Molecular Species Delimitation and Biogeography of Canadian Marine Planktonic Crustaceans

by

Robert George Young

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Integrative Biology

Guelph, Ontario, Canada

ABSTRACT

MOLECULAR SPECIES DELIMITATION AND BIOGEOGRAPHY OF CANADIAN MARINE PLANKTONIC CRUSTACEANS

Robert George Young
University of Guelph, 2016

Advisors:
Dr. Sarah Adamowicz
Dr. Cathryn Abbott

Zooplankton are a major component of the marine environment in both diversity and biomass and are a crucial source of nutrients for organisms at higher trophic levels. Unfortunately, marine zooplankton biodiversity is not well known because of difficult morphological identifications and lack of taxonomic experts for many groups. In addition, the large taxonomic diversity present in plankton and low sampling coverage pose challenges in obtaining a better understanding of true zooplankton diversity. Molecular identification tools, like DNA barcoding, have been successfully used to identify marine planktonic specimens to a species. However, the behaviour of methods for specimen identification and species delimitation remain untested for taxonomically diverse and widely-distributed marine zooplanktonic groups. Using Canadian marine planktonic crustacean collections, I generated a multi-gene data set including COI-5P and 18S-V4 molecular markers of morphologically-identified Copepoda and Thecostraca (Multicrustacea: Hexanauplia) species. I used this data set to assess generalities in the genetic divergence patterns and to determine if a barcode gap exists separating interspecific and intraspecific molecular divergences, which can reliably delimit specimens into species. I then used this information to evaluate the North Pacific, Arctic, and North Atlantic biogeography of marine Calanoida (Hexanauplia: Copepoda) plankton. My work provides essential information on the molecular data and analysis tools necessary to conduct rapid biodiversity surveys on marine plankton, facilitating the exploration of unknown faunas using DNA barcoding and providing baseline data against which future comparisons can be made.

Acknowledgements

First and foremost, I would like to thank my two co-advisors Dr. Sarah Adamowicz and Dr. Cathryn Abbott for their support through my program. The guidance from both of you gave me the confidence to pursue my own scientific interests. Cathryn, I very much appreciate your thoughtful advice in assisting me with my study designs and molecular work. Sally, I would not have been able to complete this work without your help. I very much appreciate all of our conversations which helped me work through my project. I would like to also say a special thank you to you Sally, for all of your assistance in preparing my thesis. Your help in editing and revising my message to its current state was instrumental in getting me to the finishing line. I would also like thank Dr. Thomas Therriault for his helpful discussions giving me insight into my study designs and results. To Dr. Elizabeth Boulding, I would like to say thank you for your thoughtful comments on my thesis and help in navigating the university process.

I would like to also thank all of my friends that I have had during my time here at the University of Guelph. In particular, I would like to say thank you to Gillian Martin, Anne Chambers, and Fatima Mitterboeck for their great conversations and humour. Thank you to Alex Chen, Tooba Khan, Jessmyn Niergarth, Rebeca Orue Herrera, Michelle Pyle, and Jenna Snelgrove for their assistance with specimen sorting and photography. I would also like to thank Jinal Pandya for her excellent work in helping me with sorting and molecular work. And thank you to the rest of the Adamowicz/Hebert lab members over the last four years.

During my program, I was very fortunate to have been able to participate in a collection trip to the Arctic region. My time spent in both Deception Bay and Steensby Inlet was educational and rewarding. My inclusion on this trip was generously offered by Kim Howland, and I would like to thank Kim, the rest of the collections team, and my advisors and committee for supporting my decision to go. I would also like to acknowledge expert identifiers Carol Cooper, Mary Greenlaw, Rebecca Milne, and Lou Van Guelpen. I would also like to recognize the Advanced Analysis Centre at the University of Guelph for their sequencing support. Thank you to Niels Van Steenkiste, Magalie Castelin, Geoff Lowe, and Sujeevan Ratnasingham for helpful discussions. I would like to thank the staff of the Canadian Centre for DNA Barcoding at

the Biodiversity Institute of Ontario, University of Guelph, for their help with molecular troubleshooting and analysis.

I would like to acknowledge my funding agencies. This research was made possible through funding by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Second Canadian Aquatic Invasive Species Network (CAISN II) grant to Hugh MacIsaac and colleagues and by an NSERC Discovery Grant to Sarah Adamowicz. The vouchered barcode collection was generated with partial support from Canada's Genomics Research and Development Initiative (Quarantine and Invasive Species Project), including DNA sequencing by the Applied Genomics facility of the National Research Council of Canada. I am also grateful to the CAISN collection teams for obtaining the specimens used in this study.

