DNA metabarcoding of zooplankton enhances community-level analyses of connectivity in a marine pelagic environment

By

Ashrenee Govender

MSc. Genetics (*cum laude*)

Submitted in fulfillment of the academic requirements for the degree of

Doctor of Philosophy

In the Discipline of Genetics

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

As the candidate's supervisor(s), we have approved this dissertation for submission.

Supervisor: Dr. Sandi Willows-Munro	**Date:** 25 January 2021	**Signature:** [redacted]
Co-supervisor: Prof. Johan Groeneveld	**Date:** 25 January 2021	**Signature:** [redacted]
Co-supervisor: Dr. Sohana Singh	**Date:** 25 January 2021	**Signature:** [redacted]

Preface

The research contained in this dissertation was completed by the candidate while based in the discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, under the supervision of Dr. Sandi Willows-Munro, Professor Johan Groeneveld and Dr. Sohana Singh. The research was financially supported by DSI/NRF/ACEP Captor Project (Grant 110763). The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily attributed to the NRF.

This study represents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

Student: Ashrenee Govender	**Date:** 24 January 2021	**Signature:** [redacted]
Supervisor: Dr. Sandi Willows-Munro	**Date:** 25 January 2021	**Signature:** [redacted]
Co-supervisor: Prof. Johan Groeneveld	**Date:** 25 January 2021	**Signature:** [redacted]
Co-supervisor: Dr. Sohana Singh	**Date:** 25 January 2021	**Signature:** [redacted]

Declaration 1: Plagiarism

I, Ashrenee Govender, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written, but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material prepared by myself, published as journal articles, or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

_____

Signature: Ashrenee Govender

Date: 24 January 2021

Declaration 2: Publications

Detail of contribution to publications that form part and/or include research presented in this dissertation.

Chapter 2

1. Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

Chapter 3

1. Govender A., Groeneveld J., Singh S., Pillay S. and Willows-Munro S. (2021) Experimental validation of taxon-specific mini-barcode primers for metabarcoding of zooplankton. Ecological Applications.

Signature: Ashrenee Govender

Date: 24 January 2021

Abstract

Zooplankton are abundant and diverse marine organisms that form ecologically important communities in marine pelagic ecosystems. They are well-suited for biomonitoring of ecosystem health and changes in biodiversity because their community structure and biomass respond rapidly to environmental variation. Biomonitoring of zooplankton communities using traditional morphology-based species identification methods is labor-intensive due to their cryptic morphology, high diversity and small body size. Fast-developing molecular techniques such as DNA metabarcoding (large-scale, high-throughput DNA sequencing of targeted gene regions to simultaneously identify multiple species present in samples) may provide higher resolution, accurate, faster and more cost-effective biomonitoring tools. The primary objectives of this dissertation were to develop and test a novel DNA metabarcoding approach for biomonitoring of marine zooplankton over the continental shelf of eastern South Africa. Novel taxon-specific DNA mini-barcode primers were designed to increase species identification rates of selected taxa. Artificially assembled mock communities with known composition and relative abundances were then used in an experimental setup to test detection rates and the accuracy of designed and published primers. The DNA metabarcoding protocol was then used to assess connectivity among zooplankton communities over the narrow KwaZulu-Natal continental shelf. Plankton tow nets were used to sample cross-shelf transects at three sites (uThukela, Durban and Aliwal), which are strongly influenced by the Agulhas Current but differ in shelf width, seafloor substrate and benthic habitat structures. Connectivity network analysis detected distinct clustering of zooplankton communities associated with each transect. The hypothesis that a dynamic ocean current regime associated with the offshore Agulhas Current (nearby and flowing parallel to the shelf-edge) would result in similar well-mixed alongshore zooplankton communities was rejected. A strong benthic-pelagic coupling effect was inferred based on the species composition of planktonic larvae and benthic adults occurring at the respective transects. This dissertation provides a refined and novel method for biomonitoring of marine pelagic environments in coastal waters, based on taxon-specific DNA metabarcoding of zooplankton communities. The approach is well-suited to measuring the long-term effects of climate change on marine pelagic ecosystems and ocean productivity.

Acknowledgments

This dissertation is dedicated to my dear grandparents, the late Mr. Annamalay Sadien Govender, the late Mrs. Panjalamma Govender, Mr. Krishnasamy Moodley, the late Mrs. Panjalay Moodley, and to my loving parents Anandkrishna Govender and Kubashni Govender. Without the sacrifices made by all of you, I would not be where I am today. I could not have asked for better role models. Thank you for all your unconditional love and support. I appreciate you all more than you know. This Ph.D. is as much yours as it is mine. I love you all immensely.

Thank you to my supervisor Dr. Sandi Willows-Munro for your continuous mentorship, support and patience throughout my academic journey. I appreciate your guidance and help through the research and writing of this dissertation.

Thank you to my co-supervisor Professor Johan Groeneveld for all your valuable input and guidance throughout my Ph.D. and writing of this dissertation. I am grateful for your support.

Thank you to my second co-supervisor Dr. Sohana Singh, I am happy to be your first Ph.D. student and I hope to make you proud. Thank you for always answering my out of the blue phone calls. Your constant words of encouragement, support and guidance throughout this Ph.D. is much appreciated.

I want to thank the National Research Foundation (NRF) and the Oceanographic Research Institute (ORI) for their financial support throughout the years.

To some of my closest friends and family, Phylicia Rajah, Courtnee Kleinhans, Mariam Khan, Kranti Maharaj, Krimentha Govender and Sandhani Padayachee, thank you for your friendship and support. Having you in my life has really made a difference, especially during these uncertain times.

To my brother Melashen Govender, even though you are so far away, thank you for always being there for me and pushing me to do better. I love and appreciate you.

To my best friend and partner, Nolyn John, these past few years would have been impossible without you and your love, encouragement and constant support. Thank you for always believing in me and being my pillar of strength. I love and appreciate you!

Table of Contents

Preface ii

Declaration 1: Plagiarism iii

Declaration 2: Publications iv

Abstract v

Acknowledgments vi

Table of Contents vii

List of Tables ix

List of Figures xii

List of Appendices xv

Chapter One: General Introduction 1

1.1Rationale for research (nature and scope) 1

1.2 Justification 5

1.3 Aims and objectives 7

1.4 Dissertation overview 8

1.5 References 9

Chapter Two: The design and testing of mini-barcode markers in marine lobsters 20

2.1 Abstract 20

2.2 Introduction 20

2.3 Materials and methods 23

2.4 Results 25

2.5 Discussion 28

2.6 References 30

2.7 Supplementary data 36

Chapter Three: Marine zooplankton, mini-barcodes and DNA metabarcoding: The case for taxon-specific primers 45

3.1 Abstract 45

3.2 Introduction 46

3.3 Materials and methods 49

3.4 Results 59

3.5 Discussion 65

3.6 Acknowledgments 67

3.7 Data accessibility 68

3.8 References 68

3.9 Supplementary data 75

Chapter Four: Connectivity of marine zooplankton communities in a dynamic ocean environment off eastern South Africa, inferred from DNA metabarcoding analysis 83

4.1 Abstract 83

4.2 Introduction 84

4.3 Materials and methods 87

4.4 Results 94

4.5 Discussion 100

4.6 Acknowledgments 105

4.7 References 105

4.8 Supplementary data 121

C**hapter Five: General discussion and future possibilities 132**

5.1 General discussion 132

5.2 Future possibilities 134

5.3 References 135

List of Tables

Table S2.1 Primer information table for the standard COI region, universal mini-barcode, internal COI mini-barcode and lobster mini-barcode……..36

Table S2.2 Thermal cycling conditions for the standard COI, universal mini-barcode (touch up PCR), internal COI mini-barcode (touch down PCR) primers and lobster mini-barcode……...38

Table S2.3 Summary statistics of the sliding window analysis for two selected fragments of each fragment length, showing potential segments for mini-barcodes and their position within the full alignment. Statistics include mean Kimura 2-parameter (K2P) distance, proportion of zero non-conspecific K2P distance, proportion of zero cells in K2P distance matrix and congruence of neighbour joining trees (clade composition and clade composition shallow)…….…39

Table S2.4 Summary statistics for comparison trees of all 28 fragments. K-scores and Robinson-Foulds (R-F) scores are used to identify best comparison trees. Each score is ranked based on the dataset in ascending order…….……..40

Table S2.5 A list of adults and phyllosoma samples amplified using with the lobster mini-barcode primer set (taxonomy and GenBank accession numbers)……..….43

Table S2.6 Summary of amplification and sequencing results for each primer pair. PCR products that showed a single sharp band of correct size were sent for sequencing. Sequences were then BLASTed. Blast search results, identification % and E-value were recorded…….……44

Table 3.1 The six primer cocktails used in this DNA metabarcoding study: each of the COI primer cocktails amplify different fragments of the COI-5P gene region (Figure S3.1). Illumina adapter target sequences (indicated in bold and underlined) were used in accordance with the workflow from the Illumina 16S Metagenomics protocol (Illumina, 2013). These adapter targets allow Nextera indexing and Illumina adapter addition through PCR. See Supporting Information Table S3.1 for the complete list of primer sets used for the preliminary primer testing step without Illumina adapter target sequences……...……53

Table 3.2 The seven mock communities constructed and the two natural zooplankton communities together with the primers used for each library including a rationale. 56

Table 3.3 Overall comparisons for the 14 mock community libraries and four zooplankton community libraries for the different primer sets used in this study……62

Table S3.1 The complete list of primer sets used for primer testing. The primer pairs selected and sed in the DNA metabarcoding study are in bold and red……..……75

Table S3.2 Summary statistics for the prawn, shrimp, and crab dataset for the sliding window analysis of the two selected fragments of each fragment length, showing potential segments for mini-barcodes and their position within the full alignment. Statistics include mean Kimura 2-parameter (K2P) distance, proportion of zero non-conspecific K2P distance, proportion of zero cells in K2P distance matrix, and congruence of neighbour-joining trees (clade composition and clade composition shallow) ……...……78

Table S3.3 Summary statistics for comparison trees of all 28 fragments for the prawn, shrimp, and crab datasets. K-scores and Robinson-Foulds (R-F) scores are used to identify best comparison trees. Each score is ranked based on the dataset in ascending order. The selected fragment size for each dataset is highlighted in bold and red…….80

Table 4.1 Sample collection data using a plankton ring-net (500 µm mesh; 0.8 m ring diameter) near the surface (between 0 – 5m depth) for 5 minutes at a boat speed of 2 – 3 knots. Sampling stations along transects were at depth soundings of 20, 50, 100 and 200 m, respectively. Three replicate tows were conducted per station, except at Durban, where two tows were made at 50 m and 200 m respectively…….……..90

Table 4.2 Environmental parameters collected using the on-board Seabird SBE 19 plus V2 SeaCAT CTD sensor across all sampling stations. The CTD readings for each parameter were calculated between 2 – 5 m of the depth profile and averaged……94

Table 4.3 High-throughput summary statistics across the replicates collected at uThukela, Durban and Aliwal…….……95

Table S4.1 The six primer cocktails used in this DNA metabarcoding study (first round PCR): each of the COI primer cocktails amplify different fragments of the COI-5P gene region. Illumina adapter target sequences (indicated in bold and underlined) were used in accordance with the workflow from the Illumina 16S Metagenomics protocol (Illumina, 2013). These adapter targets allow Nextera indexing and Illumina adapter addition through PCR……121

Table S4.2 High-throughput summary statistics across the replicates collected at uThukela, Durban and Aliwal……...124

Table S4.3 (a) Jaccard dissimilarity and (b) Bray-Curtis dissimilarity matrices for ring net replicates across the different sampling sites.…….…….126

Table S4.4 Shared species between uThukela, Durban and Aliwal……..129

List of Figures

Figure 1.1 Map showing the unique bathymetry (dotted lines) and the water circulation of the Agulhas Current along the east coast of South Africa…….……...…….6

Figure 2.1 Frequency distributions of intra- and interspecific pairwise K2P distances calculated using the selected mini-barcode region (Fragment 230_b). The barcode gap (insert) lies between the genetic distances of 0.01 and 0.02…….……..26

Figure 2.2 Amplification of four different cytochrome oxidase (COI) primer pairs tested in this study. Lanes 2, 3, 8, 9, 14, 15, 21, 22, 27 and 28 are PCR products recovered from DNA extracted from adult *Panulirus homarus*. Lane 4, 5, 10, 11, 16, 17, 23, 24, 30 and 31 are PCR products recovered from DNA extracted from phyllosoma samples. Lane 6, 12, 18, 25 and 31 are PCR negative controls. Lane 1, 19, 20 and 32 are 100 bp molecular weight marker (Solis Biodyne). Lane 7, 13 and 26 had no samples loaded…….……27

Figure S2.1 A graphical representation of the relative annealing sites and orientation of the different primer sets on the COI barcode region……37

Figure S2.2 Frequency distribution of intra- and interspecific pairwise K2P distances calculated using the selected mini-barcode region (Fragment 230_b). The barcode gap (insert) lies between the genetic distances of 0.02 and 0.03……..……41

Figure S2.3 Agarose gel image showing the amplification success of the new lobster mini-barcode primer to amplify a range of different adult and phyllosoma lobster samples…….……42

Figure 3.1 Frequency distribution of intra- and interspecific pairwise K2P genetic distances calculated using the selected mini-barcode regions for (a) prawn, (b) shrimp (c) and crab datasets. The barcode gap (inserts) lies between the genetic distance of 0.03 to 0.04 for prawn, 0.03 to 0.04 for shrimp and 0.05 to 0.06 for crab. The frequency data (prawn = n/3500; shrimp = n/40 000 and crab = n/40 000) was normalized to obtain a range between 0 and 1……..……60

Figure 3.2 The expected number of species vs. the observed number of species for each taxonomic group in each mock community. Expected number of species per mock community (blue), observed number of species 5 detected using the taxon-specific primer cocktails (Mock 1a – 7a;

yellow) and observed number of species 6 detected using the universal primer pair only (Mock 1b – 7b; green)……..……64

Figure 3.3 Composition bar graphs at Class levels for the 4 coastal zooplankton communities identified using BOLD and BLAST databases. The different colours indicate the different classes. The overall percentage of the different classes were calculated using the number of species identified per class. Prawns, crabs, shrimp and lobster fall under Malacostraca and fish fall under Actinopterygii……65

Figure S3.1 A graphical representation of the relative annealing sites and orientation of the different primer sets for DNA metabarcoding targeting the Folmer COI barcode region…….77

Figure 4.1 Map showing the location and sampling stations for the three sampling sites (uThukela, Durban and Aliwal) in the Indian Ocean. The main ocean current, Thukela River Mouth and Durban Eddy are depicted……..88

Figure 4.2 Rarefaction curves for the three sampling sites (uThukela, Durban and Aliwal) and their respective sampling stations…….……..95

Figure 4.3 (a) A Venn Diagram with species composition plots of shared species clusters and (b) beta-diversity indices comparing the difference between the three sites (uThukela, Durban and Aliwal)……...……97

Figure 4.4 Species composition bar graphs (a) all the sites combined (overall), (b) across each site (uThukela, Durban and Aliwal) and (c) across each sampling station……..……98

Figure 4.5 Beta-diversity analysis to estimate the dissimilarity and similarity between the uThukela, Durban and Aliwal and their respective sampling stations. (A) NMDS ordination plot based on Bray-Curtis dissimilarity, stress value = 0.095555 (B) UPGMA cluster analysis based on Bray-Curtis dissimilarity……99

Figure 4.6 Connectivity network between the sampling sites (uThukela, Durban and Aliwal) and their respective sampling stations…….……100

Figure S4.1 Depth profiles measured using the on-board Seabird SBE 19 plus V2 SeaCAT CTD sensor at the deepest station of each site (uThukela, Durban and Aliwal)……..…….123

Figure S4.2 Venn Diagrams comparing the difference between the biological replicates collected at the different sampling stations across uThukela, Durban and Aliwal……..……125

Figure S4.3 Venn Diagrams comparing the differences between shared and unique clusters across the different sampling stations for uThukela, Durban and Aliwal……128

Figure S4.4 Boxplot of observed species richness, Shannon's diversity index, Simpson's diversity index and Fishers alpha across the (a) sampling stations and (b) sampling sites……..…131

List of Appendices

All appendices can be found at: https://doi.org/10.6084/m9.figshare.14378663.v1

Appendix 2.1 Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

Appendix 2.2 Lobster COI sequences downloaded from GenBank and BOLD.

Appendix 3.1 Prawn COI sequences downloaded from GenBank (tab 1), Shrimp COI sequences downloaded from GenBank (tab 2) and Crab COI sequences downloaded from GenBank (tab 3).

Appendix 3.2 Prawn mini-barcode primer sets. Each group of prawn is on a new tab.

Appendix 3.3 Shrimp mini-barcode primer sets. Each group of shrimps is on a new tab.

Appendix 3.4 Crab mini-barcode primer sets. Each group of crab is on a new tab.

Appendix 3.5 Specimens selected to create mock communities. Each mock community (Mock 1 – Mock 7) is on a new tab.

Appendix 3.6 Chi square tests for all mock communities.

Chapter One: General Introduction

1.1 Rationale for research (nature and scope)

The oceans cover more than 70% of the Earth's surface and have higher biodiversity than terrestrial and freshwater ecosystems (Radulovici *et al.* 2010; Trivedi *et al.* 2016a; Trivedi *et al.* 2016b). Marine ecosystems are home to a large number of flora and fauna, ranging from microscopic single-celled organisms to the largest mammals on earth with the World Register of Marine Species (WoRMS Editorial Board 2021) currently listing 236,632 marine species, with perhaps more than a million species yet to be discovered (Bucklin *et al.* 2011). Marine ecosystems play an essential role in providing humans with goods, services and cultural benefits for social and economic activities. In addition, marine ecosystems assist in other vital services related to their regulatory and habitat functions, such as pollution control, storm protection, flood control, habitat for species and shoreline stabilization (Barbier 2017). Despite the grave importance of marine ecosystems, they are currently experiencing stress due to anthropogenic driven climate change, overharvesting of natural resources, habitat degradation, pollution, ocean acidification and coral bleaching (Radulovici *et al.* 2010; Bhadury and Annapurna 2011; Trivedi *et al.* 2016b). These stressors affect vertical mixing on a local scale and result in higher water column stratification, directly impacting both phytoplankton (primary producers) and zooplankton (secondary producers) (Chust *et al.* 2014). Human driven stressors such as the removal of biomass through overfishing, and warming, deoxygenation and acidification of waters through the release of carbon dioxide (CO_2) and other greenhouse gases into the atmosphere have caused undesirable changes to the oceans leading to stratification (Breitburg *et al.* 2015). Stratification can alter biodiversity patterns, species distributions, marine communities and food web dynamics, leading to decreased ocean productivity, elevated extinction rates and increased threats to marine ecosystems (Hoegh-Guldberg and Bruno 2010; Jackson 2010). Global programmes to monitor marine ecosystems (Hays *et al.* 2005; Canonico *et al.* 2019) have been established (for example IO-GOOS, GEO BON, MBON, OBIS and GBIF) and local programmes in South Africa (SAEON, IOISA and Sentinel Site programme of the Department of Science and Technology) have followed this trend, however, the data provided within these programmes is not immediately useable as formats differ across studies and requires parsing/reformatting which is time-consuming. Nevertheless, these programmes provide impetus for developing robust biomonitoring tools that can rapidly and

accurately assess biodiversity changes in marine ecosystems. Knowledge of ecological trends during a period of climate change will allow for the development of effective ocean management strategies (Berry *et al.* 2019; Holman *et al.* 2019).

Marine zooplankton are small organisms that periodically dominate the abundance and biomass of multicellular pelagic animals; they are highly diverse and occupy a range of niches (Richardson 2008). Marine zooplankton are either holoplanktonic (life cycle entirely pelagic) or meroplanktonic (life cycle partially pelagic, including eggs and larval stages of many benthic invertebrates and fish) (Huggett and Kyewalyanga 2017). Zooplankton play an essential role in transferring energy from primary producers to higher trophic levels in marine ecosystems and are mediators of biogeochemical flux in oceans (Richardson 2008). Zooplankton communities (species composition and relative abundance) respond rapidly to environmental change such as temperature (Moore and Folt 1993; Kelly *et al.* 2016), salinity (Paturej and Gutkowska 2015; Gutierrez *et al.* 2018), predation pressure (Greene 1983) and chemical stressors (Havens and Hanazato 1993; Rodgher *et al.* 2009). Zooplankton communities have been described as 'beacons of change' that indicate the level of anthropogenic stress exerted on marine ecosystems (Richardson 2009).

The geographic and depth distribution of zooplankton depends largely on ocean currents and physical oceanographic features which influence their dispersal and facilitate the connectivity of marine populations (Richardson 2008; Cowen and Sponaugle 2009; Richardson 2009; Bucklin *et al.* 2018). In marine environments, connectivity refers to the demographic linking of populations across multiple temporal and spatial scales through the movement (actively or passively) or exchange of organisms (Cowen *et al.* 2007; Almany *et al.* 2009; Cowen and Sponaugle 2009; Hidalgo *et al.* 2017). Connectivity among populations or communities is crucial for the functioning of marine ecosystems and increases their resilience to natural and anthropogenic stressors (Almany *et al.* 2009; Jones *et al.* 2009). Therefore, understanding the patterns of zooplankton dispersal and connectivity can help manage marine ecosystems to preserve biodiversity and recruitment to populations of species with meroplanktonic larvae (Richardson 2008; Cowen and Sponaugle 2009).

Biomonitoring of whole zooplankton communities using traditional methods (morphological identification based on microscopic examination) is challenging because of the small size, fragile

nature of many taxa, cryptic morphology and high species diversity (Bucklin *et al.* 2016). Traditional methods are time-consuming and require taxonomic expertise, which is often not available. Recent advances in molecular technologies offer new opportunities to study biodiversity that are more rapid, cost-effective and accurate than traditional methods (Shokralla *et al.* 2012; Thomsen *et al.* 2012). The crossover from microscopy to molecular analyses in marine biodiversity studies is reviewed by Laakmann *et al.* (2020).

DNA barcoding is well-established and has revolutionized species identification and discovery over the past two decades (Hebert *et al.* 2003; Hebert and Gregory 2005). DNA barcoding allows for the assignment of specimens or samples (e.g. a piece of tissue or contents of a gut) to species-level by sequencing short, standardized DNA fragments ('DNA barcodes') and comparing them against reference libraries such as Barcode of Life Data Systems (BOLD, www.barcodeoflife.org), International Nucleotide Sequence Database Collaboration (INSDC, www.insdc.org) and the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) (Hebert *et al.* 2003). DNA barcoding assumes that genetic variation is greater between than within species (Hebert *et al.* 2003). The standard DNA barcode used in most animal groups is a 658-base pair (bp) portion of the cytochrome c oxidase 1 (COI) mitochondrial gene. This gene has a faster substitution rate than nuclear rRNA genes, making it suitable for discrimination between species (Hebert *et al.* 2003). The mutational rate of the COI region is suitable for interspecific taxonomic assignment of sequences to species-level but can also detect intraspecific variation which can be used to study genetic structure between different populations of a single species (Hebert and Gregory 2005). The mitochondrial COI region has been used in DNA barcoding studies because it has broad and extensive barcode reference libraries that provide a powerful link to taxonomic identifications (Ratnasingham and Hebert 2007), thus decreasing the probability of false taxonomic assignments amongst closely related species (Somervuo *et al.* 2017).

Recent advances in new sequencing technologies have led to the amalgamation of DNA barcoding and high-throughput sequencing (HTS) to create a new approach known as DNA metabarcoding (Dormontt *et al.* 2018; Piper *et al.* 2019; Ruppert *et al.* 2019). DNA metabarcoding allows for DNA from entire communities to be extracted in bulk followed by mass amplification of standard genetic markers and sequencing using HTS technologies (Taberlet *et al.* 2012; Bourlat

et al. 2013; Cristescu 2014; Creer *et al.* 2016). Thousands of sequences can be compared against DNA reference libraries simultaneously, allowing for rapid species identification from mixed communities (Coissac *et al.* 2012). DNA metabarcoding utilizes the same reference libraries as DNA barcoding and has been applied to terrestrial (Gibson *et al.* 2014; Arribas *et al.* 2016; Gous *et al.* 2019; Porter *et al.* 2019; Thomsen and Sigsgaard 2019), freshwater (Hajibabaei *et al.* 2011; Carew *et al.* 2013; Elbrecht and Leese 2017; Andújar *et al.* 2018; Elbrecht and Steinke 2019; Hajibabaei *et al.* 2019) and marine ecosystems (Aylagas *et al.* 2014; Leray and Knowlton 2015; Fraija-Fernández *et al.* 2020). DNA metabarcoding has been used to analyze the taxonomic composition of zooplankton communities (Djurhuus *et al.* 2018), examine the temporal and spatial distribution of zooplankton communities (Casas *et al.* 2017) and assess the prospects of zooplankton for applications for ocean monitoring (Bucklin *et al.* 2019; Laakmann *et al.* 2020). The data generated through DNA metabarcoding of zooplankton communities not only provides valuable insight into changes in species composition under different environmental conditions but by coupling DNA metabarcoding data with oceanographic information and community analysis, processes such as dispersal and connectivity can be inferred (Macher *et al.* 2020; Pitz *et al.* 2020; Singh *et al.* 2021).

The 658 bp length of the standard cytochrome c oxidase subunit I (COI) 'Folmer' (Folmer *et al.* 1994) gene region used in classic DNA barcoding is currently beyond the reach of many HTS technology platforms (e.g., the Illumina MiSeq platform) that have limited read length (Marquina *et al.* 2019). In addition, the DNA integrity of many marine organisms is highly dependent on several factors such as sampling methods, the rapidity of sampling preservation, the type of preservatives used and how the organisms are stored. It has been recommended that DNA should be extracted immediately after tissue sampling or stored at subzero temperatures for DNA extraction to be carried out as soon as possible (Oosting *et al.* 2020). There have been many reports of DNA degradation in ethanol preserved samples with increasing loss of DNA over time at warmer temperatures and in samples with higher water content (Goetze *et al.* 2013). Due to time and resource constraints, it is often not possible to carry out DNA extractions immediately after sampling, therefore in some cases the DNA of zooplankton is often damaged or degraded due to incorrect sampling methods and possible rapid post-capture DNA degradation from poor preservation (Boyer *et al.* 2012). Therefore, the use of shorter DNA fragments (200 – 300 bp), known as mini-barcodes, are recommended for DNA metabarcoding studies (Hajibabaei *et al.*

2006; Meusnier *et al.* 2008; Leray *et al.* 2013). Mini-barcodes should retain sufficient information and preferably target hypervariable DNA regions to accurately delimit species (Hajibabaei *et al.* 2006; Leray *et al.* 2013). A 'universal' mini-barcode developed for the standard COI gene region (Meusnier *et al.* 2008) can identify a range of mammals, fishes, birds and insects from archival samples. However, the universal mini-barcode primers are not effective across all taxa due to primer mismatch (Arif *et al.* 2011). Taxon-specific mini-barcode primers improve PCR amplification, sequencing success rates and offer higher discriminating power than universal primers (Dong *et al.* 2014; Govender *et al.* 2019).

Method validation is a crucial aspect of applying DNA metabarcoding to ecological studies or long-term biomonitoring to identify potential methodological biases and shortcomings (Zhang *et al.* 2018). Some biases and shortcomings using DNA metabarcoding include false-positives (Ficetola *et al.* 2016), false-negatives (Zhang *et al.* 2018), primer bias (Clarke *et al.* 2017) and the inability to quantify taxon abundance (Elbrecht and Leese 2015). There is currently very little agreement about which laboratory protocols or bioinformatic pipelines to use for DNA metabarcoding, preventing the development of standard methods that generate comparable outcomes, irrespective of the natural system studied or geographical location. Poor standardization of markers across DNA metabarcoding studies reduces comparability among studies, limiting community metabarcoding as an efficient, universal system for biodiversity assessments and monitoring. At present, many primers or pipelines are purpose-built for specific ecosystems or taxonomic groups and might not be transferrable. The MetaZooGene Working Group (https://metazoogene.org/) is a recent global initiative to create an open-access web portal, database and atlas for DNA barcodes of marine zooplankton. MetaZooGene is actively involved in designing an optimal molecular pipeline for species identification of zooplankton and in the development of best practice guides for DNA metabarcoding of zooplankton biodiversity.

1.2 Justification

The east coast of South Africa is unique and characterized by a complex seafloor morphology that includes a very narrow continental shelf, high terrigenous input and high energy benthic boundary conditions (Sink *et al.* 2019). There are around 24 submarine canyons in this region, with many of them only indenting the continental shelf break causing a morphological disruption to the adjacent continental slope (Green and Uken 2008; Green 2009, 2011). The world's strongest

western boundary current, known as the Agulhas Current, flows strongly southward along the east coast of South Africa bringing tropical and subtropical surface waters from the equatorial and subtropical regions of the western Indian Ocean (Lutjeharms 2006, 2007). The Agulhas Current follows the shelf-edge and extends well below 1000 m with a mean width of 100 km. The continental shelf between the Agulhas Current and the KwaZulu-Natal (KZN) coast is mostly narrow (3 – 11 km wide) and slopes down steeply after reaching about 100 m depth, apart from the KZN Bight (an offset of 160 km long with a maximum width of 45 km) (Schumann 1987). The broadening of the shelf at the KZN Bight causes the Agulhas Current to divert offshore, resulting in the formation of shelf-edge upwelling cells, cyclonic lee-trapped eddies and countercurrents. The KZN coastal region displays complex physical oceanography (Guastella and Roberts 2016; Roberts *et al.* 2016) that strongly influences the dispersal of marine taxa including zooplankton over the shelf (Bustamante and Branch 1996; Collocott 2016; Pretorius *et al.* 2016).

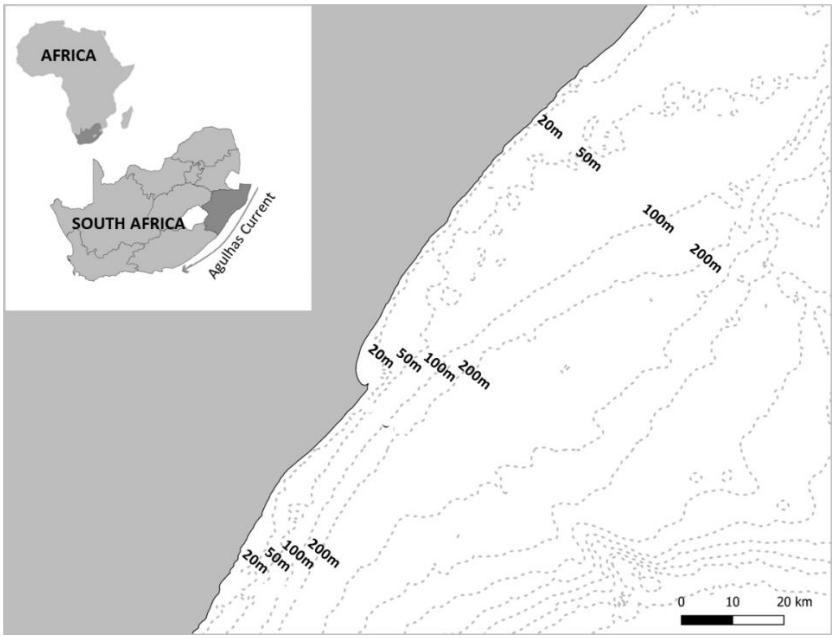


Figure 1.1 Map showing the unique bathymetry (dotted lines) and the water circulation of the Agulhas Current along the east coast of South Africa.

Zooplankton research carried out along the east coast of South Africa, particularly the KZN coastal region, stems from a small number of standalone surveys or surveys with large time gaps. These studies focused on the distribution and diversity of copepods and chaetognaths (De Decker 1964; De Decker and Mombeck 1964; Carter 1977; De Decker 1984; Schleyer 1985; Pretorius *et*

al. 2016), the description of lobster phyllosoma (Berry 1974), the assessment of siphonophores and hydromedusae assemblages (Thibault-Botha *et al.* 2004; Buecher *et al.* 2005; Thibault-Botha and Gibbons 2005) and the species composition and dispersal of fish larvae (Beckley 1986; Beckley and Hewitson 1994; Beckley 1995; Pattrick and Strydom 2014). These studies have focused on a single taxon/group using mainly morphological identification of species. DNA metabarcoding to determine the species composition of samples combined with community-level analyses can allow for long-term high-resolution biomonitoring of marine zooplankton as indicators of the effects of climate change on ocean productivity. However, developing, validating and improving DNA-based monitoring tools require that several technical hurdles be overcome. This study will focus on decapod and fish species with life-history stages important to commercial fisheries such as prawns (Dendrobranchiata), shrimps (Caridea), crabs (Brachyura), lobsters (Astacidea, Glypheidea, Achelata, and Polychelida) and fish (Actinopterygii).

Two major paradigms in high-throughput sequencing technologies are short-but-accurate read sequencing and long-but-error-prone read sequencing (Goodwin *et al.* 2016; Piper *et al.* 2019). Short-read sequencing allows for targeting of specific gene regions and whole genome analysis, especially in shotgun metagenomic sequencing followed by computational binning approaches, while long-read sequencing allows for whole-genome analysis of all genetic material in a sample (Piper *et al.* 2019). In this study, short-read sequencing using the Illumina MiSeq technology was used due to its high-quality reads, relatively inexpensive sequencing costs and shorter fragment lengths which allow for the sequencing of DNA that may be degraded and/or fragmented.

1.3 Aims and objectives

The overall objective of the dissertation was to develop, validate and apply a novel DNA metabarcoding approach for long-term biomonitoring of zooplankton species composition and relative abundance along the KZN coast.

Specific aims were:

1. To develop and test taxon-specific mini-barcode primers for use in DNA metabarcoding of selected taxa (decapods and fishes) to increase species detection rates in bulk tow-net samples.
2. To establish a standard experimental protocol and test the protocol on artificially assembled communities with known zooplankton species composition and relative

abundance to quantify amplification success rates of any combination (or cocktail) of primers selected.

3. To combine species composition information obtained from DNA metabarcoding with community analyses to assess the connectivity among geographically separated zooplankton communities over the narrow continental shelf of eastern South Africa, between the KZN coast and the upper Agulhas Current.
4. To contribute a purpose-designed and tested methodology that will enable long-term biomonitoring of pelagic ecosystems in an ocean region heavily influenced by climate change.

1.4 Dissertation overview

Chapters 2 – 4 in this dissertation have been formatted as manuscripts for publication in peer-reviewed journals, and hence each chapter has a separate introduction, materials and methods, results, discussion and conclusion section. Some overlap in the content (particularly references) was therefore unavoidable.

Chapter 1: A general introduction for the dissertation focusing on the background, rationale and justification for the intended study.

Chapter 2: A novel protocol for designing taxon-specific DNA mini-barcode primers was developed to increase species detection rates in environmental samples. An *in-silico* method was used to identify the shortest and most informative portion of the COI gene region, followed by the design and testing of mini-barcode primers against published universal COI primer sets. This chapter has been through peer review and is published in PLOS One (Govender *et al.* 2019)

Chapter 3: An experimental protocol based on artificially assembled communities with known zooplankton species composition and relative abundance was developed to quantify the species detection rates and accuracy of any combination (or cocktail) of primers selected. The experimental setup provided a rapid and cost-effective tool for optimizing primer cocktails to target selected taxa. This chapter is currently under peer review (Govender *et al.* in review).

Chapter 4: The species composition information generated with DNA metabarcoding was combined with a community-level analysis approach to infer connectivity between geographically separated zooplankton communities over the continental shelf of eastern South Africa, between

the KZN coast and the upper Agulhas Current. The techniques developed in the previous chapters were applied in a refined and novel approach as a 'proof-of-concept' and can henceforth be applied to long-term biomonitoring of marine pelagic environments during an era of global warming.

Chapter 5: The research dissertation's outcomes are integrated into a general discussion, highlighting how a successful DNA metabarcoding methodology can revolutionize biomonitoring and ecological research on zooplankton and other groups with small size and high diversity.

1.5 References

Almany G., Connolly S., Heath D., Hogan D., Jones G., McCook L., Mills M., Pressey R. and Williamson D. (2009) Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. Coral Reefs 28, 339-51.

Andújar C., Arribas P., Gray C., Bruce C., Woodward G., Yu D.W. and Vogler A.P. (2018) Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. Molecular Ecology 27, 146-166.

Arif I., Khan H., Sadoon M. and Shobrak M. (2011) Limited efficiency of universal mini-barcode primers for DNA amplification from desert reptiles, birds and mammals. Genetics and Molecular Research 10, 3559-3594.

Arribas P., Andújar C., Hopkins K., Shepherd M. and Vogler A.P. (2016) Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods in Ecology and Evolution 7, 1071-1081.

Aylagas E., Borja Á. And Rodríguez-Ezpeleta N. (2014) Environmental status assessment using DNA metabarcoding: Towards a genetics Based Marine Biotic Index (gAMBI). PLOS One 9, e90529.

Bar W., Kratzer A., Machler M. and Schmid W. (1988) Postmortem stability of DNA. Forensic Science International 39, 59-70.

Barbier E.B. (2017) Marine ecosystem services. Current Biology 27, R507-R510.

Beckley L.E. (1986) The ichthyoplankton assemblage of the Algoa Bay nearshore region in relation to coastal zone utilization by juvenile fish. South African Journal of Zoology 21, 244-252.

Beckley L.E. (1995) The Agulhas Current ecosystem with particular reference to dispersal of fish larvae. In: Status and future of large marine ecosystems of the Indian Ocean: A report of

the international symposium and workshop. Okemwa E., Ntiba M.J. and Sherman K. (eds). pp. 74-91. International Union for Conservation of Nature (IUCN), Gland.

Beckley L.E. and Hewitson J.D. (1994) Distribution and abundance of clupeoid larvae along the east coast of South Africa in 1990/91. South African Journal of Marine Science 14, 205-212.

Berry P.F. (1974) Palinurid and scyllarid lobster larvae of the Natal Coast, South Africa. Oceanographic Research Institute, Durban. Investigational report no. 34: 1-44.

Berry T.E., Saunders B.J., Coghlan M.L., Stat M., Jarman S., Richardson A.J., Davies C.H., Berry O., Harvey E.S. and Bunce M. (2019) Marine environmental DNA biomonitoring reveals seasonal patterns in biodiversity and identifies ecosystem responses to anomalous climatic events. PLOS Genetics 15, e1007943.

Bhadury P. and Annapurna C. (2011) Marine barcoding – How will it help Indian marine benthic studies? Indian Journal of Marine Sciences 40, 645-647.

Bourlat S.J., Borja A., Gilbert J., Taylor M.I., Davies N., Weisberg S.B., Griffith J.F., Lettieri T., Field D., Benzie J., Glöckner F.O., Rodríguez-Ezpeleta N., Faith D.P., Bean T.P. and Obst M. (2013) Genomics in marine monitoring: New opportunities for assessing marine health status. Marine Pollution Bulletin 74, 19-31.

Boyer S., Brown S.D.J., Collins R.A., Cruickshank R.H., Lefort M.C., Malumbres-Olarte J. and Wratten S.D. (2012) Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA. PLOS One 7, e38215.

Breitburg D.L., Salisbury J., Bernhard J.M., Cai E., Dupont S., Doney S.C., Kroeker K.J., Levin L.A., Long W.C., Milke L.M., Phelan B., Passow U., Seibel B.A., Todgham A.E. and Tarrant A.M. (2015) And on Top of All That… Coping with Ocean Acidification in the Midst of Many Stressors. Oceanography 28, 48-61.

Bucklin A., Divito K., Smolina I., Choquet M., Questel J., Hoarau G. and O'Neill R. (2018) Population Genomics of Marine Zooplankton. In: Population Genomics: Marine Organisms. Oleksiak M.F. and Rajora O.P. (eds). pp. 61-102. Springer, Cham.

Bucklin A., Lindeque P.K., Rodriguez-Ezpeleta N., Albaina A. and Lehtiniemi M. (2016) Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. Journal of Plankton Research 38, 393-400.

Bucklin A., Steinke D. and Blanco-Bercial L. (2011) DNA barcoding of marine metazoa. Annual Review of Marine Science 3, 471-508.

Bucklin A., Yeh H.D., Questel J.M., Richardson D.E., Reese B., Copley N.J. and Wiebe P.H. (2019) Time-series metabarcoding analysis of zooplankton diversity of the NW Atlantic continental shelf. ICES Journal of Marine Science 76, 1162-1176.

Buecher E., Goy J. and Gibbons M. (2005) Hydromedusae of the Agulhas Current. African Invertebrates 46, 27-69.

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

Canonico G., Buttigieg P.L., Montes E., Muller-Karger F.E., Stepien C., Wright D., Benson A., Helmuth B., Costello M., Sousa-Pinto I., Saeedi H., Newton J., Appeltans W., Bednaršek N., Bodrossy L., Best B.D., Brandt A., Goodwin K.D., Iken K., Marques A.C., Miloslavich P., Ostrowski M., Turner W., Achterberg E.P., Barry T., Defeo O., Bigatti G., Henry L.A., Ramiro-Sánchez B., Durán P., Morato T., Roberts J.M., García-Alegre A., Cuadrado M.S. and Murton B. (2019) Global observational needs and resources for marine biodiversity. Frontiers in Marine Science 6, 367.

Carew M.E., Pettigrove V.J., Metzeling L. and Hoffmann A.A. (2013) Environmental monitoring using next generation sequencing: Rapid identification of macroinvertebrate bioindicator species. Frontiers in Zoology 10, 45.

Carter R. (1977) The distribution of calanoid Copepoda in the Agulhas Current system off Natal, South Africa. MSc Thesis, University of KwaZulu-Natal, South Africa.

Casas L., Pearman J.K. and Irigoien X. (2017) Metabarcoding reveals seasonal and temperature-dependent succession of zooplankton communities in the Red Sea. Frontiers in Marine Science 4, 241.

