

**Studies on the Population Genetics of
Euphausiids: A Comparison of Patterns in
Pelagic Taxa Displaying Different
Distributions and Life – Histories**

Gordon W. Harkins



A thesis submitted in fulfilment of the requirements for the degree of Doctor Philosophiae in the faculty of Biodiversity and Conservation Biology, University of the Western Cape.

November 2006

Supervisors: Prof. M. J. Gibbons and Dr. M. E. D'Amato

KEYWORDS:

Euphausiid

Phylogeography

DNA

Population genetics

Coalescent

Effective population size

Mutation-drift equilibrium

Molecular clock

Locus

Holoplanktic crustaceans



ABSTRACT

Studies on the Population Genetics of Euphausiids: A Comparison of Patterns in Pelagic Taxa Displaying Different Distributions and Life – Histories

G. W. Harkins

PhD Thesis, Faculty of Biodiversity and Conservation Biology, University of the Western Cape.

The systematic and population genetic relationships were characterised for three ecologically related euphausiid species: *Euphausia lucens*, *E. recurva* and *E. vallentini*. These species have different geographical distributions and life histories. All three species have a circumpolar distribution in the Southern Hemisphere while *E. recurva* is also distributed in the North Pacific. DNA sequence variation was determined for three regions of mitochondrial DNA (ND1, CO1 and 16SrRNA) and a single nuclear gene (ITS-1). It was conclusively demonstrated that both *E. lucens* and *E. vallentini* represent valid taxonomic species with fixed differences observed in both the nuclear and mitochondrial genes and that the low divergences previously reported for these species with 16SrRNA and CO1 resulted from a species misidentification. It is also shown that previous attempts to date the divergence between Antarctic and Sub Antarctic euphausiid species based on 16SrRNA distances (Zane and Paternello 2000) suffer from a large overestimation due to a calculation error. The corrected estimates suggest Pliocene divergences for these groups and cast doubts on the importance of

the Antarctic Circumpolar Current as barrier to population exchange between these lineages.

Marked differences in the patterns of variation and diversity were exhibited by the three species. In *E. vallentini* sequence variation among individuals was completely absent for all the molecular markers (*COI*, *16S rDNA* and *ITS-1*) employed with the exception of the mitochondrial *NDI* fragment in which sequence heterogeneity was very low. Consequently, no population subdivision was observed in this species across its circumglobal distribution. In contrast, *E. recurva* exhibited qualitatively concordant patterns of differentiation with all of the molecular markers with three geographically delineated clades consistently recovered in the phylogenetic analyses that correspond to distinct populations the SE Atlantic and NE and IW Pacific. Coalescent analyses with MDIV indicated moderate unidirectional gene flow from the Indo-West Pacific to the SE Atlantic for this species suggesting that zooplankton can be highly geographically structured on macrogeographic scales despite experiencing gene flow. In *E. lucens* geographical structure was revealed in the mitochondrial genes *NDI* and *COI* with fixed differences observed between populations occurring in the SE and SW Atlantic basins, while the nuclear *ITS-1* sequences were invariable.

Shallow within-population phylogenies and an excess of rare alleles was observed in the Southern Hemisphere populations of *E. recurva* closely resembling the patterns observed in *E. lucens* for the mitochondrial marker

ND1. Consistent with a population expansion scenario exponential growth was positive and significant in both the SE and SW Atlantic *E. lucens* populations and for the Southern Hemisphere *E. recurva* populations using Fluctuate. The North Pacific *E. recurva* population in contrast, contained an excess of common alleles resembling the patterns that appear to characterize the Sub Antarctic species *E. vallentini*. This suggests a deviation from a constant population size in the direction of population shrinkage. The levels of differentiation observed between populations in these species appear to correlate with latitude, the size of the geographical area encompassed by the species and the hydrological conditions that characterize the area. The lower variation observed in the Sub Antarctic species *E. vallentini* compared to the cold temperate *E. lucens* was predicted by the hypothesis that states that the level of variation will reflect differences in the degree of latitudinal displacement and habitat contraction experienced during past glacial periods.

DECLARATION

I declare that *Studies on the Population Genetics of Euphausiids: A Comparison of Patterns in Pelagic Taxa Displaying Different Distributions and Life – Histories* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Gordon W. Harkins



November 2006

Signed:

UNIVERSITY of the
WESTERN CAPE

Acknowledgements

I would like to thank Dr. Gustavo Alvarez from UNIDEP, Janet-Bradford Grieve from the National Institute of Water Affairs, New Zealand, Dr. Erica Goetze from the Smithsonian Institute, USA, [Marion Island Guy](#), and Dr Alex Rogers from the British Antarctic Survey for kindly supplying euphausiid samples for use in this thesis. My appreciation is also extended to all the staff of Marine and Coastal Management in Cape Town including Professor Larry Hutchings, Janet xxx and Phoebe Wellington Mullins for all their help with the zooplankton sampling in South African waters and especially the friendly manner in which it was offered. In this respect I must also express my gratitude to the government of Norway for kindly donating ship time aboard the Research Vessel Dr. Fritzjof Nansen. My thanks also go out to several people from the Biodiversity and Conservation Biology Department at the University of the Western Cape in particular to the Secretary Mrs. Linda van Heerden, Mr. Verno Gordon, Dr. Franz Weitz, Dr. Eugene Moll and Dr. Richard Knight as well as the many other that contributed to making my time at UWC such a special experience. Special thanks in this respect must go to my close friends and colleagues Dr. Garry Russell, Dr. Touffiek Samaai, Mr. Goosain Isaacs. A special mention must be given to Mrs. Siaam Samaai from The legal Aid Clinic at the University of the Western Cape and to Mr. Molamu from the Department of Home Affairs, South Africa. I would also like to give special thanks to the following people for their unwavering support in all of my endeavors that have made a major contribution to advancing both my personal growth and my academic career: Professor Mark Gibbons and Dr. Maria Eugenia D'amato from the University of the Western Cape, Dr. Michelle Kelly from the National Institute of Water Affairs, New Zealand and Dr. Daniel Vasco from the University of Texas. I would also like to extend my sincere and everlasting

appreciation to the Geland family and to Professor Margaretha Hofmeyer and her family for providing a “home from home” during my time in South Africa. Words cannot express the love and gratitude I extend to my girlfriend Christabel Geland for her unreserved support and indefatigable belief in me. Lastly I would like to extend my love to my father Mr. W. A. Harkins for whom this thesis is dedicated for all your love and encouragement, for giving me the chance to run.



CONTENTS	PAGE NUMBERS
Title Page	i
Keywords	ii
Abstract	iii
Declaration	vi
Acknowledgements	vii
Contents	xiv
List of Figures	xv
List of Tables	xxiii
CHAPTER ONE	1
1.1. General Introduction	1
1.2. Previous Euphausiid Molecular Population Genetic Studies	5
1.3. Previous Euphausiid Molecular Systematics Studies	10
1.3.1. <i>Limitations to the Estimation of Molecular Population Parameters</i>	12
1.4. Gene Flow and Barriers within the Marine Realm	12
1.5. Euphausiid Systematic and Population Studies in Perspective	14
1.6. Relevance of Studying Temperate and Sub-Antarctic Euphausiid Species	15
1.7. Species Investigated	16
1.7.1. <i>Euphausia lucens</i> (Hansen 1905)	17

1.7.2.	<i>Euphausia vallentini</i> (Stebbing 1900)	27
1.7.3.	<i>Euphausia recurva</i> (Hansen 1905)	31
CHAPTER 2		35
Materials and Methods		35
2.1.	Sample Collection	35
2.2.	DNA Isolation	37
2.3.	Amplification of Mitochondrial DNA	39
2.3.1.	<i>NADH 1</i>	39
2.3.2.	<i>Cytochrome Oxidase I</i>	42
2.3.3.	<i>16S rRNA</i>	43
2.3.4.	<i>Sequencing Reactions</i>	43
2.4.	Amplification of Nuclear DNA	44
2.5.	Assessing Within-Individual Variation	45
2.6.	SSCP Screening of mtDNA Variation	48
2.6.1.	<i>Staining the SSCP Polyacrylamide Gels</i>	49
2.6.2.	<i>Scoring of mtDNA Haplotypes</i>	50
2.7.	Population Descriptive Summary Statistics	51
2.8.	Substitution Models	52
2.9.	Interspecific and Intraspecific Phylogenetics	52
2.9.1.	<i>Genealogical Network Construction</i>	53
2.10.	Estimation of Evolutionary Rates between Species	56
2.11.	Neutrality Tests	57
2.11.1.	<i>How the Tests are calculated</i>	63
2.12.	Estimates of Effective Population Mutation Rate (θ)	67
2.12.1.	<i>MCMC methods</i>	73

2.12.2. <i>Two-Lineage Estimator (Likelihood Method)</i>	76
CHAPTER 3	78
RESULTS <i>Euphausia lucens</i>	78
3.1. SSCP Scoring and Sequencing	78
3.2. Mitochondrial Sequence Variation	82
3.3. Phylogenetic Analysis	84
3.3.1. <i>MDIV</i>	94
3.4. Rooting of the Intraspecific Genealogy	96
3.5. Spatial and Temporal Comparisons	99
3.6. Spatial distribution and Direction of Mutations	101
3.6.1. <i>Frequencies of Mutant Nucleotides in the Sample</i>	102
3.7. Neutrality Tests	105
3.8. Estimating the Effective Population Size	109
3.9. Nuclear ITS-1 Results	118
CHAPTER 4	119
RESULTS <i>Euphausia vallentini</i>	119
4.1. SSCP Scoring and Sequencing	119
4.2. Sequence Variation	120
4.3. Phylogenetic Analysis and Rooting the Network	123
4.4. Neutrality Tests	124
4.5. Estimates of Effective Population Size	125
4.6. Demographic Hypotheses	125
4.7. Interspecific Comparisons	126

4.7.1. <i>Mitochondrial Variation in Cytochrome Oxidase I (COI)</i>	128
4.7.2. <i>COI Phylogenetics</i>	130
4.7.3. <i>Dating divergence using 16S rRNA sequences</i>	133
4.7.4. <i>Relative rates tests among species</i>	133
4.7.5. <i>Coalescent Estimate of Divergence Time</i>	137
4.8. Nuclear Sequence Variation	141
CHAPTER 5	143
5.1. Discussion <i>E. lucens</i> and <i>E. vallentini</i>	143
5.2. Systematic Comparisons	143
5.3. Species Divergence Time Estimates	144
5.4. Characterising the Mitochondrial Variation	148
5.5. Spatial and temporal genetic structure	149
5.5.1. <i>Coalescent Estimate of Population Divergence in E. lucens</i>	150
5.6. Differences between <i>E. lucens</i> and <i>E. vallentini</i>	154
5.7. Comparative Genetic Investigations	164
CHAPTER 6	162
RESULTS <i>Euphausia recurva</i>	162
6.1. Mitochondrial Variation, Phylogeography and Population Parameter Estimation	162
6.2. Mitochondrial Sequence Variation	170

6.3.	Genealogy Reconstruction	172
	6.3.1. <i>Rooting of the Intraspecific Genealogy</i>	177
6.4.	Regional and Temporal Comparisons	179
	6.4.1. <i>Spatial distribution and Direction of Mutations</i>	182
6.5.	Frequencies of Mutant Nucleotides in the Sample	183
6.6.	Neutrality Tests	185
6.7.	Effective Population Size Estimates	193
	6.7.1. <i>Coalescent Estimators in a Variable Environment (EVE)</i>	195
	6.7.2. <i>Fluctuate</i>	196
6.8.	MDIV	201
6.9.	Cytochrome Oxidase 1 Variation within <i>E. recurva</i>	203
	6.9.1. <i>Genealogical CO1 estimates</i>	205
6.10.	Nuclear ITS-1 Results	209
	6.10.1. <i>Nuclear Sequence Variation</i>	224
	6.10.2. <i>Neutrality tests</i>	227
6.11.	Effective Population Size Estimates	228
6.12.	Comparison of the different markers	232
6.13.	Interspecific <i>ITS-1</i> Comparisons	233
	 CHAPTER 7	 235
	DISCUSSION <i>Euphausia recurva</i>	235
7.1.	DNA Sequence Variation, Diversity and Phylogenetics	235
	7.1.1 <i>Mitochondrial Variation</i>	236

7.1.2. <i>Nuclear Variation</i>	236
7.2. Within-individual Variation in the Nuclear Marker <i>ITS-1</i>	237
7.3. Genetic Variation through Space and Time	238
7.4. Coalescent Estimate of <i>E. recurva</i> Population Divergence	241
7.4.1. <i>Pairwise Differences</i>	243
7.4.2. <i>Estimates of Migration Rates</i>	244
7.4.3. <i>Spatial and Temporal Structure around Southern Africa</i>	244
7.5. Congruent Geographical Distributions	245
7.6. Quantitative and Qualitative Differences among Markers	246
7.7. Appropriate Model of Population Divergence	251
7.8. Estimation of Population sizes	253
7.8.1. <i>Potential Vectors for Gene Exchange</i>	253
7.8.2. <i>Low Recruitment in a High Dispersal Species?</i>	255
7.8.3. <i>Spatial Distribution of Migrants</i>	257
CHAPTER 8	260
Conclusions	260
CHAPTER 9	265
REFERENCES	265

LIST of FIGURES

- Figure 2.1. The relative position and the orientation all of the primers used in this study. Dashed arrows indicate fragment size. bp = base pairs.
- Figure 3.1. An example of the *Euphausia lucens* PCR-SSCP profiles showing a 156-bp fragment of mitochondrial *NDI* among individuals collected from Argentina. Haplotype labels correspond to the designations from table 3.1.
- Figure 3.2. An example of the *Euphausia lucens* PCR-SSCP profiles generated for the southeast Atlantic (South African) samples. Again the haplotype designations correspond to those presented in Table 3.1.
- Figure 3.3. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. lucens* mtDNA haplotypes (filled coloured circles) based on *NDI* sequences: orange, Argentina, green, South Africa and black, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency. Red diamonds indicates missing intermediates. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Figure 3.4. The minimum evolution tree for *E. lucens* ND1 sequences using the most appropriate substitution model as determined hLRT using ModelTest (see text). Trees were rooted with four *E. vallentini* sequences. Numbers above branches represent bootstrap values.

Figure 3.5. The maximum likelihood tree for *E. lucens* ND1 sequences using the most appropriate substitution model as determined hLRT using ModelTest (see text). Trees were rooted with four *E. vallentini* sequences. Numbers above branches represent bootstrap values.

Figure 3.6. Map showing the three sampling locations for *Euphausia lucens* over the Patagonian Continental Shelf in the southwest Atlantic. The pie charts show the frequency of *NDI* haplotypes among samples. The same haplotype (B1) was the most frequent in each of the separate samples. Sample numbers shown in pie charts.

Figure 3.7. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. lucens* mtDNA haplotypes (filled coloured circles) based on *NDI* sequences with the loops resolved using various coalescent criteria (see text).

Figure 3.8. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of *E. lucens* mtDNA haplotypes (filled coloured circles) rooted with the sister group *E. vallentini* based on *NDI* sequences: grey, Argentina, red, South Africa and green, New Zealand. Red diamonds indicates missing intermediates. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Figure 3.9a-b. Haplotype frequency spectrum (*NDI*) for *E. lucens* populations a) SEA b) SWA.

Figure 3.10a-d. The estimates of θ (a-b) and g (c-d) obtained with Fluctuate for *E. lucens* with the 95% confidence limits. a-c) SWA b-d) SEA. The initial g values for runs 1-10 correspond to those in Table 3.9.

Figure 4.1. An example of the *NDI* PCR single strand conformational polymorphism profiles generated for the *E. vallentini* samples. The haplotype labels (I-IV) correspond to those in Table 4.1

Figure 4.2. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. vallentini* mtDNA haplotypes (filled coloured circles)

based on *NDI* sequences: black, Argentina, blue, Marion Island, red, Sub-Antarctica and white, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency and the lowest frequency was 1 (haplotype III). Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Figure 4.3. The *COI* maximum parsimony tree for *E. lucens* and *E.vallentini*. This was one of the 32 most parsimonious trees found of length 166 steps. The numbers at the nodes are the bootstrapped values calculated with 4000 replicates. An identical tree was obtained with the maximum likelihood method (not shown). E.val* = gi|5853132|gb|AF177188.1| and E.LUC* = |5853129|gb|AF177185.1 ELSW = SW Atlantic *E. lucens* ELSE = SE Atlantic *E. lucens*.

Figure 4.4. Neighbour-joining tree for tranversions only inferred for the mitochondrial large subunit (16S) rRNA sequences. *Squilla spp* was used to root the tree. Numbers at the nodes represent bootstrapped support calculated with 1000 replicates. The *reticulatum* and *sulcatum* groups contain species used in calibration of the molecular clock.

Figure 4.5a-b. The marginal posterior probability distributions for model parameters a) population size (θ) b) the population

divergence time (T) (scaled by the neutral mutation rate). Curves are shown for the analysis of *E. lucens*-*E. vallentini* *NDI* data.

Figure 4.6. Consensus sequences of partial 18SrRNA (-95 to 0), ITS-1 (starting from 1(and partial 5.8S rRNA underlined) from 10 individuals of each of *E. lucens* and *E. vallentini*.

Figure 6.1. PCR-SSCP profiles showing a 156-bp fragment of mitochondrial *NDI* for a single *Euphausia recurva* sample collected from the west coast of South Africa. The labels correspond to the *NDI* haplotype designations presented in Table 6.1.

Figure 6.2. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. recurva* mtDNA haplotypes (filled coloured circles) based on *NDI* sequences: black = South Africa, yellow = Indo-West Pacific, blue = northeast Pacific and red = New Zealand. The outgroup sequence *E. lucens* is shown in green. The size of each circle is proportional to the corresponding haplotype frequency. Red diamonds indicates missing intermediates. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Figure 6.3. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of all *E. recurva* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences with the loops resolved using various coalescent criteria (see text). Again: black = South Africa, yellow = Indo-West Pacific, blue = northeast Pacific and red = New Zealand. The outgroup sequence *E. lucens* is shown in green.

Figure 6.4. The distribution of *E. recurva* *ND1* haplotypes around southern Africa

Figure 6.5a-b. The results from independent runs of Fluctuate for the *E. recurva* IWP sample with differing initial values of *g*. Runs 1-11 correspond to the initial values of *g* given in the leftmost column in Table 6.10.

Figure 6.6 a-b. The results from independent runs of Fluctuate for the *E. recurva* SEA sample with differing initial values of *g* (see table 6.10. for details).

Figure 6.7. The maximum parsimony majority rule consensus tree for *COI* sequences of the three euphausiid species. Haplotype labels are as before with the prefix R = *E. recurva*, L = *E. lucens*. The numbers above the branches represent bootstrap values calculated with 1,000 replicates. The tree is rooted with the outgroup sequence *Euphausia pacifica*. The

sequence labelled *E. lucens*- again groups within the *E. vallentini* clade.

Figure 6.8. The maximum likelihood tree for *COI* sequences of the three euphausiid species. Haplotype labels are the same as those employed for ND1. With the prefix R = *E. recurva*, L = *E. lucens*. The numbers above the branches represent bootstrap values calculated with 1,000 replicates. The tree is rooted with the outgroup sequence *Euphausia pacifica*. The sequence labelled *E. lucens*- again groups within the *E. vallentini* clade.

Figure 6.9. Alignment of the *E. recurva ITS-1* sequences. The variable sites respectively are marked with a single and double asterisk.

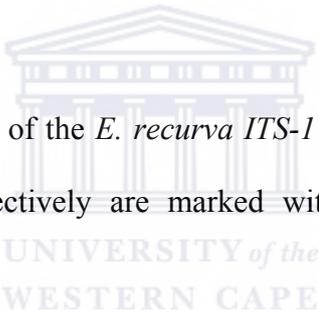
The logo of the University of the Western Cape, featuring a classical building with columns and a pediment, with the text 'UNIVERSITY of the WESTERN CAPE' below it.

Figure 6.10. Maximum parsimony tree for the *E. recurva ITS-1* sequences with the gaps treated as a fifth character state. Numbers on branches represent bootstrap values with 1000 replicates.

Figure 6.11. Maximum parsimony tree for the *E. recurva ITS-1* sequences with the gaps coded using the simple gap coding method. Numbers on branches represent bootstrap values with 1000 replicates.

Figure 6.12. Maximum parsimony tree for the *E. recurva ITS-1* sequences with the gaps omitted. Numbers on branches represent bootstrap values with 1000 replicates.

Figure 6.13. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. recurva* nuclear sequence variants (filled coloured circles) based on *ITS-1* sequences: black, South Africa, yellow, Indo-West Pacific, blue, northeast Pacific and red, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency. Red diamonds indicates missing intermediates. Numbers in red on each branch indicate the nucleotide site where the substitution occurred.

Figure 6. 14. The alignment of the *ITS-1* sequences for the three euphausiid species

LIST of TABLES

- Table 2.1. The locations, dates and number of individuals of each of the three euphausiid species (*E. lucens*, *E. recurva* and *E. vallengini*), investigated here.
- Table 3.1. Haplotype frequency distribution for ND1 sequences of *E. lucens*.
- Table 3.2. Summary statistics for the individual samples collected in the different geographical regions.
- Table 3.3. Results of the coalescent analysis of *E. lucens* populations based on NADH1 (ND1) sequence data. Theta, the migration parameter (N_eM) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see text for details).
- Table 3.4. Patterns of nucleotide change in *NDI* mtDNA sequences of *E. lucens*.

- Table 3.5. Frequency of mutations of various sizes in *NDI* mtDNA sequences of *E. lucens* for the entire sample.
- Table 3.6. Frequencies of mutations of various sizes in the *E. lucens* in the *NDI* mtDNA sequences for the regional sub-samples.
- Table. 3.7. The results of the neutrality tests based on *NDI* for the various *E. lucens* data sets. %e = percentage external mutations, %s = percentage on singleton mutations.
- Table. 3.8. The estimates of θ for the *E. lucens* samples using a range of different estimators.
- Table 3.9. Results from Fluctuate for *NDI* for the south Atlantic populations of *E. lucens* under a model of exponential growth and varying the initial growth parameter. Column one contains the initial growth values for each of the independent runs.
- Table 4.1. Positions of the variable sites in *E. vallentini NDI* sequences.
- Table 4.2. The *NDI* summary statistics of diversity for *E. vallentini* samples. None of the neutrality tests were significant with $p > 0.10$ obtained for all the tests.

- Table 4.3. The variable sites for *E. vallengini* and *E. lucens* for *COI*.
ELSA = *E. lucens* from South Africa, ELSE = *E. lucens* from Argentina.
- Table 4.4. Table showing the change in summary statistic measures of *COI* variation within *E. vallengini* when the mtDNA sequence gi 5853132 (Jarman *et al.* 2000a) is excluded.
- Table 4.5. Results of the coalescent analysis of the species *E. lucens* and *E. vallengini* populations based on NADH1 (ND1) sequence data. Theta, the migration parameter (N_eM) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see text for details).
- Table 6.1. Haplotype designations and their frequencies for *E. recurva*.
Nonsynonymous changes are marked in bold n =248
- Table 6.2. Summary statistic measures of diversity for *E. recurva* samples collected in the different geographical regions.

Table 6.3. Patterns of nucleotide change in *NDI E. recurva* sequences.

Table 6.4. Frequencies of mutations of various sizes in the total *E. recurva* sample based on the *NDI* sequences (n =247).

Table 6.5. The frequencies of mutations of various sizes in the regional subsamples of *E. recurva*. *K* is the number of haplotypes in each sample.

Table 6.6. The results of the neutrality tests for the *E. recurva NDI* sequence data. The significance was estimated with 10,000 coalescent simulations in DNAsp version 3 (Rozas and Rozas, 1999).

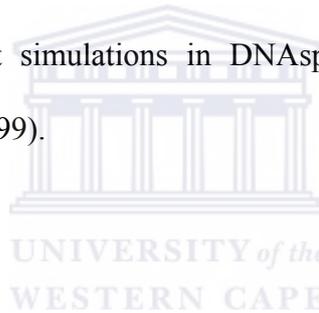


Table 6.7. Summary statistics of diversity for the *E. recurva* phylogenetic clades with Fu and Li's *F* and *D* tests with *E. lucens* as an outgroup.

Table 6.8. Estimates of θ for *E. recurva NDI* sequences of. Wattersons θ^c = Wattersons estimate with singleton mutations omitted.

Table 6.9. The effective population size estimates (in millions of individuals) for *NDI* based on the values of θ presented in Table 6.8.

Table 6.10. Results from the Fluctuate analysis varying the initial growth rate (g_0). Presented are the *NDI* results for the southern hemisphere populations of *E. recurva*. The most important results (θ and g) are highlighted in bold text.

Table 6.11. Results of the coalescent analysis of *E. recurva* populations based on NADH1 (*ND1*) sequence data. Theta, the migration parameter (N_eM) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see text for details).

Table 6.12. Variable sites for *E. recurva* using CO1. RD = *E. recurva* from the Indo-west Pacific. RA = *E. recurva* from the southeast Atlantic, RB = *E. recurva* from the northeast Pacific. Haplotype labels conform with those applied to the *ND1* sequences previously. Nonsynonymous changes are marked in bold.

- Table 6.13. Summary statistics of *ITS-1* diversity among the *E. recurva* clades.
- Table 6.14. The neutrality test results for the *ITS-1* sequences of *E. recurva*.
- Table 6.15. The theta ($\theta = 4N\mu$) estimates from the *ITS-1* sequences of *E. recurva*.
- Table 6.16. The effective population size (N_e) estimates (in millions) from the *ITS-1* sequences of *E. recurva*. S = segregating sites π = average pairwise differences.
- Table 6.17. Genetic differentiation of regional lineages of *E. recurva* based on 640 base pairs of *COI*, 156 base pairs of *NDI* and 384 base pairs of *ITS-1*. The pairwise difference between clades is given in brackets. Also given are the number of fixed differences f (brackets) and the percentage of fixed differences between clades

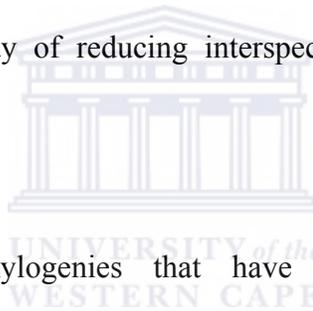
CHAPTER 1

1.1. General Introduction

Euphausiids are holoplanktic crustaceans and can be found in the pelagial throughout the world's oceans. Individual species, like many marine organisms, are often distributed in conspicuous latitudinal bands (Mauchline and Fisher 1969), the boundaries of which are thought to be strongly delimited by water temperature (Mauchline and Fisher 1969). The Order Euphausiacea comprises some 86 species, nine of which have circumglobal distributions in the Southern Hemisphere, and although most have oceanic distributions where numbers of individuals are thought to be low (Mauchline and Fisher 1969) they often occur at high abundance over continental shelves where they make an important contribution to local neritic ecosystems (Mauchline and Fischer 1969; Gibbons *et al.* 1999). One characteristic of this group is the tendency to form swarms; often of great magnitude comprising billions of individuals and consequently commercial fishing operations have existed since the 1950's to exploit these valuable resources in Antarctica (Bakun 1996).

Most euphausiids are distributed widely in the marine realm and are frequently associated with particular water masses or environments (Gibbons *et al.* 1999) and have been used to track water movement. Euphausiids are truly omnivorous although different genera tend toward carnivory or herbivory (Roger 1973). Selective feeding has been reported for some euphausiid species

(Stuart 1989; Gibbons *et al.* 1991b) and it appears that the food environment dictates switches between herbivory and carnivory. Many species of euphausiid display pronounced diel-vertical migration (DVM) and frequently traverse distances in excess of 200 m at night (Mauchline 1980). The evolution of DVM although believed to be a primary adaptation to avoid predation, is thought to confer a secondary advantage in that it also enables populations to access differential horizontal water flow in the vertical water column and so maintain themselves in productive shelf areas (Pillar *et al.* 1989; Barange and Pillar 1992). Multi-species assemblages are commonly observed at fronts and in the open ocean where it is thought that euphausiids partition the water column vertically as a way of reducing interspecific interactions (Barange 1990).



The few euphausiid phylogenies that have been constructed using morphological information (Casanova 1984; van der Spoel *et al.* 1990) produced largely incongruent results although both agree that the monospecific family Bentheuphausia is the most ancient. The deep-water cosmopolitan species *Bentheuphausia ambylops* is believed to have originated in the North Atlantic during the early Cretaceous and dispersed elsewhere at a later date (van der Spoel *et al.* 1990).

The van der Spoel *et al.* (1990) study presented an area cladogram based on morphological characters with Decapoda (which was considered to be the sister

group of the Euphausiacea in the Eucardida at the time), as an outgroup. However, a more recent systematic investigation of the sub class Eumalacostraca based on 28S rRNA phylogenies by Jarman *et al.* (2000b) has since shown that the sister group to Euphausiacea is Mysida (as was believed for most of the 19th century) and not the Decapoda as previously thought. If the relationships based on 28S rRNA are correct it implies that the polarization of many of the morphological characters in the area cladogram by van der Spoel *et al.* (1990) likely contain errors that will affect the original biogeographical conclusions of that study.

It is traditionally held that longitudinal, rather than latitudinal gene flow will dominate in circumglobal marine species that exist in a continuous environment with limited barriers to dispersal (Palumbi 1994; Jarman *et al.* 2000a; Van der Spoel and Pierrot-Bults 1990). Furthermore, the lack of physical barrier to exchange in the sea, and the fact that the larvae of euphausiids have a free-swimming stage, predict dispersal rates to be high. High levels of gene flow will act to prevent the isolation of regional populations and consequently, the degree of divergence between regional samples separated by vast distances of open ocean are expected to be low (Palumbi 1992). Although significant population structure has been observed for planktic copepod species with open ocean distributions (Bucklin 1996; Goetze 2003) species-specific differences are also common even between sibling species currently sympatric across at least part of their distributional

range (Goetze 2003). This pattern is likely reflects the extremely complex and dynamic evolutionary history of individual species.

Direct tracking of the movement of large-bodied animals such as fish using mark and recapture methods can be used to obtain direct measures of contemporary population parameters such as population size, population turnover rates and rates of emigration/migration (Kitagawa *et al.* 2000). However, the small size and planktic habit of euphausiids precludes the use of such methods and in the absence of direct information, knowledge of the spatial and temporal scales on which oceanographic processes may operate to shape the population distribution patterns in this marine group remain largely unknown. In such cases, an indirect population genetics approach may be adopted to investigate the complex interactions between genetic material in marine species and the environment. Molecular methods have been applied to investigate the demography, geographical structure, and phylogeography and speciation mode of a wide range of marine organisms from blue whales through fish to microscopic foraminiferans (Palumbi 1990; Darling *et al.* 2000, 2004; Arnason 2000; Peijnenburg *et al.* 2005; Papadopoulos *et al.* 2005; Papetti *et al.* 2005).

The field of population genetics has been revolutionized with the recent development of coalescent theory (Kingman 1982; Hudson 1990; Fu and Li 1999; Vasco *et al.* 2001; Rosenberg and Nordborg 2002). The ‘coalescent’ is

fundamentally a stochastic model that allows the construction of random genealogies to provide a coherent statistical framework for the analysis of genetic polymorphism data (Rosenberg and Nordborg 2002). Coalescent-based methods attempt to take into account the uncertainty caused by the inherent randomness of evolution and thus aim to avoid the pitfalls of data over-interpretation, which is inherent with the application of heuristic models previously borrowed from phylogenetics (Rosenberg and Nordborg 2002).

Developments in the application of the coalescent have been driven by the rapid accumulation of DNA sequence data, and enable inferences to be made of events that happened in the past (Fu and Li 1999). This theory is now recognized as the cornerstone for various statistical analyses of molecular population samples (Fu and Li 1999; Rosenberg and Nordborg 2002) and the coalescent framework is fundamental to many of the methods of analysis employed here. However, it is only recently that coalescent-based methods have been applied to estimate population parameters in marine plankton (Goetze 2003; Peijnenburg *et al.* 2005; Papadopoulos *et al.* 2005), and although several molecular studies have involved euphausiid species (covered in more detail below), this approach to population parameter estimation remains largely under-utilized.

1.2. Previous Euphausiid Molecular Population Genetic Studies

Although it is more than 30 years since the first molecular population genetic study of a euphausiid species was conducted (Ayala *et al.* 1975), subsequent investigations have been relatively few and mostly involving Antarctic species. The earliest studies used allozymes and reported significant differences in allele frequencies among different geographical locations around Antarctica (Fevolden and Ayala 1981; cited in Fevolden and Scheppenheim 1989) that were not supported by several subsequent investigations (Scheppenheim and MacDonald 1984; Fevolden 1986; MacDonald *et al.* 1986; Fevolden and Scheppenheim 1989). A similar study using allozymes compared samples of *Euphausia krohni* and *Nematoscelis megalops* in North Atlantic slope waters and reported significant among sample heterogeneity of allozyme frequencies indicating both species are genetically structured Bucklin and Weibe (1986). More recently, genetic studies of krill have involved mitochondrial DNA sequence data.

The Antarctic species *Euphausia superba* was investigated at the DNA level using part of the mitochondrial *ND1* gene, in four samples collected in the different sectors of Antarctica by Zane *et al.* (1998). Two interesting findings emerged from this research. First, despite having employed very much smaller sample sizes compared to previous studies, haplotype diversity was relatively high, and significant differences were found in pairwise tests between locations separated by the Scotia-Weddell Confluence, a well-documented putative

oceanic barrier to dispersal (Zane *et al.* 1998). Second, *E. superba* populations were not at mutation-drift equilibrium and estimates of effective population size (N_e) were several orders of magnitude lower than census sizes (interpreted as a population expansion signature). The results suggested the possibility of a retention mechanism that could generate separate stocks of krill. However, the limited number of samples used in this study and the fact they were collected in different years, weakened any conclusions that could be drawn from such findings.

Another population genetic study of the Antarctic krill species *Euphausia crystallorophias* (Jarman *et al.* 2002) was conducted using mitochondrial DNA sequence data for a 616 base pair region of cytochrome oxidase subunit 1 (*COI*), and single strand conformational polymorphism (SSCP) analysis. Specimens were collected from three different regions of the Antarctic coastline, two from the Mertz Glacier Polyna, and a single sample from both the western side of the Antarctic Peninsula and from the Davis Sea. Significant genetic differences between krill samples were identified, the extent of which did not correlate with the degree of geographical separation between sampling sites. The major source of differentiation was a sample collected from the Mertz Glacier Polyna, which differed from all other samples including a second sample taken 16 km apart in the same area. In contrast, no differences were observed between the two most distantly separated samples collected in the West Antarctic Peninsula and the Davis Sea that were separated by > 6000 km.

The authors postulated that this much-unexpected finding might be attributed to genetic differences between individual swarms. Currently investigations are underway in Antarctica to determine whether genetic differences between individual swarms can be detected in this species using nuclear markers, and the first microsatellite molecular markers for a euphausiid species have been designed to achieve this objective (T. Patarnello personal communication).

Four separate investigations of the Northern Hemisphere euphausiid species *Meganyctiphanes norvegica* (Sundt and Fevolden 1996; Bucklin *et al.* 1997; Zane *et al.* 2000; Papetti *et al.* 2005) have been conducted. The first (Sundt and Fevolden 1996) used protein polymorphisms and reported genetic homogeneity among eight different geographical sampling localities in the Northeast Atlantic Ocean. The second (Bucklin *et al.* 1997) used a fragment of the mitochondrial cytochrome *b* (*Cyt b*) and cytochrome oxidase subunit 1 (*COI*) genes and investigated the NW and NE Atlantic regions (including samples collected in a Norwegian fjord). This study also found evidence of significant spatial structure that was mainly associated with the Norwegian fjord sample, and the weak genetic differences that were observed within the NW samples were considered to be due to the temporal effects of the sampling regime. The third study (Zane *et al.* 2000) again involved the North Atlantic region but was extended to include the Mediterranean Sea. Using a fragment of the mitochondrial gene *ND1*, and 385 individuals from eight population samples, these authors reported strong deviations from panmixia. This deviation was

entirely due to two samples from the Ligurian Sea and the Cadiz Sea that differed significantly from all other samples. No evidence of any further genetic heterogeneity was observed among the North Atlantic samples or those collected in the Skaggevatn Fjord in successive years (Zane *et al.* 2000). The greatest differences between samples of *M. norvegica* were between the separate NE and NW Atlantic basins. The most recent study confirmed that the geographical patterns of differentiation observed previously were stable on both a temporal and spatial scale (Papetti *et al.* 2005). The three distinct genetic pools identified in *M. norvegica* within the North Atlantic Ocean (Bucklin *et al.* 1997; Zane *et al.* 2000; Papetti *et al.* 2005) also appear to correspond to the ocean-basin scale patterns of circulation that partition the North Atlantic into three gyres which are thought to effectively isolate and entrain populations (Papetti *et al.* 2005). A fourth population was again indicated for the Ligurian Sea samples suggesting a separate Mediterranean Sea population (Papetti *et al.* 2005).

It has been suggested by Zane and Patarnello (2000) that the greater population structure observed in the northern krill *Megacystiphanes norvegica* (Bucklin *et al.* 1997; Zane *et al.* 2000; Papetti *et al.* 2005), compared with the Antarctic krill *Euphausia superba* (Zane *et al.* 1998) is partly explained by the different geographical size of the distribution ranges and hydrological conditions. The distribution of *M. norvegica* spans a larger geographical area, characterized by an oceanographic pattern that could actually restrict the gene flow between

different regions, whereas strong circumpolar currents that provide extensive mixing of Antarctic waters characterise the Southern Ocean (Zane and Patarnello 2000). However, the differences observed among samples of *E. crystallorophias* in Antarctica by Jarman *et al.* (2002) cannot be readily explained by the geographical size of the distribution range of this species nor contemporary hydrological conditions.

1.3. Previous Euphausiid Molecular Systematic Studies

In addition to the population genetic studies discussed previously, phylogeographic investigations using DNA sequence data have been conducted in attempts to elucidate the larger biogeographic patterns in euphausiids. Unsurprisingly, all three investigations were centred upon Antarctic and sub-Antarctic species and compared phylogenies based on *16S* and *COI* subunit 1, molecular markers. The first study (Patarnello *et al.* 1996) employed a phylogenetic and molecular clock approach for the mitochondrial marker *16S rRNA*, and the results suggested that the Antarctic and sub-Antarctic lineages are genetically distinct and that the break is linked to the formation of the Antarctic Convergence around 22 MYA (Barker and Burrell 1997; Rabassa *et al.* 2005; cf. Barret 2001; Lawver and Gahagan 1998). The second investigation (Patarnello *et al.* 2000) included information from another mitochondrial marker *ND1* and was expanded to include more species but it confirmed the general findings of the Patarnello *et al.* (1996) study.

A similar study by Jarman *et al.* (2000a) investigated the speciation history of seven euphausiid species with circumglobal distributions in the Southern Hemisphere using a phylogenetic and molecular clock approach for the mitochondrial markers *16S* rRNA and *COI*. The estimates of divergence times reported in this study between Antarctic and sub-Antarctic euphausiid species (~ 15 MYA), although, close to the putative estimated time of the formation of the Antarctic Convergence displayed very wide 95% confidence limits (11-25 MYA). A phylogenetic tree constructed using *16S* rRNA and *COI* revealed that *E. vallentini* and *E. lucens* had strong statistical support for their close relatedness, giving a mean divergence time estimate of 1.04 million years ago.

Jarman *et al.* (2000a) argued that since ocean circulation conditions have been stable for at least two million years, based on evidence from fossils of microplankton foraminifera (Murray 1995), the genetic differentiation between *E. vallentini* and *E. lucens* must have occurred without the dispersal barriers associated with oceanic fronts. Thus implying that speciation was either sympatric or parapatric or, that these taxa had been incorrectly diagnosed as separate species. The authors suggested that the low molecular divergence estimates were correlated with the small number of diagnostic morphological characters separating the species in the species inference key of Mauchline and Fisher (1969), in which a single diagnostic morphological character (the shape of the antenular peduncle) is used for species identification. Quite why the authors disregarded information from a more recent species identification keys

(Baker *et al.* 1990) that included a second diagnostic morphological character (a conspicuous spine, located dorsally on the third abdominal segment, and present in *E. vallentini* only) that can be used to discriminate these species, while citing it in their reference list is unclear. It should be noted that rough handling of specimens in collection or storage could cause this spine to break off, making correct species identification difficult.

1.3.1. Limitations to the Estimation of Molecular Population Parameters

Only two of the research initiatives discussed above involving the analysis of intraspecific data, used sequence data *and* contained a temporal element (Zane *et al.* 2000; Papetti *et al.* 2005). Furthermore, none of the studies involved comparisons between independent, unlinked loci and could not therefore; reliably discriminate between the effects of population processes that should affect all loci similarly and selection, which should be locus specific. Thus, although significant deviation from neutrality has been indicated in several studies, the cause for such deviation remains ambiguous, as separating the confounding effects of demographic and selective processes is not possible with single-loci data (and sometimes not even when multiple loci are investigated (Akey *et al.* 2004).

1.4. Gene Flow and Barriers within the Marine Realm

Early suggestions that hydrological barriers act as effective barriers to exchange between different oceanic regions (Kennet 1998) are now being challenged by evidence of recent and frequent dispersal of several planktic taxa across such strong boundaries (Darling *et al.* 2000, 2004; Goetze 2003). These barriers include the formation of the Antarctic Circumpolar Current around 22 MYA (Barker and Burrell 1997; Rassaba *et al.* 2005; Patarnello *et al.* 1996, 2000; Jarman *et al.* 2000a; cf. Barret 2001; Lawver and Gahagan 1998) or tectonic barriers such as the final closure of the Isthmus of Panama around 3 MYA (Knowlton *et al.* 1993; Knowlton and Weigt 1998; Schubart *et al.* 1998; Marko 2002). The finding that supposed barriers to long-distance dispersal are far more permeable than once believed has led certain authors to argue that the role of physical and hydrological barriers to exchange as a driver of evolution has been overestimated (Norris 2000). Instead, it has become clear that continental landmasses act as a barrier to longitudinal dispersal in some zooplankton species only, and species-specific patterns of gene flow on circumglobal scales may be expected, even among sibling species (Goetze 2003).

An alternative view to the dispersal-limited models of speciation, holds that a species distribution may be more limited by its ability to successfully establish itself in new location where conditions are favourable, as evidenced by the fact that many pelagic marine species are limited to only part of the ocean despite

indications of regular dispersal outside their home range (Norris 2000). An example of such a case involves the foraminifer species *Pulleniatina obliquiloculata* that is episodically eliminated from the Atlantic during glacial cycles but reinvades around southern Africa during interglacial periods (Norris 1999). Such findings support the proposal that biogeographic ranges are determined more by where pelagic populations can maintain viable populations, than by dispersal, and speciation models than do not rely on limiting dispersal are probably more satisfactory for explaining pelagic diversity (Norris 2000). This further suggests that open-ocean taxa have narrower tolerance limits than has generally been appreciated (Norris 2000).

It is clear that a general mode of speciation is not apparent in euphausiids, with a recent phylogeographic study invoking four different modes of speciation to explain the observed patterns of relationships (Jarman *et al.* 2000a). Thus, although it is generally held that historical biogeographic factors are relatively more important in structuring genetic diversity (Knowles 2004), species-specific ecological and behavioural characteristics likely play an important role in this respect in euphausiids, and many other marine taxa (Darling *et al.* 2000, 2004; Norris 2000).

1.5. Euphausiid Systematic and Population Studies in Perspective

It can be seen that molecular investigations of Southern Hemisphere euphausiid species have not been spread evenly across the region, with the Southern Ocean species receiving the lion's share of interest. Whilst the focus on Antarctica partly reflects a management need for information on the commercially exploitable *Euphausia superba* (Miller and Hampton 1989), it can also be viewed as being driven by a political willpower to be seen to be involved in Antarctic research. This effort has unfortunately resulted in an ignorance concerning the euphausiid population genetics elsewhere in the region, and this imbalance needs to be redressed if we are to get comparable information therefrom (Gibbons *et. al.* 1999). A similar situation is seen in the Northern Hemisphere where several investigations of the species *M. norvegica* involving mtDNA have been carried out to the almost complete exclusion of any other euphausiid species from this area.

1.6. Relevance of Studying Temperate and Sub-Antarctic Euphausiid Species

The main frontal systems in various sectors of the Southern Ocean show marked spatial and temporal variability in their importance as biogeographic barriers to the distribution of plankton and as areas of high productivity (Lutjeharms 2006). This variability has been attributed to mesoscale changes in the physical environment including meanders across fronts, the formation of eddies and cross frontal mixing (Bernard and Froneman 2003). Such processes

likely facilitate the transfer of plankton across the fronts. It is generally believed that absence of biological enhancement in the frontal waters reflects the temporal variability in the stability of the water column (Lutjeharms 2006). Global climate change will likely result in shifts in the intensity and geographical position of major frontal systems and are thus likely to coincide with alterations in the distributions of species and productivity in the Southern Ocean (Lutjeharms 2006).

The recurrent advance and retreat of land and sea ice and the accompanying climatic changes have been of great influence for the present distribution pattern of organisms inhabiting terrestrial and aquatic habitats (Hewitt 1996; Brunner *et al.* 2001; Luttikhuizen *et al.* 2003) During interglacials the area of suitable habitat will have greatly expanded, creating opportunities for population subdivision. Glacial maxima will be accompanied by a compression of the distribution area. Range contractions and latitudinal displacement of populations directly linked to previous glacial dynamics have been described for other zooplankton species (McIntyre *et al.* 1989; Bucklin and Wiebe 1998; Norris 1999). Consensus has emerged that many species are expanding their range towards the poles in response to climate change (Thomas *et al.* 2006). Therefore, any contraction of euphausiid species ranges in Antarctica due to a reduction in the extent of ice cover and increases in water temperature may be expected to result in concomitant southward expansion in temperate species boundaries currently situated north of the Antarctic Convergence. Thus, the

population genetics study of circumglobally distributed euphausiid sister species might elucidate biological responses to past climate change that may help in understanding the likely response of such species to any future temperature rises due to global warming.

1.7. Species Investigated

To address some of these shortcomings, the molecular diversity was studied in three species of euphausiid (Malacostraca: Crustacea) that have a center of distribution in the Southern Hemisphere: *Euphausia lucens*, *E. recurva* and *E. vallentini* using both nuclear and mitochondrial markers previously shown to be appropriate for investigations of both inter- and intraspecific molecular systematic studies in Crustaceans. The three species differ in their latitudinal range, geographical distribution and ecological habitat “preferences” and several life history characteristics including number of generations per year and rates of egg production. A similar approach was adopted to study the molecular diversity of two copepod species in the North Atlantic (Bucklin and Weibe 1998). All three euphausiid species have a circumglobal distribution in the Southern Hemisphere; and *E. lucens* and *E. vallentini* are absent from Northern Hemisphere where *E. recurva* is distributed in the North Pacific Ocean (Mauchline and Fisher 1969; Brinton 1964).

With information from both mitochondrial and nuclear DNA markers it is hoped the problems of reliably discriminating between the effects of historical

population processes and selection that affect single-locus studies can be circumvented. More accurate estimates of population parameters may be obtained using coalescent methods for variable environments (Kuhner *et al.* 1998; Vasco *et al.* 2001) when the basic assumption of constant population size is not justified (Vasco *et al.* 2001; Hudson 1990) overcoming the potential problems associated with incorrect model specification. In the following paragraphs, I now present a separate brief introduction for each species, including specific details of the various hypotheses I would like to test.

1.7.1. *Euphausia lucens* (Hansen 1905)

Euphausia lucens is distributed in temperate waters of the Southern Hemisphere between 28 ° S and 55 ° S. The Antarctic Convergence limits its southern distribution and the Subtropical Convergence its northern distribution (Lomakina 1964; Mauchline and Fisher 1969). Although not considered an abundant species across its distributional range (Mauchline and Fisher 1969), *E. lucens* is an extremely important species in the neritic zones around the various landmasses in the Southern Hemisphere where local environmental conditions may differ considerably. It is considered a sub-Antarctic species (Mauchline and Fisher 1969), although it reaches its northernmost point in the southern Benguela region, where the upwelling waters extend the cold-water temperature regime more typical of the core of its distribution. This equatorward latitudinal extension of this species' distribution in the S. Atlantic is a general pattern for cold-water euphausiid assemblages in this region and is

thought to reflect the influence of the South Atlantic Gyre (Gibbons *et al.* 1999). In a similar fashion, *E. lucens* is also distributed further north of the core of its distribution area with the Peru and Falkland Currents (Zimmer 1914; Hansen 1915; Colossi 1917; Montu 1977, 1982; Ramirez and Dato 1983; Curtolo *et al.* 1990; Tarling *et al.* 1995).

In the southern Benguela ecosystem, *E. lucens* dominates the zooplankton assemblages in the neritic zone all year round (Boden 1954, 1955; Nepgen 1957; Pillar *et al.* 1989) playing a key role in trophic food webs and forming an important component of the diets of a number of commercially valuable fish species including Hake (Botha 1980; James 1987) and Kingklip (personal observation). In contrast, off the south coast of South Africa this species is uncommon and is confined to the cold inner-shelf regions mainly due to the influx of warm Agulhas Current waters further offshore (Gibbons 1995).

Females of this species in the Southern Benguela Region may be reproductively active for up to nine months, during which time they may release a total of 135 broods (Stuart 1992). However, the number of broods is likely overestimated, as this value was obtained from observations of a single individual maintained in an aquarium under optimal conditions (Stuart 1992). Such conditions are far removed from those experienced in the southern Benguela Region where a highly dynamic and variable seasonal upwelling

regime results in pulses of productivity and consequently, an ephemeral supply of prey species for *E. lucens* (Gibbons *et al.* 1991).

In the SW Atlantic *E. lucens* is present over most of the southern Patagonian continental shelf, except for the innermost sector north of 42 °S, and provides a valuable food resource to squid, whales, and penguins (Tarling *et al.* 1995). In contrast to the situation seen around South Africa, seasonal shifts in euphausiid species dominance are observed in this region where *E. lucens* adults and larvae dominate during the summer, and to a lesser extent, autumn, when *E. vallentini* is more abundant (Curtolo *et al.* 1990; Ramirez and Dato 1983; Tarling *et al.* 1995). Distribution patterns of euphausiids in this area are mainly influenced by the Malvinas Current (Curtolo *et al.* 1990), and the seasonal changes in species composition and abundance in this region are thought to reflect physical changes in circulation as much as they do to physiological responses by different species to changes in the environment (Gibbons *et al.* 1999). Previous studies show that the distributions of euphausiid species in this region are influenced to a variable degree by sub-Antarctic waters (Curtolo *et al.* 1990). Three main water masses can be identified, the boundaries of which vary seasonally and are difficult to define. These include the Malvinas Current, formed purely by sub-Antarctic waters that branch off from the westwind-drift. The core of this current moves along the edge of the continental slope in a NNE direction, up to 34-35 °S when part of the stream changes direction and turns eastward before mixing with tropical waters to form the transition zone.

Between this current and the coast, lie the coastal waters whose origins include the Malvinas Current and South Pacific waters advected via the Strait of Magellan.

The available evidence from studies of this species on the Patagonian shelf strongly imply that prolonged periods of reproduction unlikely occur, and that the different reproductive biology of this species in the southern Benguela region (Pillar and Stuart 1988) may be due to the greater feeding potential experienced in the more productive upwelling waters (Tarling *et al.* 1995). The average annual surface water temperature in the southern Benguela is approximately 16°C ranging between 12°C and 20°C (Shannon 1966) substantially higher than those experienced on the Patagonian shelf that average 10°C with a range of 7-13°C (Tarling *et al.* 1995).

Current closure schemes of varying dimensions have been proposed as a putative mechanism whereby *E. lucens* can remain and develop high concentrations in both regions (Pillar *et al.* 1989; Ramirez and Dato 1983). Off the west coast of South Africa, *E. lucens* blooms in the southern Benguela and exhibits a series of behaviors (diel vertical migration, selective omnivory, buoyant eggs) that should allow populations to be locally maintained. However, it is unknown whether populations in this region are actually self-maintained or whether they are continually seeded from the south, boom on the journey north, and then go bust during entrainment into the south Atlantic

Gyre. Similarly in Argentina, the Gulf of San Matias has been identified as an area where several euphausiid species may complete their life cycles without being expatriated as evidenced by hosting the most complete series of development stages (Curtolo *et al.* 1990).

Currently, there is a continuous route for gene exchange for *E. vallentini* individuals between locations separated by thousands of kilometers. A long pelagic larval phase (Mauchline 1980) coupled with strong Antarctic circumpolar currents could potentially carry individuals between shelf areas even across the enormous distances encompassed by the southern ocean. However, while the ACC could also potentially transport *E. lucens* individuals in an easterly direction across the South Atlantic, it might be situated too far south to act as an efficient dispersal vector into the lower latitudes of the Southern Benguela system around southern Africa. Although episodic inputs of sub-Antarctic water into the Southern Benguela System occur (Shannon *et al.* 1989) as a secondary circumstance of the shedding of Agulhas rings (Lutjeharms and van Balleygooyen 1988), such northward displacements of sub-Antarctic waters are rare. The ACC is therefore, probably not an efficient transport vector for *E. lucens* individuals between neritic populations in the separate South Atlantic Ocean basins. Passive transport of individuals between the Patagonia and the west coast of South Africa should however be possible at lower latitudes in the surface waters of the southern arm of the South Atlantic Gyre. The path of this current flow at latitudes midway between the subtropical

and Antarctic Convergences and feeds directly into the southeast Atlantic waters adjacent to southern Benguela Current (Pinot 1975). Movement of oceanic waters inshore occurs regularly on the narrow shelf at the Cape of Good Hope and results in short-lived increases in species richness of coastal euphausiid communities (Gibbons 1999).

Transport across the South Atlantic Ocean in the opposite direction is theoretically possible via the northern arm of the south Atlantic Gyre. However, the average position of this current follows a path between latitudes 0° S and 15° S (Pinot 1975) and is probably inaccessible to *E. lucens* flowing as it does, at latitudes outside the northernmost distributional limit of this species (Mauchline and Fisher 1969). Although the distribution of *E. lucens* in the Atlantic extends further north there are two lines of evidence that suggests the northern arm of the south Atlantic Gyre does not act as an efficient dispersal vector. First, *E. lucens* reaches its most northern position in the Benguela Current at Luderitz on the west coast of South Africa which is situated at 27° S (Lomakina 1964) and second, the northern arm of the south Atlantic Gyre connects with coastal waters around Brazil (Pinot 1975) where *E. lucens* has never been recovered. It is most likely that the higher water temperatures experienced in the northern arm of the south Atlantic Gyre preclude the presence of this species. Consequently, it would appear that opportunities for dispersal for both *E. vallentini* and *E. lucens* individuals between the south

Atlantic basins are high from west to east and likely non-existent in the opposite direction.

While tropical and subtropical waters of the Indian Ocean penetrate the South Atlantic via the Agulhas Current, tropical and subtropical waters of Pacific origin do not move into the south Atlantic via Cape Horn, owing to both restricted circulation, and the latitudinal extension of the South American Continent (Gibbons *et. al.* 1999). It is unknown whether the southern extension of the South American continental landmass represents a barrier to dispersal between the Pacific and Atlantic Oceans for individuals of *E. lucens* as the southernmost limit of this species' distribution coincides with the average position of the Antarctic Circumpolar Current (Lomakina 1964). In the South Pacific the West Wind Drift transports sub-Antarctic water towards the extreme southern coast of South America before branching northwards to form the Humboldt Current and southward the Cape Horn Current (Silva and Neshyba 1979). A portion of the surface water penetrates into the Chilean fjords and channels (Silva *et al.* 1998). Whether this results in the transport of the cold-temperate species such as *E. lucens* between oceans is unknown, and will hopefully be answered herein.

Several ecological investigations on zooplankton communities in Pacific (Antezena 1976; Guiglielmo and Ianora 1995, 1997; Palma and Silva 2004) and Atlantic waters (Montu 1977, 1982; Ramirez and Dato 1983; Curtolo *et al.*

1990; Tarling *et al.* 1995) around South America have revealed enormous differences exist in the abundances of *E. lucens* in the coastal waters in each ocean. In Chilean waters, where extreme hydrographic and climatic conditions are experienced, this species fails to maintain significant numbers of individuals (Palma and Silva 2004), but dominates zooplankton communities for most of the year in neritic waters of the SW Atlantic around Argentina (Montu 1977, 1982; Ramirez and Dato 1983; Curtolo *et al.* 1990; Tarling *et al.* 1995). The harshness of conditions in near-shore SE Pacific waters is reflected in the composition of zooplankton communities that are characterized by monospecific populations in a wide range of zooplankton taxa (Arcos 1974, 1976; Antezena 1981, 1999a; Ahummada 1976; Guiglielmo and Ianora 1995, 1997; Palma and Silva 2004). Of the euphausiid species that have been recorded in this region, only *E. vallentini* is able to maintain high abundances there (Palma and Silva 2004). The studies that have detected the occasional presence of *E. lucens* individuals in Chilean waters considered these to be immigrants from oceanic regions (Palma and Silva 2004) based on the physical and biological characteristics of the water.

