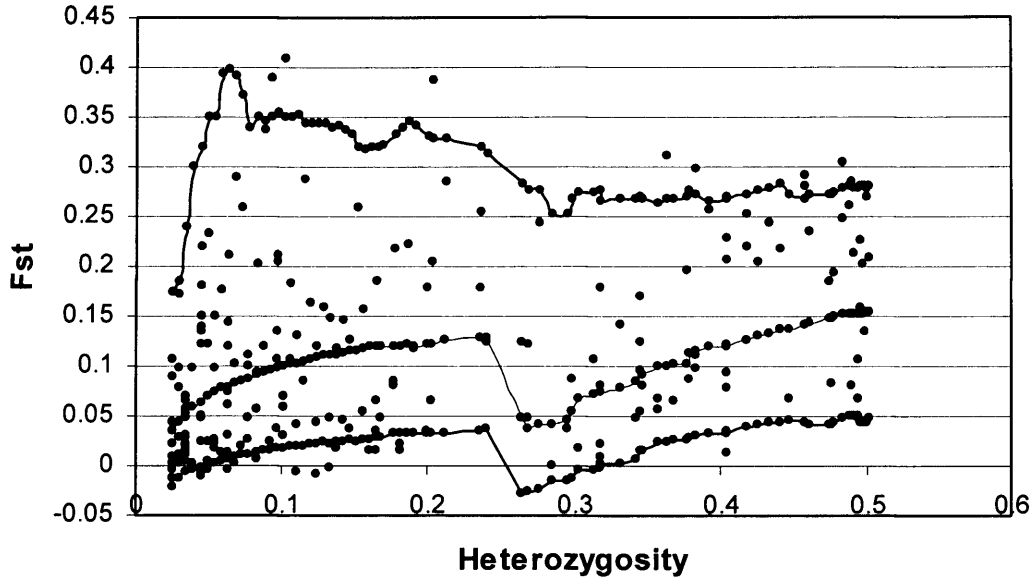

Figure 4.10. Second round DFdist results of Fst for simulated AFLP markers as a function of heterozygosity using the recalculated trimmed weighted mean F (0.134198), as calculated by Ddatacal, as the target average W&C Fst. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites.

A further eight loci out of 199 observations fell above the 97.5% quantile in the 2nd round of DFdist simulations using the recalculated trimmed weighted mean F (0.134198), (calculated by Ddatacal), as the target average W&C Fst, approximately 4%. Between the two round of DFdist simulations using the trimmed weighted mean F as the target average W&C Fst, a total of 26 loci fell above the 97.5% quantile (20 loci, 9% in the 1st round, and eight loci, 4% in the second round) indicating approximately 11.8% (out of 220 observations) of the genome fell above the 97.5% quantile.

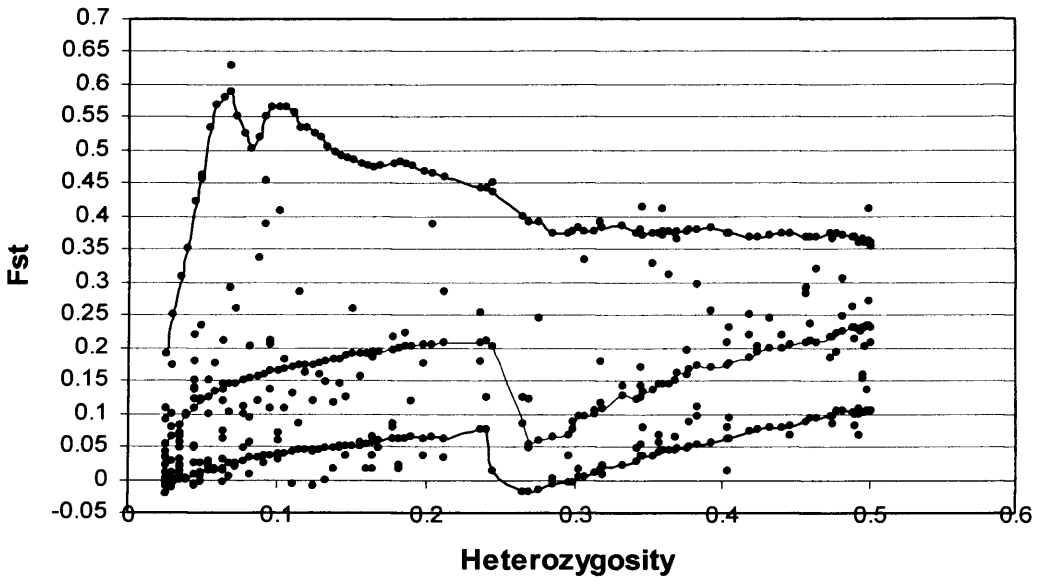

Figure 4.11. Second round DFdist results of Fst for simulated AFLP markers (290 loci) as a function of heterozygosity, using the recalculated theta B (0.2183), as calculated by Hickory, as the target average W&C Fst. The upper and lower lines indicate the simulated 97.5% and 2.5% quantiles, the middle line indicates the median Fst. The scatter points indicate the observed distribution for the values of Fst and heterozygosity for each AFLP locus in *Parvulastra exigua* in South Africa from 16 sampling sites.

Seven loci out of 212 observations fell above the 97.5% quantile in the second round of DFdist simulations using the recalculated trimmed weighted mean F (0.134198), as calculated by Ddatacal, as the target average W&C Fst, approximately 3.3%. Between the two rounds of DFdist simulations using average theta B as the target average W&C Fst, a totals of 15 loci fell above the 97.5% quantile (8 loci, 3.6% in the first round, and 7 loci, 3.3 % in the second round) indicating approximately 6.8 % (out of 220 observations) of the genome fell above the 97.5 % quantile

Loci that fall above the 97.5% quantile could be interpreted as loci (or loci linked to parts of the genome) which are subject to diverging selection. However, some loci are suspected to fall outside the 95% simulated envelope by chance. Therefore it is difficult to determine which of the outlier loci are subject to diverging selection, and which would fall outside the 95% distribution by chance. To examine this, the phenotype frequencies of the outlier loci which fell above the 97.5% quantile in both

rounds of the DFdist analyses for both the trimmed mean F (Fig.12) and the theta b mean (Fig. 13) as the target average W&C Fst values were plotted in histograms for each of the 16 sampling locations around the coast.

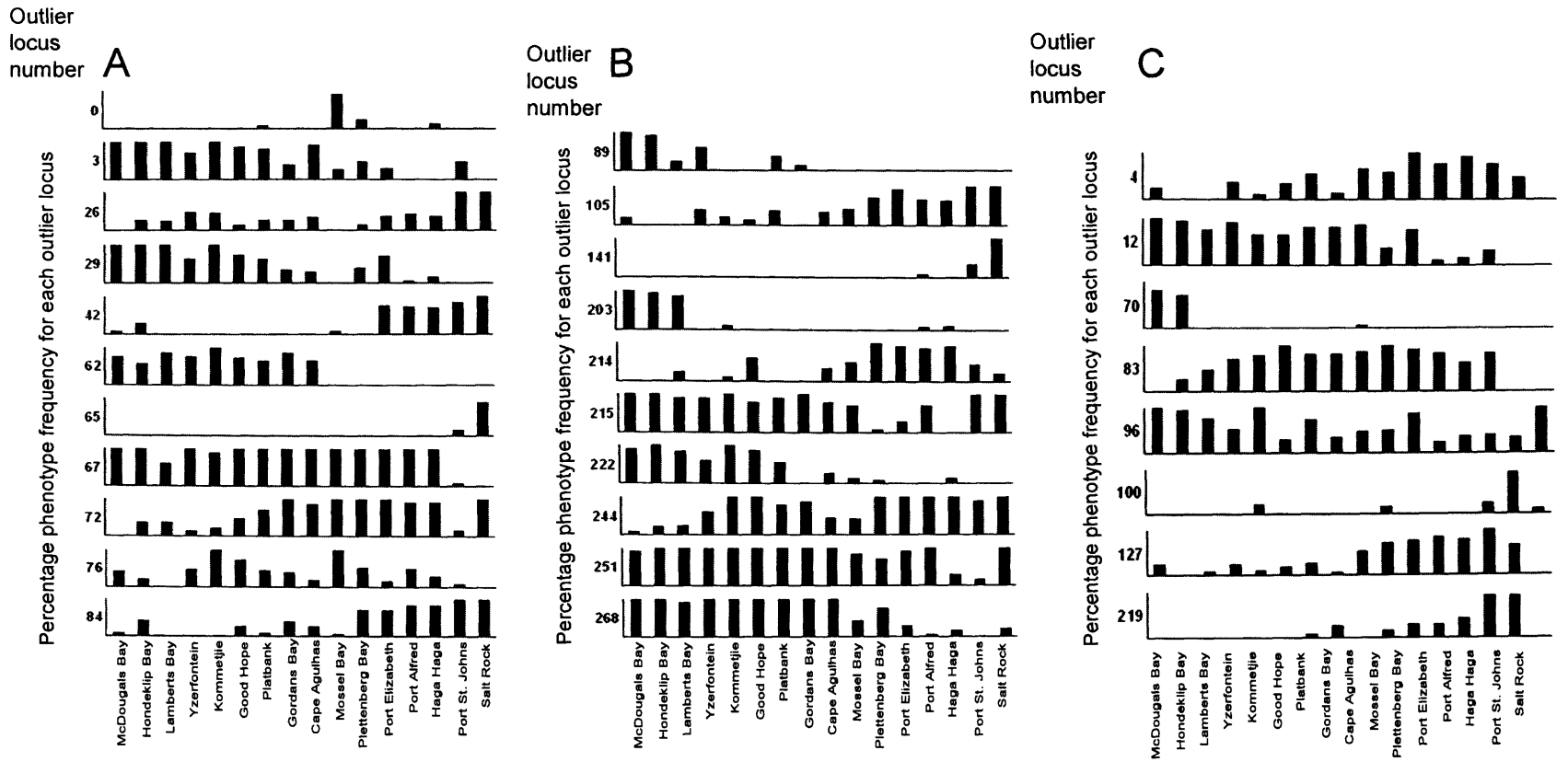

Figure 4.12. DFdist 1st (A and B) and 2nd round (C) outlier loci phenotype frequency histograms for each sampling site, using the trimmed weighted mean F as the target average W&C Fst.