Chust G., Allen J.I., Bopp L., Schrum C., Holt J., Tsiaras K., Zavatarelli M., Chifflet M., Cannaby H., Dadou I., Daewel U., Wakelin S.L., Machu E., Pushpadas D., Butenschon M., Artioli Y., Petihakis G., Smith C., Garçon V., Goubanova K., Le Vu B., Fach B.A., Salihoglu B., Clementi E. and Irigoien X. (2014) Biomass changes and trophic amplification of plankton in a warmer ocean. Global Change Biology 20, 2124-2139.

Clarke L.J., Beard J.M., Swadling K.M. and Deagle B.E. (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution 7, 873-883.

Coissac E., Riaz T. and Puillandre N. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. Molecular Ecology 21, 1834-1847.

Collocott S.J. (2016) Patterns and influencing factors of the larval fish assemblage of the KwaZulu-Natal Bight, South Africa. MSc Thesis, University of KwaZulu-Natal, South Africa.

Cowen R., Gawarkiewicz G., Pineda J., Thorrold S.R. and Werner F.E. (2007) Population connectivity in marine systems: An overview. Oceanography 20, 14-21.

Cowen R.K. and Sponaugle S. (2009) Larval dispersal and marine population connectivity. Annual Review of Marine Science 1, 443-466.

Creer S., Deiner K., Frey S., Porazinska D., Taberlet P., Thomas W.K., Potter C. and Bik H.M. (2016) The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7, 1008-1018.

Cristescu M.E. (2014) From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. Trends in Ecology and Evolution 29, 566-571.

De Decker A. (1964) Observations on the ecology and distribution of copepoda in the marine plankton of South Africa. Investigational Report, Division of Sea Fisheries, South Africa 49, 1-33.

De Decker A. (1984) Near-surface copepod distribution in the south-western Indian and south-eastern Atlantic Ocean. Annals of the South African Museum 93, 303-370.

De Decker A. and Mombeck F. (1964) South African contribution to the International Indian Ocean Expedition. A preliminary report on the planktonic Copepods. Investigational Report, Division of Sea Fisheries, South Africa 51, 10-67.

Djurhuus A., Pitz K., Sawaya N.A., Rojas-Márquez J., Michaud B., Montes E., Muller-Karger F. and Breitbart M. (2018) Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. Limnology and Oceanography: Methods 16, 209-221.

Dong W., Liu H., Xu C., Zuo Y., Chen Z. and Zhou S. (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. BMC Genetics 15, 138.

Dormontt E.E., van Dijk K.J., Bell K.L., Biffin E., Breed M.F., Byrne M., Caddy-Retalic S., Encinas-Viso F., Nevill P.G., Shapcott A., Young J.M., Waycott M. and Lowe A.J. (2018) Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections – An Australian perspective. Frontiers in Ecology and Evolution 6, 134.

Elbrecht V. and Leese F. (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass - sequence relationships with an innovative metabarcoding protocol. PLOS One 10, e0130324.

Elbrecht V. and Leese F. (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science 5, 11.

Elbrecht V. and Steinke D. (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology 64, 380-387.

Ficetola G.F., Taberlet P. and Coissac E. (2016) How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources 16, 604-607.

Folmer O., Black M., Hoeh W., Lutz R. and Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Fraija-Fernández N., Bouquieaux M.C., Rey A., Mendibil I., Cotano U., Irigoien X., Santos M. and Rodríguez-Ezpeleta N. (2020) Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. Ecology and Evolutionn 10, 7560-7584.

Gibson J., Shokralla S., Porter T.M., King I., van Konynenburg S., Janzen D.H., Hallwachs W. and Hajibabaei M. (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proceedings of the National Academy of Sciences 111, 8007-8012.

Goetze E. and Jungbluth M.J (2013) Acetone preservation for zooplankton molecular studies. Journal of Plankton Research 35, 972-981.

Goodwin S., McPherson J.D. and McCombie W.R. (2016) Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics 17, 333-351.

Gous A., Swanevelder D.Z.H., Eardley C.D. and Willows-Munro S. (2019) Plant–pollinator interactions over time: Pollen metabarcoding from bees in a historic collection. Evolutionary Applications 12, 187-197.

Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

Govender A., Singh S., Groeneveld J., Pillay S. and Willows-Munro S. (In review) Marine zooplankton, mini-barcodes, and DNA metabarcoding: The case for taxon-specific primers.

Green A. (2009) Sediment dynamics on the narrow, canyon-incised and current-swept shelf of the northern KwaZulu-Natal continental shelf, South Africa. Geo-Marine Letters 29, 201-219.

Green A. (2011) Submarine canyons associated with alternating sediment starvation and shelf-edge wedge development: Northern KwaZulu-Natal continental margin, South Africa. Marine Geology 284, 114-126.

Green A. and Uken R. (2008) Submarine landsliding and canyon evolution on the northern KwaZulu-Natal continental shelf, South Africa, SW Indian Ocean. Marine Geology 254, 152-170.

Greene C.H. (1983) Selective predation in freshwater zooplankton communities. Internationale Revue der gesamten Hydrobiologie und Hydrographie 68, 297-315.

Guastella L. and Roberts M. (2016) Dynamics and role of the Durban cyclonic eddy in the KwaZulu-Natal Bight ecosystem. African Journal of Marine Science 38, S23-S42.

Gutierrez M.F., Tavsanoglu U.N., Vidal N., Yu J., Teixeira de Mello F., Cakiroglu A., He H., Liu Z. and Jeppesen E. (2018) Salinity shapes zooplankton communities and functional diversity and has complex effects on size structure in lakes. Hydrobiologia 813, 237-255.

Hajibabaei M., Porter T.M., Wright M. and Rudar J. (2019) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLOS One 14, e0220953.

Hajibabaei M., Shokralla S., Zhou X., Singer G.A.C. and Baird D.J. (2011) Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLOS One 6, e17497.

Hajibabaei M., Smith M.A., Janzen D.H., Rodriguez J.J., Whitfield J.B. and Hebert P.D.N. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6, 959-964.

Havens K.E. and Hanazato T. (1993) Zooplankton community responses to chemical stressors: A comparison of results from acidification and pesticide contamination research. Environmental Pollution 82, 277-88.

Hays G.C., Richardson A.J. and Robinson C. (2005) Climate change and marine plankton. Trends in Ecology and Evolution 20, 337-344.

Hebert P. and Gregory T.R. (2005) The promise of DNA barcoding for taxonomy. Systematic Biology 54, 852-859.

Hebert P.D.N., Cywinska A., Ball S.L. and deWaard J.R. (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270, 313-21.

Hidalgo M., Kaplan D.M., Kerr L.A., Watson J.R., Paris C.B. and Browman H.I. (2017) Advancing the link between ocean connectivity, ecological function and management challenges. ICES Journal of Marine Science 74, 1702-1707.

Hoegh-Guldberg O. and Bruno J.F. (2010) the impact of climate change on the world's marine ecosystems. Science 328, 1523-1528.

Holman L.E., de Bruyn M., Creer S., Carvalho G., Robidart J. and Rius M. (2019) Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. Scientific Reports 9, 11559.

Huggett J. and Kyewalyanga M. (2017) Ocean productivity. In: The RV Dr Fridtjof Nansen in the Western Indian Ocean: Voyages of marine research and capacity development. Groeneveld J.C. and Koranteng K.A. (eds). pp. 55-80. Food and Agriculture Organization, Rome.

Jackson J.B.C. (2010) The future of the oceans past. Philosophical Transactions of The Royal Society B Biological Sciences 365, 3765-3778.

Jones G.P., Almany G.R., Russ G.R., Sale P.F., Steneck R.S., van Oppen M.J.H. and Willis B.L. (2009) Larval retention and connectivity among populations of corals and reef fishes: History, advances and challenges. Coral Reefs 28, 307-325.

Kelly P., Clementson L., Davies C., Corney S. and Swadling K. (2016) Zooplankton responses to increasing sea surface temperatures in the southeastern Australia global marine hotspot. Estuarine, Coastal and Shelf Science 180, 242-257.

Laakmann S., Blanco-Bercial L. and Cornils A. (2020) The crossover from microscopy to genes in marine diversity: From species to assemblages in marine pelagic copepods. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190446.

Leray M. and Knowlton N. (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences of the United States of America 112, 2076-2081.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T. and Machida R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10, 34.

Lutjeharms J. (2007) Three decades of research on the greater Agulhas Current. Ocean Science 3, 129-147.

Lutjeharms J.R.E. (2006) The Agulhas Current. Volume 329. Springer, Berlin.

Macher J.N., Hoorn B., Peijnenburg K., Walraven L. and Renema W. (2020) Metabarcoding reveals different zooplankton communities in northern and southern areas of the North Sea. bioRxiv 2020.07.23.218479.

Marquina D., Andersson A.F. and Ronquist F. (2019) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19, 90-104.

Meusnier I., Singer G.A., Landry J.F., Hickey D.A., Hebert P.D. and Hajibabaei M. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214.

Moore M. and Folt C. (1993) Zooplankton body size and community structure: Effects of thermal and toxicant stress. Trends in Ecology and Evolution 8, 178-183.

Pattrick P. and Strydom N.A. (2014) Larval fish variability in response to oceanographic features in a nearshore nursery area. Journal of Fish Biology 85, 857-881.

Paturej E. and Gutkowska A. (2015) The effect of salinity levels on the structure of zooplankton communities. Archives of Biological Sciences 67, 483-492.

Piper A.M., Batovska J., Cogan N.O.I., Weiss J., Cunningham J.P., Rodoni B.C. and Blacket M.J. (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience 8, giz092.

Pitz K.J., Guo J., Johnson S.B., Campbell T.L., Zhang H., Vrijenhoek R.C., Chavez F.P. and Geller J. (2020) Zooplankton biogeographic boundaries in the California Current System as determined from metabarcoding. PLOS One 15, e0235159.

Porter T.M., Morris D.M., Basiliko N., Hajibabaei M., Doucet D., Bowman S., Emilson E.J.S., Emilson C.E., Chartrand D., Wainio-Keizer K., Séguin A. and Venier L. (2019) Variations in terrestrial arthropod DNA metabarcoding methods recovers robust beta diversity but variable richness and site indicators. Scientific Reports 9, 18218.

Pretorius M., Huggett J. and Gibbons M. (2016) Summer and winter differences in zooplankton biomass, distribution and size composition in the KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S155-S168.

Radulovici A.E., Archambault P. and Dufresne F. (2010) DNA barcodes for marine biodiversity: Moving fast forward? Diversity 2, 450-472.

Ratnasingham S. and Hebert P.D.N. (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7, 355-364.

Richardson A. (2009) Plankton and climate. In: Elements of Physical Oceanography: A derivative of the Encyclopedia of Ocean Sciences. Steele J.H., Thorpe S.A. and Turekian K.K. (eds). pp. 397 – 406. Academic Press, Cambridge.

Richardson A.J. (2008) In hot water: Zooplankton and climate change. ICES Journal of Marine Science 65, 279-295.

Roberts M.J., Nieuwenhuys C. and Guastella L.A. (2016) Circulation of shelf waters in the KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S7-S21.

Rodgher S., Lombardi A.T. and Melão M.d.G.G. (2009) Evaluation onto life cycle parameters of *Ceriodaphnia silvestrii* submitted to 36 days dietary copper exposure. Ecotoxicology and Environmental Safety 72, 1748-1753.

Ruppert K., Kline R. and Rahman M. (2019) Past, present and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring and applications of global eDNA. Global Ecology and Conservation 17, e00547.

Schleyer M. (1985) Chaetognaths as indicators of water masses in the Agulhas Current system. Investigational Report No. 61. Durban: Oceanographic Research Institute.

Schumann E. (1987) The coastal ocean off the east coast of South Africa. Transactions of the Royal Society of South Africa 46, 215-229.

Shokralla S., Spall J., Gibson J. and Hajibabaei M. (2012) Next-generation sequencing technologies for enviromental DNA research. Molecular Ecology 21, 1794-1805.

Singh S., Groeneveld J., Huggett J., Naidoo D., Cedras R. and Willows-Munro S. (2021) DNA metabarcoding of marine zooplankton in South Africa: how good is the reference library? African Journal of Marine Science (In press).

Sink K.J., van der Bank M.G., Majiedt P.A., harris L.R., Atkinson L.J., Kirkman S.P. and Karenyi N. (2019) South African national biodiversity assessment 2018 technical report volume 4: Marine realm. South African National Biodiversity Institute, Pretoria, South Africa.

Somervuo P., Yu D.W., Xu C.C.Y., Ji Y., Hultman J., Wirta H. and Ovaskainen O. (2017) Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution 8, 398-407.

Taberlet P., Coissac E., Pompanon F., Brochmann C. and Willerslev E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21, 2045-2050.

Thibault-Botha D. and Gibbons M.J. (2005) Epipelagic siphonophores off the east coast of South Africa. African Journal of Marine Science 27, 129-139.

Thibault-Botha D., Lutjeharms J.R. and Gibbons M. (2004) Siphonophore assemblages along the east coast of South Africa; mesoscale distribution and temporal variations. Journal of Plankton Research 26, 1115-1128.

Thomsen P.F., Kielgast J., Iversen L.L., Wiuf C., Rasmussen M., Gilbert M.T., Orlando L. and Willerslev E. (2012) Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21, 2565-2573.

Thomsen P.F. and Sigsgaard E.E. (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. Ecology and evolution 9, 1665-1679.

Trivedi S., Aloufi A.A., Ansari A.A. and Ghosh S.K. (2016a) Role of DNA barcoding in marine biodiversity assessment and conservation: An update. Saudi Journal of Biological Sciences 23, 161-171.

Trivedi S., Rehman H., Saggu S., Panneerselvam C., Abbas Z.K., Ahmad I., Ansari A.A. and Ghosh S.K. (2016b) DNA barcoding in the marine habitat: An overview. In: DNA barcoding in marine perspectives: Assessment and conservation of biodiversity. Trivedi S.,

Ansari A.A., Ghosh S.K. and Rehman H. (eds). pp. 3-28. Springer International Publishing, Cham.

WoRMS Editorial Board (2021). World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2021-01-02. doi:10.14284/170

Zhang G.K., Chain F.J.J., Abbott C.L. and Cristescu M.E. (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11, 1901-1914.

Chapter Two: The design and testing of mini-barcode markers in marine lobsters

This chapter has been published as (Appendix 2.1): Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

2.1 Abstract

Full-length mitochondrial cytochrome c oxidase I (COI) sequence information from lobster phyllosoma (larvae) can be challenging to obtain when DNA is degraded or fragmented. Primers that amplify smaller fragments are more useful in DNA metabarcoding studies. This study developed and tested a method to design a taxon-specific mini-barcode primer set for marine lobsters. The shortest, most informative portion of the COI gene region was identified *in silico*, and DNA barcode gap analysis was performed to assess its reliability as a species diagnostic marker. Primers were designed, and cross-species amplification success was tested on DNA extracted from a taxonomic range of spiny-, clawed-, slipper- and blind lobsters. The mini-barcode primers successfully amplified both adult and phyllosoma COI fragments and were able to successfully delimit all species analyzed. Previously published universal primer sets were also tested and sometimes failed to amplify COI from phyllosoma samples. The newly designed taxon-specific mini-barcode primers will increase the success rate of species identification in bulk environmental samples and add to the growing DNA metabarcoding toolkit.

Keywords: Lobster phyllosoma, marine lobsters, mini-barcode, DNA metabarcoding

2.2 Introduction

Holthuis (1991) provided a detailed systematic catalogue of nearly all the marine lobsters known up to the early 1990s, based solely on the morphology of adult specimens. The traditional classification system used in the catalogue recognized the superfamilies; Nephropoidea (clawed lobsters), Palinuroidea (spiny and slipper lobsters), Eryonoidea (blind lobsters) and the living fossil Glypheoidea within the decapod suborder Macrura Reptantia (Holthuis 1991). More recently, Chan (2010) updated the list of valid species by adding several newly described taxa and organized all living marine lobsters into four infraorders: Astacidea, Glypheidea, Achelata and Polychelida.

The list recognized six families, 55 genera and 248 species (with four subspecies) of marine lobsters (Chan 2010).

Marine lobsters have cryptic early life-history stages. Larvae (called phyllosomas) hatch from eggs carried ventrally on the abdomen of the female and are then dispersed as meroplankton by water movements (Phillips and Sastry 1980). Phyllosomas are dorso-ventrally flattened, leaf-like, transparent and moult through a series of developmental stages of increasing size and morphological complexity (Booth and Phillips 1994). The final phyllosoma stage undergoes a metamorphic molt into a post-larva (or puerulus), which settles on the substrate to begin a benthic existence. Early benthic juveniles are morphologically similar to adult lobsters and can readily be identified to species level (Booth and Phillips 1994). In contrast, phyllosomas are challenging to distinguish because they are morphologically cryptic and have not yet been fully described for all extant taxa (Berry 1974; Prasad *et al.* 1975).

DNA barcoding is now well established as a species identification technique and discovery (Hebert *et al.* 2003). It relies on short, standardized nucleotide sequences (DNA barcodes) as internal species tags and searchable online sequence repositories, such as the Barcode of Life Data Systems (BOLD, www.barcodeoflife.org), International Nucleotide Sequence Database Collaboration (www.insdc.org) and the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/). DNA barcoding augments traditional taxonomy methods (Lambert *et al.* 2005) and is beneficial for distinguishing cryptic or polymorphic species such as marine lobsters and associating life history stages of unknown identities such as eggs or larvae with identifiable adult stages (Schander and Willassen 2005).

The DNA barcode used in most animal groups is a 658-base pair (bp) protein-coding region of the mitochondrial cytochrome c oxidase 1 (COI) gene (Hebert *et al.* 2003). This region has been successfully used to identify adult lobsters and phyllosomas (Jeena *et al.* 2016; Palero *et al.* 2016), but amplification success of phyllosomas is often low (Chow *et al.* 2006), possibly due to rapid post-capture DNA degradation due to poor preservation and/storage or primer bias (Bar *et al.* 1988; Boyer *et al.* 2012). Several studies have consequently relied on 16S rRNA or 18S rRNA gene regions to obtain higher amplification success rates (Palero *et al.* 2009; O'Rorke *et al.* 2013; Bracken-Grissom *et al.* 2014; Genis-Armero *et al.* 2017). A higher COI amplification success rate

can be obtained by designing a shorter informative section of the gene region known as a DNA mini-barcode.

DNA mini-barcodes need to be sufficiently informative and preferentially target hypervariable DNA regions to accurately delimit species (Hajibabaei *et al.* 2006; Leray *et al.* 2013). The reliability of mini-barcodes relies on the presence of a 'barcode gap', or the difference between inter- and intraspecific genetic distances within a group of organisms (Barrett and Hebert 2005; Meyer and Paulay 2005). Several studies have successfully designed and tested mini-barcodes in species from a wide taxonomic range, for example, moths (Meusnier *et al.* 2008), Australian mammals (Modave *et al.* 2017) and Indian snakes (Dubey *et al.* 2011). Meusnier *et al*. (2008) developed a 'universal' mini-barcode from the standard COI gene region to successfully identify a range of mammals, fishes, birds and insects from archival samples. Nevertheless, a limitation of universal primers is that the most informative portion of the COI region is not the same for all taxa (Arif *et al.* 2011). Hence, taxon-specific primers tend to have higher PCR amplification and sequencing success rates and offer higher discriminating power than universal primers (Dong *et al.* 2014).

Advancements in next-generation sequencing and DNA metabarcoding encourages the development of primers for taxon-specific mini-barcodes which will, in turn, improve the efficiency and accuracy of taxon discovery and identification (Leray *et al.* 2013), especially in bulk samples such as mixed zooplankton collected using plankton tow-nets. The taxonomic coverage of the primer sets can then be used in a tree-of-life approach for ecosystem biomonitoring (Meusnier *et al.* 2008; Stat *et al.* 2017). This study used an *in-silico* method to identify the shortest, most informative portion of the COI gene region in marine lobsters and used it to design a taxon-specific mini-barcode. The reliability of the mini-barcode as an identification tool was tested using the DNA barcode gap analysis. The cross-species amplification of the mini-barcode primers was tested on tissue samples of a broad range of lobster taxa, including species of spiny- (Palinuridae), clawed- (Nephropidae), slipper- (Scyllaridae) and blind lobsters (Polychelidae). The amplification success of the mini-barcode primer set on phyllosomas was compared with that of primers already available in the literature.

2.3 Materials and methods

A total of 350 lobster COI sequences were downloaded from GenBank and BOLD (www.ncbi.nlm.nih.gov/genbank, date accessed: 02-05-2017; http://www.boldsystems.org/, date accessed: 02-05-2017) (Appendix 2.2). Where available, individuals from different geographical regions were included in the dataset to accommodate potential phylogeographic structure within recognized species. The final dataset included 175 species belonging to 42 genera and 4 families, covering some 71% of known marine lobster species, 76% of the genera and 67% of the families listed by Chan (2010). The sequences were aligned using Clustal X2.1 (Larkin *et al.* 2007) and optimized manually to ensure homology using Bioedit 7.2.5 (Hall 1999). The number of variable- (V), parsimony informative characters (Pi) and the average nucleotide composition were estimated for the data (full-length and mini-barcode alignments) using MEGA 6.0 (Tamura *et al.* 2013).

Mini-barcode fragments were estimated using sliding window analysis (SWAN) (Proutski and Holmes 1998) in the Species Identity and Evolution (SPIDER) (Brown *et al.* 2012) package in R (http://www.r-project.org). The slideAnalyses function was used to generate windows varying in size from 100 to 230 base pairs (bp). Windows were shifted along the length of the COI alignment using 10 bp intervals. The top two mini-barcode fragments for each window length were selected for further analyses based on: (1) high mean Kimura 2-parameter (K2P) distance; (2) few zero pairwise non-conspecific distances; and (3) high proportion of clades shared between the neighbor-joining tree from the full-length DNA sequence alignment and the tree constructed using only data from selected windows. From this analysis, a total of 28 potential mini-barcode alignments were created.

Maximum likelihood analysis was conducted on the 29 datasets (1 full-length reference dataset and 28 SWAN mini-barcodes) using Garli 0.951 (Zwickl 2006). In all analyses, the K2P model of sequence evolution (Kimura 1980) was implemented as this is the model implemented on BOLD. The 28 mini-barcode maximum likelihood trees were then compared to the full-length reference tree using Ktreedist 1.0 (Soria-Carrasco *et al.* 2007). Ktreedist calculated K-scores (topology and branch length differences) and Robinson-Fouls symmetric difference (topological differences). For both methods, lower values indicated a high degree of similarity between the reference tree and the mini-barcode tree.

A DNA barcode gap analysis was conducted on the top-scoring mini-barcode dataset. Intra- and interspecific genetic distances were calculated using the K2P nucleotide substitution model in MEGA 6.0 and plotted. The maximum intraspecific distance was subtracted from the minimum interspecific distance to determine the barcoding gap (Meier *et al.* 2006) and the Jeffries-Matusita distance (J-M) statistic was used to test whether the intra- and interspecific genetic distance classes were separable. The J-M statistic considers the distance between the means of the intra- and interspecific genetic distances and the distribution of values from the mean (Dabboor *et al.* 2014). The J-M distance is asymptotic to 1.414 and as such, a value of 1.414 or greater suggests that intra- and interspecific genetic distances are statistically separable (Trigg and Flasse 2001).

Primers were designed flanking the top-scoring mini-barcode region (LobsterMinibarF:5'-GGWGATGAYCAAATTTAYAATGT–3' and LobsterMinibarR: 5'-CCWACTCCTCTTTCTACTATTCC –3'). Amplification and sequencing success were tested on both adult and phyllosoma samples of different lobster species. The adult and phyllosoma samples were obtained from the Oceanographic Research Institute, Durban, South Africa and were a mixture of freshly collected and preserved samples.The adult samples included: two from the family Nephropidae (*Metanephrops mozambicus, Nephropsis stewarti*), two from Scyllaridae (*Scyllarides elisabethae, Scyllarides squammosus*), six from Palinuridae, comprising of three genera, namely; *Panulirus* (*P. homarus, P. versicolor*), *Palinurus* (*P. gilchristi, P. delagoae*) and *Jasus* (*J. lalandii, J. paulensis*) and one from the family Polychelidae (*Polycheles typhlops*). The phyllosoma samples included three specimens from the family Palinuridae (*Panulirus ornatus, P. homarus,* and *P. homarus rubellus*) and five from the family Scyllaridae (*Scyllarus arctus, Petrarctus rugosus, Acantharctus ornatus, Scyllarus sp.* and *Petractus sp.*). These samples provided 67% coverage across the different families within the lobster taxonomy.

DNA from 17 species (adults and larvae combined = 19 samples) was extracted from pereiopod tissue using the Zymo Quick-DNA Universal Kit (Zymo Research), as per the manufacturer's protocol, which was modified to include an initial incubation step at 55 ºC overnight. PCR reactions were 25 µl in volume and contained 30 ng genomic DNA, 12.5 µl OneTaq Quick-Load Master Mix (1X, BioLabs, New England), 0.50 µl forward and reverse mini-barcode primer (10 µM each), 6.5 µl sterile nuclease-free water, 2 µl additional $MgCl_2$ (25 µM) and 2 µl Bovine Serum Albumin (BSA) (1 $mg.m^{-1}$) was added. All PCR reactions were run with a

negative control containing no DNA. The thermal cycling program included initial denaturation at 94 ºC for 2 minutes, followed by 35 cycles of denaturation at 94 ºC for 30 seconds, annealing at 46 ºC for 30 seconds and extension at 68 º C for 1 minute. The final extension step was carried out at 68 ºC for 5 minutes. PCR clean-up and sequencing reactions were performed at the Central Analytical Facilities (CAF) at the University of Stellenbosch (South Africa). All sequences were checked for their species specificity using the nucleotide BLAST tool (BLASTn) on GenBank and BOLD. Percentage identification between 92 – 100% was used to confirm the exact species match.

The amplification success of the lobster-specific mini-barcode primer set was compared to that of the standard COI primer set (expected size = 658 bp) (Folmer *et al.* 1994), a universal mini-barcode primer (expected size = 130 bp) (Meusnier *et al.* 2008) and internal COI mini-barcode primers (expected size = 313 – 319 bp) (Leray *et al.* 2013) (see Table S2.1). A graphical representation with the relative annealing sites and each primer's orientation on the COI barcode region can be seen in Figure S2.1. The internal mini-barcode primers designed by Leray *et al*. (2013) works in conjuncture with the Folmer *et al*. (1994) COI primers. PCR reactions were the same as above. Thermal cycling conditions can be found in Table S2.2. The PCR products were visualized on a 1.2% (w/v) TBE agarose gel containing 0.02% Ethidium Bromide (EtBr). A 100 bp molecular weight marker (Solis Biodyne) was used to estimate the size of PCR products. PCR clean-up and sequencing were carried out on successful amplifications.

2.4 Results

Smaller window sizes (100-160 bp from SWAN) had higher mean K2P distances and lower zero non-conspecific values in the K2P distance matrix. Larger window sizes (170-230 bp) showed better congruence of neighbor-joining trees (Table S2.3). Larger mini-barcode fragments generated lower K- and R-F scores when compared with the reference tree (Table S2.4). Based on these results, Fragment 230_b (position 109-339 of the full alignment) was selected as the best candidate for a lobster-specific mini-barcode.

The DNA barcode gap analysis was carried out on 350 DNA barcodes downloaded from Genbank and BOLD (Appendix 2.2), including two representative individuals per species. To test the impact of sample size per species, the analysis was carried out on a larger dataset with 2 – 5 individuals per species (Figure S2.2). Increasing the sample size did not significantly impact the DNA barcode gap analysis. The intra-specific K2P pairwise distances for the Fragment 230_b

alignment ranged from 0.00 to 0.01, while the inter-specific distances ranged from 0.02 to 0.36 (Figure 2.1). The position of the DNA barcode gap is between 0.01 and 0.02. The Jeffries-Matusita distance of 1.998 exceeds the significance thresholds and confirms that the intra- and interspecific distance classes are statistically separable.

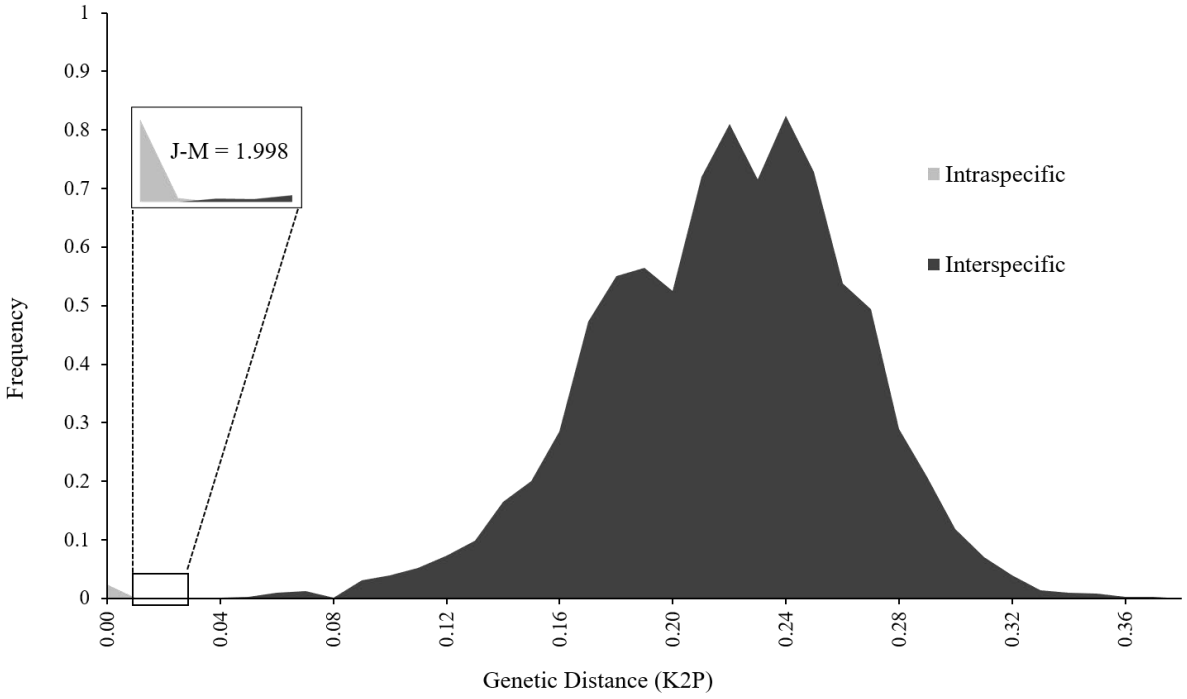


Figure 2.1 Frequency distributions of intra- and interspecific pairwise K2P distances calculated using the selected mini-barcode region (Fragment 230_b). The barcode gap (insert) lies between the genetic distances of 0.01 and 0.02.

The mini-barcode region was successfully amplified using the designed primers (LobsterMinibarF and LobsterMinibarR) across all 17 different species in both adult and phyllosoma samples (Figure S2.3). BLAST search results confirmed that 16 mini-barcode sequences were a match (percentage identification between 92 – 100% was used to confirm the exact species match) to the morphologically identified adult lobster and phyllosoma voucher specimens. In three cases (*N. stewarti, Scyllarus sp.* and *Petrarctus sp.*) no direct match could be found, because no sequences were available on GenBank or BOLD for these species. GenBank accession numbers are provided in Table S2.5.

PCR amplification was successful for all published primer pairs when tested on DNA extracted from adult *P. homarus* samples (Figure 2.2), but for phyllosomas, only the Folmer *et al*. (1994) primer set (LCOI490, HCO2198) produced a PCR product, for one of two specimens. In contrast, the lobster mini-barcode primers consistently amplified the DNA extracted from phyllosoma samples. BLAST searches performed on successfully amplified PCR products confirmed that all sequences were *P. homarus*, with a match of 92 – 99% (Table S2.6).

Given that our mini-barcode primer set was selected to amplify the most variable portion of the COI gene, the < 100% sequence match with GenBank data is probably due to a combination of genetic diversity within species and limited COI data for these species currently uploaded. As more data from more species and populations are uploaded to GenBank or BOLD species identification will become more accurate.

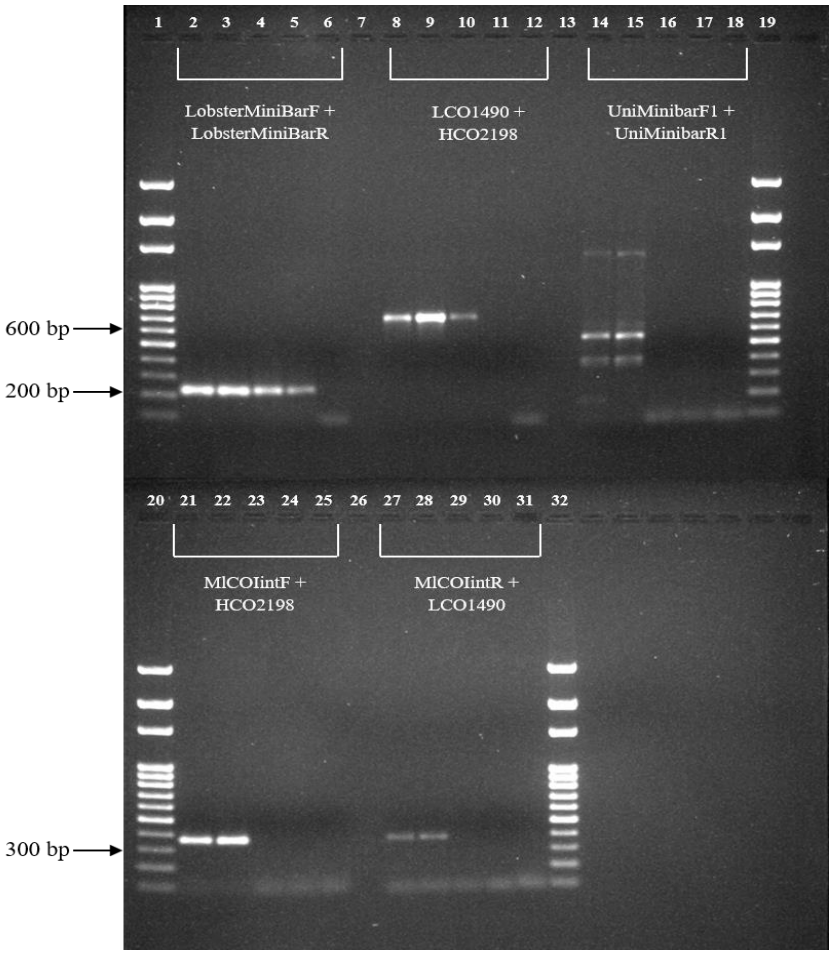


Figure 2.2 Amplification of four different cytochrome oxidase (COI) primer pairs tested in this study. Lanes 2, 3, 8, 9, 14, 15, 21, 22, 27 and 28 are PCR products recovered from DNA

extracted from adult *Panulirus homarus*. Lane 4, 5, 10, 11, 16, 17, 23, 24, 30 and 31 are PCR products recovered from DNA extracted from phyllosoma samples. Lane 6, 12, 18, 25 and 31 are PCR negative controls. Lane 1, 19, 20 and 32 are 100 bp molecular weight marker (Solis Biodyne). Lane 7, 13 and 26 had no samples loaded.

2.5 Discussion

The lobster specific mini-barcode primer pair designed in this study was tested on a taxonomically diverse set of marine lobsters from seven different genera. It consistently amplified COI from both adults and phyllosomas across all taxa and outperformed published primers. Confirmation of a barcode gap highlights its value as a diagnostic tool that can be used to match phyllosomas to species (Meier *et al.* 2006). The sliding window analysis method (Proutski and Holmes 1998) accurately identified the shortest, most informative portion of the COI gene of marine lobsters.

The taxon-specific mini-barcode primers were designed in response to the repeated low amplification success rate of the standard COI gene region in DNA extracted from lobster phyllosomas. Given our broader objective, to design primers that can account for the presence of marine lobsters using DNA metabarcoding on unsorted zooplankton samples, a high amplification success for lobster larvae was considered to be crucial. The phyllosoma samples tested in this study were obtained from plankton tows at sea and stored in 95% ethanol at -20°C. COI often failed to amplify completely using published primers, or the sequences obtained were messy and terminated abruptly. The effects of suboptimal field sampling conditions (temperature and pH fluctuations and contamination) (Bar *et al.* 1988), post-survey sorting of plankton samples, or inadequate preservation of biological material (whole phyllosomas) before DNA extraction may have contributed to the degradation or fragmentation of DNA (Boyer *et al.* 2012; Hajibabaei and McKenna 2012). Despite ethanol being used as the most common preservation technique in barcoding studies, it has some drawbacks as there have been many reports of DNA degradation in ethanol preserved samples with increasing loss of DNA over time at warmer temperatures and in samples with higher water content (Goetze *et al.* 2013). Stein *et al.* (2013) carried out a series of tests using different ethanol preservation techniques and found that if samples were initially preserved in 95% ethanol, successful sequencing of COI barcodes was not affected when transferring samples to 70% ethanol with a hold time of up to six months. However, the researchers

found varying success rates across different taxa which was most likely due to poor PCR primer efficiency.

Other studies have had similar difficulties with the amplification of COI from larval material using different organisms. A study based on Antarctic larval marine invertebrates which were alcohol-preserved encountered a low amplification success of 22%, despite using 18S RNA, COI and 16S RNA primer sets (Webb *et al.* 2006). The authors suggested using a taxon-specific primer to increase PCR amplification success rate. Baird *et al.* (2011) created a COI reference library for freshwater benthic macroinvertebrate specimens ranging in preservation time between <1 and 23 years old, but this yielded only 2.9% full-length usable barcodes (Baird *et al.* 2011). Adding a universal mini-barcode primer increased the yield to 17.5%, and it was concluded that the DNA was likely degraded, because samples were collected and fixed in formalin in the field and thereafter transferred to 70% ethanol for long-term storage. Formalin preservation is widely used for animal specimens, however, formalin fixation cross-links DNA, compromising downstream applications such as sequencing and PCR, making this method inappropriate for DNA (meta)barcoding studies (Taleb-Hossenkhan *et al.* 2013).

Hajibabaei *et al*. (2006) used both *in silico* and *in vitro* tests to examine the accuracy of mini-barcodes in species identification of century-old museum samples. Mini-barcodes of varying lengths were tested *in silico* on Australian fish and lepidoptera sequences and found to be as accurate as full-length barcodes. *In vitro* tests were subsequently carried out on museum specimens with varied age, preservation methods and taxonomic scope. Primers designed for the mini-barcodes had a success rate of > 90% after sequencing, compared to the 50% using full-length primers. Hence, mini-barcode primers that amplify a smaller COI region can improve barcoding success where DNA is degraded.

The lobster mini-barcode designed in this study returned a higher amplification success for lobster phyllosomas than the universal mini-barcode (Meusnier *et al.* 2008). Internal mini-barcodes (Leray *et al.* 2013) designed to work in conjunction with the commonly used COI primer set (Folmer *et al.* 1994) were also tested in addition to degenerate versions of the universal COI primer set (Meyer 2003; Geller *et al.* 2013). The forward internal primer combined with the reverse COI primer and its degenerate versions had the highest amplification success. Nevertheless, when tested on lobster phyllosomas in the present study, these internal primers failed to amplify COI.

The emergence of DNA metabarcoding techniques combined with high throughput next-generation sequencing provides a powerful new tool for biodiversity assessments from environmental samples (Ji *et al.* 2013). DNA metabarcoding can increase the speed, accuracy and resolution of species identification while allowing for cost-effective biodiversity monitoring. For example, zooplankton in the marine environment (including phyllosomas of various lobster species) are model organisms for monitoring trends in ecosystem health and biodiversity in the face of climate change and habitat degradation because they exhibit a rapid response to environmental change (Bucklin *et al.* 2016). Within this context, taxa that are important to fisheries (i.e., decapods such as marine lobsters, crabs and prawns, or fish) can be selected as indicator species when analyzing mixed zooplankton samples.

The efficiency and accuracy of DNA metabarcoding for taxonomic detection and identifications rely on specifically targeted barcodes that are taxonomically informative (Liu *et al.* 2008) and on suitable primer sets for amplifying hypervariable DNA regions from target organisms (Leray *et al.* 2013). The method used to develop mini-barcodes for lobsters in our study can easily be applied to other taxa - for example, crabs, prawns, shrimps, or fish. Identifying the shortest, most variable portions of the genome are particularly relevant in applications involving next-generating sequencing technologies, such as Illumina with limited read length.

To conclude, studies have highlighted the need for multiple DNA metabarcoding assays to catalogue biodiversity, including universal and multiple taxon-specific assays (Stat *et al.* 2017). From this perspective, the use of taxon-specific mini-barcodes is encouraged because in combination, they can maximize richness estimates and increase the possibility of recovering amplicons from degraded DNA (Meusnier *et al.* 2008; Stat *et al.* 2017).

Data Accessibility

All the scripts and sequences used to create the mini-barcodes in this chapter are available on figshare: https://doi.org/10.6084/m9.figshare.14378663.v1

2.6 References

Arif I., Khan H., Sadoon M. and Shobrak M. (2011) Limited efficiency of universal mini-barcode primers for DNA amplification from desert reptiles, birds and mammals. Genetics and Molecular Research 10, 3559-3594.

Baird D.J., Pascoe T.J., Zhou X. and Hajibabaei M. (2011) Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: Formalin preservation vs. specimen age. Journal of the North American Benthological Society 30, 125-130.

Bar W., Kratzer A., Machler M. and Schmid W. (1988) Postmortem stability of DNA. Forensic Science International 39, 59-70.

Barrett R.D.H. and Hebert P.D.N. (2005) Identifying spiders through DNA barcodes. Canadian Journal of Zoology 83, 481-491.

Berry P.F. (1974) Palinurid and scyllarid lobster larvae of the Natal Coast, South Africa. Oceanographic Research Institute, Durban. Investigational report no. 34: 1-44.

Booth J.D. and Phillips B.F. (1994) Early life history of spiny lobster. Crustaceana 66, 271-294.