Last, but certainly not least, I would like to most importantly thank my family. The help that my parents, parents-in-law, and siblings provided were invaluable. But the patience and support of my loving wife has been the most important on my journey. Without Kimberley I would not have been able to complete this process. Kimberley was always supportive through family losses, house renovations, and pregnancy, even while I was 4,400 km across the country. And Kimberley was still patient while I spent hours and hours writing while carrying baby number two. Because of these and so, so, so, many other things thank you very much Kim, George, and baby two, I love you all.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
General introduction to Canadian marine zooplankton biodiversity

Introduction

Plankton are a major component of the marine environment in both diversity and biomass and are a crucial source of nutrients for organisms at higher trophic levels (Bron *et al.* 2011; Blanco-Bercial *et al.* 2014). Zooplankton play an important role as consumer of the primary producers and prey for animals at higher trophic levels. However, the true zooplanktonic species diversity within global marine ecosystems is not well understood (Archambault *et al.* 2010; Bucklin *et al.* 2011). Difficulties associated with studying marine plankton include their generally fragile nature, small body size with subtle morphological differences, and lack of expert taxonomists (McManus & Katz 2009; Packer *et al.* 2009). These factors, in combination with low sample coverage due to a massive ocean habitat, make obtaining a better understanding of true zooplankton diversity difficult. Characterizing the diversity present in Canadian marine ecosystems will provide baseline data against which future comparisons can be made. Obtaining this information is important due to the alteration of marine habitats through human influences such as habitat destruction, environmental pollution, and climate change (Singh 2002; Archambault *et al.* 2010; Radulovici *et al.* 2010). The consequences of these human-mediated changes are increased rates of non-native species introductions and the increased speed at which local and global extinctions are occurring (Singh 2002; Gamfeldt *et al.* 2015).

Our lack of knowledge of plankton diversity is especially concerning in the North Pacific, Arctic, and North Atlantic waters of Canada, as these areas are relatively undisturbed by human influence compared to many other global locations, yet disturbance is increasing; this is particularly true of the Arctic region (Vermeij & Roopnarine 2008; Wassmann *et al.* 2011; Renaud *et al.* 2015). Thus, as they are one of the largest taxonomic groups in marine waters, it is necessary to characterize Canadian crustacean zooplankton. Accomplishing this through traditional morphological identifications is not feasible due to the expense and time-consuming nature of the work as well as the lack of available expertise for many groups (Bucklin *et al.* 2011).

With significant challenges to traditional morphological methods, the use of molecular techniques must be further investigated including DNA barcoding and metabarcoding approaches (Bucklin 2010a; Cristescu 2014). DNA barcoding is a standardized approach, using molecular data from one or a few specific gene regions, to identify a specimen as belonging to a particular species (Hebert *et al.* 2003a, b). Metabarcoding utilizes the DNA barcoding approach, but instead of applying the method to a single specimen at a time, it uses DNA barcoding across large collections of specimens at the same time (Cristescu 2014). A metabarcoding approach uses the high-throughput capabilities of next-generation sequencing platforms in an attempt to assess all diversity for a large taxonomic group within a mixed-species sample. The performance of DNA barcoding and metabarcoding among widely-distributed geographically and taxonomically diverse zooplankton, such as crustaceans, has been largely untested.

In addition to identifying a specimen to a particular species, DNA barcoding is also utilized to delimit likely species groups present within a particular specimen data set (Hubert & Hanner 2015). To delimit species, a data set is partitioned into molecular operational taxonomic units (MOTU) through a variety of approaches including similarity-based approaches. A similarity-based approach partitions a data set into MOTUs assuming a cut-off value or threshold for maximum within-species variation. This type of single-marker species delimitation has drawn substantial research attention regarding methods development (Hubert & Hanner 2015) as well as criticism for applying a single threshold across broad taxonomic groups (Meyer & Paulay 2005; Will *et al.* 2005; Krishnamurthy & Francis 2012; Collins & Cruickshank 2013). Nevertheless, once a suitable threshold is chosen through reproducible and justifiable methods, the threshold-based approach to delimiting species using DNA barcodes has been relatively successful in terms of recovering groups consistent with species boundaries defined through independent approaches (Blanco-Bercial *et al.* 2014; Will *et al.* 2005; Ebach & Holdrege 2005; Radulovici *et al.* 2010; Huemer *et al.* 2014). Unfortunately, it is all too common in the literature that thresholds used for species delimitation are not appropriately chosen and instead are selected because they were once used in past literature (Collins & Cruickshank 2013).

Although there has been work developing methodologies to justify the selection of a molecular threshold for larger taxonomic groups (Lefébure *et al.* 2006; Blanco-Bercial *et al.* 2014), these methods often rely solely on external comparisons, often taxonomic species identifications (Shen *et al.* 2013). It is not uncommon for these taxonomic identifications to vary

wildly depending on the veracity of the identifier (Shen *et al.* 2013). Furthermore, taxonomic revisions and morphologically cryptic species can also influence the accuracy and biological relevance of traditional taxonomic identifications. There remains a gap in the current barcoding literature with regards to applying an *internal* method to evaluate the ability of a molecular region to effectively partition data into clusters largely congruent with known taxonomic groups (Handl *et al.* 2005) or with likely evolutionarily distinct species.