Failure to maintain a viable population in a particular area does not necessarily imply dispersal across said area does not occur. All marine zooplankton populations experience expatriation of individuals from their core area of their distribution as evidenced by the occurrence of transition zones commonly found where different water masses meet. It is likely that *E. lucens* individuals

on either side of the South American continent are occasionally transported around Cape Horn in both directions. It is possible that this species can transverse these waters and thus maintain a connection between the different oceans and therefore, it is of interest to determine whether the latitudinal extension of the South American Continent results in an efficient barrier to dispersal for Atlantic and Pacific populations of *E. lucens*. These two alternative scenarios are investigated here by examining the genetic structure on both a spatial and temporal scale, and by comparing the results with material collected from the SE Atlantic and from the SW Pacific.

Although current hydrological conditions likely have a large influence upon the contemporary distribution patterns in euphausiid species, historical events such as repeated glacial periods may have left a signature in the levels of variation within and divergence between populations. During the last glacial maximum (~ 18,000 years ago) sea temperatures around the South American Continent were far colder and populations in the Pacific and Atlantic were almost certainly separated during this period. Paleoclimatic reconstructions of the Southern Hemisphere have indicated that during this colder glacial period the average position of the Polar Front was shifted northwards, while the position of the Subtropical Front did not change (CLIMAP 1967). It is therefore likely that the most southerly-distributed *E. lucens* populations in the SW Atlantic basin suffered a greater degree of areal range reduction and latitudinal displacement to the north during this period than neritic populations at lower

latitudes in the SE Atlantic basin. This may be reflected in the amounts and levels of molecular variation in samples in the separate south Atlantic basins Bucklin and Weibe (1998).

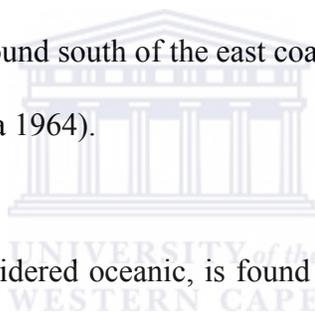
The predictions of these hypotheses are as follows:

- i) If *E. lucens* populations in the Southern Benguela Current system are seeded from the south via the southern arm of Atlantic Gyre, then fixed genetic differences between populations in the east and west boundary currents in the south Atlantic should not be observed. Furthermore, if the higher water temperatures experienced at lower latitudes precludes the transport of individuals via the northern arm of Atlantic Gyre the most ancestral sequences should be found in the source population in the SWA.
- ii) If the Southern Benguela Region acts as refuge for *E. lucens* fixed differences between these populations will be observed.
- iii) If Cape Horn represents a barrier to dispersal, then *E. lucens* populations in eastern and western boundary currents in the south Atlantic will be genetically more similar than either are to those in the SW Pacific.

1.7.2. *Euphausia vallentini* (Stebbing 1900)

Euphausia vallentini is a circumpolar species that is restricted to the sub-Antarctic zone, whose southern limit is the Antarctic Convergence and whose

northern limit lies at around 45°S (Gibbons *et. al.* 1999). In the western Pacific and the eastern Indian Ocean it is found south of the convergence (Lomakina 1964) and has also been found in the northern Antarctic region, in the Drake Passage (Lomakina 1964) and in the area of South Georgia (Mackintosh 1934, Hardy and Gunther 1935). Most reports from the Antarctic region are from the autumn period when there is a southward movement of sub-Antarctic waters in the area of the Antarctic Convergence (John 1936). A single specimen has been caught off the eastern North American coast as far north as 32°S however; all other records are south of 40°S (Mauchline and Fisher 1969). The temperature limits of this species are from 2°C to 8-10° C (Barry 1956) with the optimum around 3-6°C. It has been found south of the east coast of Australia between the limits 1° and 9°C (Lomakina 1964).



This species, although considered oceanic, is found on the continental shelf in southern Patagonia, New Zealand and the various island landmasses that occur within its circumglobal distribution. It has never been recorded in the waters around the southern African subcontinent, although questions have been raised over its validity as an independent species distinct from *E. lucens* (Jarman *et al.* 2000a), which dominates neritic euphausiid communities in this region.

In southern Patagonian shelf waters, *E. vallentini* dominates during spring, but shows equal dominance with adults of *E. lucens* in winter (Tarling *et al.* 1995). Temporal and spatial variation in the timing of reproduction appears to be a

general feature in euphausiids (Mauchline and Fisher 1969) and the results of several studies have consistently indicated that spawning off Patagonia occurs in this species during spring, after which the majority of adults are thought to die (Montu 1977, 1982; Ramirez and Dato 1983; Curtolo *et al.* 1990; Tarling *et al.* 1995). Different growth rates and generation times have been recorded between populations in the Indian Ocean, and those on the Southern Patagonian Continental Shelf (Ridoux 1988, Tarling *et al.* 1995). The average length of individuals from the Indian Ocean is much larger than those from Patagonia (17 mm vs. 24 mm respectively), and the generation times were longer in the former population. These characteristics are thought to be directly related to surface water temperature differences between the two regions (Ridoux 1988; Tarling *et al.* 1995). This species is an important food source for Blue, Fin and Sei whales in Antarctica and for the latter two species in the waters south of New Zealand (Nemoto 1962). It is also an important prey item for two of the most exploited species of cephalopods in the south Atlantic (Ivanovic and Brunetti 1994).

Johnson and Brinton (1963) considered *E. vallentini* to be a sub-Antarctic relative of *Euphausia pacifica* based on similarities in growth rates and life cycles. However, more recent phylogenies based on mtDNA molecular data for *16S* rRNA and *COI* suggest this species is more closely related to the morphologically similar *E. lucens* (Jarman *et al.* 2000a). It has been suggested that the evolution of Antarctic and sub-Antarctic euphausiid lineages may have

taken place in the mid-latitudes within the Atlantic, Pacific and Indian Oceans during period of temperature fluctuations (Brinton 1964). The landmasses would then have been barriers separating segments of the population. Alternatively, the Antarctic Convergence has been postulated as a potential barrier to dispersal that might be responsible for the separation of this and other Antarctic *Euphausia* species (Patarnello *et al.* 1996, 2000; Jarman *et al.* 2000a).

After the initial drop of temperatures during the late Pliocene, the onset of the Pleistocene (about 1.8 MYA) represents a period of large oscillations with a series of glacial and interglacial periods in a succession of 40, 80, and 120 thousand years cycles (Prof. G. Philander, Stanford University, personal communication). As *E. vallentini* has a more southern distribution compared to *E. lucens*, it is therefore likely that the former species suffered a greater degree of areal range reduction and latitudinal displacement to the north during this last glacial maxima period. This may be reflected in the amounts and levels of molecular variation in *E. lucens* and *E. vallentini*. The degree to which present day currents influence the geographical structuring of populations, and the affect of past ice ages upon the levels of molecular variation within populations, can be investigated in comparisons between *E. lucens* and *E. vallentini*. The predictions of these hypotheses are as follows:

- i) If *E. vallentini* and *E. lucens* represent separate species, fixed differences will be observed with both the mitochondrial and nuclear markers.
- ii) If contemporary hydrological conditions determine the degree of spatial genetic differentiation then *E. vallentini* should show less evidence of separate breeding pools across its distributional range than *E. lucens*.
- iii) If populations of *E. vallentini* have experienced greater range reductions and latitudinal displacement, perhaps related to past glacial dynamics, than *E. lucens*, a shallower genealogical structure and reduced molecular variation might be observed in the former species.

1.7.3. *Euphausia recurva* (Hansen 1905)

E. recurva is a warm temperate oceanic species with a bi-antitropical distribution in the Pacific; where it is restricted to the warmer regions at depths of 400m. It occurs in the central waters of the north and south Pacific between 40° and 12 ° N and 18 ° and 45 ° S (Baker *et al.* 1990). Although all euphausiids have patchy distributions where abundances differ substantially, evidence suggests that this and many other euphausiid species probably have a continuous distribution throughout their latitudinal range in the southern Atlantic (Hansen 1915). This species has not been found in the equatorial region. In a recent study of Argentinean euphausiid communities, *E. recurva* was only found in waters where the surface temperatures ranged between 17.3 and 21.8 °C in 1991 (Tarling *et al.* 1995). In the NE Pacific this species is

found in the offshore part of the California Current across the breadth of the Pacific south of 40-43° N. In the California Current this species extends to 20° N. It is numerous along the northern and eastern margins of its range and scarce in the central part of the eastern North Pacific. It was once caught at 7° N but all other records are at lower latitudes in the North Pacific (Brinton 1964). It has not been reported from the Philippines or the East Indian Archipelago. In the South Pacific it has been recorded from Chile (Brinton 1962) and New Zealand (Tattershall 1924) and from the mid-ocean (Brinton 1962).

In the southern Benguela region this species is found in greatest abundances at the shelf break, where it may dominate the euphausiid community off the west coast of South Africa (Barange *et al.* 1992). Thus, reflecting the importance of this feature as a boundary between mid-oceanic and coastal regimes (Barange *et al.* 1992). It has been hypothesized that water flow at the shelf edge around South Africa provides a means for regional communication and should result in substantial mixing of these waters, resulting in panmixia on the meso-scale (Barange *et al.* 1992; Pillar *et al.* 1991; Gibbons *et al.* 1999).

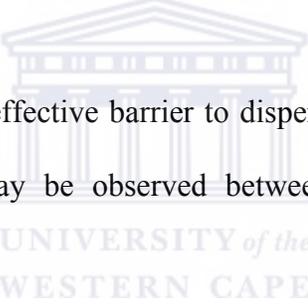
Tropical water temperatures are often lethal to temperate adapted species (Grant and Bowen 1998) forming what is generally perceived to be a formidable barrier to trans-equatorial dispersal. Antitropical distributions have been observed in a number of marine taxa including fishes (Grant and Bowen 1998), molluscs (Crame 1993, 1996), foraminiferans (Darling *et al.* 2000;

2004) and several euphausiid species (Mauchline and Fisher 1969; Brinton 1964). All antitropical euphausiids are also panoceanic occurring in Atlantic, Indian and Pacific Oceans. Three possible origins of biantitropicality have been hypothesised (Brinton 1964). The first hypothesis suggests that marine organisms traverse the tropics by submergence in deep water along isothermal surfaces with emergence in high latitudes (Brinton 1964). The second hypothesis assumes that northern and southern populations are relict bands of broad distributions that were occupied during the part of the early Tertiary epoch when the world climate was more uniform and the deep sea warmer than today (Emiliani 1954). Subsequent warming or cooling brought about a split in the latitudinal distributions. Lastly, it has been hypothesised that tropical waters were transgressed during Pleistocene periods of global cooling (Berg 1933, Hubbs 1952). Explanatory models for such distribution patterns have thus historically been either vicariant or dispersalist or some mixture of both, reflecting the prominence of this debate in the field of biogeography.

Under dispersal models, transplantation within is seen as more likely than transplantation between Hemispheres. In this interpretation *E. recurva* sequences from South Africa would be expected to be more similar to those in the South and Indo-West Pacific than either of these are to the NE Pacific. Furthermore, levels of genetic differentiation between populations are predicted to be small if dispersal involves recent or contemporary movements of individuals.

In contrast, under a vicariance scenario, Pacific *E. recurva* populations separated by the tropical equatorial waters would have been initially isolated by the breakup of Gondwanaland producing evolutionary separations of tens of millions of years. Tropical waters of the Pacific have existed for at least 20 million years (White 1994) therefore, if no contact had been made since this time, large divergences between populations centered in the Northern and Southern Hemispheres should be expected. Under this vicariant model of population separation, cryptic speciation between Northern and Southern Hemisphere Pacific *E. recurva* populations remains a possibility.

The predictions of these hypotheses are as follows:

- 
- i) If the equator is an effective barrier to dispersal the highest number of fixed differences may be observed between northern and southern Pacific samples
 - ii) If dispersal is limited by distance, fixed differences may be observed between the Southern Hemisphere populations in the Atlantic and Pacific Oceans
 - iii) If a combination of the geographical size of the distribution range, contemporary hydrological conditions and the position of the continents operate to shape the population distribution in this marine group, the greatest structure should be observed in *E. recurva*.

CHAPTER 2

Materials and Methods

2.1. Sample Collection

Samples of *Euphausia lucens*, *E. recurva* and *E. vallentini* were collected from several locations around the world and a full list of sampling locations, dates and collection methods is presented in Table 2.1. The author collected samples from neritic and offshore waters around the south and west coasts of South Africa by means of vertical Bongo nets with mesh diameters of 300 μm . The nets were deployed as close to the sea bottom as was deemed safe between the ranges of 60 m and 200 m. Immediately upon surfacing the nets were washed down with seawater to flush the catch into the collecting pouches, which were then transferred into buckets containing seawater. The seawater in the buckets was then poured through sieves (mesh size 500 μm) to remove the smaller components of the holoplankton (mainly copepods). Finally, the sieves were rinsed with a solution of 95% ethanol and the contents washed into 300 ml plastic jars containing the same for preservation, which was renewed upon arrival in the laboratory. It was important to ensure that the whole procedure was conducted as rapidly as possible to obtain samples in such a condition to facilitate successful DNA extraction. All *euphausiids* were identified in the laboratory to species using a dissecting microscope and the identification key of Baker *et al.* (1990) that uses adult morphological characters as its basis.

Location	Date	<i>n</i>	species
South Africa	2000	202	<i>E. lucens</i>
South Africa	2001	236	<i>E. lucens</i>
New Zealand	2001	3	<i>E. lucens</i>
Argentina	2003	152	<i>E. lucens</i>
California	1978	8	<i>E. recurva</i>
California	1996	33	<i>E. recurva</i>
California	1998	4	<i>E. recurva</i>
Indo-West Pacific	2001	20	<i>E. recurva</i>
South Africa	2000	115	<i>E. recurva</i>
South Africa	2001	62	<i>E. recurva</i>
Madagascar	2005	2	<i>E. recurva</i>
Sub-Antarctica	2001	125	<i>E. vallentini</i>
Argentina	2003	62	<i>E. vallentini</i>
New Zealand	2001	11	<i>E. vallentini</i>
Marion Island	2005	54	<i>E. vallentini</i>

Table 2.1. The locations, dates and number of individuals of each of the three euphausiid species (*E. lucens*, *E. recurva* and *E. vallentini*), investigated here.

A map showing the position of the sampling sites is shown in Figure 2.1.

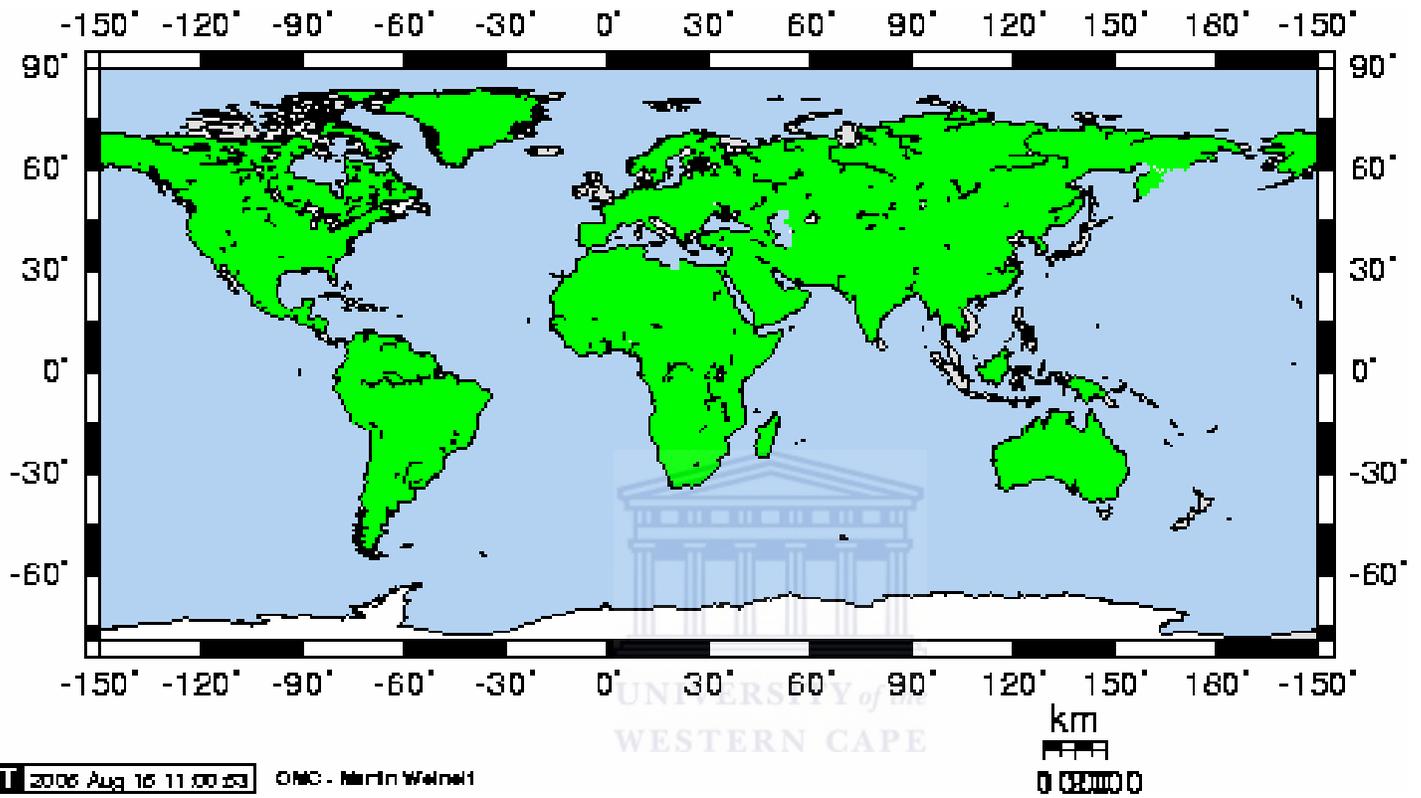
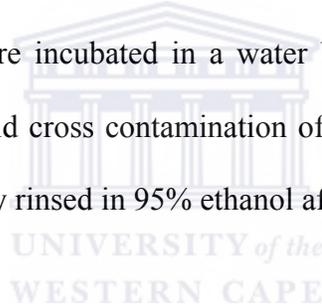


Figure 2.1 Map

2.2. DNA Isolation

DNA extraction was achieved using the CTAB (Hexadecyltrimethylammonium bromide) method (Corach 1991). DNA was purified from a portion of the abdomen and tail of adult individuals. This portion of the animal is rich in muscular tissue and should therefore represent an area of the animal's body where mitochondrial organelles are at their highest concentration. It is important to ensure the cephalothoracic region of the animals, where the stomach is located, is avoided to avoid non-specific DNA isolation involving the prey species of euphausiids (copepods and phytoplankton Gibbons *et al.* 1991, 1992). Additionally, in many of the euphausiids samples collected, a small proportion of individuals displayed the presence of what appeared to be an ellobiopsid macro parasite, thought to be the species *Thalassomyces fagei* (Mauchline 1966b). This parasite was almost always found protruding from the median dorsal surface of the cephalothorax. It has been shown that these parasites in the course of their ontogenetic development are first observed as a small body present in the region of the gonads under the carapace from where it separates into a finger like extension that penetrates the carapace and a 'root' that grows down into the gonads, often resulting in castration (Mauchline 1966b). Again to avoid non-specific DNA isolation, all infected animals were excluded from the samples prior to DNA extraction.

Individual animals were first removed from the sample jars containing the preservation medium and placed on tissue paper laid out on a horizontal surface. The animals were then pressed by hand to remove as much ethanol as possible from the tissues, as this often inhibits the DNA extraction process. Individual abdomens and tails were then removed using dissecting scissors, placed in labeled 2 ml Eppendorf tubes, and homogenized in 500 μ l of CTAB digestion buffer containing 2% CTAB, 1.4M NaCl, 0.2% (v/v) 2-mercaptoethanol, 20mM EDTA and 100mM Tris/HCl, and water. To these tubes 5 μ l of 0.5mg/ml proteinase K enzyme (Promega) was added to denature proteins, and the tubes were incubated in a water bath at 55°C overnight to facilitate digestion. To avoid cross contamination of DNA among individuals, the scissors were thoroughly rinsed in 95% ethanol after each use.



Following digestion the tubes were removed from the water bath and 500 μ l of chloroform: isoamyl alcohol (ratio 24:1) was added, and the tubes were shaken for 5 minutes. The tubes were then centrifuged at high speed (1400 rpm), for 10 minutes to separate the aqueous from the organic phase. The resulting supernatant was removed using a pipette, taking care to avoid taking the interface, which is usually milky-white in colour. This step was repeated until the interface was clear, indicating that most proteins and lipoproteins have been removed from the aqueous phase. Precipitation of the nucleic acids was achieved by adding two volumes of absolute ethanol (100%) to each tube,

which were then placed in the freezer at -20°C overnight, or, until a white precipitate was observed. The tubes were then removed from the freezer and centrifuged again at high speed for 20 minutes, during which time, a precipitated pellet of total genomic DNA was formed attached to the bottom of the tube. The absolute ethanol was then discarded. The precipitated pellet of DNA was then washed in 200 μl of 70% ethanol to remove residual salts for 5 minutes. Finally, the tubes were then centrifuged at high speed for 20 minutes and then oven dried at 55°C to remove excess alcohol and the precipitated pellet of DNA was dissolved in 50 μl of ultra pure sterile distilled water.

2.3. Amplification of Mitochondrial DNA

2.3.1. *NADH 1*

A 156 base pair region of a portion of the gene coding for subunit 1 of the mitochondrial NADH dehydrogenase gene (*ND1*), was amplified with the polymerase chain reaction (PCR) using the oligonucleotide primers *ND1f* (5'-CTTGTGTTGTACTAGTTTAGG-3') (Zane *et al.* 1998) and *ND1r* (5'-ACAATCACGCTGATAAAGAGAAAAT-3') (Zane *et al.* 1998). These primers were specifically designed for the euphausiid *Meganyctyphanes norvegica* and worked well for the three species investigated here. This gene region has been previously shown to be highly polymorphic in two other krill species *E. superba* (Zane *et al.* 1998), and *M. norvegica*, (Zane *et al.* 2000) and therefore, considered an appropriate starting point for population genetic investigation in our species. A map showing the relative position of sites and

the orientation for all of the oligonucleotide primers used in this investigation is shown in Figure 2.2.

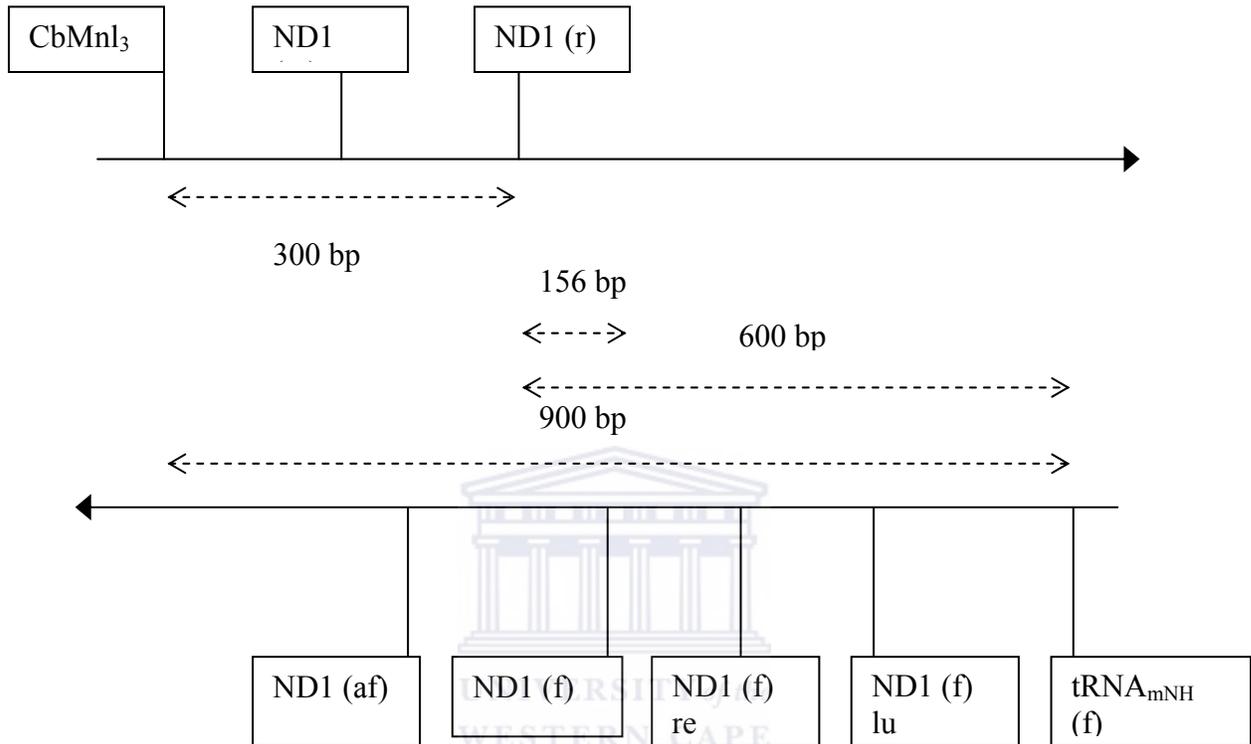


Figure 2.2. The relative position and the orientation all of the mtDNA primers used in this study. Dashed arrows indicate fragment size. bp = base pairs.

Each 15 μ l reaction contained around 20-40 *ng* of purified DNA from one individual krill as template. The reaction buffer contained 7.5 μ l of distilled water, 10x buffer, 2.5 mM MgCl₂, 0.24 mM deoxynucleotides (dNTP's), 0.45 μ l of Taq (5U/ μ l) and 0.3 μ M of each primer. The amplification reaction was run in a Perkin Elmer 9600 Gene Amp model using the following program: 94

°C (2 minutes): 94 °C (30 seconds) 50 °C (30 seconds): 72 °C (1 minute) for 37 cycles with a final extension time of 5 minutes at 72 °C.

The oligonucleotide primer tRNA_{mNH} 5'-CAGAGGGATGCTCAGGATTTA-3' (Zane *et al.* 1998) located upstream of *NDIf*, in combination with *NDIr*, was used to amplify a region of 600 base pairs in length downstream of the target region. This amplification, that encompassed the target sequence, was used to identify suitable sites for the design of species-specific primers just up/downstream of our target region for sequencing purposes. The species-specific primers *NDI-f-lu* (5'-TCCTTATTATTTGTCTCCTG-3'), for *E. lucens*, (also worked for *E. vallentini*) and *NDI (f)-re* 5' – TGTTTACAAAAGAACAGACGTACCC – 3', for *E. recurva*, I designed in my lab using this approach with the program Oligo version 1.4 (Rychlick 1989) and these were used in combination with the primer *CbMnl3* (5'-GGAGCTCGAACCTGTAGAAGA-3') designed by Zane *et al.* (1998) specific for the species *Meganyctyphanes norwegica*, to amplify a DNA segment of 600 base pairs in length that encompassed our target region.

If this reaction failed to amplify, the species-specific primers were used in combination with the primer *NDIr*, to produce a segment 300 base pairs in length. If PCR reactions with the primer combination *NDIr* and *NDIf* failed with any particular regional population of animals, I investigated whether the presence of polymorphisms in the primer sites was causing this phenomenon by

sequencing several individual using the primer combination *Cb_{Mnl3}* (5'-GAGCTCGAACCTGTAGAAGA-3'), and *ND1-f-lu* (5'-TCCTTATTATTGTCTCCTG-3'). When these sequences revealed the presence of polymorphisms near or at the 3' prime end in the primer sites, new primers were designed specific to these populations. This was successfully done and produced the primers *GH1f* 5'-TTTTTCTATGTTGTACAAGATT-3' and *GH2r* 5' ACAATCTCGCTGATATAATGA -3' and these amplified well for specimens of *E. lucens* collected from Argentina using the same reaction conditions described above.

2.3.2. *Cytochrome Oxidase I*

For a subset of individuals from each of the three species a region of subunit 1 of the cytochrome oxidase 1 (*COI*) gene 1300 base pairs long was amplified using the primers LCO (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO (5'- TAAACTTCAGGGTGACCAAAAATCA-3') of Jarman *et al.* (2000) and *COI-a* (5' – AGTATAAGCGTCTGGGTAGTC-3') and *COI-f* (5'-CCTGCAGGAGGAGGAGATCC-3') of Bucklin *et al.* (1997). The amplification reactions were run in a Perkin Elmer 9600 Gene Amp model using the following program: 94 °C (2 minutes): 94 °C (30 seconds) 50 °C (30 seconds): 72 °C (1 minute) for 37 cycles with a final extension time of 5 minutes at 72 °C. Both sets of *COI* primer pairs have previously been used for phylogeny studies in Crustacea (Bucklin *et al.* 1997; Jarman *et al.* 2000,

Ovenden *et al.* 1997; Shubert *et al.* 1998) but have never been utilized together in any published study.

2.3.3. *16S rRNA*

Finally, an approximately 570 base pair region of the *16S* rRNA mitochondrial gene was amplified in several euphausiid species using the primers *16Sa* and *16Sb* (Palumbi *et al.* 1991) for the purpose of estimation of evolutionary rates of sequence substitution (see below). The amplification reaction was run in a Perkin Elmer 9600 Gene Amp model using the following program: 94 °C (2 minutes): 94 °C (30 seconds) 50 °C (30 seconds): 72 °C (1 minute) for 37 cycles with a final extension time of 5 minutes at 72 °C. This gene fragment has also previously been used to investigate the molecular phylogenetic relationships among euphausiid species (Patarnello *et al.* 1996; Jarman *et al.* 2000). Identification and quantification of the amplification products was achieved by running 2.5 µl of PCR product on a 1% agarose gel (reference) at 85 volts for 15 minutes along with the size marker λ /Hind III (Promega) Staining of the DNA bands in the gel was achieved with the use of ethidium bromide solution (1%) for 15 minutes, which was visualized by illumination with ultraviolet light.

2.3.4. *Sequencing Reactions*

An ABI reaction kit (ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit v 3.1) was used to cycle sequence unpurified PCR products at the

reaction conditions recommended by the manufacturer. The following modified protocol was used: 1.6 pmol primer, 2 μ l ABI reaction kit; 10-15 ng PCR product; deionized water to a final volume of 10 μ l. The cycle sequence reaction was run in a Perkin Elmer 9600 Gene Amp model using the following program: 94 °C (5 minutes): 94 °C (10 seconds) 55 °C (10 seconds): 60°C (4 minutes) for 25 cycles. DNA sequencing was carried out in an Applied Biosystems. Inc; Automated DNA sequencer, Model AB3100. All amplicons were sequenced in both directions.

2.4. Amplification of Nuclear DNA

For a subset of individuals from each of the three species investigated here a ~500bp base pair region of the nuclear *ITS-1* region was amplified with the primers SP-1-5' 138 (5'-CACACCGCCCGTCGCTACTA-3) located in the 18S rDNA (sites 1737 to 1756 in the reference sequence U4837) and Sp-1-3 (5'-ATTTAGCTGCGGTCTTCATC -3') located in the 5.8S rDNA (Chu *et al.* 2001). This region is highly divergent among different crustaceans and has been shown to be an appropriate marker for molecular systematic studies at both the species and the population levels (Chu *et al.* 2001).

Each 15 μ l reaction contained around 20-40 ng of purified DNA from one individual krill as template. The reaction buffer contained 7.5 μ l of distilled water, 1.5 μ l of 10x Mg²⁺ free buffer, 2.5 mM MgCl₂, 0.24 mM deoxynucleotides (dNTP's), 0.45 μ l of Taq (*Thermus aquaticus*) and 0.3 μ M of

each primer. The amplification reaction was run in a Perkin Elmer 9600 Gene Amp model using the following program: 94 °C (3 minutes): 94 °C (30 seconds) 55 °C (30 seconds): 72 °C (1 minute) for 37 cycles and finally 5 minutes at 72 °C. The size and quality of PCR products were visualized on a 1.5% agarose gel. Prior to sequencing, PCR products were purified using QIAquick PCR purification kit, or if necessary the PCR products were excised from the gel and purified with a QIAquick gel extraction kit (Qiagen). Purified double-stranded PCR products were then used in cycle sequencing reactions to generate templates for automated nucleotide sequencing. Cycle sequencing reaction mix contained 8.0 µl of ABI Prism dRhodamine terminator (ABI, Prism Perkin-Elmer, Foster City, California), 3-5 µl of PCR products, 1 µl of primer, and ddH₂O to make up to 20 µl. The unincorporated primers were removed using an ethanol sodium acetate precipitation procedure. The purified samples were then taken and DNA sequencing carried out in an Applied Biosystems. Inc; Automated DNA sequencer, Model AB3100. ABI

2.5. Assessing Within-Individual Variation

Given the repetitive nature of rDNA genes, variation among the tandemly repeated units might be expected within individuals, and this was assessed by sequencing five clones from two individuals from each of the *E. recurva* regional populations sampled. Fresh PCR products were obtained and purified from either solution or agarose gels using the GFX™ DNA and gel band

purification kit (Amersham Biosciences) according to manufacturer's specifications.

Ligation of the purified amplicons and the linear pTZ57R vector was performed overnight (16 h) at 16°C in a total volume of 10 µl. The reaction mixture contained 1 µl of a 10 × DNA ligase buffer (60 mM Tris·HCl, pH 7.8, 20 mM MgCl₂, 20 mM DTT, 1 mM ATP, 10% PEG 6000), 50 ng of pTZ57R vector, between 4 and 7 µl insert DNA, 5 U of T4 DNA ligase (NEB 5 U·µl⁻¹) and ddH₂O. The vector to insert ratio was typically 1:3.

Chemically competent cells were prepared according to the method of Hanahan (1983) with slight modification. All glassware was thoroughly acid-washed with 30% H₂SO₄, rinsed, and autoclaved prior to use. A single colony of the *E. coli* strain was inoculated into 30 ml of LB-broth and incubated at 37°C, with shaking, until stationary phase. Then 1 ml of the culture was transferred to 100 ml of LB-broth and incubated at 30°C until mid-logarithmic phase (OD₆₀₀ of 0.5). The flasks were rapidly cooled in ice-water for 20 minutes and 60 ml of the cells were collected in polypropylene tubes by centrifugation at 1000 × g for 10 minutes in an Eppendorf 5810 R swing bucket centrifuge. After discarding the supernatant, the cells were resuspended in 0.5 × volume filter sterilized competency buffer (0.1 M CaCl₂ [w/v], 0.07 M MnCl₂ [w/v] and 0.04 M NaOAc [w/v], pH 5.5) and incubated at 4 °C for 30 minutes. Following incubation the cells were harvested by centrifugation at 1000 × g for 5 minutes

and resuspended in 7.5 ml competency buffer. 575 μ l 80% glycerol was added thoroughly mixed and the competent cells dispensed into 100 μ l aliquots and stored at -80 $^{\circ}$ C until required.

An Eppendorf tube containing 100 μ l of chemically competent cells was removed from -80 $^{\circ}$ C and allowed to thaw on ice. Then 2 μ l of ligation mix was added to the thawed cells and gently mixed. The mixture was incubated on ice for 30 minutes then heat-shocked at 42 $^{\circ}$ C for 90 seconds in a water bath. The Eppendorf tube was then returned to the ice water for 2 minutes whereafter 900 μ l of sterile LB-broth was added and the Eppendorf tube incubated at 37° C for 1 h with agitation. The cells were plated in aliquots of 100 to 200 μ l onto LB-agar plates supplemented with the appropriate antibiotic. Where applicable, recombinant transformants were selected by blue/white colour selection based on insertional inactivation of the *lacZ* gene. For this purpose, the cells were spread together with 40 μ l of X-gal (2% [v/v] stock solution) and 10 μ l IPTG (100 mM stock solution) over the surface of LB-agar plates, supplemented with the appropriate antibiotic and incubated overnight at 37 $^{\circ}$ C.

Following transformation of *E. coli* cells, colony PCR was used to screen all putative recombinant clones. The putative recombinants were aseptically inoculated into LB-broth containing the appropriate antibiotic and incubated overnight at 37 $^{\circ}$ C with agitation. Cultures were further analyzed by pipetting 200 μ l of each into 0.6 ml PCR tubes. The tubes were centrifuged at $13\ 000 \times g$

to pellet the cells and the supernatant discarded. The cells were resuspended in 200 μ l UHQ Millipore water and lysed by incubation at 98°C for 5 minutes. Tubes were then centrifuged at 13 000 \times g for 5 minutes to pellet cell debris. The DNA-containing supernatant served as template in 30 μ l PCR reactions performed essentially as described previously (the annealing temperature was lowered to 49 °C). An aliquot of each PCR reaction was analysed by gel electrophoresis as described previously. Plasmid extractions performed for subsequent nucleotide sequence analysis was performed using the Talent plasmid purification kit.

2.6. SSCP Screening of mtDNA Variation

Screening of mtDNA haplotype variants was achieved using single strand conformational polymorphism analysis (Orita *et al.* 1989; Hayashi 1991a,b). This method entails electrophoresis of single-stranded DNA fragments of suitable size through a nondenaturing polyacrylamide gel, followed by visualization. Under appropriate conditions, (low temperature and non-denaturing conditions), DNA strands fold into structures that migrate according to their shape, and so have distinct gel mobilities. Recent evidence suggests that these mobility differences are based primarily on tertiary, rather than secondary structure of the DNA molecules. The sensitivity is generally inversely proportional to the size of the fragment with single nucleotide polymorphisms resolved 99% of the time for 100-300 base pairs fragments

(Sunnucks *et al.* 2000; Girman 1996) making it particularly suitable for our fragment of 156 base pairs in length (Orita *et al.* 1989).

For each individual, 2 μ l of the PCR product amplified using the primers *NDI_f* and *NDI_r*, was added to an equal volume of formamide loading gel (REF), and heat denatured at 95 °C for 3 minutes in a Perkin Elmer 9600 Gene-Amp system. Heat and chemical denaturation causes the double stranded DNA double helix structure to disassociate into two complementary, single stranded molecules. The PCR tubes containing DNA from individual animals, immediately on ending denaturation, were then plunged into an ice bucket to avoid renaturation of the single-stranded DNA molecules. The 4 μ l of DNA and formamide loading dye mixture was then loaded on a 10% acryl amide gel (37.5:1 acrylamide / bisacrylamide) with 5% glycerol. Runs were performed at 4 °C and at 135V for 16 hours in a 20 cm high vertical apparatus.

2.6.1. Staining the SSCP Polyacrylamide Gels

The polyacrylamide SSCP gels were stained with silver salts after separation by gel electrophoresis using a modified version of the Sambrook *et al.* (1989) procedure. The polyacrylamide gels were carefully removed from the between the glass plates and fixed by incubating at room temperature in 100 ml of a buffer solution A (containing 90 ml ddH₂O, 10 ml 99% ethanol and 0.5 ml of acetic acid) and left for 4 minutes with gentle shaking. This step was then repeated. The fixing solution was then discarded and 500 ml of ddH₂O was

added and then incubated for 10 minutes at room temperature. The ddH₂O solution was then discarded again and staining solution B (containing 100 ml of ddH₂O, 100 mg of AgNO₃) was added to the container and then incubated for 10 minutes with gentle shaking. This solution was then discarded and the gel given a thorough rinse in ddH₂O for 5 minutes. Then 50 ml of a solution C (containing 150 ml ddH₂O 2.25 g NaOH, 15 mg of NaBH₄ and 0.6 ml of formaldehyde) was added and then incubated for 1 minute. This step was then repeated with 150 ml of solution C. The gel was then incubated at room temperature until the desired contrast was obtained. The reaction was then quenched by washing the gel in solution A for a few minutes and then given a final thorough rinse with ddH₂O.

2.6.2. Scoring of mtDNA Haplotypes

Samples were assigned different categories according to their electrophoresis mobility, assuming that same conformation or mobility indicates the same DNA sequence information. It was, however, necessary to rerun samples from the same gel in adjacent lanes because sometimes, similar ones were not run sufficiently close to allow easy comparison, particularly when occasional irregularities disrupted certain comparisons. When this was successfully accomplished among-gel comparisons could be made.

An initial gel was run, and two individuals of each SSCP pattern observed were sequenced to assess the accuracy of the method. When the accuracy was shown

to be high (determined by the proportion of times the SSCP gel mobility patterns discriminated the different *ND1* haplotypes) these individuals were then used as standards for subsequent comparisons. Further confirmation of the accuracy of scoring was achieved by sequencing randomly chosen individuals and comparing the observed haplotype designation based on the sequence information, with that predicted from the SSCP pattern.

2.7. Population Descriptive Summary Statistics

Mitochondrial sequences were aligned using Bioedit (Hall 1999). DNAsp v 4.51 (Rozas and Rozas 2003) was used to obtain variable sites, parsimony informative sites, population nucleotide diversity (π of Nei 1987), and haplotype diversity (h , gene diversity). The latter was calculated using the equation of Nei (1987, eq. 8.4 replacing $2n$ by n) giving:

$$H = n \sum (1 - \sum p_i^2) / (n-1)$$

where p_i = allele frequency and n = sampled individuals.

Nucleotide diversity (Nei, 1987), the average number of nucleotide differences per site between two sequences was calculated using the method of Nei (1987,eq 10.5). This method takes into account both haplotype frequencies and sequence divergence between haplotypes and is calculated using the equation:

$$\pi = \frac{n}{n-1} \sum_{ij} x_i x_j \pi_{ij}$$

Where x_i and x_j are the frequencies of the two haplotypes in a population and π_{ij} is the proportion of nucleotide differences between them (Nei 1987).

2.8. Substitution Models

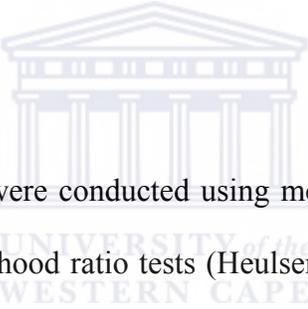
The most appropriate nucleotide substitution model was calculated for each individual population and for the combined data set for each euphausiid species using the program MODELTEST, version 3.06 (Posada and Crandall 1998). This program uses two statistical approaches to model fitting, hierarchical likelihood ratio tests, and the Akaike information criterion (Akaike 1974).

2.9. Interspecific and Intraspecific Phylogenetics

The evolutionary relationship of *E. lucens*, *E. recurva*, and *E. vallentini* were investigated using phylogenetic reconstructions. Maximum parsimony (MP), maximum likelihood (ML) and searches were conducted with the heuristic search approach of PAUP* 4.0 (Swofford 1998) while the minimum evolution (ME) trees were reconstructed using the software program Mega 2 (Kumar 2001). For maximum parsimony, the default settings were applied where all nucleotide substitutions were weighed equally regardless of type or codon position. Each base was treated as an unordered character with four alternative states. Ancestral states were determined via outgroup comparison. Only the unique mtDNA haplotypes were used for our analysis. In the interspecific data set where the number of terminal taxa was too large to permit evaluating all

trees or employing the branch-and-bound algorithm (Hendy and Penny 1982), heuristic searches were used for each tree building topology.

One hundred repeated randomized input orders of taxa were used for all MP analysis to minimize the effects of entry sequence on the topology of the resulting cladograms. MP analysis were conducted with the steepest descent option, and with accelerated character transformation (ACCTRAN) optimization, tree bisection-reconnection (TBR) branch swapping with a starting tree obtained by step-wise addition, save all multiple trees (MULPARS), and zero length branches collapsed to yield polytomies settings in place.

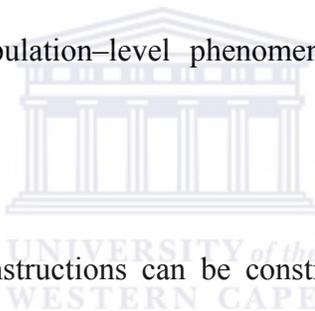


The ML and ME searches were conducted using models of DNA substitution selected by a series of likelihood ratio tests (Heulsenbeck and Crandall 1997). Statistical confidence in the stability of tree nodes was calculated by nonparametric bootstrap (Felsenstein, 1985) with 1000 replications (Hedges 1992). Nonparametric bootstrapping is considered a conservative measure of the probability that a recovered group represents a true clade (Zharkikh and Li 1992; Li 1997).

2.9.1. Genealogical Network Construction

There are several reasons why traditional methods of phylogenetic analyses are not the most appropriate representation of intraspecific polymorphic data

(Zhang and Hewitt 2003). In particular, the low levels of polymorphism typically observed within such data sets make resolving unambiguous evolutionary relationships in the data difficult, and implicitly assuming a strictly bifurcating process ignores the complications associated with reticulate relationships produced by recombination and parallel and recurrent mutations (Zhang and Hewitt, 2003; Bandelt *et al.* 1995; Smouse 1998, 2000). Also the treatment of rare polymorphism by methods such as parsimony, ignore the information contained in such occurrences for inferring population demographic processes (Zhang and Hewitt, 2003). Networks provide an alternative approach to estimating intraspecific genealogical relationships that take into account the population-level phenomena (Posada and Crandall, 2001).



Intraspecific network reconstructions can be constructed using a number of different approaches (Posada and Crandall, 2001; Cassens *et al.* 2003). I used Network version 2.0 (Bandelt *et al.* 1999), to reconstruct genealogical relationships among haplotypes as a median-joining (M-J) network, assigning equal weights to all variable sites and with default values for the epsilon parameter ($\epsilon = 0$). This method begins by combining the minimum-spanning trees (MST's) within a single network, then, with a parsimony criterion, median vectors are added to the network. A recent critical examination of several widely-used network estimation and rooting methods (Cassens *et al.* 2003) found substantial differences in the resulting intraspecific

networks, underlying the limitations of some of the algorithms. The authors concluded that the performance of the M-J network was found to be one of the most reliable of the methods investigated (Cassens *et al.* 2003) and is thus, an appropriate choice for the analysis of my euphausiid intraspecific data sets.

Rooting of the phylogenetic trees and networks was done using appropriate outgroups. As the putative sister-species to *E. lucens* (Jarman *et al.* 2000) *E. vallentini* represents an appropriate outgroup, and vice versa. Depending on the gene fragment involved, the *E. recurva* phylogenies were rooted using a range of euphausiid species. The identity of the most recent common ancestor (most basal lineage) in each of the species and regional populations was inferred from the location of the branch joining each species or population to its sister group in the majority consensus tree. Some predictions from coalescent theory were then applied to obtain the most likely connection between two haplotypes (Crandall and Templeton 1993; Crandall *et al.* 1994; Posada and Crandall 2001). The two most basic: (i) haplotypes are more likely to be connected to common than to rare haplotypes (the frequency criterion) and (ii) haplotypes are more likely to be connected to interior than to exterior haplotypes (the topology criterion). Additionally, coalescence theory predicts that the most common ancestor will be the one found in the greatest number of different populations (Crandall and Templeton 1993).

2.10. Estimation of Evolutionary Rates between Species

My rationale to estimate the coalescence time between species is as follows: first I aligned my krill *16S* rRNA sequences (one randomly selected individual per species) with sequences of other crustacean species previously used to calibrate molecular clocks based on well-dated vicariant events (Schubart *et al.* 1998; Knowlton and Weigt 1998), such as the closure of the Panama Isthmus ~3.1 million years ago.

The NJ tree was constructed with LinTree with 1000 bootstraps to obtain a balanced tree with 2 clusters rooted with the outgroup sequence of *Squilla empusa* (AF107617) (*Holocarida*, *Sromatopoda*). In order to verify whether the *16S* rRNA molecular clock calibrated for crabs (Cunningham *et al.* 1992) can be applied to krill, I compared the substitution rate between krill and six non-Jamaican grapsid crab taxa from Schubart *et al.* (1998), including congeners of *Sesarma* from the Atlantic and Pacific Oceans. The calibration point used on the krill and *Grapsidae* trees was the transisthmian vicariance set between 2.5-3.1 million years ago. Models of nucleotide substitution for each of the datasets were then obtained using a hierarchical likelihood ratio test as implemented in Modeltest v.3.06 (Posada and Crandall, 1998). Constancy of evolutionary rate between lineages (euphausiids vs other crustaceans) was then tested by the Two Cluster Test (TCT) and Branch Length Tests (BLT) with the program LinTree of Takezaki *et al.* (1995). The null hypothesis in these test is constancy of nucleotide substitution rate between each of the two species clusters in a tree

and equal branch lengths between each branch and the average of all other branches but the outgroup. Neighbour-joining trees were constructed using PAUP 4.1 with 500 bootstraps. I chose outgroups as equidistant as possible from both (Euphausiacea and Decapoda groups (Eucarida Malacostra)).

2.11. Neutrality Tests

The neutral theory of molecular evolution first proposed by Kimura (1968) maintains that most evolutionary change at the molecular level in natural populations is due to the random genetic drift of neutral mutations. Under this theory natural selection does not play an important role in the fixation of mutant alleles (Kimura 1968). Although controversial, and the subject of intense debate in the past (Hey 1999) the current utility of the strictly neutral model of evolution for evolutionary biologists is as the primary null hypothesis used to test for the effects of natural selection (Hey 1999; Ford 2002). The rationale behind such an approach to testing neutrality of molecular variants in populations is that if genes have more or less alleles than would be expected from mutation pressure alone, then other forces that operate in nature must tend to favour, or conversely, eliminate alleles (Ford 2002).

The adoption of the strictly neutral Wright-Fisher model as a null hypothesis however, involves making some explicit assumptions. These include random mating among many populations, each sex with equal size, experiencing no selection, recombination, overlapping generations, population subdivision, or

demographic change and each at mutation-drift equilibrium (when the introduction of new alleles by mutation is balanced by the loss of new alleles by random genetic drift), and migration–drift equilibrium (when the introduction of new alleles to the population by migration is balanced by the loss of new alleles random by genetic drift). It is unlikely that any natural population will meet all these assumptions, although a violation of any of these can lead to significant deviation from theoretical expectations. Theoretically, mutation-drift equilibrium should be reached if the effective population size (N_e), has remained stable/stationary for $2N_e-4N_e$ generations (Nei and Li 1976), while migration drift equilibrium should be reached if population size and migration rate are stable for $1/m$ or $2N_e$ generations, whichever is the larger (Takahata 1983). Neutrality and equilibrium-drift mutation is in fact the condition that permits an estimate of demographic parameters such as the effective population size (N_e).

Both background selection (Charlesworth *et al.* 1993, 1995) and population size changes may affect estimates of θ , and selective sweeps, perhaps related to variation in reproductive success, large changes in population size and bottlenecks have all been proposed as reasons for the previous general finding for euphausiids (and for many marine organisms), that estimates of effective population size are often several orders of magnitude smaller than corresponding estimates of census population size (Bucklin and Weibe 1998; Grant and Bowen 1998; Zane *et al.* 1998, 2000; Peijnenburg *et al.* 2005).

Many of these forces have been thought to have been particularly affected by the glacial cycles. Rejection of the null hypothesis probably means that selection and/or population level processes (expansion, contraction, subdivision etc) are responsible for the pattern of polymorphisms observed.

As the pattern of polymorphism in a sample is affected by all kinds of evolutionary forces it follows that no single statistical test may have the power required to detect all of them (Fu 1997). For this reason several so-called pure significance tests of neutrality have been designed, each one being the most powerful for detecting a particular class of departures from the neutral model (Tajima 1989; Fu & Li 1993; Fu 1996, 1997). Of particular interest is being able to distinguish between different factors that produce similar patterns in the polymorphisms. For instance, background selection, logistic population growth or genetic hitchhiking all tend to result in an excess of rare alleles or young mutations (Fu 1997). Consistency in the results of the different tests is often the only criterion for acceptance or rejection of the null hypothesis (Ford 2002). However, a fuller understanding of the reason for inconsistency in the results of the various tests may be achieved with knowledge of the underlying assumptions of the respective methods and of how various types of mutations contribute to the values of the test statistics (Rand 1996; Fu 1994a; Yu *et al.* 2002). Mutations resulting in segregating sites of a sample of DNA sequences can be classified by size and type and the frequencies of mutations of different sizes and types can be inferred from the sample (Fu 1994a). For a population

that evolves under the Wright-Fisher model, then, in each sample of n sequences has a genealogy of $2(n-1)$ branches that connect the sequences to their most common ancestor. From Fu (1995): “The size of branch is defined as the number of sequences in the tree that were the descendants of that branch Fu (1995) which, will be a positive integer less than the number of sequences in the genealogy. A mutation of size i is said to be of size i if it falls on a branch of size i . Although mutations of size and type take the sample genealogy into consideration they can be inferred without reference to the sample genealogy. In the example given below without knowledge from the outgroup sequence which allows correct polarization of character state, the first segregating site must result from a mutation of either size 1 or 4. With the outgroup information the four segregating sites correspond to mutations of size 1, 2, 4 and 3 respectively.

Consider the following example:

	Sequence								
1	---	A	---	G	---	C	---	T	--
2	---	G	---	G	---	C	---	T	--
3	---	G	---	C	---	C	---	A	--
4	---	G	---	C	---	C	---	A	--
5	---	G	---	C	---	T	---	A	--
Segregating Site	1	2	3	4					
Outgroup	G	C	T	T					

Fu (1994a,b) showed that the expected number of mutations of size i in a sample is given by:

$$E(\xi_i) = 1/i \theta$$

The probability density function for mutations of size i contains important information regarding the evolution of a sample of sequences (Fu 1994a,b). When an appropriate outgroup sequence is not available, then the size of a branch cannot be uniquely defined and can be either i or $n-i$ and cannot be distinguished. In this case mutations can be classified as belonging to $n/2$ groups so that group i consists of all mutations that are either size i or size $n-i$ (Fu 1995).”

The real challenge however, is to dissect the causes for the deviation from the null model taking into consideration various lines of evidence (Yu *et al.* 2002). I investigated departures from neutral expectations using four different tests and these can be divided into two main types and all of which were calculated using DNAsp v 4.51 (Rozas and Rozas 2003). The first including Tajima’s D (1989), and Fu and Li’s (1993) D , and F tests, generate summary statistics using information of the mutation (segregating site) frequency and are based on the difference between two alternative estimates of the population mutation parameter $\theta = 2 N_e\mu$ where N_e is the effective population size and μ is the mutation rate per generation. Fu and Li’s (1993) tests use information from the number of recent mutations, and therefore, are more accurate when the correct outgroup is available to accurately polarize the cladistic characters on the tree.

These statistics are used to distinguish population growth from constant sized populations as population growth generates an excess of mutations in external branches of the genealogy (i.e. recent), and therefore, an excess of singleton substitutions (Tajima 1989; Slatkin and Hudson 1991). Because the purpose of these tests is to detect departures characterized by an excess of the number rare alleles and a reduction in the number of common alleles, which tends to give negative values for these tests, large negative values are taken then as evidence against the neutral model.

The second type, Fu's F_s test (Fu 1997), uses information from the haplotype distribution. This summary statistic is based on Ewen's (Ewens 1972) sampling distribution and is sensitive to demographic expansion. The significance of F_s was evaluated using 1000 random permutations dependent on theta with the population genetics software program DNAsp (Rozas and Rozas 2003).

All of the tests described above use only intra-specific polymorphisms and therefore test if an observed polymorphism is consistent with the strict neutral model within a short period of $4N$ generations on average (Fu and Li 1999). To test the neutral hypothesis over a longer period I used the McDonald–Kreitman (1991) test that compares non-synonymous and synonymous substitutions both within, and between species. In a population under mutation-drift equilibrium, the numbers of substitutions in these two classes of polymorphism are expected to be equal (Kimura 1983).

2.11.1. *How the Tests are calculated*

The test of Tajima (1989) compares two estimates of the population parameter θ , based on the number of segregating sites (S), and the average number of nucleotide differences in pairwise comparisons (π); reasoning that Watterson θ_w (1975) and Tajima (1989) θ_π estimates are affected by natural selection to different extents (Tajima 1989).

Neutral theory (Kimura 1983) predicts that for a sample of sequences collected from a population that is at mutation-drift equilibrium, then $\pi = S$. After a change in population size, the estimate of θ , using the latter measure is more slowly affected by the (new) present population size, while the estimate from the former would reflect the more recent population sizes (Tajima 1989). Because π is sensitive to the frequency of a given type of sequence in the sample, and S is not, the difference between these two measures is the basis of Tajima's (1989) test of neutrality. A positive D value in this test will indicate an excess of common alleles and a negative D an excess of rare alleles.

Fu and Li's (1993) D test, like Tajima's also compares two estimates of θ but, unlike Tajima's test, incorporates information from the genealogical tree structure to compare the relative frequencies of internal and external mutations. At a given particular polymorphic site, the number of mutations in external branches is counted as the number of distinct singleton nucleotide variants (in the intraspecific data file), that are not shared with the outgroup (a singleton mutation is a nucleotide variant that appears only once among the sequences).

The total number of mutations in external branches of the genealogy is then computed as the sum of the number of mutations in external branches of every polymorphic site. This test assumes that since selection will remove deleterious mutations, those mutations present are likely to have arisen recently and will be found close to the tips of the genealogy (Fu and Li 1993). Older mutations are expected to be found internally and to be selectively neutral. However, selective sweeps would make more recent mutations occur on internal branches. This test has also been shown to be very powerful, particularly for its sensitivity in detecting the presence of background selection (Fu 1997).

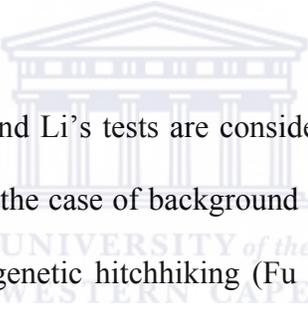
Fu and Li (1993), proposed a similar test where the test statistic F is based on the differences between the total number of mutations in external branches of the genealogy and k , the average number of nucleotide differences between pairs of sequences. Negative D and F values indicate either that purifying selection is acting on the locus, or that the population has recently undergone an expansion or selective sweep (Fu 1997).