Boyer S., Brown S.D.J., Collins R.A., Cruickshank R.H., Lefort M.C., Malumbres-Olarte J. and Wratten S.D. (2012) Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA. PLOS One 7, e38215.

Bracken-Grissom H.D., Ahyong S.T., Wilkinson R.D., Feldmann R.M., Schweitzer C.E., Breinholt J.W., Bendall M., Palero F., Chan T.Y., Felder D.L., Robles R., Chu K.H., Tsang L.M., Kim D., Martin J.W. and Crandall K.A. (2014) The emergence of lobsters: Phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). Systematic Biology 63, 457-479.

Brown S.D.J., Collins R.A., Boyer S., Lefort M.C., Malumbres-Olarte J., Vink C.J. and Cruickshank R.H. (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources 12, 562-565.

Bucklin A., Lindeque P.K., Rodriguez-Ezpeleta N., Albaina A. and Lehtiniemi M. (2016) Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. Journal of Plankton Research 38, 393-400.

Chan T.Y. (2010) Annotated checklist of the world's marine lobsters (Crustacea: Decapoda: Astacidea, Glypheidea, Achelata, Polychelida). The Raffles Bulletin of Zoology 23, 153-181.

Chow S., Suzuki N., Imai H. and Yoshimura T. (2006) Molecular species identification of spiny lobster phyllosoma larvae of the genus *Panulirus* from the Northwestern Pacific. Marine Biotechnology 8, 260-267.

Dabboor M., Howell S., Shokr M. and Yackel J. (2014) The Jeffriesa-Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. International Journal of Remote Sensing 35, 6859-6873.

Dong W., Liu H., Xu C., Zuo Y., Chen Z. and Zhou S. (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genetics 15, 138.

Dubey B., Meganathan P.R. and Haque I. (2011) DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species. Forensic Science International Genetics 5, 181-184.

Folmer O., Black M., Hoeh W., Lutz R. and Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Geller J., Meyer C., Parker M. and Hawk H. (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology Resources 13, 851-861.

Genis-Armero R., Guerao G., Abelló P., González-Gordillo J.I., Cuesta J.A., Corbari L., Clark P.F., Capaccioni-Azzati R. and Palero F. (2017) Possible amphi-Atlantic dispersal of Scyllarus lobsters (Crustacea: Scyllaridae): Molecular and larval evidence. Zootaxa 4306, 325.

Goetze E. and Jungbluth M.J (2013) Acetone preservation for zooplankton molecular studies. Journal of Plankton Research 35, 972-981.

Hajibabaei M. and McKenna C. (2012) DNA Mini-barcodes. In: DNA Barcodes. Methods in Molecular Biology (Methods and Protocols). Kress W. and Erickson D. (eds). Humana Press, Totowa.

Hajibabaei M., Smith M.A., Janzen D.H., Rodriguez J.J., Whitfield J.B. and Hebert P.D.N. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6, 959-964.

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.

Hebert P.D.N., Cywinska A., Ball S.L. and deWaard J.R. (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270, 313-321.

Holthuis L.B. (1991) Marine lobsters of the world: An annotated and illustrated catalogue of species of interest to fisheries known to date. Food and Agriculture Organization of the United Nations, Rome.

Jeena N.S., Gopalakrishnan A., Radhakrishnan E.V., Kizhakudan J.K., Basheer V.S., Asokan P.K. and Jena J.K. (2016) Molecular phylogeny of commercially important lobster species from Indian coast inferred from mitochondrial and nuclear DNA sequences. Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis 27, 2700-2709.

Ji Y., Ashton L., Pedley S.M., Edwards D.P., Tang Y., Nakamura A., Kitching R., Dolman P.M., Woodcock P., Edwards F.A., Larsen T.H., Hsu W.W., Benedick S., Hamer K.C., Wilcove D.S., Bruce C., Wang X., Levi T., Lott M., Emerson B.C. and Yu D.W. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16, 1245-1257.

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

Lambert D.M., Baker A., Huynen L., Haddrath O., Hebert P.D.N. and Millar C.D. (2005) Is a large-scale DNA-based inventory of ancient life possible? Journal of Heredity 96, 279-284.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T. and Machida R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10, 34.

Liu Z., DeSantis T.Z., Andersen G.L. and Knight R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Research 36, e120.

Meier R., Shiyang K., Vaidya G. and Ng P.K.L. (2006) DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Systematic Biology 55, 715-728.

Meusnier I., Singer G.A., Landry J.F., Hickey D.A., Hebert P.D. and Hajibabaei M. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214.

Meyer C.P. (2003) Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biological Journal of the Linnean Society 79, 401-459.

Meyer C.P. and Paulay G. (2005) DNA barcoding: Error rates based on comprehensive sampling. PLOS Biology 3, e422.

Modave E., MacDonald A.J. and Sarre S.D. (2017) A single mini-barcode test to screen for Australian mammalian predators from environmental samples. Gigascience 6, 1-13.

O'Rorke R., Jeffs A.G., Fitzgibbon Q., Chow S. and Lavery S. (2013) Extracting DNA from whole organism homogenates and the risk of false positives in PCR based diet studies: A case study using spiny lobster larvae. Journal of Experimental Marine Biology and Ecology 441, 1-6.

Palero F., Crandall K.A., Abelló P., Macpherson E. and Pascual M. (2009) Phylogenetic relationships between spiny, slipper and coral lobsters (Crustacea, Decapoda, Achelata). Molecular Phylogenetics and Evolution 50, 152-162.

Palero F., Genis R., Hall M. and F. Clark P. (2016) DNA barcoding the phyllosoma of Scyllarides squammosus (H. Milne Edwards, 1837) (Decapoda: Achelata: Scyllaridae). Zootaxa 4139, 481-498.

Phillips B.F. and Sastry A.N. (1980) Larval ecology. In: The biology and Management of Lobsters. Conn J.S. and Phillips B.F. (eds.). pp. 11-58. Academic Press, New York.

Prasad R.R., Tampi P.R.S. and George M.J. (1975) Phyllosoma larvae from the Indian Ocean collected by Dana Expedition 1928-1930. Journal of the Marine Biological Association of India 17, 56-107.

Proutski V. and Holmes E. (1998) SWAN: Sliding window analysis of nucleotide sequence variability. Bioinformatics 14, 467-468.

Schander C. and Willassen E. (2005) What can biological barcoding do for marine biology? Marine Biology Research 1, 79-83.

Soria-Carrasco V., Talavera G., Igea J. and Castresana J. (2007) The K tree score: Quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics 23, 2954-2956.

Stat M., Huggett M.J., Bernasconi R., DiBattista J.D., Berry T.E., Newman S.J., Harvey E.S. and Bunce M. (2017) Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. Scientific Reports 7, 12240.

Stein A.D., White B.P., Mazor R.D., Miller P.E. and Pilgrim E.M. (2013) Evaluating ethanol-based sample preservation to facilitate use of dna barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. PLoS ONE 8, e51273.

Taleb-Hossenkhan N., Bhagwant S. and Gourrege N. (2013) Extraction of nucleic acids from ancient formalin- and ethanol-preserved specimens of the tapeworm *Bertiella studeri:* which method works best? The Journal of Parasitology 99, 410-416.

Tamura K., Stecher G., Peterson D., Filipski A. and Kumar S. (2013) MEGA 6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30, 2725-2729.

Trigg S. and Flasse S. (2001) An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. International Journal of Remote Sensing 22, 2641-2647.

Webb K.E., Barnes D.K.A., Clark M.S. and Bowden D.A. (2006) DNA barcoding: A molecular tool to identify Antarctic marine larvae. Deep Sea Research Part II: Topical Studies in Oceanography 53, 1053-1060.

Zwickl D. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin, United States.

2.7 Supplementary data

Table S2.1 Primer information table for the standard COI region, universal mini-barcode, internal COI mini-barcode and lobster mini-barcode.

Primers	Forward	Reverse	Expected amplicon size for each primer set
Standard COI	LCO1490 5′- GGTCAACAAATCATAAAGATATTGG-3′	HCO2198 5′- TAAACTTCAGGGTGACCAAAAAATCA-3′	658 bp
Universal mini-barcode	UniMinibarF1 5′- TCCACTAATCACAARGATATTGGTAC-3’	UniMinibarR1 5′- GAAAATCATAATGAAGGCATGAGC-3’	130 bp
Internal COI mini-barcode	mlCOIintF 5′- GGWACWGGWTGAACWGTWTAYCCYCC-3′	mlCOIintR 5′- GGRGGRTASACSGTTCASCCSGTSCC-3’’	313 – 319 bp
Lobster mini-barcode	LobsterMinibarF 5′- GGWGATGAYCAAATTTAYAATGT-3′	LobsterMinibarR 5′- CCWACTCCTCTTTCTACTATTCC-3′	230 bp

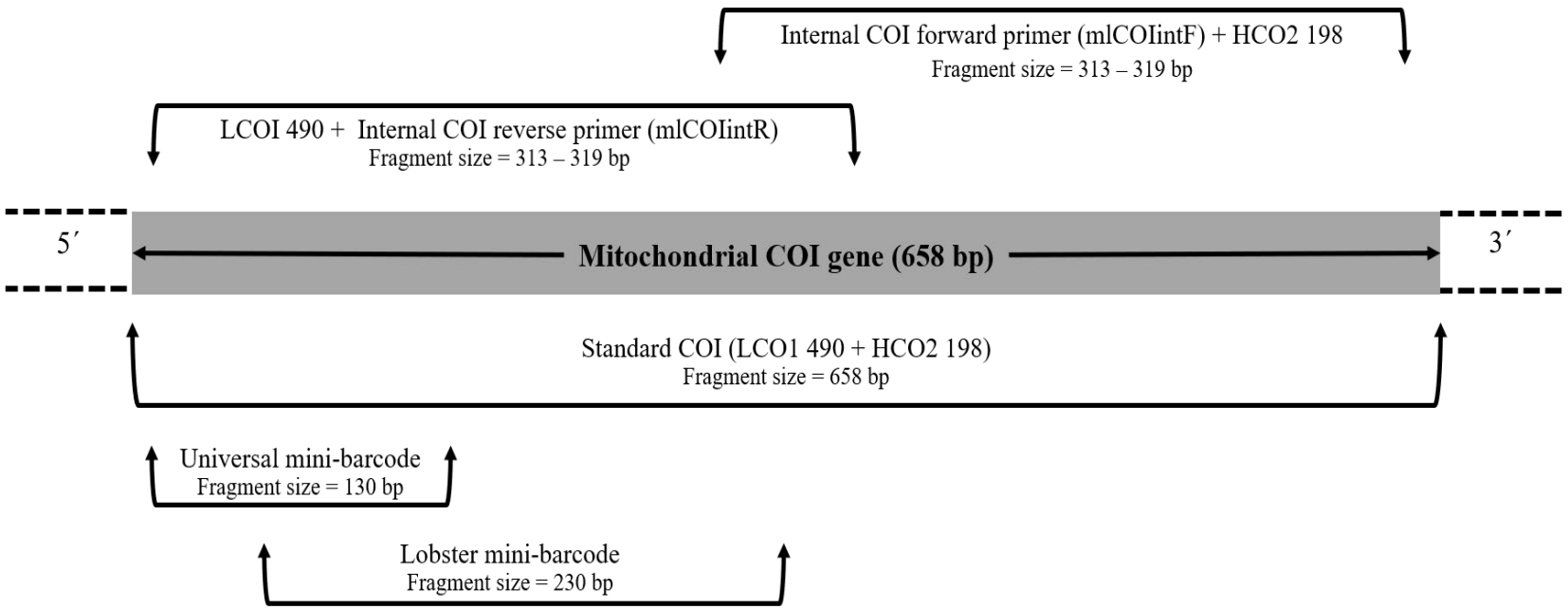


Figure S2.1 A graphical representation of the relative annealing sites and orientation of the different primer sets on the COI barcode region.

Table S2.2 Thermal cycling conditions for the standard COI, universal mini-barcode (touch up PCR), internal COI mini-barcode (touch down PCR) primers and lobster mini-barcode.

	Standard COI			Universal mini-barcode (touch up)			Internal COI mini-barcode (touch down)			Lobster mini-barcode		
	Cycles	**° C**	**Min**	**Cycles**	**° C**	**Min**	**Cycles**	**° C**	**Min**	**Cycles**	**° C**	**Min**
Initial denaturation		94	2		94	2		94	2		94	2
Denaturation		94	1		94	1		94	10 s		94	30 s
Annealing	x 35	40	1	x 5	46	1	x 16	62 **	30 s	x 35	46	30 s
Extension		72	1.5		72	30 s		72	1		68	1
Denaturation		-	-		94	1		94	10 s		-	-
Annealing	-	-	-	x 35	53	1	x 25	46	30 s	-	-	-
Extension		-	-		72	30 s		72	1		-	-
Final extension		72	7		72	5		72	5		68	5

** Temperature (-1°C per cycle = touch down PCR)

Table S2.3 Summary statistics of the sliding window analysis for two selected fragments of each fragment length, showing potential segments for mini-barcodes and their position within the full alignment. Statistics include mean Kimura 2-parameter (K2P) distance, proportion of zero non-conspecific K2P distance, proportion of zero cells in K2P distance matrix and congruence of neighbour joining trees (clade composition and clade composition shallow).

Fragment length/name	Position	K2P dist	Zero non con dist	Zero dist	Clade comp	Clade comp shallow
Fragment100_a	87	0.295	0.0714	0.00298	0.652	0.966
Fragment100_b	152	0.282	0.0800	0.00300	0.618	0.954
Fragment110_a	84	0.277	0.0714	0.00296	0.644	0.966
Fragment110_b	149	0.278	0.0686	0.00291	0.647	0.966
Fragment120_a	84	0.278	0.0714	0.00295	0.658	0.966
Fragment120_b	154	0.281	0.0600	0.00288	0.681	0.966
Fragment130_a	85	0.278	0.0714	0.00296	0.647	0.966
Fragment130_b	150	0.274	0.0600	0.00287	0.670	0.966
Fragment140_a	84	0.270	0.0714	0.00291	0.655	0.966
Fragment140_b	155	0.270	0.0600	0.00288	0.667	0.966
Fragment150_a	84	0.271	0.0714	0.00291	0.695	0966
Fragment150_b	156	0.267	0.600	0.00290	0.667	0.966
Fragment160_a	84	0.270	0.0714	0.00290	0.687	0.977
Fragment160_b	132	0.268	0.0571	0.00280	0.684	0.971
Fragment170_a	83	0.264	0.0714	0.00287	0.695	0.971
Fragment170_b	131	0.263	0.0571	0.00278	0.690	0.971
Fragment180_a	81	0.264	0.0714	0.00287	0.698	0.971
Fragment180_b	126	0.264	0.0571	0.00277	0.693	0.971
Fragment190_a	82	0.262	0.0600	0.00278	0.698	0.971
Fragment190_b	115	0.264	0.0543	0.00270	0.684	0.971
Fragment200_a	85	0.262	0.0486	0.00273	0.716	0.977
Fragment200_b	113	0.264	0.0429	0.00264	0.707	0.977
Fragment210_a	88	0.270	0.0486	0.00273	0.718	0.977
Fragment210_b	111	0.268	0.0429	0.00264	0.716	0.977
Fragment220_a	87	0.269	0.0486	0.00272	0.716	0.978
Fragment220_b	110	0.269	0.0429	0.00262	0.713	0.977
Fragment230_a	85	0.265	0.0486	0.00272	0.710	0.971
Fragment230_b	109	0.266	0.0429	0.00260	0.0716	0.977

Table S2.4 Summary statistics for comparison trees of all 28 fragments. K-scores and Robinson-Foulds (R-F) scores are used to identify best comparison trees. Each score is ranked based on the dataset in ascending order.

Comparison tree	Position	K-score	Scale factor	R-F score	K-score rank	R-F score rank
Fragment100_a	87	2.01237	0.76807	376	25	25
Fragment100_b	152	2.11378	0.12715	362	28	24
Fragment110_a	84	1.83492	1.37474	340	15	20
Fragment110_b	149	2.04018	0.12333	356	27	22
Fragment120_a	84	1.93939	0.55600	360	22	23
Fragment120_b	154	1.90771	0.26661	336	20	19
Fragment130_a	85	1.90815	0.21935	336	21	19
Fragment130_b	150	1.95615	0.16101	344	23	21
Fragment140_a	84	1.98095	0.27977	328	24	16
Fragment140_b	155	1.84495	0.58261	318	17	15
Fragment150_a	84	1.88476	0.16584	334	19	18
Fragment150_b	156	1.67155	0.54366	308	9	12
Fragment160_a	84	2.02150	0.13397	316	26	14
Fragment160_b	132	1.72608	0.57789	296	14	9
Fragment170_a	83	1.85863	0.18692	330	18	17
Fragment170_b	131	1.71813	0.25365	312	13	13
Fragment180_a	81	1.67349	0.56663	304	10	10
Fragment180_b	126	1.56501	0.88654	276	5	5
Fragment190_a	82	1.84444	0.23458	306	16	11
Fragment190_b	115	1.59154	1.00567	264	8	2
Fragment200_a	85	1.69298	0.61303	284	11	8
Fragment200_b	113	1.70125	0.88842	270	12	4
Fragment210_a	88	1.57009	0.77518	270	6	4
Fragment210_b	111	1.56155	0.82888	282	4	7
Fragment220_a	87	1.51802	0.75945	278	3	6
Fragment220_b	110	1.51785	0.80795	278	2	6
Fragment230_a	85	1.58463	0.76811	266	7	3
Fragment230_b	**109**	**1.47431**	**0.84953**	**258**	**1**	**1**

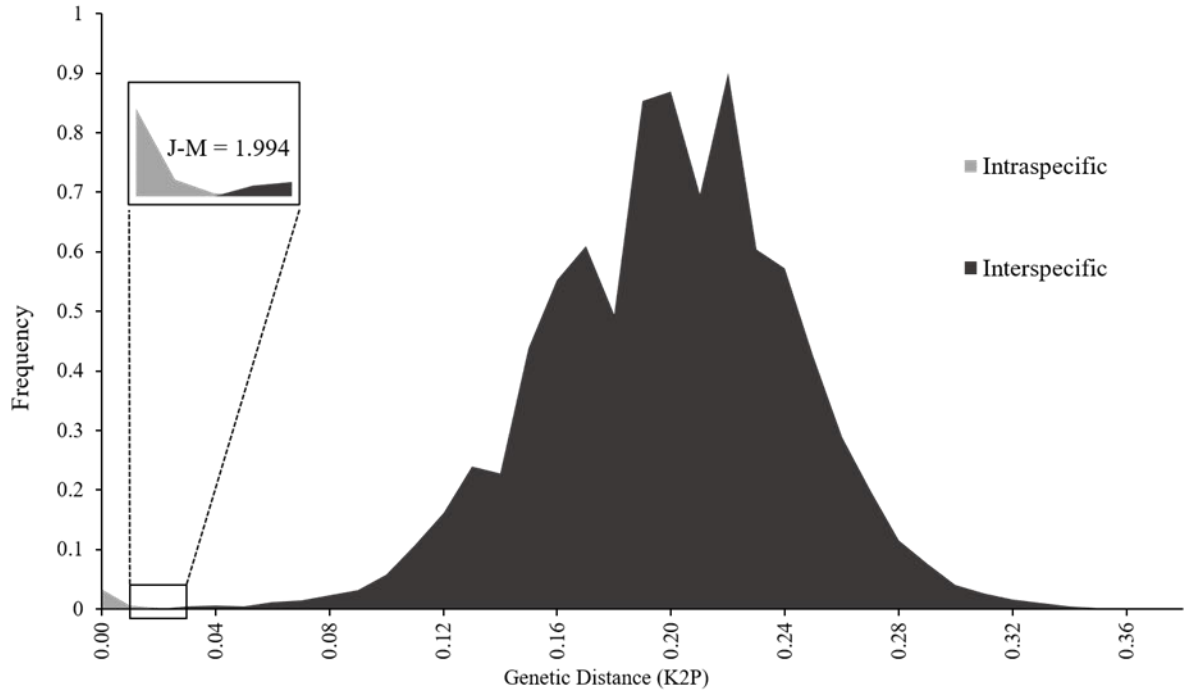


Figure S2.2 Frequency distribution of intra- and interspecific pairwise K2P distances calculated using the selected mini-barcode region (Fragment 230_b). The barcode gap (insert) lies between the genetic distances of 0.02 and 0.03.

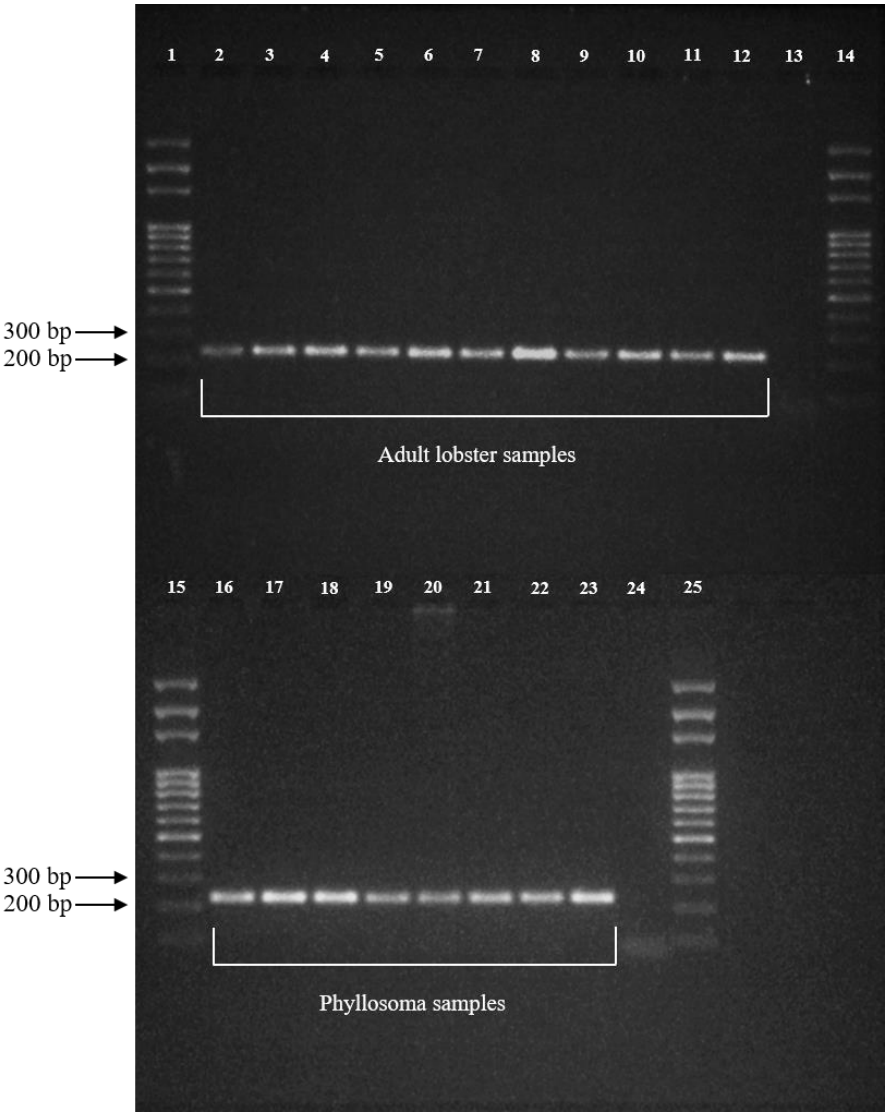


Figure S2.3 Agarose gel image showing the amplification success of the new lobster mini-barcode primer to amplify a range of different adult and phyllosoma lobster samples.

Amplification success rate of the newly designed lobster mini-barcode primer set across different adult and phyllosoma lobster samples. Lanes 2 to 12* are the PCR products recovered from the different adult lobster species. Lanes 16 to 23** are the PCR products recovered from the different phyllosoma lobster species. Lane 13 and 24 are PCR negative controls. Lane 1, 14, 15 and 25 are 100 bp molecular weight marker (Solis Biodyne).

*Adult samples: Lane 2 - *Metanephrops mozambicus*, lane 3 - *Nephropsis stewarti*, lane 4 - *Scyllarides elisabethae*, lane 5 - *Scyllarides squammosus*, lane 6 - *Panulirus homarus*, lane 7 - *Panulirus versicolor*, lane 8 - *Palinurus gilchristi*, lane 9 - *Palinurus delagoae*, lane 10 - *Jasus lalandii*, lane 11 - *Jasus paulensis*, and lane 12 - *Polycheles typhlops*.

**Phyllosoma samples: Lane 16 - *Panulirus ornatus*, lane 17 - *Panulirus homarus*, lane 18 - *Panulirus homarus rubellus*), lane 19 - *Scyllarus arctus*, lane 20 - *Petrarctus rugosus*, lane 21 - *Acantharctus ornatus*, lane 22 - *Scyllarus sp*, and lane 23 - *Petractus sp.*.

Table S2.5 A list of adults and phyllosoma samples amplified using with the lobster mini-barcode primer set (taxonomy and GenBank accession numbers).

Class	Family	Genus	Species	Accession Number on Genbank
Decapoda	Nephropidae	*Metanephrops*	*Metanephrops mozambicus*	MK113927
Decapoda	Nephropidae	*Nephropsis*	*Nephropsis stewarti*	MH428010
Decapoda	Scyllaridae	*Scyllarides*	*Scyllarides elisabethae*	MK113932
Decapoda	Scyllaridae	*Scyllarides*	*Scyllarides squammosus*	MK113933
Decapoda	Palinuridae	*Panulirus*	*Panulirus homarus*	MK113929
Decapoda	Palinuridae	*Panulirus*	*Panulirus versicolor*	MK113931
Decapoda	Palinuridae	*Palinurus*	*Palinurus gilchristi*	MK113926
Decapoda	Palinuridae	*Palinurus*	*Palinurus delagoae*	MK113925
Decapoda	Palinuridae	*Jasus*	*Jasus lalandii*	MK113924
Decapoda	Palinuridae	*Jasus*	*Jasus paulensis*	MK113928
Decapoda	Polychelidae	*Polycheles*	*Polycheles typhlops*	MK113930
Decapoda	Palinuridae	*Panulirus*	*Panulirus ornatus*	MK113919
Decapoda	Palinuridae	*Panulirus*	*Panulirus homarus*	MK113918
Decapoda	Palinuridae	*Panulirus*	*Panulirus homarus rubellus*	MK113917
Decapoda	Scyllaridae	*Scyllarus*	*Scyllarus arctus*	MK113922
Decapoda	Scyllaridae	*Petrarctus*	*Petrarctus rugosus*	MK113920
Decapoda	Scyllaridae	*Acantharctus*	*Acantharctus ornatus*	MK113916
Decapoda	Scyllaridae	*Scyllarus*	*Scyllarus sp.*	MK113923
Decapoda	Scyllaridae	*Petrarctus*	*Petractus sp.*	MK113921

Table S2.6 Summary of amplification and sequencing results for each primer pair. PCR products that showed a single sharp band of correct size were sent for sequencing. Sequences were then BLASTed. Blast search results, identification % and E-value were recorded.

Primer Pair	Sample	Blast top hit	Identification %	E-value
LobsterMiniBarF + LobsterMiniBarR	Adult 1	*Panulirus homarus*	97%	7e-80
	Adult 2	*Panulirus homarus*	96%	4e-82
	Phyllosoma 1	*Panulirus homarus*	93%	3e-74
	Phyllosoma 2	*Panulirus homarus*	92%	5e-71
LCO1490 + HCO2198	Adult 1	*Panulirus homarus*	96%	0.0
	Adult 2	*Panulirus homarus*	97%	0.0
	Phyllosoma 1	*Panulirus homarus*	97%	0.0
	Phyllosoma 2	No amplification	No amplification	No amplification
UniMinibarF1 + UniMinibarR1	Adult 1	Multiple bands	No amplification	No amplification
	Adult 2	Multiple bands	No amplification	No amplification
	Phyllosoma 1	No amplification	No amplification	No amplification
	Phyllosoma 2	No amplification	No amplification	No amplification
MlCOIintF + HCO2198	Adult 1	Multiple bands	No amplification	No amplification
	Adult 2	*Panulirus homarus*	99%	5e-148
	Phyllosoma 1	No amplification	No amplification	No amplification
	Phyllosoma 2	No amplification	No amplification	No amplification
MlCOIintR + LCO1490	Adult 1	Multiple bands	No amplification	No amplification
	Adult 2	*Panulirus homarus*	99%	5e-148
	Phyllosoma 1	No amplification	No amplification	No amplification
	Phyllosoma 2	No amplification	No amplification	No amplification

Chapter Three: Marine zooplankton, mini-barcodes and DNA metabarcoding: The case for taxon-specific primers

This chapter has been submitted for review as: Govender A., Groeneveld J., Singh S., Pillay S. and Willows-Munro S. (In review) Marine zooplankton, mini-barcodes, and DNA metabarcoding: The case for taxon-specific primers. Ecological Applications.

3.1 Abstract

DNA metabarcoding is an emerging tool for rapid species identification, with output data that can be applied to describe communities and provide biodiversity estimates. The approach utilizes high-throughput sequencing and target-specific markers to identify multiple species present in taxonomically complex samples simultaneously. Successful identification at species-level relies on genetic markers that cover a broad taxonomic range yet retain sufficient sequence divergence to resolve species and DNA reference libraries that link individual sequences to taxonomic descriptions. This study developed a versatile DNA metabarcoding protocol for biomonitoring of bulk marine zooplankton samples targeting taxa with life-history stages important to commercial fisheries, such as prawns, shrimps, crabs, lobsters and fish. Taxon-specific DNA mini-barcode primers at a family level were designed to amplify variable portions of the mitochondrial cytochrome c oxidase subunit I (COI) gene region and used together with primers already available in the literature. A series of tests were conducted on artificially assembled mock (with known species ratios) and coastal zooplankton communities to determine the efficiency of the selected protocol and primers. The use of a single primer pair versus primer cocktails (multiple primer sets) was assessed by evaluating species detection rates. Primer cocktails significantly increased the overall species detection in all samples. Taxon-specific primers increased the detection rate of target taxa when compared to using the universal primer set, highlighting the importance of accurate taxon-specific primer design for DNA metabarcoding applications. The protocol developed provides a rapid and cost-effective tool for biomonitoring of zooplankton communities in marine pelagic environments.

Keywords: DNA metabarcoding, primer cocktails, DNA mini-barcode, mitochondrial cytochrome c oxidase subunit I (COI), zooplankton.

3.2 Introduction

Marine ecosystems can shift rapidly in reaction to environmental factors caused by natural events or anthropogenic interventions (Duke and Burton 2020). Species extinction and the "global biodiversity crisis" have emphasized the importance of cataloging and understanding the distribution of biodiversity (Savard *et al.* 2000; Pimm *et al.* 2014). In marine ecosystems, it is essential to monitor changes in the abundance and distribution of marine organisms as an important element of conservation efforts and management (Duke and Burton 2020). Molecular techniques complement traditional morphology-based taxonomy by offering researchers a rapid and cost-effective approach to species identification (Cristescu 2014).

DNA barcoding is a well-established technique used for species identification (Hebert and Gregory 2005). In many animal taxa, DNA barcoding relies primarily on the standard 658 bp mitochondrial cytochrome c oxidase I (COI) gene region (Folmer *et al.* 1994). There are extensive COI-based reference libraries available for many groups of organisms (CBOL database, www.boldsystems.org). The combination of DNA barcoding and high-throughput sequencing (HTS) technology is known as DNA metabarcoding (Taberlet *et al.* 2012). DNA metabarcoding involves the extraction of DNA from bulk samples, mass amplification of standard genetic markers and sequencing using HTS technologies (Cristescu 2014). The DNA sequences generated by HTS technologies are assigned to operational taxonomic units (OTUs) or amplicon sequence variants (ASVs), that are compared against DNA barcode reference libraries, facilitating rapid and cost-effective species identification of multiple organisms present in mixed samples. This approach is commonly referred to as metabarcoding and is increasingly being applied to biodiversity monitoring and assessment (Deiner *et al.* 2017). Metabarcoding has been used to investigate communities of marine zooplankton (Djurhuus *et al.* 2018; Bucklin *et al.* 2019), marine fish (Fraija-Fernández *et al.* 2019) and freshwater communities (Elbrecht and Leese 2017; Hajibabaei *et al.* 2019).

Despite the utility of DNA metabarcoding, its efficient application is not error-free (Zhang *et al.* 2018). The errors include primer bias (Elbrecht and Leese 2015), difficulty in differentiating live cells/organisms from dead cellular material co-collected during sampling (Creer *et al.* 2016), false-positives and false-negatives. False-positive results occur when species that are not present in a sample are detected, either due to contamination during sampling or sequencing errors such

as PCR artifacts (Ficetola *et al.* 2016). False-negative results occur; (a) when species are present in a sample but are not detected, either due to the selected marker not amplifying during PCR (poor primer design) or the sequences are not variable enough to delimit species (Zhang *et al.* 2018); (b)from unbalanced sampling when a dominant taxon overwhelms the signals of taxa occurring at low frequencies (Leray and Knowlton 2017); (c) when mutationsat primer binding sites lead to differences in amplification efficiency, such that the use of "universal" primers may bias amplification towards certain groups of taxa, distorting biodiversity estimates (Clarke et al. (2014); (d) when ASVs for a given species do not match with any sequences on the available reference libraries and hence species assignment is not possible (Clarke et al. 2014); and (e) from tag jumping in illumina indexes (Schnell *et al.* 2015). Marker choice is a challenging step in DNA metabarcoding studies (Clarke *et al.* 2017), especially when analyzing taxonomically diverse samples such as bulk zooplankton (Deagle *et al.* 2014; Creer *et al.* 2016).

Four essential criteria used for selecting primers for DNA metabarcoding are the ability to: amplify short fragments to maximize recovery of amplicons especially where DNA is degraded; amplify genetic barcodes that provide an adequate taxonomic resolution; target taxa of interest to avoid amplification of non-target taxa and amplify DNA from target taxa with equal efficiency (Coissac *et al.* 2012; Elbrecht and Leese 2015; Clarke *et al.* 2017). Where species identification is the goal, evaluation of the DNA fragment chosen for sequencing can be done in advance to ensure that the genetic marker selected will accurately distinguish between closely related species. Few studies test the specificity and sensitivity of primers before use, relying on primers available in the literature that may not be appropriate for the target taxa being investigated (Govender *et al.* 2019).

Recent DNA metabarcoding studies have used conserved genetic markers such as the hypervariable regions of the small subunit (SSU) rRNA genes such as 12S rRNA (mitochondrial ribosomal gene), 16S rRNA (mitochondrial ribosomal gene) and 18S rRNA (nuclear ribosomal gene) (Epp *et al.* 2012; Lindeque *et al.* 2013; Clarke *et al.* 2014; Deagle *et al.* 2014). Nuclear rRNA and mitochondrial rRNA genetic primers are conserved enough to amplify across a wide taxonomic range of species but offer a limited taxonomic resolution for species identification (Hebert *et al.* 2003). The hypervariable regions of the SSU rRNA genes (12S rRNA, 16S rRNA and 18S rRNA) are not extensively represented in barcode reference libraries (Andújar *et al.* 2018). Poor standardization of markers across DNA metabarcoding studies reduces comparability among

studies, limiting community metabarcoding as an efficient, universal system for biodiversity assessments and monitoring.

The mitochondrial COI gene region has been used previously in DNA metabarcoding studies because it has broad and extensive barcode reference libraries that provide a powerful link to taxonomic identifications (Ratnasingham and Hebert 2007). Extensive barcode reference libraries decrease the probability of false taxonomic assignments amongst closely related species (Somervuo *et al.* 2017). Despite there being extensive COI reference libraries for some taxa (e.g., fish, insects, lepidoptera, etc.), reference sequences are almost non-existent for some groups (e.g., nematodes, microbial metazoan, etc.) leading to a huge taxonomic bias in publicly available COI reference libraries. However, the mutational rate of the COI gene region in most animal groups is suitable for interspecific taxonomic assignment of sequences to species level and it can also detect intraspecific variation (Hebert and Gregory 2005).

The 658 bp length of the standard COI gene region used in classic DNA barcoding (Folmer *et al.* 1994) is currently beyond the reach of many HTS technology platforms (e.g., the Illumina MiSeq platform) that have limited read length (Marquina *et al.* 2019), supporting the use of shorter DNA fragments (100-250 bp), known as mini-barcodes for DNA metabarcoding studies (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008). Mini-barcodes can provide a similar degree of taxonomic discrimination than the standard COI gene region while being more successful in amplifying damaged, degraded, or fragmented DNA common in environmental and bulk zooplankton samples (Govender *et al.* 2019). Mini-barcode markers are designed to target the most variable and diagnostic portion of the COI gene region, even though this may not be the same gene portion in all taxa. Primer cocktails (use of three or more primers in a single reaction) reduces amplification biases and increases success rates by improving the detection rates of targeted groups in mixed samples with high diversity (Zhang *et al.* 2018).

Artificially assembled mock communities with known species composition and relative abundance can be used to test detection rates and accuracy of primers against taxa expected to be present in living communities (Yu *et al.* 2012; Brown *et al.* 2015; Krehenwinkel *et al.* 2017; Zhang *et al.* 2018; Duke and Burton 2020). This study presents an optimized protocol for detecting decapod and fish species with life-history stages important to commercial fisheries by DNA metabarcoding marine zooplankton samples collected with tow nets. The groups considered were

prawns (Dendrobranchiata), shrimps (Caridea), crabs (Brachyura), lobsters (Astacidea, Glypheidea, Achelata, and Polychelida) and fish (Actinopterygii). The accuracy of taxon-specific mini-barcode primers and general-use universal primers to detect species was compared using artificial mock communities with known ratios and relative abundance of input species, and on tow net samples with unknown species composition collected from the intended study area – the coastal waters off eastern South Africa. Primers were either multiplexed or used independently with a universal primer pair to optimize species detection. The utility of the primers designed for future zooplankton studies in the region was assessed as a critical step in developing DNA metabarcoding protocols.

3.3 Materials and methods

Taxon-specific primer design

Taxon-specific mini-barcode primers for marine lobsters (230 bp; Govender *et al.* 2019) and fish (313 – 319 bp; Ward *et al.* 2005) were obtained from the literature. Taxon-specific mini-barcode primers for prawns (277 bp and 316 bp), shrimps (310 bp) and crabs (331 bp) were designed following the method outlined in Govender et al. (2019) with the most variable and informative COI fragment targeted. The performance of the taxon-specific primers was compared to a universal primer set designed by Leray et al. (2013) (313 – 319 bp) in the conserved regions of COI.

To determine the most variable portion of the COI for the design of taxon-specific primers, COI sequences for 105 prawn (6 families, 32 genera), 369 shrimp (22 families, 100 genera) and 382 crab species (62 families, 217 genera) were downloaded from GenBank (accessed on 20-07-2018; Appendix 3.1). All species belonging to suborders or infraorders of interest that were available on GenBank were included, irrespective of their geographic origin. Where available, two individuals per species from different geographical areas were included to capture below-species variation.

Sequences were aligned separately for the three datasets using Clustal X2.1 (Larkin *et al.* 2007). Alignments were manually optimized to ensure homology using Bioedit (Hall 1999). Potential mini-barcode fragments per dataset were estimated using sliding window analysis (SWAN) (Proutski and Holmes 1998) in the Species Identity and Evolution (SPIDER) (Brown *et al.* 2012) package in R (http:/www.r-projects.org). The *slideAnalyses* function was used to

generate windows varying in size from 100 to 230 base pairs (bp). Windows were shifted along the length of the COI alignment at ten bp intervals, and two mini-barcode fragments per window length were selected for further analyses based on (1) high mean Kimura 2-parameter (K2P) distance; (2) zero pairwise non-conspecific distances; and (3) a high proportion of clades shared between the neighbor-joining tree drawn from the mini-barcode region compared to the tree drawn from the full-length DNA sequence alignment.

A total of 32 potential mini-barcode fragments were created for each dataset. Maximum likelihood (ML) analysis was conducted on potential mini-barcode fragments and full-length sequence alignments per dataset using Garli 0.951 (Zwickl 2006). In all ML analyses, the K2P model of sequence evolution (Kimura 1980) was implemented, as it is most often used by the DNA barcode community and the Barcode of Life Data Systems (BOLD, www.barcodeoflife.org). The mini-barcode ML trees were then compared to the full-length reference trees using Ktreedist (Soria-Carrasco *et al.* 2007). K-scores (topology and branch length differences) and Robinson-Foulds symmetric differences (topological differences) were calculated for each dataset. For both methods, low values indicated high similarity between the full-length and mini-barcode trees.

A DNA barcode gap analysis was conducted on the highest-scoring mini-barcode fragment for each dataset to confirm that the selected mini-barcode fragments could statistically delimit species. Intra- and interspecific genetic distances were calculated using the K2P nucleotide substitution model in MEGA 6.0 (Tamura *et al.* 2013), and the values were plotted using Microsoft Excel. The maximum intraspecific distance was subtracted from the minimum interspecific distance to determine the barcoding gap (Meier *et al.* 2006). The Jeffries-Matusita distance (J-M) statistic was used to test whether the intra- and interspecific genetic distance classes were separable by considering the distance between their means and the distribution of values around the means (Dabboor *et al.* 2014). The J-M distance is asymptotic to 1.414, and as such, values of 1.414 or higher indicate that intra- and interspecific genetic distances are statistically separable (Trigg and Flasse 2001).

Primers were designed in regions flanking the selected mini-barcode region for each dataset. In cases where flanking regions were too variable to design a single primer pair that could be used to amplify all high-level taxa of interest, phylogenetic information from the ML trees was used to

design family-specific primers (Appendix 3.2 (prawn), Appendix 3.3 (shrimp) and Appendix 3.4 (crab)).