In this PhD thesis, I address this research gap by developing and applying methodology for internal (sequence-based) threshold validation as well as for testing for congruence among character sets. I apply these methods to a novel data set of Canadian marine planktonic crustaceans. Existing studies evaluating the effectiveness of barcode data for placing specimens into species groups is often based on success/failure of how well molecular data replicates taxonomic groupings for specific taxa. Here, I use a formal, quantitative approach to concordance testing, which has not yet been applied in barcoding studies, to assess the congruence present across whole data sets as opposed to identifying the success/failure on a group-by-group basis. The strength of this approach is in seeking generalities in the genetic divergence patterns in a large dataset with the aim of then selecting analysis methods that maximize global concordance between genetic clusters and species. Using Canadian marine planktonic crustacean species, I have also generated a novel multi-gene data set including two molecular markers and morphology-based identifications to evaluate the biogeography across North Pacific, Arctic, and North Atlantic waters. This information contributes to our understanding of the dispersal history of the marine biota as well as species boundaries and speciation mechanisms, such as allopatric divergence.

The balance of this introductory chapter reviews key background material to provide context to the overall themes addressed in the following three research chapters. This background section includes reviews of current Canadian marine biodiversity, historical Canadian glaciations and their potential impact on Canadian biodiversity and biogeography, and common species concepts important for this work. I then introduce planktonic crustaceans as target taxa for this work. Next, I provide a brief introduction to DNA barcoding as concept and to the methodology used in this study, after which I briefly discuss the future for barcoding and metabarcoding applications. I conclude this chapter by outlining my specific research objectives for this thesis.

The known versus expected diversity of Canadian marine biota

To enumerate the true diversity on the planet, we must first define biodiversity and understand how we might assess it. Biodiversity can be discussed and studied in different ways, including species diversity, genetic diversity, and ecosystem diversity (Spangenberg 2007). Most commonly, biodiversity is assessed using species as the units of measure (de Queiroz 2007). So, how many species are present on the planet? For hundreds of years scientists have endeavoured to answer this question and better understand, identify, and describe biological diversity (May 1988). Since formal Linnaean descriptions were adopted in the mid 1700's, there have been approximately 1.7 million species described, in contrast to the estimated 3 to 100 million species expected to exist on earth (May 2010; Mora *et al.* 2011). Furthermore, the total number of marine eukaryote species currently described is approximately 230,000 (Mora *et al.* 2011; Appeltans *et al.* 2012). Some estimates have suggested that the total diversity in the marine realm is as high as 10 million species (Grassle & Maciolek 1992).

If the Canadian marine regions follow the same trend suggested by the global estimates, there is a significant shortfall in the number of described, as compared to the expected, species present in Canadian waters. Archambault *et al.* (2010) compiled a list of Canadian multi- and unicellular eukaryote marine species and found nearly 16,000 currently described species. Estimates suggest that there are close to 60,000 undescribed Canadian marine eukaryote species (Archambault *et al.* 2010); however, our lack of knowledge of the Canadian Arctic region makes it difficult to know if this diversity estimate is close to accurate (Archambault *et al.* 2010). These numbers appear to confirm our general lack of understanding of Canadian marine diversity and a need for increased effort toward Canadian aquatic biodiversity research (Meyer & Paulay 2005; Archambault *et al.* 2010).

Traditional techniques for biodiversity research, including labour-intensive collection methods followed by time-consuming and difficult morphological identifications, provide real challenges to obtaining a true understanding of biodiversity and are not a practical solution for large-scale surveys. A further difficulty is that the number of trained taxonomists falls far short of the population that would be necessary to complete the monumental task of a complete understanding of global biodiversity (Giangrande 2003). This insufficient taxonomic talent, referred to as the taxonomic impediment, has been a topic of intensive scientific discussion (e.g.

Hoagland 1996; Bucklin *et al.* 2010b; Hubert & Hanner 2015). With limited resources supporting taxonomic training, the number of available 'well-trained' taxonomists has decreased, and as a consequence biological researchers are obtaining identifications from the best available sources and personnel (e.g. para-taxonomists rather than "true" taxonomists) (Godfray 2007; Bortolus 2008; Hubert & Hanner 2015). The concerns become that para-taxonomists using the best-available references, such as keys and species lists, are often not correctly identifying specimens. These errors can lead to the improper impression of well-trained taxonomist identifications in the literature (de Carvalho *et al.* 2007; Bortolus 2008).