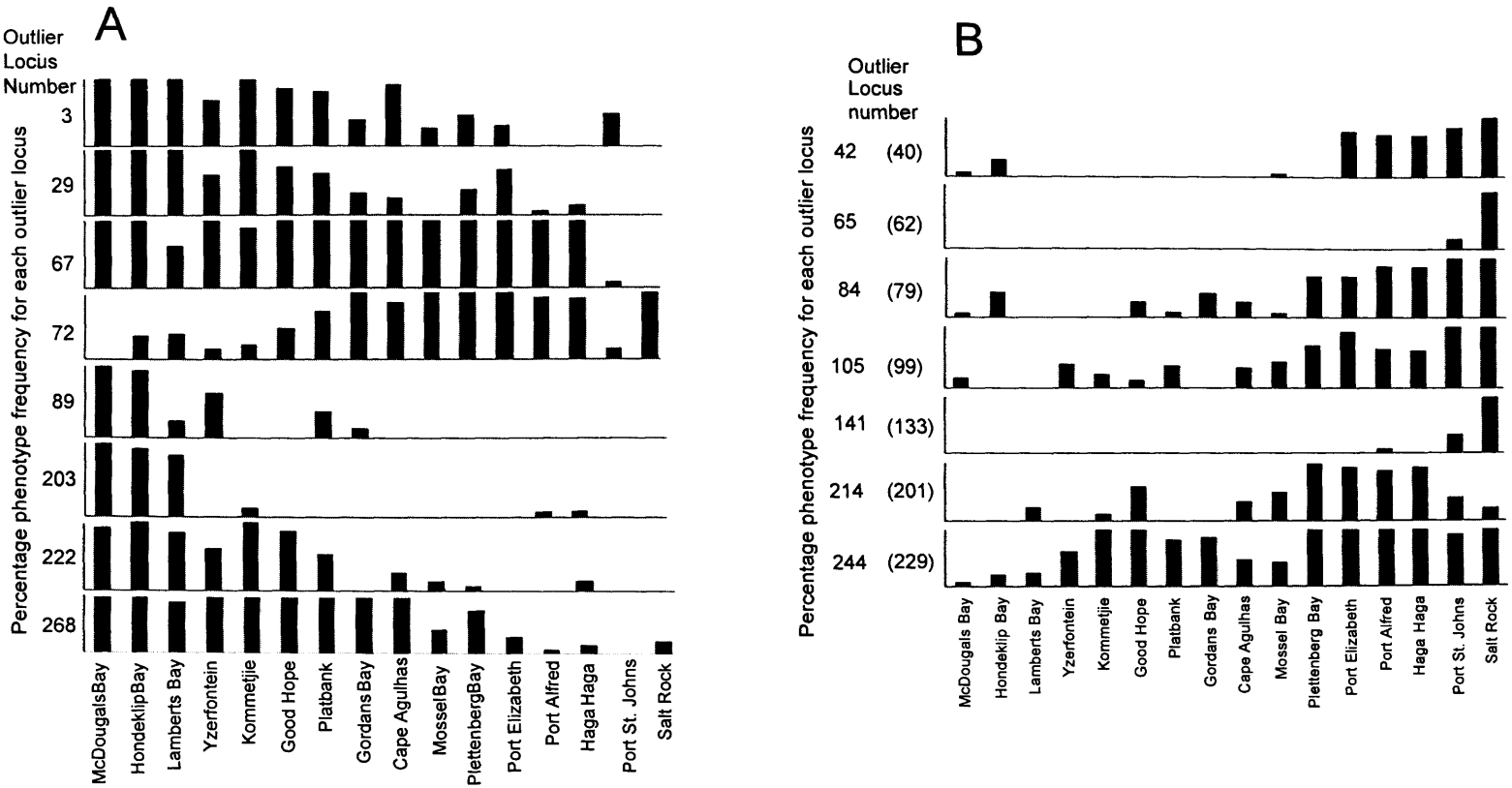

Figure 4.13. DFdist 1st (A) and 2nd round (B) outlier loci phenotype frequency histograms for each sampling site, using the average Theta B as the target average W&C Fst. Outlier locus numbers in brackets on the DFdist second round simulation histogram (B) represent outlier locus numbers that have been re-numbered for the calculations of the 2nd round, and numbers next to these (without brackets) correspond to the numbers of the outlier locus in the histograms when using the trimmed weighted mean F as the target average W&C Fst for comparison.

The trimmed mean F analyses identified more outlier loci which fell above the 97.5% quantile than the DFdist analyses when using the theta B mean. All of the theta B loci which were identified were also identified using the trimmed mean F, indicating that using the theta B mean may provide a more conservative estimate of outlier loci, and that the loci which fell above the quantile with both methods can be cautiously interpreted as outlier loci which may be subject to diverging selection (or linked to parts of the genome that may be under diverging selection).

Eight outlier loci were identified in the 1st round of the DFdist simulations using the theta B mean and seven loci identified in the 2nd round. The phenotype frequencies for each outlier loci identified above the 97.5% quantile, in each sampling location are shown in Fig. 4.13 and the outlier loci details are shown in Table 4.7. To examine the geographic distribution of these outlier loci, mantels test were conducted on the pairwise difference in phenotype frequency between the sampling locations versus the geographic distance between the locations (results shown in Table 4.7).

Outlier loci above the 97.5% quantile					Mantels test	
Outlier locus No.	Primer Pair	Size	Heterozygosity	Fst	% of pairwise difference in phenotype frequency determined by geographic distance	P value (3 d.p)
3	1	46.5	0.46	0.44	40	0.000
29	1	113	0.41	0.47	33	0.000
67	1	164	0.46	0.46	18	0.005
72	1	174	0.48	0.45	9	0.012
89	1	199	0.19	0.54	14	0.010
203	2	199	0.18	0.63	13	0.013
222	4	78.4	0.35	0.49	30	0.000
268	4	168	0.49	0.55	56	0.000
42 (40)	1	127.5	0.22	0.45	51	0.000
65 (62)	1	159	0.05	0.46	18	0.004
84 (79)	1	193.6	0.36	0.41	49	0.000
105 (99)	1	231.7	0.36	0.37	56	0.000
141 (133)	2	77.56	0.07	0.63	22	0.000
214 (201)	4	65.37	0.35	0.41	7	0.017
244 (229)	4	119.4	0.50	0.41	12	0.008

Table 4.7. Outlier loci which fell above the 97.5% quantile in the DFdist simulations for *Parvulastra exigua* in South Africa and are potentially under diverging selection.

All the outlier loci had phenotype frequencies which were significantly correlated with geographic distance. Six out of the eight diverging loci identified in the 1st round of DFdist simulations indicated directionally selective phenotype frequencies, with most displaying high, or fixed frequencies in the west coast populations, with the phenotype frequencies decreasing in more easterly samples. A further seven outlier loci were identified above the 97.5% quantile in the 2nd round of DFdist simulations. The phenotype frequencies in each sampling location for six of the seven loci identified above the 97.5% quantile in the 2nd round also indicated a strong pattern of directional selection with high phenotype frequencies in the east coast populations decreasing in the west coast populations.

There is no specific sampling location where there is an abrupt change in phenotype frequency between the west and the east coast for all identified loci. Instead, the largest change in phenotype frequency is between Lamberts Bay and Kommetjie for loci 89 and 203, and between Platbank and Mossel Bay for the remaining loci with a west to east trend (29, 222, 268). Most of the loci displaying an east to west trend indicated the largest phenotype frequency change between Mossel Bay and Gordans Bay (loci 40, 79, 99, 201). Locus 72 had high frequencies in the east (except for Port St. Johns) decreasing in a westerly direction with the largest change in phenotype frequency between Platbank and Kommetjie. Loci 62 and 133 have an almost fixed phenotype frequency in Salt Rock, a small population, and lower frequencies in Port St. Johns, then very low or absent frequencies in all other populations. Locus 229 displays a bi-modal pattern of phenotype frequency, with high frequencies from Salt Rock to Plettenberg Bay, then lower frequencies at Mossel Bay and Cape Agulhas, then again rising to high frequencies at Good Hope and Kommetjie, before tailing off westwards. These results indicate a change the selective forces between the east and west coasts, with the change occurring somewhere on the south coast approximately between Mossel Bay and Cape Agulhas.

4.5.6. *Parvulastra exigua* population genetic structure and phylogeography with outlier loci removed

A PCA plot (Fig 4.14) was constructed using a dataset after the removal of all of the outlier loci which fell above 97.5% quantile (Table 4.7) which are potentially under diverging selection. The removal of the outlier loci had no major effect on the isolation by distance pattern.

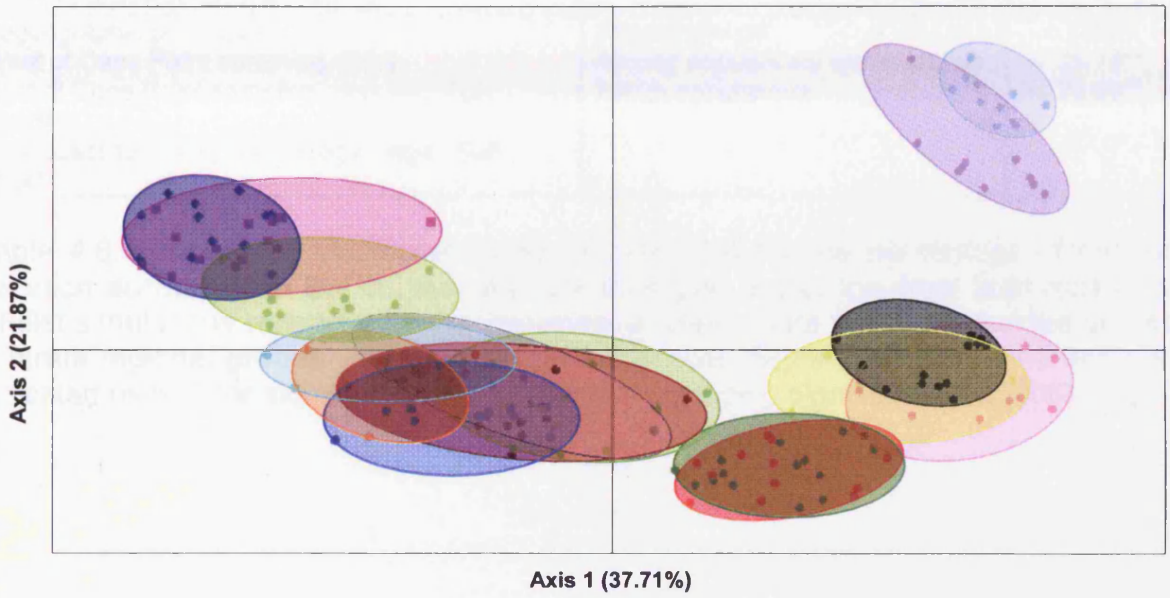

Figure 4.14. Principal Coordinate analysis (PCA) for the AFLP data from the *Parvulastra exigua* samples from 16 separate sampling locations in South Africa, with all diverging loci removed from the dataset, showing axis 1 and 2, with the percentage of genetic variation accounted for indicated in parenthesis after the axis label.

With the diverging outlier loci removed, several AMOVAs were carried out to assess the strength of the outlier loci and measure the effect of their removal on the apportioning of genetic variation when the data were divided into different geographical groups (Table 4.8). The similar values for the AMOVA conducted with (Table 4.6) and without (Table 4.8) the outlier loci, indicated no apparent effect on the apportioning of the genetic variation or genetic structure. The largest amount of genetic variation was still accounted for 'within populations' regardless of geographical grouping.