Primer testing

To test the efficiency of the newly designed primers (prawn, shrimp and crab) and those selected from literature (lobster, fish and universal) (Table S3.1), individual primer testing was carried out on adult and larval voucher specimens which were available at the Oceanographic Research Institute, Durban, South Africa. The adult and larval voucher specimens were initially stored in 95% ethanol upon collection and thereafter transferred to new Eppendorf tubes containing 95% ethanol for storage. A total of 56 voucher species were available, comprising of 3 species of prawns (representing 3 genera and 3 families), 8 shrimps (7 genera, 4 families), 15 crabs (13 genera, 11 families), 18 lobsters (9 genera, 5 families) and 12 fish species (10 genera, 10 families)DNA from the individual voucher specimens was extracted using the Qiagen DNeasy Blood and Tissue kit (Qiagen), as per the manufacturer's instructions with a slight modification to the initial incubation step at 56 ºC, where tissue samples were left overnight in the lysis buffer and Proteinase K to ensure complete digestion. PCR reactions (25 µl) contained 20 ng/µl genomic DNA, 12.5 µl OneTaq Quick- LoadMaster Mix (1X, BioLabs, New England), 0.5 µl forward and reverse primers (10 µM), 6.5 µl sterile nuclease-free water, 2 µl additional $MgCl_2$ (25 µM) and 2 µl Bovine Serum Albumin (BSA) (1 mg.m-1). Where primer cocktails were used, the 0.5 µl primer volume was divided by the number of primers for each forward and reverse primers, e.g., if there were two reverse primers, we added 0.25 of each reverse primer. All primers used the same thermal cycling program: initial denaturation at 94˚C for 2 minutes, followed by 35 cycles of denaturation at 94˚C for 30 seconds, annealing at 46˚C for 30 seconds and extension at 68˚ C for 1 minute. The final extension step was carried out at 68˚C for 5 minutes. PCR clean-up and Sanger sequencing reactions were performed at the Central Analytical Facilities (CAF) at the University of Stellenbosch (South Africa). Sequences were checked for their gene- and species-specificity using the nucleotide BLAST tool (BLASTn) on NCBI GenBank. A 95% sequence identity threshold was used for taxonomic assignment, as the voucher specimens used in this study were identified morphologically. After the initial screening of primers, primer cocktails were created to optimize the number of primers and PCR reactions needed for subsequent DNA metabarcoding applications (Table 3.1).Primers used in the DNA metabarcoding protocol included five for prawns (two

forward and three reverse primers), five for shrimps (one forward, four reverse), five for crabs (one forward, four reverse), two for lobsters (one forward, one reverse), three for fish (one forward, two reverse) and two universal COI mini-barcode primers (one forward and one reverse) (Table 3.1). The annealing sites and orientation of primers on the COI barcode region are shown in Figure S3.1.

Table 3.1 The six primer cocktails used in this DNA metabarcoding study: each of the COI primer cocktails amplify different fragments of the COI-5P gene region (Figure S3.1). Illumina adapter target sequences (indicated in bold and underlined) were used in accordance with the workflow from the Illumina 16S Metagenomics protocol (Illumina, 2013). These adapter targets allow Nextera indexing and Illumina adapter addition through PCR. See Supporting Information Table S3.1 for the complete list of primer sets used for the preliminary primer testing step without Illumina adapter target sequences.

Fragment	Primer Name	Sequence (5' - 3')	Direction	Target Taxa	Reference	Fragment Size
COI_Leray	mlCOIintF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW ACW GGW TGA ACW GTW TAY CCY CC	F	Various phyla	Leray et al., 2013	313 bp - 319 bp
	HCO2198	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** TAA ACT TCA GGG TGA CCA AAA AAT CA	R	Various phyla	Folmer et al., 1994	
COI_FISH	mlCOIintF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW ACW GGW TGA ACW GTW TAY CCY CC	F	Various phyla	Leray et al., 2013	313 bp - 319 bp
	HCO2198	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** TAA ACT TCA GGG TGA CCA AAA AAT CA	R	Various phyla	Folmer et al., 1994	
	FishR2	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** ACT TCA GGG TGA CCG AAG AAT CAG AA	R	Fish	Ward et al., 2005	
COI_LOBSTER	LobsterMinibarF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW GAT GAY CAA ATT TAY AAT G T	F	Lobster	Govender et al., 2019	230 bp
	LobsterMinibarR	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCW ACT CCT CTT TCT ACT ATT CC	R	Lobster	Govender et al., 2019	
COI_PRAWN	PrawnMiniBar1F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GCY GAA YTA GGT CAA CCA GG	F	Prawn	This study	277 bp (F1)
	PrawnMiniBar2F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGA TTT GGA AAY TGA YTA GTT CC	F	Prawn	This study	316 bp (F2)
	PrawnMiniBar1R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGA GGR TAW ACA GTT CAT CC	R	Prawn	This study	

Table 3.1 (continued).

COI_PRAWN	PrawnMiniBar2R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT ACY CCT CTT TCT ACT ATW CC	R	Prawn	This study	
	PrawnMiniBar3R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATW CGG TCT ATT GTT ATY CC	R	Prawn	This study	
COI_SHRIMP	ShrimpMiniBar6F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** CCW ATT ATA ATT GGA GGR TTY GG	F	Shrimp	This study	310 bp
	ShrimpMiniBar6R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GCT CCT ARA ATA GAA GAA ACY CC	R	Shrimp	This study	
	ShrimpMiniBar9R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT CTT CTT CGT ATR TTR ATA AC	R	Shrimp	This study	
	ShrimpMiniBar10R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT ARG ATW GAA GAR ACT CC	R	Shrimp	This study	
	ShrimpMiniBar13R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT AAY ATT GAA GAA ACW CCT GC	R	Shrimp	This study	
COI_CRAB	CrabMiniBar1F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** CCW ATT ATA ATT GGA GGA TTY GG	F	Crab	This study	331 bp
	CrabMiniBar5R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT ATW GTT ATA CC	R	Crab	This study	
	CrabMiniBar8R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT AWA GWT ATA CC	R	Crab	This study	
	CrabMiniBar11R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT AGG TCT ATT YTT ATA CC	R	Crab	This study	
	CrabMiniBar12R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT ATK GTT ATA CC	R	Crab	This study	

1

Mock and coastal zooplankton communities

Seven artificial mock communities were assembled using a total of 85 individual specimens (56 different species; Table 3.2; Appendix 3.5) consisting of 5 prawns, 8 shrimps, 22 crabs, 30 lobsters and 20 fish. The specimens were morphologically identified to the lowest taxonomic level possible using microscopy. DNA of individual specimens was extracted as above using the Qiagen DNeasy Blood and Tissue kit and quantified using a Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA). The DNA was then combined into the different mock communities using equimolar DNA concentrations (10 ng/µl) (Appendix 3.5).

Mock 1 was a taxonomically complex community that included equimolar concentrations of DNA from all 85 specimens. Mocks 2 – 6 were constructed to test the sensitivity of the primers designed for the individual taxa, i.e., lobsters, crabs, prawns, shrimps and fish, respectively, to assess whether the designed taxon-specific primers would be able to successfully detect the target taxa if a sample was overwhelmed by non-target taxa (i.e. target taxa are 'underrepresented' in a sample). Each mock community was limited to only five representatives per focal group (lobsters, crabs, prawns, shrimps and fish) depending on the target primer set being tested, for example Mock 2 had all the taxa that was added to Mock 1 but had only five lobster representatives in the sample, making lobsters 'underrepresented'.

In addition to Mocks 1 – 7, two 'coastal' zooplankton communities (collected within the intended sampling area but with unknown species composition) were included. Zoo 1 (high diversity expected) comprised of pooled zooplankton from samples collected by towing standard plankton nets behind a boat at different locations and depths along the South African east coast (5-minute tows at 2 – 3 knots; 200 – 500 um mesh size; surface to 50 m depth). Zoo 2 (lower diversity expected) comprised of a single tow net sample collected in the same manner, with 500 um mesh at a depth of 1-5 m (Table 3.2). The coastal zooplankton samples were stored in 95% ethanol upon sample collection, thereafter the samples were taken to the laboratory and the zooplankton samples transferred to fresh 95% ethanol and preserved in bulk at -20 ℃. Zooplankton samples were stored in 95% ethanol. Before DNA extraction, 2 ml of zooplankton for each sample was centrifuged at 3000 rpm for 60 seconds, and the supernatant was removed. The homogenate was centrifuged at 3000 rpm for 60 seconds, and 40 mg of tissue was transferred to a sterile Eppendorf tube. The

DNA was then extracted as above using the Qiagen Dneasy Blood and Tissue kit and quantified using a Qubit® 2.0 Fluorometer.

Table 3.2 The seven mock communities constructed and the two natural zooplankton communities together with the primers used for each library including a rationale.

Mock Community	Name	Primer cocktails used (Fragment names)	Rationale
Mock 1 **Mock community "All individuals"**	Mock 1a (Library 1)	COI_Leray COI_Fish COI_Lobster COI_Prawn COI_Shrimp COI_Crab	Testing the rate of species detection with all 6 primer sets.
	Mock 1b (Library 2)	COI_Leray only	Testing the rate of species detection with the COI_Leray primer set only.
Mock 2 **Mock community "Underrepresented lobsters"**	Mock 2a (Library 3)	COI_Leray COI_Lobster	Testing the rate of the underrepresented lobster species detection with the COI_Leray together with the COI_Lobster primer set.
	Mock 2b (Library 4)	COI_Leray only	Testing the rate of the underrepresented lobster species detection with the COI_Leray primer set only.
Mock 3 **Mock community "Underrepresented crabs"**	Mock 3a (Library 5)	COI_Leray COI_Crab	Testing the rate of the underrepresented crab species detection with the COI_Leray together with the COI_Crab primer set.
	Mock 3b (Library 6)	COI_Leray only	Testing the rate of the underrepresented crab species detection with the COI_Leray primer set only.
Mock 4 **Mock community "Underrepresented prawn"**	Mock 4a (Library 7)	COI_Leray COI_Prawn	Testing the rate of the underrepresented prawn species detection with the COI_Leray together with the COI_Prawn primer set.
	Mock 4b (Library 8)	COI_Leray only	Testing the rate of the underrepresented prawn species detection with the COI_Leray primer set only.
Mock 5 **Mock community "Underrepresented shrimp"**	Mock 5a (Library 9)	COI_Leray COI_Shrimp	Testing the rate of the underrepresented shrimp species detection with the COI_Leray together with the COI_Shrimp primer set.
	Mock 5b (Library 10)	COI_Leray only	Testing the rate of the underrepresented shrimp species detection with the COI_Leray primer set only.

Table 3.2 (continued).

Mock 6 **Mock community "Underrepresented fish"**	Mock 6a (Library 11)	COI_Leray COI_Fish	Testing the rate of the underrepresented fish species detection with the COI_Leray together with the COI_Fish primer set.
	Mock 6b (Library 12)	COI_Leray only	Testing the rate of the underrepresented fish species detection with the COI_Leray primer set only.
Mock 7 **Mock community "Single individuals per species"**	Mock 7a (Library 13)	COI_Leray COI_Fish COI_Lobster COI_Prawn COI_Shrimp COI_Crab	Testing the rate of species detection with all 6 primer sets.
	Mock 7b (Library 14)	COI_Leray only	Testing the rate of species detection with the COI_Leray primer set only.
Zoo 1 **Mixed bulk zooplankton DNA** **"Coastal zooplankton community"**	Zoo 1a (Library 15)	COI_Leray COI_Fish COI_Lobster COI_Prawn COI_Shrimp COI_Crab	Testing the rate of species detection with all 6 primer sets.
	Zoo 1b (Library 16)	COI_Leray only	Testing the rate of species detection with the COI_Leray primer set only.
Zoo 2 **Single Zooplankton sample** **"Coastal zooplankton community"**	Zoo 2a (Library 17)	COI_Leray COI_Fish COI_Lobster COI_Prawn COI_Shrimp COI_Crab	Testing the rate of species detection with all 6 primer sets.
	Zoo 2b (Library 18)	COI_Leray only	Testing the rate of species detection with the COI_Leray primer set only.

Library preparation and next-generation sequencing (NGS)

The first-round PCR was performed as three replicates per primer set per mock community (Table 3.2); thereafter, replicates were pooled for index-tag PCR. PCR reactions (25 µl) contained 0.25 µl Q5 High-Fidelity DNA Polymerase (0.02 U/ µl, New England BioLabs, Inc), 5 µl Q5 reaction buffer (1X, New England BioLabs, Inc), 5 µl Q5 high GC enhancer (1X, New England BioLabs, Inc), 0.5 µl dNTP's (10 mM of each), 1 µl forward and reverse primers (5 µM), 1 µl template DNA (10 ng/ µl), 2 µl additional $MgCl_2$ (25 µM), 2 µl Bovine Serum Albumin (BSA) (1 mg.ml) and nuclease-free water. Where primer cocktails were used, the 1 µl primer volume was divided by the number of primers for each forward and reverse primers, e.g., if there were two reverse primers, we added 0.5 of each. Thermal cycling consisted of an initial denaturation step at 98 °C for 30 seconds, and 25 cycles of denaturation at 98 °C for 10 seconds, annealing at 46 °C for 30 seconds, extension at 72 °C for 30 seconds, and a final extension step at 72 °C for 4 minutes. Each round of PCR included a negative control which had no DNA present. The use of 25 amplification cycles was used to reduce amplification bias. PCR products were visualized on a 1 % (w/v) TBE agarose gel containing 0.02% Ethidium Bromide (EtBr). A 100 bp molecular weight marker (Solis Biodyne) was used to size the PCR products.

PCR products were quantified using a Qubit 2.0 Fluorometer and pooled in equimolar concentrations (5 ng/µl) to create the 18 libraries (Table 3.2). Illumina sequencing was performed at the KwaZulu-Natal Research and Innovation Platform (KRISP), South Africa. Briefly, each library was cleaned using 1.8X AmpureXP purification beads (Beckman Coulter, High Wycombe, UK). Index PCR was performed using the Nextera XT Index Kit (Illumina, San Diego, USA). Libraries were cleaned up using 0.6X AmpureXP purification beads (Beckman Coulter, High Wycombe, UK) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 4.0 instrument (Life Technologies, California, USA). The fragment sizes were analyzed using a LabChip GX Touch (Perkin Elmer, Hamburg, Germany), with the expected fragment size being approximately 550pb. Each sample library was normalized to 4nM concentration, pooled and denatured with 0.2N sodium acetate. 5% PhiX control (PhiX Control v3) was spiked in a 12 pM library and sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) using a MiSeq Nano Reagent Kit v2 (500 cycles). All the libraries (mock and natural) were sequenced on one run.

Bioinformatics analyses

A local reference library containing 164 sequences was created for the taxonomic assignment of reads. The library included COI sequences of taxa used to build the mock communities (n = 85) that were Sanger sequenced and included sequences from GenBank (n = 79) of taxa expected in zooplankton communities in South African coastal waters.

The DADA2 algorithm (Callahan et al. 2016) implemented in QIIME2 v. 2019.4 (Bolyen et al. 2019) was used to perform quality control checks, chimera removal, filtering, trimming of primers, truncation of forward and reverse reads and merging the paired-end reads into amplified sequence variants (ASVs). The ASVs were queried against the local reference library, BOLD (www.barcodinglife.org) and GenBank. A 95% sequence identity threshold was used for taxonomic assignment to species level, as the voucher specimens used in this study were identified morphologically. Neighbor-joining (NJ) trees were constructed using the ASVs generated for the coastal zooplankton communities using the online tool for MAFFT (Misawa *et al.* 2002). ASVs were assigned to species clusters using genetic distance (DNA barcode gap) and position on NJ tree.

3.4 Results

Taxon-specific primer design and testing

The most informative portion of the COI gene region was identified in each of the three taxa. The smaller SWAN window sizes (100 – 180 bp for prawns, 100 – 130 bp for shrimps and 100-190 bp for crabs) had higher mean K2P distance and zero non-conspecific values in the K2P distance matrix. Larger window sizes (190 – 250 bp for prawns, 140 – 250 bp for shrimps and 200 – 250 bp for crabs) showed better congruence of NJ trees (Table S3.2) and generated lower K- and R-F scores when compared with reference trees (Table S3.3). Fragment 250_b was selected as the most informative region of the COI gene for the prawn dataset, 250_a for shrimps and 240_b for crabs.

The DNA barcode gap analysis was carried out on the mini-barcode regions selected for each dataset. The intra-specific K2P pairwise distances ranged from 0.01 to 0.03 for prawns, 0.01 to 0.03 for shrimps and 0.01 to 0.05 for crabs (Figure 3.1). The inter-specific distances ranged from 0.04 to 0.34 for prawns, 0.04 to 0.44 for shrimps and 0.06 to 0.51 crabs. The J-M distances for all

three groups (prawns, 1.999; shrimps, 1.999; crabs,1.997) exceeded significance thresholds and confirmed that the intra- and interspecific distance classes were statistically separable. Using mini-barcode regions, the DNA barcode gap ranged between 0.03 to 0.04 for prawns, 0.03 to 0.04 for shrimps and 0.05 to 0.06 for crabs.

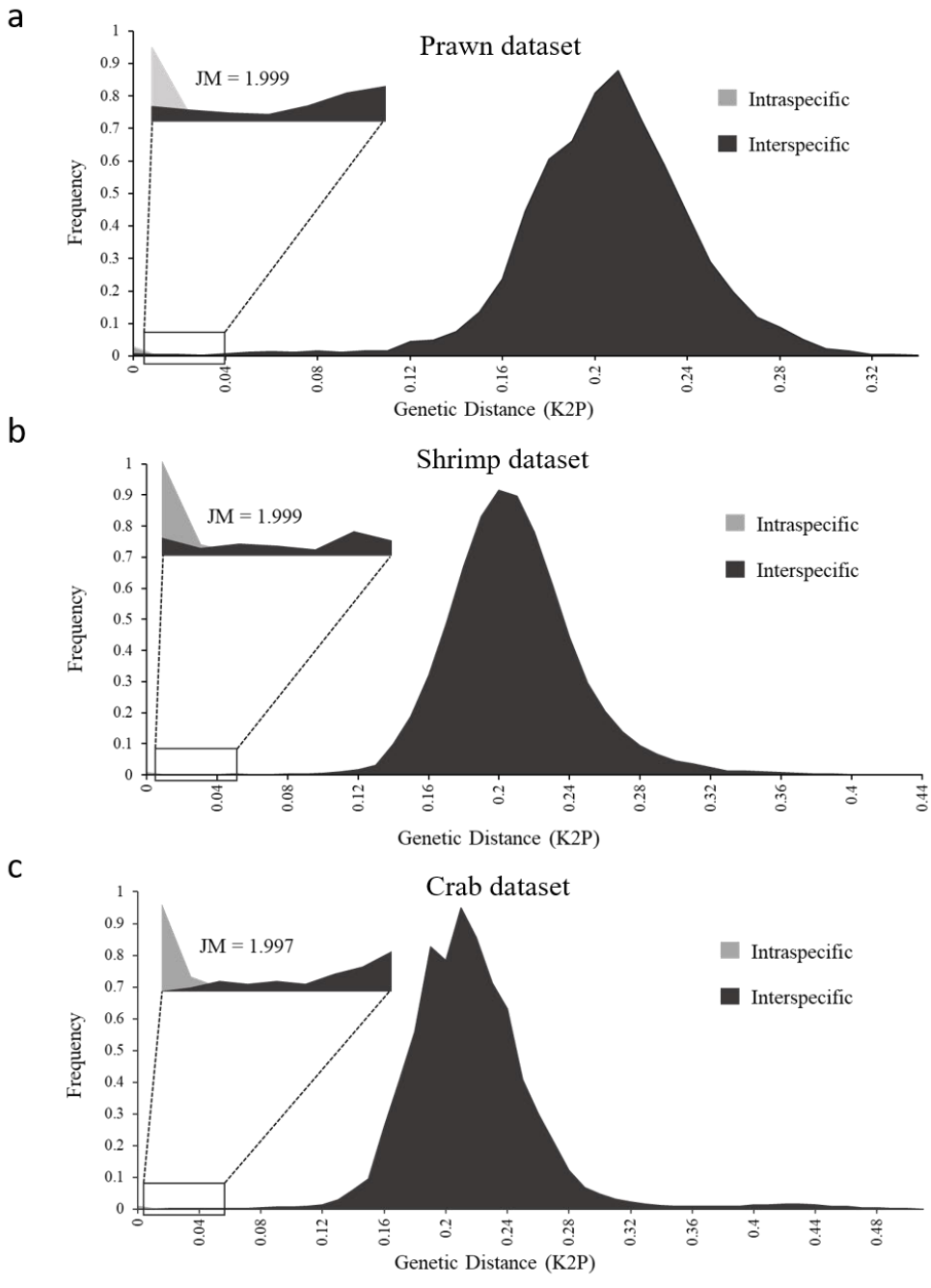


Figure 3.1 Frequency distribution of intra- and interspecific pairwise K2P genetic distances calculated using the selected mini-barcode regions for (a) prawn, (b) shrimp (c) and crab datasets.

The barcode gap (inserts) lies between the genetic distance of 0.03 to 0.04 for prawn, 0.03 to 0.04 for shrimp and 0.05 to 0.06 for crab. The frequency data (prawn = n/3500; shrimp = n/40 000 and crab = n/40 000) was normalized to obtain a range between 0 and 1.

Primers were designed within conserved regions flanking the mini-barcode regions. Two forward and three reverse primers were designed for the prawn dataset, nine forward and 14 reverse primers for the shrimp dataset and one forward and 13 reverse primers for the crab dataset (Table S3.1). The mini-barcode regions of primers were successfully amplified across all 56 adult and larval voucher specimens. BLAST search results confirmed that the mini-barcode sequences matched the morphologically identified adult and larval voucher specimens. All primers were confirmed to target the correct portion of the mitochondrial COI gene region. The final primer cocktails (Table 3.1) used for DNA metabarcoding were based on the amplification and sequencing success of the adult and larval voucher specimens.

Run quality for the mock and coastal zooplankton communities

A total of 14 mock and four coastal zooplankton community libraries were sequenced with Illumina MiSeq. The mock community libraries generated 839 438 reads (mean = 59 960, SD = 16 059). Of these, 552 156 reads were merged (mean = 39 440, SD = 11469; Table 3.3) and after trimming, quality filtering and chimera removal, there remained 276 078 reads available for analysis (mean = 19 720, SD = 5 735; Table 3.3). The coastal community libraries generated 297 310 reads (mean = 74 328, SD = 22 810; Table 3.3). Of these, 155 846 reads were merged (mean = 38 962, SD = 2979) and after quality control, 77 923 reads remained for analysis (mean = 19 481, SD = 1 489; Table 3.3). Sequencing was efficient with minimal filtering needed during merging of the paired-end reads for both the mock and coastal community libraries. Totals of 1433 ASVs (mean = 102, SD = 41) and 1396 ASV's (mean = 349, SD = 98) were identified across the mock and coastal zooplankton communities, respectively.

Table 3.3 Overall comparisons for the 14 mock community libraries and four zooplankton community libraries for the different primer sets used in this study. **PCR artifacts refer to introduced errors into sequences.**

Communities	Total read count	Total merged reads	Number of paired reads	Total ASVs	% identification	Underrepres-ented species detection	% of false positives or PCR artifacts	% of false negative
Mock communities								
Mock 1a	62728	34194	17097	193	95	-	12	5
Mock 1b	58226	40772	20386	85	61	-	23	39
Mock 2a	79494	35986	17993	70	67	5 out of 5	33	33
Mock 2b	96320	72182	36091	97	74	5 out of 5	35	26
Mock 3a	45752	36664	18332	77	67	5 out of 5	10	33
Mock 3b	64678	45624	22812	68	61	4 out of 5	28	39
Mock 4a	67126	29340	14670	136	88	5 out of 5	27	12
Mock 4b	34332	26626	13313	69	59	4 out of 5	31	41
Mock 5a	60890	44810	22405	122	83	5 out of 5	19	17
Mock 5b	40912	31258	15629	75	66	3 out of 5	30	34
Mock 6a	55500	42066	21033	104	80	4 out of 5	27	20
Mock 6b	62896	48016	24008	91	65	3 out of 5	35	35
Mock 7a	66912	30472	15236	176	88	-	22	12
Mock 7b	43672	34146	17073	70	55	-	32	45
Total	**839 438**	**552 156**	**276 078**	**1433**	-	-	-	-
Average	**59 960**	**39 440**	**19 720**	**102**	-	-	-	-
STD	**16 059**	**11 469**	**5 735**	**41**	-	-	-	-
Natural zooplankton communities								
Zoo 1a	97102	34566	17283	465	56	-	9	-
Zoo 1b	56244	39788	19894	379	67	-	13	-
Zoo 2a	90746	40398	20199	318	70	-	5	-
Zoo 2b	53218	41094	20547	234	68	-	9	-
Total	**297 310**	**155 846**	**77 923**	**1396**	-	-	-	-
Average	**74 328**	**38 962**	**19 481**	**349**	-	-	-	-
STD	**22 810**	**2 979**	**1 489**	**98**	-	-	-	-

Species detection in mock communities

The percentage of species detected in the mock communities ranged between 55 and 95% of those known to be present in each sample (Table 3.3). The highest detection rate was achieved when DNA metabarcoding the most taxonomically complex sample (n = 85 specimens) using the multiplexed taxon-specific primer cocktail approach (Mock 1a = 95% identification rate, Table 3.3). In contrast, using the universal primer pair on this sample detected only 61% of the species present. Overall, the primer cocktails including the taxon-specific primers identified more species (average identification success rate = 81%) than the single universal primer pair (average identification success rate = 63%). Mock 7, containing single representatives per species (n = 56 specimens), displayed an increase in species detection when using the primer cocktail approach (Mock 7a = 88% identification rate) as compared to using the universal primer pair (Mock 7b = 55% identification rate). The DNA metabarcoding results for Mocks 2 – 6 (Table 3.3), showed that the primer cocktails selected for each mock community detected up to 100% of the "underrepresented" taxa added to each community. The use of taxon-specific primers resulted in a lower number of false positives (not present but recorded) and negatives (not recorded but present) in comparison to using the universal primer set on its own (Table 3.3).

There was no significant difference recorded when comparing the observed and expected number of species across the mock communities when using the primer cocktails ($df = 4$, $p > 0.05$, Appendix 3.6). In contrast, Mocks 1, 3, 4 and 7 displayed a significant difference between the observed and expected number of species when using the universal primer pair only. The use of taxon-specific primer cocktails outperformed the universal primer set when detecting the different species for each taxonomic group across the mock communities (Figure 3.2).

Species detection in coastal zooplankton communities

In both Zoo 1 and Zoo 2, the multiplexed taxon-specific primer cocktail detected more species (Zoo 1a = 299, Zoo 2a = 145) than when the universal primer pair was used on its own (Zoo 2a = 221, Zoo 2b = 125). The percentage of species across the four coastal zooplankton communities that matched the different reference libraries (local library, BOLD and GenBank) ranged between 56% and 70% (Table 3.3). For both Zoo 1a and Zoo 2a, the species detection rates of the targeted taxa (prawns, crabs, shrimps, lobsters and fish; classes: Malacostraca and Actinopterygii) were increased when using the taxon-specific primer cocktail as compared to using the universal primer

pair on its own (Zoo 1b and Zoo 2b; Figure 3.3). The overall false-positive results for the coastal zooplankton communities were low, ranging between 5% and 13% (Table 3.3).

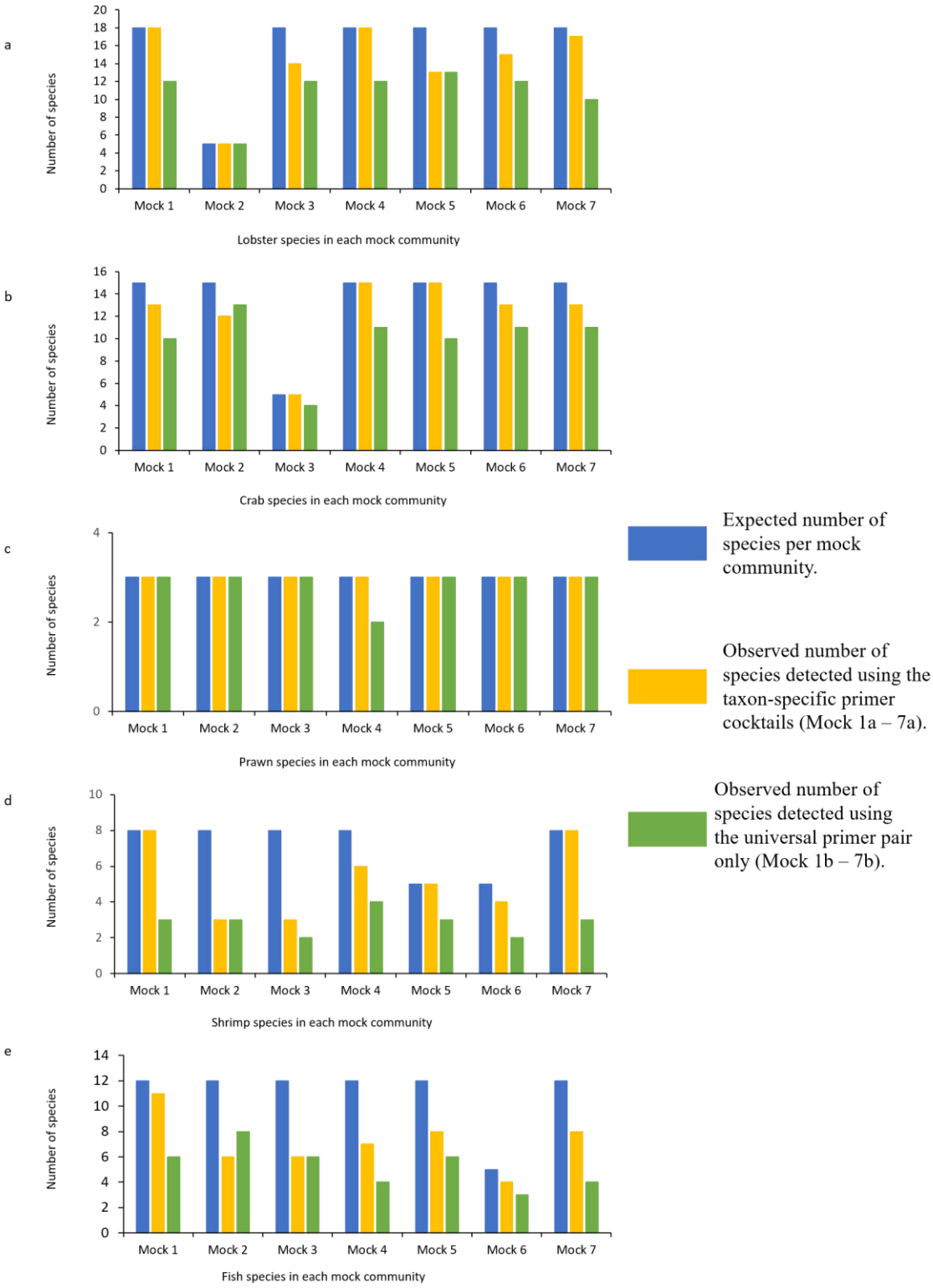


Figure 3.2 The expected number of species vs. the observed number of species for each taxonomic group in each mock community.

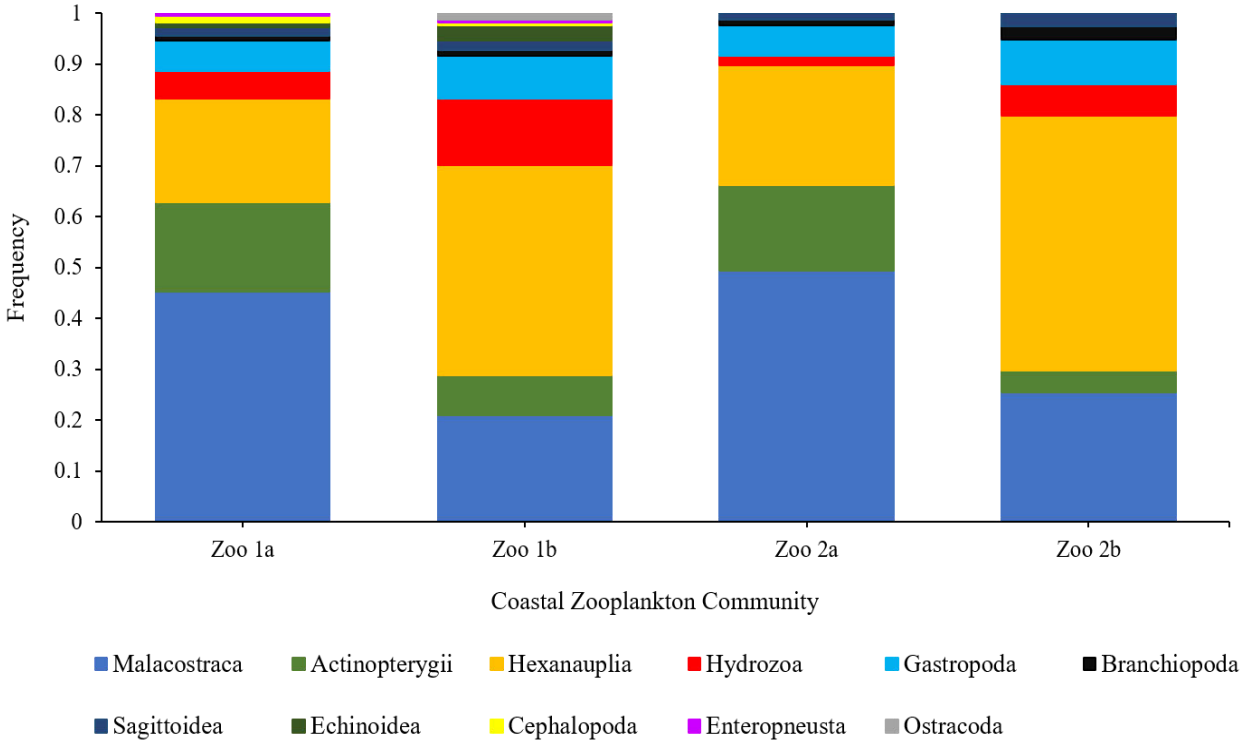


Figure 3.3 Composition bar graphs at Class levels for the 4 coastal zooplankton communities identified using BOLD and BLAST databases. Zoo 1a and Zoo 2a were carried out using the taxon-specific primer cocktails whereas Zoo 1b and Zoo 2b were carried out using the universal primer pair only. The different colours indicate the different classes. The overall percentage of the different classes were calculated using the number of species identified per class. Prawns, crabs, shrimp and lobster fall under Malacostraca and fish fall under Actinopterygii.

3.5 Discussion

The emergence of DNA metabarcoding as a technique for rapid species identification offers a powerful and cost-effective tool for large scale multi-taxon biodiversity assessments of community and environmental samples (Ji *et al.* 2013). DNA metabarcoding of marine zooplankton communities is an active field of research but recent studies have lacked consistency in the molecular markers used. The lack of standardization limits downstream meta-analyses and comparability. To improve consistency and detection rates, we developed and tested a method for DNA metabarcoding of zooplankton using a portion of the COI gene region, for which reference libraries from the global DNA barcoding initiative are available. Although the use of a single locus

has been criticized (Moritz and Cicero 2004), our results show that by using taxon-specific primer cocktails, at least 95% of targeted taxa can be identified down to a species-level.

The protein-coding COI gene is highly variable at the third position of most codons, making it difficult to design primers that can amplify a broad taxonomic sample (Deagle *et al.* 2014). In this study, we designed multiple taxon-specific mini-barcode primers within the COI gene for the prawn, shrimp and crab datasets, used the DNA barcode gap analysis to validate whether the selected mini-barcode regions could accurately delimit species and used primer cocktails for DNA metabarcoding. Confirmation of a barcode gap in all three datasets underpins the use of taxon-specific primers as a diagnostic tool in species identification. Even the smallest fragment generated (205 bp) could distinguish between species, highlighting the utility of the COI mini-barcode regions for short-read sequencing technology such as the Illumina MiSeq.

Recent zooplankton DNA metabarcoding studies have used only a single primer pair for amplification (Brown *et al.* 2015; Elbrecht and Steinke 2019; Yang and Zhang 2020). This study strongly advocates using multiple taxon-specific primers to improve COI amplification success and species detection from zooplankton community samples. In contrast to Duke and Burton (2020), where species detection decreased in taxonomically complex samples, we found that using a cocktail of taxon-specific primers succeeded in maintaining high species detection rates, even at low DNA concentrations (Mock 7) or when some taxa were underrepresented (Mocks 2 – 6). Despite using a single gene locus, the use of multiple primer sets within the COI gene region was successful in species detection, by recovering a substantial proportion of the original taxonomic information seeded in the mock communities.

Zhang et al. (2018) found that the use of primer cocktails for COI DNA metabarcoding reduced both false-positive and false-negative results. Similarly, we found a significant reduction in both false- positive and false-negative results in both the mock and coastal zooplankton analyses when using primer cocktails. The false-negative readings can be attributed to low taxonomic coverage within reference databases combined with the high level of stringency (95% match threshold) applied in our annotation pipeline (Valsecchi *et al.* 2020).

The accurate detection of species in extant zooplankton communities is critical for biodiversity monitoring and community analysis to construct long-term biological indices to track the effects of climate change on marine pelagic environments (Bourlat *et al.* 2013; Aylagas *et al.*

2014). Testing the taxon-specific primer cocktails on unaltered taxonomically diverse zooplankton samples collected at sea (*in-situ* samples Zoo 1 and Zoo 2) confirmed that the multiplexed taxon-specific primer cocktail approach significantly increased species detection when compared to using the universal primer pair. The numbers of species estimated for *in-situ* samples were comparable to those in mock communities, demonstrating that the new protocols are transferable and can be implemented in practice for biomonitoring of marine pelagic environments.

Incomplete online DNA barcoding reference libraries for South African and global marine zooplankton (e.g., BOLD and GenBank) (Singh *et al.* 2021) limited the number of taxa that could be identified to species level in the *in-situ* samples. To overcome this hurdle in future DNA metabarcoding projects of marine pelagic environments, traditional taxonomic descriptions that incorporate Sanger sequencing of full-length DNA barcodes are required for key zooplankton groups so that more ASVs can be matched with validated species-descriptions (Kvist 2013).

In conclusion, the cocktail of taxon-specific primers advocated in our study reduced PCR bias and preferential amplification compared to using a universal primer, thus improving species detection rates and diversity estimates. The methods used to design taxon-specific mini-barcodes and the mock-sampling test protocols can easily be applied to DNA metabarcoding studies in other ecosystems with multiple species that are difficult to quantify visually. Our results have important implications for choosing primers that allow for a level of standardization across biomonitoring programs. Using taxon-specific mini-barcodes maximizes richness estimates and increases the possibility of detecting underrepresented taxa overwhelmed by non-target taxa.

3.6 Acknowledgments

We would like to acknowledge funding support from DSI/NRF/ACEP Captor Project (Grant 110763). The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily attributed to the NRF. We would also like to thank David Pearton and Jennifer Giandhari for their technical assistance.

3.7 Data accessibility

The Sanger sequences generated in this study for the local reference library are available on GenBank (accession numbers: MT709164 - MT709248). The Sanger sequences downloaded from GenBank for the local reference library and the next generation sequences generated for the mock communities are available on figshare: doi.org/10.6084/m9.figshare.12615713. All the scripts and sequences used to create the mini-barcodes in this chapter are available on figshare: https://doi.org/10.6084/m9.figshare.14378663.v1

3.8 References

Andújar C., Arribas P., Yu, D.W., Vogler A.P. and Emerson, B.C. (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology 27, 3968-3975.

Aylagas, E., Borja, Á. And Rodríguez-Ezpeleta N. (2014) Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). PLOS One 9, e90529.

Bolyen E., Rideout J.R., Dillon M.R., Bokulich N.A., Abnet C.C., Al-Ghalith G.A., Alexander H., Alm E.J., Arumugam M., Asnicar F., Bai Y., Bisanz J.E., Bittinger K., Brejnrod A., Brislawn C.J., Brown C.T., Callahan B.J., Caraballo-Rodriguez A.M., Chase J., Cope E.K., Da Silva R., Diener C., Dorrestein P.C., Douglas G.M., Durall D.M., Duvallet C., Edwardson C.F., Ernst M., Estaki M., Fouquier J., Gauglitz J.M., Gibbons S.M., Gibson D.L., Gonzalez A., Gorlick K., Guo J., Hillmann B., Holmes S., Holste H., Huttenhower C., Huttley G.A., Janssen S., Jarmusch A.K., Jiang L., Kaehler B.D., Kang K.B., Keefe C.R., Keim P., Kelley S.T., Knights D., Koester I., Kosciolek T., Kreps J., Langille M.G.I., Lee J., Ley R., Liu Y.X., Loftfield E., Lozupone C., Maher M., Marotz C., Martin B.D., McDonald D., McIver L.J., Melnik A.V., Metcalf J.L., Morgan S.C., Morton J.T., Naimey A.T., Navas-Molina J.A., Nothias L.F., Orchanian S.B., Pearson T., Peoples S.L., Petras D., Preuss M.L., Pruesse E., Rasmussen L.B., Rivers A., Robeson M.S., 2nd, Rosenthal P., Segata N., Shaffer M., Shiffer A., Sinha R., Song S.J., Spear J.R., Swafford A.D., Thompson L.R., Torres P.J., Trinh P., Tripathi A., Turnbaugh P.J., Ul-Hasan S., van der Hooft J.J.J., Vargas F., Vazquez-Baeza Y., Vogtmann E., von Hippel M., Walters W., Wan Y., Wang M., Warren J., Weber K.C., Williamson C.H.D., Willis A.D., Xu Z.Z., Zaneveld J.R., Zhang Y., Zhu Q., Knight R. and Caporaso J.G. (2019) Reproducible, interactive,

scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37, 852-857.

Bourlat S.J., Borja A., Gilbert J., Taylor M.I., Davies N., Weisberg S.B., Griffith J.F., Lettieri T., Field D., Benzie J., Glöckner F.O., Rodríguez-Ezpeleta N., Faith D.P., Bean T.P. and Obst M. (2013) Genomics in marine monitoring: New opportunities for assessing marine health status. Marine Pollution Bulletin 74, 19-31.