Limits in expertise and range, taxonomically and geographically, for para-taxonomy is a further concern in understanding global biodiversity. Often, a para-taxonomist's expertise is limited to their geographic region of focus, thereby limiting their experience with intraspecific morphological variation across species distributions. Additionally, with a limited geographic focus, environmental/habitat-invoked variation may not be considered. A further challenge when assessing marine zooplankton biodiversity is that phylogenetic diversity can be quite broad, where often more than ten phyla can be found at a single site (Bucklin *et al.* 2010b). The variation in morphological characters across life stages, poorly-resolved taxonomy in some groups, and cryptic or undescribed species are just a few of the many other challenges faced by identifiers (Knowlton 1993; McManus & Katz 2009; Packer *et al.* 2009). The taxonomic impediment is a continued struggle for studying the diversity and distributions of many taxonomic groups and is especially problematic for species identifications in marine environments (Giangrande 2003; Bucklin *et al.* 2010b).

Canadian glacial cycles and biological landscape

In order to understand current biodiversity and biogeography in Canada's three oceans, historical glacial patterns must be considered. The Bering Land Bridge, a physical barrier that closed off the Bering Strait due to land mass movements, was formed near the end of Cretaceous Period 80 mya, resulting in the isolation of the Pacific and Arctic Oceans (Dunton 1992; Gladenkov *et al.* 2002). Intermittent opening and closing of this barrier started between 5.5 and 5.4 mya, likely due to a combination of tectonic events and changes in sea level, allowing the exchange of Arctic and Pacific Ocean waters (Gladenkov *et al.* 2002; Vermeij & Roopnarine 2008). During periods of open water flow, species moved into the Arctic region assisted by

northward-flowing waters (Dayton et al. 1994; Adey & Steneck 2001; Gladenkov et al. 2002). Species migrations from the Pacific into the Arctic Ocean have been supported, with fossil evidence indicating a prominent trans-Arctic movement of species approximately 3.5 MYA (Vermeij 1991). The timeframe of this major interchange was associated with an Arctic warming trend with reduced ice cover in the mid-Pliocene (approximately 3.5 mya), resulting in increased Arctic plankton productivity (Vermeij & Roopnarine 2008) and enabling colonization of the Arctic by macro-invertebrates (Vermeij 1991). After multiple cycles of glaciation during the Pleistocene, the most recent glacial retreat, approximately 18,000 to 11,000 years ago, resulted in the Arctic marine area becoming freed from glaciation, leaving unpopulated regions and enabling colonization from the Pacific through the now-open Bering Strait (Adey & Steneck 2001; Gladenkov *et al.* 2002; Carr *et al*. 2011). At the end of this period 11,000 years ago, oceans and landmasses resembled what we see today.

The Arctic environment and its present-day biota are relatively young due to the Pliocene and Pleistocene glaciations previously having limited the extent of Arctic marine habitats (Dunton 1992; Adey & Steneck 2001). At the last glacial maximum approximately 18,000 years ago, the northern glacial ice sheet extended well into what is now the Great Lakes region, to approximately 42° N latitude, and included the Arctic shores of Northern Canada as well as the majority of the Canadian Pacific and Atlantic Ocean coastlines (Wares & Cunningham 2001; Wares 2002; Hewitt 2000, 2004). This period of limited northern marine habitat occurred from approximately 5 million to 18,000 years ago (Dunton 1992; Adey & Steneck 2001).

During glaciation, Arctic marine environments experienced a freezing of a large portion of the sea, reducing sea water levels, altering ocean water salinity, and impacting the evolutionary histories of marine organisms (Dunton 1992). The continued survival of species through the length of a glacial period was dependent on their overall fitness and the availability of liquid water (Vincent *et al.* 2005). Ice-free areas present through these glacial periods are called glacial refugia (Vincent *et al.* 2005). These areas resulted in the survival of refugial populations in both marine and fresh water environments; subsequent recolonization from these regions is known to have contributed strongly to the current North American aquatic biodiversity and biogeography (Vincent *et al.* 2005). Cold-adapted species which survived in these refugial areas have resulted in the Arctic endemic species we see today (Vincent *et al.* 2005). Once the ice receded, refugial species were free to populate the newly-exposed habitat (Hewitt 2004).

Although access was available to populate the once ice-covered region, cold-adapted species may still have been contained by unique Arctic geographical barriers, including two restrictive straights, Bering and Fram, as well as the European basin and Lomonosov and Alpha-Mendeleyev Ridges (Jennings *et al.* 2010).

Understanding common species concepts applied in marine biogeography and biodiversity research

A fundamental understanding of the species concept—how organisms, based on both biological theory and observed biological attributes, are partitioned into groups signifying species units—is necessary to effectively conduct biodiversity research. The difficulty in the task of establishing a reliable definition of species for biodiversity studies is in the differing goals between taxonomy and biodiversity research. The goal of selecting a particular species concept for biodiversity studies is typically for the enumeration of interacting species across a wide range of taxonomic groups differing in morphological characters and other biological traits. This is in contrast to a taxonomist's goal of delimitation, where similar organisms, regardless of geographic location, are assessed for their possible inclusion in a known species or the formation of a new species group (Agapow *et al.* 2004). Regardless of the species concept utilized, the general notion that a species is a group of organisms sharing a similar form and function, and occupying the same territory, can be universally applied (de Queiroz 2005, 2007).