Source of variation (*P. exigua* dataset)	**Genetic variation accounted for:**	**% variation Diverging loci removed**
No regional groups:	Among populations Within populations	22.20** 77.80
MtDNA groups: [West coast MtDNA sampling sites] [East coast MtDNA sampling sites]	Among groups Among populations within groups Within populations	7.79** 17.48** 74.73**
Phenotypic distribution: [West of Cape Point sampling sites] [East of Cape Point sampling sites]	Among groups Among populations within groups Within populations	5.23** 19.01** 75.76**
Phenotypic distribution: [West of Cape Point sampling sites] [False bay sampling sites] [East of False Bay sampling sites]	Among groups Among populations within groups Within populations	6.40** 17.61** 75.98**
Biogeographic provinces: [West of Cape Point sampling sites] [East of Cape Point sampling sites incl. Port Alfred] [East coast sampling sites: Haga Haga - Salt Rock]	Among groups Among populations within groups Within populations	7.33** 16.79** 75.88**

Table 4.8. Analysis of Molecular Variance (AMOVA) for the percentage of genetic variation accounted in the dataset with the diverging outlier loci from both rounds of DFdist simulations removed for the *Parvulastra exigua* data when it is divided up into different regional groups. Fixation indices not shown. Significance of the % variation indicated using * for significant ($P < 0.05$) and ** for highly significant ($P < 0.005$).

4.6. Discussion

4.6.1. Phylogenetic structure within *Parvulastra* in South Africa

Three distinct genetic groupings within the genus *Parvulastra* were identified with the AFLP data, corresponding to *P. exigua* and *P. dyscrita* and the population sampled only from Kommetjie (Chapter 2). The AFLP data did not recapitulate the mitochondrial paraphyletic structure seen in both *P. dyscrita* and *P. exigua*. The former species displayed exceptionally high levels of polymorphism (97%) in comparison to AFLP studies in other marine invertebrates (Barki *et al.* 2000; Douek *et al.* 2002; Baus *et al.* 2005). This, coupled with the apparently panmictic population structure between the three locations where it was sampled (Buffels Bay, Gordans Bay and Cape Agulhas), indicates that a high level of gene flow is maintained within this species, consistent with its planktonic dispersal mechanism (reviewed by Hilbish 1996). In sharp contrast, the separate evolutionary entity evident at Kommetjie features a low level of polymorphism (33%). This confirms the very low level of genetic diversity suggested by the monomorphic mitochondrial DNA haplotype in this lineage (Chapter 3). It is predicted that this is a small, reproductively isolated population, and probably represents a cryptic species, confirming previous nuclear and mtDNA data (see Chapter 2). The low level of polymorphism could indicate that this lineage evolved from a founder event with population expansion from one or a few gravid females. Alternatively, the Kommetjie population could be the remnant of a larger population which experienced a bottleneck caused, for example, by competition or habitat loss. The evolution and phylogeography of the *P. dyscrita* and Kommetjie lineages, and the disparity between the mtDNA and AFLP population genetic structure observed in *P. dyscrita* will be discussed further in Chapter 5. The inclusion of these 'taxa' in this chapter was to (i) elucidate the evolutionary structure of *Parvulastra* in South Africa using a genomics approach, and (ii) to exclude separate evolutionary lineages from further *P. exigua* population analyses.

4.6.2. Gene flow and population structure within *Parvulastra exigua*

Parvulastra exigua showed polymorphism levels (63 - 84%, mean 78%) at the top end of the range seen in other sea stars (48% to 78%, Baus *et al.* 2005), and mid way between other marine invertebrates, which ranged from 56 to 89% in soft corals (Barki *et al.* 2000) and 36 to 47% in sea anemones (Douek *et al.* 2002). The AFLP data showed clear evidence for isolation by distance around the coast of South Africa, with 72% of the genetic variation (Mantel test) accounted for by geographic distance. These data contrast with the mtDNA structure (see Chapter 3), which showed two divergent reciprocally monophyletic haplogroups separated on the south coast. Further discussion of the incongruity of these datasets is detailed in Chapter 5, whereas this chapter focuses on the AFLP genetic structure.

Many authors have investigated the population genetic structure and geographic distribution of *P. exigua* worldwide (Chapters 2 and 3 and references therein) but those studies used mtDNA alone to infer genetic structure, population demography and dispersal. MtDNA is useful when inferring past processes, but the genomic approach using AFLP reflects present day gene flow and as such can highlight very different demographic processes. The current study provides evidence of dispersal distances of *P. exigua* up to a few hundred kilometres around the coasts of southern Africa, with sea stars remaining close to their natal populations and those adjacent to them. This dispersal capability is reflected in the isolation by distance genetic structure evident but does not indicate that long distance dispersal (above a few hundred kilometres) regularly occurs. However, in previous studies long distance rafting has been invoked to explain the very large anomalous global distribution of *P. exigua* (Waters and Roy 2004; Hart *et al.* 2006), despite this species' entirely benthic life cycle.

This study shows that rafting up to a few hundred kilometres on inshore currents may occur in *P. exigua*, and that its benthic life history and dispersal mechanisms may be important determinants of present day gene flow and genetic structure. The genetic

structure does not indicate that biogeographic provinces designated according to temperature gradients around the coast have an affect on dispersal of *P. exigua* as the isolation by distance pattern appears continuous around the coast and does not reflect biogeography boundaries. However, the degree of rafting and the dispersal distances may be correlated with inshore currents patterns, and concurs with the dispersal distance estimates of *P. exigua* within Australia. This study sheds doubt on the validity of rafting as a regular inter-continental dispersal mechanism (Waters and Roy 2004), therefore the wide global geographic distribution of *P. exigua* remains intriguing. The isolation by distance model of gene flow seems to reflect the organization of many coastal marine species (Hellberg *et al.* 2002; Palumbi 2003). In species, such as *P. exigua*, with large continuous distributions relative to their dispersal capability, which inhabit heterogeneous environments (Johnson and Black 1998) an isolation by distance pattern can indicate small scale localized gene flow, and open up the possibility for local adaptation in response to selection pressures (Maier *et al.* 2005).

4.6.3. Selection in *Parvulastra exigua*

Using a genome scan with a total of 307 loci (but 220 loci after the loci with frequencies of above 0.98 had been removed) and a null distribution Fst framework approach, outlier loci were identified of which 6.8% had higher than expected Fst values. The loci themselves are very unlikely to be under divergent selection, however they may be potentially linked to genomic regions that are. The phenotype frequencies for the high Fst outlier loci (or genomic regions that are linked to these) appear to indicate divergent selection between the east and the west coasts (*i.e.* phenotype frequency decreasing or increasing with an easterly or westerly direction). Similar proportions of high Fst outlier loci, interpreted as being under divergent selection (or linked to loci under selection), have been identified in other intraspecific studies using the same (or similar) methods. Wilding *et al.* (2001) found 5% of the genome in *Littorina* to be outlier loci, Mealor and Hild (2006) identified 2.6% and 8.7%

in two species of grass, Achere *et al.* (2005) found 2.5 to 3.3% in Norway Spruce, and Campbell and Bernatchez (2004) found 1.4 to 3.2% in lake whitefish ecotypes.

4.6.4. Divergent selection between the east and the west coasts

The outlier loci with values above the 97.5% quantile displayed phenotype frequencies which reflected the east - west divergence in phenotype and mitochondrial haplogroups, as well as the environmental temperature differences between the ocean basins. Although there is no clear geographic divide in phenotype frequencies between the east and west coasts, and the causative selective agents (e.g. temperature) or the likelihood that the selective agents are linked to phenotype cannot be ascertained from the generality of the phenotype frequencies, the general east / west (or vice versa) decline in phenotype frequencies implies that local adaptation is occurring, or has occurred in the past. It is interesting to note that in the first round of DFdist simulations, the divergent loci identified had a west to east frequency trend and the loci from the second round had an east to west frequency trend. This pattern may indicate that selection is acting more strongly in the west coast populations than the east coast populations.

Despite the apparent directionality of the outlier loci phenotype frequencies, the influence of these outlier loci on the genetic structure of the South African *P. exigua* populations does not appear to be strong enough to promote divergence at all loci throughout the genome between the east and the west coasts as the isolation by distance gene flow model is not disrupted by their removal. This result is not surprising, as selective forces acting on this small number of loci would have to be extremely strong to disrupt the population structure and cause divergence. Remarkably, this situation has been observed in two morphs of the intertidal snail *Littorina saxatilis,* which occur across an environmental gradient (Grahame *et al.* 2006). When all loci were included, the two morphs were divergent from one another and grouped separately in a population structure analysis. However, when the divergent outlier loci were removed the analysis merely implicated spatially-mediated

genetic differentiation, implying strong selective pressure acting on the phenotypes (Grahame *et al.* 2006).

Gene flow (indicated by the isolation by distance pattern) and divergent selection, (indicated by the phenotype frequency patterns) both appear to be operating within *P. exigua* between the east and the west coasts, however, the balance between these two processes is unclear. Wilding *et al.* (2000) considered it probable that the distribution of mitochondrial haplotypes in *L. saxatilis* indicated expansion from different glacial refugia, but haplotype distribution was unrelated to shell phenotype distribution (Wilding *et al.* 2001). The same authors concluded that it was impossible to tell if the small proportion of the genome which appeared to show selection was the result of adaptive phenotypic divergence and active non-allopatric speciation, or differential introgression following secondary contact resulting in homogenization of allele frequencies at all loci except those under selection (Wilding *et al.* 2001; Grahame *et al.* 2006).

The divergence in phenotype frequencies in *P. exigua* could have been caused by '*in situ*' selection driving adaptive differentiation (and potentially phenotypic divergence) despite the homogenizing effect of gene flow (potentially laying the ground for sympatric speciation). The possibility of divergence occurring in the face of gene flow has achieved much attention and many authors suggest that reproductive isolation can develop via pleiotropy when populations experience strong diversifying selection on multiple characters, and sympatric speciation can be the result (Rice and Hostert 1993; Danley *et al.* 2000; Irwin 2002; Jordan *et al.* 2005). Alternatively, the divergence in the outlier phenotype frequencies may have evolved during a period of allopatry between the east and west coasts, during which time ecological specialization may have occurred in the absence of gene flow, followed by secondary contact and the establishment of the isolation by distance model of gene flow (Wilding *et al.* 2001; Grahame *et al.* 2006). The same selective forces that caused the local adaptation and divergence in the outlier loci phenotype frequencies could still be present in the geographic regions that they originally evolved in, therefore these loci

may still be under selection, resulting in the outlier Fst values, and these selective forces have not yet been disrupted by gene flow.