Fu's F_s test (Fu 1997), which detects an excess of rare alleles given θ , utilizing expectations based on Ewens' sampling theory (Ewens 1972) was calculated. This test has been shown to be very powerful in determining the most likely of the three alternative scenarios of background selection, logistic population growth, or genetic hitchhiking, when recombination is absent (Fu 1997). Unlike the other tests mentioned this test is based on an analytical form for the patterns

of polymorphism found. The critical point for the F_s test is the value corresponding to the lower second percentile of its distribution. The F_s test statistic is calculated using the following equation:

$$F_s = \ln (S'/1-S')$$

Where S' is defined as the probability of having no fewer than k_0 alleles in a random sample provided that $\theta = \pi$. In a sample with an excess of recent mutations θ (estimated by θ_π) is likely to be smaller than that based on the number of alleles. F_s will be negative when there is an excess of recent mutations and a large negative value is taken as evidence against the neutrality of mutations (Fu 1997).



It has been shown that Fu and Li's tests are considerably more powerful than Tajima's test and Fu's F_s in the case of background selection, and the opposite for population growth and genetic hitchhiking (Fu 1997). So, if only Fu and Li's test is significant, then background selection has likely played an important role in shaping the genetic diversity in the sample, and, if only F_s is significant, then population growth or hitchhiking are more likely (Fu, 1997).

Because models of a very different nature can give rise to very similar patterns of polymorphism, caution must be exercised to avoid over-confidence on the specific alternative model when the strict neutral model is rejected (Fu 1997). In addition, as all the previous tests using only intraspecific polymorphism they test if an observed polymorphism is consistent with the strict neutral model

within a short period of $4N$ generations on average. To test the neutral hypothesis over a longer period, an approach considering a combination of inter and intraspecific polymorphism was employed (McDonald and Kreitman 1991).

The M-K test (McDonald and Kreitman 1991) considers the relationship between polymorphism and divergence, focusing on synonymous and nonsynonymous variation within a single gene region in two related species. Unlike Tajima's D (Tajima 1989), this test is not affected by violations of equilibrium conditions due to changes in population size. A further advantage of this method is that drift variance, which cannot be minimized by taking larger samples, should not be a major contributor to differences in diversity between these kinds of sites (Skibinski 2000).

The test is applied to a single phylogenetic tree with two major clades separated by a single internal branch linking through a common ancestor. The number of synonymous and nonsynonymous changes is counted both within the clades (polymorphisms) and on the internal branch (fixations). This test is particularly powerful when the comparisons are among closely related species, where multiple substitutions at a single site have not occurred (McDonald & Kreitman 1991). Neutral theory predicts that evolutionary forces should influence synonymous and nonsynonymous substitutions similarly and the ratio of synonymous and nonsynonymous changes should be the same for

polymorphisms and fixations (Kimura 1983). The M-K test then compares the ratios using a 2 X 2 contingency table test (McDonald and Kreitman 1991).

This method was previously used to compare the mitochondrial DNA of ten species pairs from diverse animal groups, and it revealed an excess of non-synonymous polymorphism in all tests (Rand and Kann 1998). An excess of nonsynonymous polymorphisms can be explained due to the accumulation of mutations that are slightly deleterious within populations, but because they are deleterious, they are unlikely ever to become fixed between populations. An excess of nonsynonymous fixations can be explained by positive divergent selection (McDonald and Kreitman 1991). It has been noted that evidence of purifying selection is not necessarily inconsistent with non-significant results for neutrality tests (Skibinski 2000). Strong purifying selection can remove deleterious mutations rapidly, the remaining variation being neutral. One disadvantage this method has in detecting deviation from neutrality, is that artefactual evidence of positive amino acid substitutions can be generated with this test if some amino acid mutations are slightly deleterious and there has been an increase in effective population size (Smith-Eyre *et al.* 2003).

2.12. Estimates of Effective Population Mutation Rate (θ)

The essential population parameter for mitochondrial DNA $\theta = 2N_e\mu$ (where N_e is the effective population size of females, and μ is the mutation rate), was estimated using several different methods, of varying degrees of sophistication

that utilize different properties of the data. Because all of the assumptions of these respective estimators are likely to be violated to some extent, the approach adopted here involves choosing methods that will allow adequate comparisons to be made (Yu *et al.* 2002).

The different methods of inference can be classified on whether they utilize information from the structure of genealogical trees in their estimator, and are therefore based on coalescent theory, (Kingman 1982), or not. A further subdivision of the genealogical estimators can be made between those methods that utilize least-squares (LS) and maximum likelihood (ML) approaches. Each estimator has its' own set of associated assumptions.

The non-genealogical summary statistic estimators of θ used herein were Watterson's (1975), and Tajima's (1983). The two genealogical estimators based on generalized linear model were Fu's UPBLUE (1994a), and EVE (Vasco *et al.* 2001) estimators. For comparison, an alternative genealogical estimator that applies a maximum likelihood approach using a Monte-Carlo evaluation method was employed (Kuhner *et al.* 1998). This method has been shown to produce almost identical results to that obtained by UPBLUE when applied to real data (Kuhner *et al.* 1995; Vasco *et al.* 2001), and may be the more efficient when uncertainty in the tree reconstruction exists and as such, it avoids the bias of using a single genealogical reconstruction (Kuhner *et al.* 1995).

Of the many estimators of θ available, for simplicity, θ was estimated using two of the oldest and widely used non-genealogical methods Watterson's (1975) estimator θ_w and Tajima's (1983) estimator Π . For speed of computation, I chose the genealogical estimator UPBLUE (Fu 1994a). The latter method is based on generalized linear models and obtains its estimate from a sample genealogy estimated by the unweighted pair-group method with arithmetic mean (UPGMA) method. When necessary, a long-term history of effective population size was obtained using Watterson's estimate θ_1 (Fu and Li 1993), which is estimated with mutations of size 1 excluded, resulting in:

$$\theta_1 = (K - \xi_1) / (a_n - 1)$$

ξ_1 = the number of external mutations, K = segregating sites and $a_n = 1 + \frac{1}{2} + \frac{1}{3} \dots + \frac{1}{n-1}$.

Because θ_1 can be applied to a subset of mutation classes, it is particularly useful when one suspects that certain classes of mutations are likely to be strongly affected by an evolutionary force (Fu 1994b). For example, various scenarios of euphausiid population history have a large effect on mutations of size 1 (Fu 1994a,b). If an excess of mutations of this size is apparent it is therefore sensible to obtain a long-term euphausiid effective population size from θ_1 excluding mutations of size 1 (Yu *et al.* 2002).

Watterson (1975) derived an estimate of θ using the number of polymorphic sites (K) which was, $\theta_{\omega} = K / a_n$, where a_n is a constant coefficient with a variance (assuming no recombination) given by $\text{Var}(\theta_{\omega}) = \text{Var} K / a_n$. The estimate of Tajima (1983) on the other hand uses the mean number of pair wise differences between sequences (Π) in the sample to get θ based on the finding of Watterson (1975) that $E(\Pi) = \theta_r = 2N\mu$. Although both of these moments-based summary statistics estimators have had important applications in population genetics, it has been shown that they convey only a fraction of the information available in sample (Felsenstein 1992; Fu 1994a).

An alternative approach that makes statistical inferences based on a more detailed pattern of polymorphisms, is Fu's (1994a) UPBLUE estimate, which uses recursive least squares methods incorporating the genealogical information of the sequences. This estimator is more efficient than both the Watterson (1975) and Tajima (1983) methods as its coefficients are functions of more detailed information available in the coalescent process (Fu 1994a). In fact it has been shown that BLUE estimates often have the theoretically minimum variance possible. That is the smallest possible variance assuming the coalescent tree is known (Fu 1994a). This is because the estimation depends on determining the number of mutations of state i of the coalescent (Fu 1994a,b). This state of the coalescent process can be characterized in terms of the mutations on a branch (m_i), the size of the branch (ζ_i) or the type of branch (η_i), and the topological structure of the genealogy has no effect on estimates of θ

(Vasco *et al.* 2001). Mutations in a sample therefore fall into $n-1$ different sizes and as both K and π can be computed from the frequencies of the different classes of mutations (Fu 1994a) the latter contain more information than the former (Yu *et al.* 2002). These methods measure time in the number of generations (Vasco *et al.* 2001) and assume the infinitely-many sites model for mutations with μ_g as the mutation rate and a Poisson process for mutations (Fu, 1994a).

It should be noted that both genealogical and non-genealogical estimators of θ put different weights on mutations occurring in different time periods (Crandall *et al.* 1999; Su *et al.* 2001). Fu's UPBLUE estimator places heavy emphasis on young mutations revealing relatively recent population processes whereas, Tajima's (1983) estimate places heavy emphasis on older mutations reflecting ancient population events (Fu 1997). Comparing the two estimators has been used as a way to provide clues to how population size (N), has changed over time. For example, if the estimate calculated using UPBLUE was twice that of Tajima (Su *et al.* 2001), this would be taken as an indication of a recent population doubling (Fu 1994a; Crandall *et al.* 1999; Su *et al.* 2001). Furthermore, if a hierarchical pattern in the means and variances of Watterson, (1975) and BLUE (Fu 1994ab) estimators is observed then this signature may imply rapid population expansion (Vasco *et al.* 2001). The UPBLUE calculations were done by entering the data directly onto the website: www.hgc.sph.uth.tmc.edu/cgi-bin/UPBLUE.pl.

All of the above estimators of θ assume that the coalescent evolves in a constant environment. When this assumption is violated it may greatly affect the shape of the coalescent tree (Vasco *et al.* 2001) and failure to take account of this possibility will almost certainly lead to substantial reduction in the accuracy of any population parameters estimates produced.

Although, analytical formulas exist for the mean and variance of the coalescence time distribution and can be used to derive the distance based UPBLUE estimator described previously, this is not the case for the coalescent in a varying environment (Vasco *et al.* 2001). However, an approximation of the coalescent expectations for variable environments can be obtained by estimating the coalescent times of a gene tree using summary statistics over a set of Monte-Carlo generated genealogies (Vasco *et al.* 2001). One such least-squares estimator for variable environments EVE (Vasco *et al.* 2001), that is usually available as a free package on the internet, was used to estimate population parameters. This method is very flexible, and attempts to quantify the amount of polymorphism in a sample and to explore the causal effects of neutrality, selection (sequence environment) and recombination whilst including the role of sequence ancestry (Vasco *et al.* 2001). It is known that demographic and selective events in a species or populations history will leave a distinct signature in the pattern of branching of the gene genealogy (Hudson 1990). For example in a population having undergone a recent expansion, branch length at the tips of the tree will be longer than expected relative to

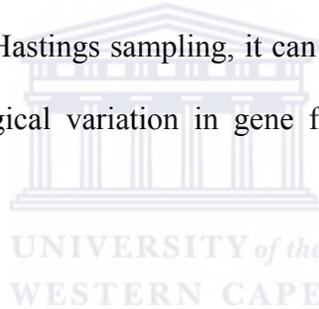
internal branches under a neutral model with constant population size (Cann *et al.* 1987; Kuhner *et al.* 1995, 1998, Vasco *et al.* 2001).

The EVE method (Vasco *et al.* 2001) compares the observed and expected pattern of branch lengths under neutrality. Utilizing the information on tree shape the EVE estimators can be derived for a variety of alternative coalescence models, including exponential and logistic population size change. It has been shown that these estimators can be applied to many different models of the coalescent in a variable environment (Vasco *et al.* 2001). This method has been shown to give a consistent and nearly unbiased estimate of demographic population parameters and will provide a more accurate estimate when compared to the non-phylogenetic summary statistic estimators based upon the average pairwise differences among sequences π (Tajima 1983), or the number of segregating sites (Watterson 1975). A major advantage of the EVE method is that it produces an unbiased estimate of the population mutation and growth rates, even when growth has been extremely rapid under which conditions the sophisticated MCMC likelihood method (Kuhner *et al.* 1998) displays substantial bias producing unreliable estimates (Vasco personal communication). The EVE program (Vasco *et al.* 2001) can usually be accessed via the Internet however, at the time of writing this thesis was unavailable. All of the EVE calculations were done on a beta version of the program kindly donated by the author and run on an IBMp655 High

Performance Computer at the Research Computing Center at the University of Georgia, Texas.

2.12.1 MCMC methods

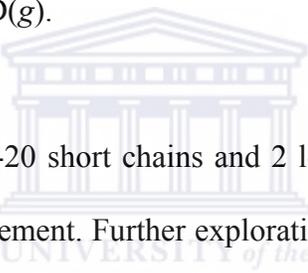
To compare the results obtained with the simple moment-based summary statistics, and least squares genealogical methods, and to assess the evidence for population expansion in euphausiid populations I used a method that employs a Markov chain Monte-Carlo approach to evaluate the integrated likelihood function for population parameters (Kuhner *et al.* 1995, 1998). Because this method utilizes the full haplotype data in samples of multiple sequences and Metropolis-Hastings sampling, it can take into account both the sampling and the genealogical variation in gene frequencies (Kuhner *et al.* 1995, 1998).



Scaled effective population size (θ), was estimated with the program Fluctuate using the Metropolis-Hastings algorithm (Kuhner *et al.* 1995, 1998) under a model of no-growth ($\theta_i = \theta_0$) and under a model of exponential growth ($\theta_i = \theta_0 e^{gt}$). Here, $\theta = 2N_e\mu$ and g is in units of μ^{-1} . This method is most sensitive to demographic change because it exploits features of the genealogy not used in other tests. This method assesses the goodness of fit of a model of exponential growth (or decline) and generates maximum likelihood estimates of the growth parameter (g) and its standard deviation. The exponential growth rate is in units of g/u_s per generation. Positive values indicate population growth and

negative values population decline. The unscaled exponential growth rate can be obtained by multiplying growth rate (g) and the mutation rate per site per generation (μ_s).

For each independent run with Fluctuate, Watterson's (1975) estimate of θ which is based on the number of segregating sites in a sample was used for the initial θ value, over various initial growth rates ranging between -100 and +100. Because these computations may show an upward bias and standard deviations are only approximate, I have conservatively used g to indicate population growth if $g > 3SD(g)$.



Individual runs involved 10-20 short chains and 2 long chains of 20000 steps with a 30 step sampling increment. Further exploration of parameter space was done including restarts at or near local peaks obtained and using the same number of steps and chains. Replicate runs with different random number seeds were conducted to check whether the chain was converging upon the ergodic stationary distribution (Kuhner *et al.* 1995, 1998).

Obtaining an unbiased estimate of growth with confidence intervals narrow enough to discriminate between for example, the case of a constant population size with coalescence estimators is notoriously difficult due to the severe lack of information about the most ancient parts of a single-gene genealogy because only a few such lineages are still present in the population (Kuhner *et al.* 1995,

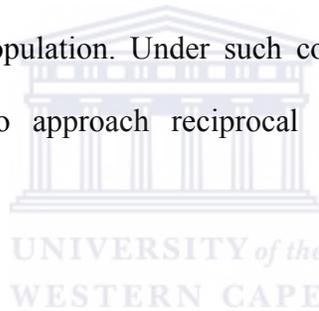
1998). It is thought that this inherent upward bias will be a common feature of all approaches to estimating this parameter (Kuhner *et al.* 1995, 1998), thus limiting the power of these methods to distinguish between alternative hypotheses such as growth vs. no growth. A comparison of the results obtained with the least squares (EVE) and maximum likelihood (Fluctuate) methods can be informative and may potentially help to better characterize this source of variation (Vasco personal comment).

For example, to assess the performance of these alternative least squares and likelihood methods, I compared the results obtained under a model of exponential growth to quantify the relative contribution of each method to the observed variances. It is, of course, quite possible that the exponential growth model is not the most appropriate for my data however, this is the only growth model currently supported by Fluctuate.

2.12.2. Two-Lineage Estimator (Likelihood Method)

All of the methods described above estimate the population parameters from a single lineage data set. For comparison, an alternative method of estimation MDIV (Nielsen and Wakeley, 2001) that uses a pair of lineages to jointly estimate multiple population parameters using Markov-Chain Monte-Carlo simulations was also employed. This method (Nielsen and Wakeley, 2001) was designed for data from one locus and derives the joint estimation of multiple parameters in a likelihood framework using the full haplotype data in samples

of multiple sequences, and assumes no intralocus recombination has occurred. This method utilizes a maximum likelihood approach to determine the probability of observing the exact sample configuration. Its use is appropriate when reciprocal monophyly of alleles is observed, as a single-gene tree is generally more informative about population structure under isolation than migration (Nielsen in Edwards and Beerli 2000). This method was recently extended to the infinite-sites model to allow for unequal population sizes between ancestral and descendant populations (Nielsen and Wakeley, 2001). However, it assumes that all populations do not change in size over time and are evolving according to selectively neutral expectations with no further subdivision within each population. Under such conditions, after separation, two taxa are expected to approach reciprocal monophyly after $> 2N_e$ generations (Wright 1953).



The simulation method used to approximate the likelihood function / posterior distribution of the parameters is related to the methods of Kuhner *et al.* (1995) and uses Metropolis-Hastings method (Metropolis *et al.* 1953; Hastings 1970). Replicate runs with different random number seeds were conducted to check whether the chain was converging upon the ergodic stationary distribution. Individual simulations were run for 40 million updates or more and took ~12 hours for completion on a reasonably fast personal computer. The program MDIV can be obtained from the website <http://www.evolutionarygenomics.dk/pgsprogram.html>

CHAPTER 3

RESULTS *Euphausia lucens*

3.1. SSCP Scoring and Sequencing

Polymerase chain reactions with the primer combination *NDI(r)* and *NDI(f)* failed for the southwest Atlantic (SWA) samples, and a further investigation revealed that mutations in the primer sites were the most likely cause. New primers were designed (Gh1 and Gh2) specific to these samples (see materials and methods), and these amplified well for specimens of *E. lucens* collected from this region. Because the addition of new primers with a different nucleotide composition from those used for the other samples, would, very likely, affect the tertiary structure of the single-stranded DNA molecules in ways not related to changes within our sequence region, the haplotype diversity in samples from this region was estimated separately, and direct comparisons among the other regional samples were made with the final sequence information. An example of the SSCP patterns produced by the experimental protocol described previously for the SWA and the southeast Atlantic (SEA) regional populations is shown in Figures 3.1 and 3.2.

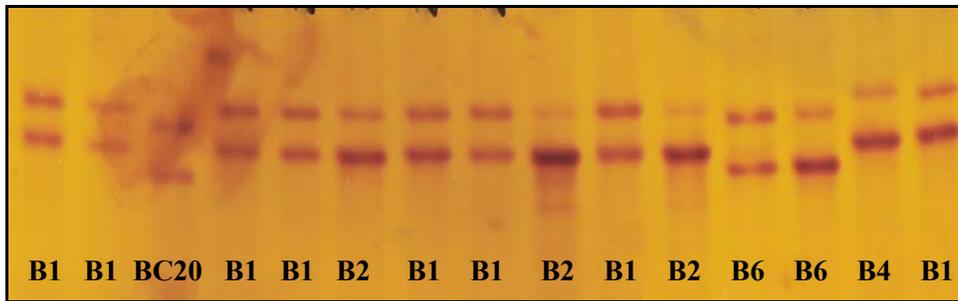


Figure 3.1. An example of the *Euphausia lucens* PCR-SSCP profiles showing a 156-bp fragment of mitochondrial *ND1* among individuals collected from Argentina.

Haplotype labels correspond to the designations from table 3.1.

Most haplotypes were sequenced more than once. Haplotypes sharing similar SSCP patterns were also mostly very similar in sequence information. However, the degree of difference was not always concordant with sequence information and often a single base substitution produced relatively large shifts in the gel mobility for a particular specimen. For example, haplotypes BC20 and B2 differed by a single substitution from haplotype B1, but only the comparison between BC20 and B1 were strikingly different gel mobility patterns observed (Figure 3.1).

In the SSCP screening of the SWA sequences amplified with the new species-specific-primers, the two major banding systems were synergistically informative on the gels. For example, haplotypes B6 and BC20 have similar upper bands, but a larger difference in migration of the lower ones: they differ by 2 bp (Figure 3.1) whereas; haplotypes B1 and B4 have similar lower bands but different upper ones and differ by 1 bp (Figure 3.1).

The same pattern was not observed in the SSCP gels for the SEA samples using the primers ND1f – ND1r and only the lower bands were of strong intensity; the upper bands included additional less intense bands (Figure. 3.2.).

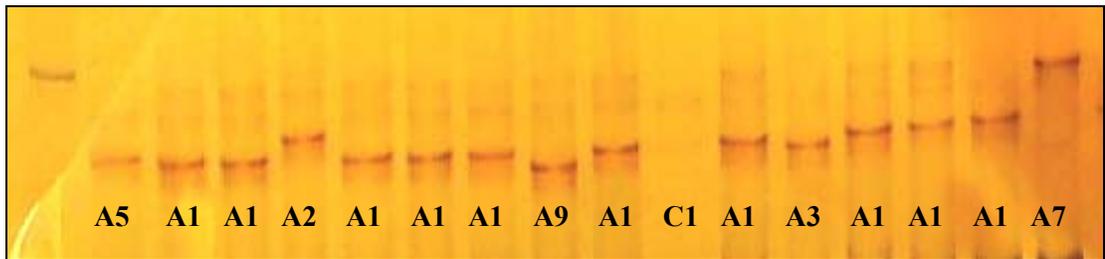


Figure 3.2. An example of the *Euphausia lucens* PCR-SSCP profiles generated for the southeast Atlantic (South African) samples. Again the ND1 haplotype designations correspond to those presented in Table 3.1.

There are several possible reasons for the presence of less intense bands: the samples may have been completely denatured or partially renatured prior to loading, there may be multiple sequences present (the result of heteroplasmy or nonspecific target amplification) or the single strands may have formed metastable conformers. Incomplete denaturation was considered to be unlikely for several reasons. The foremost being, that the presence of extra bands was reproducible between PCR's of different samples of the same species between gels. The presence of multiple sequences was an unlikely possibility, particularly, when the additional bands were of weak intensity, as no evidence of multiple banding was observed in the agarose gel PCR electrophoresis or as multiple nucleotide peaks in the chromatograms produced after sequencing.

Faint additional bands were therefore, tentatively attributed to the presence of metastable conformers (Zehbe *et al.*, Application note 384). These are identical in sequence to those of the primary bands but have an alternative conformation, which affected their mobility relative to the primary conformer. The presence of additional upper bands did not, however, affect the ability of SSCP to effect accurate identification of *E. lucens* haplotypes.

For the SEA samples, the degree of difference in the gel mobility patterns was generally concordant with sequence information. For example, haplotypes A1 and A9, A2 and A9 and A3 and A7 differed by one, two and three substitutions respectively (Figure. 3.2) and these differences are clearly correlated with the gel mobility patterns with the most striking difference displayed between the latter haplotype pair.



Some sequences showed consistently pale phenotypes (e.g. C1), even though they amplified strongly as visualized on agarose/ethidium bromide stained gels (Figure 3.2). Sequencing the haplotypes for different individuals with the same SSCP pattern revealed no undetected mutations. It was however, often necessary to rerun samples from the same gel next to each other, as very often similar ones were not sufficiently close for easy comparison and/or occasional gel irregularities particularly at the edge of the gel disrupted certain comparisons (Figure 3.2). Often, despite cooling the gel apparatus during the runs, an upward curvature of the bands towards the outer edges of the gel was

observed, indicating that a temperature differential had developed between the central and peripheral lanes in the gel (Figures 3.1 and 3.2.). When this occurred, and when some sequences displayed very similar SSCP gel mobility patterns it was necessary to load the same sample at regular intervals across the entire gel along with a marker sequence of known haplotype designation. Taken together, the results suggest that SSCP screening of *E. lucens* sequences has efficiently detected moderately high levels of variation in this DNA fragment.

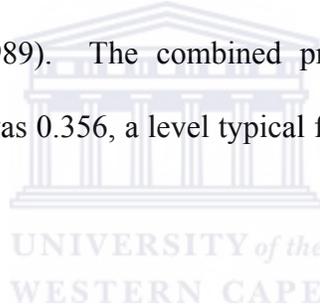
3.2. Mitochondrial Sequence Variation

Five hundred and ninety-two specimens of *E. lucens* from the southwest (SWA) and southeast (SEA) Atlantic and the southwest Pacific (SWP), were examined (Table 3.1). Thirty-eight distinct *NDI* haplotypes were identified of which, twenty-four were singletons. The highest number of haplotypes was found in Argentina (20), followed by South Africa (17), and New Zealand (2). Haplotype designation and their frequencies are shown in Table 3.1.

With the putative sister species *E. vallentini* as an outgroup, the total *E. lucens* *NDI* data set contained 39 variable sites, 31 segregating within *E. lucens* sequences, with a transition/transversion ratio of 29.52, which accounted for 30 synonymous and 2 non-synonymous substitutions. Of the 32 mutations, 11 were purine/purine transitions, 20 were pyrimidine/pyrimidine transitions, and 1 was a purine/pyrimidine transversion. The *NDI* haplotypes differed from



each other by between one to seven substitutions. The non-synonymous changes within *E. lucens* both occurred at 2nd codon positions and involved a substitution of asparagine with serine A₁₄₇ → G₁₄₇ and a substitution of serine with asparagine G₁₅₃ → A₁₅₃ (Table 3.1). In both cases, the non-synonymous changes involved the replacement of one functionally similar amino acid for another (Dayhoff *et al.* 1972), and each was found in one individual only (A3 and A16), collected from the west and south coasts of South Africa respectively. Similarly, all of the sites segregating between species involved synonymous substitutions suggesting a widespread occurrence of slightly deleterious mutations consistent with nearly neutral theory (Ohta 1992, Kimura 1963, Rand and Kahn 1989). The combined proportion of guanine and cytosine nucleotide bases was 0.356, a level typical for the A-T rich crustacean genome.



3.3 Phylogenetic Analysis

A median joining network (Bandelt *et al.* 1999) was constructed that reveals a clear geographical pattern which can be superimposed onto the network with two distinct reciprocally monophyletic clades: the first, corresponding to the SEA, the second, to the area of ocean covering the SWA and including three individuals collected in the southwest Pacific (Figure 3.3).

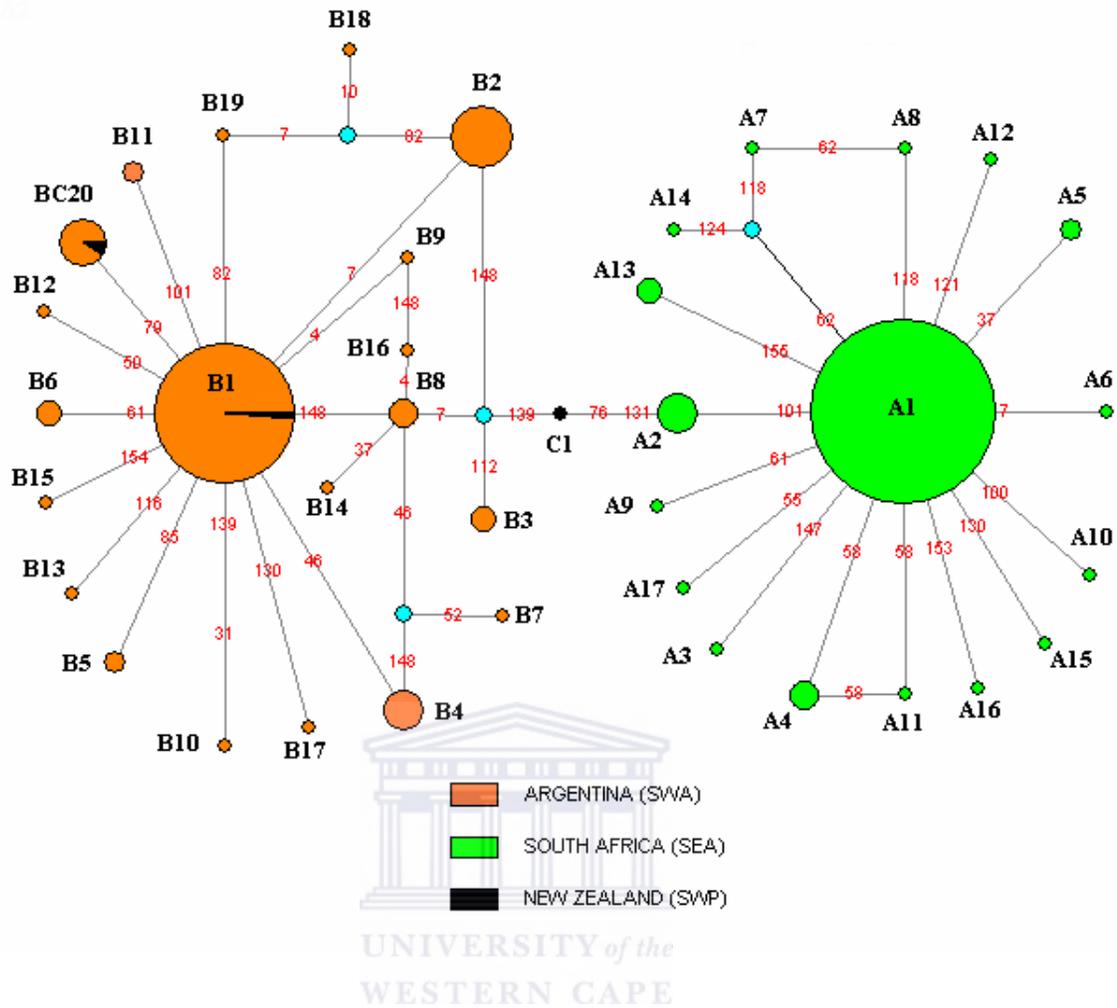


Figure 3.3. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. lucens* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences: orange, Argentina, green, South Africa and black, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency. Blue circles indicate missing intermediates. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

These same clades were consistently recovered in the majority-rule consensus trees constructed by minimum evolution (Figure 3.4) and the maximum likelihood (Figure 3.5) albeit, with moderate bootstrap support. Obtaining high

bootstrap support values is very unlikely when the divergence between clades is low (Bandelt *et al.* 1999; Zhang and Hewitt 2000), as is typically observed with intraspecific data.

Although traditional phylogenetic methods are not the most appropriate in determining genealogical relationships among sequences for a variety of reasons (Posada and Crandall 2001), this result provides some support for the suggestion that the two regional populations have been isolated long enough to become genetically isolated, and that substantial gene flow with individuals crossing the open ocean has not occurred in the recent past.

Results of the most appropriate substitution model estimated in ModelTest using a likelihood ratio test was HKY + G ($2(\ln L_1 - \ln L_0) = 0.025$, $df = 1$ $P > 0.437$) with 5 rate categories (K) and a gamma distribution shape parameter (G) of 0.40, $\text{Pinvar} = 0$ and a TRatio of 43.15.

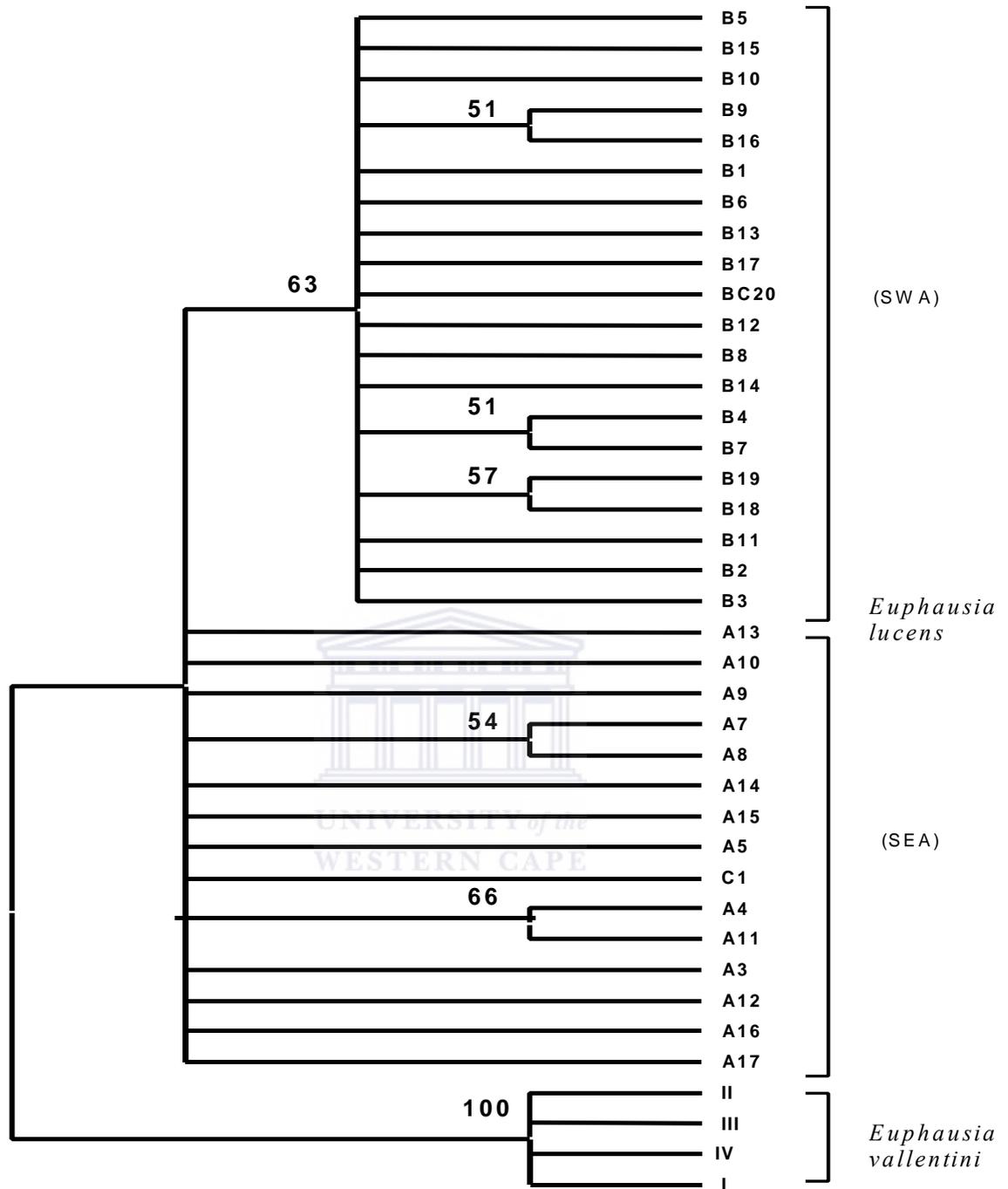


Figure 3.4. The minimum evolution trees for *E. lucens* ND1 sequences using the most appropriate substitution model as determined hLRT using ModelTest (see text). Trees were rooted with four *E. vallentini* sequences. Numbers above branches represent bootstrap values.

The high frequency haplotypes were internal to the tree, and a number of rare or singleton haplotypes were derived from each high frequency type (Figure 3.3). A shallower topology was observed in the SEA clade, where, apart from the most common haplotype, only two other haplotypes (A4, A8) were positioned internally in the network. By contrast, seven haplotypes (B1, B2, B4, B8, B9, B16, B19), including the most frequent, were placed internally in the SWA clade. Two of these however, were only placed internally because of the insertion of missing intermediates (blue circles) by the analysis program.

Although this suggests that the SWA may be the older of the two clades, both clades are still relatively shallow, perhaps implying a high turnover of variants of polymorphism. Also, the relative frequency of the rare haplotypes is lower in the SEA population even though the sample size was three times larger; suggesting that the higher variation observed in the SWA population accurately characterizes this population and if anything, has been underestimated in this investigation (Table 3.2).



Connecting the center of the two main clades is a branch six mutations long, two of which are fixed, and placed on this branch is haplotype C1, representing a single individual collected from New Zealand (Figure 3.3). The haplotypes differed from each other by between one to seven substitutions, and all but four haplotypes (A2, C1, B6, and B10), differed by a single substitution from their nearest neighbour. One haplotype (B1) was shared between New Zealand and the SWA where, in the latter region, it occurred at the highest frequency in all of the local samples (Figure 3.6).

A total of four missing intermediates have been inserted in the genealogy (Figure 3.3), one of which is positioned on the branch connecting the two major clades, while the remainder form alternate pathways represented here as loops, two in the SWA, and one in the SEA clade respectively. Ambiguous relationships among haplotypes (i.e. reticulations) were twice as common in the SWA compared to the SEA, a result of the homoplasy observed at sites 4, 7, 46, 82 and 148 (Figure 3.3). Most conflicts could be resolved with the frequency criterion of Crandall and Templeton (1993). Haplotype B18 provides an example. This haplotype shares purine/pyrimidine transitions at sites 7 and 82 with haplotypes B19 and B2. However, the latter haplotype was much more

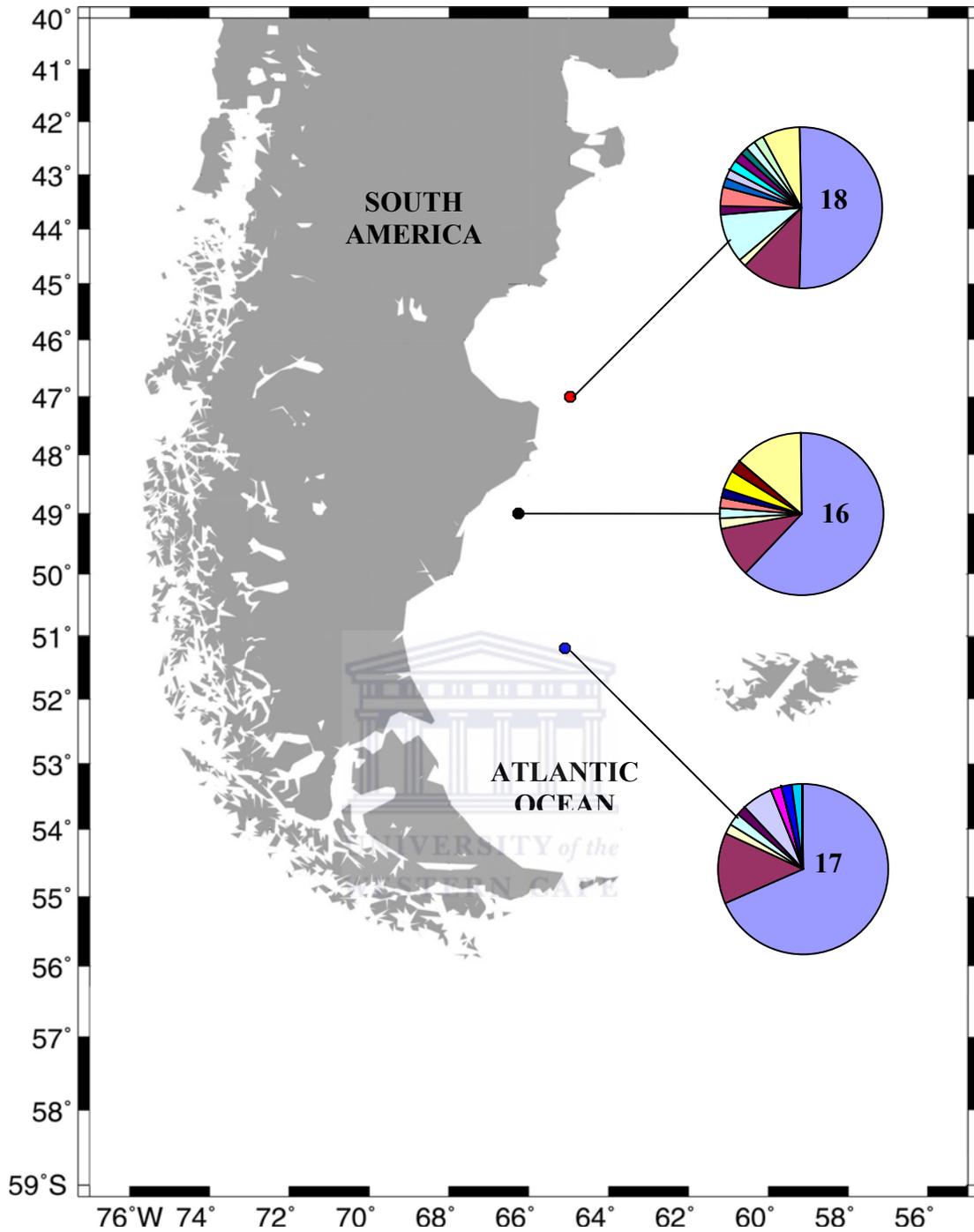


Figure 3.6. Map showing the three sampling locations for *Euphausia lucens* over the Patagonian Continental Shelf in the southwest Atlantic. The pie charts show the frequency of *NDI* haplotypes among samples. The same haplotype (B1) was the most frequent in each of the separate samples. Sample numbers shown in pie charts.

frequent than the former, which, occurs as a singleton. Therefore, the genealogy was resolved as shown in Figure 3.7. All conflicts in the data set could be resolved in this manner. Further, none of the conflicts in the network were between haplotypes from different populations, precluding the use of the geographical criterion to provide additional support to any particular topology produced for this data set. When the loops have been resolved this way the SWA, unlike the SEA clade, contains several sub-clades, each dominated by a single haplotype at an overall intermediate frequency, with rare haplotypes as descendants (Figure 3.7).

The level of divergence between the two main clades involved four substitutions, all non-synonymous, representing a 4.8% (p-uncorrected) difference that indicates that either these lineages split relatively recently or that the mutation rate in this mtDNA fragment is relatively low. The excess of rare *NDI* alleles that are singletons is less pronounced in the SWA population, where several haplotypes achieve intermediate frequency (Figure 3.7). It is interesting to note that one of the New Zealand haplotypes is identical to the most common haplotype in the SWA.

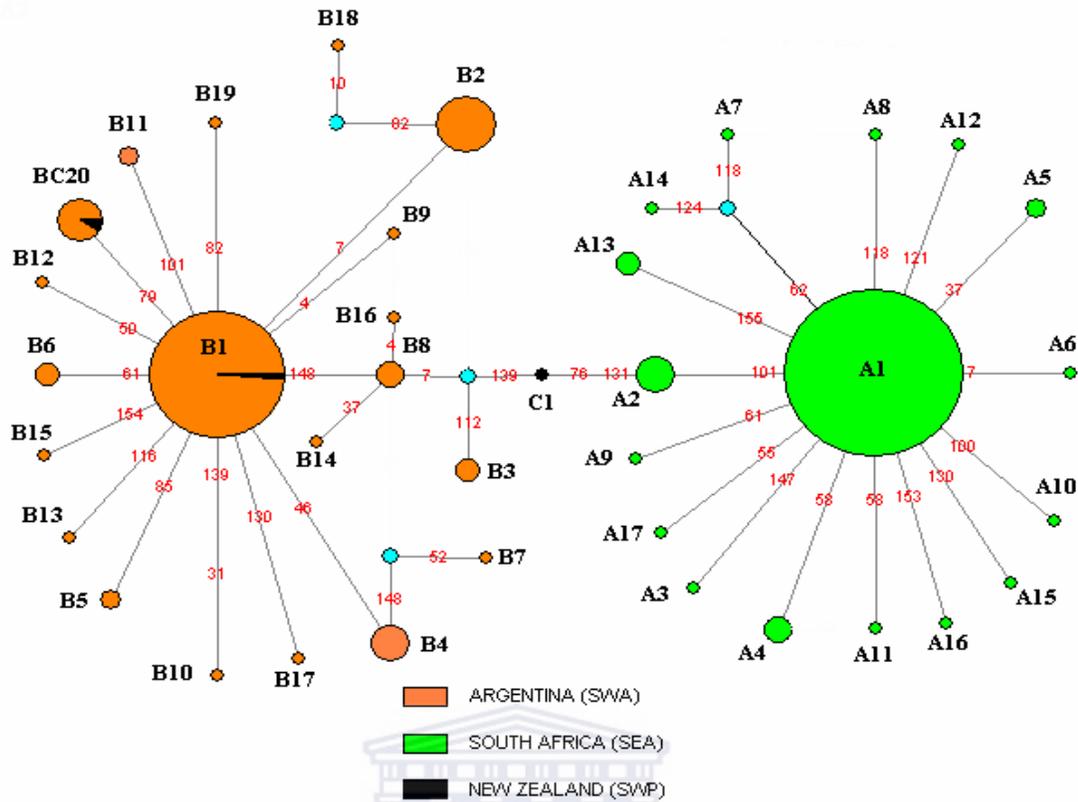


Figure 3.7. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. lucens* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences with the loops resolved using various coalescent criteria (see text).

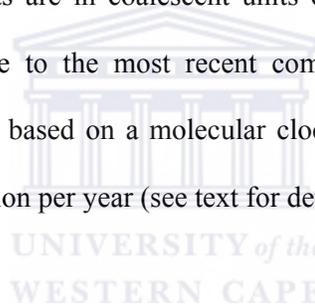
3.3.1. MDIV

Recently developed methods for fitting the full isolation model with migration were applied to the data for the two regional Atlantic populations using the program MDIV (Nielsen and Wakeley 2001). The coalescence based population parameters, including the maximum likelihood estimators of divergence times (T), number of migrants (M), and θ_{ML} for the two regional

euphausiid populations were estimated with MDIV and the results for the southeast and southwest Atlantic clades are listed in [Table 3.3](#).

Comparison	Theta	N _m	PDT	TMRCA	PDT (MYA)	TMRCA (MYA)
SWA-SEA	2.262	0.004	1.42	2.05	1.14	1.65

Table 3.3. Results of the coalescent analysis of *E. lucens* populations based on NADH1 (ND1) sequence data. Theta, the migration parameter (N_m) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see text for details).



To convert to estimates of percentage divergence time (PDT) and expected time to the most recent common ancestor (TMRCA) to actual time (millions of years ago, MYA) I used an averaged molecular clock estimate based on the mitochondrial CO1 sequences of another crustacean, the snapping shrimp *Alpheus spp.*; (Knowlton *et al.* 1993; Knowlton and Weigt 1998), which ranged from 1.4 to 2.2% sequence divergence between pairs of lineages per million years. I also assumed generation times of one year for *Euphausia lucens* (Stuart and Pillar 1988). To convert coalescence-based parameters to estimates in millions of years, I assumed an average of 1.8% sequence divergence per

million years and for a fragment 156 base pairs in length thus obtained a value for the mutation rate, u of 2.81×10^{-6} mutations per sequence per generation. Using this estimate, the effective population size (N_e) was calculated from the estimated θ_{ML} using the formula $N_e = \theta_{ML} / 2\mu$ and then used to convert the T values scaled by N_e , in number of generations. Applying the previously described mutation rate to estimates of PDT estimated the time of separation between the southeast and southwest Atlantic basins to be 1.14 MYA, indicating middle Pleistocene divergence with the most common ancestor suggested to have occurred in the early Pleistocene (1.65 MYA). Although models of gene flow between populations vary in their assumptions, it has been shown that practically any migration, however small, will be sufficient to preclude genetic divergence (Maynard-Smith 1989; Slatkin 1985); the results of this analysis imply that *Euphausia lucens* populations separated by the Atlantic Ocean have been isolated without gene flow since they began to diverge.

3.4. Rooting of the Intraspecific Genealogy

All outgroup analyses under the two phylogenetic methods employed (Figure 3.4 and 3.5), confirm the reciprocal monophyly of *NDI* sequences from *E. lucens* and *E.vallentini*. More surprisingly, using ML, and ME, the most basal lineage in the *Euphausia lucens* clade is located in South Africa, suggesting that the Argentinean population form a monophyletic group nested among Atlantic haplotypes.

With an outgroup sequence (*E. vallentini*), which allows us to infer the ancestral nucleotides of each segregating site; a haplotype evolutionary network was constructed (Figure 3.8), using the median-joining method (Bandelt *et al.* 1999). Including all the haplotypes in the outgroup analysis however, leads to the insertion of numerous putative missing intermediates using this network construction method. This results in an increase in reticulations, rendering data visualization and interpretation difficult. This is presumably due to the increase in alternative pathways (homoplasy), between a relatively distant outgroup, and the respective ingroups where sequence similarity is high (within clades). Indeed this lack of power to resolve among alternative connections is a problem inherent in a star-shaped phylogeny where the average number of mutational steps between the constituent sequences within the ingroup is low (Figure 3.3). It has been shown that haplotypes of low frequency usually occur at the tips of cladograms, while those of high frequency occur in the interior (Excoffier and Langaney 1989, Crandall and Templeton 1993), and that there is an age to frequency relationship (Donnelly and Tavaré 1986) with older alleles (high frequency) having a greater probability of producing mutational derivatives and therefore, becoming internal, than do younger haplotypes (low frequency). Therefore only those haplotypes whose frequencies were > 1 were included in the outgroup analysis using a M-J network

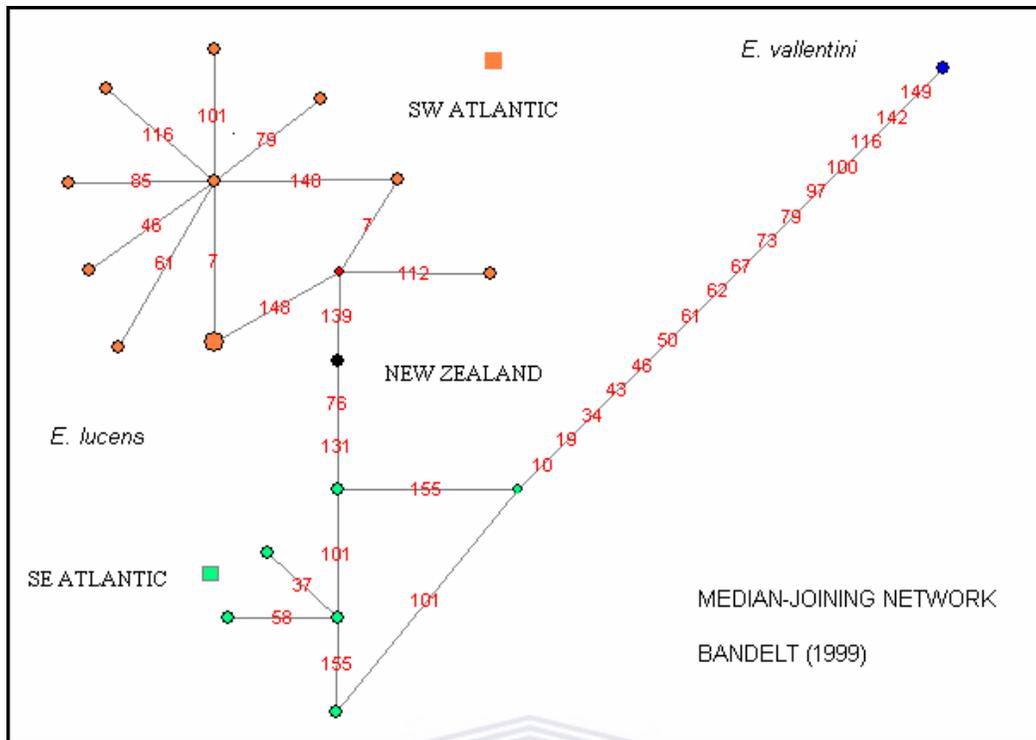


Figure 3.8. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of *E. lucens* mtDNA haplotypes (filled coloured circles) rooted with the sister group *E. vallentini* based on *ND1* sequences: orange, SWA, green, SEA and black, SWP. A red diamond indicates a missing intermediate. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Again, the MRCA of the Atlantic *E. lucens* sequences is inferred to have occurred in the SEA, with 17 and 19 substitutions respectively, representing the minimum and maximum distances between the *E. vallentini* outgroup sequence, and this clade (Figure 3.8). For the SWA clade, this distance was 21 and 22 substitutions respectively, and for New Zealand 20.

Using the frequency criterion, the position of the MRCA in the Atlantic for this species would be in the SEA, although this estimate is biased towards this population due to the relatively large differences in sample sizes employed (153 and 438) between regions. With the topology criterion, the SWA or the SEA could be the root, and using the total number of occurrences in different populations, either New Zealand or the SWA. The fact that the SWA region displayed the highest diversity, as evidenced by the h , π and θ statistics (Table 3.2), despite comprising the smallest sample, provides additional support for its candidacy as the site of the MCRA. From the phylogenetic evidence provided here for the Atlantic Ocean, the southeast basin appears to be the more likely location of the most recent common ancestor (MRCA) in *E. lucens*. However, the differences between the numbers of steps separating the two clades from the *E. vallentini* outgroup are not large (Figure 3.8). Additionally, the shared haplotypes between the SWA and the SWP suggests that a Pacific origin for this species cannot be ruled out at this time. As only two individuals were collected from Pacific waters, further data will be required for inferences regarding the location of the MRCA in this species to be made with any confidence.

3.5. Spatial and Temporal Comparisons

As well as harbouring the highest number of *NDI* haplotypes, the SWA population had the highest haplotype diversity (h), nucleotide diversity (π), and

number of segregating sites (S) (Table 3.2). This pattern was consistently observed in all of the separate samples as well as the combined data set. The high and dominant frequency of haplotype A1 (Table 3.1) in the SEA was responsible for the low haplotype and nucleotide diversities there because frequencies of haplotypes influence these statistics. The differences between countries were smaller in their estimates of θ (Table 3.7) based on the number of segregating sites, which are governed by sample size and not by frequencies of haplotypes. However, nucleotide diversity was still relatively low in both south Atlantic basins reflecting the fact that most haplotypes differ by a single substitution .

No evidence of any subdivision is apparent among the three SWA samples, with four haplotypes present in all the samples, the most common (B1) ranging in frequency from 2 to 68% (Figure 3.6). However the number of “private alleles” (Slatkin 1985) was relatively high with three found in samples 16 and 17 and seven in sample 18 (Figure 3.6). The latter sample also had the highest number of haplotypes (14), and this is reflected in the high average pair-wise differences between sequences and haplotype and nucleotide diversity values (Table 3.2). Estimating the ratio of transitions to transversions was difficult for the SWA sequences as all of the observed polymorphisms were transitions, twelve of which were pyrimidine/pyrimidine, and four, purine/purine.

Temporal comparisons between samples collected from the SEA in the years 2000, and 2001, revealed an absence of any significant inter-annual variation in diversity as estimated by summary statistics (Table 3.2). In the years 2000 and 2001, eight and eleven *NDI* haplotypes respectively were found. The most frequently observed haplotype was the same in both years, and sample-specific haplotypes were common due to an excess of rare alleles, with 72.22 % of haplotypes occurring as singletons. Of the 17 *NDI* haplotypes recorded for the SEA, only three were shared between years (A1, A2 and A14). Of the remainder, 11 were singletons, and 3 occurred at very low frequencies (Table 3.1). In the SEA data there were a total of 15 transitions a single transversion. Eight of these transitions were pyrimidine/pyrimidine, and seven were purine/purine. The single transversion involved the substitution of G₅₈ - T₅₈, and occurred in two individuals, at a first codon position (Table 3.1). The absence of significant temporal or spatial genetic structuring within the SEA population enabled combining samples for increased statistical power.

3.6. Spatial distribution and Direction of Mutations

I examined the pattern of mutations to see if there is any unusual feature that could affect the subsequent analyses (Yu *et al.*, 2003). All of the 32 mutations in all the sequences could be inferred for the direction of mutation i.e. which nucleotide is ancestral and which is the mutant. Table 3.4 shows the pattern of mutations. For the *E lucens* sequences the number of mutations from x to y (x, y = A, G, C, or T) shows an excess T to C.

anc mutation	A	G	C	T
A....	0	8	0	2
G....	8	0	0	2
C....	0	1	0	4
T....	0	1	13	0

Table 3.4. Patterns of nucleotide change in *ND1* sequences of *E. lucens*.
Anc = Ancestral mutation derived from *E. vallentini*

I examined the spatial distribution of mutations. The entire *ND1* region of 156 nucleotides was divided into 10 regions of equal length (16 base pairs) and the occurrences of mutations in the ten regions were 3, 2, 4, 5, 5, 3, 3, 5, 6, and 2 among the sequences. A chi-square test cannot reject the hypothesis that mutation rates in all regions were the same ($\chi^2 = 4.72$, $df = 9$, $p > 0.5$).

3.6.1. Frequencies of Mutant Nucleotides in the Sample

The frequencies of mutations in a sample can be used for estimating population parameters and for inferring the evolutionary forces that have operated on the locus (Yu *et al.* 2003). Mutations can be classified into different sizes and the observed and expected frequencies under the neutral Wright-Fisher (1959) model can be compared (Table 3.5). An excess or deficiency of a particular class of mutations in a sample can provide clues to which type of evolutionary process may be responsible for the observed pattern of polymorphisms. A conspicuous excess of singletons was found in the total (Table 3.5) and separate *E. lucens* sub-samples (Table. 3.6).

Total		
Mutation Size	Occurrences	Expectation
1	21	4.934
2	7	2.467
3	4	1.645
7	1	0.704
10	1	0.493
18	1	0.274
150	2	0.033
430	1	0.011
438	2	0.011
591	1	0.008
others.....	0	30.43
Total.....	41	41

Table 3.5. Frequency of mutations of various sizes in *NDI* mtDNA sequences of *E. lucens* for the entire sample.

Size	SEA (k = 17)		SWA (k = 20)	
	Count	Expect	Count	Expect
1	9	2.298	13	3.084
2	3	1.149	3	1.542
3	1	0.766	2	1.028
7	0	0.328	1	0.441
8	2	0.287	0	0.386
18	0	0.128	1	0.171
142	0	0.016	1	0.022
149	0	0.015	1	0.021
152	0	0.015	1	0.020
437	1	0.005	0	0.007
others	0	10.993	0	16.278
Total	16	16	23	23

Table 3.6. Frequencies of mutations of various sizes in the *E. lucens* for the *NDI* mtDNA sequences for the regional sub-samples. K = the number of haplotypes.