The works of de Queiroz (1998, 2005, 2007) have focused on the complexities of defining species. These works have asserted that all modern species concepts are variants of a common element, and therefore a unified species concept exists which is robust in both conceptual and methodological arguments (de Queiroz 2007). Applying different species concepts can yield different numbers of taxonomic units, with major conservation implications, such as for understanding the number of endangered species (Agapow *et al.* 2004). As well, to enable comparisons across studies, providing an explicit definition of species is necessary for each individual study to understand the implications of study methods, data, and the scope of the author's conclusions. At least 22 distinct notions on how to define a species have been proposed (Mayden 1997), and these can be grouped into several distinct fundamental concepts that are commonly used.

The Biological Species Concept (BSC) is most commonly applied among studies of easily identifiable, terrestrial complex organisms (Garnett & Christidis 2007). The BSC uses the sexual success, as measured by production of fertile offspring, between two organisms of a distinct population to define a species (Mayr 1942, 1996). With the BSC there are several significant gaps when applied to nature: the exclusion of asexual organisms (Mayr 1996), the inability for the BSC to evaluate extinct organisms (Agapow *et al.* 2004), and the complications associated with the hybridization of two distinct species, as previously defined by the BSC, resulting in fertile offspring (de Queiroz 1998).

The Morphological Species Concept has direct origins from the Linnaean system of classification and was the primary method of classification up to the middle of the 20th century (Hillis 1987; de Queiroz 1998; Agapow *et al.* 2004). Also referred to as the phenetic species concept, it establishes taxonomic divisions based on a group of organisms sharing similar character states (de Queiroz 1998). Similar to the BSC, the exclusive use of the Morphological Species Concepts presents problems with taxonomic classification and identification. There are difficulties in establishing a reliable delimitation among species having few distinct traits or species with traits which are difficult to differentiate (Hebert *et al.* 2003a; Agapow *et al.* 2004). In addition, species possessing two or more separate morphological character states, such as a larval life stage, will not have the same set of morphological criteria throughout individuals' lifespans, further complicating morphological identification (Tang *et al.* 2010). Furthermore, fundamental characters that differ between related species may not be seen through phenotypic evaluations but evident in chemical differences or other variation with no morphological differentiation (Hillis 1987; Bickford *et al.* 2007).

The Ecological Species Concept delineates species based on the survival of organisms within a particular niche, emphasizing the natural selection pressures of the environment (de Queiroz 1998). Cavender-Bares *et al.* (2009) note that variation in habitat influences regional species composition. They also suggest that a species' ecologically relevant traits, as determined through phylogenetic analysis, are as responsible for preserving regional ecology as other taxonomically relevant traits. As an example, the high influence of ecological adaptations is observed in zooplankton, where salinity gradients, depth, and water currents influence regional species composition (Archambault *et al.* 2010; Laakmann *et al.* 2012). The practicality of the Ecological Species Concept as a sole method of species delimitation is questionable, given that

determining a species based on an ecological species concept must take into account a nearly limitless set of parameters and conditions present in a complex biogeographic region. Furthermore, these parameters can change with outside influences such as the introduction of non-indigenous species.

The Phylogenetic Species Concept delineates species using cladistics, where all specimens that are considered a species are monophyletic in relation to each other (Hillis 1987; Nixon & Wheeler 1990). A monophyletic group is a clade, on a phylogenetic tree, that consists of an ancestor and all of its descendants (Freeman & Herron 2007). These monophyletic clades are representative of species that make up a distinct group or population of organisms that have shared a similar "evolutionary fate" (de Queiroz 1998). Although the phylogenetic species concept only needs a single shared derived character to distinguish between two species—whether morphological, chemical, or molecular—molecular data provide a greater number of informative characters, yielding a more accurate representation of relationships, and is the predominant choice in current phylogenetics (Garnett & Christidis 2007). One of the strengths of the phylogenetic species concept is the ability to compare species across disparate phylogenetic groups and studies, allowing evolutionary and biogeographic hypotheses to be addressed (Cracraft 1987). Furthermore, when using the phylogenetic species concept one can also investigate relationships below the species level in biodiversity and biogeography studies (Agapow *et al.* 2004). A further advantage with the Phylogenetic Species Concept is the ability to use secondary analysis techniques in combination with the phylogenetic analysis. One such technique is the use of molecular clocks to place evolutionary time on the ancestral nodes of phylogenetic trees (Doyle & Gaut 2000). Finally, unlike the Biological Species Concept, the Phylogenetic Species Concept is able to evaluate both extant and extinct species as well as sexually reproducing and asexual species (Agapow *et al.* 2004).