Murray and Hare (2006) investigated whether a 'genome scan' is able to detect locus specific selection in secondary contact zones, using a continuously distributed oyster species in Florida which has a secondary contact zone between the Atlantic and Gulf of Mexico and displays reciprocal monophyly for mtDNA either side of the contact zone (Reeb and Avise 1990). They suggested that genome scans may be unreliable in identifying outlier loci if populations are in secondary contact because a combination of pre-contact differentiation variance and post-contact differential introgression could inflate the Fst variance over neutral expectations, causing the null Fst distribution to not be as robust, therefore making outlier identification unreliable (Bierne *et al.* 2003). However, Murray and Hare (2006) concluded that they found no statistically significant distinction between the Fst distributions simulated under the contrasting models of secondary contact after allopatry, and *in situ* divergence. Therefore, despite the theoretical expectation that secondary contact in clinal populations would distort the Fst neutral model, outlier loci can still be identified under conditions of secondary contact. It therefore seems possible to detect locus specific selection using a genome scan, but it may be difficult to distinguish between selection caused during allopatry, or '*in situ*' selection resulting in divergence at the outlier loci. The isolation by distance pattern seen in the AFLP data could reflect the present introgression of the neutral portion of the genome after secondary contact, but gene flow between the east and west coasts is not extensive enough to dilute the effects of selection, or that selection is strong enough to resist the homogenising effects of gene flow. Considering the reciprocally monophyletic mitochondrial structure in *P. exigua*, secondary contact after a period of allopatry seems a more plausible explanation than selection causing divergence *in situ* resulting in reciprocally monophyletic mitochondrial haplogroups.

4.7. Conclusions

The data presented here confirmed the genetic distinction of *P. exigua* and *P. dyscrita* as separate taxa, and provided further evidence that there is a cryptic species of *Parvulastra* at Kommetjie on the west coast of the Cape Peninsula. Furthermore, the AFLP data indicated that *P. dyscrita* has an exceptionally high level of polymorphism, the Kommetjie lineage has a low level of polymorphism, possibly reflecting a founder event origin, and the polymorphism level in *P. exigua* is at the high end of the range seen in other sea stars (Baus *et al.* 2005). The current study has confirmed that AFLP is a useful tool in population genetic studies of closely related species (Giannasi *et al.* 2001; Ogden and Thorpe 2002). The AFLP data strongly indicated a pattern of isolation by distance between the *P. exigua* populations around the coast, reflecting the benthic life cycle of this species and possible rafting over short to medium distances. The AFLP results suggest that the dispersal ability of *P. exigua* is limited to adjacent sampling locations and provides no evidence for long distance (over a few hundred kilometres) rafting being an important dispersal mechanism. The AFLP structure contrasted with previously identified mtDNA structure which indicated two reciprocally monophyletic groups separated on the south coast, highlighting the importance of using a genomic approach to infer genetic connectivity and separate past processes from present day gene flow patterns. The genome scan approach combined with Beaumont and Nichols' (1996) approach to identify outlier loci potentially under selection revealed that some loci were found to differ from the neutral model of genetic variation, and the phenotype frequencies of these loci were found to display strongly directional trends which corresponded to the divergence seen in phenotype, mitochondrial haplogroups and abiotic variables between the east and the west coast. It is concluded that these loci may have become locally adapted during a period of allopatry between the east and west coasts, and that upon secondary contact, the divergence has persisted due the balance between gene flow and strong selection pressures. Future investigation into the outlier loci would include isolating these loci and sequencing for BLAST identification to investigate gene function and expression.

4.8. References

Achere V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology* **14**, 3191-3201.

Ajmone-Marsan P, Valentini A, Cassandro M, *et al.* (1997) AFLP (TM) markers for DNA fingerprinting in cattle. *Animal Genetics* **28**, 418-426.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* **12**, 1805-1814.

Andres JA, Sanchez-Guillen RA, Rivera AC (2000) Molecular evidence for selection on female color polymorphism in the damselfly *Ischnura graellsii*. *Evolution* **54**, 2156-2161.

Andres JA, Sanchez-Guillen RA, Rivera AC (2002) Evolution of female colour polymorphism in damselflies: testing the hypotheses. *Animal Behaviour* **63**, 677-685.

Barki Y, Douek J, Graur D, Gateno D, Rinkevich B (2000) Polymorphism in soft coral larvae revealed by amplified fragment-length polymorphism (AFLP) markers. *Marine Biology* **136**, 37-41.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551-568.

Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Baus E, Darrock DJ, Bruford MW (2005) Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star *Asterina gibbosa*. *Molecular Ecology* **14**, 3373-3382.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**, 1619-1626.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: Why so few animals? *Molecular Ecology* **14**, 2899-2914.

Bensch S, Irwin DE, Irwin JH, Kvist L, Akesson S (2006) Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus warblers*. *Molecular Ecology* **15**, 161-171.

Bierne N, Daguin C, Bonhomme F, David P, Borsa P (2003) Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology* **12**, 2505-2510.

Bonin A, Bellemain E, Eidesen PB, *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Booth JD, Ovenden JR (2000) Distribution of *Jasus spp.* (*Decapoda : Palinuridae*) *phyllosomas* in southern waters: implications for larval recruitment. *Marine Ecology-Progress Series* **200**, 241-255.

Bowcock AM, Kidd JR, Mountain JL, *et al.* (1991) Drift, admixture, and selection in human-evolution - a study with DNA polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 839-843.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* **21**, 945-956.

Carisio L, Cervella P, Palestrini C, DelPero M, Rolando A (2004) Biogeographical patterns of genetic differentiation in dung beetles of the genus *Trypocopris* (*Coleoptera, Geotrupidae*) inferred from mtDNA and AFLP analyses. *Journal of Biogeography* **31**, 1149-1162.

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54**, 1725-1737.

Darling JA, Reitzel AM, Finnerty JR (2004) Regional population structure of a widely introduced estuarine invertebrate: *Nematostella vectensis* Stephenson in New England. *Molecular Ecology* **13**, 2969-2981.

Degen B, Petit RJ, Kremer A (2001) SGS - Spatial genetic software: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. *Journal of Heredity* **92**, 447-449.

Despres L, Gielly L, Redoutet W, Taberlet P (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution* **27**, 185-196.

Dodd RS, Afzal-Rafii Z, Kashani N, Budrick J (2002) Land barriers and open oceans: effects on gene diversity and population structure in *Avicennia germinans L.* (*Avicenniaceae*). *Molecular Ecology* **11**, 1327-1338.

Douek J, Barki Y, Gateno D, Rinkevich B (2002) Possible cryptic speciation within the sea anemone *Actinia equina* complex detected by AFLP markers. *Zoological Journal of the Linnean Society* **136**, 315-320.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics* **131**, 479-491.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online, Application Note* **2005**, 47-50.

Giannasi N, Thorpe RS, Malhotra A (2001) The use of amplified fragment length polymorphism in determining species trees at fine taxonomic levels: analysis of a medically important snake, *Trimeresurus albolabris*. *Molecular Ecology* **10**, 419-426.

Gillespie RG, Oxford GS (1998) Selection on the color polymorphism in Hawaiian happy-face spiders: Evidence from genetic structure and temporal fluctuations. *Evolution* **52**, 775-783.

Golding B (1992) The prospects for polymorphisms shared between species. *Heredity* **68**, 263-276.

Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology* **15**, 1759-1768.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution* **60**, 268-278.

Grosberg R, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine community ecology.* (eds. Bertness MD, Hay ME, Gaines SD), p. 61–84. Sinauer, Sunderland, Massachusetts.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Hart MW, Keever CC, Dartnall AJ, Byrne M (2006) Morphological and genetic variation indicate cryptic species within Lamarck's little sea star, *Parvulastra (=Patiriella) exigua*. *Biological Bulletin* **210**, 158-167.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.

Hilbish TJ (1996) Population genetics of marine species: The interaction of natural selection and historically differentiated populations. *Journal of Experimental Marine Biology and Ecology* **200**, 67-83.

Hoffman EA, Schueler FW, Jones AG, Blouin MS (2006) An analysis of selection on a colour polymorphism in the northern leopard frog. *Molecular Ecology* **15**, 2627-2641.

Holsinger KE, Lewis PO, Dey DK (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* **11**, 1157-1164.

Holsinger KE, Wallace LE (2004) Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (*Orchidaceae*). *Molecular Ecology* **13**, 887-894.

Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes - implications for the evolution of fundamental niches. *Evolutionary Ecology* **6**, 433-447.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution* **56**, 2383-2394.

Johnson MS, Black R (1998) Effects of habitat on growth and shape of contrasting phenotypes of *Bembicium vittatum* Philippi in the Houtman Abrolhos Islands, Western Australia. *Hydrobiologia* **378**, 95-103.

Jordan MA, Snell HL, Snell HM, Jordan WC (2005) Phenotypic divergence despite high levels of gene flow in Galapagos lava lizards (*Microlophus albemarlensis*). *Molecular Ecology* **14**, 859-867.

Kai Y, Nakayama K, Nakabo T (2002) Genetic differences among three colour morphotypes of the black rockfish, *Sebastes inermis*, inferred from mtDNA and AFLP analyses. *Molecular Ecology* **11**, 2591-2598.

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution* **20**, 893-900.

Kreitman M, Akashi H (1995) Molecular evidence for natural-selection. *Annual Review of Ecology and Systematics* **26**, 403-422.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**, 175-195.

Lowe AJ, Gillies ACM, Wilson J, Dawson IK (2000) Conservation genetics of bush mango from central / west Africa: implications from random amplified polymorphic DNA analysis. *Molecular Ecology* **9**, 831-841.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.

Maier E, Tollrian R, Rinkevich B, Nurnberger B (2005) Isolation by distance in the scleractinian coral *Seriatopora hystrix* from the Red Sea. *Marine Biology* **147**, 1109-1120.

Mealor BA, Hild AL (2006) Potential selection in native grass populations by exotic invasion. *Molecular Ecology* **15**, 2291-2300.

Mishmar D, Ruiz-Pesini E, Golik P, *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 171-176.

Monsen KJ, Blouin MS (2004) Extreme isolation by distance in a montane frog *Rana cascadae*. *Conservation Genetics* **5**, 827-835.

Moore WS (1995) Inferring phylogenies from MtDNA variation - mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**, 718-726.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology* **15**, 4229-4242.

Nei M, Maruyama T (1975) Lewontin-Krakauer test for neutral genes - comment. *Genetics* **80**, 395-395.

Ogden R, Thorpe RS (2002) The usefulness of amplified fragment length polymorphism markers for taxon discrimination across graduated fine evolutionary levels in Caribbean Anolis lizards. *Molecular Ecology* **11**, 437-445.

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology & Evolution* **13**, 502-506.

Oxford GS (2005) Genetic drift within a protected polymorphism: Enigmatic variation in color-morph frequencies in the candy-stripe spider, *Enoplognatha ovata*. *Evolution* **59**, 2170-2184.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, S146-S158.