In addition to the excess of mutations of size 1 the SWA sample displayed an excess of mutations of high frequency, and this is reflected in the shapes of the allele frequency spectra for these populations (Figure. 3.9a-b).

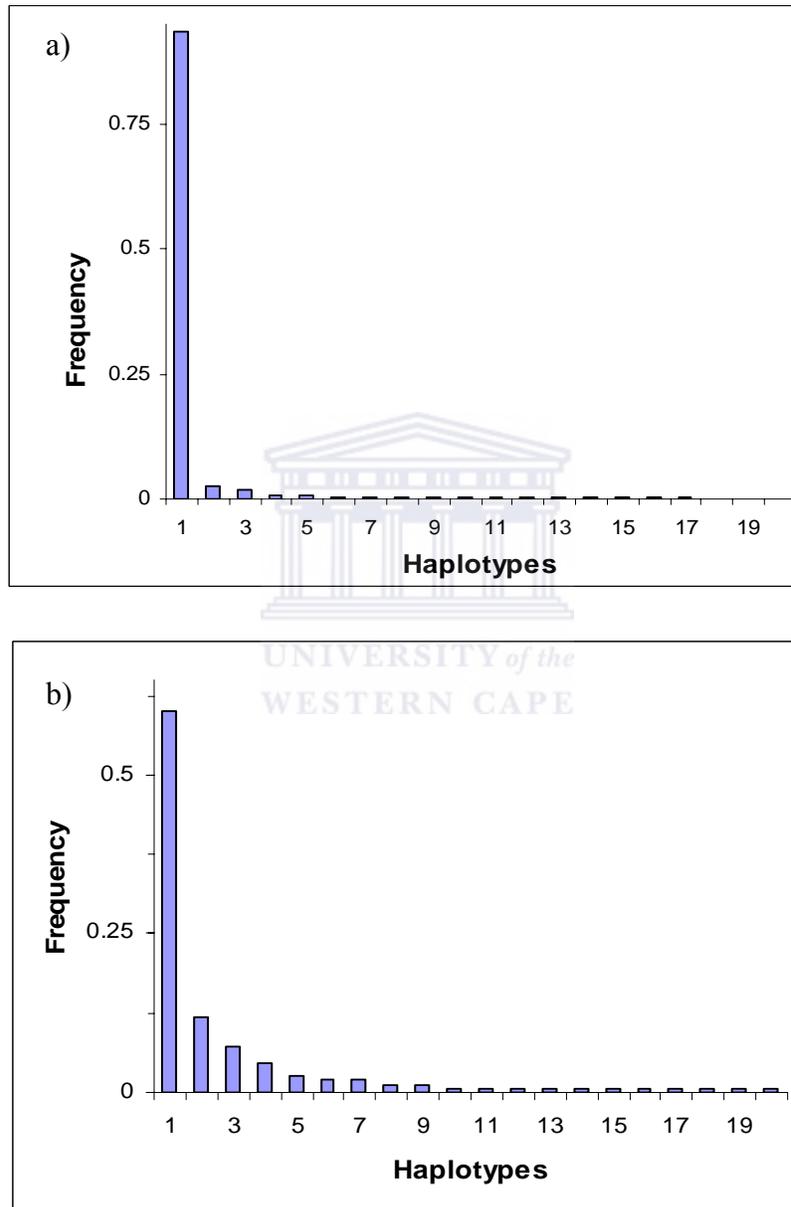


Figure 3.9. Haplotype frequency spectrum (*NDI*) for *E. lucens* populations a) SEA b) SWA

3.7. Neutrality Tests

All of the neutrality tests employed implicitly assume that the data fit an infinite-sites model of nucleotide substitution. Therefore, when multiple hits at some sites of DNA sequences are evident, certain corrections are required before applying these tests (Fu 1997). All of the SEA samples violate this assumption. Therefore, the values of the variables used in the tests for these samples were calculated from the sample genealogy estimated by maximum parsimony, an efficient way to deal with the problem of multiple hits (Fu 1994b, 1996). Calculating the F and D values (Fu and Li 1993) requires the inclusion of an outgroup sequence which should be from a closely related population or species to avoid the complication caused by parallel and back mutations (Fu and Li 1993). The values in these tests for the separate south Atlantic populations of *E. lucens* were calculated with *E. vallentini* as an outgroup.

A significant departure from mutation-drift equilibrium was indicated for the SEA populations using all of the four tests (Tajima's D , Fu and Li's F and D and Fu's F_s) whereas, in the SWA population, Fu and Li's F and D tests were not significant. The values of all the neutrality test statistics were negative for both regional populations (Table 3.7) indicating an excess of rare alleles confirming the impression from a visual inspection of Table 3.6 and Figure 3.9a-b. The relative proportions of external and singleton mutations in the different populations are also presented in Table 3.7 and a careful examination

of these statistics can help us to understand the reasons why the F and D neutrality tests for the SWA data set failed to detect a significant excess of rare mutants.

Samples	n	%e	%s	Fu & Li's D	Fu & Li's F	Fu's F_s	Tajima's D
SEA Total	438	53	60	-3.909 ($P < 0.001$)	-3.928 ($P < 0.001$)	-35.719 ($P < 0.0001$)	-2.245 ($P < 0.01$)
SEA 2000	202	57	71	-2.708 ($P < 0.05$)	-2.907 ($P < 0.02$)	-13.104 ($P < 0.03$)	-1.929 ($P < 0.05$)
SEA 2001	236	64	64	-3.880 ($P < 0.02$)	-3.876 ($P < 0.02$)	-17.475 ($P < 0.0001$)	-2.125 ($P < 0.01$)
SWA	152	29	47	-1.097 ($P > 0.10$)	-1.706 ($P > 0.10$)	-18.929 ($P < 0.001$)	-2.004 ($P < 0.05$)
Total	590	30	40				

Table. 3.7. The results of the neutrality tests based on NDI for the various *E. lucens* data sets. %e = percentage external mutations, %s = percentage of singleton mutations.

Because D and F tests compare the relative proportion of external mutations to the total number of mutations and the average difference between pairs of sequences respectively, a significantly negative test result indicates an excess of external mutations (Fu and Li 1993). From Tables 3.2 and 3.7 it can be seen that the SWA population which had the highest number of mutations and nucleotide diversity also had lowest proportion of external and singleton mutations producing the least negative values of D and F and explaining the

non-significant results obtained. Although the result is non-significant for the SWA data set, the test is skewed towards an excess of mutations of size 1, as evidenced by the negative values produced for both the D and F statistics (Table 3.7).

For all the data sets, with the exception of South Africa 2001, F values were more negative than D due to the higher proportions of singleton compared to external mutations. In the 2001 data set, both types of mutations were in similar proportions, and this is reflected in the similar values observed with the F and D tests (Table 3.7). For the SWA, all of the different test statistics values are less negative than those obtained for the SEA (Table 3.7), and this may be due, in part, to the greater nucleotide diversity in the former population, that was at least one order of magnitude higher than the latter (Table 3.2).

Tajima's test (Tajima 1989) produced similar values for the separate regional populations with the most negatively significant values observed in the SEA 2001 data set where size 1, and intermediate frequency mutations (25-75%), were in excess (data not shown). A polymorphism of intermediate frequency appears as a difference between many pairs of sequences and contributes more to Π than does a low frequency mutation (Hudson 1990). In contrast, mutations of low frequency contribute more to θ_w ; (Hudson 1990, Fu 1994a). When both low and intermediate frequency mutations are in excess, one inflates the value of the numerator (Π) and the other inflates the value of the denominator (θ_w)

resulting in a small difference between θ_w and Π . To obtain a negatively significant value in this sample the excess of low size mutations would have to be relatively large.

Fu's (1997) F_s statistic, which is sensitive to the presence of rare mutations relative to expectation based on nucleotide diversity, was highly significant for both regional populations with the most negative values displayed for the SWA (Table 3.7). All four neutrality tests conducted separately on the SEA data for the different years were negatively significant, more so in 2001 compared to 2000 (Table 3.7),

For all the tests, the magnitude of difference between the two main regional populations is less affected by the contribution of the year 2000, compared to 2001 (Table 3.7). A separate analysis revealed that in both years, the value of F_s increases (becomes less negative), relative to the SWA population, and in the year 2000 this effect is also seen in Tajima's statistic. As the D and F tests of Fu and Li (1993) were only marginally insignificant, negative selection cannot be ruled out.

It is expected that selection would produce similar polymorphism patterns in the South African and Argentinean populations (Yu *et al.* 2001), and that is the case here. However, although both regional populations show an excess of size 1 or young mutations, the proportion of older, larger size mutations is greater in

the SWA (Table 3.6). It should be noted that no shared haplotypes were observed between the separate regional populations. This is not likely the effect of limited sampling effort given the sample sizes employed and implies recent genetic exchange has not occurred between the two regional populations, as indicated by the results of the phylogenetic analysis.

The probability of observing reciprocal monophyly of sequences between populations' increases with time since the most common recent ancestor, and under the nearly neutral mutation model, should be achieved when the $4N$ generations have passed since these populations diverged. However, if the genetic variation in the sampled population is organized into separate demes, and one samples only two of these, this probability will also be inflated (Arbogast *et al.* 2002). The data required to test this hypothesis would necessarily involve collecting samples at regular intervals along a longitudinal transect across the south Atlantic Ocean that would involve substantial financial resources not currently available to this study.

3.8. Estimating the Effective Population Size

An essential parameter of a population is θ . For a mitochondrial locus, $\theta = 2k N_e \mu$ where N_e is the effective population size, and μ is the mutation rate per generation ($k = 1$ haploid, $k = 2$ diploid) and is relevant to almost all the statistics one can compute from the polymorphism in a sample (Fu and Li 1993). Thus, its value is critical in understanding how the population has

evolved. Estimates of effective population size N_e , calculated from the θ values are presented in Table 3.8.

θ estimated from several estimators			
Methods	Total Sample	South Africa	Argentina
Tajima	2.433	1.045	1.316
Watterson	4.934	2.298	3.085
Fu's UPBLUE	10.472	4.546	6.056
Watterson ^c	1.418	1.155	1.259
EVE	120.12	80.112	110.020
Fluctuate	n/a	30.264	10.842

vc^c singleton mutations removed

Table. 3.8. The estimates of the population mutation rate parameter θ for the *E. lucens* *NDI* samples using a range of different estimators. $\theta = 2Nu$.

For the entire sample there is a considerable range in the θ estimates. For $\mu = 2.81 \times 10^{-6}$ per generation, the Tajima (1983), and Watterson (1975) estimates of θ give effective population sizes (N_e) of 0.43×10^6 and 0.88×10^6 individuals respectively (Table 3.8). The UPBLUE (Fu 1994a) estimator produced even higher values (1.87×10^6). When singleton mutations are excluded the Watterson estimate θ_1 , yields an estimate of 1.418, that corresponds to an effective population size of 0.25×10^6 individuals.

Considering each sub sample separately: with all the estimators the highest θ values were obtained for the SWA population, providing support for its candidacy as the site of the MRCA. For this population, effective population size estimates of 0.234×10^6 , 0.549×10^6 and 1.08×10^6 individuals were obtained with the Tajima (1983), Watterson (1975) and UPBLUE (Fu 1994a) estimates respectively. For the SEA population the same estimators produced estimates of effective population size of 0.186×10^6 , 0.409×10^6 and 0.81×10^6 . As was observed for the entire sample, estimates of θ for separate sub samples are even lower, indicating effective population sizes of 0.23×10^6 and 0.21×10^6 individuals for the SWA and SEA populations respectively (Thus, reflecting the relative contribution to these statistics from mutations of size 1). Note that all of θ estimators vary consistently for the two subsamples and the total sample. For example, the UPBLUE estimate of θ (Table 3.8) for the entire population, is four times as large as the Tajima estimate, a pattern also seen in the separate SW (UPBLUE $\theta = 6.056$ / Tajima's $\theta = 1.316$), and SE (UPBLUE $\theta = 4.54$ / Tajima's $\theta = 1.041$) Atlantic populations.

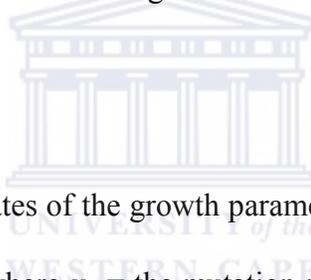
All of the methods of estimating θ employed previously implicitly assume that the coalescent evolves in a constant environment. However, the accuracy and reliability of coalescence-based parameter estimates will be markedly affected when this basic assumption is violated (Vasco *et al.* 2001; Hudson 1990). The

results of the neutrality tests indicate that this assumption may not hold for *E. lucens* populations, therefore, θ was also calculated using an estimator for variable environments with the EVE package (Vasco *et al.* 2001) that obtains approximations to the coalescent expectations using Monte-Carlo generated genealogies under a variety of alternative coalescent models.

Under the exponential model of population growth, the EVE (Vasco *et al.* 2001) estimates of θ are higher than the UPBLUE ones (Table 3.8), for both the total sample ($\theta = 120$, $r = 143$) and the separate sub samples, and rapid growth is indicated in both the combined, and separate SW ($\theta = 110$, $r = 149$), and SE ($\theta = 80$, $r = 142$) populations. Applying the same mutation rate as before (page 96) these estimates give effective population sizes of 1.96×10^7 and 1.42×10^7 individuals for the SWA and SEA populations respectively.

A major criticism of the UPBLUE methods is that they rely on a single genealogy, although Fu (1994a) analyzed 100 most parsimonious trees and concluded that the variance differed little among trees. To provide a comparison, θ was estimated using the program Fluctuate (Kuhner *et al.* 1995, 1998) that takes a maximum likelihood approach that is evaluated by Markov Chain Monte Carlo methods. This method uses the Metropolis-Hastings algorithm to evaluate the likelihood's and estimate parameters, while the topology of the trees is treated as a nuisance variable.

Under the model of no growth, the Fluctuate maximum likelihood estimates of θ were substantially higher in the SE ($\theta_{SEA} = 0.139$), compared to the SW Atlantic ($\theta_{SWA} = 0.0520$). Giving effective population sizes of 3.86×10^6 and 1.44×10^6 individuals respectively for these populations. The highest log-likelihood likelihood's were obtained under the exponential growth model for both south Atlantic populations however, with even higher estimates of θ produced ($\theta_{SEA} = 0.388$ and $\theta_{SWA} = 0.110$) corresponding to effective population sizes of 5.39×10^6 and 1.92×10^6 individuals respectively. In contrast to the previous results (Table 3.8), the maximum likelihood estimates of θ for the Atlantic suggest that the highest effective population size is in the SEA.

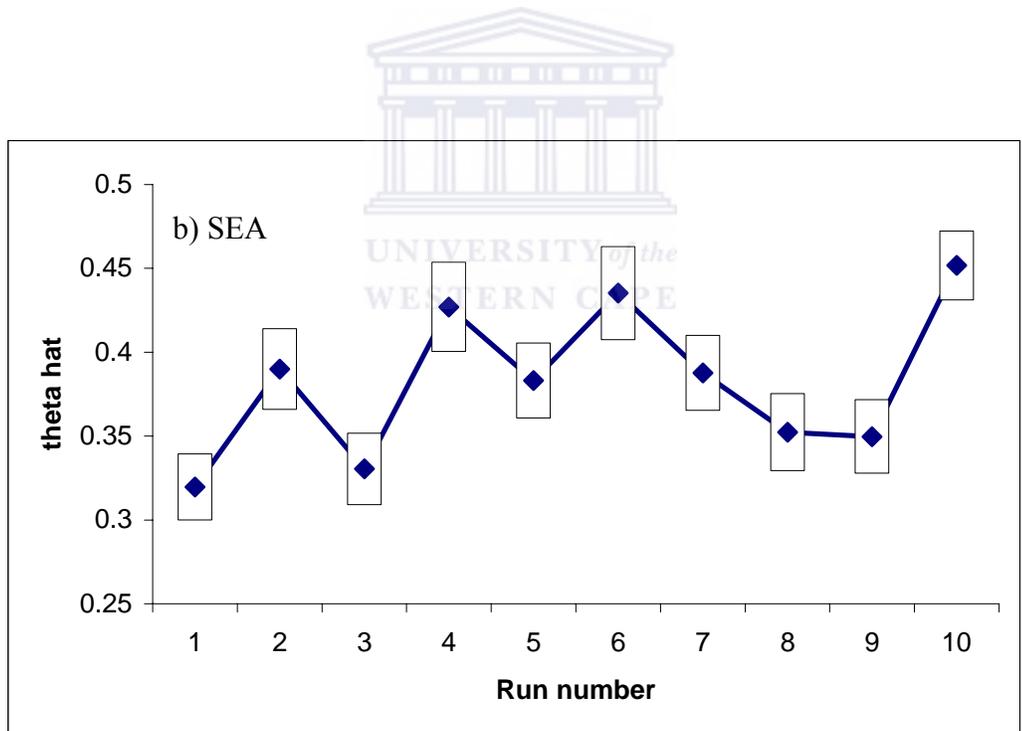
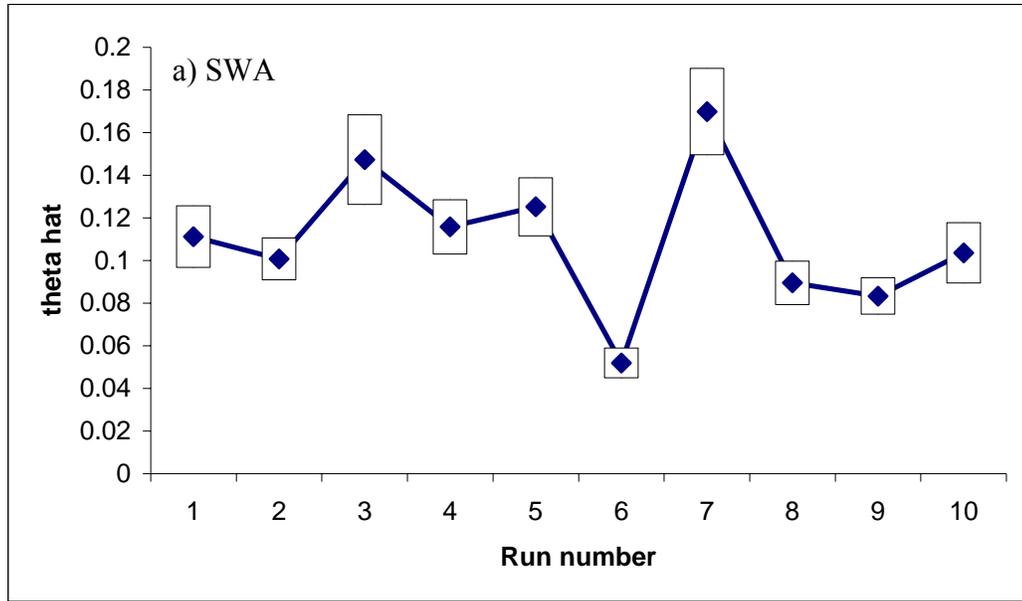


Maximum likelihood estimates of the growth parameter g under an exponential model (expressed as $1/u_s$) where u_s = the mutation rate per site per generation, were significantly positive and high for both clades. The unscaled exponential growth rate per generation was 6.65×10^{-6} and 7.77×10^{-6} for the SE and SW Atlantic populations respectively. Figure 3.10 plots the approximate 95% confidence intervals of θ and g and show how variable they ($\theta \pm 1.96^* \text{ s.d.}$). Growth was considered significant if g was larger than three standard deviations from zero, and as such, both south Atlantic populations of *E. lucens* show evidence of high and significant growth with the highest rates displayed in the SWA. However, it should be noted the standard deviations are approximations only (Kuhner *et al.* 1998).



g_0	Southeast Atlantic					Southwest Atlantic				
	θ	s.d. θ	g	s.d. g	cov (θ, g)	θ	s.d. θ	g	s.d. g	cov (θ, g)
5	0.320	0.0101	376.7	15.10	-2.22E-07	0.111	0.0074	546.9	39.44	-1.82E-04
10	0.390	0.0123	328.0	12.03	-1.70E-05	0.101	0.0050	193.6	19.26	-9.01E-08
20	0.330	0.0109	356.9	13.28	-1.41E-07	0.147	0.0107	506.9	32.99	1.49E-05
50	0.427	0.0136	370.0	13.65	-1.49E-06	0.116	0.0065	662.8	33.90	-7.32E-09
100	0.383	0.0114	374.6	12.35	6.64E-06	0.125	0.0069	506.5	32.50	-1.99E-09
-5	0.435	0.0142	374.4	14.13	-1.33E-10	0.052	0.0036	197.7	37.05	8.55E-05
-10	0.388	0.0114	432.6	14.04	-1.96E-07	0.170	0.0104	568.1	37.10	-2.01E-07
-20	0.352	0.0117	338.0	13.85	-1.27E-05	0.089	0.0052	302.7	23.14	2.36E-15
-50	0.350	0.0112	374.1	13.79	-2.07E-05	0.083	0.0044	364.7	29.25	-6.63E-10
-100	0.452	0.0105	372.1	14.12	-3.40E-07	0.104	0.0072	287.1	27.80	2.27E-07
mean	0.383	0.0141	369.7	13.63	-6.10E-06	0.110	0.0067	413.7	31.24	-8.13E-06
	0.443	0.0131	292.0	10.25	-6.73E-10	0.149	0.0090	478.6	32.62	-3.35E-06

Table 3.9 The results for *ND1* for the south Atlantic populations of *E. lucens* under a model of exponential growth and varying the initial growth parameter. Column one contains the initial growth values for each of the independent runs.



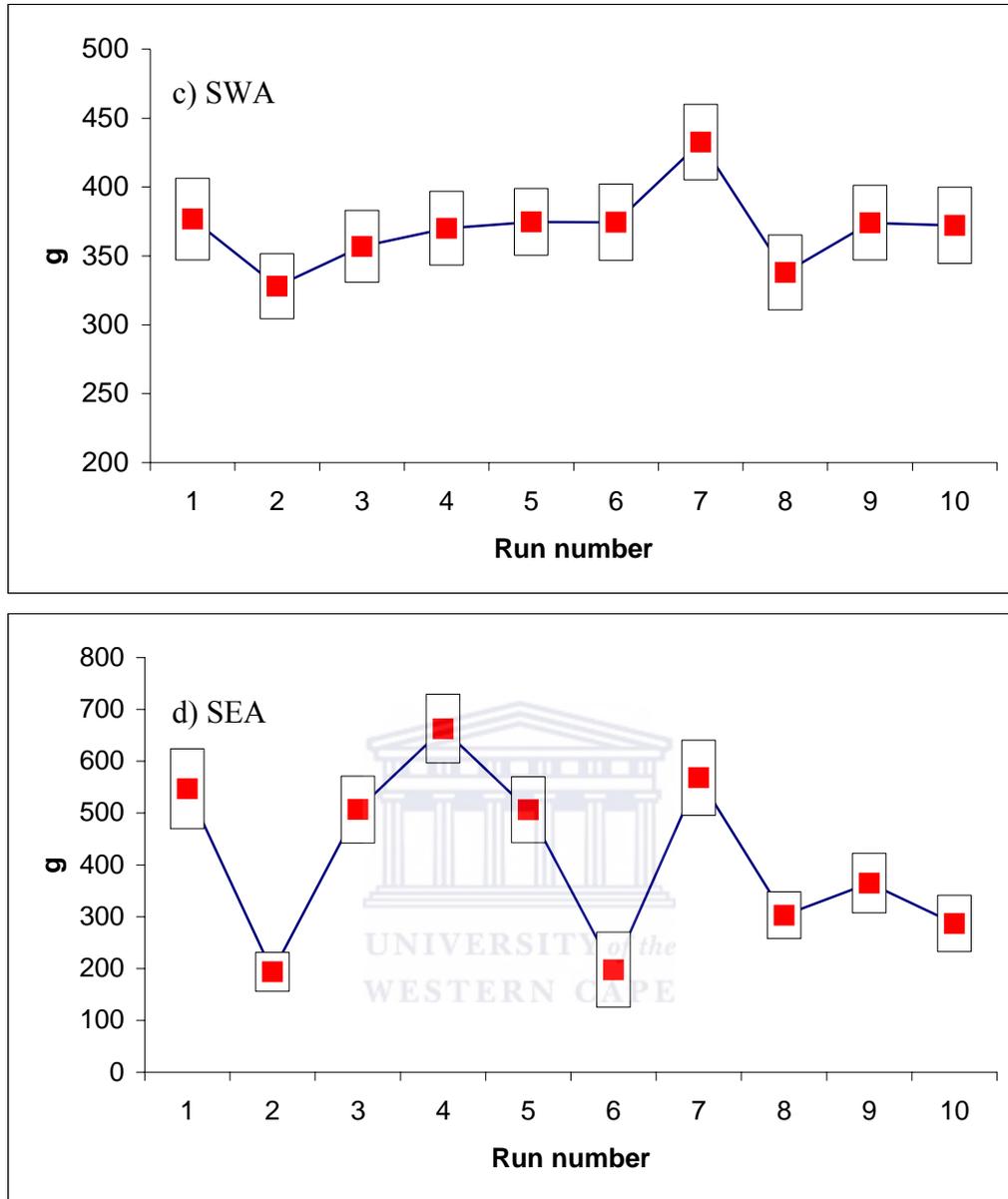


Figure 3.10a-d. The estimates of θ (a-b) and g (c-d) obtained with Fluctuate for *E. lucens* with the 95% confidence limits. a-c) SWA b-d) SEA. The initial g values for runs 1-10 correspond to those in Table 3.9.

3.9. Nuclear ITS-1 Results

The nuclear marker *ITS-1* was sequenced in a random selection of 10 individuals including specimens from each of the regional locations previously recovered as distinct clades in the phylogenetic analysis of the mtDNA data. Wide length variation in a variety of Malacostracans has previously been reported (Chu *et. al.* 2001; Murphy and Goggin 2000; Harris and Crandall 2000), and at 438 base pairs, the length of *ITS-1* in *Euphausia lucens* was longer than that of the mitten crabs *Eriocheir sinensis*, *E. leptognathus* and *E. Formosa* (Decapoda, Grapsidae), but shorter than that of the penaeid prawn *Penaeus japonicus* (Decapoda, Denauidae), and almost half the length reported for the freshwater crayfish species *Orconectes luteus* (Decapoda, Cambaridae) (Chu *et. al.* 2001; Harris and Crandall 2000). The GC content at 59.9% was similar to that observed in abalone species of the genus *Haliotis* (Mollusca, Gastropoda, Halrotridae) (Coleman and Vacquier 2002), higher than that observed in several freshwater crayfish species (Harris and Crandall 2000), and more than twice that found in the tiger beetle (Vogler and DeSalle 1994). All of *ITS-1* sequences including those from New Zealand, Argentina and South Africa were identical with no evidence of geographic subdivision apparent, a pattern somewhat at odds with the result obtained from the mtDNA data. The *E. lucens ITS-1* sequences were, however, too divergent from all other available crustacean amplicons to be reliably aligned.

CHAPTER 4

RESULTS *Euphausia vallentini*

4.1. SSCP Scoring and Sequencing

An example of the SSCP patterns produced by the experimental protocol described previously for the *E. vallentini* *NDI* sequences is shown in Figure 4.1.

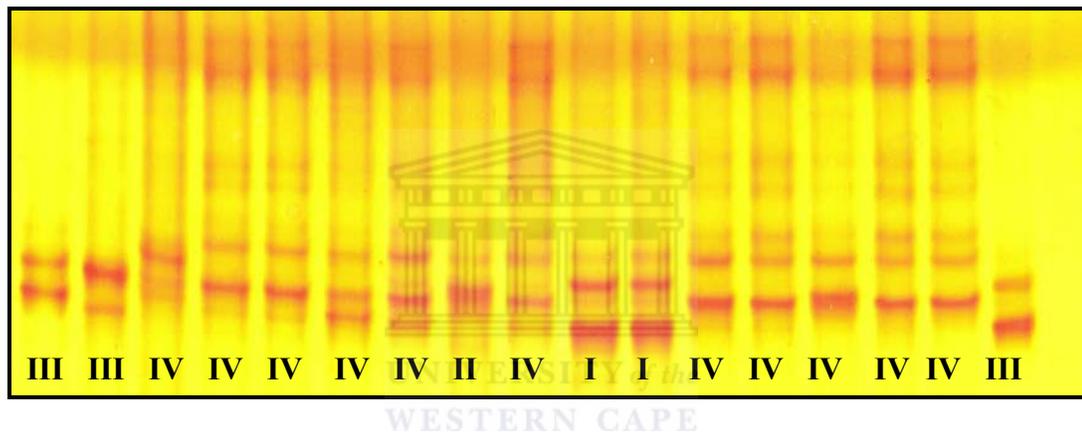


Figure 4.1. An example of the *NDI* PCR single strand conformational polymorphism profiles generated for the *E. vallentini* samples. The haplotype labels (I-IV) correspond to those in Table 4.1

For this species the two major banding systems in the SSCP gels were synergistically informative and both bands could be used to accurately score the different haplotypes. Both bands were strong and unambiguous, greatly facilitating the haplotype screening process. As was observed for *E. lucens* (Figure. 3.1b) however, additional, less intense upper bands were occasionally observed in *E. vallentini*. As before (page 81) the faint additional bands were

tentatively attributed to the presence of metastable conformers (Zehbe *et al.* Application note 384) and did not affect the ability of SSCP to effect accurate identification.

All haplotypes were sequenced more than once and sequencing the haplotypes for different individuals with the same SSCP pattern discovered no undetected mutations. Further, a random sample of specimens was sequenced and no unexpected haplotypes were recovered. Taken together, this shows that SSCP screening has again efficiently detected low levels of variation in this mtDNA fragment in *E. vallentini*.

4.2. Sequence Variation

Two hundred and fifty-two specimens of *E. vallentini* from Argentina, New Zealand, the south Indian Ocean, and sub-Antarctic waters in the SE Atlantic were examined (Table 1). Haplotype designations are shown in table 4.1.

	Polymorphic Sites		
	1	1	1
	1	3	3
Haplotypes	2	1	9
I	C	T	G
II	.	C	.
III	.	.	A
IV	T	.	.
Codon Position	3	3	3

Table. 4.1. Positions of the variable sites in *E. vallentini* *NDI* sequences

Four distinct haplotypes were identified, one of which was a singleton. Two haplotypes were recovered from certain localities; haplotype III, which was found only in the Argentinean Falkland Shelf sample, and haplotype II, recovered from a single sub-Antarctica sample. The remaining two haplotypes, IV, and I were found in all other samples where they differed only in their relative frequencies, with the latter haplotype, always the most frequent. Genetic variation in *NDI* was relatively low in this species (Table 4.2) and the combined data set contained only three polymorphic sites, all of which involved synonymous transitions. Consequently, the average number of nucleotide differences between sequences and nucleotide and haplotype diversities were low ($K= 0.486$ $\pi = 0.0031$ $h = 0.428$). This is expected as all of the haplotypes recovered differ by a single substitution with respect to haplotype I. The genetic variation observed in *E. vallentini* at this locus is relatively low compared to *E. lucens*, the putative sister species (Jarman *et al.* 2000a), and suggests that very different evolutionary forces may have been responsible for shaping the patterns of polymorphism in the two descendant species since their split from an ancestral taxon.

	<i>n</i>	<i>Haplotypes</i>	<i>S</i>	<i>h</i>	π (per site)	<i>D</i>	<i>F_s</i>
Sub Antarctica	125	3	2	0.335	0.0022	-0.088 (<i>P</i> > 0.10)	0.086 (<i>P</i> > 0.10)
Argentina	62	3	2	0.565	0.0045	1.0865 (<i>P</i> > 0.10)	1.278 (<i>P</i> > 0.10)
Marion Island	54	2	1	0.509	0.0033	0.5303 (<i>P</i> > 0.10)	2.026 (<i>P</i> > 0.10)
New Zealand	11	2	1	0.509	0.0033	n/a	n/a
Total	252	4	3	0.428	0.0031	-0.079 (<i>P</i> > 0.10)	-0.092 (<i>P</i> > 0.10)

Table 4.2. The *NDI* summary statistics of diversity for *E. vallentini* samples. None of the neutrality tests were significant with *p* > 0.10 obtained for all the tests. *S* = segregating sites, *h* = haplotype diversity, π = nucleotide diversity, *D* = Tajima's test, *F_s* = Fu's neutrality test

The regional distinctness of the sub-Antarctic and Argentinean samples was based solely on the presence of the haplotypes II and III (Table 4.1), and is most likely not an artefact given the relatively large sample sizes employed. Given the low levels of variation observed for this locus, and the relatively large sample sizes employed, many more samples would be required to determine, with certainty, whether haplotype II, which occurred as a singleton, is actually restricted to sub-Antarctic waters. A similar situation applies to the

intermediate frequency haplotype III, as only a single location was sampled over the Falkland Shelf, and several additional sampling sites in this area would have to be investigated before population subdivision could be reliably inferred.

4.3. Phylogenetic Analysis and Rooting the Network

A median-joining network (M-J) was constructed and, not surprisingly, revealed the presence of one distinct clade, containing the four haplotypes (Figure 4.2). Rooting of the intraspecific network could not be unambiguously achieved using the previously constructed ML, and ME majority consensus phylogenetic trees (Figures 3.4 and 3.5) due to the lack of resolving power between ingroup alternatives. If one were willing to assume that present distributions reflect those of the past, then, using the frequency criterion and topology criterion, the location of the MRCA in this species would be undetermined as the most frequent haplotype (IV), and the one with the highest number of connections to other haplotypes (I), occur in all of the samples. Using the total number of occurrences in different populations, this would be either the Falkland Shelf area around Argentina or the Atlantic sector of sub-Antarctica. It should be noted that although genetic variation in this species is low for *NDI*, the highest diversity was observed in the neritic samples. Again, further sampling over the southern Patagonian Continental Shelf region will be required to determine the generality of this observation.

MEDIAN-JOINING NETWORK

BANDELT (1999)

Euphausia vallentini

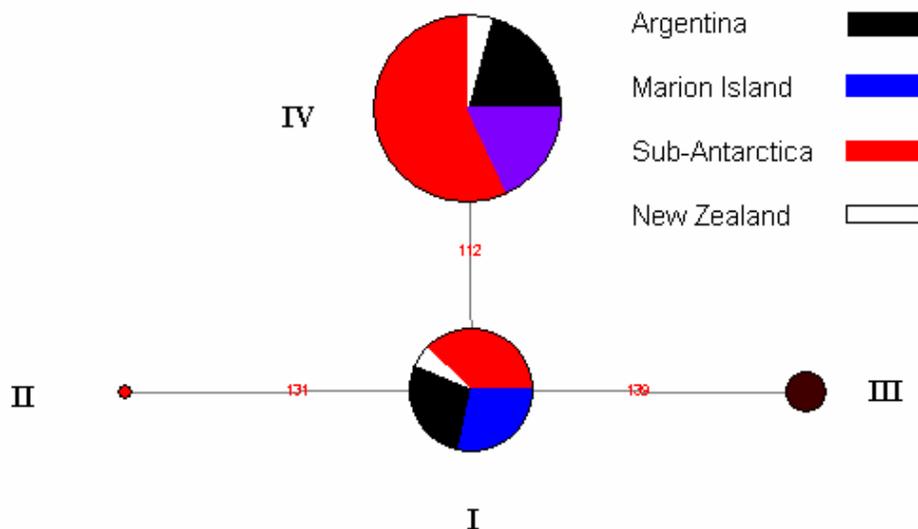


Figure 4.2 Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. vallentini* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences: black, Argentina, blue, Marion Island, red, Sub-Antarctica and white, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency and the lowest frequency was 1 (haplotype III). Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

4.4. Neutrality Tests

Neutrality tests conducted on the combined data set failed to detect any significant deviation from neutral expectations for mutation-drift equilibrium with all the tests, returning positive values with the exception of the sub-

Antarctica samples, where the value was minimally negative for Tajima's D (Table 4.2). As expected, all tests indicate that there is an excess of common alleles in all the samples, as evidenced by the positive values obtained for these statistics. Fu and Li's D test (1993) could not be calculated due to the absence of any internal or older mutations in this data set.

4.5 Estimates of Effective Population Size

Very similar estimates of θ for the total population were produced with the Watterson ($\theta = 1.59$) Tajima ($\theta = 1.5$) and Fu's (1994b) UPBLUE ($\theta = 2.095$) methods. For $\mu = 2.81 \times 10^{-6}$ the Watterson (1975) and Tajima (1983) and Fu's (1994b) estimates translate to effective population sizes (N_e) of 141,459 133,451 and 163,887 individuals respectively. Again these estimates are far lower than would be expected based on potential census sizes in this species however, the low levels of variation observed for this marker limits my ability to provide detailed inferences about the population mutation rate parameter and the results should be viewed as tentative.

4.6. Demographic Hypotheses

The absence of significant temporal or spatial genetic structuring enabled samples to be combined for increased statistical power. However, the limited genetic NDI variation in *E. vallentini* precludes the use of some of the more sophisticated genealogy-based population parameter estimators (Kuhner *et al.* 1998; Vasco *et al.* 2001) used previously in the analysis of the *E. lucens* data.

4.7. Interspecific Comparisons

A phylogenetic tree was constructed for the combined *E. lucens* and *E. vallentini* *NDI* haplotypes (Figures 3.4 and 3.5) and revealed far greater divergences than had been previously reported for these species with two other mitochondrial markers, cytochrome oxidase subunit 1 (*COI*) and the large subunit ribosomal RNA (*16S* rRNA) (Jarman *et al.* 2000a). The average percentage base pair distance between these species for *NDI* in this study was 19.27% with a net divergence of 17.79%, an order of magnitude greater than previously reported (1.25%) for *16S* and *COI* combined (Jarman *et al.* 2000a). The interspecific genetic distance reported by these authors, was lower than the average *NDI* intraspecific percentage base pair distances for *E. lucens* (2.47%) but higher than that observed within *E. vallentini* (0.49%). The relatively small difference between the estimates of the average and net pairwise distances reflects the pattern of low variation within species relative to the divergence between them. The net pairwise distance between the *E. lucens*-*E. vallentini* species pair was then used to obtain a rough estimate of divergence times under the assumption of a molecular clock. This is an appropriate method of estimating divergence time that corrects for the effects of ancestral polymorphism, which can be substantial especially in large subdivided populations (Nei and Li 1979; Edwards and Beerli 2000; Wakeley 2000). Applying the mutation rate previously calculated for this *NDI* fragment of 2.81×10^{-6} per sequence per generation (M and M) the split between these species was estimated to have occurred around 6.33 million years ago. The previous

estimate using *16S* rRNA and *COI* placed this event at less than 1 million years ago (Jarman *et al.* 2000a) albeit, using a different method of dating the divergences.

Although equivocal divergence dates are not necessarily expected when different molecular markers and dating methods are employed, the magnitude of the difference between the two divergence time estimates is too great to suggest methodological differences alone are likely the cause. Mitochondrial DNA can be thought of as a single non-recombining locus, and the relative differences between a pair of genes in a pair of species, although subject to the stochastic forces of random genetic drift and mutation, should be correlated unless forces such as selection are shaping the pattern of polymorphism at one, or both markers in either species. This expectation is in fact the null hypothesis in tests of neutrality such as the Hudson-Kreitman-Aguade (HKA) test of linkage disequilibrium (Hudson *et al.* 1987).

To investigate the apparent inconsistency between the results obtained here with the previous study of Jarman *et al.* (2000a), and to provide additional information regarding the estimates of the TMRCA of both these species, four *E. vallentini* individuals from Marion Island, and eight individuals of *E. lucens*, including representatives of the regional clades recovered in the M-J network for the *NDI* data were sequenced with the *COI* primers used in the Jarman *et al.* (2000a) study. Haplotype designation and their frequencies for this marker

are shown in Table 4.3.

4.7.1. Mitochondrial Variation in Cytochrome Oxidase I (COI)

The total 640 base pair *COI* data set for *E. lucens* had eighteen variable sites and defined five haplotypes. All substitutions were synonymous and the transition/transversion ratio for this locus was 17:1. Of the eighteen substitutions, nine were purine/purine transitions, eight were pyrimidine/pyrimidine transitions, and one was a purine/pyrimidine transversion. Genetic diversity was highest in SW Atlantic basin and twelve fixed differences separated this and the SE Atlantic population. The average nucleotide difference between these *E. lucens* populations was high ($K = 15.5$), while in the *E. vallentini* data set, the four sequences were identical. Both results are consistent with those obtained from the *NDI* locus for these species here.

The *COI* data for the *E. lucens* –*E. vallentini* species pair had a total of fifty-nine variable sites all involving synonymous substitutions. Forty fixed differences separated the two species and the average number of nucleotide differences between populations was forty-eight (Table 4.3). Also included in Table 4.3 is one *COI* sequence each of *E. lucens* (gi|5853129|gb|AF177185.1|) and *E. vallentini* (gi|5853132|gb|AF177188.1|) recovered from an earlier study (Jarman *et al.* 2000a) and previously deposited in GenBank.

	POLYMORPHIC SITE POSITION																																							
	1 1 1 1 1 1 2 3																																							
	1 1 1 3 4 4 4 6 7 9 9 1 4 4 6 7 9 0 0 0 0 1 1 2 2 4 4 5 5 6 7 9 0																																							
	2 3 6 9 0 5 9 6 0 5 6 3 8 0 6 4 4 7 5 4 5 1 2 4 7 5 8 1 8 0 6 2 3 8 3 1 0																																							
ELSE1	T	C	C	T	G	A	G	A	T	G	T	A	G	A	C	A	T	T	A	A	C	G	T	G	G	A	T	C	C	T	G	A	C	C	T	C	G	A		
ELSE2	G
ELSE3
ELSE4
ELSW1	.	T	.	.	G	G	.	.	.	A	.	.	T	.	.	.	T	.	T	.	T	.	C		
ELSW2	.	T	.	.	G	G	.	.	.	A	T	A	A	T	.	.	.	T	.	T	.	T	.	C	
ELSW3	.	T	.	.	G	G	.	.	.	A	.	.	T	.	.	.	T	.	T	.	T	.	C		
ELSW4	.	T	.	.	G	G	.	.	.	A	.	.	T	.	.	.	T	.	T	.	T	.	C		
EVAL	.	.	.	C	.	G	.	G	C	A	C	G	A	G	T	G	C	C	G	.	T	A	C	A	.	.	.	T	C	A	G	T	.	C	A	A	C	A		
E. LUC*	.	.	.	C	.	G	T	G	C	A	C	G	A	G	T	G	C	C	G	.	T	A	C	A	.	.	.	T	C	A	G	T	.	C	A	A	C	A		
Eval*	A	A	G	G	A	G	.	G	.	A	C	G	A	G	T	G	C	C	G	.	T	A	C	A	.	.	.	T	C	A	G	T	.	C	A	A	C	A		
Codon	2	3	3	3	1	3	1	3	1	3	1	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	1	3	3	3	1	1	3	3	3			
Position																																								

Table. 4.3 The variable sites for *E. vallentini* and *E. lucens* for the *CO1* sequences amplified with the primers HCO and COI. Argentina EVAL = *E. vallentini* from Marion Island in the Prince Edward Island group. E. LUC* = sequence AF177188.1. The grey filled boxes mark the nonsynonymous changes

4.7.2. COI Phylogenetics

Phylogenetic analysis was conducted on the *E. lucens*–*E. vallentini* species pair data set that included all of the sequences in Table 4.3. with both maximum parsimony (MP) and maximum likelihood (ML) methods. Molecular evolutionary relationships were consistent across the two linked gene loci and tree construction methods employed (Figure 4.3a-b). It can be seen that the sequence previously identified as *E. lucens* (Jarman *et al.* 2000a) (gi|5853129|gb|AF177185.1) groups in the *E. vallentini* clade with high bootstrap support and not with the *E. lucens* sequences generated in this investigation thus, explaining the low divergences between these species reported in that study (Jarman *et al.* 2000a). Two amino acid changes separate the two species (Table 4.3), the first occurred at a 1st codon position and involved the substitution of alanine with serine $G_{19} \rightarrow T_{19}$ the other, occurred at a 2nd codon position and involved a substitution of tryptophan with phenalanine $A_{635} \rightarrow T_{635}$ (Table 4.3). Neither of these was fixed between species and indeed, both occurred in the misidentified *E. lucens* sequence previously deposited in GenBank (gi|5853129|gb|AF177185.1|).

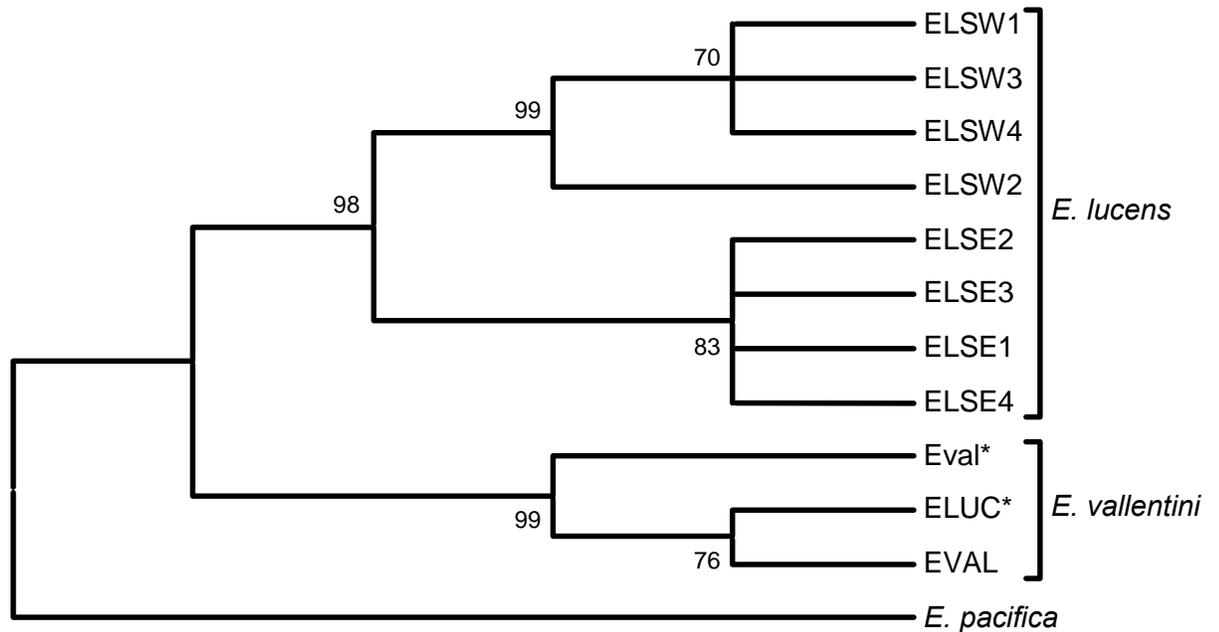


Figure 4.3. The *COI* maximum parsimony tree for *E. lucens* and *E. vallentini*. This was one of the 32 most parsimonious trees found of length 166 steps. The numbers at the nodes are the bootstrapped values calculated with 4000 replicates. An identical tree was obtained with the maximum likelihood method (not shown). E.val* = gi|5853132|gb|AF177188.1| and E.LUC* = |5853129|gb|AF177185.1 ELSW = SW Atlantic *E. lucens* ELSE = SE Atlantic *E. lucens*.

Five additional non-synonymous changes are observed between these species with the inclusion of the *E. vallentini* sequence recovered from GenBank gi|5853132|gb|AF177188.1| (Jarman *et al.* 2000a) (Table 4.3) however, these all occur within a section at the 3' end of this fragment nine base pairs in length, and probably represent uncertainty in scoring the nucleotide composition for these bases in the original chromatogram. Excluding this sequence from the *E. vallentini* data set results in a significant decrease in all of

the measures of *COI* genetic variation within *E. vallentini* and diversity between the *E. lucens* –*E. vallentini* species pair (Table 4.4). Given the low levels of divergence in *E. vallentini* observed for both the *ND1* and *COI* markers (this study) substitutions in sequence gi|5853132|gb|AF177188.1| (Jarman *et al.* 2000a), should be viewed with suspicion until such time that the original chromatograms can be obtained and checked for accuracy. To avoid uncertainty in my conclusions this sequence was excluded from any further analysis.

gi 5853132	<i>S</i>	<i>h</i>	π	θ
Included	8	0.7	0.0050	0.0060
Excluded	2	0.5	0.0016	0.0017

Table 4.4. Table showing the change in summary statistic measures of *COI* variation within *E. vallentini* when the mtDNA sequence gi 5853132 (Jarman *et al.* 2000a) is excluded. *S* = segregating sites, *h* = haplotype diversity, π = nucleotide diversity, θ = Wattersons theta.

Consequently, the actual mean pairwise distance for the *E. lucens* –*E. vallentini* species pair for this *COI* fragment was 7.6%. However, the lack of variation in *E. vallentini* for this gene fragment precluded calculation of the net pairwise distances. This level of divergence is more consistent with that commonly observed for congeneric plankton species (Bucklin *et al.* 1997; Zane and

Patarnello 2000; Goetze 2003; Peijnenburg *et al.* 2005) and with *NDI* results previously presented. It can therefore be reliably concluded, that the extremely small divergence estimates previously reported for this group (Jarman *et al.* 2000a) were based on a misidentification of an individual specimen of *E. vallengini* as *E. lucens* and when correct identification is achieved the two species actually form reciprocally monophyletic clades using the three tree reconstruction methods deployed in this study, with very high bootstrap support (Figure 4.3).

4.7.3. Dating divergence using 16S rRNA sequences

All three previously published attempts to date the divergence of southern hemisphere euphausiid species employed sequence information from the 16S rRNA mitochondrial gene using information from both transitions and transversions (Jarman *et al.* 2000a), or transversions only (Zane *et al.* 1998; Zane and Patarnello 2000). Therefore, for comparative purposes, this gene fragment was also obtained for all of the euphausiid species investigated here, and used to date the splits between species under the assumption of a molecular clock with the same methods applied in the previous studies by Zane *et al.* (1998 and; Zane and Patarnello (2000).

4.7.4. Relative rates tests among species

I assessed the homogeneity of the mutation rate between krill taxa using the comparative method software package LinTree of Takezaki *et al.* (1995)

implemented in the that assumes a molecular clock and requires known calibration points in a neighbour-joining (NJ) tree. Two relative rates tests were applied to the NJ tree; the two-cluster test (TCT) and the branch-length test (BLT) to identify nodes and branches that do not conform to the molecular clock. The null hypothesis in the TCT is equal branch lengths in each of the two clusters from a given node (branch point) in a tree, excluding the outgroup. The null hypothesis for the BLT test is equal root to tip distances among sequences and the average root to tip distance. A confidence probability higher than 95% indicates that the molecular clock should be rejected for a pair of branches. A chi-squared test with “Q” with $n-1$ (n = number of sequences under the root) degrees of freedom was conducted to test rate constancy for all sequences under the root. The NJ tree was constructed with LinTree with 1000 bootstraps to obtain a balanced tree with two clusters rooted with the outgroup sequence of *Squilla empusa* (Af107617) (*Holocarida*, *Stomatopoda*). In order to verify whether the *16S* rRNA molecular clock calibrated for crabs (Cunningham *et al.* 1992) can be applied to krill, I compared the substitution rate between krill and six non-Jamaican grapsid crab taxa from Shubart *et al.* (1998), including congeners of *Sesarma* from the Atlantic and Pacific Oceans. The calibration points used on the krill and *Grapsidae* trees was the transisthmian vicariance set at 2.5- 3.1 million years ago. Neither the two-cluster ($Q = 8.4$, $df = 11$ $p > 0.05$), or branch-length test ($Q = 17.53$, $df = 11$, $p > 0.05$) detected significant deviation from rate constancy in any of the

branches thus allowing the use of *16S* rRNA corrected genetic distances to date the nodes on the tree.

This N-J tree is presented in Figure 4.4 and although not shown for the sake of brevity, the *E. lucens* *16S* rRNA sequence (gi|5853129|gb|AF177185.1) used in the previous study (Jarman *et al.* 2000a), grouped within the *E. vallengini* clade (that comprised a single individual collected by the author *E. vallengini* 1, and another *E. vallengini* 2, obtained in a previous study by Zane and Patarnello 1998), with high bootstrap support, and distinct from the *E. lucens* sequence (*E. lucens*) also collected by the author. This result is expected given that both the *COI* and *16S* rRNA sequences produced in the study by Jarman *et al.* (2000a) were obtained from the same individual krill.

Applying the same molecular clock previously used for krill *16S* rRNA rate for transversions of 0.155-0.177% per million years (Cunningham *et al.* 1992; Zane and Paternello 2000), and considering the corrected genetic divergences between *E. lucens*-*E. vallengini*, the split between these two species from the common ancestor is of the order 2.91-3.19 MYA. A previous estimate of the divergence time for this species pair calculated using a different method to that applied above (Jarman *et al.* 2000a), gave a time of 1.04 MYA.

In turn, the split between the Antarctic species *E. superba* and the clade containing the sub-Antarctic species *E. lucens* and *E. vallengini* is of the order

of 10.98-12.04 MYA. It can also be seen that *E. recurva* groups with *E. krohni* on the linearised tree, and using the same rate as applied above, produces an identical divergence date to the *E. lucens*-*E. vallentini* split of the order 2.91-3.19 MYA. The most basal node on the euphausiid side of the tree that separates the *E. recurva*-*E. krohni* clade from the one containing *E. superba*, *E. lucens* and *E. vallentini*, is dated around 14.98-16.43 MYA.

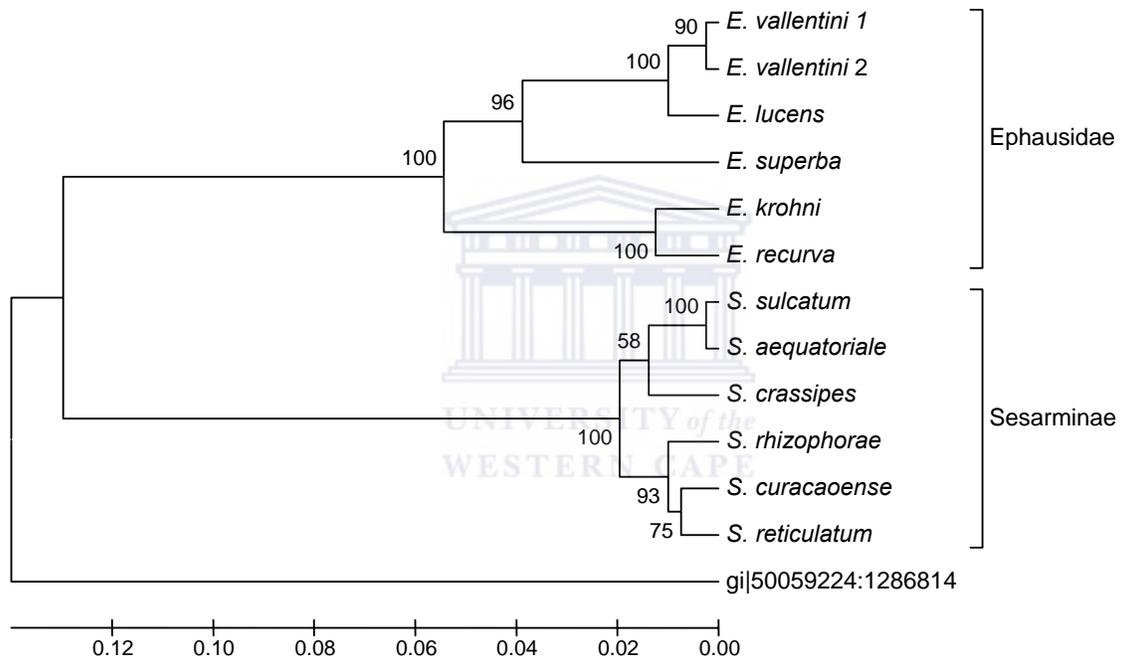


Figure 4.4. Neighbour-joining tree for tranversions only inferred for the mitochondrial large subunit (16S) rRNA sequences. *Squilla spp* was used to root the tree. Numbers at the nodes represent bootstrapped support calculated with 1000 replicates. The *reticulatum* and *sulcatum* groups contain species used in calibration of the molecular clock.

To obtain an unbiased result, with the relative rates tests (BLT and TCT) applied earlier, equivalent numbers of species of grapsid crabs and euphausiids in each cluster are required. To achieve this, I arbitrarily omitted two other *Euphausia* species, *E. frigida*, and *E. crystallophorias*, which were included in the previous study by Zane and Patarnello (2000). When these species are included here, the split between *E. vallentini* and *E. frigida* is of the order of 3.66-3.95 MYA, and that between *E. superba* and *E. crystallophorias* around 1.125-1.457 MYA. Rate constancy among *16S* rRNA sequences in the tree with these additions was not tested however, and must be assumed.

These new *16S* rRNA divergence dates are exactly half those previously reported for these lineages (Zane and Patarnello 2000) due to a simple methodological oversight perpetuated in that study. Specifically, these authors omitted to divide the observed corrected pairwise genetic distances between species (obtained from the distance matrix) by a factor of *two* to account for the number of separate branches leading to the descendent lineages. Thus, producing estimates of divergence time, twice that of the actual value.

Furthermore, when a correctly identified specimen of *E. lucens* is included in the analysis, the smallest divergence among the southern hemisphere species is between the Antarctic species *E. superba* and *E. crystallophorias* and not between *E. lucens* and *E. vallentini* as previously thought (Jarman *et al.* 2000a).