The evolutionary species concept, rooted in phylogenetics, delineates species as a group of organisms that has a particular identity which is separate from other such groups in both time and space (Wiley & Mayden 2000). The evolutionary species concept is based on the notion that species have their own history and evolutionary fate (Wiley & Mayden 2000). Within the context of biological diversity, the evolutionary species concept is unique in that it can integrate all aspects of biological diversity as evidence of evolutionary isolation, including ecological, physiological, and genetic data (Mayden 1997). This concept may be the most appropriate and

general concept described to date, as it relates to the study of biological diversity, because it can take into account species groups as defined by all other species concepts and character systems (Mayden 1997). For example, a species as defined when applying the morphological species concept, with a particular morphological trait distinguishing the species, would also be defined as a species through the evolutionary species concept as the characteristics of the species can be identified as a particular morphology that has occurred through an evolutionary path in time (Mayden 1997). Finally, this concept does not need to rely on a single set of information, such as reproductive isolation or specific morphological characteristic, allowing very similar-looking sets of morphological species as well as asexual species to potentially be recognized as distinct (Wiley & Mayden 2000).

The taxonomy of Canadian marine planktonic crustaceans

Marine plankton are comprised of organisms living in the water column, which are very diverse (Hays *et al.* 2005; Bucklin *et al.* 2011; de Vargas *et al.* 2015). Although most species that make up marine plankton have some mobility, organisms are only considered to be planktonic if they cannot move against ocean water currents or if they are unable to remain stationary in ocean currents (Hays *et al.* 2005). Numerous groups make up the plankton, including single- and multi-celled phytoplankton and zooplankton, with zooplankton having 11 phyla (Bucklin *et al.* 2010b). Zooplankton are made up of holoplankton, animals that live their entire lives in the water column, and meroplankton, species which are part of the plankton for only a portion of their lives, usually during their juvenile stage (Bucklin *et al.* 2010b; Renaud *et al.* 2015). This work focuses on the metazoan zooplankton, a major component of the plankton (Bron *et al.* 2011). Zooplankton feed on other plankton, including phytoplankton, making them an important link in marine food webs, passing energy from primary producers up to higher trophic levels (Bron *et al.* 2011).

Marine planktonic crustaceans, including copepods, make up a large portion of the specimens by abundance in the marine plankton (Longhurst 1985). Crustaceans are a subphylum within the phylum Arthropoda (Regier *et al.* 2005). The focus of this work is on a taxonomic group within the marine plankton which contains the Copepoda and Thecostraca. This taxonomic group was selected as its members are abundant and diverse and make up a large portion of the zooplankton in Canadian marine waters. Past work had treated these two taxonomic groups as

classes under the superclass Maxillopoda, which also contained the classes Orstenocarida, Ostracoda, Branchiura, Skaracarida, and Mystacocarida (Newman 1992). However, more recent work has indicated that the Maxillopoda is not monophyletic (Regier *et al.* 2005). Therefore, the taxonomic placement of Copepoda and Thecostraca within crustaceans has been redefined (Newman 1992; Regier *et al.* 2010; Oakley *et al.* 2013). Recent work has supported the Copepoda and Thecostraca as sister lineages within the Multicrustacea (Regier *et al.* 2010; Oakley *et al.* 2013). Oakley *et al.* (2013) conducted an investigation of the phylogeny of crustaceans; to provide well-supported phylogenetic conclusions, they applied phylogenetic methods using multiple lines of evidence, or total evidence (including transcriptome data, morphological data including fossil records, available expressed sequence tag data, mitochondrial DNA sequence data, and nuclear DNA sequence data). Their results suggest that the Copepoda/Thecostraca lineage, which they name Hexanauplia, is sister to all Malacostraca and that these three classes make up the Multicrustacea (Figure 1.1). Although the name is no longer accepted, it was often necessary for this thesis research to use the term Maxillopoda when searching for public sequence data, as many databases have not been updated to the currently-accepted taxonomy.

DNA barcoding: applications and challenges when studying marine plankton

DNA barcoding, analogous to the universal product codes (UPCs) used to identify products in stores, is a means by which specimens can be identified to a specific species using short DNA segments (Mitchell 2008). The term barcoding, with respect to biological identifications, was initially used by Arnot *et al.* (1993) followed by Floyd *et al.* (2002) and later described by Hebert *et al.* (2003a), where it was suggested that a standard system using a 648 base-pair segment of DNA from the 5' end of the cytochrome *c* oxidase subunit I (COI) mitochondrial gene was a reliable means of identification for animal species. The DNA barcoding method is not a species concept but was originally framed as a way of identifying specimens to known species and is an extension of molecular phylogenetics and morphological species concepts (Hebert *et al.* 2003a; Pons *et al.* 2006).