Palumbi SR (2004) Marine reserves and ocean neighborhoods: The spatial scale of marine populations and their management. *Annual Review of Environment and Resources* **29**, 31-68.

Parchman TL, Benkman CW, Britch SC (2006) Patterns of genetic variation in the adaptive radiation of New World crossbills (*Aves : Loxia*). *Molecular Ecology* **15**, 1873-1887.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Porter AH (2003) A test for deviation from island-model population structure. *Molecular Ecology* **12**, 903-915.

Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species - mitochondrial-DNA in the American oyster, *Crassostrea Virginica*. *Genetics* **124**, 397-406.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation - what have we learned in 40 years. *Evolution* **47**, 1637-1653.

Richman A (2000) Evolution of balanced genetic polymorphism. *Molecular Ecology* **9**, 1953-1963.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12242-12245.

Riginos C, Victor BC (2001) Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 1931-1936.

Salvato P, Battisti A, Concato S, *et al.* (2002) Genetic differentiation in the winter pine processionary moth (*Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers. *Molecular Ecology* **11**, 2435-2444.

Savolainen V, Anstett MC, Lexer C, *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.

Schlotterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753-763.

Schluter D (1996) Ecological causes of adaptive radiation. *American Naturalist* **148**, S40-S64.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution* **16**, 372-380.

Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science* **276**, 1855-1857.

Sponer R, Roy MS (2002) Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (*Echinodermata*) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution* **56**, 1954-1967.

Steinger T, Haldimann P, Leiss KA, Muller-Scharer H (2002) Does natural selection promote population divergence? A comparative analysis of population structure using amplified fragment length polymorphism markers and quantitative traits. *Molecular Ecology* **11**, 2583-2590.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Storz JF, Dubach JM (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution* **58**, 1342-1352.

Sullivan JP, Lavoue S, Arnegard ME, Hopkins CD (2004) AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (*Mormyroidea: Teleostei*). *Evolution* **58**, 825-841.

Timmermans M, Ellers J, Marien J, *et al.* (2005) Genetic structure in *Orchesella cincta* (*Collembola*): strong subdivision of European populations inferred from mtDNA and AFLP markers. *Molecular Ecology* **14**, 2017-2024.

Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Stable genetic polymorphism in heterogeneous environments: balance between asymmetrical dispersal and selection in the acorn barnacle. *Journal of Evolutionary Biology* **19**, 589-599.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.

Waters JM, Roy MS (2004) Out of Africa: The slow train to Australasia. *Systematic Biology* **53**, 18-24.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Wilding CS, Grahame J, Mill PJ (2000) Mitochondrial DNA COI haplotype variation in sibling species of rough periwinkles. *Heredity* **85**, 62-74.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8**, 907-913.

Chapter 5

Discussion

Chapter 5: Discussion

As the first genetic analysis of South African populations of the genus *Parvulastra*, this study has both resolved a number of key problems and identified new questions about the phylogeny and phylogeography of this well studied sea star taxon. For example, this is the first study to address the genetic distinction of *Parvulastra exigua* and *P. dyscrita*, resolving the longstanding controversy surrounding these two taxa. Furthermore, a new highly localised, divergent, and potentially 'at risk' lineage within the *Parvulastra* genus is identified. In terms of South African intertidal biogeography, this is one of only a few studies examining phylogeography and population genetic structure on a large spatial scale in a continuously distributed species. These data were also related to historical abiotic processes and present day biogeographic provinces. This is also the first study to use Amplified Fragment Length Polymorphism (AFLP) and a coalescent simulation approach to identify markers under selection for an echinoderm species. The following sections aim to synthesise the data from Chapters 2 - 4 and highlight gaps in our knowledge and key questions that remain to be answered.

5.1. Kommetjie

Three different genetic approaches (mtDNA and nuclear DNA sequences, and AFLP) all showed divergence between the Kommetjie population and the *P. exigua* and *P. dyscrita* lineages. The presence of a unique four base pair indel in the actin intron and nuclear copies/heteroplasmy in mtDNA sequences suggests that this group is reproductively isolated from other *Parvulastra* lineages. MtDNA and AFLP showed low levels (or an absence of) genetic variation within the divergent Kommetjie group. Maximum likelihood analysis of mtDNA, grouped the Kommetjie lineage within a clade of the genus *Parvulastra* containing *P. vivipara* and *P. parvivipara* (Fig. 5.1) not with *P. exigua* or *P. dyscrita*. The Kommetjie samples had oral gonopores and could be distinguished by their reddish orange irregular colour (1 and 2 in Fig. 5.1), in contrast to morphs of *P. exigua* (3 and 4 in Fig. 5.1). The divergent Kommetjie group has a sympatric distribution to *P. exigua* and appears to have an extremely restricted

geographic range which may indicate that it should receive conservation status, as has been suggested for the relatively recently discovered *P. vivipara* (see Dartnall 1969, cited in Keough and Dartnall 1978) and *P. parvivipara* (see Keough and Dartnall 1978) species in Australasia (Dartnall *et al.* 2003). The current genetic data provide compelling evidence that the Kommetjie lineage should be classed as a new species, *Parvulastra kommetjia* sp. nov. Holotype and paratypes will be submitted to the Natural History Museum, London, and a full species description prepared.

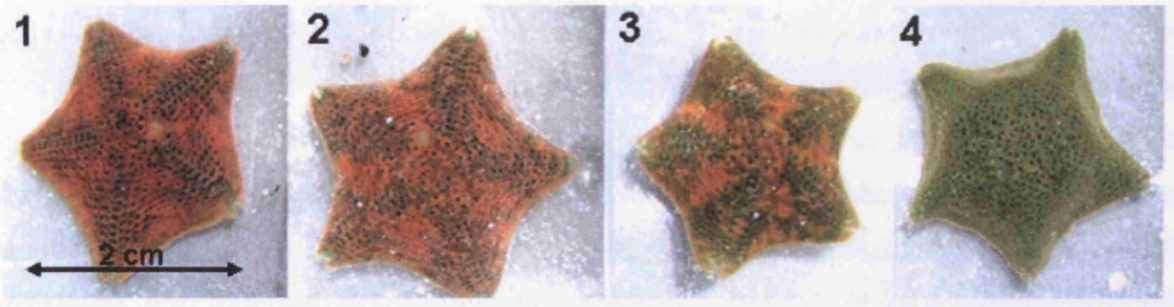

Figure 5.1. Photographs of the three South African *Parvulastra* samples. 1 & 2: *P. kommetjia* sp. nov., 3: intermediate phenotype of *P. exigua*, and 4: unmottled phenotype of *P. exigua*.

Four explanations are proposed for the existence and distribution of this new species: the first and second, based on demography, and the third and fourth, on a speciation (either sympatric or allopatric) event between *P. exigua* and *P. kommetjia* sp. nov. (see below).

5.1.1. Remnant of an ancient lineage

The current distribution of *P. kommetjia* sp. nov. represents a remnant population of an ancient lineage which previously occupied a much wider range. Due to competition for resources, disease, changing habitats etc., this species may have undergone a dramatic reduction in size and genetic bottlenecking resulting in low genetic variability. Such an explanation has been suggested for other marine intertidal taxa occupying 'island-like' habitats (Bucklin and Wiebe 1998; Espinosa and Ozawa 2006) it has also been proposed to explain the restricted distribution of *P. vivipara*, which occupies only two locations either side of the Eyre Peninsula in Australia (Byrne and Cerra 1996). If this proves to be the cause of the restricted distribution in this *P.*

kommetjia sp. nov it might be expected that other isolated pockets of this species may exist, but these remain currently undetected. Further analysis of the geographic distribution of this taxon is now an urgent priority to address this question.

5.1.2. Founder effect speciation

Parvulastra kommetjia sp. nov. may have arisen from a founder event at Kommetjie. According to the founder effect model (Mayr 1963), speciation can occur when a few migrant individuals colonise a new habitat and genetic structuring ensues under the conditions of low genetic variation caused by the founder event (Templeton 1980 1981). This theoretical model has not been corroborated using laboratory experiments (Ringo *et al.* 1985; Moya *et al.* 1995; Templeton 1999) and has been criticized by several authors in the past (Lande 1980; Barton and Charlesworth 1984; Rice and Hostert 1993; Coyne 1994). However, Paulay and Meyer (2002) argue that founder speciation in marine environments is more important than previously considered, and this method of speciation has been suggested for a number of South African intertidal taxa (Evans *et al.* 2004; Teske *et al.* 2005). If this scenario was responsible for the presence and limited distribution of *P. kommetjia* sp. nov., this poses the problem of identifying the 'parental' species. *Parvulastra kommetjia* sp. nov. appears to be closely related to (according to mtDNA) *P. parvivipara* and *P. vivipara*. However, their viviparous life cycle and extremely restricted dispersal ability (see Dartnall 1969; Keough and Dartnall 1978; Chia and Walker 1991; Byrne and Cerra 1996) make it unlikely that *P. parvivipara* and *P. vivipara* are parental lineages. The possibility that *P. dyscrita* or another asterinid species are the parental species can not be excluded. Finally, the parental species could be *P. exigua* from another global location, or from the east coast of South Africa. However, this explanation is unlikely for *P. kommetjia* sp. nov. because it is more difficult to imagine reproductive isolation evolving and speciation occurring, unless there were no existing populations of *P. exigua* at Kommetjie at the time of arrival of the founding individuals of *P. kommetjia* sp. nov. Moreover, if there was an existing population of *P. exigua* at Kommetjie at the time of the founder event, sympatric speciation would be the most appropriate model for the divergence of the two species.

5.1.3. Ecological or adaptive (sympatric) speciation

Adaptive speciation which occurs sympatrically is considered a more controversial phenomenon than allopatric speciation, but nevertheless many authors advocate its existence (Knowlton 1993; Danley *et al.* 2000; Briggs 2006 Savolainen *et al.* 2006). Sympatric adaptive speciation implies that disruptive selective forces drive adaptive differentiation and assortative mating resulting in reproductive isolation, despite gene flow (Rice and Hostert 1993) occurring between the diverging taxa (reviewed in Doebeli *et al.* 2005). According to Doebeli *et al.* (2005) adaptive speciation is a plausible evolutionary process in many different evolutionary scenarios. Doebeli and colleagues (2005) argue that adaptive speciation is likely to occur in response to many different selection scenarios including competitive speciation or frequency dependent disruptive selection on the phenotype. Knowlton (1993) in a review of marine sibling species, suggests that sexual selection on reproductive recognition systems on the adults or gametes (using chemical cues) may play a role in driving divergence. Knowlton (1993) also suggests that sibling species which occur at different depth or salinity distributions, which may partially disrupt gene flow, could reflect clinal speciation. Kommetjie shows geological discontinuity (Brown and Jarman 1978) and has been suggested to be the boundary between the Namaqua and Agulhas biogeographic provinces (Stephenson 1944), although the location of this boundary is disputed (Jackson 1976; Brown and Jarman 1978; Emanuel *et al.* 1992; Hiller 1994; Bustamante *et al.* 1997; Neraudeau and Mathey 2000). However, if the boundary does occur at Kommetjie, this may suggest unusual or intense environmental selection pressures at this location – another issue requiring further investigation. It seems a remarkable coincidence that this species occurs on a biogeographic boundary, suggesting that this may provide the ecological selection pressures driving adaptive speciation.