4.7.5. Coalescent Estimate of Divergence Time

Recently developed methods for fitting the full isolation model with and without migration were applied to the *E. lucens*-*E. vallentini* NDI data using the program MDIV (Nielsen and Wakeley 2001). Unlike the single-lineage methods used previously, this maximum likelihood based method compares pairs of lineages to estimate population divergence times while taking into account the critical effects of genetic polymorphism in the ancestral species or population, which may comprise over 50% of the total divergence (Edwards and Beerli 2000) even in recently separated species.

The coalescence based population parameters, including the maximum likelihood estimators of divergence times (T), number of migrants (M), and ancestral population size (θ_{ML}) were estimated with MDIV and the results listed in Table 4.5. Repeated runs of the MDIV program with 40 million steps and a burn-in of 10% this value, revealed unambiguous marginal posterior probability distributions for the *E. lucens*-*E. vallentini* data regardless of the initial values used and the model parameters chosen. This is good evidence that convergence to the ergodic averages was achieved (Nielsen and Wakeley 2001). To convert coalescence-based parameters to estimates in millions of years, the mutation rate estimated previously (page 96 for details) where $\mu = 2.81 \times 10^{-6}$ mutations per sequence per generation was applied. Using this estimate, the effective population size (N_e) was calculated from the estimated θ_{ML} using the formula $N_e = \theta_{ML} / 2\mu$ and then used to convert the T values

scaled by N_e , in number of generations. The peaks of the primary two parameters were confined to fairly narrow ranges with corresponding credibility intervals (Figure 4.5).

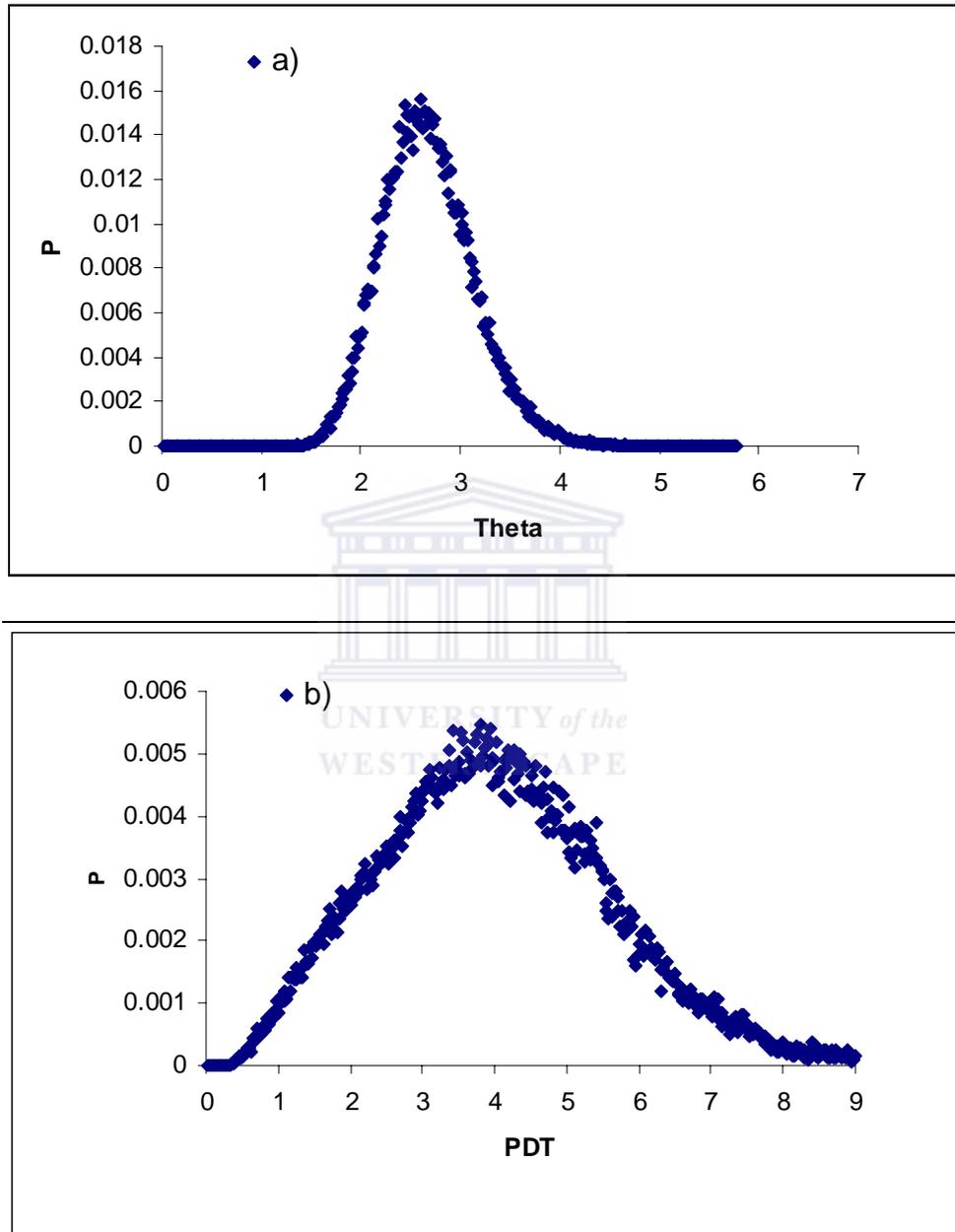


Figure 4.5a-b. The marginal posterior probability distributions for model parameters a) population size (θ) b) the population divergence time (T) (scaled by the neutral mutation rate). Curves are shown for the analysis of *E. lucens*-*E. vallentini* *ND1* data.

Comparison	Theta	Nm	PDT	TMRCA	PDT (MYA)	TMRCA (MYA)
<i>E. lucens</i> - <i>E. vallentini</i>	2.205	0.004	3.94	5.42	3.16	4.36

Table 4.5. Results of the coalescent analysis of the species *E. lucens* and *E. vallentini* populations based on NADH1 (ND1) sequence data. Theta, the migration parameter (Nm) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see page 96 for details).

Applying the previously described mutation rate to estimates of PDT the estimated time of separation between *E. lucens* and *E. vallentini* was 3.16 MYA, indicating middle Pliocene divergence, with the most common ancestor suggested to have occurred in the early Pliocene (4.36 MYA). There is a relatively large difference between the PDT and TMRCA estimates (28%) that may be expected when population divergence is relatively recent (Arbogast *et al.* 2002). This coalescence-based estimate of the divergence time between *E. lucens* and *E. vallentini* was very similar to the one based on 16S rRNA obtained previously.

The posterior distribution of m_2 for *ND1* revealed a very narrow peak at the lower limit of the resolution and the probability that the migration rate for this

locus is zero, is zero (not shown). Consequently the estimate of the effective migration rate between species ($2MN_e$) is very low (Table 4.5) implying that these species have been isolated without gene flow since they began to diverge. It should be noted however, that MDIV (Nielsen and Wakeley 2001) assumes that both the ancestral and descendent populations are of similar size, which may not be reasonable given the very different levels of mtDNA variation observed in each species.

4.8. Nuclear Sequence Variation

Twelve *E. vallentini* individuals, including representatives from all the regional samples were sequenced for *ITS-1*, and, as was observed in *E. lucens*, the sequences were observed to be monomorphic. When these *ITS-1* sequences were aligned with those of *E. lucens* (Figure 4.6) it was revealed that four indels of two of three and one of four and five base pairs in length and two fixed substitutions separate the two species. The indels are fixed in *E. vallentini* relative to *E. lucens* and the blunt end of these indels suggests that each one was a result of a single mutation event. Consequently, the *E. vallentini ITS-1* sequences (at 426 base pairs in length) were 12 base pairs shorter than those of *E. lucens*. The GC content was very similar at 63% in the former and 60% in the latter species. Furthermore, no substitutions were observed, in either species in the highly conserved *18S* and *5.8S* rRNA flanking regions (Figure 4.6.).

	-95	-85	-75	-65	-55	-45	-35	-25	-15	-5	
<i>E. lucens</i>	GGCGGCCTTG	GTCGCCTCGA	GCTGCCGGAA	ACATGTCCAA	ACTTGATCAT	TTAGAGGAAG	TAAAAGTCGT	AACAAGGTTT	CCGTAGGTGA	ACCTGCCGAA	
<i>E. lucens</i>	GGCGGCCTTG	GTCGCCTCGA	GCTGCCGGAA	ACATGTCCAA	ACTTGATCAT	TTAGAGGAAG	TAAAAGTCGT	AACAAGGTTT	CCGTAGGTGA	ACCTGCCGAA	
<i>E. vallentini</i>	GGCGGCCTTG	GTCGCCTCGA	GCTGCCGGAA	ACATGTCCAA	ACTTGATCAT	TTAGAGGAAG	TAAAAGTCGT	AACAAGGTTT	CCGTAGGTGA	ACCTGCCGAA	
<i>E. vallentini</i>	GGCGGCCTTG	GTCGCCTCGA	GCTGCCGGAA	ACATGTCCAA	ACTTGATCAT	TTAGAGGAAG	TAAAAGTCGT	AACAAGGTTT	CCGTAGGTGA	ACCTGCCGAA	
	0	5	15	25	35	45	55	65	75	85	
<i>E. lucens</i>	GGATCATTA	CGTGACCACG	ACCACTTGCT	GGCGTGAACA	CAAAAACCTCC	TTCTTGGGTG	CCGTGCTCGG	TTCGGAATGT	CCGACTGGTT	<u>GTTGGCTTGG</u>	
<i>E. lucens</i>	GGATCATTA	CGTGACCACG	ACCACTTGCT	GGCGTGAACA	CAAAAACCTCC	TTCTTGGGTG	CCGTGCTCGG	TTCGGAATGT	CCGACTGGTT	<u>GTTGGCTTGG</u>	
<i>E. vallentini</i>	GGATCATTA	CGTGACCACG	ACCACTTGCT	GGCGTGAACA	CAAAAACCTCC	TTCTTGGGTG	CCGTGCTCGG	TTCGGAATGT	CCGAC--GTT	G---GCTTGG	
<i>E. vallentini</i>	GGATCATTA	CGTGACCACG	ACCACTTGCT	GGCGTGAACA	CAAAAACCTCC	TTCTTGGGTG	CCGTGCTCGG	TTCGGAATGT	CCGAC--GTT	G---GCTTGG	
	95	105	115	125	135	145	155	165	175	185	
<i>E. lucens</i>	AACCGCAAGG	TCCCTAGCCC	ATCCTT GTGCC	ACTGCCCTCC	GATGCCGGCC	GGCCCCAATC	ACTTGGCGTA	CCCTCCGAAA	GAACCCCGGC	AACGGGGTAC	
<i>E. lucens</i>	AACCGCAAGG	TCCCTAGCCC	ATCCTT GTGCC	ACTGCCCTCC	GATGCCGGCC	GGCCCCAATC	ACTTGGCGTA	CCCTCCGAAA	GAACCCCGGC	AACGGGGTAC	
<i>E. vallentini</i>	AACCGCAAGG	TCCCTAGCCC	---- GT GTGCC	ACTGCCCTCC	GATGCCGGCC	GGCCCCAATC	ACTTGGCGTA	CCCTCCGAAA	GAACCCCGGC	AACGGGGTAC	
<i>E. vallentini</i>	AACCGCAAGG	TCCCTAGCCC	---- GT GTGCC	ACTGCCCTCC	GATGCCGGCC	GGCCCCAATC	ACTTGGCGTA	CCCTCCGAAA	GAACCCCGGC	AACGGGGTAC	
	195	205	215	225	235	245	255	265	275	285	
<i>E. lucens</i>	AAGAGGAGGG	AGACCGCGGC	CCGGTCTAAG	ACCAAAGCTC	CTAACGGAGC	AAAGCGTCTA	ACGGGTCCAC	GGGTCGCCGA	CCGAGCCTAA	CA ACCCCTCA	
<i>E. lucens</i>	AAGAGGAGGG	AGACCGCGGC	CCGGTCTAAG	ACCAAAGCTC	CTAACGGAGC	AAAGCGTCTA	ACGGGTCCAC	GGGTCGCCGA	CCGAGCCTAA	CA ACCCCTCA	
<i>E. vallentini</i>	AAGAGGAGGG	AGACCGCGGC	CCGGTCTAAG	ACCAAAGCTC	CTAACGGAGC	AAAGCGTCTA	ACGGGTCCAC	GGGTCGCCGA	CCGAGCCTAA	---CCCCTCA	
<i>E. vallentini</i>	AAGAGGAGGG	AGACCGCGGC	CCGGTCTAAG	ACCAAAGCTC	CTAACGGAGC	AAAGCGTCTA	ACGGGTCCAC	GGGTCGCCGA	CCGAGCCTAA	---CCCCTCA	
	295	305	315	325	335	345	355	365	375	385	
<i>E. lucens</i>	CCGGGTGGG	CTCAACCCTG	GCGACGAAAA	<u>GTCCTCTCCG</u>	TGGGTGCCGC	CCGCCTCCCC	CAGATTGGCG	GGCTTTCTGT	CGGGTTTATC	<u>AGTCTCAAAC</u>	
<i>E. lucens</i>	CCGGGTGGG	CTCAACCCTG	GCGACGAAAA	<u>GTCCTCTCCG</u>	TGGGTGCCGC	CCGCCTCCCC	CAGATTGGCG	GGCTTTCTGT	CGGGTTTATC	<u>AGTCTCAAAC</u>	
<i>E. vallentini</i>	CCGGGTGGG	CTCAACCCTG	GCGACGAAAA	<u>GTCCTCTCCG</u>	TGGGTGCCGC	CCGCCTCCCC	CAGATTGGCG	GGCTTTCTGT	CGGGTTTATC	<u>AGTCTCAAAC</u>	
<i>E. vallentini</i>	CCGGGTGGG	CTCAACCCTG	GCGACGAAAA	<u>GTCCTCTCCG</u>	TGGGTGCCGC	CCGCCTCCCC	CAGATTGGCG	GGCTTTCTGT	CGGGTTTATC	<u>AGTCTCAAAC</u>	
	395	405	415	425	435						
<i>E. lucens</i>	<u>CCCTTGATTG</u>	<u>TGTCTCCCCT</u>	<u>AGTGGAGTCT</u>	<u>AAACAACCAA</u>	<u>AAATACAAC</u>	<u>CT</u>					
<i>E. lucens</i>	<u>CCCTTGATTG</u>	<u>TGTCTCCCCT</u>	<u>AGTGGAGTCT</u>	<u>AAACAACCAA</u>	<u>AAATACAAC</u>	<u>CT</u>					
<i>E. vallentini</i>	<u>CCCTTGATTG</u>	<u>TGTCTCCCCT</u>	<u>AGTGGAGTCT</u>	<u>AAACAACCAA</u>	<u>AAATACAAC</u>	<u>CT</u>					
<i>E. vallentini</i>	<u>CCCTTGATTG</u>	<u>TGTCTCCCCT</u>	<u>AGTGGAGTCT</u>	<u>AAACAACCAA</u>	<u>AAATACAAC</u>	<u>CT</u>					

Figure 4.6. Consensus sequences of partial 18SrRNA (-95 to 0), ITS-1 (starting from 1(and partial 5.8S rRNA underlined) from 10 individuals of each of *E. lucens* and *E. vallentini*.

CHAPTER 5

5.1. Discussion *E. lucens* and *E. vallentini*

This study is one of the few that has examined Southern Hemisphere zooplankton, and that has compared the same genes for three systematically and ecologically related species, which differ mainly in distributional range. Several questions have been addressed at both the population and species level, thereby providing insights into the factors influencing the evolution of genetic diversity in these species. Including two closely related species and a more distant one, has allowed me to capture some of the temporal and spatial dynamics with respect to genetic material in species formation.

5.2. Systematic Comparisons

It has been conclusively demonstrated that both *E. lucens* and *E. vallentini* represent valid taxonomic species: fixed differences between species were consistently observed in both the nuclear and mitochondrial genomes to indicate a total absence of gene flow between the different lineages. The observed levels of genetic differentiation are well within the range of those previously reported both for other euphausiid species (Zane and Patarnello 2000, Jarman *et al.* 2000a) and for a wide range of pelagic organisms such as copepods (Bucklin and Weibe 1998; Bucklin *et al.* 1995; Goetze 2003), chaetognaths (Peijnenburg *et al.* 2005); and fish (Arnason *et al.* 2000) as well

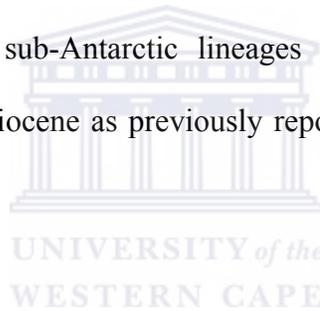
as benthic marine organisms such as the snapping shrimp *Alpheaus* (Knowlton *et al.* 1993) and penaeid prawns (Benzie 2000).

5.3. Species Divergence Time Estimates

I have also shown that a previous attempt to quantify the degree of divergence for *E. lucens* and *E. vallentini* using *16S* rRNA (Jarman *et al.* 2000a) is an underestimate, caused by the misidentification of a single specimen of *E. vallentini* as *E. lucens*. This error should be corrected by removing or renaming the relevant sequences currently held in GenBank records. In that study the low genetic distances observed between this species pair, combined with a lack of evidence for a candidate isolating mechanism around the time of the split, led these authors to invoke a sympatric speciation scenario to explain the result. The corrected *16S* rRNA dates presented here imply a separation around 2.91-3.19 MYA, placing this event in the mid- to late Pliocene. To the author's knowledge, the new dates still cannot be explained by reference to any known hydrographic barrier to dispersal existing at that time; so sympatric speciation remains a plausible scenario to explain the species divergence time estimates obtained with *16S* rRNA (as also Jarman *et al.* 2000a).

Although not part of the original aims of this study, I have also shown that previous attempts to date the divergence between *E. vallentini* and several other euphausiid species, based on the *16S* rRNA distances (Zane and Patarnello 2000), suffer from a large overestimation due to a calculation error. The

corrected estimates presented here cast doubt on the importance of the formation of the Antarctic Convergence as a barrier to population exchange between Antarctic and sub-Antarctic euphausiid species lineages; questioning the appropriateness of applying dispersal-limiting models of speciation in these groups (Zane and Patarnello 2000; Jarman *et al.* 2000a; Norris 2000). The Antarctic Circumpolar Current was installed in the early Miocene (~22 mya) as a consequence of the opening of the Drake Passage between the Antarctic Peninsula and Tierra del Fuego in southernmost South America (Barker and Burrell 1997; Rabassa *et al.* 2005; c.f. Barret 2001; Lawver and Gahagan 1998). These new results also indicate that many of the euphausiid speciation events within the Antarctic and sub-Antarctic lineages occurred in the Pliocene (Figure 4.4) and not the Miocene as previously reported (Zane and Patarnello 2000).



It should be recognised that single-locus estimates such as those based on the calibration for *16S* rRNA (Cunningham *et al.* 1992) fail to incorporate the intrinsic coalescent and mutational variance in the inferences of divergence times (Edwards and Beerli 2000; Arbogast *et al.* 2002; Hickerson *et al.* 2006). This type of variation can be substantial and is most apparent in single-locus data sets (Hickerson *et al.* 2006), particularly when populations may be subdivided into multiple demes (Wakeley 2000). The divergence times thus estimated will be upwardly biased, and should be considered as approximations

of the earliest possible time of separation (Edwards and Beerli 2000; Arbogast *et al.* 2002).

Gene-tree species-tree discordances can arise due to incomplete lineage sorting. Even under complete genealogical concordance there can be discordance between times of gene and organismal divergence (Knowles 2004). This divergence arises because prior to species divergence a degree of gene divergence has already occurred as accrued polymorphism in the ancestral population/species. It has an expectation of $2N$ generations (Wright 1951). For recently diverged species pairs the ancestral gene divergence can comprise a substantial proportion of the total divergence observed between species; a fraction that will increase if the ancestral population was further subdivided. This is not thought to be a problem for ancient divergences but, depending on the size of the ancestral population, can have a significant effect for the first several millions of years! (Arbogast *et al.* 2002).

The estimates are nevertheless useful as they allow a direct comparison to be made with the results of previously published studies that have adopted this approach to dating divergence times (Zane and Patarnello 1996, 2000; Jarman *et al.* 2000a), allowing methodological errors to be identified, while permitting a comparison with the estimates produced using some of the more sophisticated coalescent-based estimators under neutral (Nielsen and Wakeley 2001) and non-equilibrium models (Kuhner *et al.* 1998; Vasco *et al.* 2001).

The coalescence-based divergence time estimate (calculated with MDIV) for *NDI* does not depend on any single genealogy (Nielsen and Wakeley 2001), and the gene trees function as theoretical tools for deriving parameters of interest rather than as the basis for inference (Hudson 1990; Rosenberg and Nordborg 2002; Knowles 2004). By integrating over all possible genealogies, this method should be less biased than those that rely exclusively on the structure of a single gene tree. By using a common mutation rate for a particular molecular marker, relative comparisons can be made between the different methods.

The MDIV results for *NDI* indicated that the most recent common ancestor of *E. lucens* and *E. vallentini* occurred in the early Pliocene around 4.36 MYA and the descendent species lineages split around 1.2 million years later. This is very similar to the single-tree estimate based on corrected *16S* rRNA distances (Figure 4.4). There is a small difference between population divergence time (PDT) and time to the most recent common ancestor (TMRCA) consistent with a relatively ancient separation (Arbogast *et al.* 2002). Thus the separation of these species was likely contemporaneous with other sister-taxa lineages of the Antarctic and sub-Antarctic groups (page 136). A weakness of this method is that it assumes ancestral and descendent populations sizes are the same; this may not be reasonable.

5.4. Characterising the Mitochondrial Variation

In both species investigated here most of the mitochondrial variation occurred primarily at silent third codon position sites and involved nonsynonymous mutations, far more than would be expected from the nature of the genetic code (King and Jukes 1969). This suggests that the observed mitochondrial variation is neutral, remaining after the selective removal of disadvantageous mutations. Consistent with the general pattern for the evolution of mitochondrial proteins in animals, where deleterious mutations form an important component of polymorphic amino acids, both *ND1* and *COI* genes exhibit a relative excess of replacement polymorphisms compared with replacement divergence suggesting that some weak deleterious selection still operates (Nachman *et al.* 1996; Ballard and Kreitman 1994; Rand and Kann 1998; and Rand 2001). The excess of synonymous mutations between species was observed in comparisons of the both recently separated species *E. lucens* and *E. vallentini* and between these species and the relatively distant congener *E. recurva* (Chapter 6). All of the amino acid variants in *E. lucens* (and *E. recurva*) are functionally conservative amino acid substitutions and occur as singletons. Therefore, selection operating against them would have little effect on the interpretation of population structure because they occur at such low relative frequencies (Arnason *et al.* 2000). A similar substitution pattern has been revealed in the North Atlantic euphausiid species *Meganyctiphanes norvegica* with *ND1* (Bucklin *et al.* 1997; Zane *et al.* 2000; Papetti *et al.* 2005) and for the Antarctic species *Euphausia*

crystallophorias (Jarman *et al.* 2002) and *E. superba* (Zane *et al.* 1998) for both *COI* and *NDI*.

However, despite the lack of evidence for positive selection the possibility that the over-represented haplotype in each of the geographical regions may be linked to another mitochondrial region that is under selection cannot be excluded since the entire region acts as a single locus. Under such a selective sweep the mtDNAs through the entire population will be replaced with a phenotypically advantageous allele, followed by the accumulation of neutral variants (Maruyama and Birky 1991).

5.5. Spatial and temporal genetic structure

Both *E. lucens* and *E. vallentini* exhibited concordant patterns of variation with the nuclear marker *ITS-1*. However, qualitative and quantitative differences in the levels of mitochondrial variation both within, and between these species were evident. The geographical structuring of *NDI* and *COI* haplotypes at trans-ocean basin scales in *E. lucens*, with fixed differences observed between samples in eastern and western margins of the South Atlantic, is neither mirrored in *E. vallentini* samples separated by equivalent distances, nor by *E. lucens ITS-1* sequences. The greater degree of population structure observed in the neritic species compared to the oceanic one was predicted by the hypothesis that states the degree of population structuring will reflect differences in the extent of the geographical area inhabited by a species and the different

hydrographic current regimes that characterise these areas (Zane and Patarnello 2000).

No further spatial subdivision of *E. lucens* *NDI* haplotypes was observed at the meso-scale on either the Patagonian or the South African continental shelves, and no temporal differences were observed in the haplotype composition among samples recovered during 2001 and 2002 in the latter region. A similar result was reported for the Northern Hemisphere species *Meganyctiphanes norvegica* (Papetti *et al.* 2005). This implies that substantial mixing of continental shelf waters occurs and that inter-annual genetic variation is stable. The low levels of within-population variation, as evidenced by the shallow *NDI* phylogeny (Figure 3.3), implies either that the mutation rate for the mitochondrial markers is relatively slow, or that the observed polymorphisms are relatively recent, suggesting rapid population turnover of genetic variation.

5.5.1. Coalescent Estimate of Population Divergence in *E. lucens*

The coalescent-based estimate for *NDI* calculated with MDIV (Nielsen and Wakeley 2001) indicated that the *E. lucens* populations in the SW and SE Atlantic basins separated between 1.14-1.65 MYA. This places the split in the mid-to-late Pleistocene and subsequent to the estimated times of separation for *E. lucens* and *E. valleritini* presented previously (page 136). The population divergence estimates for the *E. lucens* populations coincide with the Great Patagonian Glaciations (GPG) when ice sheets extended to their maximum

proportions (Rabassa *et al.* 2005). Over the last few million years the earth's climate has alternated between ice ages and warmer intervals. Although the process of climate deterioration probably initiated around the Mesozoic, it culminated in the Miocene cold-warm climatic cycles that led to the development of ice ages (Rabassa *et al.* 2005). The last glacial maximum (18,000 B.P.) is thought to have resulted in the steepening of thermal gradients along polar frontal systems around Antarctica, an equatorial displacement of polar frontal systems, general cooling of surface waters, and low temperatures extending equatorial ward along the western coast of Africa, Australia and South America that indicate increased upwelling (CLIMAP 1976). If the genetic differentiation between *E. lucens* populations in the separate South Atlantic Ocean basins originated in the Pleistocene it has apparently been maintained through a subsequent succession of glacial and interglacial periods. Similar results have been obtained for the foraminifer species *Neoglobobadrina pachyderma* in the Atlantic (Darling *et al.* 2004) and the euphausiid species *Meganyctiphanes norvegica* (Zane *et al.* 1998; Bucklin *et al.* 1997; Zane *et al.* 1999; Papetti *et al.* 2004) and the chaetognath species *Sagitta setosa* (Peijnenburg *et al.* 2005) in both the Atlantic and the Mediterranean.

Of course, identical patterns of geographical differentiation cannot strictly be observed with these species, as *E. vallentini* co-occurs with *E. lucens* over the Patagonian continental shelf, but has never been recorded in the neritic waters

around the southern Africa. However, samples of *E. vallentini* were collected from neritic waters separated by equivalent distances in the south Indian, Atlantic and Pacific Oceans, and all contained the same two most frequently occurring *NDI* haplotypes (Figure 4.2), while the sequences of *COI* and *ITS-1* were monomorphic. This indicates either that the mutation rate is not fast enough in these gene fragments to have been affected by genetic drift, or, that the *E. vallentini* population is panmictic.

Expatriation is likely to be a general feature of marine plankton populations, a fact especially true for the smaller-bodied euphausiid species investigated here. Although euphausiids are strong swimmers and can maintain their position in the water column (Sundt and Fevolden 1996; Mauchline and Fisher 1969) by exploiting different current velocities and directions at different depths, the long-distance dispersal of adults and particularly, larvae, more likely involves passive transport in the prevailing currents (Mauchline and Fisher 1969).

Regarding the question of whether the southern extension of the South American continent represents a barrier to dispersal for *E. lucens* given that the southernmost limit of this species distribution coincides with the average position of the Antarctic Circumpolar Current (Lomakina 1964). The presence of two *NDI* haplotypes shared among *E. lucens* samples collected from New Zealand and Argentina suggest that recent contact between these regions has occurred. However, given the low level of sampling from the SW Pacific

waters around New Zealand, it is impossible to determine the extent of this process or in which direction it has likely occurred. It is possible that these individuals represent occasional expatriates from Argentinean waters and that substantial migration of animals in either direction around Cape Horn does not occur. This might explain why only occasional individuals of *E. lucens* have been recorded in the coastal waters around Chile (Palma and Silva 2004) and it implies that Cape Horn may be an effective barrier to dispersal into the south Pacific for this species.

On the other hand, *E. lucens*, like many euphausiid species, undergoes diel vertical migration (Gibbons 1993), and is thus able to tolerate large vertical gradients in environmental properties. This should predispose the species to being able to survive in a variety of oceanographic environments (Mauchline and Fisher 1969). It is possible that individuals do disperse around Cape Horn, but fail to establish populations in the Pacific coastal waters (e.g. due to competition with *E. vallentini*) which is the only euphausiid species able to maintain high abundances in this area (Palma and Silva 2004). I consider the latter explanation less likely as the studies of *E. lucens* in neritic waters off the Pacific coast of South America have only ever detected occasional individuals of this species and considered these to be immigrants from oceanic regions transported in sub-Antarctic waters via the West Wind Drift (Palma and Silva 2004).

It should be noted that *E. lucens* has been recovered in the Magellan Strait (Palma and Silva 2004) albeit at rather low densities, and transport of this cold-temperate species across the South American continental landmass, while avoiding the apparently harsh environmental conditions typically experienced around Cape Horn, may be achieved via this route.

5.6. Differences between *E. lucens* and *E. vallentini*

As well as displaying differences in the degree of geographical structuring of mitochondrial haplotypes, the shape of the *NDI* frequency spectra differed substantially between these species. In *E. lucens* populations for example, a different haplotype dominates all of the samples in each of the geographic regions to produce a negatively skewed frequency spectrum; in *E. vallentini* two haplotypes are maintained in all the samples with almost identical relative frequencies and the spectrum is bimodal. As a result, genetic diversity was higher in *E. lucens* with all the summary statistic estimators. An excess of rare and common *NDI* alleles was observed in *E. lucens* and *E. vallentini* respectively and consequently, all of the neutrality tests for *NDI* returned significantly negative results for the former species and positive but insignificant results for the latter. The lower variation observed in the sub-Antarctic *E. vallentini* compared to the cold-temperate *E. lucens* was predicted by the hypothesis that states the level of within-population variation will reflect differences in the degree of latitudinal displacement and habitat contraction experienced during the past glacial periods.

While the patterns of polymorphism in the different marker systems (mitochondria vs. nuclear) were congruent in *E. vallentini* (suggestive of a demographic or population level processes as being the cause) a selective sweep cannot be entirely ruled out. A selective sweep would explain the observed pattern where molecular variation within this species was low while the divergences between *E. vallentini* and other euphausiid species were similar to those for *E. lucens* (Nielsen 2005). Under this explanation the two high frequency *NDI* haplotypes could have resulted from recurrent selective sweeps neither of which has led to complete fixation.

Similarly in *E. lucens*, the incongruent patterns in the different marker systems (mitochondria vs. nuclear) and the excess of mutations of size 1 in *NDI* suggest that the departure from neutrality may also have involved natural selection. Two types of natural selection are known to affect an excess of mutations of size 1: genetic hitchhiking and background selection (Fu 1997). The former assumes that the locus is genetically linked to another locus at which an advantageous allele has been fixed recently, whereas the latter assumes that at the linked locus many mutations were deleterious. Because mitochondrial DNA lacks recombination it is expected to be subject to the effects of linkage and selection that reduces effective population size and increases the number of low frequency polymorphisms (Hill and Robertson 1966). Importantly however, population expansion (Fu 1994a) and sweepstakes recruitment (Hedgecock 1994) also produce an excess of rare alleles. Distinguishing between the

patterns left by a selective sweep, negative selection acting on slightly deleterious mutations across multiple loci, sweepstakes recruitment and population expansion is a difficult challenge, particularly with data from only two independent loci when the levels of polymorphism are low.

Neither genetic hitchhiking nor background selection are likely to be the cause of the excess of singletons in *E. lucens* because there are several lines of evidence against the presence of these two forms of natural selection. First, the *NDI* locus appears to behave normally in terms of the directions of mutations, as well as the spatial distributions of mutations. Second, if genetic hitchhiking or background selection has been operating, the level of genetic diversity at this locus would have been considerably reduced. A similar discrepancy between nuclear and mitochondrial markers occurs in humans (Akashi 1995; Hamblin and Di Renzo 2000; Harding *et al.* 1997; Hey 1997; Nickerson *et al.* 1998; Zietkiewicz *et al.* 1998), which it has been argued may be compatible with a population bottleneck (Fay and Wu 1999; cf. Hey and Harris 1999).

To reiterate: although variation within *E. vallentini* was low with all the markers compared to *E. lucens* the divergence between the former species and *E. recurva* (Chapter 6) is very similar to that between *E. recurva* and *E. lucens*. As negative selection acting on multiple loci will tend to reduce variability between species more drastically than variability within species (Nielsen 2005)

background selection may not be a plausible scenario to explain the patterns of polymorphisms observed in *E. lucens*.

For *E. lucens* the estimates of θ that take into account genealogical information were higher than those derived from non-genealogical summary statistics for both South Atlantic clades and the entire data set. This can be interpreted to reflect recent demographic expansion in this species (Crandall *et al.* 1999; Villa' *et al.* 1999; Su *et al.* 2001). For *E. vallentini* in contrast, similar θ values were obtained both among and between samples with both types of estimators and generally, the deviation from a constant-population size model was in the direction of population shrinkage (Tajima 1989; Fu and Li 1993; Fu 1995). Consistent with a population expansion scenario in *E. lucens*, the maximum likelihood estimates for the growth parameter under an exponential model, expressed, as $1/u_s$ was positive for both the Argentinean and South African clades. Furthermore, Fu's F_s test (1997), which has been shown in simulation studies to have the greatest power to detect growth in a wide variety of cases (Fu 1997; Ramos-Onsins and Rozas 2002), was significant for all the samples. Therefore, a more plausible explanation for the excess of rare *NDI* alleles in *E. lucens* is a recent population expansion. However, any expansion hypothesis would still have to explain the discrepancy in the patterns of polymorphism observed between the mitochondrial and nuclear markers in this species.

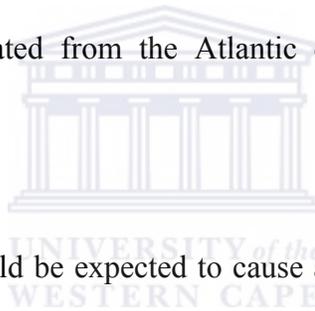
If the expansion indicated by the results of the EVE and Fluctuate analysis was preceded by a contraction in population size that was of sufficiently short duration and recent timing, a plausible demographic scenario for *E. lucens* can still be proposed that can explain the quantitative and qualitative polymorphism differences observed between the nuclear and mitochondrial markers. During and immediately after a bottleneck, a deficiency of rare variants is expected but as time passes the accumulation of recent mutations leads to an excess of low frequency segregating sites. Fay and Wu (1999) have shown that the time over which the pattern of polymorphism responds to change in population size will be greater for larger populations and the slower rate of drift for the nuclear gene relative to the mitochondria genes causes the former to reflect the effects of the population reduction (first stage of the bottleneck) and causes the latter to reflect the effects of the population expansion (the second stage of the bottleneck). Therefore, in a general way, both a bottleneck and an expansion model can be fitted to the *E. lucens* data. This scenario has been proposed to explain the conflicting patterns of polymorphism observed between nuclear and mitochondrial genes in the human lineage (Fu and Way 1999; Hey and Harris 1999).

It is possible that speciation between *E. lucens* and *E. vallentini* was associated with a bottleneck, perhaps of a different severity or duration from which only the former species has partly recovered. Alternatively, if the bottleneck was of equal magnitude, then the signatures of expansion indicated by the EVE and

Fluctuate results for *E. lucens*, and the contraction indicated for *E. vallentini*, may reflect differential species-specific rates of recovery from this event, possibly implying differences in the *NDI* substitution rate.

The magnitude of the difference between the estimates of divergence times for the *E. lucens* populations in the South Atlantic (1.14-1.65 MYA), and between this species and *E. vallentini* (3.65-4.25 MYA) perhaps implies that if a bottleneck was responsible for the observed pattern of polymorphism in either species, it must have begun much later than the time of the initial split from the common ancestor. The extremely low levels of mitochondrial variation displayed in *E. vallentini* are consistent with a more recent bottleneck and, as expected, all neutrality tests produced positive values. Lower levels of variation in *E. vallentini* were predicted by the hypothesis that the most southerly distributed species in the Southern Hemisphere would likely have experienced a greater degree of range contraction during the last glacial maximum (LGM). While the Drake Passage has never been completely covered by ice-sheets during any of the known Patagonian glaciations (Rabassa *et al.* 2005) surface ice was present in this channel during these periods (Rabassa pers.com). The associated water temperature reductions around Patagonia were marked during previous ice ages (CLIMAP 1967), which, almost certainly precluded the presence of either of these species during such periods. Exclusion from the waters of the Drake Passage during glacial periods could have led to the historical isolation of *E. vallentini* populations in the Atlantic and Pacific

Oceans (CLIMAP 1967; Rabassa *et al.* 2005). If this isolation was accompanied by a reduction in population size, the elevated levels of random genetic drift experienced in the smaller populations could have led to the fixation of a different haplotype in each region. During warmer interglacial periods contact between the Pacific and Atlantic may have been re-established leading to an admixture of populations driven by the Antarctic Circumpolar Current (Zane and Patarnello 2000). This could explain the presence of the two high frequency *NDI* haplotypes in all of the *E. vallentini* samples. A striking example of such a case involves the foraminifer *Pulleniatina obliquiloculata* that has repeatedly reinvaded the tropical Atlantic from refugia in the Indo-Pacific after being eliminated from the Atlantic during Pleistocene glacial cycles (Norris 1999).



However, a bottleneck would be expected to cause a more severe reduction in variation in the mitochondria genome due to the smaller effective population size relative to the nuclear genome. It is, therefore, possible that the complete absence of variation observed with the nuclear *ITS-1* within *E. vallentini* and *E. lucens* could be the result of limited sampling effort, as only ten *E. lucens* and twelve *E. vallentini* individuals were sampled this marker.

Two fixed polymorphisms and four fixed indels with a total length of 16 nucleotide base pairs separate the *E. lucens* and *E. vallentini ITS-1* sequences suggesting that the absence of any differences between *E. lucens* samples

collected in the separate South Atlantic basins may simply reflect slower mutation rate of the nuclear gene and a recent separation. Although there is a large variance around the single locus *NDI* divergence dates estimates, if the degree of differentiation is roughly proportional to time since separation, then one may expect to observe at least a single fixed polymorphism and a single indel in the nuclear gene *ITS-1* separating the regional South Atlantic *E. lucens* samples.

Previous studies on chaetognaths (Peijnenburg *et al.* 2005) and copepods (Goetze 2003) have shown that even the presence of several deep mitochondrial lineages may not be adequately detected using nuclear markers. Therefore, far larger sample sizes may be required to detect any variation in *ITS-1* in these species. However, given that the sampling locations were separated by thousands of kilometers of open ocean habitat, if the large-scale spatial organization of genetic variation revealed for the mtDNA in *E. lucens* is characteristic of the Atlantic population it should have been detected with the nuclear *ITS-1* gene with the sample sizes employed.

An alternative explanation for the patterns of polymorphism observed in *E. lucens* that is consistent with a shallow multifurcating genealogy, an excess of low frequency alleles and large differences between estimates of effective and census population sizes is if reproductive success was sweepstakes and depended on assembling a highly fit nuclear genotype (Williams 1975). The

inter-annual variability in krill abundance at certain locations in Antarctica (Brierley *et al.* 1997) has been explained in terms of recruitment variability (Seigel and Loeb 1995) and certainly the potential for sweepstakes reproductive success (Hedgecock 1994) exists in *E. lucens*. This species is considered to be highly fecund (Stuart 1992) and the early life stages likely suffer high mortality rates. Undoubtedly, these organisms reproduce under spatially and temporally varying oceanographic conditions that may affect their sexual maturation, choice of mate, fertilization success, survival of larvae and recruitment (Hedgecock 1994) about which we are largely ignorant.

However, unlike marine fish such as the Atlantic cod *Gadus morhua* L (Arnason 2004), *E. lucens* and *E. vallentini* do not display broadcast spawning and external fertilization, and thus sperm limitation and hence, competition among eggs for sperm, is unlikely. Instead reproduction occurs when the male passes a spermatophore to the female and fertilization is internal. The larvae are then released into the water and no parental care has been observed (although limited parental care is a feature of some other euphausiid genera).

Sweepstakes recruitment causes huge variances in the reproductive success and would also produce shallow topologies, as the turnover of genetic variance would be expected to change rather rapidly over time due the stochastic nature of the process. However, significant differentiation of genetic variation between cohorts may be difficult to observe in nature unless samples over

many generations are employed. The reason for this is that the “system” itself will have some memory. For instance, when the population is dominated by individuals bearing a single haplotype at high frequency with several other rare ones present the probability that any particular individual will leave offspring in the next generation will be proportional to its frequency in the current generation. In this way high frequency haplotypes have the highest likelihood of being the ones leaving offspring in the next generation. Therefore, structure may be observed with synoptic investigations but it may be ephemeral (Hedgecock 1994). While this scenario may be plausible, the data necessary to test the predictions of the sweepstakes reproductive success hypothesis would require sequence information from recruits and adults collected over several years if not decades, from the same location. Consequently, in *E. lucens* a sweepstakes reproductive success model cannot be ruled out with the *NDI* data presented here as the temporal stability in genetic turnover observed in successive years in the SEA may be expected over such short time scales under both historical and contemporary hypotheses. A study of the benthic sea urchin species *Strongylocentrotus purpuratis*, which displays broadcast spawning, tested the expectations of the sweepstakes theory (Hedgecock 1994) and revealed no evidence of large deviations in the variances in reproductive success and little evidence of reduced genetic variation in recruits of *S. purpuratis* relative to a previously reported sample of 145 *S. purpuratis* adults. Therefore, it remains debatable how general this mechanism is in marine

organisms particularly regarding those species such as euphausiids that do not display broadcast spawning.

5.7. Comparative Genetic Investigations

Relatively few studies of Southern Hemisphere populations of marine organisms have been conducted with which to compare the phylogeographic patterns observed in the current investigation. However, a recent study of the foraminifer *Neogloboquadrina pachyderma* (Darling *et al.* 2004), collected from the same general locations as those reported here for *E. lucens*, revealed congruent patterns of differentiation to those observed in the present study. These authors also found separate populations of *N. pachyderma* existed in the SW and SE Atlantic basins that must have been formed after the initial split from the Northern Hemisphere, which was estimated to have occurred around 1.5-1.8 MYA. The close relationship observed among *N. pachyderma* haplotypes collected in Antarctica and the northern Benguela system in the study by Darling *et al.* (2004) indicated to the authors that a Benguela relict population of *N. pachyderma* had been seeded from the Southern Ocean some time between 0.5-1.1 million years ago. This upper limit is very close to the coalescent-based estimate of the divergence time obtained with MDIV (Nielsen and Wakeley 2001) for *E. lucens* populations from the SW and SE Atlantic basins with the mitochondrial marker *ND1*.

Other authors have suggested that during glacial periods colder sub-polar water from the Southern Ocean could have been advected into the Benguela system (McIntyre *et al.* 1989), and since *N. pachyderma* had been present in the region for 420,000 years (Ulkes *et al.* 2000), they thought it likely that such incursions had occurred repeatedly. It has since been shown that episodic incursions of cold sub-Antarctic water into the Southern Benguela region do indeed occur as secondary circumstances of the regular shedding of Agulhas Current Rings into the SE Atlantic (Shannon *et al.* 1989). However, complete lineage sorting of the mitochondrial genes between *E. lucens* populations was observed suggesting either that a rapid population turnover of genetic variation has occurred, or that such recent incursions were unlikely responsible for transporting *E. lucens* individuals between the respective south Atlantic basins. Migration between the separate South Atlantic basins is not supported by the estimate of the effective migration rate (Nm) obtained with MDIV (Table 3.3).

If similar mechanisms to those proposed for *N. pachyderma* have been responsible for the patterns of differentiation observed for *E. lucens*, the location of the MRCA in this species should be somewhere in the SW Atlantic. This scenario would be compatible with levels of *NDI* nucleotide and haplotype diversity revealed in this study: the ancestral population as expected, exhibited higher levels of genetic diversity than that of a more recently founded descendent population centered in the SE Atlantic waters around South Africa (Table 3.2). However, if the data fit the model of a continuous, finite, linear

population greater genetic diversity will be expected near the center of the habitat (Wilkins and Wakeley 2002). It should be noted however; that all of substitutions in *NDI* separating individuals within the Argentinean population were transitions, and may therefore, be recent. The higher diversity in the SWA may have resulted from recent mixing of Pacific and SW Atlantic animals as evidenced by the shared mitochondrial haplotypes observed between New Zealand and Argentinean samples.

In contrast, the phylogenetic analyses indicated that the most basal, and therefore, ancient mtDNA haplotypes, were recovered from the SE Atlantic. Because mtDNA does not undergo recombination, independent replicated data cannot be obtained about the history of a population because anything that affects one part of the molecule directly affects all other parts of the molecule. However, by sequencing the larger *COI* fragment, one can rule out the possibility that the observed phylogeographic subdivision was a sampling artifact produced by chance in the smaller *NDI* fragment.

The relative reduction in the aerial extent of sub-Antarctic waters due to the equatorial displacement of the Polar Front during the LGM was greater for the western south Atlantic (CLIMAP 1976). This in turn might have resulted in a greater reduction in population size and loss of variability via extinction in the Patagonian populations compared to those in the South Africa. As multi-factorial causes of extinction are likely to be typical during periods of climate

change, there may be a low correlation between the amount of habitat reduction experienced by a population and the rate of extinction of alleles within that population. If the SE Atlantic population had previously been seeded by one of these now extinct lineages, then the most basal sequences could be found in this population.

No evidence supporting the seeding of the neritic *E. lucens* southern Benguela population from oceanic waters was observed and therefore, this population is likely self-maintained, at least on a short-time scale. The absence of any significant inter-annual *NDI* variation in southern Benguela samples suggests a stationary allele frequency distribution consistent with this scenario. The levels of variation are, however, relatively low and consequently the phylogenetic trees have very shallow topologies, suggesting either a very slow substitution rate for this population or a fairly rapid turnover of molecular variation. The marked differences between the mitochondrial and nuclear in levels of variation suggests that a selective sweep of the mitochondrial gene cannot be ruled out though low sampling intensity may be responsible for the low nDNA variation. The geographical patterns observed for *E. lucens* mtDNA haplotypes are therefore strongly supportive of a scenario where a recent allopatric separation of SE and SW Atlantic populations had occurred with subsequent independent evolution within each population.

The diversity and potentially complex configuration of processes operating at the population level presents a major challenge to testing historical hypotheses. Identifying the cause for a significant result in a neutrality test can be difficult to due to the confounding effects of selection and demography. Current methods to circumvent the problem of demographic confounding involve comparing multiple unlinked single copy nuclear loci using simulations under various demographic models (Wall *et al.* 2002; Akey *et al.* 2004; Hickerson *et al.* 2006 Estoup *et al.* 2004). The power to distinguish between hypotheses can be sensitive to how well a particular summary statistic extracts information relevant to the question of interest (Beaumont *et al.* 2002; Hickerson *et al.* 2006). Many of these studies go to great lengths in trying to exclude the possibility that rejection of a neutral model may be caused by demographic effects. To unambiguously determine which processes are the most likely to have produced the very different patterns of variation and geographical structure in *E. lucens* and *E. vallentini* may require far longer mtDNA sequences than were used here as well as data from multiple single copy nuclear genes (Hey 2005).

CHAPTER 6

RESULTS *Euphausia recurva*

6.1. Mitochondrial Variation, Phylogeography and Population Parameter Estimation

An example of the *E. recurva* SSCP patterns for *NDI* sequences with the experimental protocol described previously is shown in Figure 6.1.

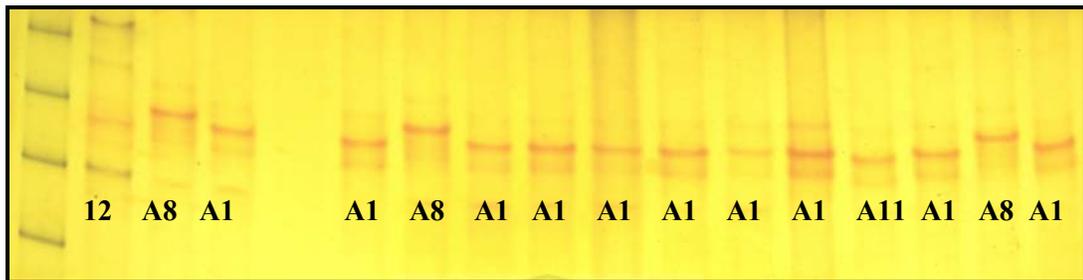


Figure 6.1. PCR-SSCP profiles showing a 156-bp fragment of mitochondrial *NDI* for a single *Euphausia recurva* sample collected from the west coast of South Africa. The labels correspond to the *NDI* haplotype designations presented in Table 6.1.

Most haplotypes were sequenced more than once. Sequencing the haplotypes for different individuals with the same SSCP pattern revealed no undetected mutations. Further, haplotypes with similar SSCP patterns were also mostly very similar in sequence information. However, the degree of difference in SSCP pattern was not always concordant with sequence information. For example, haplotypes A11 and A8 both differed by two substitutions from haplotype A1 but displayed very different patterns of gel mobility (Figure 6.1). Common for SSCP applications (Sunnucks *et al.* 2000), the two major banding systems were synergistically informative, however, and the silver staining procedure always produced strong and unambiguous upper bands only and

consequently the usefulness of the lower banding system to efficiently discriminate between different haplotypes was substantially reduced. As before, when gel irregularities (particularly in the lanes at the periphery of the gel), disrupted certain comparisons (Figure 6.1), it was necessary to rerun these PCR products in more centrally positioned lanes to reliably achieve the correct phenotype identification. When very similar SSCP patterns such as those between haplotypes A1 and A11 were observed (Figure 6.1), it was necessary to load the same sample at regular intervals across the gel and adjacent to a known standard for which sequence information had previously been confirmed. The most frequently observed SSCP profile was most commonly used for this purpose. To test the accuracy of the SSCP method, a random sample of individuals were sequenced, turning up no unexpected haplotypes.

6.2. Mitochondrial Sequence Variation

A total of two hundred and forty-eight specimens of *E. recurva* from the SE Atlantic (SEA), the Indo-West Pacific (IWP) and the NE Pacific (NEP) were examined. Thirty-five distinct *NDI* haplotypes were identified, of which 15 were singletons with the highest number of haplotypes found in the SEA (21), followed by the IWP (12), NEP (4) and New Zealand (2). Haplotype designation and their frequencies are shown in Table 6.1. With *E. lucens* as an outgroup, the total *E. recurva* *NDI* data set contained 40 variable sites, 27 segregating within *E. recurva* sequences.



UNIVERSITY *of the*
WESTERN CAPE

The transition/transversion ratio was 2.2 that accounted for 30 synonymous and 2 non-synonymous substitutions. Of the 32 mutations 10 were purine/purine transitions, 12 were pyrimidine/pyrimidine transitions and 10 were pyrimidine/purine transversions. The haplotypes differed from each other by between one to 19 substitutions. The non-synonymous changes involved a substitution of isoleucine with methionine at the third codon position $T_{127} \rightarrow A_{127}$ represented by a single individual from South Africa (A5), and valine with alanine at a second codon position $T_{135} \rightarrow C_{135}$, in two individuals from the Indo-West Pacific (D1 and D11). As observed in the previous analysis of *E. lucens*, all the nonsynonymous changes involved the replacement of one functionally similar amino acid for another (as Dayhoff *et al.* 1972). Percentages of adenine and thiamine were similar for sequences from the three geographical regions and averaged sixty-five percent. The most appropriate substitution model estimated in Modeltest using a likelihood ratio test was HKY + G ($2(\ln L1 - \ln L0) = 46.9219$, $df = 1$ $P = < 0.001$) with 5 rate categories (K) and a gamma distribution shape parameter (G) of 0.0623 and a Ti/Tv ratio of 2.6493. Using the Akaike (Akaike, 1974) information criterion the TVM + I substitution model was selected as the most appropriate with $K = 6$ and $P_{inv} = 0.7883$ (AIC).

6.3. Genealogy Reconstruction

A median joining network revealed a clear geographical pattern that can be superimposed onto the network with three distinct clades corresponding to the SEA and the NEP and IWP regions respectively (Figure 6.2). Strict reciprocal

monophyly however, is not observed with three haplotypes (ABC1, D2, D7) shared between SEA and IWP, and a single haplotype (ABC1) shared among all three southern hemisphere regions. The latter haplotype differed by 15 substitutions from the nearest individual in the southeast Atlantic representing an average uncorrected divergence of $\sim 9.6\%$, and by one substitution from the nearest individual in New Zealand and the IWP representing an uncorrected divergence of 0.64%. A further two *NDI* haplotypes (A17, A18) that were collected in the SEA also group in the IWP clade. The estimated number of mutations in the shortest network was 73.

Similar to the pattern observed previously for *E. lucens* (Chapter 3) the high frequency *E. recurva* haplotypes were internal to the networks and a number of rare or singleton haplotypes were derived from each high frequency type (Figure 6.2). Despite the large difference in sample sizes, the number of internal haplotypes (4) was the same in both the SEA, and IWP clades and higher than the NEP where a single haplotype (B3) occupied an internal position in the network. Also, the relative frequency of the rare haplotypes is lower in the South African population even though the combined sample size was nine times that of the IWP, suggesting that the higher diversity observed in latter sample is likely to be characteristic for this geographical region, and, if anything, has been underestimated in this study. A branch of ten substitutions in length connects the SEA and NEP clades, and the latter, connected to the IWP clade by a branch a single substitution in length. All except three

haplotypes (A17, D6 and C1) differed by a single substitution from their nearest neighbour. A total of five missing intermediates have been inserted into the genealogy, forming alternate pathways represented here as loops.

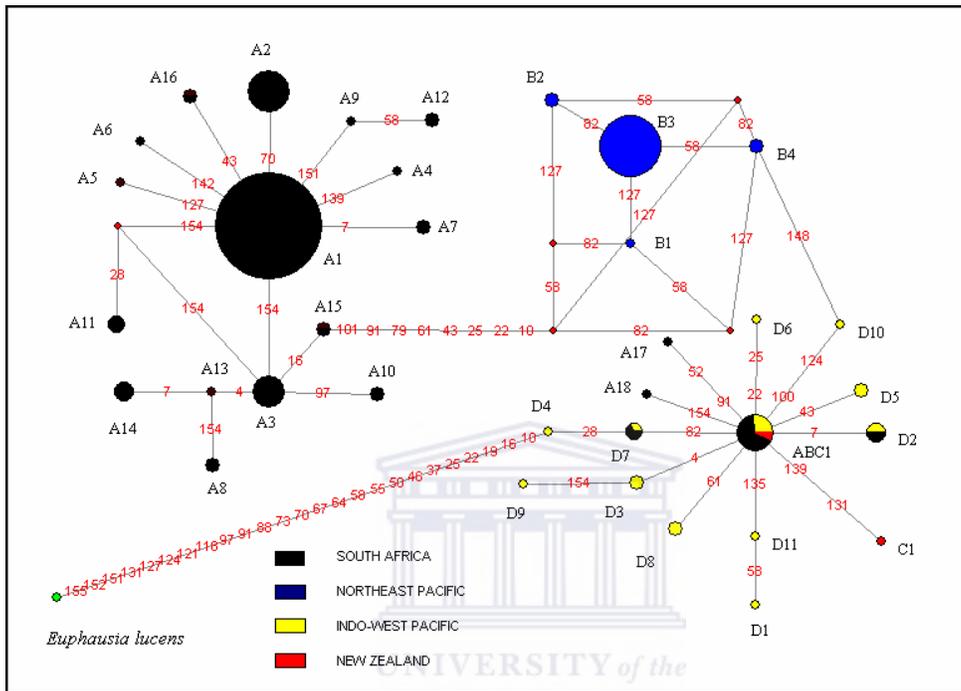


Figure 6.2. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. recurva* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences: The outgroup sequence *E. lucens* is shown in green. The size of each circle is proportional to the corresponding haplotype frequency. Red diamonds indicates missing intermediates. Numbers in red on each branch indicates the nucleotide site where the substitution occurred.

Ambiguous relations among haplotypes (i.e. reticulations) were most common in the NEP clade and they were absent in the IWP. This is the result of the homoplasy observed at sites 58, 82 and 127 which results in a lack of resolving

power between alternative connections. All conflicts could be resolved with the frequency criterion. Haplotype A11 provides an example. This haplotype shares a pyrimidine/purine transversion with haplotype A1, and a purine/purine transition with haplotype A3 both at site 154. However, haplotype A1 is more frequent than haplotype A3; therefore, the genealogy was resolved as shown in Figure 6.3. All conflicts could be resolved in this manner. Further, none of the conflicts in the network were between haplotypes from different populations, precluding the use of the geographical criterion (Posada and Crandall 2001) to provide additional support to correctly resolve the topology.