Brown E.A., Chain F.J.J., Crease T.J., MacIsaac H.J. and Cristescu M.E. (2015) Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities? Ecology and Evolution 5, 2234-2251.

Brown S.D.J., Collins R.A., Boyer S., Lefort M.C., Malumbres-Olarte J., Vink C.J. and Cruickshank R.H. (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources 12, 562-565.

Bucklin A., Yeh H.D., Questel J.M., Richardson D.E., Reese B., Copley N.J. and Wiebe P.H. (2019) Time-series metabarcoding analysis of zooplankton diversity of the NW Atlantic continental shelf. ICES Journal of Marine Science 76, 1162-1176.

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A. and Holmes S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13, 581-583.

Čandek K. and Kuntner M. (2015) DNA barcoding gap: Reliable species identification over morphological and geographical scales. Molecular Ecology Resources 15, 268-277.

Clarke L.J., Beard J.M., Swadling K.M. and Deagle B.E. (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution 7, 873-883.

Clarke L.J., Soubrier J., Weyrich L.S. and Cooper A. (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources 14, 1160-1170.

Coissac E., Riaz T. and Puillandre N. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. Molecular Ecology 21, 1834-1847.

Creer S., Deiner K., Frey S., Porazinska D., Taberlet P., Thomas W.K., Potter C. and Bik H.M. (2016) The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7, 1008-1018.

Cristescu M.E. (2014) From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. Trends in Ecology and Evolution 29, 566-571.

Dabboor M., Howell S., Shokr M. and Yackel J. (2014) The Jeffries and Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. International Journal of Remote Sensing 35, 6859-6873.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F. and Taberlet P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. Biology Letters 10, 20140562.

Deiner K., Bik H.M., Machler E., Seymour M., Lacoursiere-Roussel A., Altermatt F., Creer S., Bista I., Lodge D.M., de Vere N., Pfrender M.E. and Bernatchez L. (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26, 5872-5895.

Djurhuus A., Pitz K., Sawaya N.A., Rojas-Márquez J., Michaud B., Montes E., Muller-Karger F. and Breitbart M. (2018) Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. Limnology and Oceanography: Methods 16, 209-221.

Duke E.M. and Burton R.S. (2020) Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. Ecology and Evolution 10, 3463-3476.

Elbrecht V. and Leese F. (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLOS One 10, e0130324.

Elbrecht V. and Leese F. (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science 5, 11.

Elbrecht V. and Steinke D. (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology 64, 380-387.

Epp L.S., Boessenkool S., Bellemain E.P., Haile J., Esposito A., Riaz T., Erseus C., Gusarov V.I., Edwards M.E., Johnsen A., Stenoien H.K., Hassel K., Kauserud H., Yoccoz N.G., Brathen K.A., Willerslev E., Taberlet P., Coissac E. and Brochmann C. (2012) New environmental

metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. Molecular Ecology 21, 1821-1833.

Ficetola G.F., Taberlet P. and Coissac E. (2016) How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources 16, 604-607.

Folmer O., Black M., Hoeh W., Lutz R. and Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Fraija-Fernández N., Bouquieaux M.C., Rey A., Mendibil I., Cotano U., Irigoien X., Santos M. and Rodríguez-Ezpeleta N. (2019) Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. Ecology and Evolution 10, 7560-7584.

Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

Hajibabaei M., Porter T.M., Wright M. and Rudar J. (2019) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLOS One 14, e0220953.

Hajibabaei M., Smith M.A., Janzen D.H., Rodriguez J.J., Whitfield J.B. and Hebert P.D.N. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6, 959-964.

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.

Hebert P. and Gregory T.R. (2005) The promise of DNA barcoding for taxonomy. Systematic Biology 54, 852-859.

Hebert P.D.N., Cywinska A., Ball S.L. and deWaard J.R. (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270, 313-321.

Ji Y., Ashton L., Pedley S.M., Edwards D.P., Tang Y., Nakamura A., Kitching R., Dolman P.M., Woodcock P., Edwards F.A., Larsen T.H., Hsu W.W., Benedick S., Hamer K.C., Wilcove D.S., Bruce C., Wang X., Levi T., Lott M., Emerson B.C. and Yu D.W. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16, 1245-1257.

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

Krehenwinkel H., Wolf M., Lim J.Y., Rominger A.J., Simison W.B. and Gillespie R.G. (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports 7, 17668.

Kvist S. (2013) Barcoding in the dark: A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Molecular Phylogenetics and Evolution 69, 39-45.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Leray M., and Knowlton N. (2017) Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5, e3006.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T. and Machida R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10, 34.

Lindeque P.K., Parry H.E., Harmer R.A., Somerfield P.J. and Atkinson A. (2013) Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PLOS One 8, e81327.

Marquina D., Andersson A.F. and Ronquist F. (2019) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19, 90-104.

Meier R., Shiyang K., Vaidya G. and Ng P.K.L. (2006) DNA Barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Systematic Biology 55, 715-728.

Meusnier I., Singer G.A., Landry .J.F., Hickey D.A., Hebert P.D. and Hajibabaei M. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214.

Misawa K., Katoh K., Ki K. and T M. (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Research 30, 3059-3066.

Moritz C. and Cicero C. (2004) DNA Barcoding: Promise and pitfalls. PLOS Biology 2, e354.

Pimm S.L., Jenkins C.N., Abell R., Brooks T.M., Gittleman J.L., Joppa L.N., Raven P.H., Roberts C.M. and Sexton J.O. (2014) The biodiversity of species and their rates of extinction, distribution, and protection. Science 344, 1246752.

Proutski V. and Holmes E. (1998) SWAN: Sliding window analysis of nucleotide sequence variability. Bioinformatics 14, 467-468.

Ratnasingham S. and Hebert P.D.N. (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7, 355-364.

Savard J.P.L., Clergeau P. and Mennechez G. (2000) Biodiversity concepts and urban ecosystems. Landscape and Urban Planning 48, 131-142.

Schnell I.B., Bohmann K. and Gilbert M.T.P. (2015) Tag jumps illuminated--reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources 15, 1289-1303.

Singh S., Groeneveld J., Huggett J., Naidoo D., Cedras R. and Willows-Munro S. (2021) DNA metabarcoding of marine zooplankton in South Africa: How good is the reference library? African Journal of Marine Science (Accepted).

Somervuo P., Yu D.W., Xu C.C.Y., Ji Y., Hultman J., Wirta H. and Ovaskainen O. (2017) Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution 8, 398-407.

Soria-Carrasco V., Talavera G., Igea J. and Castresana J. (2007) The K tree score: Quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics 23, 2954-2956.

Taberlet P., Coissac E., Pompanon F., Brochmann C. and Willerslev E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21, 2045-2050.

Tamura K., Stecher G., Peterson D., Filipski A. and Kumar S. (2013) MEGA 6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30, 2725-2729.

Trigg S. and Flasse S. (2001) An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. International Journal of Remote Sensing 22, 2641-2647.

Valsecchi E., Bylemans J., Goodman S.J., Lombardi R., Carr I., Castellano L., Galimberti A. and Galli P. (2020) Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. Environmental DNA 00, e72.

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

Yang J. and Zhang X. (2020) eDNA metabarcoding in zooplankton improves the ecological status assessment of aquatic ecosystems. Environment International 134, 105230.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C. and Ding Z. (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3, 613-623.

Zhang G.K., Chain F.J.J., Abbott C.L. and Cristescu M.E. (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11, 1901-1914.

Zwickl D. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin, United States.

3.9 Supplementary data

Table S3.1 The complete list of primer sets used for primer testing. The primer pairs selected and used in the DNA metabarcoding study are in bold and red.

Fragment	Primer Name	Sequence (5' - 3')	Direction	Target Taxa	Reference	Fragment Size
COI_Leray	**mlCOIintF**	**GGW ACW GGW TGA ACW GTW TAY CCY CC**	F	Various phyla	Leray et al., 2013	313 bp -
	HCO2198	**TAA ACT TCA GGG TGA CCA AAA AAT CA**	R	Various phyla	Folmer et al., 1994	319 bp
COI_FISH	**mlCOIintF**	**GGW ACW GGW TGA ACW GTW TAY CCY CC**	F	Various phyla	Leray et al., 2013	313 bp -
	HCO2198	**TAA ACT TCA GGG TGA CCA AAA AAT CA**	R	Various phyla	Folmer et al., 1994	319 bp
	FishR2	**ACT TCA GGG TGA CCG AAG AAT CAG AA**	R	Fish	Ward et al., 2005	
COI_LOBSTER	**LobsterMinibarF**	**GGW GAT GAY CAA ATT TAY AAT GT**	F	Lobster	Govender et al., 2019	230 bp
	LobsterMinibarR	**CCW ACT CCT CTT TCT ACT ATT CC**	R	Lobster	Govender et al., 2019	
COI_PRAWN	**PrawnMiniBar1F**	**GCY GAA YTA GGT CAA CCA GG**	F	Prawn	This study	277 bp (F1)
	PrawnMiniBar2F	**GGA TTT GGA AAY TGA YTA GTT CC**	F	Prawn	This study	316 bp (F2)
	PrawnMiniBar1R	**GGA GGR TAW ACA GTT CAT CC**	R	Prawn	This study	
	PrawnMiniBar2R	**CCT ACY CCT CTT TCT ACT ATW CC**	R	Prawn	This study	
	PrawnMiniBar3R	**GGT ATW CGG TCT ATT GTT ATY CC**	R	Prawn	This study	
COI_SHRIMP	ShrimpMiniBar1F	CCW ATT WTA ATT GGA GGA TTT GG	F	Shrimp	This study	310 bp
	ShrimpMiniBar2F	CCA ATT ATA ATT GGR GGC TTY GG	F	Shrimp	This study	
	ShrimpMiniBar3F	CCY ATT ATA ATT GGA GGR TTT GG	F	Shrimp	This study	
	ShrimpMiniBar4F	CCW ATT ATA ATT GGR GGR TTT GG	F	Shrimp	This study	
	ShrimpMiniBar5F	CCW ATT ATA ATY GGA GGA TTY GG	F	Shrimp	This study	
	ShrimpMiniBar6F	**CCW ATT ATA ATT GGA GGR TTY GG**	F	Shrimp	This study	
	ShrimpMiniBar7F	CCT ATT ATA ATW GGA GGA TTT GG	F	Shrimp	This study	
	ShrimpMiniBar8F	CCA ATT ATA ATY GGA GGA TTY GG	F	Shrimp	This study	
	ShrimpMiniBar9F	CCW ATT ATA ATT GGW GGR TTT GG	F	Shrimp	This study	
	ShrimpMiniBar1R	GTA GWT ATA AAA TTA GCG GCW CC	R	Shrimp	This study	

Table S3.1 (continued).

	ShrimpMiniBar2R	GTA GTT ATA AAG TTA ACK GCT CC	R	Shrimp	This study	
	ShrimpMiniBar3R	CCT AGY ATT GAG GAT ACY CCT GC	R	Shrimp	This study	
	ShrimpMiniBar4R	GCC CCT AAR ATA GAA GAM ACT CC	R	Shrimp	This study	
	ShrimpMiniBar5R	CCT ARA ATT GAA GAR ACW GG	R	Shrimp	This study	
	ShrimpMiniBar6R	**GCT CCT ARA ATA GAA GAA ACY CC**	R	Shrimp	This study	
	ShrimpMiniBar7R	GTT ATA AAR TTA ACK GCT CC	R	Shrimp	This study	
	ShrimpMiniBar8R	GTT GTT ATR AAA TTA ACT GCY CC	R	Shrimp	This study	
	ShrimpMiniBar9R	**CCT CTT CTT CGT ATR TTR ATA AC**	R	Shrimp	This study	
	ShrimpMiniBar10R	**CCT ARG ATW GAA GAR ACT CC**	R	Shrimp	This study	
	ShrimpMiniBar11R	CCT AGR ATA GAW GAW ACA CCT GC	R	Shrimp	This study	
	ShrimpMiniBar12R	CCT ARG ATT GAW GAW ACW CC	R	Shrimp	This study	
	ShrimpMiniBar13R	**CCT AAY ATT GAA GAA ACW CCT GC**	R	Shrimp	This study	
	ShrimpMiniBar14R	CCT ARR ATA GAA GAA ACT CCT GC	R	Shrimp	This study	
COI_CRAB	**CrabMiniBar1F**	**CCW ATT ATA ATT GGA GGA TTY GG**	F	Crab	This study	331 bp
	CrabMiniBar1R	GGT ATT TGG TCT ATW GWT ATW CC	R	Crab	This study	
	CrabMiniBar2R	GGT ATT TGG TCT ATT ATT ATA CC	R	Crab	This study	
	CrabMiniBar3R	GGT ATT AGG TCY ATT GTT ATW CC	R	Crab	This study	
	CrabMiniBar4R	GGT ATT TGR TCT ATW GTT ATW CC	R	Crab	This study	
	CrabMiniBar5R	**GGT ATT TGR TCT ATW GTT ATA CC**	R	Crab	This study	
	CrabMiniBar6R	GGT ATT AGR TCT ATA GTT ATA CC	R	Crab	This study	
	CrabMiniBar7R	GGT ATT TGA TCT AWA KTT ATT CC	R	Crab	This study	
	CrabMiniBar8R	**GGT ATT TGR TCT AWA GWT ATA CC**	R	Crab	This study	
	CrabMiniBar9R	GGT ATT TGR TCT ATR GGT ATA CC	R	Crab	This study	
	CrabMiniBar10R	GGT ATT TGR TCT ATW GTT ATA CC	R	Crab	This study	
	CrabMiniBar11R	**GGT ATT AGG TCT ATT YTT ATA CC**	R	Crab	This study	
	CrabMiniBar12R	**GGT ATT TGR TCT ATK GTT ATA CC**	R	Crab	This study	
	CrabMiniBar13R	GGT ATT TGR TCT AGA RTT ATA CC	R	Crab	This study	

1

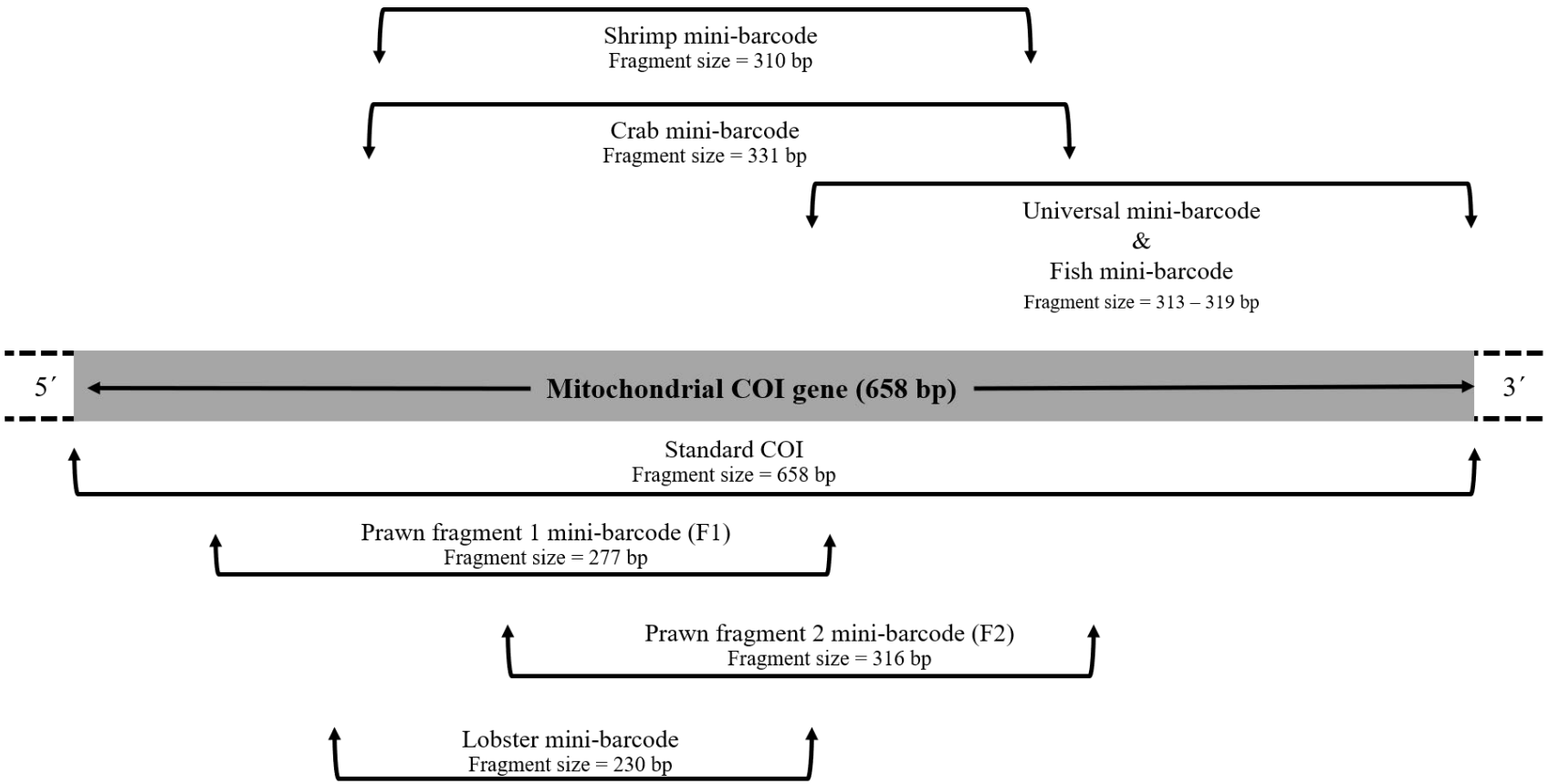


Figure S3.1 A graphical representation of the relative annealing sites and orientation of the different primer sets for DNA metabarcoding targeting the Folmer COI barcode region.

Table S3.2 Summary statistics for the prawn, shrimp, and crab dataset for the sliding window analysis of the two selected fragments of each fragment length, showing potential segments for mini-barcodes and their position within the full alignment. Statistics include mean Kimura 2-parameter (K2P) distance, proportion of zero non-conspecific K2P distance, proportion of zero cells in K2P distance matrix, and congruence of neighbour-joining trees (clade composition and clade composition shallow).

Fragment length/name	window shift	K2P dist	Zero non con dist	Zero dist	Clade comp	Clade comp shallow
Prawn dataset						
Fragment100_a	110	0.2591	0.1157	0.005168	0.6028	0.8899
Fragment100_b	153	0.2671	0.1343	0.005125	0.6682	0.8990
Fragment110_a	152	0.2603	0.1296	0.004995	0.6822	0.9174
Fragment110_b	172	0.2641	0.1250	0.005426	0.6121	0.8991
Fragment120_a	151	0.2560	0.1296	0.004995	0.6963	0.9358
Fragment120_b	172	0.2681	0.1250	0.005383	0.6262	0.8991
Fragment130_a	153	0.2579	0.1296	0.004996	0.6963	0.9174
Fragment130_b	172	0.2651	0.1157	0.005038	0.6542	0.9083
Fragment140_a	152	0.2522	0.1296	0.004952	0.7009	0.9266
Fragment140_b	166	0.2563	0.1250	0.005297	0.6542	0.8991
Fragment150_a	83	0.2491	0.1296	0.005082	0.6729	0.9174
Fragment150_b	161	0.2544	0.1019	0.004952	0.6729	0.9083
Fragment160_a	83	0.2449	0.1111	0.004909	0.6682	0.8991
Fragment160_b	161	0.2506	0.1019	0.004952	0.6589	0.9083
Fragment170_a	82	0.2465	0.1111	0.004479	0.7009	0.9266
Fragment170_b	153	0.2493	0.09259	0.004521	0.7196	0.9358
Fragment180_a	82	0.2446	0.1111	0.004823	0.7103	0.9083
Fragment180_b	153	0.2521	0.09259	0.004522	0.7056	0.9266
Fragment190_a	82	0.2407	0.1111	0.004823	0.7149	0.9174
Fragment190_b	152	0.2473	0.07407	0.004306	0.7149	0.9449
Fragment200_a	110	0.2414	0.08333	0.004263	0.7009	0.9174
Fragment200_b	130	0.2417	0.09259	0.004349	0.7476	0.9266
Fragment210_a	107	0.2385	0.08333	0.004220	0.7383	0.9174
Fragment210_b	121	0.2390	0.08333	0.004263	0.7476	0.9357
Fragment220_a	105	0.2395	0.08333	0.004220	0.7242	0.9174
Fragment220_b	114	0.2420	0.08333	0.004263	0.7523	0.9174
Fragment230_a	104	0.2440	0.08333	0.004220	0.7383	0.9282
Fragment230_b	112	0.2474	0.08333	0.004220	0.7429	0.9174
Fragment240_a	103	0.2454	0.08333	0.004220	0.7289	0.9174
Fragment240_b	113	0.2438	0.08333	0.004220	0.7663	0.9266
Fragment250_a	102	0.2409	0.08333	0.004220	0.7196	0.9174
Fragment250_b	110	0.2441	0.08333	0.004177	0.7336	0.9266

Table S3.2 (Continued).

Shrimp dataset						
Fragment100_a	146	0.3274	0.08961	0.001266	0.6536	0.9611
Fragment100_b	154	0.3421	0.1000	0.001297	0.6315	0.9585
Fragment110_a	147	0.3335	0.08961	0.001259	0.6536	0.9637
Fragment110_b	154	0.3262	0.1000	0.001290	0.6367	0.9611
Fragment120_a	141	0.3218	0.07532	0.001222	0.6471	0.9559
Fragment120_b	154	0.3270	0.09480	0.001273	0.6532	0.9585
Fragment130_a	113	0.3167	0.07792	0.001236	0.6536	0.9689
Fragment130_b	140	0.3204	0.07532	0.001212	0.6575	0.9715
Fragment140_a	113	0.3176	0.07142	0.001178	0.6744	0.9740
Fragment140_b	154	0.3141	0.09480	0.001259	0.6549	0.9637
Fragment150_a	113	0.3169	0.06623	0.001148	0.6757	0.9792
Fragment150_b	131	0.3168	0.07012	0.001175	0.6601	0.9766
Fragment160_a	111	0.3163	0.06233	0.001117	0.6770	0.9844
Fragment160_b	130	0.3158	0.07012	0.001165	0.6614	0.9792
Fragment170_a	113	0.3101	0.0662	0.001128	0.6835	0.9844
Fragment170_b	127	0.3081	0.07012	0.001155	0.6666	0.9792
Fragment180_a	112	0.3113	0.05974	0.001097	0.6888	0.9792
Fragment180_b	125	0.3093	0.0701	0.001141	0.6757	0.9740
Fragment190_a	113	0.3149	0.06363	0.001094	0.6992	0.9818
Fragment190_b	120	0.3088	0.06493	0.001111	0.6888	0.9766
Fragment200_a	114	0.3181	0.06363	0.001097	0.6979	0.9844
Fragment200_b	117	0.3059	0.06493	0.001101	0.6940	0.9818
Fragment210_a	111	0.3215	0.05974	0.001063	0.7122	0.9844
Fragment210_b	116	0.3078	0.06363	0.001077	0.6914	0.9792
Fragment220_a	110	0.3188	0.05974	0.001053	0.7057	0.9766
Fragment220_b	115	0.3087	0.06363	0.001074	0.6927	0.9766
Fragment230_a	109	0.3125	0.05974	0.001036	0.7044	0.9818
Fragment230_b	115	0.3074	0.06363	0.001070	0.6914	0.9792
Fragment240_a	112	0.3125	0.05974	0.001047	0.7083	0.9766
Fragment240_b	113	0.3120	0.06363	0.001060	0.7083	0.9818
Fragment250_a	113	0.3092	0.6363	0.001053	0.7148	0.9766
Fragment250_b	114	0.3055	0.6363	0.001060	0.7187	0.9766
Crab dataset						
Fragment100_a	82	0.2707	0.02678	0.001065	0.6125	0.9693
Fragment100_b	162	0.2800	0.04591	0.001172	0.6035	0.9668
Fragment110_a	83	0.2665	0.02678	0.001058	0.6227	0.9744
Fragment110_b	162	0.2715	0.04209	0.001166	0.6138	0.9693
Fragment120_a	83	0.2615	0.02678	0.001039	0.6304	0.9744
Fragment120_b	171	0.2702	0.02168	0.001110	0.6074	0.9744
Fragment130_a	82	0.2629	0.02168	0.001019	0.6342	0.9719
Fragment130_b	171	0.2677	0.01785	0.001094	0.6125	0.9821
Fragment140_a	83	0.2620	0.01403	0.0009904	0.6419	0.9821

Table S3.2 (Continued).

Fragment140_b	165	0.2625	0.02168	0.001097	0.6253	0.9770
Fragment150_a	82	0.2593	0.01403	0.0009839	0.6470	0.9821
Fragment150_b	162	0.2602	0.02168	0.001075	0.6265	0.9744
Fragment160_a	84	0.2638	0.01275	0.0009676	0.6470	0.9846
Fragment160_b	161	0.2545	0.01785	0.001062	0.6304	0.9821
Fragment170_a	82	0.2595	0.01275	0.0009611	0.6457	0.9846
Fragment170_b	153	0.2534	0.01020	0.001023	0.6585	0.9795
Fragment180_a	83	0.2562	0.01275	0.0009480	0.6700	0.9948
Fragment180_b	152	0.2514	0.01020	0.001016	0.6624	0.9821
Fragment190_a	84	0.2527	0.01275	0.0009513	0.6739	0.9897
Fragment190_b	151	0.2478	0.01020	0.001006	0.6611	0.9821
Fragment200_a	82	0.2490	0.01275	0.0009415	0.6636	0.9897
Fragment200_b	128	0.2467	0.01785	0.001013	0.6764	0.9846
Fragment210_a	84	0.2478	0.01275	0.0009448	0.6828	0.9923
Fragment210_b	112	0.2482	0.007653	0.0009578	0.7020	0.9948
Fragment220_a	82	0.2470	0.01275	0.0009252	0.6777	0.9948
Fragment220_b	111	0.2506	0.007653	0.0009513	0.7046	0.9948
Fragment230_a	82	0.2451	0.01275	0.0009220	0.6879	0.9923
Fragment230_b	108	0.2474	0.007653	0.000938	0.6892	0.9974
Fragment240_a	102	0.2438	0.007653	0.0009122	0.7084	0.9974
Fragment240_b	111	0.2472	0.007653	0.0009448	0.7135	0.9948
Fragment250_a	100	0.2432	0.007653	0.0009122	0.7097	0.9974
Fragment250_b	108	0.2478	0.00765	0.0009317	0.6982	0.9948

Table S3.3 Summary statistics for comparison trees of all 28 fragments for the prawn, shrimp, and crab datasets. K-scores and Robinson-Foulds (R-F) scores are used to identify best comparison trees. Each score is ranked based on the dataset in ascending order. The selected fragment size for each dataset is highlighted in bold and red.

Comparison tree	Position	K-score	Scale factor	R-F score
Prawn Dataset				
Fragment100_a	110	0.46658	1.00410	208
Fragment100_b	153	0.49801	1.01505	212
Fragment110_a	152	0.48742	0.92233	200
Fragment110_b	172	0.47384	0.81887	194
Fragment120_a	151	0.46305	0.93523	194
Fragment120_b	172	0.45342	0.85127	198
Fragment130_a	153	0.42192	0.91966	174
Fragment130_b	172	0.45151	0.85837	190
Fragment140_a	152	0.45125	0.89697	180
Fragment140_b	166	0.46948	0.85185	200
Fragment150_a	83	0.41938	1.08495	182

Table S3.3 (Continued).

Fragment150_b	161	0.44512	0.88860	186
Fragment160_a	83	0.43090	1.04615	176
Fragment160_b	161	0.40208	0.97034	166
Fragment170_a	82	0.37669	1.08922	164
Fragment170_b	153	0.37786	0.97089	154
Fragment180_a	82	0.38383	1.03377	170
Fragment180_b	153	0.39283	0.96617	158
Fragment190_a	82	0.38397	1.01203	174
Fragment190_b	152	0.39345	0.91803	164
Fragment200_a	110	0.35977	1.00917	150
Fragment200_b	130	0.40221	0.95314	172
Fragment210_a	107	0.36038	0.99832	136
Fragment210_b	121	0.35881	0.99408	158
Fragment220_a	105	0.35310	1.00253	136
Fragment220_b	114	0.36888	0.98108	172
Fragment230_a	104	0.34809	0.99933	142
Fragment230_b	112	0.33121	0.98452	132
Fragment240_a	103	0.33275	0.99241	142
Fragment240_b	113	0.38839	0.92597	164
Fragment250_a	102	0.31449	0.97843	136
Fragment250_b	**110**	**0.28795**	**0.97888**	**134**
		Shrimp dataset		
Fragment100_a	146	1.24956	1.05498	754
Fragment100_b	154	1.32258	1.06793	780
Fragment110_a	147	1.14398	1.15345	756
Fragment110_b	154	1.27265	0.95678	760
Fragment120_a	141	1.25327	1.05636	758
Fragment120_b	154	1.04605	1.05948	714
Fragment130_a	113	1.23280	1.03862	752
Fragment130_b	140	1.20353	0.96759	738
Fragment140_a	113	1.15392	1.09147	722
Fragment140_b	154	1.05238	0.97502	698
Fragment150_a	113	1.10746	1.09792	720
Fragment150_b	131	1.09305	1.02712	720
Fragment160_a	111	1.13203	1.03735	704
Fragment160_b	130	1.07018	1.01149	694
Fragment170_a	113	1.01646	1.00583	674
Fragment170_b	127	1.05032	1.00287	696
Fragment180_a	112	0.92479	1.05826	666
Fragment180_b	125	1.03423	1.03211	686
Fragment190_a	113	0.98728	1.02752	660
Fragment190_b	120	0.92349	1.06048	654
Fragment200_a	114	1.14628	0.95850	652

Table S3.3 (Continued).

Fragment200_b	117	1.01476	1.00461	650
Fragment210_a	111	1.03074	0.99312	644
Fragment210_b	116	0.94767	1.01873	664
Fragment220_a	110	0.88179	1.06066	638
Fragment220_b	115	1.05618	0.95559	660
Fragment230_a	109	0.91004	1.01686	622
Fragment230_b	115	0.90876	0.98792	614
Fragment240_a	112	0.92704	0.93735	604
Fragment240_b	113	0.98035	0.94147	610
Fragment250_a	**113**	**0.90345**	**0.93853**	**602**
Fragment250_b	114	0.88041	0.96921	618
		Crab dataset		
Fragment100_a	82	1.12594	0.95073	778
Fragment100_b	162	1.06661	0.92993	758
Fragment110_a	83	1.00668	0.97871	742
Fragment110_b	162	1.01927	0.82066	726
Fragment120_a	83	1.00263	0.95016	736
Fragment120_b	171	0.94287	0.85619	676
Fragment130_a	82	0.95793	1.03810	720
Fragment130_b	171	1.09345	0.72427	690
Fragment140_a	83	0.93746	1.01232	682
Fragment140_b	165	1.09807	0.72155	704
Fragment150_a	82	0.92167	1.03619	678
Fragment150_b	162	0.93630	0.84665	668
Fragment160_a	84	0.94517	1.01887	650
Fragment160_b	161	0.92752	0.88076	652
Fragment170_a	82	0.86126	1.04485	650
Fragment170_b	153	0.96209	0.88056	638
Fragment180_a	83	0.90461	1.02482	668
Fragment180_b	152	0.85319	0.92048	634
Fragment190_a	84	0.91615	0.94307	634
Fragment190_b	151	0.85783	0.92410	600
Fragment200_a	82	0.91858	0.94290	654
Fragment200_b	128	0.82476	1.01654	616
Fragment210_a	84	0.82393	0.96839	600
Fragment210_b	112	0.97683	0.91083	644
Fragment220_a	82	0.77195	1.01372	582
Fragment220_b	111	0.88054	0.95303	604
Fragment230_a	82	0.90185	0.91983	598
Fragment230_b	108	0.92952	0.89155	630
Fragment240_a	102	0.78939	0.96297	596
Fragment240_b	**111**	**0.74837**	**0.98659**	**582**
Fragment250_a	100	0.76982	0.96524	586
Fragment250_b	108	0.82293	0.92964	568

1

Chapter Four: Connectivity of marine zooplankton communities in a dynamic ocean environment off eastern South Africa, inferred from DNA metabarcoding analysis

This chapter is preparation for peer review as: Govender A., Groeneveld J., Singh S., Pillay S. and Willows-Munro S. (In prep) Connectivity of marine zooplankton communities in a dynamic ocean environment off eastern South Africa, inferred from DNA metabarcoding analysis.

4.1 Abstract

Zooplankton form abundant and ecologically important multi-species communities in marine pelagic ecosystems. Zooplankton communities are well-suited for monitoring ecosystem health and biodiversity changes because their relative abundance and species composition are sensitive to environmental change. In this study, surface waters (1 – 5 m depth) were sampled with plankton tow nets (0.8 m diameter; 500 um mesh size) at cross-shelf transects (between 20 and 200 m depth soundings) at three sites (uThukela, Durban and Aliwal) along the east coast of South Africa. Although connected by the strong Agulhas Current, the three sites differ in shelf width, seafloor substrate and benthic habitat structures. DNA metabarcoding (high-throughput DNA sequencing of the mitochondrial cytochrome c oxidase I gene region) was used to determine the species composition and relative abundance of samples, and to infer connectivity between sampling sites. Connectivity network analysis detected distinct clustering of zooplankton communities by site. The hypothesis that strong ocean currents would result in similar well-mixed zooplankton communities over the KZN continental shelf was rejected. Similarities in the species composition of zooplankton and benthic life history stages at sampling sites suggested a benthic-pelagic coupling effect facilitated by bottom topography and water movements. This study provides a novel approach to biomonitoring of marine pelagic environments in coastal waters off southern Africa, based on DNA metabarcoding of zooplankton communities.

Keywords: Zooplankton, DNA metabarcoding, biomonitoring, species composition, connectivity.

4.2 Introduction

Ecological connectivity can be defined as the movement of organisms, resources and energy between habitats (Taylor *et al.* 1993; Bishop *et al.* 2017). Connectivity in marine environments encapsulates the demographic linking of populations across multiple temporal and spatial scales through the active or passive movement and exchange of organisms (Almany *et al.* 2009; Cowen and Sponaugle 2009; Hidalgo *et al.* 2017). Connectivity is fundamental in maintaining biological productivity and marine biodiversity with connected ecosystems showing resilience to natural and anthropogenic stressors such as climate change, habitat degradation, pollution, ocean acidification and coral bleaching (Almany *et al.* 2009; Jones *et al.* 2009). Connectivity maintains biological productivity and marine biodiversity by influencing gene flow through the dispersal and survival of individuals (Kinlan and Gaines 2003; Singh *et al.* 2019), species interactions (Gaines and Lafferty 1995), population dynamics (Roughgarden *et al.* 1988), patterns of distribution (Reed *et al.* 2000; Singh *et al.* 2020), community composition (Carr *et al.* 2017) and the functioning of ecosystems (Sheaves 2009; Jeltsch *et al.* 2013). Examples of the close linkages between movement (hence connectivity) and biodiversity are provided by the studies highlighting the importance for species distributions (Bonte et al. 2004; Yu et al. 2012) and metapopulation dynamics (Hanski et al. 1994). Understanding the physical and biological factors contributing to marine connectivity is essential for marine conservation and management (Roberts 1997; Cowen *et al.* 2000).

Zooplankton are the most abundant and ecologically important animals in marine pelagic ecosystems. Zooplankton communities are highly diverse and play an essential role in the transfer of energy from primary producers (phytoplankton) to higher trophic levels (larger organisms) and mediate biogeochemical flux in oceans (Richardson 2008; Huggett and Kyewalyanga 2017). Zooplankton community structure and biomass respond rapidly to environmental changes such as temperature (Moore and Folt 1993; Kelly *et al.* 2016), salinity (Paturej and Gutkowska 2015; Gutierrez *et al.* 2018), predation pressure (Greene 1983) and chemical stressors (Havens and Hanazato 1993; Rodgher *et al.* 2009). Zooplankton comprises of holoplankton (whole life-cycle planktonic) and meroplankton (life cycle partially planktonic) (Huggett and Kyewalyanga 2017) and are model organisms for monitoring the effects of climate change on biodiversity and species distributions in pelagic environments (Hays *et al.* 2005; Lin *et al.* 2017).

Zooplankton dispersal in marine ecosystems is governed by water movements, such as ocean currents, tides, fronts, waves and sub-mesoscale processes such as eddies and countercurrents (Roberts 1997; Richardson 2008; Hays 2017; Singh *et al.* 2018). Dispersal is further determined by behavioral adaptations such as vertical migrations of larval stages, the orientation of organisms relative to currents (Shanks and Brink 2005; Knights *et al.* 2006; Butler *et al.* 2011) and swimming behavior (Wilkin and Jeffs 2011). Dispersal between adjacent sites (and hence connectivity) may be hindered by ocean fronts (Gilg and Hilbish 2003; Singh *et al.* 2018), or in contrast, geographically distant sites may be connected by drift in currents (Treml *et al.* 2008; Mitarai *et al.* 2009; Groeneveld *et al.* 2012; Ockhuis *et al.* 2017; Noyon *et al.* 2018).

Biomonitoring of whole zooplankton communities using morphological identification (traditional microscopic methods) is challenging because of the high species diversity, small size, fragile nature and cryptic morphology of many taxa (Bucklin *et al.* 2016). As an alternative, DNA metabarcoding which relies on high throughput sequencing (HTS) can allow for large-scale taxonomic identification of complex samples using DNA barcodes (Taberlet *et al.* 2012; Bucklin *et al.* 2016; Creer *et al.* 2016; Laakmann *et al.* 2020). DNA metabarcoding uses DNA reference libraries such as Barcode of Life Data Systems (BOLD, www.barcodeoflife.org), International Nucleotide Sequence Database Collaboration (INSDC, www.insdc.org) and the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) to link DNA barcodes to verified taxonomic records describing individual species. DNA metabarcoding plays a role in assisting zooplankton biodiversity assessments as it reduces the reliance on expert taxonomists and is under development in many parts of the world (Zhang *et al.* 2018; Govender *et al.* 2019; Laakmann *et al.* 2020) (see https://metazoogene.org/).

The combination of DNA metabarcoding outputs (species composition and relative abundance) with community analyses to compare species diversity or richness at different sites lends itself to studies of connectivity in marine environments (Yamanaka and Minamoto 2016). Zooplankton species that are dispersed passively by water movements are more likely to form relatively homogenous communities in interconnected environments (Watson *et al.* 2011; Niebuhr *et al.* 2015). Other species (such as lobster phyllosomas; see Butler et al. 2011) undertake diurnal vertical migrations or can swim directionally during later larval stages (Leis 2007; Shanks 2009), thus playing a more active role in dispersal processes. Seasonality also influences community

composition and connectivity because of seasonal life-history cycles and variable abundance of certain taxa (Smeti *et al.* 2015; Neumann Leitão *et al.* 2019). Fundamental assumptions of null hypotheses (no difference between zooplankton communities, signifying high connectivity) are therefore that passive dispersal occurs (different species in the community are dispersed similarly) and that samples are not influenced by seasonal differences.

The KwaZulu-Natal (KZN) coastal region in South Africa has complex physical oceanography (Guastella and Roberts 2016; Roberts *et al.* 2016) that strongly influences the dispersal of zooplankton over the shelf (Bustamante and Branch 1996; Collocott 2016; Pretorius *et al.* 2016). Based on particle dispersal models and gene flow estimates of a spiny lobster with long-lived phyllosoma larvae, Singh *et al.* (2018, 2019) demonstrated the uncertain fate of drifting larvae in this dynamic ocean environment – larvae can be retained over the shelf by sub-mesoscale processes, dispersed downstream along the coast, or become entrained in the western boundary Agulhas Current at the shelf-edge and presumably lost. There is also evidence of northward dispersal of larvae against the predominant direction of flow (Teske *et al.* 2008; von der Heyden *et al.* 2008) facilitated by countercurrents between the Agulhas Current and the coast (Roberts *et al.* 2010; Guastella and Roberts 2016; Roberts *et al.* 2016).

Zooplankton research carried out in KZN marine waters have been mostly descriptive and stem from a small number of standalone surveys or surveys with large time gaps that focused on: distribution and diversity of copepods and chaetognaths (De Decker 1964; De Decker and Mombeck 1964; Carter 1977; De Decker 1984; Schleyer 1985; Pretorius *et al.* 2016), lobster phyllosomas (Berry 1974; Govender *et al.* 2019), siphonophores and hydromedusae (Thibault-Botha *et al.* 2004; Buecher *et al.* 2005; Thibault-Botha and Gibbons 2005) and fish larvae (Beckley 1986; Beckley and Hewitson 1994; Beckley 1995; Beckley and Leis 2000; Pattrick and Strydom 2014; Collocott 2016). Pretorius *et al.* (2016) investigated the seasonality of zooplankton biomass, distribution and size composition in the KZN Bight. These studies have focused on a limited number of taxa using mainly morphological identification of species, with varying results. DNA metabarcoding to rapidly and accurately identify species present in tow-net samples is expected to exponentially increase the knowledge of zooplankton biodiversity in the region (Laakmann *et al.* 2020; Singh *et al.* 2021), allowing for novel studies of connectivity in this highly dynamic marine environment. In addition, zooplankton DNA metabarcoding data coupled with analysis of physical

oceanographic features can contribute to our understanding of the physical/environmental processes that drive dispersal and connectivity (Smeti *et al.* 2015).

This study used DNA metabarcoding together with community-level analyses to assess the connectivity of zooplankton communities over the continental shelf of KZN. Geographically separated sampling sites differed in shelf width, seafloor substrate and benthic habitat structures. Meroplanktonic taxa important to fisheries (mainly decapods and fish) were sequenced using DNA metabarcoding and species were identified using online DNA barcode reference libraries. The species composition and relative abundance of zooplankton in samples were used to investigate connectivity between sampling sites, based on the hypotheses of no difference in species richness and diversity.