5.1.4. Allopatric speciation

The same geological and theoretical arguments outlined in Section 5.1.3. could be suggested for allopatric speciation, except without gene flow occurring between the

diverging taxa. A vicariant event may have caused some *P. exigua* individuals to be isolated, causing a genetic bottleneck resulting in low genetic variation in the isolated population, which then diverged in allopatry, accruing genetic differences and reproductive isolation. Secondary contact could have ensued causing geographic introgression but not genetic introgression. The Cape Peninsula is a hotspot for species endemism (Branch and Branch 1981) and sea level rises in interglacial periods over the past 300,000 years would have repeatedly caused the Cape Peninsula to form an island (Branch and Branch 1981). This does not explain why *P. kommetjia* sp. nov. is found only at Kommetjie, and not on the rest of the Peninsula. However, Kommetjie may have been repeatedly exposed and submerged, potentially creating the vicariant event necessary for allopatry.

Due to the location and geographic distribution of *P. kommetjia* sp. nov. a speciation event is more likely than a demographic explanation, however, whether *P. kommetjia* sp. nov. diverged in sympatry or allopatry is open to debate. Sympatric speciation is thought to occur much more quickly than allopatric speciation (Briggs 2006). As there are no known geological events in the area that could have caused a population to be isolated for a long time period, then sympatric speciation seems the most plausible explanation for the evolution of *P. kommetjia* sp. nov. However, currently there is no compelling evidence for or against any of the four explanations, and further intensive surveys and study will be required to resolve the origin of this enigmatic taxon.

5.2. Parvulastra dyscrita

Although it was not a primary aim of this study to investigate the population genetics of *P. dyscrita*, molecular investigation was necessary in order to clarify its relationship with *P. exigua*. Morphological examination of gonopore location revealed that *P. dyscrita* lacked oral gonopores, confirming that it is a distinct species from both *P. exigua* and *P. kommetjia* sp. nov. which both have oral gonopores. AFLP data indicated that *P. dyscrita* is a cohesive species, genetically divergent from the other South African *Parvulastra*. The actin intron data did not separate *P. exigua* from *P. dyscrita* into two monophyletic clades, however this marker is probably not sensitive

enough to distinguish between closely related species. The mitochondrial data for *P. dyscrita* indicated two parapatric reciprocally monophyletic mitochondrial clades. The same mitochondrial structure was observed in *P. exigua*. The inferred contact zone between these two clades of *P. dyscrita* is on the Agulhas Bank, approximately 200 km west of the contact zone between the two *P. exigua* clades observed. The formation of these clades may be explained by a vicariant event, as discussed in Section 5.3.3, in relation to the two *P. exigua* mitochondrial clades. Although a planktonic life history (suspected in *P. dyscrita*; see Branch *et al.* 1994) is considered ancestral to *P. exigua's* benthic life history (Byrne 1995; Hart *et al.* 1997; Hart 2000; Byrne 2006), the divergence event between these two species was not investigated, and therefore it remains unclear which is the ancestral species in South Africa, and by what speciation mechanism (e.g. sympatric, allopatric or founder) they diverged. However, this is the first time *P. dyscrita* has been genetically analysed and this species can now be included in the family phylogenetic tree with confidence (Fig. 5.2).

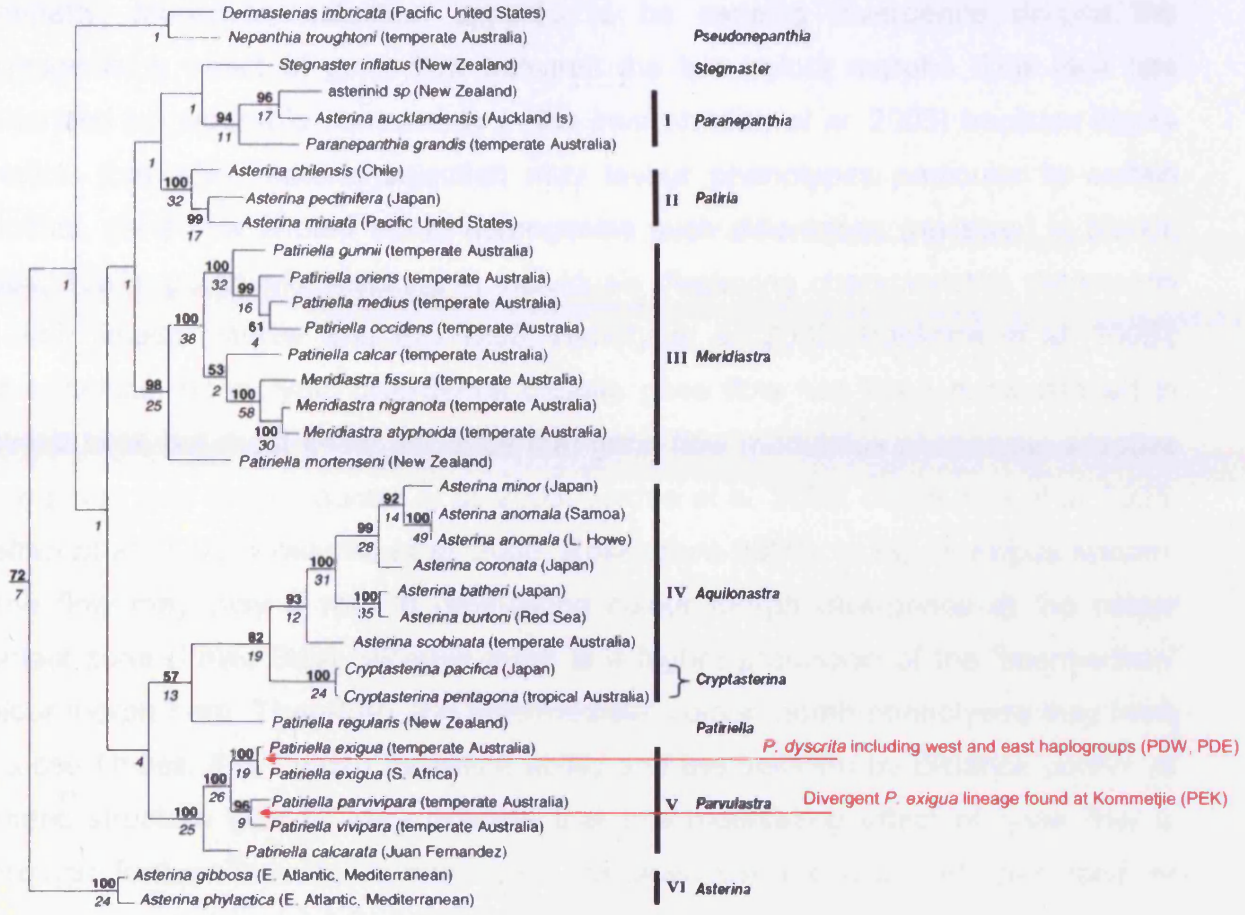

Figure 5.2. Maximum parsimony tree of asterinid mtDNA sequences. Geographic origin is indicated to the right of the taxa. Values in bold above nodes indicate bootstrap support (>50%), and values in italics (below nodes) are decay indices. MtDNA clades I – VI and genus names to the right of the numerals are newly classified genera (O'Loughlin and Waters 2004). Tree taken from Waters et al. 2004. Red arrows represent position of *Parvulastra dyscrita* and *P. kommetjia* sp. nov. identified from mtDNA during this study.

5.3. *Parvulastra exigua*

The remarkable *P. exigua* colour morph polymorphism appears to be decoupled from both mtDNA and AFLP neutral genetic structure indicating demographic or neutral processes are not the driving mechanisms behind its polyphenism. Phenotypic plasticity in response to ecological specialization (diet or water temperature) could be responsible for the colour morphs. However, there is more compelling evidence for selection being the driving force behind the phenotypic polymorphism, namely (a)

diverging outlier loci, corresponding to the east / west colour morph distribution, and (b) both *P. exigua* and *P. dyscrita* displaying the mottled morph in the area of sympatry. Moreover, selection appears to be causing divergence despite the homogenising effect of gene flow between the two colour morphs. This idea has generated considerable controversy in the past (Jordan *et al.* 2005) because theory predicts that while natural selection may favour phenotypes particular to certain habitats, gene flow should act to homogenise such differences (reviewed in Slatkin 1985, 1987), potentially resulting in individuals displaying characteristics detrimental to their fitness (Storfer and Sih 1998; Hendry *et al.* 2002; Hoekstra *et al.* 2005). Nevertheless, phenotypic divergence despite gene flow has been demonstrated in several taxa, but most show evidence that gene flow modulates phenotypic adaptive divergence (e.g. Saint-Laurent *et al.* 2003; Jordan *et al.* 2005; Guillaumet *et al.* 2005; Petren *et al.* 2005; Hoekstra *et al.* 2005; Rosenblum 2006). In the *P. exigua* system, gene flow may play a role in modulating colour morph divergence at the colour contact zone (False Bay) because there is a higher proportion of the 'intermediate' colour morph here. Therefore, the 'intermediate' colour morph phenotypes may have reduced fitness. The limited dispersal ability and the isolation by distance pattern of genetic structure (AFLP) may indicate that this modulating effect of gene flow is obscured further from the contact zone, because the proportion of 'unmottled' or 'mottled' sea stars increases.

Three genetic approaches (mtDNA and Actin intron sequencing, and AFLP) showed different patterns of population structure in *P. exigua* on the South Africa coast. The mtDNA data revealed two monophyletic allopatric east and west clades, and indicated a population expansion event within the west coast clade. The Actin intron sequences data showed a single *P. exigua* group, but was not sensitive enough to differentiate the east and west coast mtDNA haplogroups. The AFLP data showed a strong pattern of isolation by distance, with dispersal distances in the region of 10's to 100's of kilometers. Approximately 6.8 % of the genome was potentially subject to diversifying selection, indicated by the AFLP 'outlier' alleles which show a pattern of directionality with phenotype frequencies declining from east to west (or vice versa,

see loci phenotype frequency histograms in Chapter 4) along the coast, possibly coinciding with the phenotypic patterns / biogeographic provinces or mitochondrial structure. Three hypotheses are suggested to explain the genetic structure evident in *P. exigua* in South Africa. Although a combination of any of the forces associated with the three hypotheses could have contributed to evolution of the *P. exigua* genetic structure, a vicariance event, followed by secondary contact is the most plausible.