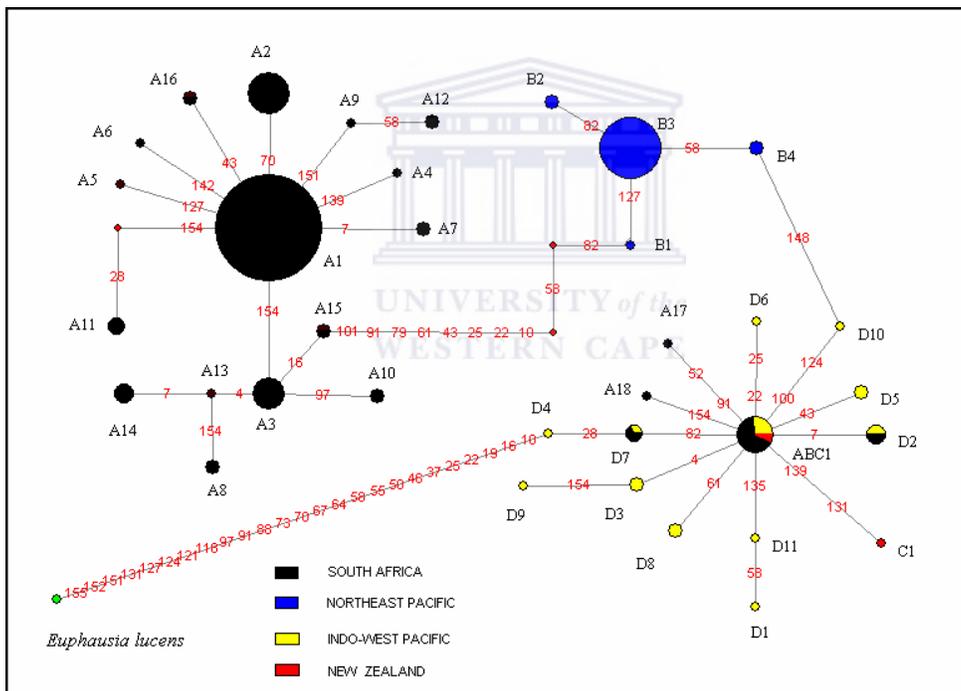


Figure 6.3. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of all *E. recurva* mtDNA haplotypes (filled coloured circles) based on *ND1* sequences with the loops resolved using various coalescent criteria (see text). Colours as in Figure 6.2.

The minimum level of divergence among the three *NDI* clades, disregarding the shared haplotypes, was as follows. Between SEA and NEP the level of divergence involved 10 synonymous substitutions (7 fixed) representing a 6.4% difference. Between SEA and IWP there were 13 substitutions (5 fixed), representing a difference of 8.3% and between NEP and IWP this was 0.64% (1 fixed). It should be noted that all of the fixed differences between the Atlantic and Pacific populations fall away if I adopt the geographical grouping scheme (and assume the presence of shared haplotypes to be due to ancestral retention in the SEA), as apposed the genealogical one proposed here.

The relative excess of rare alleles that were singletons is most pronounced in the IWP population, less so in the SEA population, and least in the NEP population (Figure 6.3). However, this result is heavily influenced by sample size differences among the three regions, and the fact that the SEA data represent the combination of two successive years sampling. Nevertheless, the SEA had the highest sample size and one expects a larger sample to pick up rare haplotypes if levels of molecular variation within populations are the same in both sampling regions. The fact that the smaller IWP sample contained a relatively greater number of singletons perhaps indicates that *NDI* variation is much higher in the IWP region.

It is notable that of the five *NDI* haplotypes shared among the southern hemisphere populations, one is relatively rare in the SEA, and relatively

common in the IWP (ABC1) supporting the hypothesis that present populations are still connected by gene flow over the alternative one of ancestral polymorphism retention in the SEA. The South American continent, represents a barrier to dispersal from the IWP to the Atlantic via Cape Horn that is unlikely to have been breached by *E. recurva*, where even the maximum annual temperature of 7°C (Palma and Silva 2004) would be presumably be lethal to this subtropical species given the latitudinal boundaries that characterize the range of its present geographical distribution.

6.3.1. Rooting of the Intraspecific Genealogy

With the outgroup sequence (*E. lucens*), which allows one to infer the ancestral nucleotides of each segregating site, the location of the MRCA is indicated to have been in the IWP (Figure 6.3). This region displayed the highest haplotype diversity (Table 6.2), and the highest number of occurrences in different populations. Furthermore, the haplotype shared among the three Southern Hemisphere regions (ABC1) was also the most frequently observed haplotype (13%) in the IWP sample, whereas in the SEA samples it was relatively rare (4.97%) (Table 6.1). Although, the absolute number of connections to other haplotypes is higher for the most frequent haplotype in the SEA (A1), proportionally this is higher for the most frequently observed haplotype in the IWP sample (ABC1) (Figure 6.3). On the available evidence the IWP is well supported as the location of the MRCA for *Euphausia recurva*.



UNIVERSITY *of the*
WESTERN CAPE

6.4. Regional and Temporal Comparisons

The average number of nucleotide differences between *NDI* sequences, was highest in the SEA, followed by the IWP, and the NEP. Haplotype diversity was highest in the IWP (Table 6.2). In four of the six samples collected in the NEP data all of the sequences recovered were monomorphic, and this is reflected in the low nucleotide diversity values for the combined NEP data set (Table 6.2). In the IWP population, the most frequent haplotype (ABC1) was shared with New Zealand and the SEA where, in the latter region, it occurred at a relatively low frequency (despite the large sampling effort!). In New Zealand, frequency information is absent, due to the sample size limitations, but it should be noted that this haplotype (ABC1) was only two substitutions removed from a second New Zealand haplotype (C1) that also placed in the IWP clade in the phylogenetic analysis (Figure 6.3).

No further geographic subdivision of *NDI* haplotypes within the southern Benguela region and no temporal differences were apparent between years with seventeen and thirteen haplotypes, found in the years 2000 and 2001 respectively (Figure 6.4). Of the 21 haplotypes recorded for South Africa, 10 were recovered in successive years (Figure 6.4), while the remainder occurred as singletons. A total of 32 mutations defined the 21 *NDI* haplotypes, made up of 31 synonymous and a single non-synonymous substitution, in the combined SEA sample of 181 individuals. The most frequent haplotype (A1) was the same in both years ranging in frequency between 53-84% in individual samples. Thirty-eight



UNIVERSITY *of the*
WESTERN CAPE

percent of the haplotypes were found in single samples only, the majority of which occurred as singletons. Genetic diversity among the South African samples was variable (Table 6.2), with the highest number of unique haplotypes (10), found in a single sample (6), collected during 2001. In addition, three samples (1, 3 and 6; Table 6.2) contained individuals bearing the haplotypes shared among Southern Hemisphere populations that clustered within the IWP sample in the median-joining network (Figure 6.3). This is reflected in the higher average pairwise distances among sequences (K) within these samples (Table 6.2).

It should be noted that individuals bearing the haplotypes shared among Southern Hemisphere populations were not randomly distributed among sampling sites in the southern Benguela region, but rather were concentrated in the three samples only (Figure 6.4) collected during successive years (2000 and 2001). The maximum genetic divergence within a single sample in this mitochondrial protein-coding gene was around 9%: a degree of difference more usually associated with among-species comparisons (Bradley and Baker 2001; Goetze 2003), and the highest yet recorded for any euphausiid species. This is reflected in nucleotide diversity values (Table 6.2). The absence of significant temporal genetic structuring within the SEA population enabled me to combine samples for increased statistical power.

Samples of *E. recurva* collected in the NEP between the years 1978-2001 were almost entirely dominated by a single haplotype (B3), suggesting stability in genetic turnover for this region. This low variation probably characterizes this region and is not a result of limited sampling, given the levels of polymorphism observed in the *NDI* sequences in the IWP population where the sample size employed was smaller.

6.4.1. *Spatial distribution and Direction of Mutations*

I examined the pattern of mutations to see if there is any unusual feature that could affect the subsequent analyses (Yu *et al.* 2000). All of the 40 mutations in the *E. recurva NDI* sequences could be inferred for the direction of mutation i.e. which nucleotide is ancestral and which the mutant. Table 6.3 shows the pattern of mutations in this gene fragment.

anc mutation	A	G	C	T
A....	0	12	1	5
G....	5	0	1	4
C....	0	0	0	7
T....	4	2	7	0

Table 6.3 Patterns of nucleotide change in *NDI E. recurva* sequences.

For the *E. recurva* sequences the number of mutations from x to y (x, y = A, G, C, or T) shows an excess of changes from A to G. The transition transversion ratio for this data was 2.649.

I also examined the spatial distribution of mutations. The entire 156 base pair region was divided up into 10 regions of equal length (16 base pairs) and the occurrences of mutations in each of the ten regions are 3, 4, 1, 3, 2, 2, 3, 2, 3 and 4. A chi-square test cannot reject that the mutation rates in all the regions are not the same ($\chi^2 = 3.0$, $df = 9$, $p > 0.5$)

6.5. Frequencies of Mutant Nucleotides in the Sample

As was observed in *E. lucens*, and again as already indicated by the phylogenetic analysis, Tables 6.4 and 6.5 reveal a conspicuous excess of singletons in both the total *E. recurva* sample and in the subsamples, with the exception of the NEP and combined SEA data sets.

Mutation Size	Total	
	Occurrences	Expectation
1	14	5.845
2	9	2.922
3	3	1.982
4	2	1.487
5	1	1.169
6	1	0.974
35	2	0.167
36	1	0.162
81	1	0.072
others.....	1	20.22
Total.....	35	35

Table 6.4. Frequencies of mutations of various sizes in the total *E. recurva* sample based on the *ND1* sequences (n =247).

Unlike the pattern observed in *E. lucens*, there was also an excess of size two mutations in all the subsamples with the exception of the SEA 2001 data set

(Table. 6.5).

Size	SEA ($k = 16$)		SEA 2000($k = 14$)		SEA 2001 ($k = 9$)		IWP ($k = 15$)		NEP ($k = 4$)	
	Count	Expect	Count	Expect	Count	Expect	Count	Expect	Count	Expect
1	3	2.200	4	2.020	7	1.864	11	2.626	1	0.795
2	6	1.100	5	1.010	1	0.932	3	1.313	2	0.375
3	2	0.733	1	0.673	0	0.621	0	0.875	0	0.265
4	2	0.550	0	0.550	0	0.466	2	0.656	0	0.199
5	0	0.440	1	0.404	0	0.373	0	0.525	0	0.159
6	0	0.367	0	0.337	2	0.311	0	0.438	0	0.133
7	1	0.314	0	0.289	0	0.266	0	0.375	0	0.113
13	0	0.169	1	0.155	0	0.143	0	0.202	0	0.061
19	1	0.116	0	0.106	0	0.098	0	0.138	0	0.042
26	0	0.085	0	0.078	0	0.072	1	0.101	0	0.031
35	0	0.063	0	0.058	0	0.053	1	0.075	0	0.023
36	0	0.061	0	0.056	0	0.052	0	0.073	0	0.022
56	0	0.039	0	0.036	1	0.033	0	0.047	0	0.015
60	0	0.037	0	0.034	1	0.031	0	0.044	0	0.013
81	0	0.027	0	0.025	0	0.023	0	0.032	0	0.009
148	0	0.015	0	0.014	0	0.013	0	0.018	0	0.005
172	1	0.013	1	0.012	0	0.011	0	0.015	0	0.004
Other	0	9.671	0	7.143	0	6.638	0	10.447	0	0.736
Total	16	16	13	13	12	12	18	18	3	3

Table 6.5. The frequencies of mutations of various sizes in the regional subsamples of *E. recurva*. K is the number of haplotypes in each sample.

Distinguishing between the retention of ancestral polymorphisms and migration becomes vitally important in order to derive the correct interpretation of population parameters estimates such as N_e and θ , as well as the results of the neutrality tests. If not accounted for, non-negligible migration can substantially bias estimates of N_e , either upward, or downward (Wang and Whitlock 2003). Although this distinction cannot be made with certainty here, the previous results lend greater support for an isolation-with-migration model for *E. recurva* populations and I therefore, present the following results for the genealogical clades identified in the phylogenetic analysis (unless otherwise stated).

6.6. Neutrality Tests

Of the five *E. recurva* data sets analyzed, three (total, combined SEA, and SEA 2000) have sites where more than two nucleotide bases are segregating and therefore the data violate the assumptions of the infinite-sites model (Table 6.1). To correct for this, the values of the variables used in the tests for these three samples were calculated from the sample genealogy estimated by maximum parsimony (Fu 1994b, 1996). Calculating the F and D values requires the inclusion of an outgroup sequence, which should be from a closely related population or species to avoid the complication caused by parallel and back mutations (Fu and Li 1993). The F and D values for the total *E. recurva* data set were calculated with *E. lucens* as the outgroup, while for each of the

regional sub samples, the most frequent haplotype from the most divergent clade in the m-j network was used.

Considering the total data set first, all four tests failed to detect a significant departure from mutation-drift equilibrium with Tajima's D (1989) and the Fu and Li's F (1993) test producing positive values, and both Fu and Li's D (1993) and Fu's F_s (1997) tests, negative ones (Table 6.6). In contrast, and Fu's F_s test was negatively significant for the all the subsamples as was Tajima's D with the exception of the NEP and SEA 2000 data sets, which were marginally non-significant. Notably, all of the test statistics were negative for the regional subsamples, with the exception of the F and D tests for the NEP sample, and the Fu and Li's D test for the combined SEA data set (Table 6.6). Thus, indicating a general excess of rare alleles and confirming the impression from a visual inspection of Tables 6.1 and 6.5. The relative proportions of external mutations in the different populations are also presented (Table 6.7), and a careful examination of these statistics can help us to understand the reasons why the F and D neutrality tests data only detected a significant excess of rare mutants for the SEA 2001 sample.

	Total Population	SEA combined	SEA 2000	SEA 2001	IWP	NEP
	248	167	105	62	36	45
Fu and Li's D	-0.25 (P > 0.10)	0.58 (P > 0.10)	-0.20 (P > 0.10)	-1.5 (P > 0.10)	-0.53 (P > 0.10)	0.90 (P > 0.10)
Fu and Li's F	0.20 (P > 0.10)	-0.58 (P > 0.10)	-1.08 (P > 0.10)	-2.0 (P > 0.05)	-1.02 (P > 0.10)	0.24 (P > 0.10)
Tajima's T	0.68 (P > 0.10)	-2.41 (p < 0.05)	-1.83 (P > 0.05)	-2.00 (p < 0.05)	-2.14 (p < 0.05)	-1.44 (P > 0.10)
Fu's F _s	-4.53 (P > 0.10)	-13.00 (P < 0.0001)	-10.29 (P < 0.002)	-5.79 (p < 0.05)	-11.04 (P < 0.0001)	-2.855 (P < 0.05)

Table 6.6. The results of the neutrality tests for the *E. recurva* *NDI* sequence data. The significance was estimated with 10,000 coalescent simulations in DNAsp version 3 (Rozas and Rozas, 1999).

Fu and Li's (1993) D and F tests compare the relative proportion of external mutations in a sample to the total number of mutations and the mean-pairwise difference between sequences respectively, a significantly negative test result indicates an excess of external, size 1 mutations. In all sub samples, the F values are more negative than D , reflecting the low levels of divergence between pairs of sequences relative to the total number of mutations, as evidenced by the shallow topologies recovered in the phylogenetic analysis (Figure 6.3). Comparing the results obtained for the separate sub samples, the most negative value of F was obtained for the SEA 2001 sub-sample, which had the second lowest K (Table 6.7) and the second highest excess of size 1 mutations after the IWP (Table 6.5). Similarly, the NEP population withstanding, the least negative F value was seen in the IWP (Table 6.6) where the high nucleotide diversity (Table 6.7) presumably offsets the contribution of size 1 mutations to this estimate (Table 6.5). Combining the separate SEA subsamples acts to reduce the excess of size 1 mutations by decreasing the number of singleton haplotypes while at the same time increasing the number of mutations of size 2 and 3 and higher (Table 6.5). Consequently, the F values produced for the combined data were less negative than those for the separate SEA subsamples (Table 6.6). A similar effect can be observed in the results for the D tests where, only in the 2001 data set is the value highly negative. The values obtained for both the D and F test for *E. recurva* are less negative than those previously observed for *E. lucens* (Chapter 3) where significant deviation from the assumption that all mutations in this DNA fragment are selectively

neutral and evolve according to the Wright-Fisher model were detected by both tests in each of the regional south Atlantic populations. Comparing the two species, the mean number of pair-wise differences between sequences (K) is much higher, and the percentage of external mutations lower in *E. recurva* samples, while the total number of mutations is similar in both species explaining the less negative F and D values produced for this species.

	Total	SEA (combined)	SEA 2000	SEA 2001	IWP	NEP
	253	172	110	62	36	45
K	5.65	0.78	0.84	0.68	1.55	0.22
S	23	12	10	9	17	3
% singleton	22	14	25	64	66	1
% external	18	14	25	64	33	0
# mutations	27	14	12	11	18	3

By Fu and Li's Test with outgroup

Table 6.7. Summary statistics of diversity for the *E. recurva* phylogenetic clades with Fu and Li's F and D tests with *E. lucens* as an outgroup. K = average nucleotide difference between sequences, S = segregating sites.

The test of Tajima (1989) compares two estimates of the population parameter θ , the first based on the mean number of pair-wise differences between sequences (K) and the second, on the number of segregating sites (S). A

positive value in this test will indicate an excess of common alleles, and a negative value an excess of rare alleles. Interestingly, the Tajima values were similar in all of the sub samples but were only negatively significant for the combined SEA and the 2001 SEA data sets (Table 6.6). The values of D for the 2000 data set were less negative than 2001, and marginally insignificant due to the high K value observed, relative to S in this sample (Table 6.7) where, although size 1 mutations are in excess, size 2 mutations are more so (Table 6.5). Combining the data sets collected in successive year's results in an increase in the number of segregating sites and K relative to the 2001 data set only (Table 6.7). Relative to the year 2000, combining the separate data sets results in a reduction in K values. The overriding effect of combining the separate SEA sub samples is a relatively greater increase in S compared to K , thus resulting in a more negative value for Tajima's D statistic than either of those produced for the two separate samples (Table 6.6). A large sample picks up rare haplotypes which increases S , compared to K , resulting in a more negative D value. Several NDI haplotypes were shared between samples collected in successive years in the SEA, many of which occurred as singletons. The act of combining the data from the two years simultaneously decreases the number of singletons, while increasing the number of mutations of high size (>2). So again one can see that with regards to Tajima's test, the contribution to K from mutations of sizes > 1 in the combined data set offsets the contribution to θ_w from mutations designated as size 1 in the data for the separate years. Fu's (1997) F_s statistic, which is sensitive to the presence of rare mutations

relative to expectation based on nucleotide diversity, was highly significant for all the regional subsamples except the NEP, with the most negative values produced for the combined SEA sample followed by the IWP (Table 6.6). Again the effect of combining the data sets collected in successive years in the SEA results in a much more negative value. As the F_s test was significant while the D and F tests of Fu and Li (1993) were not, the implication is that background or purifying selection is unlikely to be responsible for the excess of rare alleles, and instead is probably due population expansion or genetic hitchhiking (Fu 1994b).

It was shown previously that no replacement substitutions were observed between *E. lucens* and *E. vallentini* (Chapter 3) for NDI , precluding the use of the McDonald–Kreitman (1991) test for selective neutrality for these putative sister taxa (Jarman *et al.*, 2000). However, it was possible to apply this neutrality test for the *E. recurva* (Chapter 5) and *E. lucens* data. Twelve fixed differences were observed in a comparison between these species, eleven of which were synonymous changes (sites 19, 22, 55, 64, 67, 70, 73, 88, 97, 118, 121), and one non-synonymous change (site 152). A total of 46 polymorphic sites existed where synonymous substitutions had occurred, and 11 of these were fixed between species, giving a ratio of $46/11 = 4.18$ while for the non-synonymous this was $5/1 = 5$. The Neutrality Index (NI), was 1.196, and again neither Fisher's exact test ($P = 1.00$) nor a G test ($P = 0.87405$) with Williams' or Yates' correction ($P = 0.705$), were significant. Although insignificant this result

indicates that a general pattern of excess synonymous mutations occurs, not just between recently separated species (Jarman *et al.*, 2000) such as *E. lucens* and *E. vallentini*, but is also maintained between relatively distant euphausiid congeners. It should be remembered that as the McDonald-Kreitman (1991) test compares polymorphisms both within and between species, it tests for the deviation from neutrality expectations over a longer time period than the single-lineage tests applied previously. According to neutrality theory (Kimura 1983, Wright-Fisher 1953) this should be around $4N$ generations (Kimura 1983; McDonald-Kreitman, 1991). It is also robust to with respect to demography.

Therefore, if the divergent sequences found in the SEA result from migration, most of the genealogical populations investigated here deviate from the assumption that all mutations in this DNA fragment are selectively neutral and evolve according to the Wright-Fisher model, with a constant effective population size since the MRCA of the sample. However, it should be noted, that if the divergent haplotypes are due to the retention of ancestral polymorphism in the SEA population, and the geographical grouping of samples is justified, deviation from the neutral model is much less severe in all the subsamples (results not shown) which is consistent with the result obtained for the entire *E. recurva* data set (Table 6.6). It should be remembered that the McDonald-Kreitman (1991) test compares polymorphisms both within and between species, and therefore tests for the deviation from neutrality

expectations over a longer time period than those based on single species lineages. Consequently, it is unable to discriminate between recent and ancient selection.

6.7. Effective Population Size Estimates

Again in samples where multiple hits are evident, the number of mutations separating a pair of sequences was calculated from the maximum parsimony tree (Fu 1994b). Taking the total sample first, it can be seen that there is a considerable range in the estimates of θ among the different estimators (Table 6.8).

Methods	Total Sample	Northeast Pacific	Southeast Atlantic	Indo-West Pacific
Tajima	4.538	0.750	1.267	1.509
Watterson	5.845	0.795	2.200	2.626
Fu's UPBLUE	10.486	1.047	5.123	4.695
Watterson ^c	3.030	n/a	3.468	1.780
EVE	n/a	1.734	10.001	39.943
Fluctuate	n/a	n/a	6.450	46.640

Table 6.8. Estimates of θ for *E. recurva* *ND1* sequences. Wattersons^c = Wattersons estimate with singleton mutations omitted. $\theta = 2Nu$.

For $u = 2.81 \times 10^{-6}$ per sequence, per generation, the Watterson (1975) Tajima (1983) and UPBLUE (1994a) estimates of θ give effective population sizes (N_e) of 1.04, 0.81 and 1.86 million individuals respectively (Table 6.9). When singleton mutations are excluded (θ_l) this is substantially reduced (Table 6.8) and corresponds to a N_e of 0.54 million individuals (Table 6.9). When singleton mutations are in excess, it is sensible to obtain an estimate that does not rely on singleton mutations (Yu *et al*, 2000). For this reason Tajima's estimate and θ_l are the methods of choice for this data set.

Consider each subsample separately; the highest and lowest estimates of θ were observed in the IWP and the NEP clades respectively, with all of the methods except θ_l , (Table 6.8) which is consistent with the candidacy of the former region as the site of the MRCA. As was observed for *E. lucens*, the UPBLUE estimate of θ for the total *E. recurva* population (Table 6.8), is more than twice as large as the Tajima (1983) estimate. A similar situation is seen in the IWP (UPBLUE $\theta = 4.695$ / Tajima's $\theta = 1.509$), and SEA (UPBLUE $\theta = 5.123$ / Tajima's $\theta = 1.267$), but in the NEP similar values were obtained with the different estimators (UPBLUE $\theta = 1.047$ / Tajima's $\theta = 0.750$).

The Watterson (1975) Tajima (1983), and UPBLUE (1994a) estimates correspond to effective population sizes in the IWP of 0.467, 0.269 and 0.835 million individuals respectively (Table 6.9). The same methods applied to the SEA data set produces estimates of effective population sizes of 0.391, 0.225

and 0.912 million individuals. In the NEP this was, 0.141, 0.133 and 0.186 million individuals (Table 6.9). Estimates of θ_1 for separate sub samples are closer to the Tajima (1983) and Watterson (1975) estimates, reflecting the relative contribution to these statistics from mutations of size 1 (Table 6.8).

Methods	Total Sample	Northeast Pacific	Southeast Atlantic	Indo-West Pacific
Tajima	0.807	0.133	0.225	0.269
Watterson	1.040	0.141	0.391	0.467
Fu's UPBLUE	1.860	0.186	0.912	0.835
Watterson ^c	0.540	n/a	0.617	0.320
EVE	n/a	0.308	1.780	7.107
Fluctuate	n/a	n/a	1.147	8.298

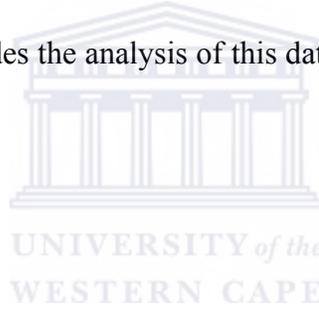
^c Watterson estimate with singleton mutations omitted

Table 6.9. The effective population size estimates (in millions of individuals) for *NDI* based on the values of θ presented in Table 6.8.

The Watterson estimate of θ with mutations of size 1 excluded (θ_1) gives effective population sizes in the IWP and SEA respectively of 0.32 and 0.62 million individuals (Table 6.9). It should be noted that the effective population size of the total sample is not equal to the sum of the separate subsamples (Table 6.9). This is natural because the populations are not isolated from each other.

6.7.1. Coalescent Estimators in a Variable Environment (EVE)

As was observed for *E. lucens* (Chapter 3), the results of the neutrality tests indicate that the constant population size assumption may not hold for *E. recurva* populations, therefore, θ was also estimated EVE package (Vasco *et al.* 2001). The EVE estimates of the population mutation rate (θ) were very much larger than those produced by UPBLUE for both the SEA and the IWP (Table 6.8). Positive growth is also indicated for both the IWP ($g = 37$) and SEA ($g = 8$) populations, but not in the combined data set ($g = 0.25$). The EVE estimates of g are much lower than those presented earlier for the separate *E. lucens* populations (Chapter 3). The low diversity in *NDI* among the NEP sequences however, precludes the analysis of this data with the EVE coalescent estimator.



6.7.2. Fluctuate

Under the model of no growth, the fluctuate (Kuhner *et al.* 1995, 1998) maximum likelihood estimates of θ were greater for, the IWP ($\theta_{IWP} = 0.498$), than the SEA $\theta_{SEA} = 0.057$ giving effective population sizes of 1.382×10^7 and 1.582×10^6 individuals respectively for these populations. The highest log-likelihoods however, were obtained under the exponential growth model for both the IWP ($\theta_{IWP} = 0.598$) and SEA ($\theta_{SEA} = 0.083$), corresponding to effective population sizes of 8.30×10^6 and 1.15×10^6 individuals respectively (Table 6.9). Maximum likelihood estimates of the growth parameter under an exponential model expressed as $1/\mu_s$, where μ_s represents the mutation rate per

site per generation, were positive for both clades, giving an unscaled exponential growth rates per generation of 0.00001204 and 0.000002178 for the IWP and SEA clades respectively. Figures 6.5 and 6.6 plot the approximate 95% confidence intervals of θ and g for these populations and show how variable they are ($\theta \pm 1.96^* \text{ s.d.}$). Growth was considered significant if g was larger than three standard deviations from zero and, under this criterion, both the SEA and IWP populations of *E. recurva* show evidence of significant growth with the highest rates displayed in the IWP (Table 6.10).

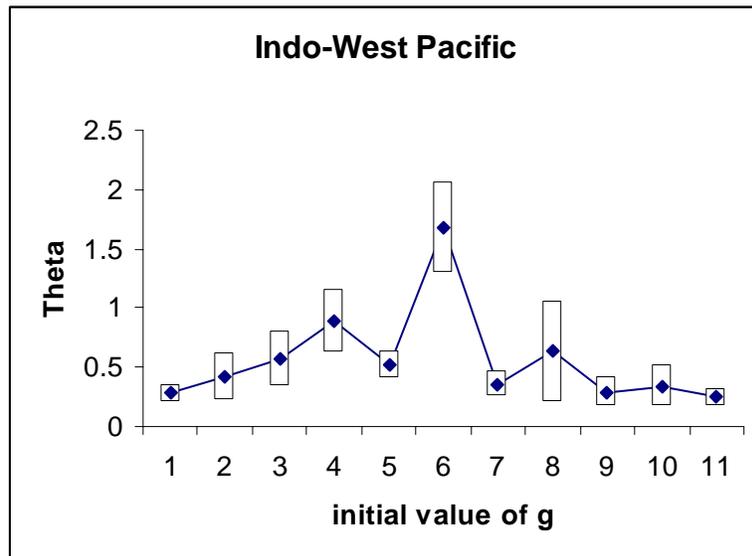
It should be noted, that the Fluctuate estimates of g show an upward bias (Kuhner *et al.* 1998), a property the authors claim will be inherent in any method estimated from the genealogical structure. The same authors also reported a smaller upward bias in θ due to the correlation between the two parameters. The best way to reduce this bias is to sample additional unlinked loci (Kuhner *et al.* 1998), the results obtained with *NDI* should be viewed as tentative until this can be accomplished.



g0	IWP					SEA				Cov (θ , g)
	θ	s.d. θ	g	s.d. g	cov (θ , g)	θ	s.d. θ	g	s.d. g	
5	0.277	0.035	588.16	34.78	3.65E -6	0.112	0.006	149.11	15.73	2.21E -5
10	0.416	0.102	595.74	54.63	9.02E -3	0.075	0.005	107.10	18.43	1.42E -6
15	0.571	0.117	692.41	40.66	7.35E -5	0.057	0.004	50.42	18.88	4.57E -5
20	0.891	0.139	1129.3	44.11	9.02E -5	0.080	0.004	149.34	18.05	6.51E -6
50	0.524	0.060	586.11	23.64	1.57E -5	0.114	0.006	251.87	22.88	1.39E -6
100	1.679	0.195	688.69	19.94	2.78E -6	0.059	0.003	77.88	16.66	1.07E -5
-5	0.359	0.054	570.21	32.84	3.61E -4	0.113	0.006	152.17	20.14	9.27E -8
-10	0.631	0.220	721.07	137.37	1.61E 0	0.101	0.006	162.01	18.25	5.77E -5
-20	0.287	0.064	553.39	50.47	6.89E -2	0.053	0.003	66.04	15.43	7.51E -5
-50	0.342	0.090	561.36	85.55	4.15E +2	0.054	0.003	47.00	18.15	3.65E -5
-100	0.246	0.040	525.70	36.85	3.11E -5	0.100	0.005	117.51	16.48	1.15E -7
mean	0.598	0.108	668.65	52.40	5.1E -1	0.083	0.005	121.035	18.184	0.086
Final	0.620	0.085	827.74	34.37	-3.29E -12	0.091	0.005	69.00	12.16	1.96E -5

Table 6.10. Results from the Fluctuate analysis varying the initial growth rate (g0). Presented are the *NDI* results for the southern hemisphere populations of *E. recurva*. The most important results (θ and g) are highlighted in bold text

a)



b)

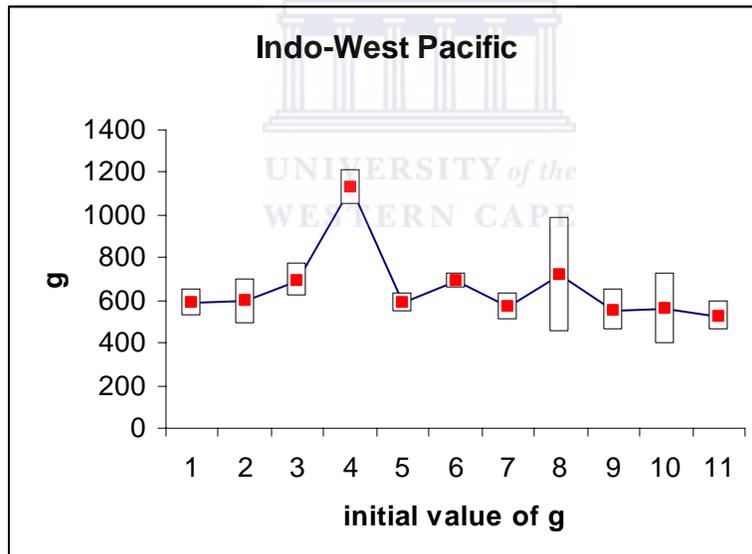
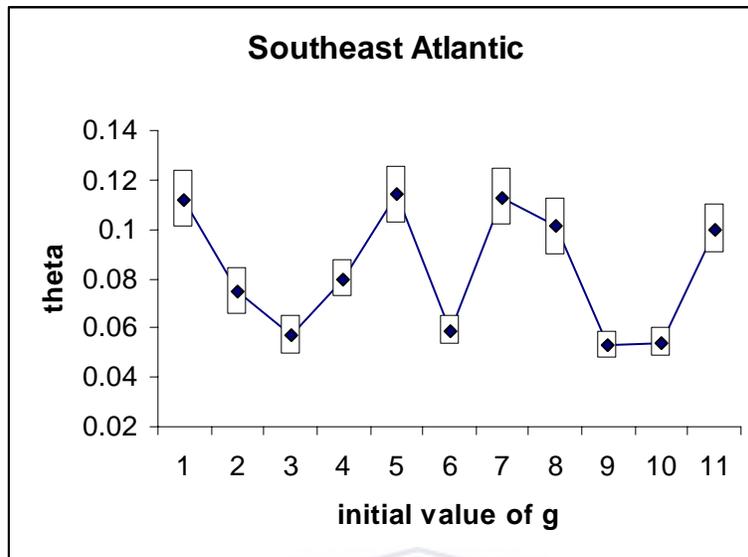


Figure 6.5a-b. The results from independent runs of Fluctuate for the *E. recurva* IWP sample with differing initial values of g . Runs 1-11 correspond to the initial values of g given in the leftmost column in Table 6.10.

a)



b)

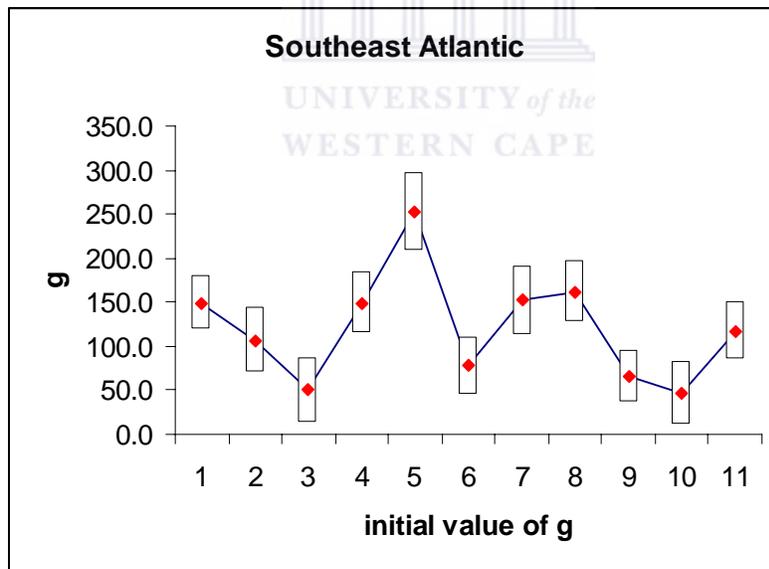


Figure 6.6a-b. The results from the independent runs of Fluctuate for the *E. recurva* SEA sample with differing initial values of g (see table 6.10. for details).

6.8. MDIV

To distinguish between isolation and migration as explanations for the observed pattern of genetic divergence between Atlantic and Pacific *E. recurva* populations recently developed methods for fitting the full isolation model with migration were applied to the data using the program MDIV (Nielsen and Wakeley 2001). The coalescence-based population parameters, including the maximum likelihood estimators of divergence times (T), number of migrants (M), and θ_{ML} for the three regional *E. recurva* populations were estimated with MDIV and the results are listed in Table 6.11.

Comparisons	Theta	2Nm	PDT	TMRCA	PDT (MYA)	TMRCA (MYA)
IWP-SEA	2.364	0.361	3.15	3.8113	2.65	3.21
NEP-IWP	1.147	0.016	1.64	2.5792	0.67	1.05
SEA-NEP	1.618	0.056	1.14	4.8121	0.71	2.77

Table 6.11. Results of the coalescent analysis of *E. recurva* populations based on NADH1 (ND1) sequence data. Theta, the migration parameter (2Nm) and population divergence time are all modes of the posterior probability distributions generated in the programme MDIV. All results are in coalescent units except the last two columns, which list PDT and the time to the most recent common ancestor (TMRCA) in millions of years ago (MYA) based on a molecular clock of 1.8% per million years and assuming a single generation per year (see text for details).

To convert to estimates of percentage divergence time (PDT) and expected time to the most recent common ancestor (TMRCA) to actual time (millions of years ago, MYA) I applied the previously described mutation rate of u of 2.81×10^{-6} mutations per sequence per generation. Using this estimate of u the effective population size (N_e) was calculated from the estimated θ_{ML} using the formula $N_e = \theta_{ML} / 2\mu$ and then used to convert the T values scaled by N_e , in number of generations.

The estimates of the population divergence time (PDT) place the time of separation between the IWP and the SEA at around 2.65 MYA indicating late Pliocene divergence with the most common ancestor suggested to have occurred around 3.21 MYA (Table 6.11). The effective migration rate ($2Nem$) for these populations was 0.381 implying that these *Euphausia recurva* populations have been isolated with moderate levels of gene flow since they began to diverge (Nielsen and Wakeley 2001). A more recent separation was indicated for the NEP and IWP populations with a PDT of 0.67 MYA and a TMRCA of around 1 MYA and little evidence of any migration (Table 6.11). The estimate of the timing of the split between the NEP and the SEA populations was almost identical to that of the Pacific populations (0.71 MYA) while the TMRCA estimate (2.77 MYA), was very similar to that for the SEA-IWP comparison and again negligible migration was indicated (Table 6.11). The maximum likelihood effective population size estimates for the three

comparisons were 1.13, 0.408 and 0.58 million individuals respectively (Table 6.11) indicating the highest population size in the SEA and lowest in the NEP.

6.9. Cytochrome Oxidase 1 Variation within *E. recurva*

For completeness in comparisons involving the three congeners investigated here, four individuals of *E. recurva* from the NEP, and six individuals from both the IWP and SEA, were sequenced with the *COI* primers used in the previously cited study (Jarman *et al.*, 2000). The *COI* sequences were 640 bp in length and the total data set for *E. recurva* had 65 polymorphic sites and defined 14 haplotypes (Table 6.12). The total of sixty-nine substitutions accounted for 67 synonymous and 2 non-synonymous substitutions. The non-synonymous changes involved a substitution of histidine with glycine at the third codon position $C_{72} \rightarrow A_{72}$ represented in two individuals from South Africa, and isoleucine with valine at a first codon position $A_{628} \rightarrow G_{628}$, represented by a single individual from the same region (Table 6.12). As was found in the previous analysis of this species and *E. lucens* using *NDI*, all the non-synonymous changes involved the replacement of one functionally similar amino acid for another (Dayhoff *et al.* 1972). Percentages of adenine and thiamine were similar for the three regions and averaged sixty-five percent. Again more non-synonymous mutations were segregating within species than between, with a single fixed non-synonymous change observed between *E. lucens* and *E. recurva* and *E. recurva* and *E. vallentini* that involved a substitution of histidine with glutamine at the third codon position $C_{72} \rightarrow G_{72}$.

6.9.1. The Genealogical *COI* Estimates

The evolutionary relationships of *E. lucens*, *E. recurva* and *E. vallentini* based on the *COI* sequence data were investigated using phylogenetic reconstructions. Maximum parsimony (MP) and maximum likelihood (ML) searches were conducted with the heuristic search approach of PAUP* 4.0 (Swofford 1998). One hundred repeated randomized input orders of taxa were used for all MP analysis to minimize the effects of entry sequence on the topology of the resulting cladograms. MP analysis were conducted with the steepest descent option, and with accelerated character transformation (ACCTRAN) optimization, tree bisection-reconnection (TBR) branch swapping with a starting tree obtained by step-wise addition, save all multiple trees (MULPARS), and zero length branches collapsed to yield polytomies settings in place. The ML searches were conducted using models of DNA substitution selected by a series of likelihood ratio tests (Heulsenbeck and Crandall 1997). The 640 base pair data set contained 146 parsimony informative sites and under parsimony produced eight most parsimonious trees of 364 steps in length (Figure 6.7). The ML tree constructed using the GTR + G substitution model had a $-\ln L$ 2009.77 (Figure 6.8). Statistical confidence in the stability of tree nodes was calculated by nonparametric bootstrap (Felsenstein, 1985) with 1000 replications (Hedges 1992). Employing *Euphausia pacifica* as an outgroup, sequence molecular evolutionary relationships for the three species investigated here were consistent across the two linked gene loci and tree construction methods employed.

In all of the trees recovered, the relationships within the genus *Euphausia* were consistent with the taxonomic classification based on morphological characters and the same three geographically delineated *E. recurva* clades identified in the phylogenetic analysis of *NDI* (Figure 6.3) were also recovered using *COI* with extremely high bootstrap support (Figures 6.7 and 6.8). Furthermore, the two individuals collected in the SEA and previously identified as putative migrants based on the information from *NDI* (Figure 6.3) show congruent patterns with the *COI* fragment (Figures 6.7 and 6.8). Genetic diversity in *E. recurva* for *COI* was highest in the IWP and thirty-four fixed differences separated the Atlantic and Pacific clades; the average nucleotide difference between these populations was high ($K = 43.5$). In the IWP population every sequence investigated recovered a unique *COI* haplotype. Similarly, genetic diversity for *COI* was lowest in the NEP, with four fixed differences separating this clade from that of the IWP, with an average nucleotide difference between these populations of 9.375. These results are again quantitatively and qualitatively compatible with those obtained from the *NDI* for these populations (Table 6.2) as expected from a linked locus. The average *COI* divergence between *E. recurva* and both *E. lucens* and *E. vallentini* respectively, was 96 and 101 substitutions, 70 of which were fixed between the former species pair and 74 between the latter.

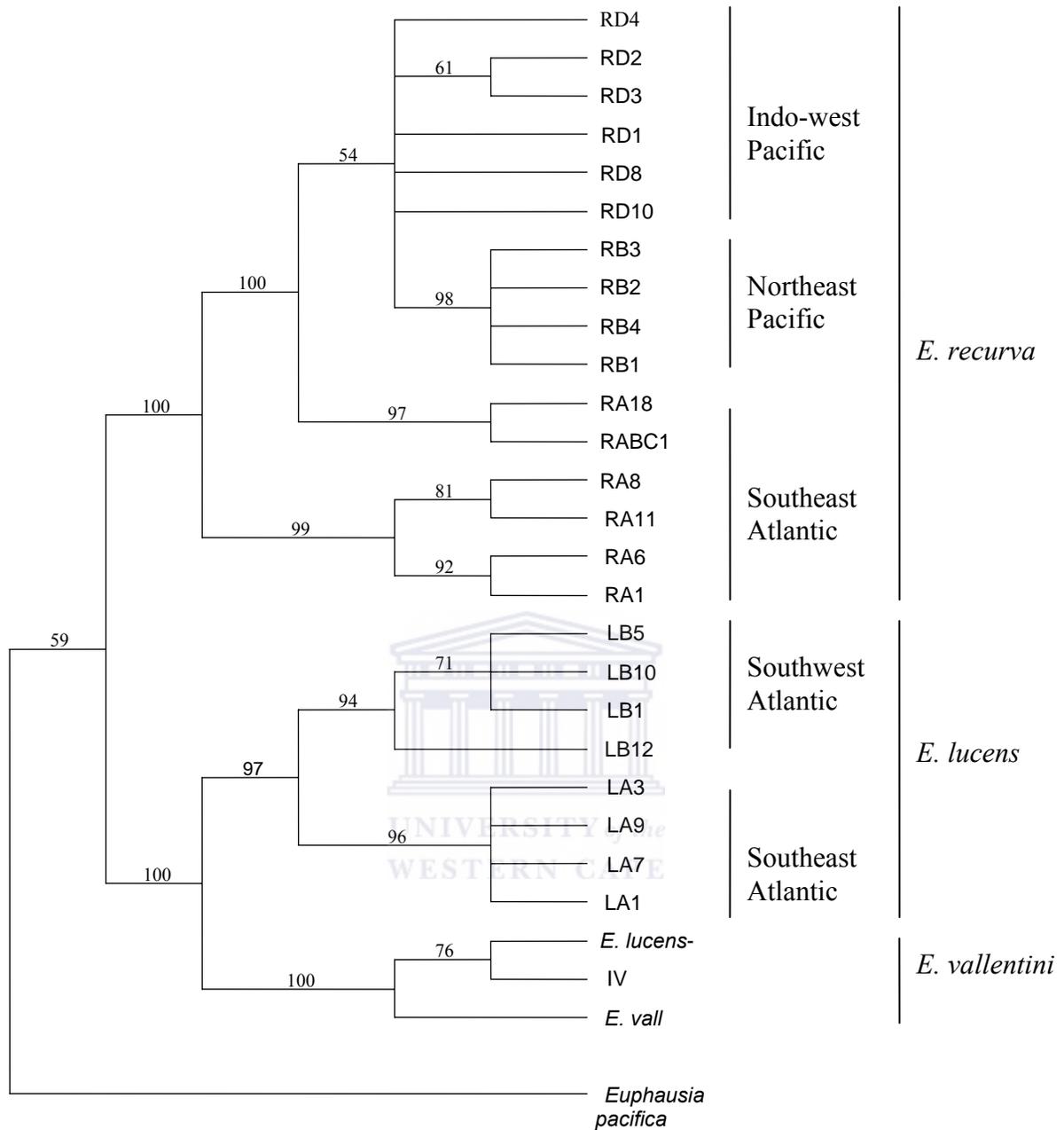


Figure 6.8. The maximum likelihood tree for the *COI* sequences of the three euphausiid species. Haplotype labels are the same as those employed for *NDI* as before with the prefix R = *E. recurva*, L = *E. lucens*. The numbers above the branches represent bootstrap values calculated with 1,000 replicates. The tree is rooted with the outgroup sequence *Euphausia pacifica*. The sequence labelled *E. lucens*- again groups within the *E. vallentini* clade.

Consistent with the morphological taxonomy *E. lucens* and *E. vallentini* form a sister group in all the trees with high bootstrap support (Figures 6.7 and 6.8). As previously demonstrated, incongruent (Chapter 2) and congruent (Chapter 3) phylogeographic patterns between mitochondrial and nuclear markers have been observed in other euphausiid species. This provides a reminder of the need to include data from independent unlinked loci if robust inferences about the evolutionary and demographic processes that have shaped extant patterns of genomic variation in a species are to be made with any level of confidence. Something that is particularly relevant when attempting to separate the possible confounding effects of population demography from those of natural selection. The nuclear marker investigated here (*ITS-1*) provides an opportunity to test whether the patterns of polymorphism in the different marker systems (mitochondria vs. nuclear) are concordant in *E. recurva* and therefore likely to be caused by demographic or population level processes; or incongruent whereby, a departure from neutrality may have involved natural selection.

6.10. Nuclear ITS-1 Results

The nuclear marker *ITS-1* was sequenced in 31 individuals including representatives from each of the geographical clades previously recovered in the phylogenetic analysis of the mtDNA data. At 384 base pairs, the length of *ITS-1* in *Euphausia recurva* was shorter than the other two crustacean congeners investigated here, and the average GC content at 59.9% was also slightly lower in this species relative to the other two. Alignment of *E. recurva*

sequences was relatively straightforward since there is great similarity among the sequences and no evidence of saturation is apparent in the data. No nucleotide positions in the alignment have all four nucleotide bases represented, and only two sites 152 and 253, have three.

To assess the degree of intragenomic variation, two individuals were chosen from each of the three regional samples and five independent clones covering the *ITS-1* region were sequenced from these individuals. Clones recovered from a single individual were labelled with the same letter code according to the location in Figure 6.3. Ten clones from each of the regional populations were analysed. Of the 30 clones, 15 were different. In the 10 clones recovered from the NEP, one (NEP1) was identical to the *ITS-1* sequence obtained via direct PCR, while the remainder differed by between one and five polymorphisms from this sequence. Of the 10 clones recovered from the IWP, three occurred as singleton variants, while three others were recovered at higher frequencies. In this population the maximum divergence from the original sequence obtained via direct PCR was relatively high due to the inclusion of a single clone (IWP1) that differed by a total of 16 base pairs from the *ITS-1* sequence obtained from direct PCR. Seven of these consist of the four-base pair indel between bases 315-318, and three consecutive polymorphisms at bases 210-213. The remainder represent nucleotide polymorphisms resulting from single base mutations that are transformed by another base or to a gap. When this clone was omitted the maximum divergence among these clones was four

substitutions. In the SEA population, four of the five clones (SEA2-5) were identical to sequence obtained via direct PCR, while the fifth (SEA1), differed by eight substitutions. Five clones (SEA 3:1 SEA 3:2, SEA 2:6-8) were inferred to be the result of recombination and were omitted from the phylogenetic analyses (Templeton *et al.* 1987; Crandall and Templeton 1993). The aligned sequences for the remaining 25 clones along with those obtained via PCR and direct sequencing are given in Figure 6.9.

					**	*
	*				**	*
NEP1:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
B4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
B3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
B1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP2:3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP2:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
B2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP1:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP1:4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP3:6	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP3:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP3:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP2:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
NEP2:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP2:3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	ACTCCTTCTT	GGGTGCCGCG
IWP2:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	ACTCCTTCTT	GGGTGCCGCG
IWP2:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	ACTCCTTCTT	GGGTGCCGCG
IWP1:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP2:4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	ACTCCTTCTT	GGGTGCCGCG
IWP2:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	ACTCCTTCTT	GGGTGCCGCG
ABC1:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP1:4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP1:3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP1:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
ABC1:4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
ABC1:3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
IWP1:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
ABC1:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
ABC1:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
D5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
C1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A18	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
SEA1:1	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A12	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A6	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A10	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A11	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A14	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
A15	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
SEA1:3	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
SEA1:4	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
SEA1:5	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG
SEA1:2	ACGTGACCAC	GTCCACTCTG	CTGGCGTGAA	CACAAAAAAA	-TCCTTCTT	GGGTGCCGCG

```

*           *           ***           *           **           ***   *
*           *           ***           *           **           ***   *
NEP1:1      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
B4          CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
B3          CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT-TCCC CC-TCAGCCC
B1          CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
NEP2:3      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
NEP2:1      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
B2          CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
NEP1:2      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
NEP1:4      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CC-TCAGCCC
NEP3:6      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CCATCAGCCC
NEP3:2      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CCATCAGCCC
NEP3:5      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CCATCAGCCC
NEP2:5      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CCATCAGCCC
NEP2:2      CCCGGTTCGG AATACAGTCC GGGG--GAG ACTCCAGCAG TCTCT--CCC CCATCAGCCC
IWP2:3      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGTCC
IWP2:2      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGTCC
IWP2:5      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGTCC
IWP1:1      TTCGGTTCGT AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
A4          CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTCCC  CCATCAGCCC
IWP2:4      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
IWP2:1      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
ABC1:5      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
IWP1:4      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
IWP1:3      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
IWP1:5      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
ABC1:4      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
ABC1:3      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
IWP1:2      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
ABC1:2      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
ABC1:1      CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
D5          CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
C1          CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
A18         CTCGGTTCGG AAAACAGTCC GGGGCGAGAG ACTCCAGCAG TCTCTTTCCC CCATCAGCCC
SEA1:1      CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTTTCCCCC CC---AGCCC
A12         CTCGGTTCGG AA-CCAGTCC GGGGCGAGAG ACTCCAGCGG TCTCTCCCCC CCATAGCCC
A6          CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
A10         CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
A11         CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
A14         CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
A15         CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
SEA1:3      CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
SEA1:4      CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
SEA1:5      CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC
SEA1:2      CTCGGTTCGG AA-CCAGTCC GGGG-GAGAG ACTCCAGCGG TCTCTCCCCC CC---AGCCC

```

		*	*	*	* * *	*	
	*	*	*	*	* * *	*	*
NEP1:1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
B4	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
B3	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
B1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP2:3	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCCC	TAGACCTGAC	CGTCTAGAGG	
NEP2:1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCCC	TAGACCTGAC	CGTCTAGAGG	
B2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP1:2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGA	
NEP1:4	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP3:6	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP3:2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP3:5	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP2:5	TCCGACTGCC	GGCCAGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
NEP2:2	TCCGACTGCC	GGCCAGCCCC	AATCCTTTGC	GCTACCGCTC	TAGACCTGAC	CGTCTAGAGG	
IWP2:3	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP2:2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP2:5	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP1:1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTATC-CTC	TAGGCCTGAC	CGTCTAGAGG	
A4	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP2:4	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP2:1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
ABC1:5	TCCGACTGCC	GGCCGGCCCC	AAACCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP1:4	TCCGACTGCC	GGCCGGCCCC	AAACCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP1:3	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	C-TCTAGAGG	
IWP1:5	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	C-TCTAGAGG	
ABC1:4	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
ABC1:3	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
IWP1:2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
ABC1:2	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
ABC1:1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
D5	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
C1	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
A18	TCCGACTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
SEA1:1	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGACCTGAC	CGTCTAGAGG	
23-7-9	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGACCTGAC	CGTCTAGAGG	
A6	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGGCCTGAC	CGTCTAGAGG	
A10	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
A11	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
A14	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGACCTGAC	CGTCTAGAGG	
A15	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
SEA1:3	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
SEA1:4	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
SEA1:5	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	
SEA1:2	TCCGATTGCC	GGCCGGCCCC	AATCCTTTGC	GCTACC-CTC	TAGTCCTGAC	CGTCTAGAGG	

```

          *          *          *          *          *          *          *          *
          *          *          *          *          *          *          *          *
NEP1:1    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
B4        GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
B3        GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
B1        GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP2:3    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP2:1    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
B2        GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP1:2    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP1:4    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP3:6    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP3:2    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP3:5    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP2:5    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
NEP2:2    GAGACCGCGG CCCGGTCCAG TA-CACGTCC CATCCTCTCT CTTCA-CCGA G-AGGGGAAG
IWP2:3    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA G-AGGG---
IWP2:2    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA G-AGGG---
IWP2:5    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA G-AGGG---
IWP1:1    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC CACCCTC--- -TTCA-CCGA GG-AGG---
A4        GAGACCGCGG CCCGGTCCAG TA-CCGCTCC CACCCTC--- -TTCA-CCGA GGGCGG---
IWP2:4    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA G-AGGG---
IWP2:1    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA G-AGGG---
ABC1:5    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
IWP1:4    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
IWP1:3    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
IWP1:5    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
ABC1:4    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
ABC1:3    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
IWP1:2    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
ABC1:2    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
ABC1:1    GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
D5        GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
C1        GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
A18       GAGACCGCGG CCCGGTCCAG TA-CCGCTCC ACTCCTC--- -TTCA-CCGA GGAGGG---
SEA1:1    GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG--AG
A12       GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT CTT-AACCGA ---GGG-AAG
A6        GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --T-AACCGA ---GGG-AAG
A10       GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCA-CCGA ---GGG-AAG
A11       GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --T-AACCGA ---GGG-AAG
A14       GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --T-AACCGA ---GGG-AAG
A15       GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG-AAG
SEA1:3    GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG-AAG
SEA1:4    GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG-AAG
SEA1:5    GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG-AAG
SEA1:2    GAGACCGCGG CCCGGTCCAG TA-CC--TCC --TCCTCTCT --TCAACCGA ---GGG-AAG

```

```

*           *
**  * *           *           *           **  ***  ****  *****  *
**  * *           *           *           **  ***  ****  *****  **

NEP1:1  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
B4      GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
B3      GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
B1      GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CC-TA
NEP2:3  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AATCCTCA--  -----CCCTA
NEP2:1  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AATCCTCA--  -----CCCTA
B2      GCGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CC-TA
NEP1:2  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
NEP1:4  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
NEP3:6  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
NEP3:2  GCGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
NEP3:5  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
NEP2:5  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
NEP2:2  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP2:3  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP2:2  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP2:5  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP1:1  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGAGTT  AAACCTCA--  -----CCCTA
A4      -ACGACTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP2:4  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP2:1  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
ABC1:5  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP1:4  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP1:3  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP1:5  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
ABC1:4  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
ABC1:3  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
IWP1:2  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
ABC1:2  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
ABC1:1  -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
D5      -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
C1      -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
A18     -ACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTGA-TT  AAACCTCA--  -----CCCTA
SEA1:1  GACGGTCTTC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CTT---G  GAACACCCTA
A12     GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CTGA--  --ACACCCTA
A6      GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CTCA--  -----CCCTA
A10     GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
A11     GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCA-A  GGACACCCTA
A14     GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
A15     GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
SEA1:3  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
SEA1:4  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
SEA1:5  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA
SEA1:2  GACGGTCTGC  CGGGTCCACG  GGTGCGCCGAC  CGGGTG--TT  ---CCTCACA  GGACACCCTA

```



```

* *
* * *
NEP1:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
B4       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
B3       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
B1       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP2:3   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP2:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
B2       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP1:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP1:4   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP3:6   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP3:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACGAC
NEP3:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
NEP2:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAGCAAC
NEP2:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAGCAAC
IWP2:3   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP2:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP2:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP1:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A4       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP2:4   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP2:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
ABC1:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP1:4   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP1:3   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP1:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
ABC1:4   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
ABC1:3   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
IWP1:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
ABC1:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
ABC1:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
D5       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
C1       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
A18      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTAAACAAC
SEA1:1   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A12      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A6       TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A10      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A11      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A14      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
A15      TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
SEA1:3   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
SEA1:4   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
SEA1:5   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC
SEA1:2   TGTCGGGTTT TATCAGTGTC AAACCCCTTG ATTGCGTCTC CTTACGGAG TCTGAACAAC

```

Figure 6.9. Alignment of the *E. recurva ITS-1* sequences. The variable and parsimony informative sites respectively are marked with a single and double asterix.