4.3 Materials and methods

Study area

Geographic description and ocean environment: The south-westward flowing Agulhas Current along the east coast of South Africa is considered the world's strongest western boundary current and forms part of the anticyclonic Indian Ocean gyre (Lutjeharms 2006). The current follows the shelf-edge and extends well below a depth of 1000 m with a mean width of 100 km (Bang and Pearce 1976). The Agulhas Current waters are reported to be warming (Rouault *et al.* 2009) and recent evidence suggests an increase in the turbulence of the current caused by intensifying winds (Beal and Elipot 2016).

The continental shelf between the Agulhas Current and the KZN coast is mostly narrow (3 – 11 km wide) and slopes down steeply after reaching a depth of 100 m. An exception is the broader KZN Bight (an offset of 160 km long with a maximum width of 45 km) (Schumann 1988). The broadening of the shelf at the KZN Bight diverts the Agulhas Current offshore and diminishes the currents velocity gradient, giving rise to shelf-edge upwelling cells, cyclonic lee-trapped eddies and countercurrents, creating a semi-closed circulation system within the Bight (Lutjeharms *et al.* 1989; Lutjeharms *et al.* 2000; Meyer *et al.* 2002; Lutjeharms 2006). Large singular meanders known as the Natal Pulse originate within the KZN Bight and grow in amplitude as they move south towards Durban (Lutjeharms 2006). A mesoscale semi-permanent cyclonic eddy to the south of Durban with a strong north-eastward counter-current retains shelf waters inshore (Roberts *et al.*

2010; Guastella and Roberts 2016; Roberts *et al.* 2016). Inner-shelf circulation over the central part of the KZN Bight (near the mouth of the Thukela River) is weak and highly variable. The circulation of shelf waters is also strongly influenced by wind (Roberts *et al.* 2016).

Biogeography: Spalding *et al.* (2007) divided coastal and shelf waters into a hierarchical nested system of 12 realms, 62 provinces and 232 ecoregions. The present study area is located at the boundary of two realms (Western Indo Pacific and Temperate Southern Africa), two provinces (Western Indian Ocean and Agulhas Provinces) and two ecoregions (Delagoa and Natal ecoregions). The region has high species diversity, comprising tropical, subtropical and temperate water species and a transitional zone for several invertebrate taxa (Teske *et al.* 2011; Jooste *et al.* 2018), including estuarine prawns (Teske *et al.* 2007c; Teske *et al.* 2009), intertidal limpets (Ridgway *et al.* 1998; Teske *et al.* 2007a), spiny lobsters (Gopal *et al.* 2006; Singh *et al.* 2019) and mussels (Zardi *et al.* 2007). High biodiversity in the study area is further enhanced by diverse coastal and shelf habitat structures such as canyons, coral reefs, mangrove forests, sand dunes, kelp forests, estuaries and seagrass beds (Keesing and Irvine 2005; Teske *et al.* 2011).

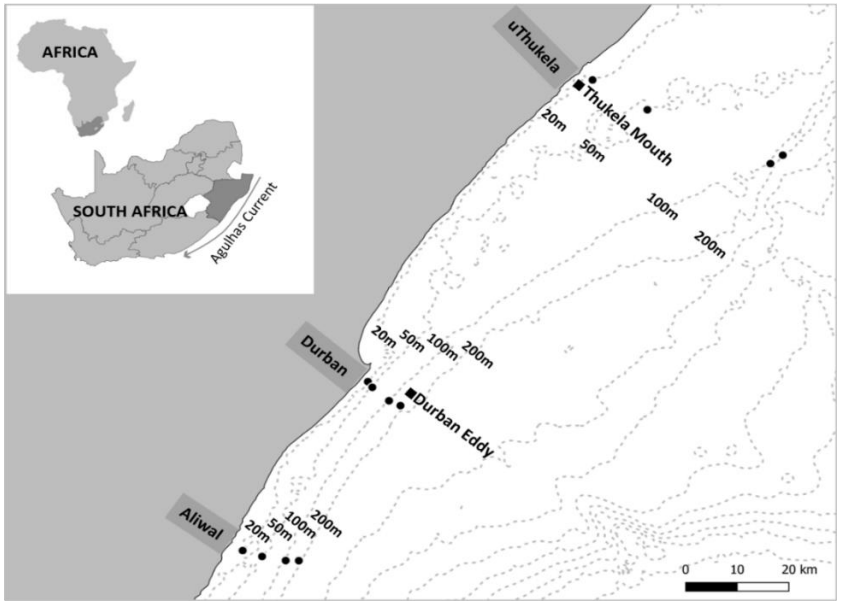


Figure 4.1 Map showing the location and sampling stations for the three sampling sites (uThukela, Durban and Aliwal) in the Indian Ocean. The main ocean current, Thukela River Mouth and Durban Eddy are depicted. The dotted lines depict the bathymetry along the east coast of South Africa.

Locations/characteristics of the three sampling sites: Cross-shelf transects were sampled at Thukela, Durban and Aliwal (Figure 4.1). Two transects fall within the KZN Bight (Thukela and Durban), while Aliwal is approximately 50 km south of the Bight. The recorded macrofaunal composition within the KZN Bight included: Annelida, Arthropoda, Mollusca, Echinodermata, Cnidaria, Sipuncula, Brachiopoda, Bryozoa, Chordata, Echiura, Nematoda, Nemertea and Platyhelminthes (MacKay *et al.* 2016). The recorded pelagic organisms were ichthyoplankton, fish eggs and juveniles from pelagic spawners (Hutchings *et al.* 2002). The mesoscale semi-permanent cyclonic eddy (see above) is responsible for the intrusion of cold, low-salinity and well-mixed nutrient-enriched water over the KZN Bight, and this region has been described as oligotrophic, mesotrophic and eutrophic, depending on the proximity to the coast, upwellings, Thukela River mouth, Agulhas Current edge and season (Bustamante and Branch 1996; Lutjeharms *et al.* 2000; Meyer *et al.* 2002; Fennessy *et al.* 2016; MacKay *et al.* 2016).

The Thukela site is near the center of the KZN Bight, at the mouth of the Thukela River (one of the largest rivers in South Africa), and forms part of the Thukela Marine Protected Area (MPA). The benthic habitat is comprised of sand, silt, muddy substrates, sandbanks, gravel beds and rocky pinnacles (Lutjeharms 2006; Untiedt 2013). The freshwater input from the Thukela River provides prompts for the spawning and recruitment of juvenile fish and invertebrates that use the KZN Bight as a nursery area (Forbes and Demetriades 2005; Lamberth *et al.* 2009). A highly diverse and abundant macrobenthic community structure has been reported for the uThukela site (Untiedt 2013; MacKay *et al.* 2016; Untiedt and MacKay 2016).

The Durban site is located at the southern region (downstream edge) of the KZN Bight, where the shelf has narrowed, and reefs dominate the benthic habitats at depths shallower than 30 m and a mixture of muddy, sandy and reef substrates in deeper waters (Pillay 2002; Pillay *et al.* 2008). Durban Bay is a functional nursery area for juvenile fish, particularly Clupeidae, Gobiidae, Engraulidae, Blenniidae, Tripterygiidae, Sparida, Myctophidae, Haemulidae and Leiognathidae (Beckley and Fennessy 1996; Harris and Cyrus 1999).

The Aliwal site is a subtidal, subtropical reef-forming part of the Aliwal MPA to the south of the KZN Bight, where the shelf has narrowed. The benthic habitat comprises of both hard and soft coral reefs, algae and sponges. The recorded macrofaunal composition at Aliwal includes Cnidaria,

Polychaeta, Bryozoa, Mollusca, Echinodermata, tunicates and fishes (Brash 2006; Schleyer *et al.* 2006; Olbers *et al.* 2009).

Sampling strategy

Zooplankton sampling was conducted at night (zooplankton migrate towards the surface at night) in September (uThukela and Durban) and November 2018 (Aliwal). The on-board Seabird SBE 19 plus V2 SeaCAT CTD sensor was deployed to measure salinity (ppt), temperature (°C), oxygen (ml/l) and pH vertical profiles of the water column, to approximately 10 m above the seafloor at each station. Samples were collected by towing a plankton ring-net (500 µm mesh; 0.8 m ring diameter) near the surface (<5 m depth) for 5 minutes at a boat speed of 2 – 3 knots. Cross-shelf transects were sampled at four stations per site (depth soundings of 20, 50, 100 and 200 m) with three replicate tows made at each station except at Durban, where two tows were made at 50 m and 200 m, respectively due to technical miscalculations (Figure 4.1, Table 4.1). Replicate tows were used to quantify the variability inherent in the method and sampling gear at the scale of the sampling station. Samples from each tow were immediately washed from the net's cod-end to a jar with 95% ethanol and stored at -20 °C until further processing.

Table 4.1 Sample collection data using a plankton ring-net (500 µm mesh; 0.8 m ring diameter) near the surface (between 0 – 5m depth) for 5 minutes at a boat speed of 2 – 3 knots. Sampling stations along transects were at depth soundings of 20, 50, 100 and 200 m, respectively. Three replicate tows were conducted per station, except at Durban, where two tows were made at 50 m and 200 m, respectively.

Sampling site	Sampling station(m)	Date	Time	Latitude	Longitude	Tow net replicates
uThukela	20	14/09/2018	20:17	-29.2605	31.5093	3
	50	14/09/2018	21:32	-29.3243	31.6199	3
	100	14/09/2018	23:42	-29.4395	31.8681	3
	200	15/09/2018	00:58	-29.4214	31.8932	3
Durban	20	06/09/2018	18:33	-29.9058	31.056	3
	50	06/09/2018	19:55	-29.9179	31.0648	2
	100	06/09/2018	21:10	-29.9466	31.0986	3
	200	06/09/2018	22:13	-29.9571	31.1216	2
Aliwal	20	15/11/2018	19:43	-30.267	30.8033	3
	50	15/11/2018	21:01	-30.2797	30.8425	3
	100	15/11/2018	22:10	-30.2883	30.8902	3
	200	15/11/2018	23:25	-30.2886	30.9166	3

Extraction and quantification of genomic DNA

Individual samples (n = 34) were homogenized for 45s using a consumer blender (Defy PB7354X, 350 W and 22000 rpm). The blender blades and container were sterilized with 4% industrial bleach and washed with 95% ethanol between applications. Three subsamples (triplicates) from each homogenized sample were taken to increase sequencing depth and improve diversity estimates (Lanzén *et al.* 2017). Subsamples consisted of 10 ml of zooplankton and were centrifuged at 3000 rpm for 1 minute; centrifugation was repeated to remove excess ethanol; thereafter, 40 mg of tissue was transferred to a sterile tube. DNA was extracted using the QIAGEN DNeasy Blood and Tissue Kit by adding 180 µl buffer ATL and 40 µl proteinase K to the tissue for overnight lysis at 56 ℃ and following the manufacturer's standard protocol for the purification of DNA from animal tissue. DNA extracted from the three subsamples (triplicates) were pooled for each sample and stored at -20 ℃.

PCR amplification, library preparation, and high-throughput DNA sequencing

First-round PCRs were performed in triplicate to address potential biases, artifacts and errors (Dopheide *et al.* 2019). Taxon-specific mini-barcode primer cocktails (n = 6 primer cocktails) were formulated after experimentation on samples with known zooplankton species composition and relative abundance (Chapter 3; Govender *et al*. (in review); Table S4.1). PCR reactions (25 µl) contained 0.25 µl Q5 High-Fidelity DNA Polymerase (0.02 U/ µl, New England BioLabs Inc), 5 µl Q5 reaction buffer (1X, New England BioLabs Inc), 5 µl Q5 high GC enhancer (1X, New England BioLabs Inc), 0.5 µl dNTP's (10 mM of each), 1 µl forward and reverse primers (5 µM), 1 µl template DNA (10 ng/ µl), 2 µl additional $MgCl_2$ (25 µM), 2 µl Bovine Serum Albumin (BSA) (1 mg.m-1) and nuclease-free water. Thermal cycling consisted of an initial denaturation step at 98 ℃ for 30 seconds, and 25 cycles of denaturation at 98 ℃ for 10 seconds, annealing at 46 ℃ for 30 seconds, extension at 72 ℃ for 30 seconds, and a final extension step at 72 ℃ for 4 minutes. Each round of PCR included a negative control which had no DNA present. PCR products were visualized on a 1 % (w/v) TBE agarose gel containing 0.02% Ethidium Bromide (EtBr), and the size was determined using a 100 bp molecular weight marker (Solis Biodyne). The three PCR products (triplicates) for each of the six primer cocktails were pooled and quantified using a Qubit 2.0 Fluorometer to create 34 libraries with equimolar concentrations (5 ng/µl).

Illumina sequencing was performed at the KZN Research and Innovation Platform (KRISP), South Africa. Each library was purified using 1.8X AmpureXP purification beads (Beckman Coulter, High Wycombe, UK). Index PCR was performed using the Nextera XT Index Kit (Illumina, San Diego, USA). Libraries were further purified using 0.6X AmpureXP purification beads (Beckman Coulter, High Wycombe, UK) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 4.0 instrument (Life Technologies, California, USA). The fragment sizes were analyzed using a LabChip GX Touch (Perkin Elmer, Hamburg, Germany) with an expected fragment size of 550 bp. Each sample library was normalized to 4nM concentration and denatured with 0.2N sodium acetate. A 5% PhiX control (PhiX Control v3) was spiked in each 12 pM library and sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) using a MiSeq Nano Reagent Kit v2 (500 cycles).

Taxonomic assignment of Amplicon Sequence Variants (ASVs)

The DADA2 algorithm (Callahan *et al.* 2016) implemented in QIIME2 v. 2019.10 (Bolyen *et al.* 2019) was used for quality control checks, chimera removal, filtering, trimming of primers, truncation of forward and reverse reads and merging of paired-end reads. Reference sequences of all Amplicon Sequence Variants (ASVs) generated using QIIME2 were queried against BOLD and GenBank. A 95% sequence identity threshold was used for taxonomic assignment to species level, taking the top hit for each species identification. To investigate the taxonomic identity with matches below 95% sequence identity, a multiple sequence alignment using all ASVs generated in this study was performed with the default settings in MAFFT v 7.470 (Katoh *et al.* 2019). A neighbor-joining (NJ) tree was constructed in MAFFT. The ASVs were assigned to species clusters using genetic distance (using the DNA barcode gap position as a threshold; Chapter 3; Govender *et al*. (in review)) and position on the NJ tree. Species clusters that could not be assigned to a species level (<95% sequence identity) were assigned to either a genus, family or order level. Before statistical analysis, clusters assigned to the same species were merged into single taxonomic units by summing up reads using phyloseq (McMurdie and Holmes 2013) in R v 4.0.2 (R development Core team) to prevent analyses of intraspecific variation. Rarefaction curves were generated using vegan v 2.5.6 (Oksanen *et al.* 2018) in R to determine whether samples were sequenced to a sufficient depth.

Analysis of species richness and relative abundance

Species richness (presence/absence) and relative abundance (read counts per species) were analyzed. Read counts were transformed to relative abundance per sample using the "conceptually simpler" total-sum scaling (TSS) normalization method (McMurdie and Holmes 2014) and was then used as a response variable in subsequent analyses. Differences in species composition between replicate tows were assessed visually using VennDiagram (Chen and Boutros 2011) drawn in R, and statistically with Jaccard similarity and Bray-Curtis dissimilarity matrices estimated in vegan v 2.5.6. For the Jaccard index, 0 signifies no overlap of species between communities, and 1 a perfect overlap, and the opposite for Bray-Curtis. Clusters obtained for replicate tows were combined for each sampling station for further analysis.

Species richness and relative abundance were compared between sites (uThukela, Durban and Aliwal) and between sampling stations (20, 50, 100 and 200 m depth soundings) using Phyloseq and plotted with ggplot2 (Wickham 2009). Several alpha-diversity (differences within a community) estimates, including richness (ACE and Chao1) and diversity (Shannon, Simpson, and Fisher's alpha) were calculated using phyloseq and plotted using ggplot2 for each site and their respective sampling stations.

The species richness and diversity between sites was assessed using the Bray-Curtis dissimilarity matrix to explore beta-diversity (differences between communities). Non-metric multidimensional scaling (NMDS) analyses based on the Bray-Curtis dissimilarity index was performed in phyloseq using benthic habitat structures as an explanatory variable. Permutational analyses of variance (PERMANOVA) was used to assess whether the benthic habitat structure at each site significantly affected zooplankton community composition. PERMANOVA was conducted using the Bray-Curtis dissimilarity matrix and the 'adonis' PERMANOVA function in vegan. In addition to ordination, a clustering analyses using the unweighted pair group average (UPGMA) in phyloseq was conducted using the Bray–Curtis dissimilarity matrix to visualize how sampling stations clustered. Connectivity between the sites was assessed using connectivity network analyses in phyloseq.

4.4 Results

Environmental parameters at each sampling site

Across all sampling stations for the three sites (n = 12), the surface (2-5 m depth) salinities oscillated between 35.1 – 35.4 ppt (mean ± SD = 35.3 ± 0.1 ppt), temperature ranged between 19.2 – 22.8 ℃ (20.9 ± 1.2℃), pH between 7.8 – 8.9 (8.7 ± 0.4) and oxygen between 5.8 – 6.7 mg/l (6.2 ± 0.3 mg/l) (Table 4.2). Depth profiles measured at each site's deepest station (Figure S4.1) showed that salinity remained constant between the surface and approximately 200 m below the surface at all three sites; temperature remained similar in surface waters, with an intense thermocline at approximately 80 m depth at Durban (water temp decreased from 21 to 15.5 ℃); a less intense thermocline at Thukela between 80 and 200 m depth; and well-mixed waters at Aliwal with the temperature declining gradually from the surface to deeper water. Surface waters were well-oxygenated up to at least 50 m depth at all three sites (>6 mg/l).

Table 4.2 Environmental parameters collected using the on-board Seabird SBE 19 plus V2 SeaCAT CTD sensor across all sampling stations. The CTD readings for each parameter were calculated between 2 – 5 m of the depth profile and averaged.

Sampling Site	Sampling station (m)	Salinity (ppt)	Temperature (℃)	pH	Oxygen (ml/l)
uThukela	20	35.1	21.1	8.9	6.5
	50	35.3	21.7	8.9	6.1
	100	35.3	22.5	8.9	6.0
	200	35.4	22.8	8.9	6.4
Durban	20	35.4	21.5	7.8	6.7
	50	35.4	21.3	7.8	6.7
	100	35.3	21.1	8.8	6.3
	200	35.3	21.1	8.8	6.3
Aliwal	20	35.4	19.6	8.9	6.1
	50	35.4	19.4	8.9	6.0
	100	35.4	19.2	8.9	6.0
	200	35.4	19.9	8.9	5.8
	Mean	**35.3**	**20.9**	**8.7**	**6.2**
	SD	**0.1**	**1.2**	**0.4**	**0.3**

High-throughput sequencing results

Sequencing was efficient across all three sampling sites with minimal filtering when merging the paired-end reads for all 34 zooplankton libraries (Table S4.2). High-throughput summary statistics (numbers of reads, ASVs and species clusters; Table 4.3) increased from north to south, with the fewest ASVs (274) and species clusters (50) identified at uThukela, intermediate at Durban (459 ASVs; 81 species clusters) and the most at Aliwal (593 ASVs; 132 species clusters). Species rarefaction curves showed adequate sequencing depth for all libraries (Figure 4.2). A total of 186 species clusters were found across sampling sites. Identified taxa were classified as Malacostraca (117 species), Actinopterygii (29 species), Copepoda (23 species), Gastropoda (8 species), Hydrozoa (4 species), Sagittoidea (2 species), Ostracoda (1 species), Thaliacea (1 species) and Branchiopoda (1 species). Some 70.4% (131 of 186) of species clusters could be matched with >95% sequence similarity to sequences on BOLD or GenBank.

Table 4.3 High-throughput summary statistics across the replicates collected at uThukela, Durban and Aliwal.

Site	Read Count	Merged reads	Paired reads	ASVs	Species per site (shared and unique)
uThukela	476 032	32 824	16 412	274	50
Durban	1 354 634	37 744	18 872	459	81
Aliwal	112 416	12 893	6 447	593	132
Total	1 943 082	83 461	41 731	1 326	-

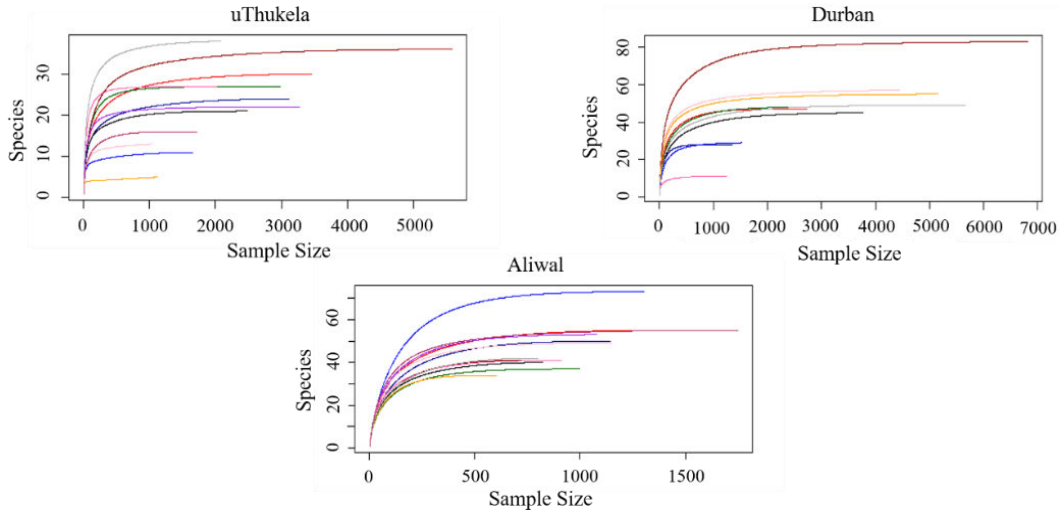


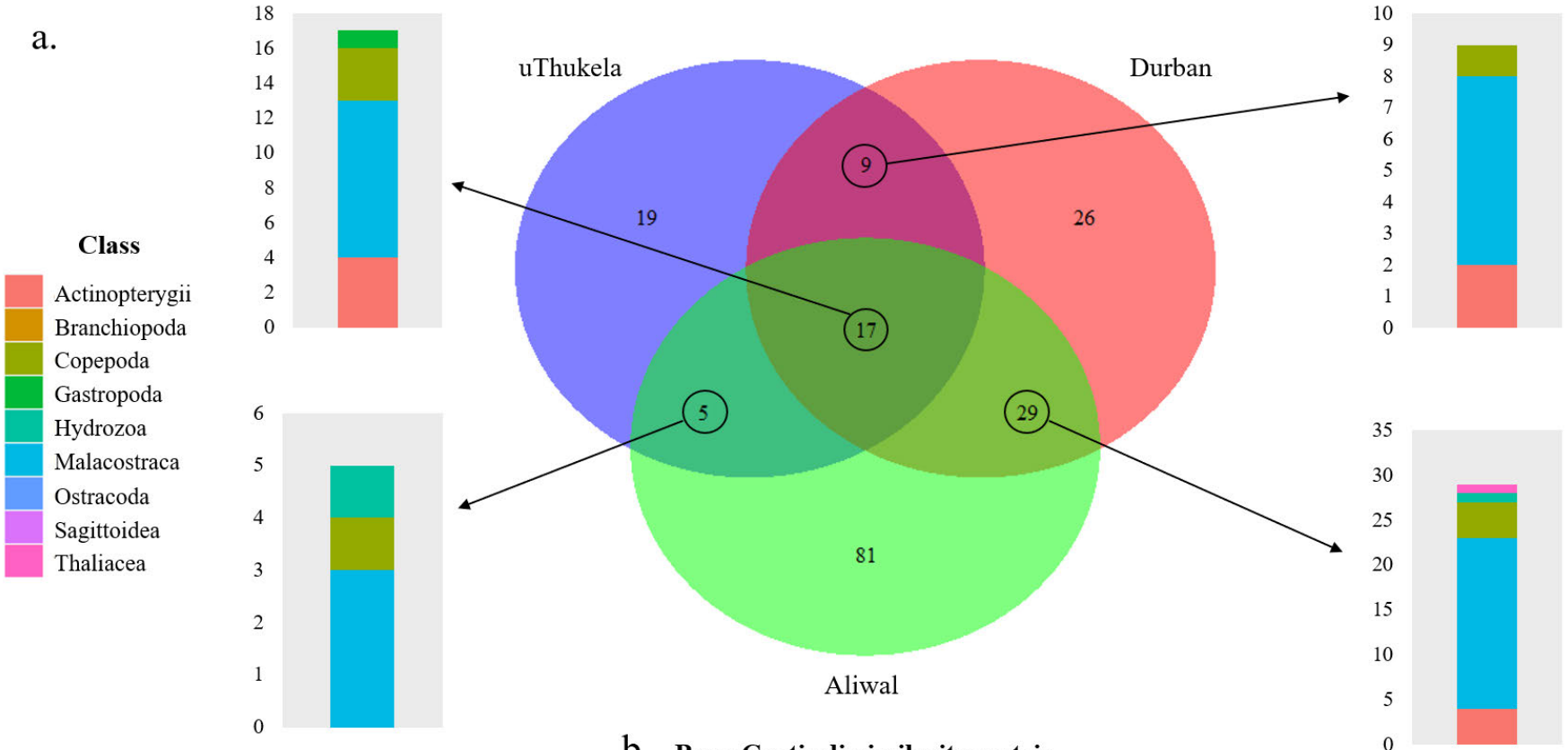
Figure 4.2 Rarefaction curves for the three sampling sites (uThukela, Durban and Aliwal) and their respective sampling stations.

Community composition

Samples collected from replicate tows at each station shared 18 – 55% of species clusters at uThukela; 25 – 58% at Durban and 45 – 48% at Aliwal (Figure S4.2) and were supported by high Jaccard similarity and low Bray-Curtis dissimilarity values per station (Table S4.3). Samples collected from each sampling station at the different depths shared 20% of species clusters at uThukela, 44% at Durban and 44% at Aliwal (Figure S4.3). Only 9% of species clusters were shared between the three sampling sites (taxa belonging to Actinopterygii, Malacostraca, Copepoda and Gastropoda), and based on the Bray-Curtis dissimilarity index there was a significant difference between species richness observed at uThukela and Aliwal; and Durban and Aliwal (Figure 4.3). Out of the 186 identified species clusters, 10% were found exclusively at uThukela, 14% at Durban, but this increased nearly 4-fold to 44% at Aliwal. uThukela and Durban shared 5% of species clusters, uThukela and Aliwal shared 3% of the species clusters, and Durban and Aliwal shared 16% of the species clusters (Figure 4.3, Table S4.4).

The three most sequenced zooplankton groups based on species richness across all three sites were Malacostraca (63%), Actinopterygii (16%) and Copepoda (12%). However, Actinopterygii accounted for 52% of reads based on the overall relative abundance (Figure 4.4 a). Based on the species composition plots (Figure 4.4 b, c) and the alpha-diversity estimates (Figure S4.4), significant differences in species richness, diversity and abundance were found among all three sites. Overall, the lowest species richness (Chao1 = 59.43, ACE = 63.55) and diversity (Shannon = 0.97, Simpson = 0.40, Fisher = 6.86) was recorded at uThukela, where Actinopterygii accounted for 76% of the reads based on relative abundance. An intermediate species richness (Chao1 = 92.14, ACE = 90.85) and diversity (Shannon = 2.27, Simpson = 0.79, Fisher = 12.05) was recorded for Durban, where both Malacostraca and Actinopterygii dominated the proportion of reads. The highest species richness (Chao1 = 132, ACE = 132) and diversity (Shannon = 3.41, Simpson = 0.94, Fisher = 21.48) was recorded at Aliwal, which was highly dominated by Malacostraca and Copepoda (Figure 4.4 b, c).

a.



b. **Bray-Curtis dissimilarity matrix**

	uThukela	Durban	Aliwal
uThukela	-	0.40	0.90
Durban	0.40	-	0.79
Aliwal	0.90	0.79	-

Figure 4.3 (a) A Venn Diagram with species composition plots of shared species clusters and (b) beta-diversity indices comparing the difference between the three sites (uThukela, Durban and Aliwal).

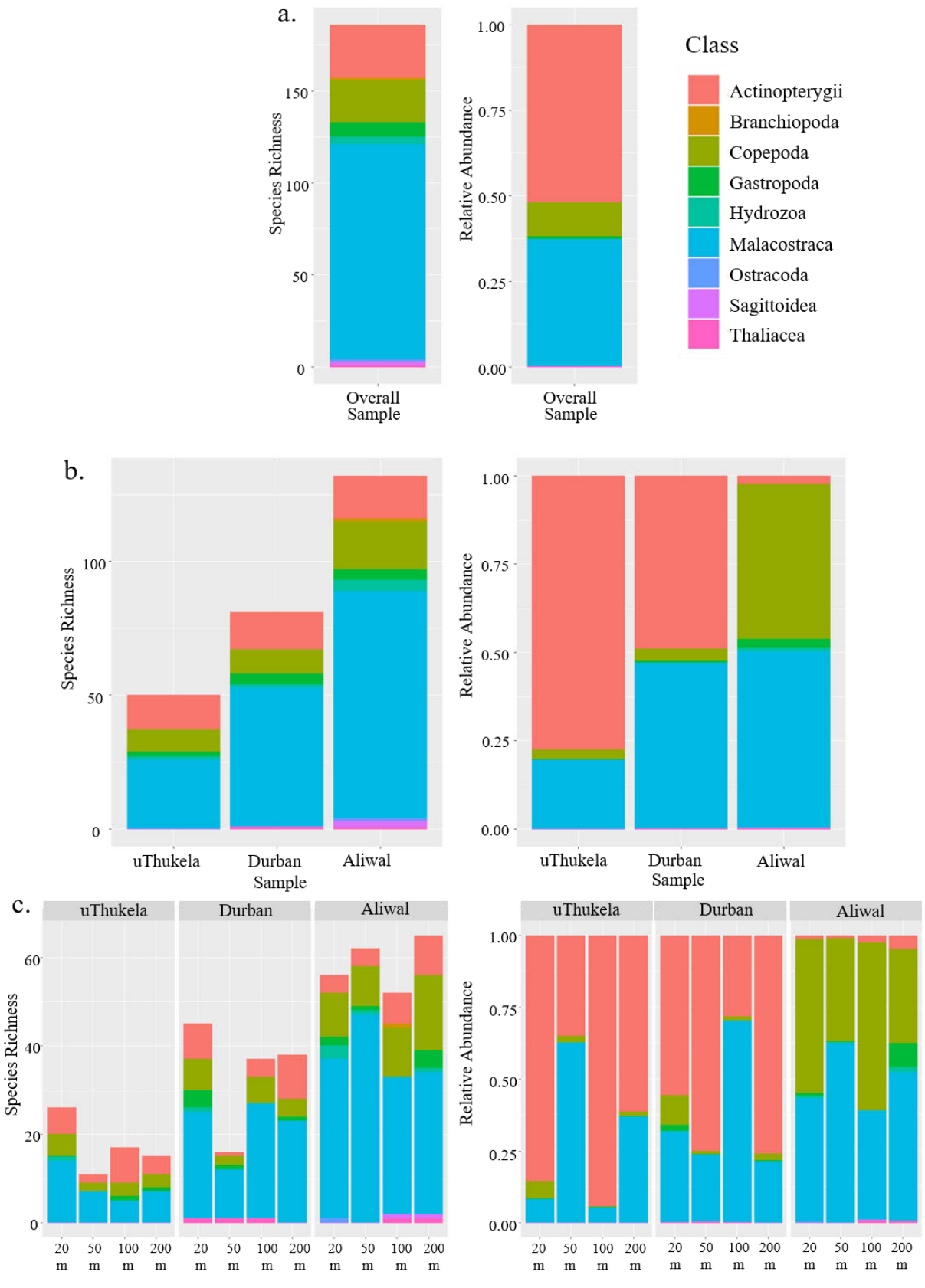


Figure 4.4 Species composition bar graphs (a) all the sites combined (overall), (b) across each site (uThukela, Durban and Aliwal) and (c) across each sampling station.

Community connectivity

The NMDS ordination plot (Figure 4.5 a, stress value = 0.095) indicated an overlap between the uThukela and Durban communities while Aliwal is very different. Similar results were seen in the UPGMA cluster analysis (Figure 4.5 b). The PERMANOVA test indicated that the community composition differed significantly between the three sampling sites (p = 0.002**, R2 = 0.569, F = 5.945). Connectivity network projections using the Bray-Curtis dissimilarities connected uThukela and Durban, while Aliwal is isolated (Figure 4.6).

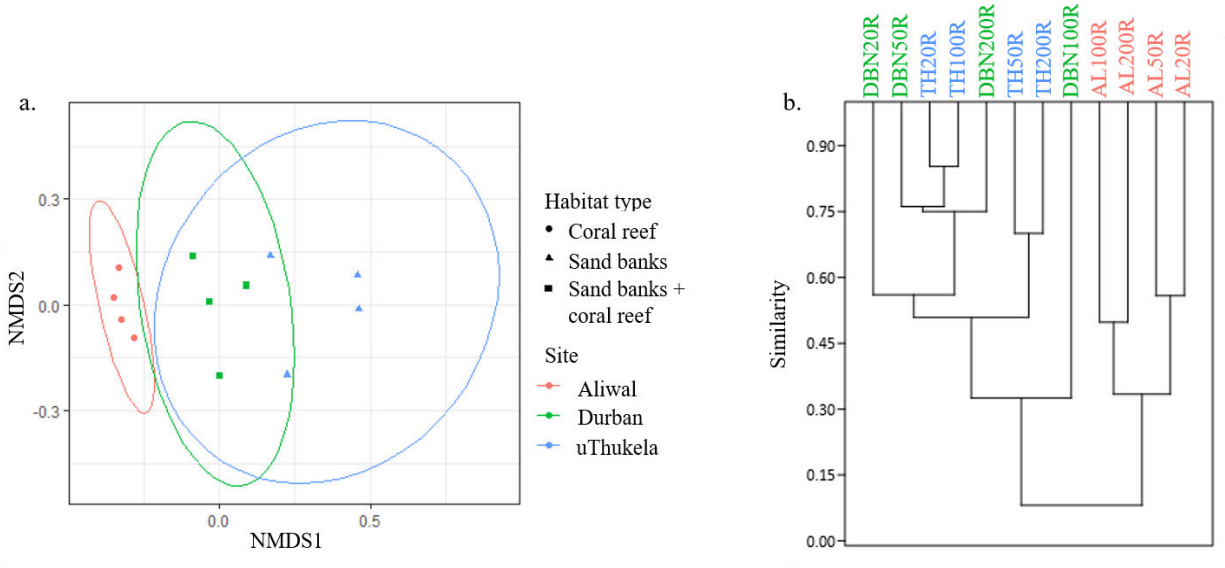


Figure 4.5 Beta-diversity analysis to estimate the dissimilarity and similarity between the uThukela, Durban and Aliwal and their respective sampling stations. (A) NMDS ordination plot based on Bray-Curtis dissimilarity, stress value = 0.095555 (B) UPGMA cluster analysis based on Bray-Curtis dissimilarity.

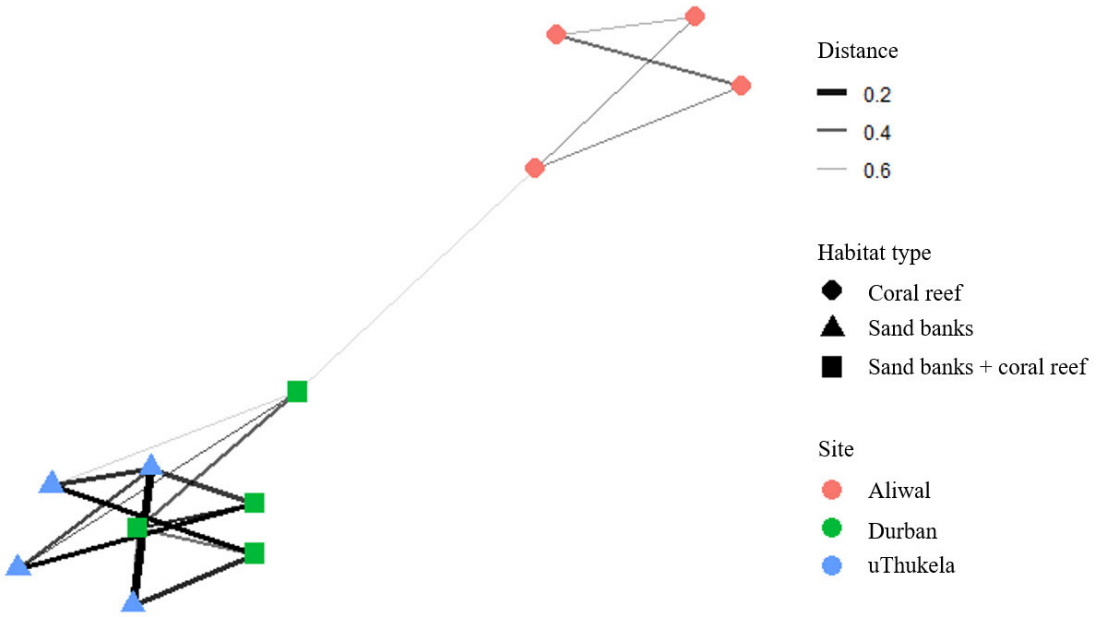


Figure 4.6 Connectivity network between the sampling sites (uThukela, Durban and Aliwal) and their respective sampling stations.

4.5 Discussion

In this study, the diversity and connectivity of marine zooplankton across three sites situated along the east coast of South Africa was assessed. DNA metabarcoding was used to help understand the influence of the Agulhas Current, seafloor substrates and benthic habitat structures on zooplankton community structure and connectivity at the different sites. DNA metabarcoding has recently been used in other studies (see Macher et al. 2020; Pitz et al. 2020), but to the best of our knowledge, this is the first time it has been applied to zooplankton communities occurring over the KZN continental shelf.

The results in this study were represented using species richness and relative abundance, where the proportion of reads was set as the response variable. Proportional data is used here as an effective way of controlling for differences in read counts. Many studies have only assessed the species richness, as PCR-based techniques may not sufficiently approximate each taxon's relative abundance (Elbrecht and Leese 2015; Piñol *et al.* 2015; Jusino *et al.* 2019). However, Nichols *et al.* (2018) emphasized the advantage of taxon abundance estimates in DNA metabarcoding and biodiversity studies, as abundance estimates are essential for assessing diversity indices and

ecosystem health status. Studies have shown that DNA metabarcoding can successfully determine the relative abundance of complex nematode communities (Schenk *et al.* 2019) and arthropod taxa in songbird diets (Verkuil *et al.* 2020)... In addition, the use of relative abundance has been applied to assessing the differences in natural zooplankton communities in the North Sea (Macher *et al.* 2020) and California Current System (Pitz *et al.* 2020). Both studies used the same universal COI mini-barcode primer set by Leray *et al.* (2013); the results from these studies showed that the abundance data inferred through DNA metabarcoding were mostly matched with the known distribution of species described in previous studies. Therefore, we used the relative abundance of species-specific data in a semi-quantitative manner in this study, albeit with caution.

DNA metabarcoding combined with community-level analyses has previously been applied to biogeographical studies of marine zooplankton. Macher *et al.* (2020) applied COI DNA metabarcoding to zooplankton samples collected from the North Sea to assess community composition in the northern (influence by an inflow of oceanic Atlantic waters) and southern (mostly made up of coastal waters) regions. Pitz *et al.* (2020) applied both 18S rRNA and COI DNA metabarcoding to zooplankton samples collected within the California Current System, which runs southward off the western coast of North America. Their study confirmed a clear shift in community composition congruent with two well-documented biogeographic boundaries separating strikingly different biological communities.

The species rarefaction curves in this study leveled off as the sequencing depth increased at all three sites, implying that the data approached saturation and that sampling stations were sufficiently sampled. A total of 186 species clusters were detected across the three sampling sites of which 70.4% could be assigned to species-level after comparison with publicly available reference libraries. Singh *et al.* (in press) could only assign 32% of detected species clusters to species-level despite using similar gear and samples collected along the KZN coast. The present study used taxon-specific primers for PCR amplification, sequences were queried against GenBank and BOLD and a 95% cut-off for sequence identity was used. In contrast, Singh *et al.* (in press) used a universal primer set, sequences were queried only against BOLD and a stricter 97% cut-off was used for sequence identity. Therefore, using both GenBank and BOLD is important as it increases the number of sequences assigned to a correct species-level. Identifying and barcoding locally collected species, including species endemic to South Africa, should be a

priority research area to allow for successful DNA metabarcoding and mass species-level identifications.

Patchiness is one of the most prominent characteristics of zooplankton populations in marine environments (McGillicuddy 2001), and the level of heterogeneity can cover many orders of magnitude across spatio-temporal scales. The effects of patchiness on estimates of species richness and relative abundance can be partially overcome by replicate sampling at smaller spatial scales. In the present study, biological replicates collected at individual stations showed 18 – 58% overlap in species clusters across sites. An increase in the numbers of replicates per station is encouraged for future studies to increase the numbers of species clusters and reduce the effects of patchiness on species richness and relative abundance estimates.

Overall, the highest relative abundance (proportional number of reads) was recorded for Actinopterygii, Malacostraca and Copepoda. The Actinopterygii cluster comprised of a very high proportion of one species, *Scomber japonicus*, a small pelagic fish that spawns mainly during winter and spring (August to November) (Beckley and Leis 2000; Connell 2001). *S. japonicus* is a broadcast spawner that produces large numbers of small eggs and larvae (Hutchings *et al.* 2002) that drift in the water column, where they form dense patches in the zooplankton. The high relative abundance of *S. japonicus* (and hence Actinopterygii) in this study can therefore be explained by large numbers of fish eggs collected in tow nets during field sampling in September and November 2018 – during the spawning season. The relative abundance of Actinopterygii dominated the uThukela site, where ichthyoplankton and juveniles of pelagic spawners are abundant over the KZN Bight and near the Thukela River mouth (Beckley 1993; Hutchings *et al.* 2002; Lamberth *et al.* 2009).