One of the important attributes of the COI barcode region is that it is flanked by two nucleotide regions which are fairly conserved in most animal species (Folmer *et al.* 1994; Hebert *et al.* 2003a,b; Bucklin *et al.* 2011). These primer regions, short 10-20 nucleotide sequences

which are the initiation point of DNA synthesis using Taq polymerase, can then be used to PCR amplify the COI barcode region across a large breadth of taxonomic diversity (Hebert *et al.* 2003a,b; Bucklin *et al.* 2011). Additionally, the COI region is a good choice as a molecular region for species-level identification due to its lack of introns, high cell copy numbers, limited exposure to recombination, and its nearly universal uniparental maternal inheritance (Saccone *et al.* 1999; Bucklin *et al.* 2011; Breton & Stewart 2015). Using this region, Hebert *et al.* (2003a) demonstrated that the COI 5' region displayed sufficient variability necessary to identify animal specimens to a species level.

The barcoding approach is focused on creating and maintaining a specific procedure that provides standardization through quality control with required specimen vouchering. These efforts provide comparable and verifiable data even when collected across various laboratories and by different researchers (Mitchell 2008). This standardization is important due to potential variations in identifications across specialists and the application of differing species concepts to describe species (Zhou *et al.* 2010). The ultimate outcome of the barcoding project is to create a complete species library, against which researchers are able to query the sequence data from an unidentified specimen for information on specimen identity (Meyer & Paulay 2005). This tool also needs to enable researchers to flag previously unknown species as distinct, due to recent speciation or morphologically cryptic speciation (Meyer & Paulay 2005; Frézal & Leblois 2008) or due to the collection of a previously unsampled species.

The utility of barcoding comes from the COI region and its high interspecific variability and lower intraspecific variability (Dettai *et al.* 2011; Hubert & Hanner 2015). This high interspecific and low intraspecific variability may be used for two applications within DNA barcoding: identification or delimitation. The first, identification, uses the barcode region for an unknown specimen to compare against a data base of previously-identified DNA barcodes to identify the specimen to a species. The use of DNA barcodes as a means of species-level identification, initially tested primarily in Lepidoptera (Hebert *et al.* 2003a, b), has since been successfully applied to many taxonomic groups (Costa *et al.* 2007; Bucklin *et al.* 2011; Raupach & Radulovici 2015). Using DNA barcodes to identify specimens as belonging to a particular species is dependent on authoritative species identifications and subsequent sequencing of the barcode region and uploading to a sequence library that can be queried (Archambault *et al.* 2010).

Delimitation is a method where the variation present in the in COI-5P region is used to cluster sequences into species-like groups, which can suggest the expected number of species present in a collection based on the barcode data alone. The use of DNA barcoding in this respect extends past identification and has potential applications for biodiversity and biogeographical studies through monitoring programs, including characterizing species richness, evenness, and geographical distribution (Bucklin *et al.* 2011). Although distinct separation between interspecific and intraspecific variation has been documented across a range of taxonomic diversity, and a 2% threshold has been applied to many taxa, there have been some exceptions noted where a 2% threshold does not provide groups consistent with morphological species identifications (e.g. cowries, turbinids, and limpets: Meyer & Paulay 2005; *Daphnia*: Jeffery *et al.* 2011; spiders: Nadolny *et al.* 2016). With the discovery of these exceptions, it was also noted that genetic divergence thresholds have a lower error rate when a set of specimens is comprised of phylogenetically more distant species compared to closely related species (Meyer & Paulay 2005).

In addition, early criticism of a threshold approach to species delimitation stemmed from including relatively few species and few specimens per species within the taxonomic groups evaluated (Meyer & Paulay 2005). Meyer and Paulay (2005) suggested that the barcoding gap, the difference between intraspecific and interspecific divergences, will shrink, or may possibly disappear, with more sampling. More recent work has shown that although the barcode gap may decrease with increased sampling, the disappearance of the gap entirely, and loss of the utility of the DNA barcoding approach, does not always occur (Bergsten *et al.* 2012; Huemer *et al.* 2014). Despite the criticisms for using barcodes and a threshold approach to delimit species-like units, different approaches have been investigated to identify the most suitable thresholds for major taxonomic groups (Lefébure *et al.* 2006; Ratnasingham & Hebert 2013; Chapter 2 of this thesis), and in the years since its inception, barcoding has proven to be reliable at delimiting species-like groups for a wide range of taxa (Meyer & Paulay 2005; Bucklin *et al.* 2011; Ratnasingham & Hebert 2013).

One of the benefits of a barcoding approach is in the independent evaluation of previously-established species groups, which could potentially lead to taxonomic revisions, and the elucidation of synonymous or cryptic species (Bickford *et al.* 2007). Cryptic species are two or more genetic or evolutionarily distinct species present under a single species description

(Bickford *et al.* 2007). Since the beginning of barcoding, the detection of potential cryptic species, even within well-described species groups, has been a frequent finding (Hebert *et al.* 2004; Bickford *et al.* 2007; Kerr *et al.* 2007; Hubert & Hanner 2015). Once potential cryptic species are elucidated, further investigation as to the cause of the discrepancy is necessary to establish whether it is an error in barcode results (such as a case of contamination or amplification of a pseudogene), a potential incorrect identification of the specimen in question, or a morphologically cryptic but evolutionarily distinct species (Dettai *et al.* 2011). In the 13 years since large-scale DNA barcoding was established, there have been many potentially cryptic species revealed across a broad range of taxonomic diversity (Hubert & Hanner 2015).