The bases in the data matrix were coded in three different ways. First, each single position of a gap was treated as an independent character (Vogler and

DeSalle 1994). Under this treatment there were 62 parsimony informative sites that recovered 60 trees (Figure 6.10) of 125 steps (CI = 0.832; RI = 0.971) excluding uninformative characters. An appropriate outgroup sequence was not available with which to root the *ITS-1* trees. Second, the Simple indel coding method (Simmons and Ochoterena 2000) was applied to the sequences where gaps were coded as separate presence/absence characters. Under this coding scheme it is assumed that contiguous gap positions originate as single indel events, and as such, each gap is treated as a single character regardless of the length of the gap. If the length of a given gap was not identical in all sequences, each length variant was coded as a single character. This treatment assumes that length variants in a given gap result from more than one mutational event. Under this treatment there were 50 parsimony informative sites and a total of 62,000 cladograms of 99 steps in length were recovered (CI = 0.869, RI = 0.972) excluding uninformative characters. The majority rule consensus tree is presented in Figure 6.11. Third, I completely eliminated the gaps from the alignment prior to the phylogenetic analysis. Under this treatment a total of eight cladograms of 40 steps in length were recovered (CI = 0.950, RI = 0.987), excluding uninformative characters. The majority rule consensus tree is presented in Figure 6.12. Confidence in resulting nodes was assessed using the bootstrap technique (Felsenstein 1985) with 1000 replicates.

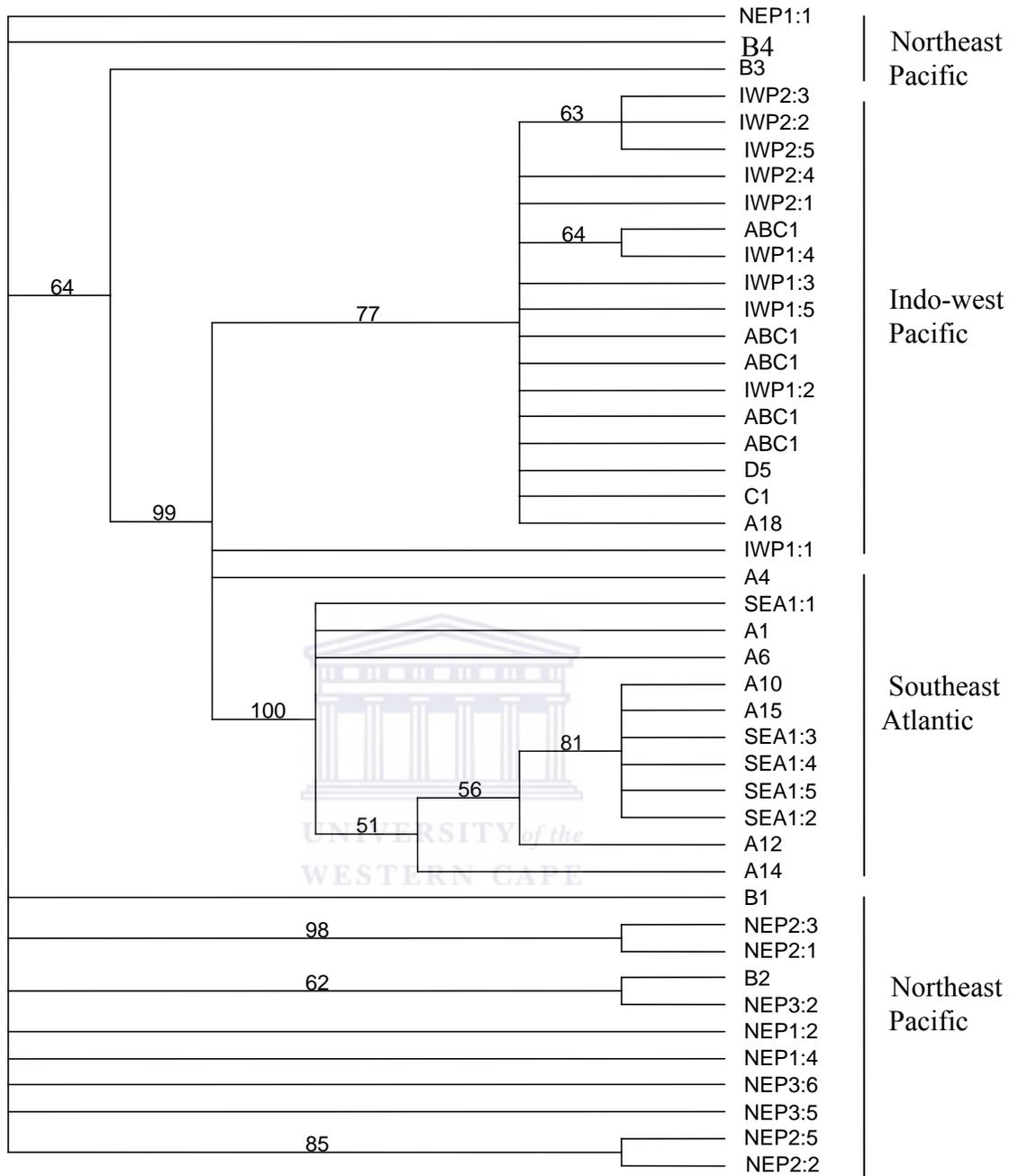


Figure 6.10. Maximum parsimony tree for the *E. recurva* ITS-1 sequences with the gaps treated as a fifth character state. Numbers on branches represent bootstrap values with 1000 replicates.

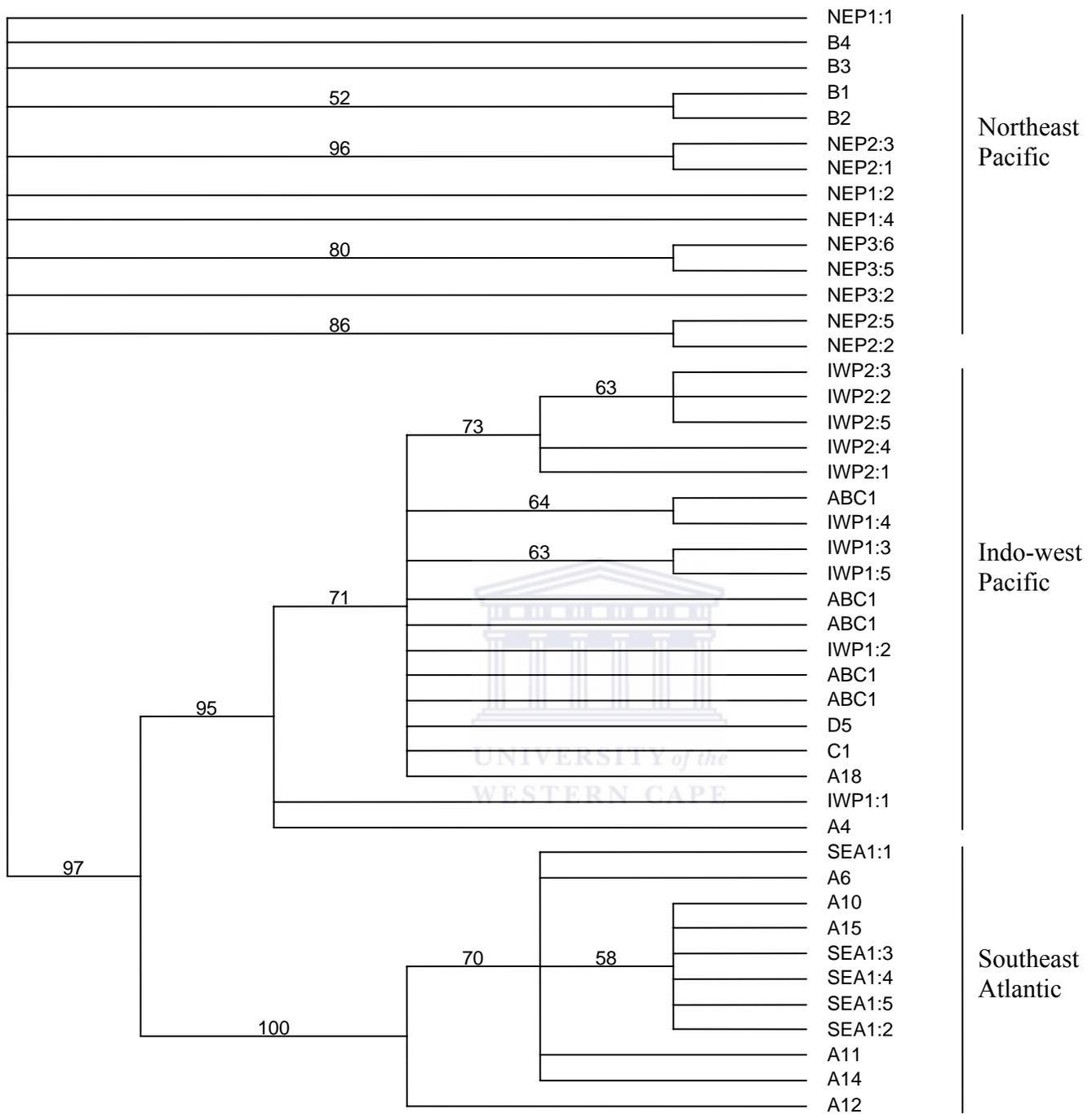


Figure 6.11 Maximum parsimony tree for the *E. recurva* ITS-1 sequences with the gaps coded using the simple gap coding method. Numbers on branches represent bootstrap values with 1000 replicates.

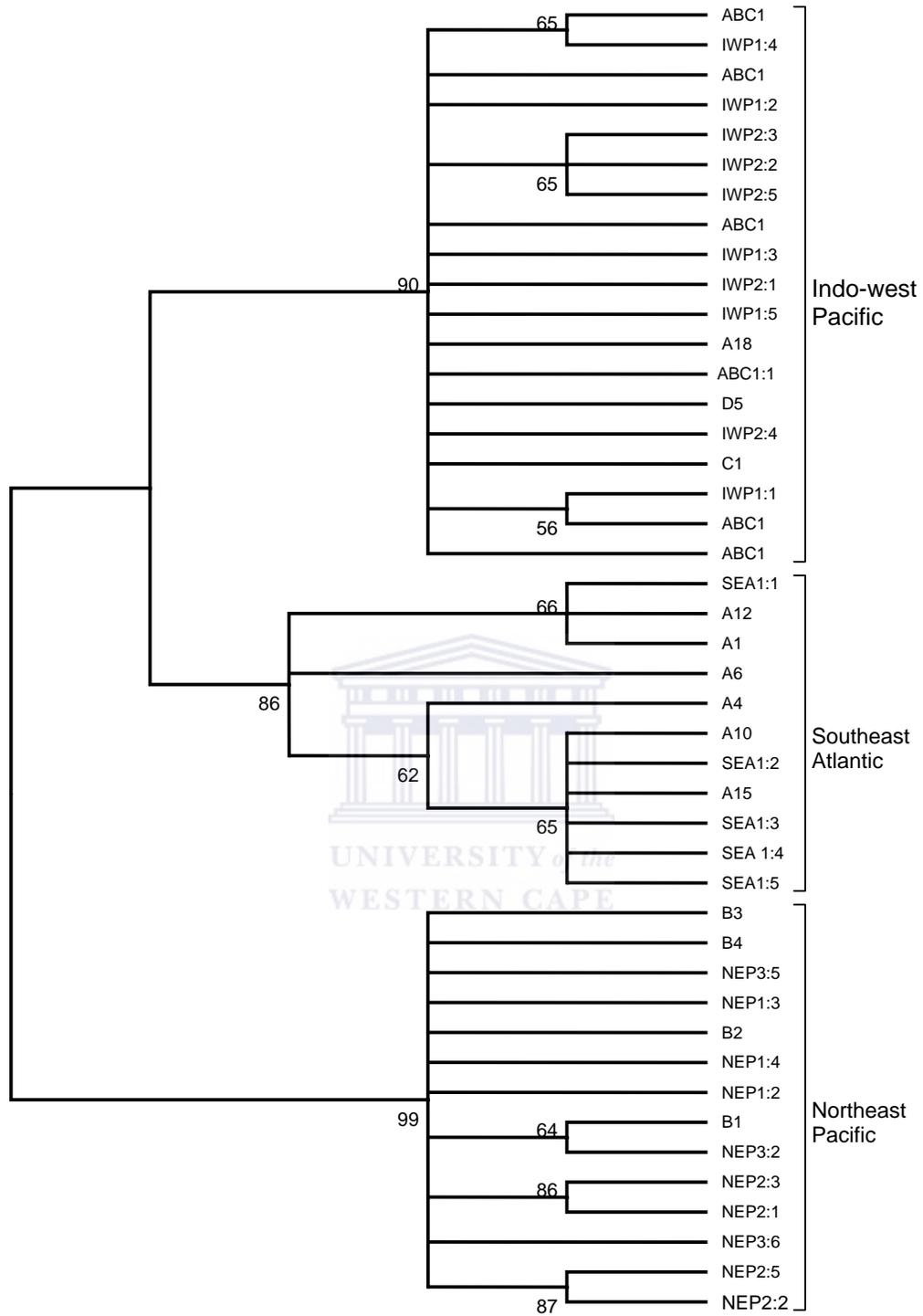


Figure 6.12 Maximum parsimony tree for the *E. recurva* ITS-1 sequences with the gaps omitted. Numbers on branches represent bootstrap values with 1000 replicates.

Phylogenetic relationships in the most parsimonious trees and in particular, the strict consensus tree derived from these trees, were entirely congruent with those observed in the trees that include the information from gaps (Figures 6.10-6.12). Thus, the inclusion of gaps as characters substantially increases the amount of phylogenetic information. Measures of homoplasy were similar under all coding methods indicating that the evolutionary dynamics of gaps is mostly similar to other positions in the data matrix (Vogler and De Salle 1994; cf. Zhang and Hewitt 2003).

For comparison, phylogenetic trees were also constructed on the *ITS-1* data set with the gaps omitted, using both the minimum evolution method (Rzhetsky and Nei 1992) and the maximum likelihood method (Felsenstein 1981). Unlike parsimony, these methods take into account the inherent characteristics of the data set (e.g. differences in substitution rates). Using the approach outlined by Heulsenbeck and Crandall (1997) alternative models of evolution were tested employing PAUP* and Modeltest (Posada and Crandall 1998). The most appropriate substitution model estimated in Modeltest using a likelihood ratio test was HKY85 ($2(\ln L_1 - \ln L_0) = 18.863$, $df=1$ $P = < 0.00001$), with 2 rate categories and a transversion:transition ratio of 2.04 with zero proportion of invariable sites. Using the Akaike information criterion (Akaike 1974), the HKY + G substitution model was selected as the most appropriate with 4 rate categories, a transition: transversion ratio of 2.08 and a gamma (G) shape parameter of 0.3191 with zero invariable sites. Confidence in resulting nodes

was assessed using the bootstrap technique (Felsenstein 1985) with 1000 replicates. The separate searches under both the maximum likelihood and minimum evolution methods incorporating these substitution models resulted in trees with scores of $-\ln L$ 786.51 and $-\ln L$ 784.29 respectively for the former method and 0.1135 and 0.1027 for the latter. The bootstrapped trees recovered using maximum likelihood and minimum evolution methods were identical to each other, and to the one recovered with maximum parsimony (not shown).

Notably, the *ITS-1* phylogenetic analysis consistently recovered three major subclades labelled I, II and III with high bootstrap support, representing respectively, the NEP, the IWP and the SEA (Figures 6.10-6.12). These same subclades are supported by all analyses, whether mitochondrial (*NDI*, *COI*) or nuclear markers (*ITS-1*) are included, regardless of gap treatment or tree construction method employed. Six fixed differences in *ITS-1* separated the NEP from the IWP and SEA populations, while three fixed differences were observed between SEA and the IWP (Table 6.13). As was observed in mtDNA data transitions and transversions accounted for almost equal numbers of unambiguous character state changes with all the tree reconstruction methods employed. Importantly, all of the cloned sequences grouped together with the sequences originally obtained by direct PCR, and all of the individuals collected in the SEA and previously identified as putative migrants based on the information from *NDI* (Figure 6.3) and *COI* (Figures 6.7 and 6.8), show congruent patterns with *ITS-1* (Figures 6.10-6.12). The average pairwise

number of nucleotide substitutions between *ITS-1* sequences within the separate clades was generally low (Table 6.13). Therefore, intragenomic variation in *ITS-1* occurs in *E. recurva*, but is not sufficient to obscure phylogenetic relationships at higher taxonomic levels and it may thus be an appropriate marker for systematic studies at the species and population levels.

6.10.1. Nuclear Sequence Variation

From the original 31 sequence *E. recurva* data set, omitting the clones, twelve distinct *ITS-1* sequence variants were identified of which eight were singletons. Five *ITS-1* variants were recovered from the IWP and SEA and two variants from the NEP and New Zealand. The SEA clade contained the highest *ITS-1* diversity and the NEP the lowest, largely concordant with the result obtained previously with the mitochondrial marker *NDI* (Table 6.7). In the SEA clade all of the nuclear sequence variants recovered were unique and this is reflected in the values obtained with the summary statistic diversity measures (Table 6.13). The frequencies of the *ITS-1* sequence variants among the three phylogeographic clades are shown in the median-joining network presented Figure 6.13. The estimated number of mutations in the shortest network was 21.

Sample	<i>n</i>	<i>S</i>	variants	<i>h</i>	π	θ_w
SEA	5	3	5	1.000 (0.016)	0.00417 (0.00090)	1.920
IWP	21	11	5	0.495 (0.129)	0.00319 (0.00143)	3.350
NEP	5	1	2	0.400 (0.231)	0.00104 (0.00062)	0.480
Total	31	16	12	0.757 (0.077)	0.01051 (0.00150)	4.756

Table 6.13. Summary statistics of *ITS-I* diversity among the *E. recurva* clades. *S* = segregating sites, *h* = haplotype diversity, π = nucleotide diversity, θ_w = Wattersons estimate of theta ($4N\mu$)

Four *ITS-I* sequence variants (ABC1:4, ABC1:2, A4 and ABC1:3) were shared between the SEA and IWP and of the twenty sequences collected in the SEA eleven grouped within the IWP clade. A single variant (C1) was shared between the three Southern Hemisphere populations. The most frequent sequence was placed internal within the IWP clade and a number of rare or singleton haplotypes were derived from this high frequency type. This variant differed by six fixed substitutions from the nearest individual in the SEA representing an average uncorrected divergence of ~ 1.5 %, and by one substitution from the nearest individual in New Zealand and the IWP representing an uncorrected divergence of 0.26% (Figure 6.13). It should be noted that, as before, all of the fixed differences between the Atlantic and Pacific populations fall away if I adopt the geographical grouping scheme (and

assume the presence of shared sequence variants to be due to ancestral retention in the SEA), as apposed to the genealogical one proposed here.

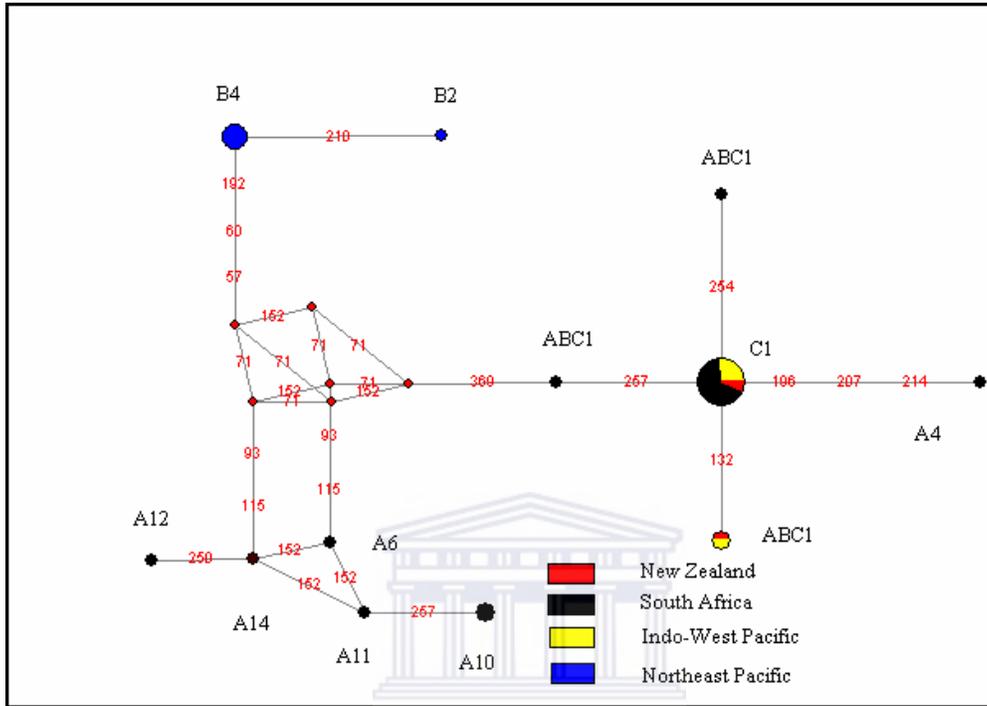


Figure 6.13. Median-joining network depicting the phylogenetic relationships among and the geographical assignment of, all *E. recurva* nuclear sequence variants (filled coloured circles) based on *ITS-I* sequences: black, South Africa, yellow, Indo-West Pacific, blue, northeast Pacific and red, New Zealand. The size of each circle is proportional to the corresponding haplotype frequency. Red diamonds indicates missing intermediates. Numbers in red on each branch indicate the nucleotide site where the substitution occurred.

Similar to the result obtained with *NDI*; only a single sequence variant (A4) differed by more than a single substitution from their nearest neighbour (Figure 6.12). The relative excess of rare alleles that were singletons is most

pronounced in the IWP and SEA populations, and least in the NEP population (Figure 6.13). However, this result is heavily influenced by migration among the three regions and the sample size differences. A branch six substitutions long connects the NEP to both the IWP and SEA clades, while the latter clades were connected by a branch four substitutions in length. A total of six missing intermediates have been inserted into the genealogy resulting from homoplasy at sites 77 and 152 forming alternate pathways represented here as loops. On the available evidence, and consistent with the results obtained previously with *NDI*, the IWP is further supported as the location of the MRCA of *Euphausia recurva*.

6.10.2. Neutrality tests

No appropriate outgroup sequence was available for the *ITS-1* sequences of *E. recurva*, precluding the use of the *F* and *D* tests (Fu and Li 1993). However, the *F** and *D** tests (Fu and Li 1993) have no such requirements and could be calculated. All four neutrality tests produced values that were negatively insignificant for the total data set, whereas Fu's (1997) *F_s* test was negatively significant for the SEA sub-sample, and Tajima's *D* (1989) and Fu and Li's *D** and *F** (1993) tests were negatively significant for the IWP. Consistent with the result obtained previously with *NDI* all of the test statistics were positive for the NEP sample (Table 6.14).

n	Total 31	SEA 5	IWP 21	NEP 5
Fu and Li's D*	-0.357 (P > 0.10)	-0.410 (P > 0.10)	-3.130 P < 0.02	0.901 (P > 0.10)
Fu and Li's F*	-0.482 (P > 0.10)	-0.557 (P > 0.10)	-3.334 P < 0.02	0.236 (P > 0.10)
Tajima's D	-0.521 (P > 0.10)	-1.094 P > 0.10	-2.138 (p < 0.05)	-0.817 (P > 0.10)
Fu's F _s	-1.722 (P > 0.10)	-3.578 P < 0.01	-1.576 P > 0.10	0.09 (P > 0.10)

Table 6.14. The neutrality test results for the *ITS-1* sequences of *E. recurva*

6.11. Effective Population Size Estimates

Taking the total sample first, it can be seen that there is a considerable range in the estimates of θ among the different estimators (Table 6.15). As was observed for mtDNA, the UPBLUE estimate of θ (Fu 1994b) for the total population (Table 6.15) is more than twice as large as the Tajima (1983) estimate. A similar situation is seen for *E. recurva* in the IWP and SEA. In all of the regional subsamples the Watterson (1975) estimate of θ (based on the number of segregating sites) was higher than the Tajima (1983) estimate (based on the average pairwise difference between sequences).

Methods	Total Sample	Southeast Atlantic	Indo-West Pacific
Tajima	4.04	1.61	1.22
Watterson	4.76	1.92	3.35
Fu's UPBLUE	9.240	2.430	3.440

Table 6.15 The theta ($\theta = 4Ne\mu$) estimates from the *ITS-1* sequences of *E. recurva*.

Estimates of effective population size calculated using the relationship $\theta = 2kNe\mu$ (where $k = 1$ for haploid and $k = 2$ for diploid individuals) require an estimate of the substitution rate μ for this marker. Estimation of the time to the most recent common ancestor under the assumption of a molecular clock further requires that the substitution rate has been calibrated with the timing of a well known geological event, such as has been done for the mitochondrial genes *16S* (Cunningham *et al.* 1992), and *COI* (Knowlton *et al.* 1998) and the closure of Panamanian Isthmus. In the absence of such information regarding *ITS-1*, estimates of effective population size can be obtained by accepting the mutation rate used previously for the mitochondrial genes (of 1.8 % per million years) which translates into a mutation rate of $\mu = 1.8 \times 10^{-8}$ per site per generation .

Although the substitution rate for *ITS-1* is unknown, the theoretical expectation that nuclear markers evolve at a slower rate relative to mitochondrial ones due

to the relatively higher effective number of nuclear alleles may be an unreasonable one in this case, considering that all of the mitochondrial coding genes (*16S*, *COI* and *NDI*) may be under far greater evolutionary constraints compared to the non-coding nuclear *ITS-1* fragment. The rate of mtDNA substitution in non-mammalian animals is not generally higher than single-copy nuclear DNA (Zhang and Hewitt 2003). Whether this also applies to multiple copy nuclear markers is uncertain, however, a high degree of *ITS-1* diversity among crustaceans above the genus level has previously been reported (Chu *et al.* 2001) which could imply a rapid rate of substitution in this marker. Thus, assuming equivalence of rates may not bias the estimates of effective population size too severely.

These caveats noted, for the entire *E. recurva* sample the Tajima (1983) and Watterson (1975) estimators suggest that $N_e = 586,000$ and $695,000$ individuals respectively. UPBLUE suggests even higher values with N_e estimated as $1,342,000$. The effective population size estimates also concur with the estimates obtained earlier from the mitochondrial genes (Table 6.9).

Methods	Total Sample	Southeast Atlantic	Indo-West Pacific
Tajima	0.58	0.23	0.18
Watterson	0.69	0.28	0.48
Fu's UPBLUE	1.340	0.350	0.500

^c Wattersons with singleton mutations omitted

Table 6.16. The effective population size (N_e) estimates (in millions) from the *ITS-1* sequences of *E. recurva*. S = segregating sites π = average pairwise differences.

Combining the entire sample of sequences from the various populations ignores the obvious subdivision in this species, and this type of population differentiation is expected to lead to an overestimation of effective population size (Wright 1943). However, as congruent phylogenetic patterns were observed with mitochondrial and nuclear markers, estimates of θ from the different genetic systems will be equally affected by the geographical subdivision and should therefore, produce comparable estimates for this population parameter.

The effective population size estimates calculated using the same mutation rate applied to the mtDNA for the nuclear marker *ITS-1* were consistently smaller than those based on the *NDI* mtDNA. This is not surprising given that the nuclear marker data set comprised a subset of the individuals that made up the

NDI data set and higher estimates of θ would be expected if a larger sample size had been employed with *ITS-1*. This would also be expected if the mutation rate had been overestimated. Using a mutation rate half of that employed for the mtDNA the estimates of N_e for the nDNA and mtDNA markers are more similar. The effective population size estimates for the entire sample are larger than the sum of the SEA and IWP subpopulations as the former data set also included the NEP sequences.

6.12. Comparison of the different markers

Combined analysis of mitochondrial and nuclear data for the euphausiid species investigated here revealed qualitative and quantitative differences in the patterns of concordance among the different species data sets. Although equivalent patterns of geographical subdivision were recorded in *E. recurva* with all of the different markers employed, quantitative differences between regional populations were observed. The highest divergences (π_{net}) observed with *ITS-1* were between populations from the NEP and both the IWP and the SEA, with six fixed differences (1.56%) separating these populations. The lowest at 0.78% was between the IWP and SEA (Table 6.17). In contrast, with both *NDI* and *COI* data sets, the divergence between the NEP and IWP populations was relatively small (0.64 and 0.63%), and the largest differences were instead observed between the NEP and the SEA (4.4 and 5.3%), followed by IWP and SEA populations (3.2 and 5.31%).

clades	ITS-1 (%)	<i>f</i>	ND1 (%)	<i>f</i>	CO1 (%)	<i>f</i>
IWP-NEP	1.6-2.6 (1.7)	1.56 (6)	1.9-5.1 (2.1)	0.64 (1)	0.93-2.2 (0.9)	0.63 (4)
IWP-SEA	1.3-2.4 (1.3)	1.04 (4)	6.4-11.5 (7.9)	3.20 (5)	6.5-7.5 (5.8)	5.31 (34)
SEA-NEP	1.6-2.4 (1.8)	1.56 (6)	6.4-8.9 (6.7)	4.40 (7)	6.7-7.2 (6.3)	5.31 (34)

Table 6.17. Genetic differentiation of regional lineages of *E. recurva* based on 640 base pairs of *COI*, 156 base pairs of *NDI* and 384 base pairs of *ITS-1*. The pairwise difference between clades is given in brackets. Also given are the number of fixed differences *f* (brackets) and the percentage of fixed differences between clades

6.13. Interspecific *ITS-1* Comparisons

The *ITS-1* sequences from the different euphausiid species were aligned using MALIGN version 1.87 (Wheeler and Gladstein 1993) which optimizes the alignment according to parsimony criteria. A gap-to-change of 5:10 was chosen. The cost of extra gaps was set to 9. These parameters were chosen because they resulted in a meaningful alignment of the most distant species and the closely related ones. The alignment reveals the conspicuous presence of several indels separating the species pair *E. lucens*-*E. vallentini* from *E. recurva*. These were largely concentrated around the central bases and both the 5' and 3' prime ends were highly conserved among species (Figure 6.14). The nuclear *ITS-1* sequences in *E. recurva* (at 404 base pairs in length) were the shortest among the three species (Figure 6.14).



UNIVERSITY *of the*
WESTERN CAPE

CHAPTER 7

DISCUSSION *Euphausia recurva*

7.1. DNA Sequence Variation, Diversity and Phylogenetics

Although all of the earliest of euphausiid studies were conducted using allozymes, no species have previously been studied using nuclear markers that produce information in the form of DNA sequence data. Of the three euphausiid species investigated here molecular variation was highest in *E. recurva* with all of the molecular markers employed although the within-population variation was still relatively low (h and π and θ) in this species. As was observed for *E. lucens* (Chapter 3), most of the mitochondrial variation in *NDI* and *COI* within *E. recurva* occurred primarily at silent, third codon position sites and involved synonymous mutations and this excess was observed in comparisons with *E. lucens* and *E. vallentini*. As a result the M-K test (McDonald-Kreitman 1991) applied to the *E. lucens* and the *E. recurva* *NDI* data was not significant. Furthermore, with all the mitochondrial markers the amino acid variants within *E. recurva* involved functionally conservative amino acid substitutions and occurred as singletons and therefore, likely have little effect on the observed population structure. This pattern has been observed in a wide range of animal groups in both terrestrial and marine environments (Chapter 5), and was taken as strong evidence for the dominance

of negative selection in protein evolution in animals (Nachman *et al.* 1996; Ballard and Kreitman 1994; Rand and Kahn 1996; and Rand 2001).

7.1.1. Mitochondrial Variation

Nucleotide and haplotype diversity in *NDI* were highest in *E. recurva* in the SE Atlantic followed by the IW Pacific and the NE Pacific (Table 6.2) However, both measures of diversity are inflated in the SEA by the presence of highly divergent sequences that have a closer phylogenetic affinity with other *NDI* sequences collected in the IWP. When a genealogical grouping scheme is applied where populations are defined as monophyletic clades in the phylogenetic trees, the highest nucleotide and haplotype diversity is displayed in IWP clade (Table 6.7). In *E. recurva* an excess of rare *NDI* alleles with respect to mutation-drift equilibrium was observed in both Southern Hemisphere clades and consequently both Tajima's (1989) and Fu's (1997) neutrality tests were negatively significant for these clades (Table 6.6).

7.1.2. Nuclear Variation

Nucleotide and haplotype diversity in *ITS-1* were highest in the SE Atlantic clade (Table 6.13) although it can be seen that this was due to the different frequencies of sequence variants in the respective clades as all of these diversity statistics are affected by the frequencies of alleles in the sample. For example, in the IW Pacific the most frequent sequence variant was carried by 71% of the individuals in that clade whereas in the SEA this was 33% (Figure

6.13). It has been shown that when a few individuals are sampled from many different populations and pooled artifactual evidence of an excess of rare alleles can be produced (Hammer *et al.* 2003). However, as an excess of rare *ITS-1* alleles was displayed in both the Southern Hemisphere samples, as well as the combined data set, this excess is unlikely to be an artifact of the sampling scheme. In contrast to the results for mitochondrial data, significant values were obtained for the nuclear data with all the neutrality tests for the IWP clade only. The failure of these tests to detect a significant departure from neutrality for the *ITS-1* SEA data may be due to a lack of power due to the small sample sizes used (Table 6.14). For example, although Fu's F_s test has been shown to be the most powerful statistic for detecting departure from neutrality under a wide range of conditions, it suffers from a lack of power when samples sizes are small (Ramos-Osnins and Rozas 2002). Although a non-significant result was obtained for the SEA *ITS-1* data set, most of the tests produced negative values perhaps suggestive of a common cause. It is possible that direct purifying selection may be responsible for the excess of rare alleles revealed in the both the mitochondrial and nuclear *E. recurva* genes however, other explanations such as genetic hitchhiking, background selection, sweepstakes recruitment or demographic expansion (perhaps) after a population bottleneck are also plausible alternatives.

7.2. Within-individual Variation in the Nuclear Marker *ITS-1*

Cloning experiments undertaken for *ITS-1* showed that within-individual genetic variation was negligible and with the exception of one highly divergent clone, the within-individual sequence variation in the 25 clones differed by 1-8 substitutions; a level of divergence slightly high for this molecular marker in animals (Coleman and Vaquier 2002). It is possible that some of the observed variation is the result of mistakes introduced by the *Taq* polymerase during amplification. However, I regard this unlikely as the amplification appears to be accurate for the more conserved 158 base pairs of *18S* rRNA coding region that did not differ in any of the clones analysed (Figure 6.9). Even if errors introduced by the *Taq* polymerase during amplification were responsible for some of this variation, the expectation is that it will be randomly distributed across nucleotide sites and that it would have little effect on the population structure inferred from the fixed differences. Taken together, it appears that concerted evolution (Hillis *et al.* 1991) is sufficient to maintain all but a few *ITS-1* positions in *E. recurva*. Despite exhibiting moderate levels of within individual sequence variation the *E. recurva ITS-1* sequences still exhibited phylogenetic separation coinciding with geographic separation of populations. The pattern where a single genome contains several divergent paralogs is thought to occur whenever concerted evolution is slower than speciation (Vogler and De Salle 1994).

7.3. Genetic Variation through Space and Time

Three distinct *E. recurva* subclades were supported by all the phylogenetic analyses, whether mitochondrial (*ND1*, *COI*), or nuclear markers (*ITS-1*) were included, and regardless of the gap treatment or tree construction method employed. These clades correspond to geographically distinct populations in the IWP, the NEP and the SEA. Reciprocal monophyly was not observed with any of the markers, with individuals collected in the SEA that grouped within the IWP clade in all the phylogenetic trees implying migration of individuals from the IWP to the SEA or retention of ancestral polymorphism in the SEA.

The presence of highly divergent mtDNA haplotypes that do not appear concordant with morphology or geography have been observed in other zooplankton, including the euphausiid *Meganyctiphanes norvegica* (Papetti *et al.* 2005), the calanoid copepods species *Calanus helgolandicus* and *C. euxinus* (Papadopoulos *et al.* 2005), and the chaetognath *Sagitta setosa* (Peijnenburg *et al.* 2006). In all three studies, the authors concluded that the highly divergent mtDNA haplotypes likely represent ancestral haplotypes that have been retained in a particular population. However, they could not rule out the possibility that the large divergences between individuals within a single sample represented haplotypes that had immigrated from undiscovered populations that were perhaps separated in the distant past. In contrast, the potential source population (IWP) for the highly divergent *E. recurva* sequence variants recovered in the SEA has been sampled in this species.

Although gene trees and species trees are not always the same (Donnelly and Tavaré 1995; Edwards and Beerli 2000; Knowles 2004), the congruent patterns of geographical subdivision revealed in the phylogenetic analyses for the independent loci with high bootstrap support, suggests the inferences of population divergence history are reliable. It is notable in this respect that sequences of *NDI*, *COI* and *ITS-1* amplified from the same SEA specimens also showed consistent phylogeographic patterns of affinity, grouping with sequences collected in the IWP rather than with those from the same region.

The low within-population variation was reflected in the shallow depth of the branches in each of the regional clades in the median-joining networks (Figures 6.2 and 6.13). This implies either that the mutation rate for these molecular markers is relatively slow, or as was indicated previously for *E. lucens* and *E. vallentini*, that the observed polymorphisms are relatively recent. If the latter were true, this would suggest that the turnover of genetic variation in local *E. recurva* populations might occur at a relatively rapid rate on an evolutionary time scale given the relatively large divergences between these populations and between *E. recurva* and the other euphausiid species

Shallow within-population mtDNA phylogenies and an excess of rare alleles may be a general feature of populations of marine organisms and similar results have been reported for several other euphausiids with *NDI* and *COI* (including the Antarctic species *E. superba* (Zane *et al.* 1998) and *E. crystallophorias*,

(Jarman *et al.* 2002), the North Atlantic species *Meganyctiphanes norvegica* (Papetti *et al.* 2005), as well as in a range of other marine organisms with a wide selection of mitochondrial markers (Palumbi 1997; Shields and Gust 1995; Grant and Bowen 1998; Arnason 2000; Cassens *et al.* 2003; Papadopoulos *et al.* 2005; Tolley *et al.* 2005). Similar patterns have also been observed in a wide range of terrestrial organisms and Pleistocene expansions have been postulated to explain the genealogical and phylogenetic patterns (Peilou 1999; Hewitt 2000; Lessa *et al.* 2003; Hugall *et al.* 2002). However, recovery from a recent population bottleneck (Tajima 1989a,b), or selection against linked deleterious mutation can also lead to an excess of rare alleles when effective population sizes are small (*e.g.*, Charlesworth *et al.* 1993). More recently, Fay and Wu (2000) suggested that an excess of high-frequency-derived alleles in a sample is more likely due to hitchhiking than to other scenarios. However, they also pointed out that if there are many fixed differences between populations that exchange rare migrants, polymorphisms in the population would tend to be at very low or high frequencies. A selective sweep will also result in shallow genetic separations within species while maintaining relatively large divergences between sister species (Nielsen 2005), and cannot be ruled out.

There are several lines of evidence that suggest that the most likely location of the MRCA is in the IWP. First, this region displayed the highest diversity with both mitochondrial and nuclear and molecular markers and proportionally the

highest number of connections to other *NDI* haplotypes (Figure 6.3). Second, the IWP clade also had the highest number of occurrences of *NDI* and *ITS-1* sequence variants in different populations and third, the branch joining the outgroup species (*E. lucens*) to the *E. recurva* ingroup sequences in the *NDI* and *COI* trees connected via a branch in the IWP clade.

7.4. Coalescent Estimate of *E. recurva* Population Divergence

The coalescent-based estimate for *NDI* calculated with MDIV (Nielsen and Wakeley 2001) indicated Pleistocene divergences of populations from the NEP and IWP (0.67 MYA) and NEP and SEA (0.71 MYA) while the divergence of the SEA and IWP (2.65 MYA) dates back to the late-Pliocene with the most common ancestor suggested to have occurred around 3.21 MYA (Table 6.11). The estimate of the TMRCA of the SEA and IWP populations is almost identical to the estimated *16S* rRNA divergence time for *E. recurva* and *E. krohni* (Figure 4.3), perhaps suggesting that the split between the Southern Hemisphere populations was contemporaneous with the speciation event between these species. Similar divergence times were indicated for the *E. lucens*-*E. vallentini* species pair with *16S* rRNA (Figure 4.4) and *NDI*. Testing for simultaneous vicariance across comparative phylogeographic data sets is a notoriously difficult problem that requires information from multiple nuclear loci and substantial computer simulations under a range of conditions (Hickerson *et al.* 2006). The difference between the TMRCA and PDT estimates for the comparison of Southern Hemisphere *E. recurva* populations

was relatively small (17%), consistent with a relatively ancient separation (Arbogast *et al.* 2002; Edwards and Beerli 2000).

Although the more recent divergences indicated for the Pacific populations in the different Hemispheres may be reasonable, the split between the NEP and SEA comparison should be accepted with some caution as the 67% of the variation in the PDT is taken up by the ancestral population (Table 6.11). The MDIV method assumes a demographic model with the two populations of equal size (Nielsen and Wakeley 2001) however, all the estimators of θ indicated much smaller effective sizes for the Northern Hemisphere population (Table 6.8). This might explain the large difference between estimates of the PDT and the TMRCA in comparisons involving this population. Unequal effective population sizes would likely affect a reliable estimation of divergence times by underestimating the actual dates (Nielsen and Wakeley 2001). In contrast, because the MDIV method uses the entire data set to calculate the likelihood of the underlying genealogies, the effect of demographic expansion (i.e. the concentration of coalescent intervals in a short time frame and the excess of rare alleles) is incorporated in the model. For this reason the inclusion of the SEA and IWP populations, which show a significant excess of rare alleles, should not affect a reliable estimation of divergence times.

7.4.1. Pairwise Differences

Estimates of divergence based on the average pairwise differences are known to have a large variance (Tajima 1983) and to be sensitive to the site frequency distribution of polymorphic sites (Tajima 1989). The date of divergence estimated this way is simply the estimated average time of the ancestral sequence for pairs of sampled sequences and takes no account of the variation between populations that is caused by variation in the ancestral population. This method can be extended to make a correction for ancestral portion of the polymorphism using an estimate from the descendent species. This measure is called the net nucleotide diversity and is unbiased if all the populations are of the same size ($N_1 = N_2 = N_A$) and $t = 2ut$ (Nei and Li 1979; Takahata 1986). Of all the methods of dating divergence the net nucleotide diversity always gave the highest estimates, which because of the large variances associated with this method, should be viewed as upper limits.

7.4.2. Estimates of Migration Rates

Negligible migration was indicated with MDIV for all comparisons except between the two Southern Hemisphere populations (Table 6.11). The effective migration rate ($2Nem$) was 0.381 implying that the Southern Hemisphere populations of *Euphausia recurva* have been isolated with moderate levels of gene flow since they began to diverge (Nielsen and Wakeley 2001). This method has been shown to be very powerful in discriminating between

migration rates of $M = 0$ and $M = 1$ even with a single locus (Nielsen and Wakeley 2001). Although models of gene flow between populations vary in their assumptions, it has been shown that practically any migration, however small, will be sufficient to preclude genetic divergence (Maynard-Smith 1989; Slatkin 1985). Although the direction of migration is not given with MDIV, the presence of very divergent sequences in samples collected in the SEA implies migration from the IWP to the SEA but not in the opposite direction.

7.4.3. Spatial and Temporal Structure around Southern Africa

As observed previously for *E. lucens*, no further subdivision of *E. recurva* *NDI* haplotypes was detected at the meso-scale in the SE Atlantic waters around South Africa, and no temporal differences were observed in the haplotype composition among samples collected during 2001 and 2002. Similarly, no differences were observed in this mitochondrial marker among *E. recurva* samples collected 23 years apart in the same area of the NE Pacific, suggesting that the genetic variation has been stable during the sampling periods in these regional populations. Similar temporal comparisons could not be made with the nuclear marker *ITS-1* as only a subset of the specimens investigated with *NDI* were sequenced with *ITS-1*.

7.5. Congruent Geographical Distributions

Congruent geographical patterns of genetic differentiation to those observed within *E. recurva* have been reported in another holoplanktic crustacean group

(Goetze 2003). In an investigation of eight circumglobally distributed copepod species, three (*Paraeucalanus sewelli*, *Subeucalanus pilattu* and *Rhincalanus nasutus*) showed fixed differences between Pacific and Atlantic populations at both *16S* ribosomal RNA and the nuclear marker *ITS-2* suggesting a total absence of gene flow between the different ocean basins, while populations of another three (*Eucalanus hyalinus* 1, *E. hyalinus* 2 and *Paraeucalanus* spp) appear to have shared alleles between oceans in the recent past (Goetze 2003). Divergence levels for the former group of species ranged between 2.4 and 10% for *16S* ribosomal RNA, and 0.2 and 0.6% for *ITS-2* (Goetze 2003).

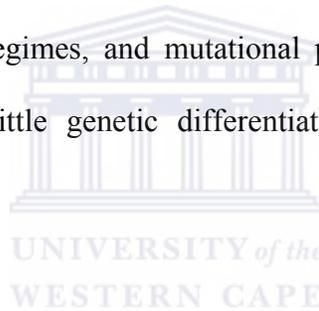
Similarly, studies that employed allozymes and micro-satellites have demonstrated that western Australian populations of the shrimp *Paeneas monodon* were significantly different from other Australian populations (Benzie 1999, 2000), and differed markedly from Indian Ocean populations collected from Southern Africa and Madagascar. Interestingly, the phylogenetic results suggested separation and subsequent divergence of the African and southeast Asian/Australian populations followed by a reinvasion of the Indonesian region by a group of African or West Indian origin. Studies of the coconut crab *Birgus latro* (Lavery *et al.* 1995, 1996) that employed both nuclear and mitochondrial markers have also revealed differences between Indian and Pacific Ocean species.

Qualitatively and quantitatively congruent patterns of differentiation similar to those displayed within *E. recurva* have also been observed between Atlantic and Pacific populations of the sea cucumber species *Holothuria nobilis* (Ulthicke and Benzie 2003). For example, the level of *COI* differentiation between samples of *H. nobilis* collected in the Reunion Islands and Australia (7.7-8.9%) is very similar to the values reported for the same molecular marker for the *E. recurva* clades in this study (5%). The maximum divergence for this gene among other *H. nobilis* samples collected throughout Australasia was low (0.75%) (Ulthicke and Benzie 2003).

7.6. Quantitative and Qualitative Differences among Markers

Although equivalent patterns of geographical subdivision were recorded in *E. recurva* with both the nuclear and cytoplasmic markers, quantitative differences between regional clades were observed. The highest divergences (π_{net}) observed with *ITS-1* were between the NEP and both the IWP and SEA clades with six fixed differences (1.56%) separating these populations while the lowest was between the IWP and SEA (0.78%) (Table 6.17). In contrast with the mitochondrial *ND1* and *COI* data sets, the divergence between NEP and IWP clades was relatively small (0.64 and 0.63%) while the largest separation was between the NEP and the SEA (4.4 and 5.3%) followed by IWP and SEA clades (3.2 and 5.31%).

Several explanations have been proposed for large differences in divergence levels between nuclear and mitochondrial markers including purifying selection of mtDNA, sex biased dispersal, smaller effective population size of mitochondria and variable reproductive success. It has also been suggested that a higher rate of mitochondrial *versus* nuclear evolution in species may be inherent in species with large populations due to diminished effects of genetic drift and consequent enhancement of the importance of selection (Peijnenburg *et al.* 2006). Therefore, the quantitative differences observed between the nuclear and cytoplasmic markers within the same species may simply reflect the random sampling variance in levels of polymorphism among different loci under different selective regimes, and mutational processes resulting in one genetic system showing little genetic differentiation while another shows extreme differentiation.



In practice, although levels of differentiation between populations and species with nuclear and mitochondrial molecular markers are often highly correlated, exceptions to this pattern are not uncommon. In a recent study of copepod populations Goetze (2003) revealed similar discrepancies between the levels of quantitative differentiation at nuclear and mitochondrial markers that were often species-specific. For example, this author found differences between lineages of the copepod species *Rhincalanus rostrifons* in the E and W Pacific of 6.7-7.6% at *16S* mt rRNA but no differences in the nuclear marker *ITS-2*,

whilst, in another copepod species *Pareucalanus sewelli*, variation was low in *16S* mtRNA and relatively high in *ITS-2*.

It has already been established that strong purifying selection may have played a major role in the evolution of these mitochondrial genes in the three euphausiid species investigated here. Although, many gaps exist in our understanding of the ecology and population dynamics of krill, to the author's knowledge, no evidence has ever been published to indicate that sex-specific differences in dispersal for any euphausiid species: strongly skewed sex ratios have been observed within populations of several euphausiid species (Mauchline and Fisher 1969).

If in the past an obstruction to dispersal between Hemispheres existed in the tropical Pacific that had more recently been breached by *E. recurva*, there might be a greater degree of structure at some nuclear markers than mtDNA. If so, the observed differences in divergence levels between nuclear and mitochondrial markers in *E. recurva* may simply reflect the fact that mitochondrial markers reach equilibrium between drift and migration faster than nuclear markers due to the lower effective population size of the former marker (Neigel 1994). In this way, information in the nuclear and the mitochondrial markers may be out of phase with one another reflecting ancient and recent population history respectively.

This explanation has been invoked to explain the results obtained for the sea urchin species *Linckia laevigata* (Williams and Benzie 1998; Williams *et al.* 2002) and the snapping shrimp species *Alpheus lottini* (Williams *et al.* 2002). In *Linckia laevigata* a distinct genetic break between the Indian Ocean and the Indo-West Pacific was indicated with allozymes but not detected with the mitochondrial markers (Williams and Benzie 1998; Williams *et al.* 2002). Similarly, in *Alpheus lottini* nuclear variation was shown to have clinal structure in the IWP, whereas the mitochondrial *COI* gene indicated high gene flow within this region (Williams *et al.* 2002).

The quantitative differences displayed by nuclear and mitochondrial markers would only be expected if the rates of gene flow across the equatorial Pacific region were sufficiently low, as the time required to reach mutation-drift equilibrium is less when gene flow is high (Williams *et al.* 2002). However, even though the mtDNA divergences between the Pacific populations were low (Table 6.17), complete lineage sorting of mtDNA haplotypes was observed, implying that present-day gene flow between Hemispheres may not be occurring. This pattern could be explained if the barrier formed by the tropical waters of the central Pacific was permeable to migrants episodically, and that migration between Hemispheres has been achieved relatively recently, but is perhaps not currently ongoing. If, as suspected, the long-distance dispersal of adults and (particularly) larvae involves passive transport in the prevailing currents (Mauchline and Fisher 1969) then the timing of migratory events

across such long-established barriers as the tropical equatorial waters may be highly stochastic; occurring only when relaxation of prevailing water currents permits.

It has been pointed out (Bowen and Grant 1997) that the tropical zone is fairly narrow in the eastern Pacific, along the west coast of Mexico, suggesting that genetic contact between the California and Chile marine populations is or has been possible. In the western Pacific, the tropical zone is much broader, making genetic exchange between Japan and Australia unlikely. A similar conclusion was drawn by Goetze (2003) in a study of the in the copepod *Rhincalanus nasutus* species complex who observed sister lineages in the California-Peru boundary currents. During the last glacial maximum a marked cooling has been recorded in the equatorial regions of the Pacific and the Atlantic (CLIMAP 1976). This feature may have had as important an impact on climate as the equatorial displacement of the Polar Front. Apparently increased upwelling along the Pacific equatorial divergence produced temperatures that were as much as 6 °C cooler than today's. The westward limit of this effect is unknown but the cooler waters did not reach the west Pacific where glacial temperatures are similar to today's (CLIMAP 1976).

The large quantitative differences in the divergence levels between nuclear and mitochondrial markers for the NEP-IWP split might reflect the action of a rapid sweep of a haplotype linked to a selected site somewhere else on the

mitochondrial genome. However, while a selective sweep will reduce within-population variation it is not expected to reduce divergences between populations unless the same sweep had spread to both Pacific regions.

The same arguments applied previously against the likelihood of sweepstakes recruitment being responsible for the localized shifts in allele frequencies observed in *E. lucens* (Chapter 5), apply equally for *E. recurva* and will not be repeated in full here. No evidence of reduced mtDNA genetic variation was observed in adults of *E. recurva* in the 2001 samples relative to the 2000. Although the SEA data are insufficient to definitively rule out the possibility of sweepstakes recruitment, the temporal stability among samples collected in the NEP suggests that sweepstakes reproduction is unlikely to be the factor responsible for the large differences in divergence levels among regional observed between nuclear and mitochondrial markers. Therefore, in *E. recurva* a sweepstakes reproductive success model cannot be unambiguously ruled out to explain the quantitative differences between the nuclear and mitochondrial markers with the data presented here, as the temporal stability in genetic turnover observed in successive years in the SEA may be expected over such short time scales under both historical and contemporary hypotheses.

7.7. Appropriate Model of Population Divergence

Although it is possible that purifying selection may be responsible for the excess of rare alleles observed in the both the mitochondrial and nuclear genes

in *E. recurva* alternative explanations such as genetic hitchhiking, background selection or demographic expansion are equally plausible. As with *E. lucens*, the evidence against the presence of these two types of selection were that the *NDI* locus appears to behave normally in terms of the direction as well as the spatial distribution of mutations. The reduction in diversity expected by background selection was not apparent with particularly high values observed in the IWP clade in particular (Table 6.8). Furthermore, as both the nuclear and mitochondrial data sets show congruent qualitative patterns of geographic subdivision, it seems unlikely that either selection has acted identically on the mtDNA and nuclear locus to produce this geographic pattern, or that selection has acted strongly enough on the mtDNA locus alone to produce a selective sweep, without at the same time altering the geographic pattern of variation compared to the other independent genetic marker (Lavery *et al.* 1996). It is possible that a selective mtDNA sweep occurred long enough ago that the geographic pattern may be approaching a new equilibrium that reflects the same isolation-by-distance pattern seen in *ITS-1*.

The estimates of θ that take genealogical information into account were higher than those derived from non-genealogical summary statistics for both Southern Hemisphere clades but not the NE Pacific one. This could be interpreted to reflect recent demographic expansion in the former, but not the latter, population (Crandall *et al.* 1999; Villa' *et al.* 1999; Su *et al.* 2001). Consistent with this scenario, the maximum likelihood estimates for the growth parameter

under an exponential model, expressed, as $1/u$ is positive for both southern hemisphere clades. This might explain why the neutrality test of Fu (1997) was significant for the mtDNA data while those of Fu and Li (1993) were not, as simulation studies have shown the former test has the greatest power for detecting growth in a wide variety of cases (Fu 1997; Ramos-Onsins and Rozas 2002). Consistent with this idea, the M-K test (McDonald-Kreitman 1991), which has increased robustness with regards to the effect of demographic assumptions compared to the single lineage tests applied previously was not significant.

7.8. Estimation of Population sizes

For *NDI* all of the estimators of θ with the exception of the Watterson^c (1975) and Fu (1994b) UPBLUE estimate indicated the largest effective population size in the IWP and smallest in the NEP. These estimates ranged between 0.32 and 17.2 and 0.45 and 2.11 million individuals in the IWP and SEA respectively. Larger population sizes in the IWP were also indicated with the nuclear marker *ITS-1* with all of the methods except that of Tajima (1983). Although census size estimates are not available for these species, these effective population sizes implied here are likely several orders of magnitude smaller than potential census sizes that can be attained by these species.

7.8.1. Potential Vectors for Gene Exchange

If migration from the IWP is responsible for the presence of very divergent *NDI*, *COI* and *ITS-1* sequences in the SEA, two puzzles remain; why migration between these populations might have occurred in only one direction given that contemporary surface current patterns suggest transport in both directions should be theoretically possible, and why the putative migrants from the Pacific fail to achieve high frequencies in the SEA if reproduction and recruitment are random?

Currently, there is a continuous route for gene exchange for individuals of *E. recurva* between the Indo-west Pacific and the east Coast of Africa in both directions in the wind-driven circulation of the North Indian Ocean. This current system is forced by monsoonal winds and reverses radically with season (Lutjeharms 2006). Individuals from the IWP entrained in this current system during the Northeast Monsoon season, would be transported southward in the Somali Current and could easily be mixed into adjacent coastal waters along the east African Coast by the physical mixing action at the edges of the gyral current stream. Transport to the west coast of South Africa would then be achieved via the southward flowing Agulhas Current (Lutjeharms 2006). A similar dispersal route was implied by the presence of the Indo-Pacific chaetognath species *Sagitta bedoti* off the coast of Angola in the northern

Benguela region (Neto 1961) and the Pacific species *Sagitta pacifica* and *Sagitta robusta* in the southern Benguela (Duró and Gili 1996).

Transport from the Atlantic to the IWP should also theoretically possible via the eastward flowing surface waters of Agulhas Return Current and the South Indian Ocean Current (Lutjeharms 2006). However, the average path of these currents lies at latitude 40° S, which coincides with the southern limit of the distributional range of *E. recurva* (Tattershall 1924). The surface water temperature in the Agulhas Return Current and the South Indian Ocean Current averages between 20-25°C annually and are likely a limiting factor to migration for individuals of *E. recurva* (Lutjeharms 2006).