5.3.1. Founder events

The two reciprocally monophyletic haplogroups could have resulted from two separate founder events, on the east and west coasts of South Africa. The founders could have come from (i) different localities (e.g. Australia seeded the east coast and Amsterdam Island, the west coast), (ii) the same locality (possibly Australia) but at different times, or (iii) one founder event that seeded a population on the east coast (which may have undergone a more ancient expansion event) then seeded a west coast population in a second founder event. Scenario (iii) was recently proposed as an explanation for two separate reproductive stocks of the abalone *Haliotis midae* which occur on either side of Cape Agulhas (Evans et al. 2004). Evidence from three different genetic markers indicated an isolated introduction event to the east of Cape Agulhas, and subsequent range expansion in an easterly direction (Evans et al. 2004). These authors also noted that the area of transition between the stocks coincides with oceanographic features of the region. It is plausible that if a small number of *P. exigua* individuals from the east coast were caught up in the Agulhas current, they could travel southwards, reaching the Benguela current which may have deposited them on the west coast.

Following the putative founder events, population expansion could have ensued on the west coast. Ancestral monophyly (the presence of a single basal haplotype which gives rise to several derived haplotypes, as seen in both the east and west clade of *P. exigua*) implies that the populations were founded by a few individuals which then rapidly multiplied (Teske et al. 2005; Park and Foighil 2000). Partial introgression may have occurred when the *P. exigua* populations reached the contact zone at the

Agulhas Bank. Introgression between the two populations may explain the nuclear isolation by distance patterns of genetic structure. However, persistence of the reciprocal monophyly in the mtDNA is expected to have been eroded without some partial reproductive isolation, hybrid breakdown or strong selection on the mitochondrial genome (Wilson and Bernatchez 1998; Bensch et al. 2002; Helbig et al. 2005; Bensch et al. 2006; Ellison and Burton 2006; Secondi et al. 2006). The sporadic nature of founder events may explain why *P. exigua* does not occur in South America or New Zealand. If the *P. exigua* haplogroups in South Africa were the result of founder events, this would directly contradict the Out of Africa hypothesis proposed by Waters and Roy (2004) which suggests that Africa is the ancestral origin and a founder event was responsible for the establishment of a *P. exigua* population in Australia.

5.3.2. Phylogeographic break without barriers to gene flow

In continuously distributed species, intraspecific mitochondrial monophyletic groups separated by large genealogical gaps, can potentially develop as a result of low individual dispersal distances and small population sizes (Neigel and Avise 1993; Irwin 2002; Kuo and Avise 2005). This pattern can arise because mtDNA is inherited from only one parent (although not necessarily matriarchally; Gyllensten *et al.* 1991; Bromham *et al.* 2003; Kvist *et al.* 2003; Rokas *et al.* 2003) and evolves through bifurcating genealogies (individuals belong to one genealogical clade, but cannot be genealogically intermediate; Irwin 2002). These conditions lead to phylogeographic structure and under extreme conditions of low dispersal or isolation by distance, one pair of adjacent sampling locations can have individuals that belong to different genealogical clades but other pairs of adjacent locations may belong to the same genealogical clade (Irwin 2002). This divergent mtDNA pattern is theoretically not mirrored in nuclear DNA because unlike mtDNA, which is descended from a single ancestral individual, different sections of the recombining nuclear genome can be inherited from different ancestral individuals (Irwin 2002). Disruptive selection (Kuo and Avise 2005) could cause the phylogeographic structure of mtDNA to persist despite gene flow if different mtDNA haplotypes are favoured in different ecological

conditions, or across ecological gradients (Irwin 2002). Kuo and Avise (2005) suggest that this model of mtDNA divergence may not hold (or may do so with less force) if strong balancing selection has acted on the loci. However, these authors conclude that it is unlikely that such strong balancing (or disruptive) selection would characterize mitochondrial genomes.

The conditions of this model of mtDNA divergence with gene flow lend themselves to the phylogeographic structure observed in the South African *P. exigua* populations, which have restricted dispersal of up to a few hundred kilometres, show an isolation by distance structure and occur over an environmental temperature gradient. Past *P. exigua* populations may have been smaller and mtDNA divergence could have occurred via the above processes around the contact zone. This may have been reinforced by selection favouring particular genotypes on either side of the contact zone. This would mean that the two mtDNA groups diverged in parapatry and that the divergence occurred along a cline without a physical barrier to gene flow. However, two lines of evidence suggest that this is not the most likely explanation for the observed mtDNA divergence and the nuclear isolation by distance pattern in *P. exigua*. Firstly, in contrast to other systems showing divergence with gene flow (Smith *et al.* 2005; Grahame *et al.* 2006;), the mtDNA divergence does not coincide with the phenotypic divide. Secondly, genetic divides in other intertidal species over the Agulhas Bank have been reported (Evans *et al.* 2004; Teske *et al.* 2006; *P. dyscrita,* this study), which suggests that the phylogeographic divide evident in *P. exigua* has not arisen in this location by chance, and it is more likely to be the result of a geographic barrier to gene flow (Avise 1992; Neigel and Avise 1993; Kuo and Avise 2005). However, the same characteristics of mtDNA that make this model of divergence plausible, also pre-dispose it to evolving reciprocal monophyly and showing phylogeographic breaks caused by barriers to gene flow. Therefore, inferences of the origin of phylogeographic structure in mtDNA must be made with caution (Irwin 2002; Grahame *et al.* 2006).

5.3.3. Vicariance (allopatry)

Geographic patterns of genetic differentiation often reflect events in the past more accurately than current patterns of gene flow (Hellberg *et al.* 2002). Significant genetic structure can result when an historical vicariance event interrupts gene flow between populations, followed by a combination of mutation, adaptation and lineage sorting (drift) eventually resulting in reciprocal monophyly (Grosberg and Cunningham 2001). A vicariance event caused by sea level changes and exposure of the Agulhas Bank continental shelf may have caused allopatry between the east and west *P. exigua* populations. During the allopatric period, mtDNA reciprocal monophyly and adaptive divergence in some parts of the nuclear genome could have evolved in response to the different temperature regimes on the coasts. Subsequent sea level changes may have resulted in secondary contact and the formation of the isolation by distance pattern evident from the AFLP data. Population expansion in the west could have occurred either during the allopatric period, when the east and west populations were adapting to the ecological conditions or after secondary contact.

MtDNA drifts to reciprocal monophyly more quickly than nuclear DNA because of its smaller effective population size (Birky *et al.* 1989, cited in Hellberg *et al.* 2002; summarized by Ballard and Whitlock 2004), therefore the period of allopatry may have been only long enough for some adaptive divergence to occur, but not long enough for complete reproductive isolation between the east and west populations. However, the mtDNA reciprocal monophyly implies a lack of mitochondrial introgression (Grosberg and Cunningham 2001), despite evidence of nuclear introgression. Nuclear but not mitochondrial introgression has been observed in other hybrid zones (Bensch *et al.* 2002; Haig *et al.* 2004; Helbig *et al.* 2005; Secondi *et al.* 2006), and could potentially be the result of three different processes. (1) *P. exigua's* benthic life cycle, limited dispersal ability, the opposing inshore current directions and the isolation by distance AFLP structure indicate that gene flow occurs only between adjacent locations. Therefore, if secondary contact was very recent, the mtDNA may

not have had time to introgress (except at the contact zone) throughout the population, and may not have been evident at the contact zone due to limited sampling. (2) Selection on mtDNA itself, either directly (nucleo-cytoplasmic incompatibility (Secondi *et al.* 2006) e.g. co-adaptation between mitochondrial and nuclear genomes causing functional incompatibilities resulting in hybrid breakdown (Ellison and Burton 2006)), or indirectly (e.g. differential survival of the heterogametic sex (Haldane's Rule), hybrid reduced fitness or different spawning times) causing a partial reproductive barrier therefore preventing mitochondrial introgression (Bensch *et al.* 2002; Helbig *et al.* 2005; Secondi *et al.* 2006). (3) As the effective population size of mtDNA is only approximately ¼ of the size of nuclear autosomal genes, mtDNA haplotypes may disappear rapidly from secondary contact populations by drift (Secondi *et al.* 2006). Therefore, using only mitochondrial markers to detect introgression has been criticized (Dasmahapatra *et al.* 2002; Mishmar *et al.* 2003; Ellison and Burton 2006; Secondi *et al.* 2006). Furthermore, the directionality of the phenotype frequencies of the nuclear 'divergent outlier loci' indicate a lack of introgression between the east and west coast. This is consistent with the suggestion that selectively neutral sections of the genome are more likely to introgress (Secondi *et al.* 2006). This also adds weight to the suggestion that the divergent outlier loci (and possibly the mtDNA) are under directional selection.

The allopatric explanation for this deep phylogenetic break is more plausible than an explanation of the phylogeographic break without a barrier to gene flow (Section 5.3.2) because of the location of the divide. If selective adaptation was responsible, it would be expected that the divide would coincide with a biogeographic boundary (Dawson 2001). However, the effects of sea level changes in the Agulhas Bank area, as well as the phylogenetic divide evident only in the mtDNA, and not the nuclear DNA, make a historical vicariance event followed by secondary contact more plausible (Dawson 2001). Further, support for the vicariance hypothesis comes from *P. dyscrita* and other estuarine and marine species (Evans *et al.* 2004; Teske *et al.* 2006), which also shows phylogeographic divides over the Agulhas Bank area. This may indicate that the whole Agulhas Bank region is a phylogeographic (but not

biogeographic) 'transition zone' (Dawson 2001) where many marine taxa with different life histories may show adjacent phylogeographic structure, similar to the phylogeographic structure attributed to vicariance events seen in many coastal marine taxa within the Californian transition zone and around Point Conception (Burton 1998; Bernardi 2000; Dawson *et al.* 2001; Sotka *et al.* 2004). Several marine invertebrates and fish show genetic divergence across an intertidal phylogeographic break at Cape Canaveral, Florida (Reeb and Avise 1990; reviewed in Avise 1992; Grosberg and Cunningham 2001; Hellberg *et al.* 2002; Palumbi 2003). It is not certain whether local adaptation or oceanographic barriers maintain the genetic divisions between the divergent populations within these species (see Hare and Avise 1996). However, it is clear that a major vicariance event interrupted gene flow for many taxa, and that larval dispersal has yet to restore genetic homogeneity between the populations (Grosberg and Cunningham 2001).

5.4. Additional work and future directions

Research, especially where there is limited prior knowledge of the system, often raises more questions than answers and this study is no exception. Further investigations are recommended into the ancestral origins; the global genetic structure; dispersal; mechanisms responsible for the *P. exigua* colour morphs; effects of the intertidal temperature gradient on selection and on the phylogeography of *Parvulastra* species in South Africa. The current study highlights the necessity of utilising both nuclear genomic and mtDNA markers, as well as the need for extensive sampling in order to identify cryptic taxa. Additional to the work reported in Chapters 2 - 4, the following methods were developed but are not reported in this thesis as they were not pursued in detail and are not central to the main theme of the project. However, to our knowledge, this was the first study to show that the ribosomal Internal Transcribed Spacer regions (ITS) are problematic markers for *Parvulastra* as they reveal divergent sequences from multiple copies within individuals. Furthermore, 30 anonymous *P. exigua* nuclear sequences were generated which can be used in the future for single nucleotide polymorphism (SNP) identification. Finally, to our

knowledge, this was the first study to isolate DNA sequence containing microsatellites using a non-enriched library method in a related species, *Asterina gibbosa*, despite several previous attempts using enriched microsatellite library methods.