This study hypothesized that zooplankton communities over the KZN shelf would have similar species richness and diversity because of the homogenizing (mixing) influence of the strong Agulhas Current. We assumed that larval dispersal of zooplankton communities would be linked to major physical drivers resulting in the connectivity between these three geographically separated regions. In contrast, we found that species richness and diversity differed across all three sampling sites and the null hypothesis was therefore rejected.

Alternatively, we proposed that a strong benthic-pelagic coupling effect would result in heterogenous species richness and diversity between uThukela (low richness), Durban

(intermediate) and Aliwal (high). Baustian *et al.* (2014) identified three significant mechanisms of benthic-pelagic coupling: (1) organism movement, (2) trophic interactions and (3) biogeochemical cycling. Griffiths *et al.* (2017) defined benthic-pelagic coupling as processes that connect the bottom substrate and the water column habitats through the exchange of mass, energy and nutrients. Therefore, nearby benthic habitats are the main drivers of species composition in pelagic waters (Heck Jr *et al.* 2003; Dahlgren *et al.* 2006). A study focusing on animal forests (soft and hard corals, sponges, bryozoans and other animals that are considered eco-engineering species) suggests that habitat structures play an essential role in benthic-pelagic coupling processes (Rossi *et al.* 2017).

Benthic-pelagic coupling processes can explain the lower zooplankton species richness and diversity at uThukela compared to the other two sites, because its benthic environment is comprised of muddy substrates (Untiedt 2013). Muddy substrates are relatively unstable with few areas where sessile invertebrates can anchor and few niches for larger benthic invertebrates and fishes, hence a lower species diversity is expected in the benthic water body (Branch *et al.* 1981). Freshwater, suspended sediments, nutrients and pollutants originating from the uThukela River can further affect nearshore bentho-pelagic habitats (Meyer *et al.* 2002), by reducing ecosystem health and the diversity and abundance of benthic taxa with pelagic life-history phases (Thrush and Dayton 2002; Lohrer *et al.* 2004).

The benthic environment could explain an increase in zooplankton species richness and diversity at Durban. The benthic environment was more complex and stabilized by rocky outcrops and reefs, providing firmer substrates and structures for benthic organisms to anchor themselves on (Branch *et al.* 1981). A study on zooplankton biomass (Pretorius *et al.* 2016) recorded low zooplankton biomass at the Thukela River mouth and higher biomass near Durban, suggesting that nutrient enrichment from the quasi-permanent upwelling off Durban had a more significant influence on zooplankton biomass than the seasonal nutrient input originating from the Thukela River. Zooplankton species richness and overall biomass, therefore, appeared to be higher near Durban than at uThukela.

The benthic environment could explain a further increase in zooplankton species richness and diversity at Aliwal. Aliwal the most complex benthic habitat structure, dominated by a highly variable coral reef (Brash 2006; Schleyer *et al.* 2006; Olbers *et al.* 2009). The benthic ecosystem's

high biodiversity leads to increased larvae diversity being released into the pelagic water column. Furthermore, Aliwal has been described as a marginal environment that hosts a broad range of highly diverse organisms from tropical, subtropical and warm-temperate regions (Schleyer 2000; Olbers *et al.* 2009).

In addition to benthic environments, the coupling of benthic and pelagic life stages of marine invertebrates is central to understanding their population dynamics and is influenced by hydrodynamic processes such as water movements at multiple spatial scales (Porri *et al.* 2014). Previously, the north to south Agulhas Current was seen as a major mechanism for southward dispersal of recruits along the east coast of South Africa (Heydorn *et al.* 1978; Garratt 1988). This study (and others) suggest that it was not the Current itself but rather other sub-mesoscale processes at its shoreward edge that facilitated larval retention over the shelf or southward dispersal (Beckley 1993; Hutchings *et al.* 2002). Inshore countercurrents between Durban and uThukela in the KZN Bight associated with the semi-permanent Durban Eddy can expedite northward dispersal of marine fauna, including zooplankton (Pearce 1977; Beal and Bryden 1997; Lutjeharms 2006; Roberts *et al.* 2010; Guastella and Roberts 2016; Roberts and Nieuwenhuys 2016), resulting in a semi-closed circulation system within the KZN Bight (Lutjeharms *et al.* 1989; Lutjeharms *et al.* 2000; Meyer *et al.* 2002). The semi-closed circulation system in the KZN Bight can explain the much stronger connectivity signal between uThukela and Durban, while Aliwal which lies outside the KZN Bight was the least connected site.

The concordance between marine biogeographic and phylogeographic boundaries is increasingly recognized in the literature (Teske *et al.* 2006; Spalding *et al.* 2007; Teske *et al.* 2007b; Teske *et al.* 2009; Bowen *et al.* 2016). Therefore, the differences in species composition at Aliwal could further be attributed to the site being situated in a subtropical reef within a transition zone between the tropical/subtropical Maputaland reefs (Schleyer 2000; Brash 2006; Olbers *et al.* 2009) and warm-temperate Pondoland reefs (Brash 2006; Mann *et al.* 2006; Celliers *et al.* 2007), whereas uThukela and Durban are situated solely within the subtropical biogeographic region. However, further sampling would need to be carried out to support these findings.

In conclusion, DNA metabarcoding of marine zooplankton samples combined with community-level analyses provided valuable insights into the species richness, diversity, dispersal and connectivity of planktonic animals in a biodiversity rich marine system. The use of

community-level analysis provides more informative assessments on the connectivity between marine environments rather than focusing on a single species. We suggest that benthic-pelagic coupling structured zooplankton communities, rather than homogenization caused by mixing through water movements. Our results were consistent with previous studies (Hutchings *et al.* 2002; Pretorius *et al.* 2016) and provides more evidence highlighting the connectivity between uThukela and Durban and the isolation of Aliwal in terms of planktonic larval distribution. This study also provides important baseline data for future biodiversity monitoring campaigns which are valuable for biodiversity studies in times of rapid ocean and climate change.

4.6 Acknowledgments

We would like to acknowledge funding support from DSI/NRF/ACEP Captor Project (Grant 110763). The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily attributed to the NRF. We would also like to thank David Pearton, Jennifer Giandhari and Simon Creer for their assistance and advice.

Data Accessibility

All the scripts used in this chapter are available on figshare: https://doi.org/10.6084/m9.figshare.14378663.v1

4.7 References

Almany G., Connolly S., Heath D., Hogan D., Jones G., McCook L., Mills M., Pressey R. and Williamson D. (2009) Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. Coral Reefs 28, 339-351.

Bang N. and Pearce A. (1976) Large-scale circulation of surface water of the south Indian Ocean. In: Ecology of the Agulhas current region - An assessment of biological responses to environmental parameters in the South-West Indian Ocean. Heydorn A.E.F. (eds). pp. 4-10. Proceedings of the Marine Freshwater Conference, CSIR, Pretoria.

Baustian M., Hansen G., De Kluijver A., Robinson K., Henry E., Knoll L., Rose K. and Carey C. (2014) Linking the bottom to the top in aquatic ecosystems: Mechanisms and stressors of benthic-pelagic coupling. In: Eco-DAS X Symposium Proceedings. Kemp P.F. (eds). pp. 25-47. Association for the Sciences of Limnology and Oceanography, Waco.

Beal L.M. and Bryden H.L. (1997) Observations of an Agulhas undercurrent. Deep Sea Research Part I: Oceanographic Research Papers 44, 1715-1724.

Beal L.M. and Elipot S. (2016) Broadening not strengthening of the Agulhas Current since the early 1990s. Nature 540, 570-573.

Beckley L.E (1993) Linefish larvae and the Agulhas Current. In: Fish, fishers and fisheries Proceedings of the second South African marine linefish symposium, Durban. Beckley L.E. and van der Elst R.P. (eds). pp. 57-63. Oceanographic Research Institute, Durban.

Beckley L.E. and Leis J. (2000) Occurrence of tuna and mackerel larvae (Family: Scombridae) off the east coast of South Africa. Marine and Freshwater Research 51, 777-782.

Beckley L.E. (1986) The ichthyoplankton assemblage of the Algoa Bay nearshore region in relation to coastal zone utilization by juvenile fish. South African Journal of Zoology 21, 244-252.

Beckley L.E. (1995) The Agulhas Current ecosystem with particular reference to dispersal of fish larvae. In: Status and future of large marine ecosystems of the Indian Ocean: A report of the international symposium and workshop. Okemwa E., Ntiba M.J. and Sherman K. (eds). pp. 74-91. International Union for Conservation of Nature (IUCN), Gland.

Beckley L.E. and Fennessy S. (1996) The beach-seine fishery off Durban, KwaZulu-Natal. South African Journal of Zoology 31, 186-192.

Beckley L.E. and Hewitson J.D. (1994) Distribution and abundance of clupeoid larvae along the east coast of South Africa in 1990/91. South African Journal of Marine Science 14, 205-212.

Berry P.F. (1974) Palinurid and scyllarid lobster larvae of the Natal Coast, South Africa. Oceanographic Research Institute, Durban. Investigational report no. 34: 1-44.

Bishop M.J., Mayer-Pinto M., Airoldi L., Firth L.B., Morris R.L., Loke L.H.L., Hawkins S.J., Naylor L.A., Coleman R.A., Chee S.Y. and Dafforn K.A. (2017) Effects of ocean sprawl on ecological connectivity: Impacts and solutions. Journal of Experimental Marine Biology and Ecology 492, 7-30.

Bolyen E., Rideout J.R., Dillon M.R., Bokulich N.A., Abnet C.C., Al-Ghalith G.A., Alexander H., Alm E.J., Arumugam M., Asnicar F., Bai Y., Bisanz J.E., Bittinger K., Brejnrod A., Brislawn C.J., Brown C.T., Callahan B.J., Caraballo-Rodriguez A.M., Chase J., Cope E.K., Da Silva R., Diener C., Dorrestein P.C., Douglas G.M., Durall D.M., Duvallet C.,

Edwardson C.F., Ernst M., Estaki M., Fouquier J., Gauglitz J.M., Gibbons S.M., Gibson D.L., Gonzalez A., Gorlick K., Guo J., Hillmann B., Holmes S., Holste H., Huttenhower C., Huttley G.A., Janssen S., Jarmusch A.K., Jiang L., Kaehler B.D., Kang K.B., Keefe C.R., Keim P., Kelley S.T., Knights D., Koester I., Kosciolek T., Kreps J., Langille M.G.I., Lee J., Ley R., Liu Y.X., Loftfield E., Lozupone C., Maher M., Marotz C., Martin B.D., McDonald D., McIver L.J., Melnik A.V., Metcalf J.L., Morgan S.C., Morton J.T., Naimey A.T., Navas-Molina J.A., Nothias L.F., Orchanian S.B., Pearson T., Peoples S.L., Petras D., Preuss M.L., Pruesse E., Rasmussen L.B., Rivers A., Robeson M.S., 2nd, Rosenthal P., Segata N., Shaffer M., Shiffer A., Sinha R., Song S.J., Spear J.R., Swafford A.D., Thompson L.R., Torres P.J., Trinh P., Tripathi A., Turnbaugh P.J., Ul-Hasan S., van der Hooft J.J.J., Vargas F., Vazquez-Baeza Y., Vogtmann E., von Hippel M., Walters W., Wan Y., Wang M., Warren J., Weber K.C., Williamson C.H.D., Willis A.D., Xu Z.Z., Zaneveld J.R., Zhang Y., Zhu Q., Knight R. and Caporaso J.G. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37, 852-857.

Bonte D., Baert L., Lens L. and Maelfait J.P. (2004) Effects of aerial dispersal, habitat specialisation, and landscape structure on spider distribution across fragmented grey dunes. Ecography 27, 343 - 349.

Bowen B.W., Gaither M.R., DiBattista J.D., Iacchei M., Andrews K.R., Grant W.S., Toonen R.J. and Briggs J.C. (2016) Comparative phylogeography of the ocean planet. Proceedings of the National Academy of Sciences 113, 7962-7969.

Branch G., Branch M. and Bannister A. (1981) The living shores of Southern Africa. Struik Publishers, Cape Town.

Brash J.M. (2006) Zonation of the benthic communities on Aliwal Shoal. MSc Thesis, University of KwaZulu-Natal, South Africa.

Bucklin A., Lindeque P.K., Rodriguez-Ezpeleta N., Albaina A. and Lehtiniemi M. (2016) Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. Journal of Plankton Research 38, 393-400.

Buecher E., Goy J. and Gibbons M. (2005) Hydromedusae of the Agulhas Current. African Invertebrates 46, 27-69.

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

Butler M.I., Paris C.B., Goldstein J.S., Matsuda H. and Cowen R.K. (2011) Behavior constrains the dispersal of long-lived spiny lobster larvae. Marine Ecology Progress Series 422, 223-237.

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A. and Holmes S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13, 581-583.

Carr M.H., Robinson S.P., Wahle C., Davis G., Kroll S., Murray S., Schumacker E.J. and Williams M. (2017) The central importance of ecological spatial connectivity to effective coastal marine protected areas and to meeting the challenges of climate change in the marine environment. Aquatic Conservation: Marine and Freshwater Ecosystems 27, 6-29.

Carter R. (1977) The distribution of calanoid Copepoda in the Agulhas Current system off Natal, South Africa. MSc Thesis, University of KwaZulu-Natal, South Africa.

Celliers L., Mann B.Q., Macdonald A.H.H. and Schleyer M.H. (2007) A benthic survey of the rocky reefs off Pondoland, South Africa. African Journal of Marine Science 29, 65-77.

Chen H. and Boutros P.C. (2011) VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35.

Collocott S.J. (2016) Patterns and influencing factors of the larval fish assemblage of the KwaZulu-Natal Bight, South Africa. MSc Thesis, University of KwaZulu-Natal, South Africa.

Connell A.D. (2001) Pelagic eggs of marine fishes from Park Rynie, KwaZulu-Natal, South Africa: seasonal spawning patterns of the three most common species. African Zoology 36, 197-204.

Cowen R.K., Lwiza K.M.M., Sponaugle S., Paris C.B. and Olson D.B. (2000) Connectivity of marine populations: Open or closed? Science 287, 857-859.

Cowen R.K. and Sponaugle S. (2009) Larval dispersal and marine population connectivity. Annual Review of Marine Science 1, 443-466.

Creer S., Deiner K., Frey S., Porazinska D., Taberlet P., Thomas W.K., Potter C. and Bik H.M. (2016) The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7, 1008-1018.

Dahlgren C.P., Kellison G.T., Adams A.J., Gillanders B.M., Kendall M.S., Layman C.A., Ley J.A., Nagelkerken I. and Serafy J.E. (2006) Marine nurseries and effective juvenile habitats: Concepts and applications. Marine Ecology Progress Series 312, 291-295.

De Decker A. (1964) Observations on the ecology and distribution of copepoda in the marine plankton of South Africa. Investigational Report, Division of Sea Fisheries, South Africa 49, 1-33.

De Decker A. (1984) Near-surface copepod distribution in the south-western Indian and south-eastern Atlantic Ocean. Annals of the South African Museum 93, 303-370.

De Decker A. and Mombeck F. (1964) South African contribution to the International Indian Ocean Expedition. A preliminary report on the planktonic Copepods. Investigational Report, Division of Sea Fisheries, South Africa 51, 10-67.

Dopheide A., Xie D., Buckley T.R., Drummond A.J. and Newcomb R.D. (2019) Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution 10, 120-133.

Elbrecht V. and Leese F. (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass - sequence relationships with an innovative metabarcoding protocol. PLOS One 10, e0130324.

Fennessy S.T., Roberts M.J. and Paterson A.W. (2016) A brief overview of the ACEP project: Ecosystem processes in the KwaZulu-Natal Bight. African Journal of Marine Science 38, S1-S6.

Forbes A.T. and Demetriades N.T. (2005) A review of the commercial, shallow water penaeid prawn resource of South Africa: Status, fisheries, aquaculture and management. Specialist report for Ezemvelo KZN Wildlife, KwaZulu-Natal, South Africa.

Gaines S. and Lafferty K. (1995) Modeling the dynamics of marine species: The importance of incorporating larval dispersal. In: Ecology of marine invertebrate larvae. McEdward L.R. (eds). pp. 389-412. CRC Press, Cleveland.

Garratt P.A. (1988) Notes on seasonal abundance and spawning of some important offshore linefish in Natal and Transkei waters, southern Africa. South African Journal of Marine Science 7, 1-8.

Gilg M. and Hilbish T. (2003) The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. Ecology 84, 2989-2998.

Gopal K., Tolley K., Groeneveld J. and Matthee C. (2006) Mitochondrial DNA variation in spiny lobster *Palinurus delagoae* suggests genetically structured populations in the southwestern Indian Ocean. Marine Ecology Progress Series 319, 191-198.

Govender A., Groeneveld J., Singh S. and Willows-Munro S. (2019) The design and testing of mini-barcode markers in marine lobsters. PLOS One 14, e0210492.

Govender A., Singh S., Groeneveld J., Pillay S. and Willows-Munro S. (In review) Marine zooplankton, mini-barcodes, and DNA metabarcoding: The case for taxon-specific primers.

Greene C.H. (1983) Selective predation in freshwater zooplankton communities. Internationale Revue der gesamten Hydrobiologie und Hydrographie 68, 297-315.

Griffiths J.R., Kadin M., Nascimento F.J.A., Tamelander T., Törnroos A., Bonaglia S., Bonsdorff E., Brüchert V., Gårdmark A., Järnström M., Kotta J., Lindegren M., Nordström M.C., Norkko A., Olsson J., Weigel B., Žydelis R., Blenckner T., Niiranen S. and Winder M. (2017) The importance of benthic–pelagic coupling for marine ecosystem functioning in a changing world. Global Change Biology 23, 2179-2196.

Groeneveld J.C., von der Heyden S. and Matthee C.A. (2012) High connectivity and lack of mtDNA differentiation among two previously recognized spiny lobster species in the southern Atlantic and Indian Oceans. Marine Biology Research 8, 764-770.

Guastella L. and Roberts M. (2016) Dynamics and role of the Durban cyclonic eddy in the KwaZulu-Natal Bight ecosystem. African Journal of Marine Science 38, S23-S42.

Gutierrez M.F., Tavsanoglu U.N., Vidal N., Yu J., Mello F.T., Cakiroglu A., He H., Liu Z. and Jeppesen E. (2018) Salinity shapes zooplankton communities and functional diversity and has complex effects on size structure in lakes. Hydrobiologia 813, 237-255.

Hanski I., Kuussaari M. and Nieminen M. (1994) Metapopulation structure and migration in the butterfly Melitaea cinxia. Ecology 75, 747-762.

Harris S.A. and Cyrus D.P. (1999) Composition, abundance and seasonality of fish larvae in the mouth of Durban Harbour, KwaZulu-natal, South Africa. South African Journal of Marine Science 21, 19-39.

Havens K.E. and Hanazato T. (1993) Zooplankton community responses to chemical stressors: A comparison of results from acidification and pesticide contamination research. Environmental pollution 82, 277-288.

Hays G.C. (2017) Ocean currents and marine life. Current Biology 27, R470-R473.

Hays G.C., Richardson A.J. and Robinson C. (2005) Climate change and marine plankton. Trends in Ecology and Evolution 20, 337-344.

Heck Jr K., Hays G. and Orth R.J. (2003) Critical evaluation of the nursery role hypothesis for seagrass meadows. Marine Ecology Progress Series 253, 123-136.

Heydorn A., Bang N., Pearce A., Flemming B., Carter R., Schleyer M., Berry P., Hughes G., Bass A., Wallace J., Elst R. and Crawford R. (1978) Ecology of the Agulhas Current region: An assessment of biological responses to environmental parameters in the south-west Indian Ocean. Transactions of The Royal Society of South Africa 43, 151-190.

Hidalgo M., Kaplan D.M., Kerr L.A., Watson J.R., Paris C.B. and Browman H.I. (2017) Advancing the link between ocean connectivity, ecological function and management challenges. ICES Journal of Marine Science 74, 1702-1707.

Huggett J. and Kyewalyanga M. (2017) Ocean productivity. In: The RV Dr Fridtjof Nansen in the Western Indian Ocean: Voyages of marine research and capacity development. Groeneveld J.C. and Koranteng K.A. (eds). pp. 55-80. Food and Agriculture Organization, Rome.

Hutchings L., Beckley L., Griffiths M., Roberts M., Sundby S. and van der Lingen C. (2002) Spawning on the edge: Spawning grounds and nursery areas around the southern African coastline. Marine and Freshwater Research 53, 307-318.

Jeltsch F., Bonte D., Pe'er G., Reineking B., Leimgruber P., Balkenhol N., Schröder B., Buchmann C.M., Mueller T., Blaum N., Zurell D., Böhning-Gaese K., Wiegand T., Eccard J.A., Hofer H., Reeg J., Eggers U. and Bauer S. (2013) Integrating movement ecology with biodiversity research - exploring new avenues to address spatiotemporal biodiversity dynamics. Movement Ecology 1, 6.

Jones G.P., Almany G.R., Russ G.R., Sale P.F., Steneck R.S., van Oppen M.J.H. and Willis B.L. (2009) Larval retention and connectivity among populations of corals and reef fishes: History, advances and challenges. Coral Reefs 28, 307-325.

Jooste C.M., Oliver J., Emami-Khoyi A. and Teske P.R. (2018) Is the Wild Coast in eastern South Africa a distinct marine bioregion? Helgoland Marine Research 72, 6.

Jusino M.A., Banik M.T., Palmer J.M., Wray A.K., Xiao L., Pelton E., Barber J.R., Kawahara A.Y., Gratton C. and Peery M.Z. (2019) An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Molecular Ecology Resources 19, 176-190.

Katoh K., Rozewicki J. and Yamada K.D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20, 1160-1166.

Keesing J. and Irvine T. (2005) Coastal biodiversity in the Indian Ocean: The known, the unknown and the unknowable. Indian Journal of Marine Sciences 34, 11-26.

Kelly P., Clementson L., Davies C., Corney S. and Swadling K. (2016) Zooplankton responses to increasing sea surface temperatures in the southeastern Australia global marine hotspot. Estuarine, Coastal and Shelf Science 180, 242-257.

Kinlan B.P. and Gaines S.D. (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. Ecology 84, 2007-2020.

Knights A., Crowe T. and Burnell G. (2006) Mechanisms of larval transport: Vertical distribution of bivalve larvae varies with tidal conditions. Marine Ecology Progress Series 326, 167-174.

Laakmann S., Blanco-Bercial L. and Cornils A. (2020) The crossover from microscopy to genes in marine diversity: From species to assemblages in marine pelagic copepods. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190446.

Lamberth S., Drapeau L. and Branch G. (2009) The effects of altered freshwater inflows on catch rates of non estuarine dependent fish in multispecies nearshore fisheries. Estuarine Coastal and Shelf Science 84, 527-538.

Lanzén A., Lekang K., Jonassen I., Thompson E.M. and Troedsson C. (2017) DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. PLOS One 12, e0179443.

Leis J. (2007) Behaviour as input for modelling dispersal of fish larvae: Behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. Marine Ecology Progress Series 347, 185-193.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T. and Machida R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology 10, 34.

Lin Q., Xu L., Hou J., Liu Z., Jeppesen E. and Han B.P. (2017) Responses of trophic structure and zooplankton community to salinity and temperature in Tibetan lakes: Implication for the effect of climate warming. Water Research 124, 618-629.

Lohrer A.M., Thrush S.F., Hewitt J.E., Berkenbusch K., Ahrens M. and Cummings V.J. (2004) Terrestrially derived sediment: Response of marine macrobenthic communities to thin terrigenous deposits. Marine Ecology Progress Series 273, 121-138.

Lutjeharms J., Catzel R. and Valentine H. (1989) Eddies and other boundary phenomena of the Agulhas Current. Continental Shelf Research 9, 597-616.

Lutjeharms J., Cooper J. and Roberts M. (2000) Upwelling at the inshore edge of the Agulhas Current. Continental Shelf Research 20, 737-761.

Lutjeharms J.R.E. (2006) The Agulhas Current. Springer, Berlin.

Macher J.N., Hoorn B., Peijnenburg K., Walraven L. and Renema W. (2020) Metabarcoding reveals different zooplankton communities in northern and southern areas of the North Sea. bioRxiv 2020.07.23.218479.

MacKay C.F., Untiedt C.B. and Hein L. (2016) Local habitat drivers of macrobenthos in the northern, central and southern KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S105-S121.

Mann B., Celliers L., Fennessy S., Bailey S. and Wood A.D. (2006) Towards the declaration of a large marine protected area: A subtidal ichthyofaunal survey of the Pondoland coast in the Eastern Cape, South Africa. African Journal of Marine Science 28, 3-4.

McGillicuddy D.J. (2001) Small-scale patchiness, models of. In: Encyclopedia of Ocean Sciences (Second Edition). Steele J.H. (eds). pp. 474-87. Academic Press, Oxford.

McMurdie P.J. and Holmes S. (2013) Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLOS One 8, e61217.

McMurdie P.J. and Holmes S. (2014) Waste not, want not: Why rarefying microbiome data is inadmissible. PLOS Computational Biology 10, e1003531.

Meyer A.A., Lutjeharms J.R.E. and de Villiers S. (2002) The nutrient characteristics of the Natal Bight, South Africa. Journal of Marine Systems 35, 11-37.

Mitarai S., Siegel D.A., Watson J.R., Dong C. and McWilliams J.C. (2009) Quantifying connectivity in the coastal ocean with application to the Southern California Bight. Journal of Geophysical Research: Oceans 114, C10026.

Moore M. and Folt C. (1993) Zooplankton body size and community structure: Effects of thermal and toxicant stress. Trends in Ecology and Evolution 8, 178-183.

Neumann Leitão S., Melo Junior M.d., Porto Neto F.d.F., Silva A.P., Diaz X.F.G., Silva T.d.A.e., Nascimento Vieira D.A.d., Figueiredo L.G.P., Costa A.E.S.F.d., Santana J.R.d., Campelo R.P.d.S., Melo P.A.M.d.C., Pessoa V.T., Lira S.M.d.A. and Schwamborn R. (2019) Connectivity between coastal and oceanic zooplankton from Rio Grande do Norte in the Tropical Western Atlantic. Frontiers in Marine Science 6, 287.

Nichols R.V., Vollmers C., Newsom L.A., Wang Y., Heintzman P.D., Leighton M., Green R.E. and Shapiro B. (2018) Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources 18, 927-939.

Niebuhr B.B.S., Wosniack M.E., Santos M.C., Raposo E.P., Viswanathan G.M., da Luz M.G.E. and Pie M.R. (2015) Survival in patchy landscapes: The interplay between dispersal, habitat loss and fragmentation. Scientific Reports 5, 11898.

Noyon M., Morris T., Walker D. and Huggett J. (2018) Plankton distribution within a young cyclonic eddy off south-western Madagascar. Deep Sea Research Part II: Topical Studies in Oceanography 166, 141-150.

Ockhuis S., Huggett J., Gouws G. and Sparks C. (2017) The 'suitcase hypothesis': Can entrainment of meroplankton by eddies provide a pathway for gene flow between Madagascar and KwaZulu-Natal, South Africa?. African Journal of Marine Science 39, 435-451.

Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P., O'Hara R., Simpson G., Solymos P., Stevens H., Szöcs E. and Wagner H. (2018) Vegan: Community ecology package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. Version 2.5-1.

Olbers J., Celliers L. and Schleyer M. (2009) Zonation of benthic communities on the Subtropical Aliwal Shoal, Durban, KwaZulu-Natal, South Africa. African Zoology 44, 8-23.

Pattrick P. and Strydom N.A. (2014) Larval fish variability in response to oceanographic features in a nearshore nursery area. Journal of Fish Biology 85, 857-881.

Paturej E. and Gutkowska A. (2015) The effect of salinity levels on the structure of zooplankton communities. Archives of Biological Sciences 67, 483-492.

Pearce A.E. (1977) The shelf circulation off the east coast of South Africa. CSIR Research Report No. 361. Council for Scientific and Industrial Research, Stellenbosch.

Pillay D. (2002) The macrobenthos of the Little Lagoon, Durban Bay. MSc Thesis, University of KwaZulu-Natal, South Africa.

Pillay D., Branch G.M. and Forbes A.T. (2008) Habitat change in an estuarine embayment: anthropogenic influences and a regime shift in biotic interactions. Marine Ecology Progress Series 370, 19-31.

Piñol J., Mir G., Gomez-Polo P. and Agustí N. (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources 15, 819-830.

Pitz K.J., Guo J., Johnson S.B., Campbell T.L., Zhang H., Vrijenhoek R.C., Chavez F.P. and Geller J. (2020) Zooplankton biogeographic boundaries in the California Current System as determined from metabarcoding. PLOS One 15, e0235159.

Porri F., Jackson J.M., Von der Meden C.E.O., Weidberg N. and McQuaid C.D. (2014) The effect of mesoscale oceanographic features on the distribution of mussel larvae along the south coast of South Africa. Journal of Marine Systems 132, 162-173.

Pretorius M., Huggett J. and Gibbons M. (2016) Summer and winter differences in zooplankton biomass, distribution and size composition in the KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S155-S168.

Reed D.C., Raimondi P.T., Carr M.H. and Goldwasser L. (2000) The role of dispersal and disturbance in determining spatial heterogeneity in sedentary organisms. Ecology 81, 2011-2026.

Richardson A.J. (2008) In hot water: Zooplankton and climate change. ICES Journal of Marine Science 65, 279-295.

Ridgway T.M., Stewart B.A., Branch G.M. and Hodgson A.N. (1998) Morphological and genetic differentiation of Patella granularis (Gastropoda: Patellidae): Recognition of two sibling species along the coast of southern Africa. Journal of Zoology 245, 317-333.

Roberts C. (1997) Connectivity and management of Caribbean coral reefs. Science 278, 1454-1457.

Roberts M., van der Lingen C., Whittle C. and Berg M. (2010) Shelf currents, lee-trapped and transient eddies on the inshore boundary of the Agulhas Current, South Africa: Their relevance to the KwaZulu-Natal sardine run. African Journal of Marine Science 32, 423-447.

Roberts M.J. and Nieuwenhuys C. (2016) Observations and mechanisms of upwelling in the northern KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S43-S63.

Roberts M.J., Nieuwenhuys C. and Guastella L.A. (2016) Circulation of shelf waters in the KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S7-S21.

Rodgher S., Lombardi A.T. and Melão M.d.G.G. (2009) Evaluation onto life cycle parameters of Ceriodaphnia silvestrii submitted to 36 days dietary copper exposure. Ecotoxicology and Environmental Safety 72, 1748-1753.

Rossi S., Coppari M. and Viladrich N. (2017) Benthic-pelagic coupling: New perspectives in the animal forests. In: Marine Animal Forests: The Ecology of Benthic Biodiversity Hotspots. Rossi S., Bramanti L., Gori A. and Orejas C. (eds). pp. 855-85. Springer International Publishing, Cham.

Rouault M., Penven P. and Pohl B. (2009) Warming in the Agulhas Current system since the 1980's. Geophysical Research Letters 36, L12602.

Roughgarden J., Gaines S. and Possingham H. (1988) Recruitment dynamics in complex life cycles. Science 241, 1460-1466.

Schenk J., Geisen S., Kleinboelting N. and Traunspurger W. (2019) Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth. Metabarcoding and Metagenomics 3, e46704.

Schleyer M. (1985) Chaetognaths as indicators of water masses in the Agulhas Current system. Investigational Report No. 61. Oceanographic Research Institute, Durban.

Schleyer M. (2000) South African coral communities. In: Coral reefs of the Indian ocean: Their ecology and conservation. McClanahan T.R., Sheppard C.R.C and Obura D.O. (eds). pp. 83-98. Oxford University Press, New York.

Schleyer M., Heikoop J. and Risk M. (2006) A benthic survey of Aliwal Shoal and assessment of the effects of a wood pulp effluent on the reef. Marine Pollution Bulletin 52, 503-514.

Schumann E.H. (1988) Physical oceanography off Natal. In: Coastal Ocean Studies off Natal, South Africa. Schumann E.H. (eds). pp. 101-30. Springer, New York.

Shanks A. (2009) Pelagic larval duration and dispersal distance revisited. The Biological Bulletin 216, 373 - 385.

Shanks A.L. and Brink L. (2005) Upwelling, downwelling, and cross-shelf transport of bivalve larvae: Test of a hypothesis. Marine Ecology Progress Series 302, 1-12.

Sheaves M. (2009) Consequences of ecological connectivity: the coastal ecosystem mosaic. Marine Ecology Progress Series 391, 107-115.

Singh S.P., Groeneveld J.C., Hart-Davis M.G., Backeberg B.C. and Willows-Munro S. (2018) Seascape genetics of the spiny lobster *Panulirus homarus* in the Western Indian Ocean: Understanding how oceanographic features shape the genetic structure of species with high larval dispersal potential. Ecology and Evolution 8, 12221-12237.

Singh S., Groeneveld J., Huggett J., Naidoo D., Cedras R. and Willows-Munro S. (2021) DNA metabarcoding of marine zooplankton in South Africa: How good is the reference library? African Journal of Marine Science (In press).

Singh S.P., Groeneveld J.C. and Willows-Munro S. (2019) Between the current and the coast: Genetic connectivity in the spiny lobster *Panulirus homarus rubellus*, despite potential barriers to gene flow. Marine Biology 166, 36.

Singh S.P., Groeneveld J.C. and Willows-Munro S. (2020) Genetic structure and life history are key factors in species distribution models of spiny lobsters. Ecology and Evolution 10, 14394-14410.

Smeti H., Pagano M., Menkes C., Lebourges-Dhaussy A., Hunt B.P.V., Allain V., Rodier M., de Boissieu F., Kestenare E. and Sammari C. (2015) Spatial and temporal variability of zooplankton off New Caledonia (Southwestern Pacific) from acoustics and net measurements. Journal of Geophysical Research: Oceans 120, 2676-2700.

Spalding M.D., Fox H.E., Allen G.R., Davidson N., Ferdaña Z.A., Finlayson M., Halpern B.S., Jorge M.A., Lombana A., Lourie S.A., Martin K.D., McManus E., Molnar J., Recchia C.A. and Robertson J. (2007) Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. BioScience 57, 573-583.

Taberlet P., Coissac E., Pompanon F., Brochmann C. and Willerslev E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21, 2045-2050.

Taylor P.D., Fahrig L., Henein K. and Merriam G. (1993) Connectivity is a vital element of landscape structure. Oikos 68, 571-573.

Teske P.R., Barker N.P. and McQuaid C.D. (2007a) Lack of genetic differentiation among four sympatric southeast African intertidal limpets (Siphonariidae): Phenotypic plasticity in a single species? Journal of Molluscan Studies 73, 223-228.

Teske P.R., Froneman P.W., Barker N.P. and McQuaid C.D. (2007b) Phylogeographic structure of the caridean shrimp Palaemon peringueyi in South Africa: Further evidence for intraspecific genetic units associated with marine biogeographic provinces. African Journal of Marine Science 29, 253-258.

Teske P.R., McQuaid C.D., Froneman P.W. and Barker N.P. (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. Marine Ecology Progress Series 314, 283-293.

Teske P.R., Papadopoulos I., Newman B.K., Dworschak P.C., McQuaid C.D. and Barker N.P. (2008) Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn. BMC Evolutionary Biology 8, 341.

Teske P.R., Papadopoulos I., Zardi G.I., McQuaid C.D., Edkins M.T., Griffiths C.L. and Barker N.P. (2007c) Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: Planktonic, abbreviated and direct development. Marine Biology 152, 697-711.

Teske P.R., von der Heyden S., McQuaid C.D. and Barker N.P. (2011) A review of marine phylogeography in southern Africa. South African Journal of Science 107, 43-53.

Teske P.R., Winker H., McQuaid C.D. and Barker N.P. (2009) A tropical/subtropical biogeographic disjunction in southeastern Africa separates two evolutionarily significant units of an estuarine prawn. Marine Biology 156, 1265-1275.

Thibault-Botha D. and Gibbons M.J. (2005) Epipelagic siphonophores off the east coast of South Africa. African Journal of Marine Science 27, 129-139.

Thibault-Botha D., Lutjeharms J.R. and Gibbons M. (2004) Siphonophore assemblages along the east coast of South Africa: Mesoscale distribution and temporal variations. Journal of Plankton Research 26, 1115-1128.

Thrush S.F. and Dayton P.K. (2002) Disturbance to marine benthic habitats by trawling and dredging: Implications for marine biodiversity. Annual Review of Ecology and Systematics 33, 449-473.

Treml E.A., Halpin P.N., Urban D.L. and Pratson L.F. (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. Landscape Ecology 23, 19-36.

Untiedt C. (2013) Community structure and function of macrobenthos in three feature areas of the Natal Bight, South Africa. MSc Thesis, University of KwaZulu-Natal, South Africa.

Untiedt C.B. and MacKay C.F. (2016) Distribution and feeding modes of macrobenthos within three oceanographic feature areas of the KwaZulu-Natal Bight, South Africa. African Journal of Marine Science 38, S91-S104.

Verkuil Y.I., Nicolaus M., Ubels R., Dietz M.M., Samplonius J.M., Galema A., Kiekebos K., de Knijff P. and Both C. (2020) DNA metabarcoding successfully quantifies relative abundances of arthropod taxa in songbird diets: A validation study using camera-recorded diets. bioRxiv, 2020.11.26.399535.

von der Heyden S., Prochazka K. and Bowie R.C.K. (2008) Significant population structure and asymmetric gene flow patterns amidst expanding populations of *Clinus cottoides* (Perciformes, Clinidae): Application of molecular data to marine conservation planning in South Africa. Molecular Ecology 17, 4812-4826.

Watson J.R., Siegel D.A., Kendall B.E., Mitarai S., Rassweiller A. and Gaines S.D. (2011) Identifying critical regions in small-world marine metapopulations. Proceedings of the National Academy of Sciences 108, E907-E913.

Wickham H. (2009) ggplot2: Elegant graphics for data analysis. Springer, New York.

Wilkin J.L. and Jeffs A.G. (2011) Energetics of swimming to shore in the puerulus stage of a spiny lobster: Can a postlarval lobster afford the cost of crossing the continental shelf?. Limnology and Oceanography: Fluids and Environments 1, 163-175.

Yamanaka H. and Minamoto T. (2016) The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. Ecological Indicators 62, 147-153.

Yu M., Hu G., Feeley K.J., Wu J. and Ding P. (2012) Richness and composition of plants and birds on land-bridge islands: effects of island attributes and differential responses of species groups. Journal of Biogeography 39, 1124-1133.

Zardi G.I., McQuaid C.D., Teske P.R. and Barker N.P. (2007) Unexpected genetic structure of mussel populations in South Africa: Indigenous *Perna perna* and invasive *Mytilus galloprovincialis*. Marine Ecology Progress Series 337, 135-144.

Zhang G.K., Chain F.J.J., Abbott C.L. and Cristescu M.E. (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11, 1901-1914.

4.8 Supplementary data

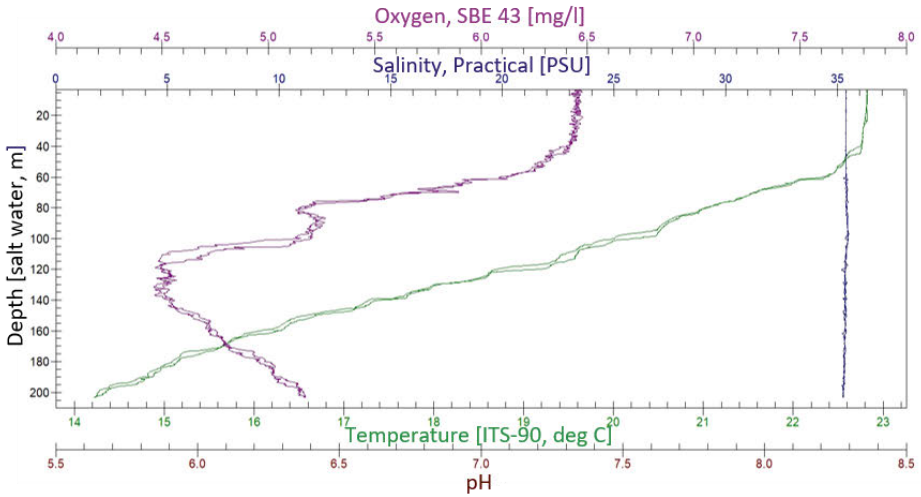
Table S4.1 The six primer cocktails used in this DNA metabarcoding study (first round PCR): each of the COI primer cocktails amplify different fragments of the COI-5P gene region. Illumina adapter target sequences (indicated in bold and underlined) were used in accordance with the workflow from the Illumina 16S Metagenomics protocol (Illumina,2013). These adapter targets allow Nextera indexing and Illumina adapter addition through PCR.