The far larger sample sizes employed in the SEA compared to the IWP bias the potential for detecting such rare migratory events into the former region reducing the reliability of any inferences concerning the direction of migration between the southern hemisphere populations. It is possible that by taking a larger sample size in the IWP migration between Southern Hemisphere populations may have been detected in both directions and this scenario cannot therefore, be ruled out with the present data.

7.8.2. Low Recruitment in a High Dispersal Species?

Now to address the second puzzle; why the putative migrants from the Pacific fail to achieve high frequencies in the SE Atlantic if reproduction and

recruitment are random? It is possible that the migrant haplotypes from the Pacific fail to develop high frequencies in the Atlantic because they are at a selective disadvantage in a locally adapted population that could involve either, timing or location of reproduction in different environments. It has been suggested that since environmentally coordinated reproduction is common in marine plankton (Bijma *et al.* 1990), genetic switches in the timing of reproduction could explain such findings (Norris 2000). Seasonal changes in vertical structure of the upper ocean occurs in all open water masses and the differences between the Atlantic and Pacific Oceans may provide opportunities for population divergence in *E. recurva* and eventually speciation, through the timing of reproduction and the production of diverse habitats over the course of a year. In this way eco-ethological differences might have developed between the separate Oceans that limit the ability of the migrants from the Pacific to successfully reproduce in SE Atlantic waters. Population divergence can occur along environmental gradients even when no barrier to gene flow exists (Endler 1977), and if timing of reproduction in *E. recurva* differed between ocean basins this could produce effective genetic isolation even in the face of continued dispersal. In this way separation may appear to involve allopatric processes but may actually occur in the face of sustained gene flow that is rendered ineffectual by changes in mating recognition cues or reproductive timing (Norris 2000).

Support for this mechanism comes from the large body of evidence that suggests that many marine species are able to easily circumvent formidable hydrographic and tectonic barriers to dispersal but still fail to establish viable populations outside their main ranges (Norris 2000). Examples include the black marlin *Makaira indica* (Mooney-Seus and Stone 1997; Norris 2000), and the tuna *Thunnus thynnus* (Mooney-Seus and Stone 1997). These findings have been explained by a combination of high dispersal and regional limits on reproduction and have often invoked depth or seasonal parapatric models of speciation to explain the observed patterns (Norris 2000).

The sympatric occurrence of the highly divergent mitochondrial lineages in the SE Atlantic hints at secondary admixture although no samples from intermediate locations were collected to test for the co-occurrence Pacific and Atlantic sequences in increasing frequencies of the most common haplotypes in the other clades with distance from these clades. I cannot therefore determine which is the most parsimonious explanation for the shared haplotypes between Southern Hemisphere populations. If secondary admixture has occurred this would imply that migration has not been constant through time.

7.8.3. *Spatial Distribution of Migrants*

Another question that may be asked about these migrant individuals is; what, if anything, can be inferred from the spatial distribution and frequency

information of the putative migrant individuals among samples collected in the SE Atlantic? The non-random distribution of putative migrant sequences among the SE Atlantic samples is intriguing, and could be a sampling artifact given the difficulty of obtaining a sample truly representative of the population of an organism with such potentially massive census population size. Alternatively, it is possible I have sampled different euphausiid swarms that may have different genetic composition (Jarman *et al.* 2000). I consider the latter explanation unlikely, given that all of the SE Atlantic samples also contained a single shared *NDI* haplotype which was also the most frequent, whereas, the putative migrant haplotypes that define the differences between samples, were relatively rare overall. If migrants are entering the coastal waters around East Africa, travel southward in the boundary currents, and are transported into the Southern Benguela region via eddies and other frontal features shed from the southward flowing Agulhas Current, then the observed spatial distribution of migrants among samples may be explained by contemporary oceanographic features characteristic of the Southern Cape Region around Southern Africa (Shannon 1966).

The Brazil and Agulhas Currents are the major western boundary currents of the Southern Hemisphere and the latter plays a key role in the global ocean circulation, in the exchange of water, heat and salt between the Indian Ocean and Atlantic Oceans (Shannon 1996; Lutjeharms 2006). A possible candidate for the transport of migrants from the Indian Ocean into the northern reaches of

the southern Benguela Current System would be a detached Agulhas ring moving up the West Coast of South Africa (Lutjeharms and Gordon 1987, Duncombe Rae *et al.* 1989). These rings are periodically shed off the Agulhas Current System retroflexion and often maintain their hydrographic structure and biological communities of species from the south and east coasts of South Africa during passage into the SE Atlantic. Although most of these rings dissipate rapidly in the Cape Basin and around 40 percent never leave the SE Atlantic, in a few instances rings have been tracked all the way to the coast of Brazil (Garnier *et al.* 2003). These rings have been invoked to explain the observations of the tropical species of Indian Ocean origin *Euphausia diomedea* found in South Atlantic waters off the east coast of Brazil (Ponomavera 1990). It is possible that the migrants are transported from the south coast in one of these rings that broke off from the Agulhas Current retroflexion and were transported in the northward flowing jet-stream along the edge of the continental shelf off the west coast of South Africa.

If this scenario were correct, it might imply that the number of *E. recurva* individuals from the IWP being transported episodically into the waters off the west coast of South Africa in a single event might be relatively high, much larger than the estimates of the effective number of migrants produced by MDIV. This in turn might imply that large numbers of individuals may be transported into the Southern Benguela System but few are actually recruited into the west coast gene pool thus, explaining why the relative frequency of

migrants among these samples was low overall or that migration has not been constant through time.

Even when taking into account the possibility that the estimations of PDT obtained with MDIV may be accompanied by a large error, it seems unlikely that the strong population subdivision is the results of events during the last 18,000 years. A substantial amount of subdivision must have occurred earlier and survived several ice ages. Therefore the population subdivision has a long history predating the Pleistocene glaciations. Apparently, populations of *E. recurva* can remain highly subdivided in spite of the potential for gene flow, implying that their population and evolutionary dynamics can be independent. The pattern of divergence suggests the Southern Hemisphere clades may have involved a vicariant event that was perhaps associated with speciation. If separation were maintained for long enough the independent evolution of lineages in each region would occur. It appears that migration from the IWP to the SEA occurs but recruitment rates into the SEA may be low as evidenced by the low relative frequencies of highly divergent haplotypes in the SEA.

CHAPTER 8

8.1. Conclusions

Markedly contrasting patterns of differentiation among, and variation within populations were observed for *E. vallentini* and *E. recurva*. In the former species, sequence variation among individuals was completely absent with all of the molecular markers employed (*COI*, *16S rDNA* and *ITS-1*) with exception of the mitochondrial *NDI* fragment in which sequence heterogeneity was still very low. Consequently, no population subdivision was observed in this species with any of the molecular markers despite the enormous physical distances separating sampling sites. In contrast, *E. recurva* exhibited qualitative concordant patterns of differentiation for all the molecular markers with three geographically delineated clades consistently recovered in the phylogenetic analysis with high bootstrap support. In *E. lucens* a third pattern was observed where geographical structure was revealed in the mitochondrial genes *NDI* and *COI* with separate populations occurring in the SE and SW Atlantic basins but not for the nuclear *ITS-1*.

The levels of differentiation observed between populations in these species appears to be correlated with the latitude, the size of the geographical area encompassed by the species distribution and the hydrological conditions that characterize this area. For example, the warm-temperate species *E. recurva* occurring as it does in both hemispheres spans a larger geographical and experiences a range of complex hydrological conditions that likely restrict the

gene flow between different regions. Consequently, opportunities for dispersal between oceans are likely fewer in this species due to the influence of the continental landmasses compared to the other two species. For example, *E. recurva* populations in Atlantic waters around Brazil and Pacific waters off Peru almost certainly do not maintain contact around Cape Horn as the annual maximum temperatures in the Drake Passage would be lethal this warm-temperate species. Similarly transport between the Indian and South Pacific Oceans in the waters south of Australia may be curtailed for the same reason. Compare this with *E. vallentini*, which occurs in Southern Ocean where strong circumpolar currents provide extensive mixing of sub-Antarctic waters (Zane and Paternello 2000) and where the influence of the continental landmasses is likely small. Intermediate between these two species is *E. lucens*, in which inter-ocean connectivity may be limited by the southerly extension of the South American continent Cape Horn but not Australia based on distributional records of appearance.

The levels of within-sample polymorphism in *E. recurva* were almost always the highest among all three species investigated with all of the molecular markers employed. The excess of rare alleles observed in the Southern Hemisphere populations of *E. recurva* resemble more closely the patterns observed in *E. lucens* at least for the mitochondrial marker, whereas the *E. recurva* population in the North Pacific bears many similarities to the patterns

of polymorphism that appears to characterize *E. vallentini*. While the excess of rare alleles observed in *E. lucens* and *E. recurva* could be caused by population expansion, selective sweeps, background selection or sweepstakes recruitment, the excess of common alleles in *E. vallentini* hints at a population bottleneck or selective sweep. The estimated divergence times between geographical populations of *E. lucens* in the SE and SW Atlantic and *E. recurva* populations in the NE and IW Pacific and the SE Atlantic predate the last maximal glaciation (18,000 years ago) by quite a large margin which is not untypical for krill (Zane *et al.* 1998) and many other marine organisms. The high level of inter-population divergence and intra-population genetic cohesiveness of the Atlantic and Pacific *E. recurva* lineages suggest that these populations are distinct on both an ecological and evolutionary scale. Under the phylogenetic species concept (Cracraft 1989) all three *E. recurva* lineages would be considered new species and further work on the Euphausiacea should include an examination of the genetically distinct lineages for differences in morphological characters.

Migration between populations was inferred for the two Southern Hemisphere *E. recurva* populations only however, the question perhaps remains why the large divergences between these clades persist despite moderate levels of migration?

If we are to gain a better understanding how differentiation processes operate in these zooplankton then future research initiatives should endeavour to obtain *E. recurva* animals from the Indian Ocean, Kurishiwo Sea, South Pacific waters around Peru and New Zealand and central Atlantic populations from around Brazil. The analyses of several single or low copy nuclear genes such as the Myocin heavy chain may enable more accurate estimation of population parameters under non-equilibrium models with some of the more sophisticated coalescent analytical methods that could potentially circumvent the confounding effects of selection and demography.



CHAPTER 9

9.1. REFERENCES

Ahumada, R. 1976. Nota sobre los quetognatos capturados en la Expedición Hero 72-4 segunda etapa. *Boletín de la Sociedad de Biología de Concepción* 50: 27 – 34.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716 – 723.

Akashi, H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139: 1067 – 1076.

Akashi, H. 1997. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Gene* 205: 269 – 278.

Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A. and Kruglyak, L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PloS Biology* 2 (10): e 286.

Ansorge, I. J., Speich, S., Lutjeharms, J. R. E., Göni, G. J., de W. Rautenbach, C. J., Froneman, P. W., Rouault, M. and Garzoli, S. 2005. Monitoring the oceanic flow between Africa and Antarctica: Report of the first Good Hope cruise. *South African Journal of Science* 101: 29 – 35.

Antezana, T. 1976. Diversidad y equilibrio ecológico en comunidades pelágicas. Orrego, F. (ed.), *Preservación del medio ambiente marino*, pp. 40 – 54. Instituto de Estudios Internacionales, Universidad de Chile: Santiago.

Antezana, T. 1981. Zoogeography of euphausiids of the South Eastern Pacific Ocean. *Memorias del seminario sobre indicadores biológicos del plankton*, pp. 5 – 23. UNESCO, Montevideo.

Antezana, T. 1999a. Plankton of southern Chilean fjords: trends and linkages. *Scientia Marina* 63 (1): 69 – 80.

Antezana, T. 1999b. Hydrographic features of Magallanean and Fuegian inland passages and adjacent subantarctic waters. *Scientia Marina* 63 (1): 23 – 34.

Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P. and Slowinski, J. B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* 33: 707 - 740.

Arcos, D. 1974. Los copépodos calanoídeos colectados en la región Magallánica por la Expedición Hero 72 – 4b. *Boletín de la Sociedad de Biología de Concepción* 47: 215 – 225.

Arcos, D. 1976. Los copépodos calanoídeos de la región Magallánica. Expedición Hero 72 – 4. *Revista de la Comisión Permanente del Pacífico Sur* 5: 85 – 100.

Arhan, M., Naveira Garabato, A. C., Heywood, K. J. and Stevens, D. B. 2002. The Antarctic circumpolar current between the Falkland Islands and South Georgia. *Journal of Physical Oceanography* 32: 1914 – 1931.

Árnason, E. 2004. Mitochondrial cytochrome *b* DNA variation in the high-fecundity Atlantic cod: Trans-Atlantic clines and shallow gene genealogy. *Genetics* 166: 1871 – 1885.

Árnason, E., Petersen, P. H., Kristinsson, K., Sigurgíslason, H. and Pálsson, S. 2000. Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod from Iceland and Greenland. *Journal of Fish Biology* 56: 409 – 430.

- Ayala, F. J., Valentine, J. W. and Zumwalt, G. S. 1975. An electrophoretic study of the Antarctic zooplankter *Euphausia superba*. *Limnology and Oceanography* 20: 635 – 640.
- Baker, A. D. C., Boden, B. P. and Briton, E. 1990. *A practical guide to the euphausiids of the world*. Natural History Museum Publications: London, UK.
- Baker, P. F. and Burrell, J. 1977. The opening of Drake Passage. *Marine Geology* 25: 15 – 34.
- Bakun, A. 1996. Patterns in the ocean. Ocean processes and marine population dynamics. California Sea Grant College System, pp. 1 – 323. Centro de Investigaciones Biológicas: del Noroeste, Mexico.
- Ballard, J. W. O., and Kreitman, M. 1994. Unravelling selection in the mitochondrial genome of *Drosophila*. *Genetics* 138: 757 – 772.
- Ballard, J. W. O., and Whitlock, M. C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* 13: 729 – 744.
- Bandelt, H. J., Forster, P. and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37 – 48.

Bandelt, H., Forster, P., Sykes, B. C. and Richards, M. B. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141: 743 – 753.

Barange, M. 1990. Vertical migration and habitat partitioning of six euphausiid species in the northern Benguela upwelling system. *Journal of Plankton Research* 12: 1223 – 1237.

Barange, M. and Boyd, A.J. 1992. Life history, circulation and maintenance of *Nyctiphanes capensis* (Euphausiacea) in the northern Benguela upwelling system. *South African Journal of Marine Science* 12: 95 – 106.

Barange, M. and Pillar, S. C. 1992. Cross shelf circulation, zonation and maintenance mechanisms of *Nyctiphanes capensis* and *Euphausia hanseni* in the northern Benguela upwelling system. *Continental Shelf Research* 12: 1027 – 1042.

Barrett, P. J. 2001. Climate change – an Antarctic perspective. *New Zealand Science Review* 58: 18 – 23.

Beaumont, B.A., Zhang, W., Balding, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025 – 2035.

Benzie, J. A. H. 1999. Major genetic differences between crown-of-thorns starfish (*Acanthaster planci*) populations in the Indian and Pacific Oceans. *Evolution* 53: 1782 – 1795.

Benzie, J. A. H. 2000. Population genetic structure in penaeid prawns. *Aquaculture Research* 31: 95 – 119.

Berg, L. S. 1933. Les poisons des eaux douces de l'U.R.S.S. et des pays limitrophes. 3-e édition, revue et augmentée. Leningrad. *Poisons des Eaux Douces URSS*: 545 – 903.

Bernard, K. S., and Froneman, P. W. 2003. Mesozooplankton community structure and grazing impact in the Polar Frontal Zone of the south Indian Ocean during austral autumn 2002. *Polar Biology* 26: 268-275.

Bijma, J. Erez, J., and Hemiben, C. 1990. Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifers. *Journal of Foraminifera Research*. 20:117 – 127.

Boden, B. P. 1954. The euphausiid crustaceans of southern African waters. *Transactions of the Royal Society of South Africa* 34: 181 – 243.

Boden, B. P. 1955. Euphausiacea of the Benguela Current survey R. R. S. William Scoresby March 1950. *Discovery Report* 27: 337 – 376.

Botha, L. 1980. The biology of the Cape hakes *Meluccius capensis* (Cast.) and *M. paradoxus* (Franca) in the Cape of Good Hope area. Ph. D. thesis, University of Stellenbosch.

Bradley, R. D. and Baker, R. J. 2001. A test of the genetic species concept: Cytochrome-b sequences and mammals. *Journal of Mammalogy* 82: 960 – 973.

Brierley, A. S., Watkins, J. L. and Murray, A. W. A. 1997. Interannual variability in krill abundance at South Georgia. *Marine Ecology Progress Series* 150: 87 – 98.

Brinton, E. 1962a. The distribution of Pacific euphausiids. *Bulletin of the Scripps Institute of Oceanography* 8: 51 – 270.

Brunner, P.C., Douglas, M. R., Osinov, A., Wilson, C. C. and Bernatchez, L. 2001. Holarctic phylogeography of arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution* 55: 573 – 586.

Bucklin, A. and Wiebe, P. H. 1986. Genetic heterogeneity in euphausiid populations: *Euphausia krohnii* and *Nematoscelis megalops* in North Atlantic Slope Water. *Limnology and Oceanography* 31 (6): 1346 – 1352.

Bucklin, A. and Wiebe, P. H. 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. *The Journal of Heredity* 89 (5): 383 – 392.

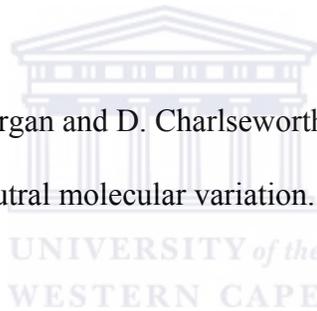
Bucklin, A., Frost, B. W. and Kocher, T. D. 1995. Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). *Marine Biology* 121: 655 – 664.

Bucklin, A., LaJeunesse, T. C., Curry, E., Wallinga, J. and Garrison, K. 1996. Molecular genetic diversity of the copepod, *Nannocalanus minor*: Genetic evidence of species and population structure in the N. Atlantic Ocean. *Journal of Marine Research* 54: 285 – 310.

Bucklin, A., Smolenack, S. B., Bentley, A. M. and Wiebe, P. H. 1997. Gene flow patterns of the euphausiid, *Meganyctiphanes norvegica*, in the NW Atlantic based on mtDNA sequences for cytochrome *b* and cytochrome oxidase I. *Journal of Plankton Research* 19 (11): 1763 – 1781.

Cann, R., Stoneking, M., and Wilson, A. C. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31 – 36.

- Casanova, B. 1984. Phylogenie des Euphausiacea (Crustaces Eucarides). *Bulletin du Muséum National d'Histoire Naturelle de Paris* 4 (6): 377 – 394.
- Cassens, I., Van Waerebeek, K., Best, P. B., Crespo, E. A., Reyes, J. and Milinkovitch, M. C. 2003. The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. *Molecular Ecology* 12: 1781 – 1792.
- Chappell, J. and Shackleton, N. J. 1986. Oxygen isotopes and sea level. *Nature* 324: 137 – 140.
- Charlesworth. B., M. T. Morgan and D. Charlseworth, 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289 – 1303.
- Charlesworth, B., Charlseworth, D. and M. T. Morgan and 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141: 1619 – 1632.
- Chu, K. H., Li, C. P. and Ho, H. Y. 2001. The first internal transcribed spacer (ITS-1) of ribosomal DNA as a molecular marker for phylogenetic and population analyses in Crustacea. *Marine Biotechnology* 3: 355 – 361.



CLIMAP Project Members. 1976. The surface of the Ice-Age earth. *Science* 191 (4234): 1131 – 1137.

Colburn, J., Crabtree, R. E., Shaklee, J. B., Pfeiler, E. and Bowen, B. W. 2001. The evolutionary enigma of bonefishes (*Albula* spp.): Cryptic species and ancient separations in a globally distributed shorefish. *Evolution* 55: 807 – 820.

Coleman, A. W. and Vacquier, V. D. 2002. Exploring the phylogenetic utility of ITS sequences for animals: A test case for abalone (*Haliotis*). *Journal of Molecular Evolution* 54: 246 – 257.

Colossi, G. 1917. Crostacei. Part II. Eufausiacei. Raccolte Planctoniche fatte dalla R. Nave *Ligura* 2: 165 – 205.



Corach, D. 1991. A reliable, rapid and simple method for DNA extraction from frozen sperm cells. *Fingerprint News* 3: 13.

Crame, J. A. 1993. Bipolar molluscs and their evolutionary significance. *Journal of Biogeography* 20: 145 – 161.

Crame, J. A. 1996. Evolution of high-latitude molluscan faunas. Taylor, J. (ed.), *Origin and Evolutionary Radiation of the Molluscan*, pp. 119 –131.

Crandall, K. A. and Templeton, A. R. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959 – 969.

Crandall, K. A., Posada, D. and Vasco, D. 1999. Effective population sizes: missing measures and missing concepts. *Animal Conservation* 2: 317 – 319.

Cunningham, C. W., Blackstone, N. W. and Buss, L. W. 1992. Evolution of king crabs from hermit crab ancestors. *Nature* 355: 539 – 542.

Curtolo, L. M., Dadon, J. R. and Mazzoni, H. E. 1990. Distribution and abundance of Euphausiids off Argentina in spring 1978. *Neritica* 5 (1): 1 – 14.

Darling, K. F., Kucera, M., Pudsey, C. J. and Wade, C. M. 2004. Molecular evidence links cryptic diversification in polar planktonic protists to Quaternary climate dynamics. *Proceedings of the National Academy of Science* 101 (20): 7657 – 7662.

Darling, K. F., Wade, C. M., Stewart, I. A., Kroon, D., Dingle, R. and Brown, A. J. L. 2000. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* 405: 43 – 47.

Dayhoff, M. O., Eck, R. V. and Park, C. M. 1972. A model of evolutionary change in proteins. Dayhoff, M. O. (ed.), *Atlas of Protein Sequence and Structure* 1972, Vol. 5, pp.89-99. Silver Springs, MD: The National Biomedical Research Foundation.

Donnelly, P. and Tavaré, S. 1986. The ages of alleles and a coalescent. *Advanced Applications of Probabilities* 18: 1 – 19.

Donnelly, P. and Tavaré, S. 1995. Coalescents and genealogical structure under neutrality. *Annual Review of Genetics* 29: 401 – 421.

Duda, T. F. Jr., and Palumbi, S. R. 1999. Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Marine Biology* 134: 705 – 710.

Duke, N. C., Benzie, J. A. H., Goodall, J. A. and Ballment, E. R. 1998. Genetic structure and evolution of species in the mangrove genus *Avicennia* (Avicenniaceae) in the Indo-West Pacific. *Evolution* 52: 1612 – 1626.

Duncombe Rae, C. M., Shannon, L. V., and Shillington, F. A. 1989. An Agulhas ring in the South Atlantic Ocean. *South African Journal of Science* 85 (11): 747 – 748.

Duró, A., and Gili, J. –M. 1996. Mesoscale spatial heterogeneity in chaetognath populations during upwelling abatement in the northern Benguela region.

Marine Ecology Progress Series. 140: 41 – 58.

Edwards, S. V. and Beerli, P. 2000. Perspective: Gene divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54 (6): 1839 – 1854.

Emiliani, C. 1954. Temperature of Pacific bottom waters and polar superficial waters during the Tertiary. *Science* 119: 853 – 855.

Endler, J. A. 1977. Geographic variation, speciation and clines. Monographs in *Population Biology* 10. Princeton University Press: Princeton, New Jersey.

Estoup, A., Beaumont, M., Sennedot, F., Moritz, C. and Cornuet, J. 2004. Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* 58 (9): 2021 – 2036.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles.

Theoretical Population Biology 3: 87 –112.

Excoffier, L. and Langaney, A. 1989. Origin and differentiation of human mitochondrial DNA. *American Journal of Human Genetics* 44: 73 – 85.

Fay, J. C. and Wu, C. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* 16 (7): 1003 – 1005.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Ecology* 17: 368 – 376.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783 – 791.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregation sites as compared to phylogenetic estimates. *Genetic Research* 56: 139 – 147.

Fevolden, S. E. 1986. Genetic variation of *Euphausia superba* Dana in the Antarctic Peninsula waters. *Sarsia* 71: 169 – 175.

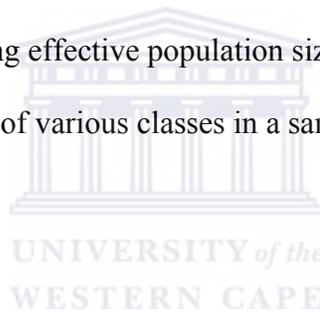
Fevolden, S. E., and Ayala, F. J. 1981. Enzyme polymorphism in Antarctic krill (Euphausiacea): Genetic variation between populations and species. *Sarsia* 66: 167 – 181.

Fevolden, S. E. and Schneppenheim R. 1989. Genetic homogeneity of krill (*Euphausia superba* Dana) in the Southern Ocean. *Polar Biology* 9: 533 – 539.

Ford, M. J. 2002. Application of selective neutrality tests to molecular ecology. *Molecular Ecology* 11: 1245 – 1262.

Fu, Y –X. 1994a. A phylogenetic estimator of effective population size or mutation rate. *Genetics* 136: 685 – 692.

Fu, Y –X. 1994b. Estimating effective population size or mutation rate using the frequencies of mutation of various classes in a sample of DNA sequences. *Genetics* 138: 1375 – 1386.



Fu, Y –X. 1995. Statistical properties of segregating sites. *Theoretical Population Biology* 48: 172 – 197.

Fu, Y –X. 1996. Estimating the age of the common ancestor of a DNA sample using the number of segregating sites. *Genetics* 144: 829 – 836.

Fu, Y –X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915 – 925.

Fu, Y –X. and Li, W –H. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693 – 709.

Fu, Y –X. and Li, W –H. 1999. Coalescing into the 21st century: An overview and prospectus of Coalescent Theory. *Theoretical Population Biology* 56: 1 – 10.

Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D. and Pettener, D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean Highlanders. *Molecular Biology and Evolution* 20 (10): 1682 – 1691.

Garnier, E. Verron, J. and Barnier, B. 2003. Variability in the South Atlantic open ocean circulation: a data assimilation experiment with 5 years of TOPEX/POSEIDON altimeter observations. *International Journal of Remote Sensing* 24 (5): 911 – 934.

Gibbons, M. J. 1993. Vertical migration and feeding of *Euphausia lucens* at two 72h stations in the southern Benguela upwelling region. *Marine Biology* 116: 257 – 268.

Gibbons, M. J. and Barange, M. 1994. Zoogeography and diversity of euphausiids around southern Africa. *Marine Biology* 123 (2): 257 – 268.

Gibbons, M. J. and Hutchings, L. 1996. Zooplankton diversity and community structure around southern Africa, with special attention to the Benguela upwelling system. *South African Journal of Science* 92: 63 – 76.

Gibbons, M. J., Barange, M., and Pillar, S. C. 1991a. Vertical migration and feeding of *Euphausia lucens* (Euphausiacea) in the southern Benguela. *Journal of Plankton Research*. 13: 473 – 486.

Gibbons, M. J., Pillar, S. C., and Stuart, V. 1991b. Selective carnivory by *Euphausia lucens*. *Continental Shelf Research*. 11: 625 – 640.

Gibbons, M. J., Spiridonov, V. A. and Tarling, G. A. 1999. Euphausiacea. Boltovsky, D. (ed.), *South Atlantic Zooplankton* Vol. 2, pp. 1241 – 1279. Backhuys Publishers: Leiden, The Netherlands.

Girman, D. 1996. The use of PCR-based single-stranded conformation polymorphism analysis (SSSCP-PCR) in conservation genetics. Smith, T. B. and Wayne, R. K. (eds.), *Molecular Genetic Approaches in Conservation*, pp. 167 – 182. Oxford University Press: Oxford.

Goetze, E. 2003. Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society of London B* 270: 2321 – 2331.

Gopurenko, D., Hughes, J. M. and Keenan, C. P. 1999. Mitochondrial DNA evidence for rapid colonisation of the Indo-West Pacific by the mudcrab *Scylla serrata*. *Marine Biology* 134: 227 – 233.

Grant, W. S. and Bowen, B. W. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* 89: 415 – 426.

Guglielmo, L. and Ianora, A. (Eds.). 1995. *Atlas of marine zooplankton. Straits of Magellan. Copepods*. Springer: Berlin.

Guglielmo, L. and Ianora, A. (Eds.). 1995. *Atlas of marine zooplankton. Straits of Magellan. Amphipods, Euphausiids, Mysids, Ostracods and Chaetognaths*. Springer: Berlin.

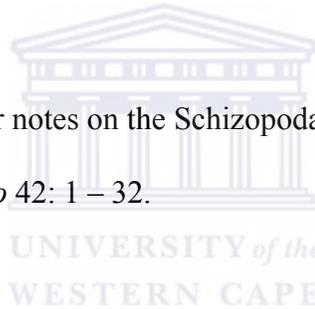
Hall, T. A. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41: 95 – 98.

Hamblin, M. T. and Di Rienzo, A. 2000. Detection of the signature of natural selection in humans: Evidence from the Duffy blood group locus. *American Journal of Human Genetics* 66: 1669 – 1679.

Hammer, M. F., Blackmer, F., Garrigan, D., Nachman, M. W. and Wilder, J. A. 2003. Human population structure and its effects on sampling y chromosome sequence variation. *Genetics* 164: 1495 – 1509.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 136: 557 – 580.

Hansen, H. G. 1905. Further notes on the Schizopoda. *Bulletin du Musée d'Océanographie de Monaco* 42: 1 – 32.



Hansen, H. G. 1915. The Crustacea Euphausiacea of the United States Museum. *Proceedings of the U.S. National Museum* 48: 59 – 114.

Harding, R. M., Fullerton S. M., Griffiths, R. C., Bond, J. Cox, M. J., Schneider, J. A., Moulin, D. S., and Clegg, J. B. 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. *American Journal of Human Genetics*. 60: 722-789.

Hardy, A. C. and Gunther, E. R. 1935. The plankton of the South Georgia whaling grounds and adjacent waters 1926 – 1927. *Discovery Report* 11: 1 – 456.

Harris, D. J. and Crandall, K. 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies. *Molecular Biology and Evolution* 17(2): 284 – 291.

Hasegawa, M., Cao, Y. and Yang, Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: Nonsynonymous / synonymous rate ratio is much higher within species than between species. *Molecular Biology and Evolution* 15 (11): 1499 – 1505.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97 – 109.

Hayashi, K. 1991. PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods and Applications* 1 (1): 34 – 38.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? Beaumont, A. R. (ed.), *Genetics and Evolution of Aquatic Organisms*, pp.122 - 134. London: Chapman & Hall.

Hendy, M. D. and Penny, D. 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Biosciences* 59: 277 – 290.

Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247 – 276.

Hey, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. *Molecular Biology and Evolution* 14: 166 – 172.

Hey, J. 1999. The neutralist, the fly and the selectionist. *Trends in Ecology and Evolution* 14 (1): 35 – 38.



Hey, J. 2005. On the number of New World founders: A population genetic portrait of the peopling of the Americas. *PloS Biology* 3: 6. e93.

Hey, J. and Harris, E. 1999. Population bottlenecks and patterns of human polymorphism. *Molecular Biology and Evolution* 16 (10): 1423 – 1426.

Heulsenbeck, J.P. and Crandall, K.A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28: 437 – 466.

Hickerson, M. J., Dolman, G., and Moritz, C. 2006. Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology*. 15: 209 – 223.

Hill, W. G., and Robertson, A. 1966. The effect of linkage on limits to artificial selection. *Genetics Research*. 8: 269 – 294.

Hillis, D.M., Moritz, C. Porter, A. and Baker, R.J. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251: 308 – 310.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Futuyma, D. and Antonovics, J. (eds.), *Oxford Surveys in Evolutionary Biology*, pp. 1 – 44. Oxford University Press: Oxford.

Hudson, R. R., Krietman, M. and Aguade, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153 – 159.

Hugall, A., Moritz, C., Moussalli, A. and Stanisic, J. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Science* 99 (9): 6112 – 6117.

Ivanovic, M. L. and Brunetti, N. E. 1994. Food and feeding of *Illex argentinus*. *Antarctic Science* 6 (2): 185 – 193.

James, A. G. 1987. Feeding ecology, diet and field-based studies on feeding selectivity of the Cape anchovy *Engraulis capensis* Gilchrist. In: *The Benguela and comparative ecosystems* (Payne, A. I. L., Gulland, J. A. and Brink, K. H. eds.). *South African Journal of Marine Science* 5: 673 – 692.

Jarman, S. N., Elliott, N. G., Nicol, S. and McMinn, A. 2000a. Molecular phylogenetics of circumglobal *Euphausia* species (Euphausiacea: Crustacea). *Canadian Journal of Fish and Aquatic Science* 57 (Suppl. 3): 51 – 58.

Jarman, S. N., Elliott, N. G., Nicol, S. and McMinn, A. 2002. Genetic differentiation in the Antarctic coastal krill *Euphausia crystallorophias*. *Heredity* 88: 280 – 287.

Jarman, S. N., Nicol, S., Elliott, N. G. and McMinn, A. 2000b. 28S rDNA evolution in the Eumalacostraca and the phylogenetic position of krill. *Molecular Phylogenetics and Evolution* 17 (1): 26 – 36.

John, D. 1936. The southern species of the genus *Euphasia*. *Discovery Report* 14: 193 – 324.

Johnson, M.W. and Brinton E. 1963. Biological species, water masses and currents. Hill, M. (ed.), *The Sea*, pp. 381 – 403.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624 – 626.

Kimura, M. 1983. *The neutral theory of Molecular Evolution*. Cambridge University Press: Cambridge, Massachusetts.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications* 13: 235 – 248.

Kitagawa, T., Nakata, H., Kimura, S., Itoh, T., Tsuji, S. and Nitta, A. 2000. Effect of ambient temperature on the vertical distribution and movement of Pacific bluefin tuna *Thunnus thynnus orientalis*. *Marine Ecology Progress Series* 206: 251 – 260.

Knowles, L. L. 2004. The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* 17: 1 – 10.

Knowlton, N. and Weigt, L. A. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B Biological Science* 265: 2257 – 2263.

Knowlton, N. Weigt, L.A. Sol'orzano, D. Mills, K. and Brmingham, E. 1993. Divergence in proteins, mitochondrial DNA and reproductive compatability across the Isthmus of Panama. *Science* 260: 1629 – 1632.

Kuhner, M. K., Yamato, J. and Felsenstein, J. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings Sampling. *Genetics* 140: 1421 – 1430.

Kuhner, M. K., Yamato, J. and Felsenstein, J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149: 429 – 434.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. 2001. Mega 2: Molecular Evolutionary Genetics Analysis. *Bioinformatics* 17: 1244 – 1245.

Lacson, J. M. and Clark, S. 1995. Genetic divergence of Maldivian and Micronesian demes of the damselfishes *Stegastes nigricans*, *Chrysiptera biocellata*, *C. glauca* and *C. leucopoma* (Pomacentridae). *Marine Biology* 121: 585 – 590.

Lavery, S., Moritz, C. and Fielder, D. R. 1995. Changing patterns of population structure and gene flow at different spatial scales in *Bigus latro* the coconut crab. *Heredity* 74: 531 – 541.

Lavery, S., Moritz, C. and Fielder, D. R. 1996. Genetic patterns suggest exponential population growth in a declining species. *Molecular Biology and Evolution* 13 (8): 1106 – 1113.

Lawver, L. A. and Gahagan, L. M. 1998. Opening of Drake Passage and its impact on Cenozoic ocean circulation. Crowley, T. J. and Burke, K. C. (eds.), *Tectonic boundary conditions for climate reconstruction*, pp. 212 – 223. Oxford University Press: Oxford.

Lessa, E. P., Cook, J. A. and Patton, J. L. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Science* 100 (18): 10331 – 10334.

Lessios, H. A., Kessing, B. D. and Pearse J. S. 2001. Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution* 55: 955 – 975.

Lessios, H. A., Kessing, B. D., Robertson, D. R. and Paulay, G. 1999.

Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution* 53: 806 – 817.

Lomakina, N.B. 1964. The euphausiid fauna of the Antarctica and notal regions. *Issledovaniya Fauny Morei*: Nauka, Leningrad 2: 260 – 342.

Lutjeharms, J. R. E. and Gordon, A.L. 1987. Shedding of an Agulhas Ring observed at sea. *Nature* 325 (7000): 138 – 140.

Lutjeharms, J. R. E. and Gordon, A.L. 2006. *The Agulhas Current*. Springer: Berlin.

Luttikhuisen, P. C., Drent, J. and Baker, A. J. 2003. Disjunct distribution of highly diverged mitochondrial lineage clade and population subdivision in a marine bivalve with pelagic larval dispersal. *Molecular Ecology* 12: 2215 – 2229.

Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution* 19 (11): 2005 – 2021.

Mauchline, J. 1966b. *Thalassomyces fagei*, an ellobiopsid parasite of the euphausiacea crustacean, *Thysanoessa raschii* (M.sars). *Journal of the Marine Biological Association of the United Kingdom* 46: 531 – 539.

Mauchline, J. and Fisher, L. R. 1969. The biology of euphausiids. *Advances in Marine Biology* 7. Academic Press: London.

Mauchline, J. 1980. The biology of mysids and euphausiids. *Advances in Marine Biology* 18: 1 – 623.

McDonald, C. M., Williams, R. and Adams, M. 1986. Genetic variation and population structure of krill (*Euphausia superba* Dana) from the Prydz Bay region of Antarctic waters. *Polar Biology* 6: 233 – 236.

McDonald, J. H. and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 354: 114 – 116.

McIntyre, A. Ruddiman, W. F., Karlin, K. and Mix, A. C. 1989. Surface water response of the equatorial Atlantic Ocean to orbital forcing. *Paleoceanography* 4(1): 19-56.

Mackintosh, N. A. 1934. Distribution of the macro-plankton in the Atlantic sector of the Antarctic. *Discovery Report* 9: 65 – 100.

Metropolis, N. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemistry and Physics* 21: 1087 – 1091.

Miller, D. G. M. and Hampton, I. 1989. Biology and ecology of the Antarctic krill (*Euphausia superba* Dana): A review. *BIOMASS Newsletter* 9: 1 – 165.

Miya, M. and Nishida, M. 1997. Speciation in the open ocean. *Nature* 389: 803 – 804.

Montú, M. 1977. Eufáusidos de la plataforma Argentina y adyacencias I. Distribución estacional en al sector Patagónico. *Ecosur* 4: 187 – 225.

Mooney-Sues, M.L. and Stone, G.S. 1997. *The forgotten giants*. Ocean Wildlife Campaign: Boston.

Murphy, N. E., and Goggin, C. L. 2000. Genetic discrimination of sacculinid parasites (Cirripedia, Rhizocephala): Implication for control of introduced green crab (*Carcinus maenas*). *Journal of Crustacean Biology* 20: 153 – 157.

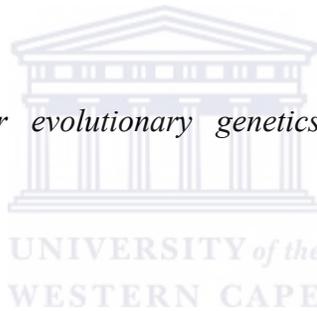
Murray, J. W. 1995. Microfossil indicators of ocean water masses, circulation and climate. In: *Marine palaeoenvironmental analysis from fossils* (Bosence,

D. W. J. and Allison, P. A., eds.). *Geological Society Special Publication* 83: 245 – 264.

Nachman, M. W., Boyer, S. N. and Aquadro, C. F. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proceedings of the National Academy of Science USA* 91: 6364 – 6368.

Nachman, M. W., Brown, W. M., Stoneking, M. and Aquadro, C. F. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* 142: 953 – 963.

Nei, M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.



Nei, M. and Li, W –H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* 76: 5269 – 5273.

Nepgen, C. S. 1957. The euphausiids of the west coast of South Africa. *Invertebrate Report, Division Sea Fisheries, South Africa* 21: 1 – 30.

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. and Sing, C. F. 1998.

- DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genetics* 19: 233 – 240.
- Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197 – 218.
- Nielsen, R. and Wakeley, J. 2001. Distinguishing migration from isolation: A Markov Chain Monte Carlo Approach. *Genetics* 158: 885 – 896.
- Norris, R. D. 1999. Hydrographic and tectonic control of plankton distribution and evolution. Abrantes, F. and Mix, A. (eds.), *Reconstructing ocean history: a window into the future*, pp. 173 – 193. Plenum: London.
- Norris, R. D. 2000. Pelagic species diversity, biogeography, and evolution. *Paleobiology* 26 (4): 236 – 258.
- Ohta, T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* 21: 263-286.
- Orita, M., Iwahana, H., Kanazawa, H. Hayashi, K. and Sekiya, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the Natural Academy of Science USA*. 86: 2766 – 2770.

Ovenden, J. R., Booth, J. D., and Smolenski, A. J. 1997. Mitochondrial DNA phylogeny of red and green rock lobsters (genus *Jasus*). *Marine and Freshwater Research* 48: 1131 – 1136.

Palma, S. and Silva, N. 2004. Distribution of siphonophores, chaetognaths, euphausiids and oceanographic conditions in the fjords and channels of southern Chile. *Deep-Sea Research II*. 51:513 – 535.

Palumbi, S. R. 1992. Marine speciation on a small planet. *Trends in Ecology and Evolution* 7: 114 – 118.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation and marine speciation. *Annual Review of Ecology and Systematics* 24: 547 – 572.

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L. and Tachina, N. 1997. Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution* 51 (5): 1506 – 1517.

Palumbi, S. R., Martin, A. Romano, S. McMillan, W.O. Stice, L. and Grabowsky, G. 1991. A simple fools guide to PCR. Version 2.0 University of Hawaii: Honolulu.

Papadopoulos, L. P., Peijnenburg, K. T. C. A. and Luttkhuizen, P. C. 2005. Phylogeography of the calanoid copepods *Calanus helgolandicus* and *C. euxinus* suggests Pleistocene divergences between Atlantic, Mediterranean and Black Sea populations. *Marine Biology* 147: 1353 – 1365.

Papetti, C., Zane, L., Bortolotto, E., Bucklin, A. and Patarnello, T. 2005. Genetic differentiation and local temporal stability of population structure in the euphausiid *Meganyctiphanes norvegica*. *Marine Ecology Progress Series* 289: 225 – 235.

Patarnello, T., Bargelloni, L., Varotto, V. and Battaglia, B. 1996. Krill evolution and the Antarctic Ocean currents: evidence of vicariant speciation as inferred by molecular data. *Marine Biology* 126: 603 – 608.

Peijnenburg, K.T.C.A., Breeuwer, J.A.J., Peirrot-Bults, A.C. and Menken, S.B.J. 2004. Phylogeography of the planktonic chaetognath *Sagitta setosa* reveals isolation in European seas. *Evolution* 58: 1472 – 1478.

Peijnenburg, K.T.C.A., Fauvelot, C., Breeuwer, J. A. J. and Menken, S. B. J. 2006. Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology* 15 (11): 3319 – 3338.

Peijnenburg, K.T.C.A., van Haastrecht E.K.V. and C. Fauvelot. 2005. Present day genetic composition suggests contrasting demographic histories of two dominant chaetognaths of the North East Atlantic. *Sagitta elegans* and *S. setosa*. *Marine Biology* 147 (6): 1279 – 1289.

Pillar, S. C. and Stuart, V. 1988. Population structure, reproductive biology and maintenance of *Euphausia lucens* in the southern Benguela current. *Journal of Planktonic Research* 10 (6): 1083 – 1098.

Pillar, S. C., Armstrong, D. A. and Hutchings, L. 1989. Vertical migration, dispersal and transport of *Euphausia lucens* in the southern Benguela current. *Marine Ecology Progress Series* 53: 179 – 190.

Pillar, S. C., Barange, M. and Hutchings, L. 1991. Influence of the frontal system on the cross-shelf distribution of *Euphausia lucens* and *Euphausia recurva* (Euphausiacea) in the southern Benguela system. *South African Journal of Science* 11: 475 – 482.

Ponomavera, L. A. 1969. Investigations on some tropical euphausiid species of the Indian Ocean. *Marine Biology* 3: 81 – 86.

Posada, D. and Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14: 817 – 818.

Posada, D. and Crandall, K. A. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* 16 (1): 37 – 45.

Posada, D. and Crandall, K. A. 2001. Selecting models of nucleotide substitution: An application to Human Immunodeficiency Virus 1 (HIV-1). *Molecular Biology and Evolution* 18 (6): 897 – 906.

Quesada, H., Ramos-Onsins, S. E., Rozas, J. and Aguadé, M. 2006. Positive selection versus demography: Evolutionary inferences based on an unusual haplotype structure in *Drosophila simulans*. *Molecular Biology and Evolution* 23 (9): 1643 – 1647.

Rabassa, J., Coronato, A. M. and Salemme, M. 2005. Chronology of the Late Cenozoic Patagonian glaciations and their correlation with biostratigraphic units of the Pampean region (Argentina). *Journal of South American Earth Sciences* 20: 81 – 103.

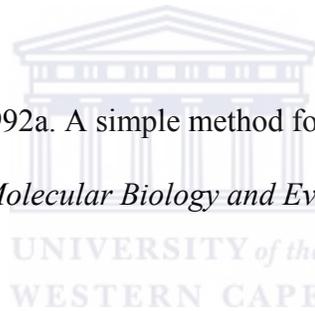
Ramirez, F. C. and Dato, C. 1983. Seasonal changes in population structure and gonadal development of three euphausiid species. *Oceanologica Acta Paris* 6 (4): 427 – 433.

Ramos – Onsins, S. E. and Rozas, J. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* 19 (12): 2092 – 2100.

Rand, D. M. 2001. The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* 32: 415 – 448.

Rand, D. M. and Kann, L. M. 1998. Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. *Genetica* 102 / 103: 393 – 407.

Rhetsky, A., and Nei, M. 1992a. A simple method for estimating and testing minimum evolution trees. *Molecular Biology and Evolution* 9: 945 – 967.



Ridoux, V. 1988. Subantarctic krill, *Euphasia vallentini* Stebbing, preyed upon by penguins around Crozet Islands (Southern Indian Ocean): Population structure and annual cycle. *Journal of Plankton Research* 10 (4): 675 – 690.

Roger, C. 1973. Recherches sur la situation trophique d'un groupe d'organismes pélagiques (Euphasiacea). II. Comportements nutritionnels. *Marine Biology* 18: 317 – 320.

Rosenberg, N. A. and Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3: 380 – 390.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. and Rozas, R. 2003. DNASP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496 – 2497.

Rychlick, W. 1989-1992. Oligo version 4. Primer analysis software. Copyright 1989-1992. National Biosciences Inc.: Plymouth, USA.

Saitou, N. and Nei, M. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406 – 425.



Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory: Cold Spring Harbour, New York.

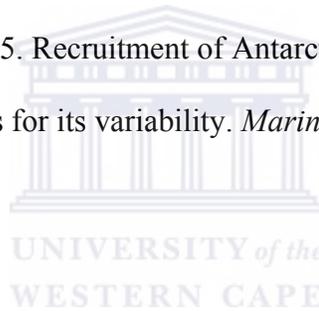
Schneppenheim, R. and McDonald, C. M. 1984. Genetic variation and population structure of krill (*Euphausia superba*) in the Atlantic sector of the Antarctic Peninsula. *Polar Biology* 6: 215 – 225.

Schubart, C. D., Diesel, R. and Hedges, S. B. 1998. Rapid evolution to terrestrial life in Jamaican crabs. *Nature* 393: 363 – 365.

Shannon, L. V. 1966. Hydrology of the south and west coasts of South Africa. *Investigation report. Division of Sea Fisheries* 58: 1 – 52.

Shields, G. F., and Gust, J. R. 1995. Lack of geographic structure in mitochondrial DNA sequences of Bering Sea walleye Pollock *Theragra chalcogramma*. *Molecular Marine Biology and Biotechnology* 4:69 – 82.

Siegel, V. and Loeb, V. 1995. Recruitment of Antarctic krill *Euphausia superba* and possible causes for its variability. *Marine Ecology Progress Series* 123: 45 – 56.



Silva, N. and Neshyba, N. 1979. On the southernmost extension of the Perú – Chile undercurrent. *Deep – Sea Research* 26A: 1387 – 1393.

Silva, N., Calvete, C. and Sievers, H. 1998. Masas de agua y circulación general para algunos canales australes entre Puerto Montt y Laguna San Rafael, Chile (Crucero Cimar Fiordio 1) *Ciencia y Tecnología del Mar* 21: 17 – 48.

Simmons, M. P. and Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology* 49 (2): 369 – 381.

Simmons, M. P., Ochoterena, H. and Carr, T. G. 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analysis. *Systematic Biology* 50 (3): 454 – 462.

Simonsen, K. L., Churchill, G. A. and Aquadro, C. F. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141: 413 – 429.

Skibinski, D. O. F. 2000. DNA tests of neutral theory: application in marine genetics. *Hydrobiologia* 420: 137 – 152.

Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16: 393 – 430.

Slatkin, M. and Hudson, R. R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555 – 562.

Smith, N. G. C. and Eyre-Walker, A. 2003. Partitioning the variation in mammalian substitution rates. *Molecular Biology and Evolution* 20 (1): 10 – 17.

Stebbing, T. R. R. 1900. On some crustaceans from the Falkland Islands collected by Mr Rupert Vallentin. *Proceedings of the Zoological Society of London* 3: 517 – 568.

Stuart, V. 1992. Fecundity of *Euphausia lucens* (Hansen) – laboratory evidence for multiple broods. *Journal of Experimental Marine Biology and Ecology* 160: 221 – 228.

Stuart, V. and Pillar, S. C. 1988. Growth and production of *Euphausia lucens* in the southern Benguela Current. *Journal of Plankton Research* 10: 1099 – 1112.

Stuart, V. and Pillar, S. C. 1990. Diel grazing patterns of all ontogenetic stages of *Euphausia lucens* in the southern Benguela Current. *Journal of Planktonic Research* 10 (6): 1099 – 1112.

Su, B., Fu, Y., Wang, Y., Jin, L. and Ranajit, C. 2001. Genetic diversity and population history of the red panda (*Ailurus fulgens*) as inferred from mitochondrial DNA sequence variations. *Molecular Biology and Evolution* 18 (6): 1070 – 1076.

Sundt, R. C. and Fevolden, S. E. 1996. Homogenous genetic structure of *Meganyctiphanes norvegica* (Euphausiacea) in the North East Atlantic Ocean as interpreted from allozymic variation. *Sarsia* 81: 155 – 159.

Sunnucks, P., Wilson, A. C. C., Beheregaray, L. B., Zenger, K., French, J. and Taylor, A. C. 2000. SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology* 9: 1699 – 1710.

Swafford, D. L. 1998. PAUP*, *Phylogenetic analysis using parsimony (*And Other Methods)*, Version 4. Sinauer Associates: Sunderland, MA.

Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite populations. *Genetics* 105: 437 – 460.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585 – 595.

Takahata, N. 1983. Gene identity and genetic differentiation of populations in the finite island model. *Genetics* 104: 497 – 512.

Takahata, N. 1986. An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. *Genetics Research* 48: 187 – 190.

Takezaki, N., Rzhetsky, A. and Nei, M. 1995. Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution* 12 (5): 823 – 833.

Tamura, K. and 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.

Tarling, G. A., Ward, P., Shearer, M., Williams, J. A. and Symon, C. 1995. Distribution patterns of macrozooplankton assemblages in the southwest Atlantic. *Marine Ecology Progress Series* 120: 29 – 40.

Tattersall, W. M. 1924. Euphausiacea. Report of the British Antarctica Terra Nova Expedition, Natural History, Zoology 13: 1 – 36.

Templeton, A. R., Crandall, K. A. and Sing, C. F. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117: 343 – 351.

Thomas, C. D., Franco, A. M. A. and Hill, J. K. 2006. Range retractions and extinction in the face of climate warming. *Trends in Ecology and Evolution* 21(8): 415 – 418.

- Tolley, K. A., Groeneveld, J. C., Gopal, K. and Matthee, C. A. 2005. Mitochondrial DNA panmixia in spiny lobster *Palinurus gilchristi* suggests a population expansion. *Marine Ecology Progress Series* 297: 225 – 231.
- Ulkes, F., Jansen, J. H. F. and Schneider, R. R. 2000. Anomalous occurrences of *Neogloboquadrina pachyderma* (left) in a 420-ky upwelling record from Walvis Ridge (SE Atlantic). *Marine Micropaleontology* 40: 23 – 42.
- Uthicke, S. and Benzie, A. H. 2003. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology* 12: 2635 – 2648.
- Van der Spoel, S., Pierrot-Bults, A. C. and Schalk, P. H. 1990. Probable Mesozoic vicariance in the biogeography of Euphausiacea. *Bijdragen tot de Dierkunde* 60 (3/4): 155 – 162.
- Vasco, A. D., Crandall, K. A. and Fu, Y –X. 2001. Molecular population genetics: coalescent methods based on summary statistics. Rodrigo, A. G. and Learn, G. H. Jr. (eds.), *Computational and Evolutionary Analysis of HIV Molecular Sequences*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Vilà, C., Amorim, I. R., Leonard, J. A., Posada, D., Castroviejo, J., Petrucci-Fonseca, F., Crandall, K. A., Ellegren, H. and Wayne, R. K. 1999.

Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology* 8 (12): 2089 – 2103.

Vogler, A. P. and DeSalle, R. 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. *Molecular Biology and Evolution* 11 (3): 393 – 405.

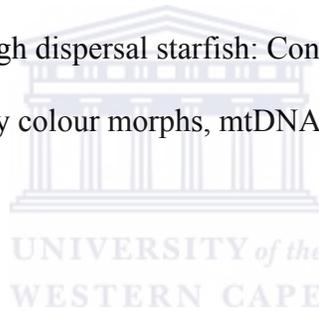
Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. *Evolution* 54 (4): 1092 – 1101.

Wakeley, J. and Hey, J. 1997. Estimating ancestral population parameters. *Genetics* 145: 847 – 855.

Wall, J. D., Andolfatto, P. and Pezeworski, M. 2002. Testing models of selection and demography in *Drosophila simulans*. *Genetics* 162: 203 – 216.

Wang, J. and Whitlock, C. 2003. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163: 429 – 446.

- Waples, R. S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity* 89: 438 – 450.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256 – 276.
- Wilkins, J. F. and Wakeley, J. 2002. The Coalescent in a Continuous Finite, Linear Population. *Genetics* 161: 873-888.
- Williams, D. and Benzie, J. A. H. 1998. Evidence of a biogeographic break between populations of a high dispersal starfish: Congruent regions within the Indo-west Pacific defined by colour morphs, mtDNA, and allozyme data. *Evolution* 52: 87 – 89.
- Williams, G. C. 1975. *Sex and Evolution*. Princeton University Press: Princeton, NJ.
- Williams, S. T., Jara, J., Gomez, E. and Knowlton, N. 2002. The marine Indo-West Pacific break: Contrasting the resolving power of mitochondrial and nuclear genes. *Integrative and Comparative Biology* 42: 941 – 952.
- Wright, S. 1943. Isolation by distance. *Genetics* 28: 114 – 138.



Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15: 323 – 354.

Yu, N., Fu, Y –X. and Li, W –H. 2002. DNA polymorphism in a worldwide sample of human X chromosomes. *Molecular Biology and Evolution* 19 (12): 2131 – 2141.

Zane, L. and Patarnello, T. 2000. Krill: a possible model for investigating the effects of ocean currents on the genetic structure of a pelagic invertebrate. *Canadian Journal of Fish and Aquatic Science* 57 (3): 16 – 23.

Zane, L., Ostellari, L., Maccatrozzo, L., Bargelloni, L., Battaglia, B. and Patarnello, T. 1998. Molecular evidence for genetic subdivision of Antarctic krill (*Euphausia superba* Dana) populations. *Proceedings of the Royal Society of London B* 265: 2387 – 2391.

Zane, L., Ostellari, L., Maccatrozzo, L., Bargelloni, L., Cuzin-Roudy, J. Bucholtz, P. Patarnello, T. 2000. Genetic differentiation in a pelagic crustacean (*Meganyctiphanes norvegica*: Euphausiacea) from the North East Atlantic and the Mediterranean Sea. *Marine Biology* 136: 191 – 199.F

Zehbe, I., Wilander, E. and Sallstrom, J. Pharmacia Application Note 384,
Phast System pp. 13 14, 18-1108-53.

Zhang, D. and Hewitt, G. M. 2003. Nuclear DNA analyses in genetic studies of
populations: practice, problems and prospects. *Molecular Ecology* 12: 563 –
584.

Zharkikh, A. and Li, W –H. 1992. Statistical properties of bootstrap estimation
of phylogenetic variability from nucleotide sequences: II. Four taxa without a
molecular clock. *Journal of Molecular Evolution* 35 (4): 356 – 366.

Zietkiewicz, E. Yotova, V. Jarnik, M. Korab-Laskowska, M. and Kidd, K.K.,
Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M. and Labuda,
D. 1998. Genetic structure of the ancestral population of modern humans.
Journal of Molecular Evolution 47 (2): 146-155.

Zimmer, C. 1914. Die Schizopoden der Deutschen Südpolar-Expedition 1901 –
1903. *Deutsch. Südpolar-Expedition 1901 – 1903*, 15 (Zool. 7): 377 – 445.