5.5. References

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna - a case history with lessons for conservation biology. *Oikos* **63**, 62-76.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133-164.

Bensch S, Helbig AJ, Salomon M, Siebold I (2002) Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Molecular Ecology* **11**, 473-481.

Bensch S, Irwin DE, Irwin JH, Kvist L, Akesson S (2006) Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. *Molecular Ecology* **15**, 161-171.

Bernardi G (2000) Barriers to gene flow in *Embiotoca jacksoni*, a marine fish lacking a pelagic larval stage. *Evolution* **54**, 226-237.

Branch G, Branch M, Bannister A (1981) *The living shores of Southern Africa*, 5th edn. Struik, Cape Town.

Branch GM, Griffiths CL, Branch ML, Beckley LE (1994) *Two Oceans, A guide to the marine life of southern Africa* David Philip Publishers (Pty) Ltd., Cape Town.

Briggs JC (2006) Proximate sources of marine biodiversity. *Journal of Biogeography* **33**, 1-10.

Bromham L, Eyre-Walker A, Smith NH, Smith JM (2003) Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology & Evolution* **18**, 2-4.

Brown AC, Jarman N (1978) Coastal Marine Habitats. In: *Biogeography and Ecology of Southern Africa* (eds. Werger MJA, Bruggen acV), pp. 1239 - 1227. The Hague, Cape Town.

Bucklin A, Wiebe PH (1998) Low mitochondrial diversity and small effective population sizes of the copepods *Calanus finmarchicus* and *Nannocalanus minor*: Possible impact of climatic variation during recent glaciation. *Journal of Heredity* **89**, 383-392.

Burton RS (1998) Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* **52**, 734-745.

Bustamante RH, Branch GM, Eekhout S (1997) The influences of physical factors on the distribution and zonation patterns of South African rocky shore communities. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **18**, 119-136.

Byrne M (1995) Changes in larval morphology in the evolution of benthic development by *Patiriella exigua* (*Asteroidea, Asterinidae*), a comparison with the larvae of *Patiriella* species with planktonic development. *Biological Bulletin* **188**, 293-305.

Byrne M (2006) Life history diversity and evolution in the *Asterinidae*. *Integrative and Comparative Biology* **46**, 243-254.

Byrne M, Cerra A (1996) Evolution of intragonadal development in the diminutive asterinid sea stars *Patiriella vivipara* and *P. parvivipara* with an overview of development in the *Asterinidae*. *Biological Bulletin* **191**, 17-26.

Coyne JA (1994) Mayr, Ernst and the Origin of Species. *Evolution* **48**, 19-30.

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54**, 1725-1737.

Dartnall AJ (1969) *The taxaonomy and biogeography of the sea star genus Patiriella in Tasmania* Unpublished MSc thesis, University of Tasmania.

Dartnall AJ, Byrne M, Collins J, Hart MW (2003) A new viviparous species of asterinid (*Echinodermata, Asteroidea, Asterinidae*) and a new genus to accommodate the species of pan-tropical exiguoid sea stars. *Zootaxa* **359**, 1 - 14.

Dasmahapatra KK, Blum MJ, Aiello A, *et al.* (2002) Inferences from a rapidly moving hybrid zone. *Evolution* **56**, 741-753.

Dawson MN (2001) Phylogeography in coastal marine animals: a solution from California? *Journal of Biogeography* **28**, 723-736.

Dawson MN, Staton JL, Jacobs DK (2001) Phylogeography of the tidewater goby, *Eucyclogobius newberryi* (*Teleostei, gobiidae*), in coastal California. *Evolution* **55**, 1167-1179.

Doebeli M, Dieckmann U, Metz JAJ, Tautz D (2005) What we have also learned: Adaptive speciation is theoretically plausible. *Evolution* **59**, 691-695.

Ellison CK, Burton RS (2006) Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* **60**, 1382-1391.

Emanuel BP, Bustamante RH, Branch GM, Eekhout S, Odendaal FJ (1992) A zoogeographic and functional approach to the selection of marine reserves on the west coast of South Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* **12**, 341-354.

Espinosa F, Guerra-Garcia JM, Fa D, Garcia-Gomez JC (2006) Effects of competition on an endangered limpet Patella *Ferruginea* (*Gastropoda : Patellidae*): Implications for conservation. *Journal of Experimental Marine Biology and Ecology* **330**, 482-492.

Evans BS, Sweijd NA, Bowie RCK, Cook PA, Elliott NG (2004) Population genetic structure of the perlemoen *Haliotis midae* in South Africa: evidence of range expansion and founder events. *Marine Ecology-Progress Series* **270**, 163-172.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution* **60**, 268-278.

Grosberg R, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine community ecology.* (eds. Bertness MD, Hay ME, Gaines SD), p. 61–84. Sinauer, Sunderland, Massachusetts.

Guillaumet A, Crochet PA, Godelle B (2005) Phenotypic variation in Galerida larks in Morocco: the role of history and natural selection. *Molecular Ecology* **14**, 3809-3821.

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**, 255-257.

Haig SM, Mullins TD, Forsman ED, Trail PW, Wennerberg L (2004) Genetic identification of Spotted Owls, Barred Owls, and their hybrids: Legal implications of hybrid identity. *Conservation Biology* **18**, 1347-1357.

Hare MP, Avise JC (1996) Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution* **50**, 2305-2315.

Hart M (2000) Phylogenetic analyses of mode of larval development. *Seminars in Cell & Developmental Biology* **11**, 411-418.

Hart MW, Byrne M, Smith MJ (1997) Molecular phylogenetic analysis of life-history evolution in Asterinid starfish. *Evolution* **51**, 1848-1861.

Helbig AJ, Seibold I, Kocum A, *et al.* (2005) Genetic differentiation and hybridization between greater and lesser spotted eagles (*Accipitriformes : Aquila clanga, A. pomarina*). *Journal of Ornithology* **146**, 226-234.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.

Hendry AP, Taylor EB, McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: Lake and stream stickleback in the misty system. *Evolution* **56**, 1199-1216.

Hiller N (1994) The environment, biogeography, and origin of the southern African recent brachiopod fauna. *Journal of Paleontology* **68**, 776-786.

Hoekstra HE, Krenz JG, Nachman MW (2005) Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity* **94**, 217-228.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution* **56**, 2383-2394.

Jackson L (1976) Aspects of the intertidal ecology of the east coast of South Africa, pp. 1 - 72. Oceanographic Research Institute, Durban.

Jordan MA, Snell HL, Snell HM, Jordan WC (2005) Phenotypic divergence despite high levels of gene flow in Galapagos lava lizards (*Microlophus albemarlensis*). *Molecular Ecology* **14**, 859-867.

Keough MJ, Dartnall AJ (1978) A new species of viviparous asterinid asteroid from Eyre Peninsula, South Australia. *Records of the South Australian Museum* **17**, 407 - 416.

Knowlton N (1993) Sibling Species in the Sea. *Annual Review of Ecology and Systematics* **24**, 189-216.

Kuo CH, Avise J (2005) Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. *Genetica* **124**, 179-186.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution* **20**, 243-247.

Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist* **116**, 463-479.

Mayr E (1963) *Animal Species and Evolution* Harvard University Press, Cambridge MA.

Mishmar D, Ruiz-Pesini E, Golik P, *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 171-176.

Moya A, Galiana A, Ayala FJ (1995) Founder effect speciation theory - failure of experimental corroboration. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3983-3986.

Neigel JE, Avise JC (1993) Application of a random-walk model to geographic distributions of animal mitochondrial DNA variation. *Genetics* **135**, 1209-1220.

Neraudeau D, Mathey B (2000) Biogeography and diversity of South Atlantic Cretaceous echinoids: implications for circulation patterns. *Palaeogeography Palaeoclimatology Palaeoecology* **156**, 71-88.

O'Loughlin PM, Waters JM (2004) A molecular and morphological revision of the genera of *Asterinidae* (*Echinodermata: Asteroidea*). *Memoirs of Museum Victoria* **61**, 1 - 40.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, S146-S158.

Park JK, Foighil DO (2000) Genetic diversity of Oceanic Island Lasaea (*Mollusca : Bivalvia*) lineages exceeds that of continental populations in the northwestern Atlantic. *Biological Bulletin* **198**, 396-403.

Paulay G, Meyer C (2002) Diversification in the tropical pacific: Comparisons between marine and terrestrial systems and the importance of founder speciation. *Integrative and Comparative Biology* **42**, 922-934.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Molecular Ecology* **14**, 2943-2957.

Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species – mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**, 397-406.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation - what have we learned in 40 years. *Evolution* **47**, 1637-1653.

Ringo J, Wood D, Rockwell R, Dowse H (1985) An experiment testing 2 hypotheses of speciation. *American Naturalist* **126**, 642-661.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology & Evolution* **18**, 411-417.

Rosenblum EB (2006) Convergent evolution and divergent selection: Lizards at the White Sands ecotone. *American Naturalist* **167**, 1-15.

Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology* **12**, 315-330.

Savolainen V, Anstett MC, Lexer C, *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.

Secondi J, Faivre B, Bensch S (2006) Spreading introgression in the wake of a moving contact zone. *Molecular Ecology* **15**, 2463-2475.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Smith TB, Calsbeek R, Wayne RK, *et al.* (2005) Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. *Journal of Evolutionary Biology* **18**, 257-268.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology* **13**, 2143-2156.

Stephenson (1944) The constitution of the intertidal fauna and flora of South Africa. *Annals of the natal museum* **10**, 261-358.

Storfer A, Sih A (1998) Gene flow and ineffective antipredator behavior in a stream-breeding salamander. *Evolution* **52**, 558-565.

Templeton AR (1980) The theory of speciation via the founder principle. *Genetics* **94**, 1011-1038.

Templeton AR (1981) Mechanisms of speciation - a population genetic approach. *Annual Review of Ecology and Systematics* **12**, 23-48.

Templeton AR (1999) Experimental tests of genetic transilience. *Evolution* **53**, 1628-1632.

Teske PR, Hamilton H, Palsboll PJ, *et al.* (2005) Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. *Marine Ecology-Progress Series* **286**, 249-260.

Teske PR, McQuaid CD, Froneman PW, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology-Progress Series* **314**, 283-293.

Waters JM, Roy MS (2004) Out of Africa: The slow train to Australasia. *Systematic Biology* **53**, 18-24.

Wilson CC, Bernatchez L (1998) The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology* **7**, 127-132.