Fragment	Primer Name	Sequence (5' - 3')	Direction	Target Taxa	Reference	Fragment Size
COI_Leray	mlCOIintF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW ACW GGW TGA ACW GTW TAY CCY CC	F	Various phyla	Leray et al., 2013	313 bp - 319 bp
	HCO2198	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** TAA ACT TCA GGG TGA CCA AAA AAT CA	R	Various phyla	Folmer et al., 1994	
COI_FISH	mlCOIintF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW ACW GGW TGA ACW GTW TAY CCY CC	F	Various phyla	Leray et al., 2013	313 bp - 319 bp
	HCO2198	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** TAA ACT TCA GGG TGA CCA AAA AAT CA	R	Various phyla	Folmer et al., 1994	
	FishR2	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** ACT TCA GGG TGA CCG AAG AAT CAG AA	R	Fish	Ward et al., 2005	
COI_LOBSTER	LobsterMinibarF	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGW GAT GAY CAA ATT TAY AAT G T	F	Lobster	Govender et al., 2019	230 bp
	LobsterMinibarR	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCW ACT CCT CTT TCT ACT ATT CC	R	Lobster	Govender et al., 2019	
COI_PRAWN	PrawnMiniBar1F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GCY GAA YTA GGT CAA CCA GG	F	Prawn	This study	277 bp (F1)
	PrawnMiniBar2F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** GGA TTT GGA AAY TGA YTA GTT CC	F	Prawn	This study	316 bp (F2)

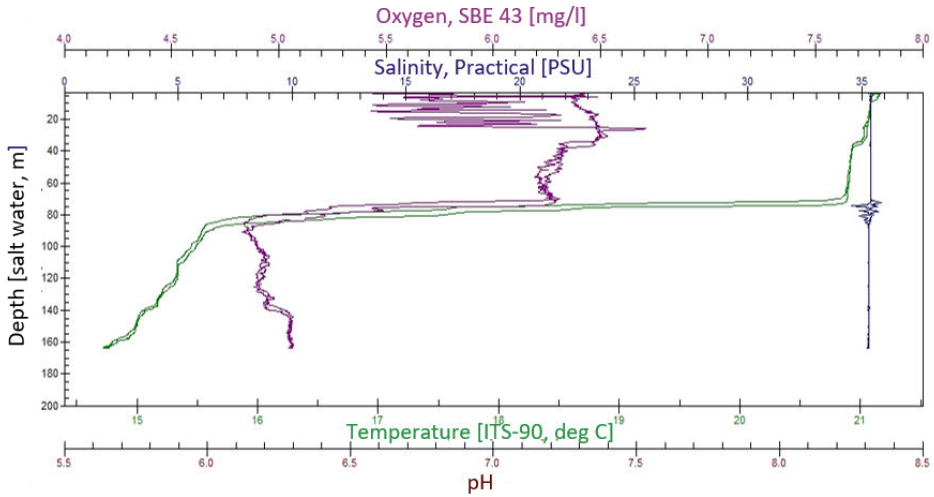
Table S4.1 (continued).

COI_PRAWN	PrawnMiniBar1R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGA GGR TAW ACA GTT CAT CC	R	Prawn	This study	
	PrawnMiniBar2R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT ACY CCT CTT TCT ACT ATW CC	R	Prawn	This study	
	PrawnMiniBar3R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATW CGG TCT ATT GTT ATY CC	R	Prawn	This study	
COI_SHRIMP	ShrimpMiniBar6F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** CCW ATT ATA ATT GGA GGR TTY GG	F	Shrimp	This study	310 bp
	ShrimpMiniBar6R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GCT CCT ARA ATA GAA GAA ACY CC	R	Shrimp	This study	
	ShrimpMiniBar9R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT CTT CTT CGT ATR TTR ATA AC	R	Shrimp	This study	
	ShrimpMiniBar10R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT ARG ATW GAA GAR ACT CC	R	Shrimp	This study	
	ShrimpMiniBar13R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** CCT AAY ATT GAA GAA ACW CCT GC	R	Shrimp	This study	
COI_CRAB	CrabMiniBar1F	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNNNN** CCW ATT ATA ATT GGA GGA TTY GG	F	Crab	This study	331 bp
	CrabMiniBar5R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT ATW GTT ATA CC	R	Crab	This study	
	CrabMiniBar8R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT AWA GWT ATA CC	R	Crab	This study	
	CrabMiniBar11R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT AGG TCT ATT YTT ATA CC	R	Crab	This study	
	CrabMiniBar12R	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G NNNNN** GGT ATT TGR TCT ATK GTT ATA CC	R	Crab	This study	

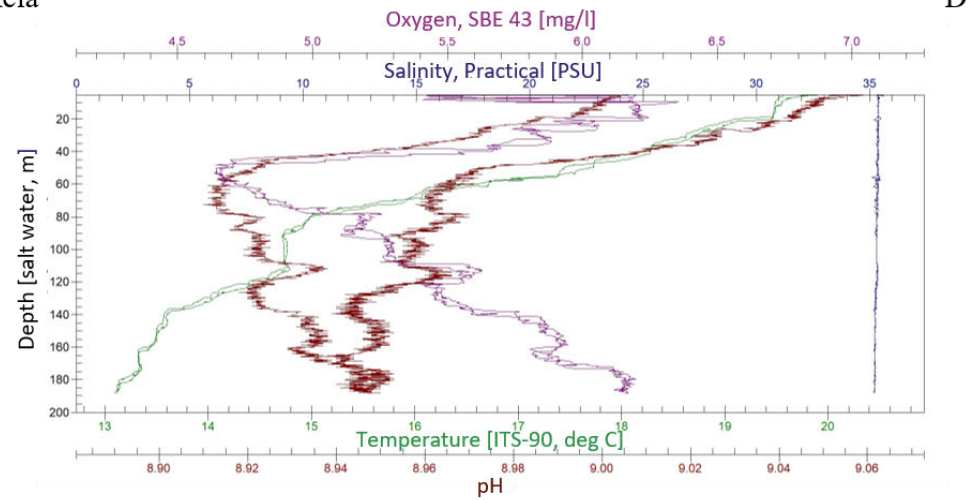
1



uThukela



Durban



Aliwal

Figure S4.1 Depth profiles measured using the on-board Seabird SBE 19 plus V2 SeaCAT CTD sensor at the deepest station of each site (uThukela, Durban and Aliwal).

Table S4.2 High-throughput summary statistics across the replicates collected at uThukela, Durban and Aliwal.

Communities	Total Read Count	Total Merged reads	Number of paired reads	ASVs
		uThukela 1		
uThukela 1 20R1	39726	3945	1973	30
uThukela 1 20R2	41016	3212	1606	27
uThukela 1 20R3	40712	3349	1675	25
uThukela 1 50R1	14300	176	88	9
uThukela 1 50R2	50978	2355	1178	28
uThukela 1 50R3	38852	817	409	13
uThukela 1 100R1	86206	7641	3821	36
uThukela 1 100R2	3004	39	20	5
uThukela 1 100R3	37046	2747	1374	22
uThukela 1 200R1	63210	3094	1547	39
uThukela 1 200R2	36352	3754	1877	23
uThukela 1 200R3	24630	1695	848	17
Total	**476032**	**32824**	**16412**	**274**
Mean	**39669**	**2735**	**1368**	**23**
SD	**21497**	**2046**	**1023**	**10**
		Durban		
Durban 1 20R1	113554	2692	1346	49
Durban 1 20R2	204514	2729	1365	48
Durban 1 20R3	221502	1350	675	29
Durban 1 50R1	94724	1578	789	30
Durban 1 50R2	24078	334	167	9
Durban 1 100R1	108740	5447	2724	60
Durban 1 100R2	211608	7394	3697	84
Durban 1 100R3	188422	6279	3140	56
Durban 1 200R2	71660	3906	1953	45
Durban 1 200R3	115832	6035	3018	49
Total	**1354634**	**37744**	**18872**	**459**
Mean	**135463**	**3774**	**1887**	**46**
SD	**67074**	**2406**	**1203**	**20**
		Aliwal		
Aliwal 1 20R1	14414	1573	787	57
Aliwal 1 20R2	10482	956	478	38
Aliwal 1 20R3	11702	1240	620	52
Aliwal 1 50R1	12868	1331	666	76
Aliwal 1 50R2	8574	1041	521	43
Aliwal 1 50R3	12228	1239	620	51
Aliwal 1 100R1	6102	651	326	43
Aliwal 1 100R2	6428	591	296	36
Aliwal 1 100R3	5642	765	383	41
Aliwal 1 200R1	6348	731	366	43
Aliwal 1 200R2	7808	917	459	54
Aliwal 1 200R3	9820	1858	929	59
Total	**112416**	**12893**	**6447**	**593**
Mean	**9368**	**1074**	**537**	**49**
SD	**2987**	**387**	**194**	**11**

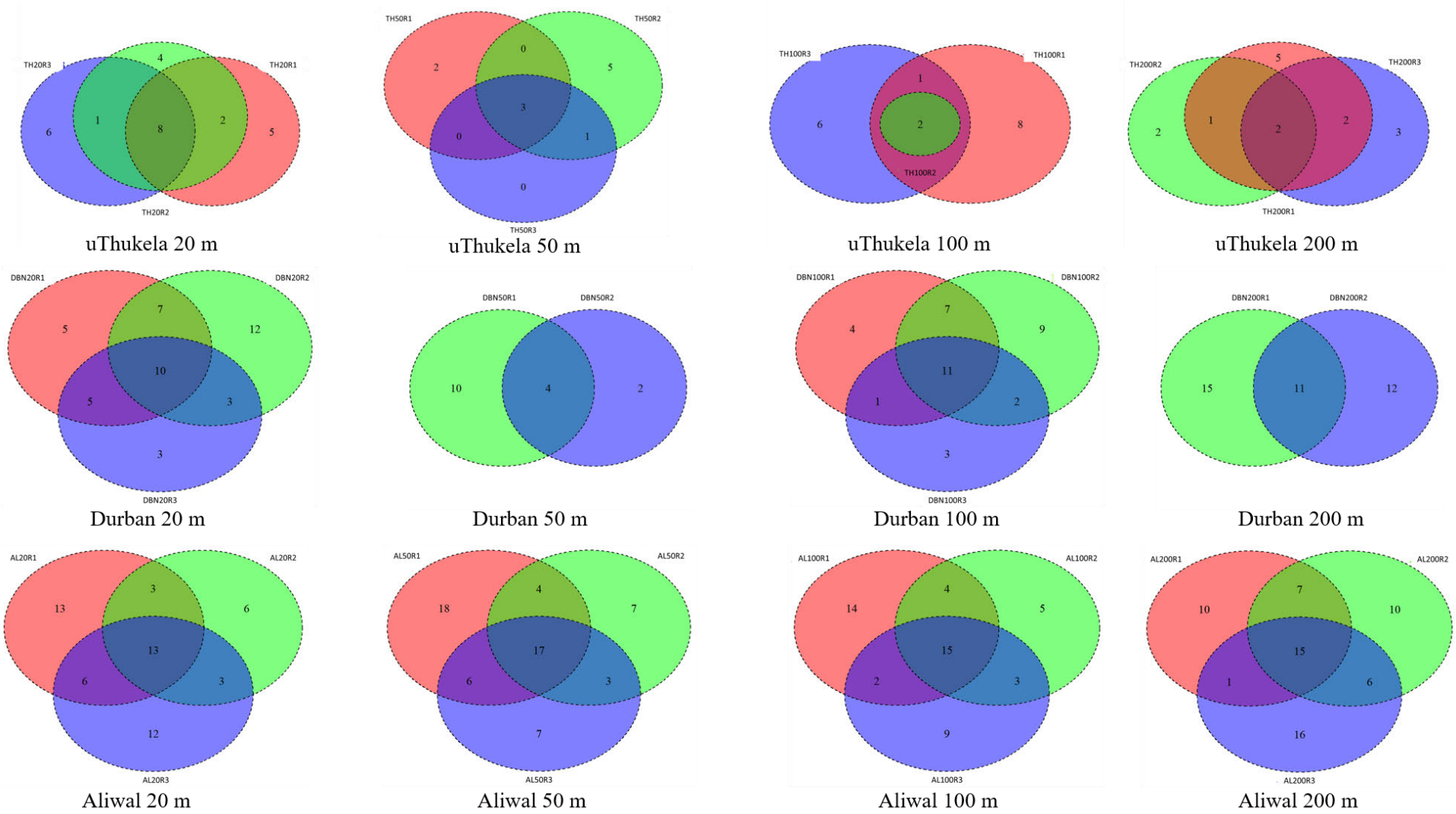


Figure S4.2 Venn Diagrams comparing the difference between the biological replicates collected at the different sampling stations across uThukela, Durban and Aliwal.

Table S4.3 (a) Jaccard dissimilarity and (b) Bray-Curtis dissimilarity matrices for ring net replicates across the different sampling sites.

a. Jaccard dissimilarity

uThukela			
uThukela 20m			
	TH20R1	TH20R2	TH20R3
TH20R1	-	0.5	0.636
TH20R2	0.5	-	0.571
TH20R3	0.636	0.571	-

Durban			
Durban 20m			
	DBN20R1	DBN20R2	DBN20R3
DBN20R1	-	0.595	0.545
DBN20R2	0.595	-	0.675
DBN20R3	0.545	0.675	-

Aliwal			
Aliwal 20m			
	AL20R1	AL20R2	AL20R3
AL20R1	-	0.636	0.62
AL20R2	0.636	-	0.628
AL20R3	0.62	0.628	-

uThukela 50m			
	TH50R1	TH50R2	TH50R3
TH50R1	-	0.727	0.5
TH50R2	0.727	-	0.556
TH50R3	0.5	0.556	-

Durban 50m		
	DBN50R1	DBN50R2
DBN50R1	-	0.75
DBN50R2	0.75	-

Aliwal 50m			
	AL50R1	AL50R2	AL50R3
AL50R1	-	0.618	0.582
AL50R2	0.618	-	0.545
AL50R3	0.582	0.545	-

uThukela 100m			
	TH100R1	TH100R2	TH100R3
TH100R1	-	0.818	0.824
TH100R2	0.818	-	0.778
TH100R3	0.824	0.778	-

Durban 100m			
	DBN100R1	DBN100R2	DBN100R3
DBN100R1	-	0.471	0.571
DBN100R2	0.471	-	0.606
DBN100R3	0.571	0.606	-

Aliwal 100m			
	AL100R1	AL100R2	AL100R3
AL100R1	-	0.558	0.638
AL100R2	0.558	-	0.526
AL100R3	0.638	0.526	-

uThukela 200m			
	TH200R1	TH200R2	TH200R3
TH200R1	-	0.75	0.692
TH200R2	0.75	-	0.8
TH200R3	0.692	0.8	-

Durban 200m		
	DBN200R1	DBN200R2
DBN200R1	-	0.711
DBN200R2	0.711	-

Aliwal 200m			
	AL200R1	AL200R2	AL200R3
AL200R1	-	0.551	0.709
AL200R2	0.551	-	0.618
AL200R3	0.709	0.618	-

b. Bray-Curtis dissimilarity

uThukela			
uThukela 20m technical replicates			
	TH20R1	TH20R2	TH20R3
TH20R1	-	0.112	0.0975
TH20R2	0.112	-	0.075
TH20R3	0.0975	0.075	-

Durban			
Durban 20m technical replicates			
	DBN20R1	DBN20R2	DBN20R3
DBN20R1	-	0.353	0.366
DBN20R2	0.353	-	0.287
DBN20R3	0.366	0.287	-

Aliwal			
Aliwal 20m technical replicates			
	AL20R1	AL20R2	AL20R3
AL20R1	-	0.465	0.378
AL20R2	0.465	-	0.262
AL20R3	0.378	0.262	-

uThukela 50m technical replicates			
	TH50R1	TH50R2	TH50R3
TH50R1	-	0.197	0.163
TH50R2	0.197	-	0.0415
TH50R3	0.163	0.0415	-

Durban 50m technical replicates		
	DBN50R1	DBN50R2
DBN50R1	-	0.22
DBN50R2	0.22	-

Aliwal 50m technical replicates			
	AL50R1	AL50R2	AL50R3
AL50R1	-	0.428	0.593
AL50R2	0.428	-	0.388
AL50R3	0.593	0.388	-

uThukela 100m technical replicates			
	TH100R1	TH100R2	TH100R3
TH100R1	-	0.139	0.0939
TH100R2	0.139	-	0.123
TH100R3	0.0939	0.123	-

Durban 100m technical replicates			
	DBN100R1	DBN100R2	DBN100R3
DBN100R1	-	0.32	0.364
DBN100R2	0.32	-	0.382
DBN100R3	0.364	0.382	-

Aliwal 100m technical replicates			
	AL100R1	AL100R2	AL100R3
AL100R1	-	0.245	0.411
AL100R2	0.245	-	0.37
AL100R3	0.411	0.37	-

uThukela 200m technical replicates			
	TH200R1	TH200R2	TH200R3
TH200R1	-	0.554	0.761
TH200R2	0.554	-	0.224
TH200R3	0.761	0.224	-

Durban 200m technical replicates		
	DBN200R1	DBN200R2
DBN200R1	-	0.123
DBN200R2	0.123	-

Aliwal 200m technical replicates			
	AL200R1	AL200R2	AL200R3
AL200R1	-	0.353	0.639
AL200R2	0.353	-	0.575
AL200R3	0.639	0.575	-

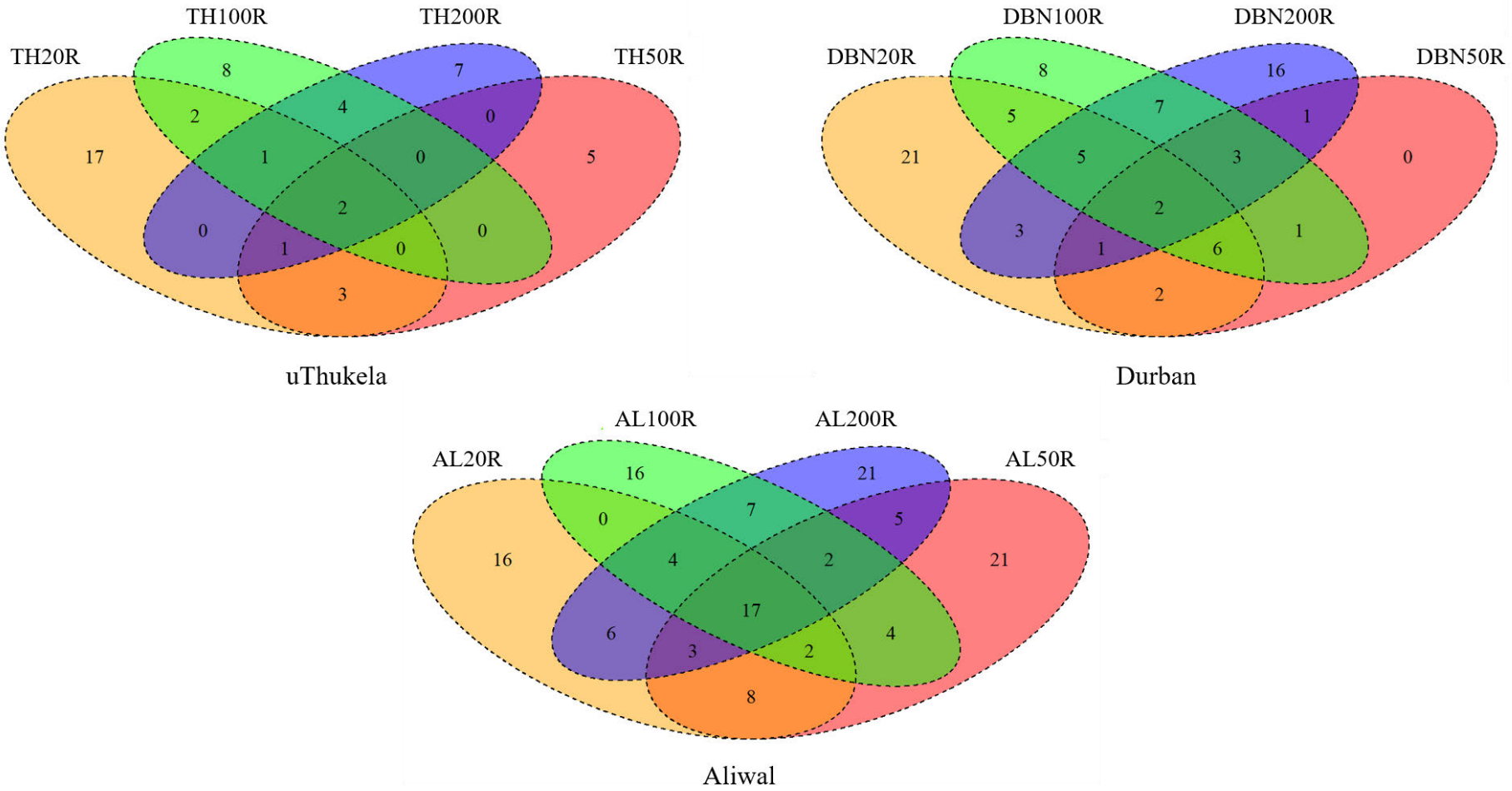


Figure S4.3 Venn Diagrams comparing the differences between shared and unique clusters across the different sampling stations for uThukela, Durban and Aliwal.

Table S4.4 Shared species between uThukela, Durban and Aliwal.

17 shared species clusters between uThukela, Durban and Aliwal						
Phylum	**Class**	**Order**	**Family**	**Genus**	**Species**	**% similarity**
Chordata	Actinopterygii	Scombriformes	Scombridae	Scomber	colias	100
Arthropoda	Malacostraca	Decapoda	Luciferidae	Lucifer	intermedius	100
Arthropoda	Malacostraca	Euphausiacea	Euphausiidae	Euphausia	recurva	100
Arthropoda	Copepoda	Calanoida	Eucalanidae	Subeucalanus	subcrassus	100
Arthropoda	Malacostraca	Decapoda	Polybiidae	Liocarcinus	corrugatus	98.53
Arthropoda	Copepoda	Calanoida	Calanidae	Calanus	sinicus	100
Arthropoda	Malacostraca	Decapoda	Portunidae	Thalamita	gatavakensis	99.63
Arthropoda	Copepoda	Calanoida	Eucalanidae	Subeucalanus	mucronatus	97
Arthropoda	Malacostraca	Decapoda	Xanthidae	Medaeops	neglectus	100
Chordata	Actinopterygii	Perciformes	Serranidae	Serranus	knysnaensis	100
Arthropoda	Malacostraca	Stomatopoda	Nannosquillidae	UK_Nannosquillidae	sp.	88.89
Mollusca	Gastropoda	Pteropoda	Creseidae	Creseis	acicula	98.51
Chordata	Actinopterygii	Perciformes	Haemulidae	Pomadasys	olivaceus	100
Arthropoda	Malacostraca	Stomatopoda	Squillidae	Pterygosquilla	sp.	86.03
Chordata	Actinopterygii	Carangiformes	Carangidae	Trachurus	delagoa	100
Arthropoda	Malacostraca	Decapoda	Palinuridae	Panulirus	homarus	99.63
Arthropoda	Malacostraca	Decapoda	Portunidae	Lupocyclus	sp.	100
9 shared species clusters between uThukela and Durban						
Phylum	**Class**	**Order**	**Family**	**Genus**	**Species**	**% similarity**
Arthropoda	Malacostraca	Euphausiacea	Euphausiidae	Euphausia	brevis	98.53
Arthropoda	Copepoda	Calanoida	Pontellidae	Pontellina	plumata	95.79
Chordata	Actinopterygii	Scombriformes	Scombridae	Scomber	japonicus	100
Arthropoda	Malacostraca	Decapoda	Portunidae	Thalamita	gloriensis	100
Chordata	Actinopterygii	Perciformes	Haemulidae	Pomadasys	striatus	100
Arthropoda	Malacostraca	Decapoda	Crangonidae	UK_Crangonidae	sp.	93.01
Arthropoda	Malacostraca	Decapoda	Luciferidae	Lucifer	typus	99.48
Arthropoda	Malacostraca	Decapoda	Ovalipidae	Ovalipes	trimaculatus	100
Arthropoda	Malacostraca	Amphipoda	Lestrigonidae	UK_Lestrigonidae	sp.	89.05
5 shared species clusters between uThukela and Aliwal						
Phylum	**Class**	**Order**	**Family**	**Genus**	**Species**	**% similarity**
Arthropoda	Malacostraca	Euphausiacea	Euphausiidae	Stylocheiron	carinatum	100

Table S4.4 (continued).

Arthropoda	Malacostraca	Decapoda	UK_Decapod	UK_Decapod	sp.	85.42
Arthropoda	Copepoda	Calanoida	Candaciidae	Candacia	curta	100
Cnidaria	Hydrozoa	Siphonophorae	Agalmatidae	Nanomia	bijuga	98
Arthropoda	Malacostraca	Decapoda	Alpheidae	Synalpheus	sp.	96.32
29 shared species clusters between Durban and Aliwal						
Phylum	**Class**	**Order**	**Family**	**Genus**	**Species**	**% similarity**
Arthropoda	Malacostraca	Decapoda	Diogenidae	Diogenes	aff. pallacens	100
Arthropoda	Malacostraca	Decapoda	Callianassidae	Callichirus	gilchristi	98.43
Arthropoda	Copepoda	Calanoida	Temoridae	Temora	turbinata	99.67
Arthropoda	Copepoda	Calanoida	Temoridae	Temora	discaudata	100
Chordata	Actinopterygii	Clupeiformes	Clupeidae	Etrumeus	whiteheadi	100
Arthropoda	Malacostraca	Decapoda	UK_Decapod	UK_Decapod	sp.	86.84
Arthropoda	Malacostraca	Decapoda	Portunidae	Xaiva	mcleayi	99.48
Arthropoda	Malacostraca	Decapoda	Euryplacidae	UK_Euryplacidae	sp.	88.24
Arthropoda	Malacostraca	Decapoda	Pilumnidae	Trachysalambria	curvirostris	86.76
Chordata	Thaliacea	Doliolida	Doliolidae	UK_Doliolida	sp.	82
Arthropoda	Malacostraca	Decapoda	Pilumnidae	UK_Pilumnidae	sp.	89.58
Arthropoda	Malacostraca	Decapoda	Diogenidae	Diogenes	costatus	100
Arthropoda	Copepoda	Calanoida	Eucalanidae	Subeucalanus	pileatus	100
Chordata	Actinopterygii	Clupeiformes	Clupeidae	Etrumeus	sadina	100
Arthropoda	Malacostraca	Decapoda	Cancridae	UK_Cancridae	sp.	91.27
Arthropoda	Malacostraca	Decapoda	UK_Decapod	UK_Decapod	sp.	86.76
Chordata	Actinopterygii	Blenniiformes	Blenniidae	Parablennius	pilicornis	100
Arthropoda	Malacostraca	Stomatopoda	Squillidae	UK_Squillidae	sp.	88.54
Arthropoda	Malacostraca	Decapoda	UK_Decapod	UK_Decapod	sp.	86.03
Cnidaria	Hydrozoa	Siphonophorae	Diphyidae	Muggiaea	atlantica	97.69
Arthropoda	Malacostraca	Decapoda	Xanthidae	Etisus	sp.	97.43
Arthropoda	Malacostraca	Decapoda	Galatheidae	Galathea	nexa	97.4
Arthropoda	Malacostraca	Decapoda	Penaeidae	Marsupenaeus	japonicus	100
Arthropoda	Copepoda	Calanoida	Eucalanidae	Rhincalanus	sp.	90.67
Arthropoda	Malacostraca	Decapoda	Alpheidae	Alpheus	dolerus	100
Arthropoda	Malacostraca	Decapoda	Palaemonidae	Gnathophyllum	americanum	99.26
Arthropoda	Malacostraca	Decapoda	Porcellanidae	UK_Porcellanidae	sp.	90.67
Arthropoda	Malacostraca	Decapoda	Palaemonidae	Palaemonella	pottsi	95.31
Chordata	Actinopterygii	Blenniiformes	Blenniidae	Scartella	emarginata	100

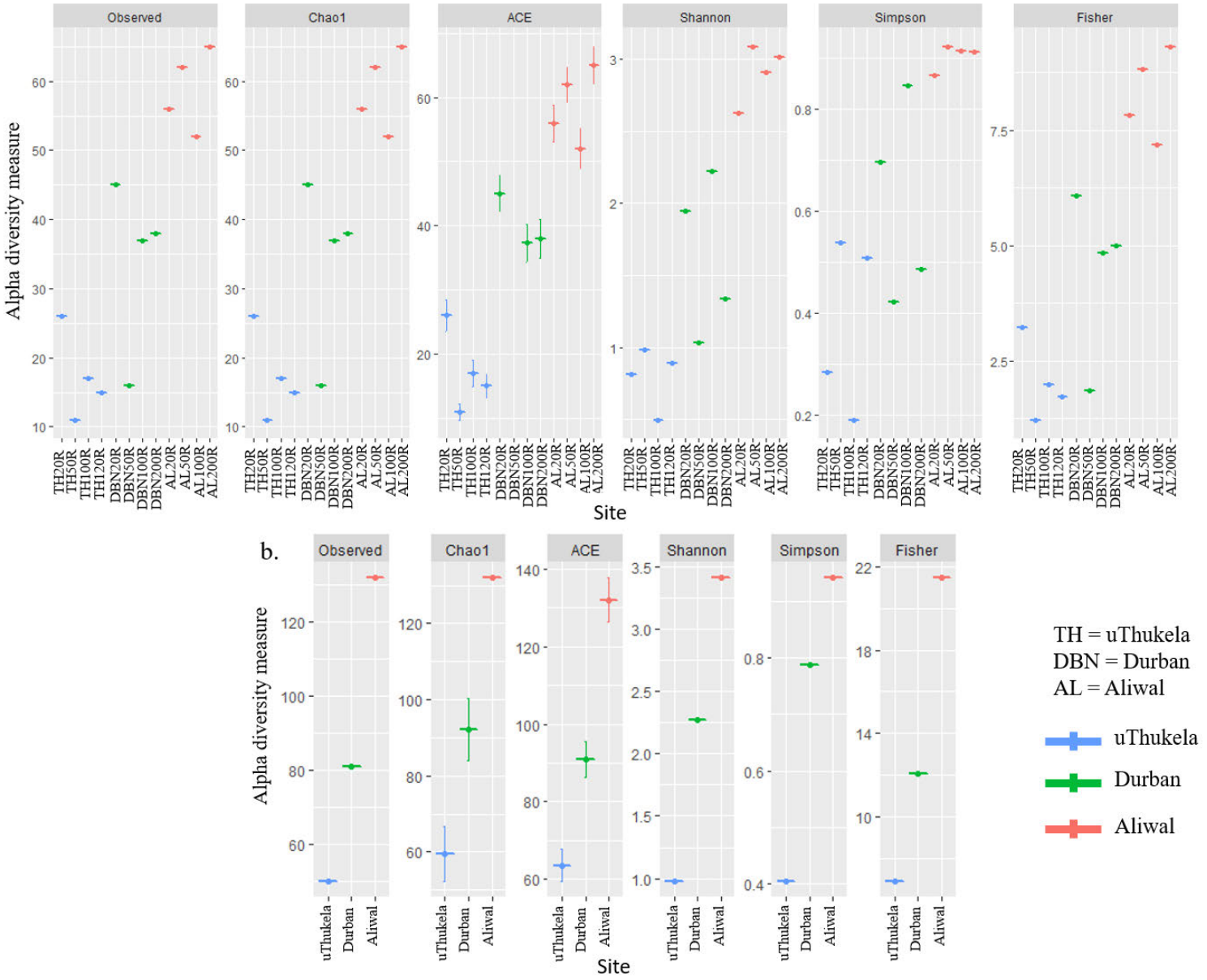


Figure S4.4 Boxplot of observed species richness, Shannon's diversity index, Simpson's diversity index and Fishers alpha across the (a) sampling stations and (b) sampling sites.

Chapter Five: General discussion and future possibilities

5.1 General discussion

This Ph.D study presents a novel and validated DNA metabarcoding approach for ecological studies and long-term biomonitoring of marine zooplankton. This methodology was tested in a highly dynamic ocean region off eastern South Africa. The approach integrates (1) the development of taxon-specific mini-barcodes to increase amplification rates; (2) experimental validation of detection rates for abundant and rare taxa; (3) state-of-the-art DNA metabarcoding techniques using high-throughput sequencing technologies; and (4) community-level analysis to compare marine zooplankton samples. Marine zooplankton are regarded as important indicators for monitoring ecosystem health and biodiversity of oceans because their species composition and relative abundance are sensitive to environmental change, such as the effects of climate change, overharvesting of natural resources, habitat degradation and pollution (Bucklin *et al.* 2016). In this 21st century period of climate change and increased anthropogenic disturbances of natural environments, this Ph.D contributes towards a growing body of literature presenting sophisticated tools designed to measure changes in the diversity of marine pelagic ecosystems.

The use of DNA metabarcoding as a biomonitoring tool of marine ecosystems is a recent advance, with the method able to provide accurate and high-resolution community composition data rapidly (Lacoursière-Roussel *et al.* 2018; Closek *et al.* 2019; Pearman *et al.* 2020). Nevertheless, many DNA metabarcoding pipelines are purpose-built for specific ecosystems or taxonomic groups and might not be directly transferrable to other systems. A validated DNA metabarcoding approach for zooplankton in the coastal region off eastern South Africa focusing on specific taxonomic groups was deemed an essential step towards future ecological studies and long-term biomonitoring of changes in this pelagic ecosystem.

A technical hurdle that had to be overcome during the Ph.D was that high-throughput sequencing technology platforms such as the Illumina MiSeq used for DNA metabarcoding have limited read lengths that are less than the 658 base pairs (bp) of the standard mitochondrial cytochrome oxidase I (COI) barcode region (Marquina *et al.* 2019). In addition, the DNA of zooplankton are often damaged and degraded; therefore, the use of shorter DNA fragments (mini-barcodes; 200 – 300 bp) was used to improve amplification success. Meusnier *et al.* (2008) designed universal mini-barcode primers, but they are not equally effective across all taxa due to

primer mismatch (Arif *et al.* 2011). Therefore, in Chapter 2 of this Ph.D study, a method to develop and test taxon-specific mini-barcode primer sets was established to improve PCR amplification and sequencing success rates (Dong *et al.* 2014). Primer sets were designed *in-silico* to amplify the shortest and most informative portion of the COI gene region for marine lobsters (Chapter 2), prawns, shrimps and crabs (Chapter 3). DNA barcode gap analysis was performed to ensure that there was statistically no overlap between inter-and intra-specific genetic distances. The utility of the newly designed primers was tested on both adult and larval samples and outperformed previously published COI primers.

Method validation is a crucial aspect of applying DNA metabarcoding to ecological studies or long-term biomonitoring and is an important first step in identifying potential methodological biases and shortcomings (Zhang *et al.* 2018). In Chapter 3, an experimental approach was used to test the efficiency of the designed taxon-specific primers against published universal primers using artificially assembled zooplankton communities with known species composition and relative abundance. The results indicated that taxon-specific primers increased detection rates of target taxa and confidently identified rare species. Using primer cocktails (multiple primer sets) increased detection rates and reduced preferential amplification compared to universal primers. This finding supports Zhang *et al.* (2018), who found that the use of primer cocktails for COI DNA metabarcoding reduced both false-positive and false-negative results. Similarly, this chapter found a significant reduction in false-positive results.

To demonstrate proof of concept, in Chapter 4, the newly developed DNA metabarcoding approach was used to test the hypothesis that ocean currents over the shelf of eastern South Africa between the Agulhas Current and the coast will have a strong mixing effect and act to homogenize zooplankton communities. Ocean currents play an essential role in marine pelagic ecosystems by redistributing water, heat, oxygen, biological propagules (such as zooplankton) and enhancing primary productivity critical for marine life (Hays 2017). Plankton tow nets were used to sample zooplankton along cross-shelf transects for three sampling sites (uThukela, Durban and Aliwal). Sampling sites differed in shelf width, seafloor substrate and benthic habitat structures. The new methodology was used to generate species composition information. This information was then used in community-level analyses to infer connectivity between geographically separated zooplankton communities. An alternative hypothesis of a patchy zooplankton distribution with

some connectivity was inferred. It was concluded that a strong benthic-pelagic coupling effect has given rise to patchiness rather than homogenization of zooplankton communities by water movements. The benthic-pelagic coupling effect could be related to benthic habitat structure, seafloor substrates and biota. Connectivity network analysis further detected distinct clustering of zooplankton communities by site, attributed to sub-mesoscale processes formed by the Agulhas Current, especially where the shelf broadened to form the KZN Bight. Chapter 4 demonstrated the utility of the DNA metabarcoding approach for ecological studies, setting the stage for long-term biomonitoring of trends in marine pelagic ecosystems over the continental shelf of eastern South Africa.

5.2 Future possibilities

Biomonitoring is an integral component of marine ecosystem management; it provides important data to evaluate changes in marine ecosystems related to climate change and anthropogenic disturbances. Biomonitoring programs that can link biological changes of zooplankton to physio-chemical changes of the ocean will help identify ecological trends and predict future trajectories over the long-term. Bean *et al.* (2017) characterized the current equipment and technology used to monitor oceans into two categories: (1) a platform from which measurements are taken, i.e., research vessels, static observatories, or automated vehicles; and (2) the actual sensor or methodology used to take measurements, i.e., multibeam sonar arrays, underwater cameras, or analyses of physical samples. By overcoming methodological and technical hurdles and improving the DNA metabarcoding approach on zooplankton along the east coast of South Africa, this Ph.D study creates many new research opportunities for future monitoring of coastal marine biodiversity in various highly vulnerable ecosystems.

An ongoing hurdle in ecological studies is the inability to collect data with high precision and accuracy at a reasonable rate to detect and manage critical global change processes (Bonada *et al.* 2006). DNA metabarcoding has been shown to be useful in many ecological studies (Baird and Hajibabaei 2012; Bush *et al.* 2019), as such, the methods and protocols developed in this Ph.D study are shown to perform large-scale biodiversity screening and allow for detailed ecological investigations into the species richness, diversity, dispersal and connectivity of planktonic animals in a biodiversity-rich marine system. Combining DNA metabarcoding data from marine organisms along with physical oceanographic data has the potential to detect the impacts of climate change,

the monitoring and assessment of ecosystem health, characterization of food webs and detection of introduced and non-indigenous species (Bucklin *et al.* 2016; Goodwin *et al.* 2017; Deagle *et al.* 2018). The protocol and methods used in this Ph.D study can easily be applied to other DNA metabarcoding studies to enable comparative studies among different ecosystems. In addition, the techniques developed for decapods (lobsters, crabs, prawns, and shrimp) and fish can easily be transferred to other groups of marine organisms worldwide.

The development of new molecular tools as a complement to morphology-based assessments for species identification of zooplankton will not only allow for rapid detection of changes in zooplankton community structure and connectivity but will also accelerate the description of undescribed species. Currently, the number of species occurring in marine pelagic ecosystems remains unknown. The combination of DNA barcoding and morphology-based species identification is required to assist in expanding DNA barcode reference libraries, to allow for the successful identification of species and increase the scope and reach of DNA metabarcoding applications in the future. DNA metabarcoding can accelerate the assessments of biodiversity and contribute towards a better description of marine zooplankton's biogeographic distributions globally. Such information is essential to a better understanding of pelagic ecosystems' functioning in the face of climate change and anthropogenic disturbances.

5.3 References

Arif I., Khan H., Sadoon M. and Shobrak M. (2011) Limited efficiency of universal mini-barcode primers for DNA amplification from desert reptiles, birds and mammals. Genetics and Molecular Research 10, 3559-3594.

Baird D.J. and Hajibabaei M. (2012) Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21, 2039-2044.

Bean T.P., Greenwood N., Beckett R., Biermann L., Bignell J.P., Brant J.L., Copp G.H., Devlin M.J., Dye S., Feist S.W., Fernand L., Foden D., Hyder K., Jenkins C.M., van der Kooij J., Kröger S., Kupschus S., Leech C., Leonard K.S., Lynam C.P., Lyons B.P., Maes T., Nicolaus E.E.M., Malcolm S.J., McIlwaine P., Merchant N.D., Paltriguera L., Pearce D.J., Pitois S.G., Stebbing P.D., Townhill B., Ware S., Williams O. and Righton D. (2017) A review of the tools used for marine monitoring in the UK: Combining historic and

contemporary methods with modeling and socioeconomics to fulfill legislative needs and scientific ambitions. Frontiers in Marine Science 4, 263.

Bonada N., Prat N., Resh V.H. and Statzner B. (2006) Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. Annual Review of Entomology 51, 495-523.

Bucklin A., Lindeque P.K., Rodriguez-Ezpeleta N., Albaina A. and Lehtiniemi M. (2016) Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. Journal of Plankton Research 38, 393-400.

Bush A., Compson Z.G., Monk W.A., Porter T.M., Steeves R., Emilson E., Gagne N., Hajibabaei M., Roy M. and Baird D.J. (2019) Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. Frontiers in Ecology and Evolution 7, 434.

Closek C.J., Santora J.A., Starks H.A., Schroeder I.D., Andruszkiewicz E.A., Sakuma K.M., Bograd S.J., Hazen E.L., Field J.C. and Boehm A.B. (2019) Marine vertebrate biodiversity and distribution within the Central California Current using environmental DNA (eDNA) metabarcoding and ecosystem surveys. Frontiers in Marine Science 6, 732.

Deagle B.E., Clarke L.J., Kitchener J.A., Polanowski A.M. and Davidson A.T. (2018) Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. Molecular Ecology Resources 18, 391-406.

Dong W., Liu H., Xu C., Zuo Y., Chen Z. and Zhou S. (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. BMC Genetics 15, 138.

Goodwin K.D., Thompson L.R., Duarte B., Kahlke T., Thompson A.R., Marques J.C. and Caçador I. (2017) DNA sequencing as a tool to monitor marine ecological status. Frontiers in Marine Science 4, 107.

Hays G.C. (2017) Ocean currents and marine life. Current Biology 27, R470-R473.

Lacoursière-Roussel A., Howland K., Normandeau E., Grey E.K., Archambault P., Deiner K., Lodge D.M., Hernandez C., Leduc N. and Bernatchez L. (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. Ecology and Evolution 8, 7763-7777.

Marquina D., Andersson A.F. and Ronquist F. (2019) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19, 90-104.

Meusnier I., Singer G.A., Landry J.F., Hickey D.A., Hebert P.D. and Hajibabaei M. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214.

Pearman J.K., von Ammon U., Laroche O., Zaiko A., Wood S.A., Zubia M., Planes S. and Pochon X. (2020) Metabarcoding as a tool to enhance marine surveillance of nonindigenous species in tropical harbors: A case study in Tahiti. Environmental DNA.

Zhang G.K., Chain F.J.J., Abbott C.L. and Cristescu M.E. (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11, 1901-1914.