The use of DNA barcoding has also been valuable when studying difficult-to-identify organisms such as those present in the marine plankton. Extensive work using the COI-5P animal barcode region has been conducted on marine taxa (Bucklin *et al.* 2011). The proper identification of marine zooplankton to established species using DNA barcoding has been successful across many taxonomic groups, including gastropods, copepods, and marine benthic crustaceans with zoea larvae (Bucklin *et al*. 2007; Costa *et al.* 2007; Puillandre *et al.* 2009; Blanco-Bercial *et al.* 2014).

There are numerous advantages of using barcoding to identify marine plankton specimens. The identification of specimens across complex life histories and across male/female morphological variations is expedited using a barcoding method (Puillandre *et al*. 2009). Species with a high degree of phenotypic plasticity are easily identified using barcodes (Radulovici *et al.* 2010). In addition to identifying whole specimens, the barcoding procedure can be used to identify damaged specimens, portions of specimens, or consumed material in the gut of specimens (Frézal & Leblois 2008). Each of these advantages of using barcoding is further reinforced by the speed of molecular DNA acquisition compared to the time-intensive traditional taxonomic method (Ratnasingham & Hebert 2007).

Even with the clear advantages of DNA barcoding, it still has limitations, acknowledged by barcode researchers, including biological difficulties such as: low resolution among recently-diverged species; no resolution or very little resolution with hybrids and slowly-evolving species; and the presence of nuclear pseudogenes which can provide incorrect sequence data, leading to possible misidentifications or incorrect recognition of cryptic species (Moritz & Cicero 2004). Low resolution using the mtDNA COI region has been reported among certain marine

invertebrates, including the phyla Porifera (sponges) and Ctenophora (comb jellies) and the Cnidaria class Anthozoa (corals and sea anemones), and these cases are attributed to a lower mitochondrial evolutionary rate, resulting in insufficiently variable mitochondrial sequences to discriminate between closely-related species (Bucklin *et al.* 2011; Hotke 2015). Hybridization events between two species provide further challenges and yield a pattern of low interspecific diversity, resulting in difficult species delimitations or entirely overlooking hybrid individuals (Galtier *et al.* 2009). Finally, balancing selection could occur following reintroduction of alleles into a population from a related species. This reintroduction could result in a single species with two mtDNA sequences, one of which is identical to another species, thereby confusing barcode analysis and underestimating the species diversity (Galtier *et al.* 2009).

Nuclear pseudogene amplification is problematic due to analysing barcode data with potentially different rates of molecular evolution (Bensasson *et al.* 2001; Richly & Leister 2004). Furthermore, the accumulation of multiple pseudogenes in the nuclear DNA can further complicate the assumed uniparental barcode analysis (Richly & Leister 2004). Additional examples of problematic situations in DNA barcoding are biparental and doubly uniparental inheritance of mitochondrial DNA, where mitochondrial sequences can potentially be amplified from two sources (Barr *et al.* 2005; Passamonti *et al.* 2011; Teske *et al.* 2012; Zouros 2013). Analysis of multiple molecular regions could provide further evidence to discriminate between alternative modes of inheritance among species suspected of biparental mitochondrial inheritance; such additional data may be particularly necessary in specific marine taxa, such as some fish and bivalve species (Song *et al.* 2008; White *et al.* 2008; Breton & Stewart 2015).

Thesis Overview

This PhD represents novel research contributions that increase our understanding of marine plankton distributions and biodiversity across Canadian marine waters as well as provide a better understanding of methodologies to facilitate research of this biodiversity. There are two main elements to this research. The first focuses on testing and developing analysis approaches to enable greater applications of DNA barcoding for marine plankton research, thereby contributing to our understanding of the current plankton diversity in Canadian marine waters. This is approached through the analysis and subsequent justification for using two possible

molecular markers in delimiting and identifying marine planktonic crustacean species. I evaluate two markers: the 5' end of the cytochrome c oxidase subunit I gene (COI-5P), the "animal barcode", and the V4 region of the nuclear small subunit ribosomal RNA gene (18S), both of which have been used in DNA barcoding, metabarcoding, and/or molecular systematics in the marine plankton literature (Bucklin *et al.* 2010a, 2011; Cristescu 2014; Zhan *et al.* 2013). I evaluate the effectiveness of these molecular markers for species delimitation using both internal (sequence-based) and external (morphology-based Linnaean identifications) verification. My focus for this research is with similarity-based delimitation approaches, as these methods are computationally efficient and therefore necessary for use with the large data sets that are increasingly available. I then apply a bi-directional concordance analyses to determine the degree to which these different data sets agree. The second element to this work uses the COI-5P molecular markers to investigate the biogeographical patterns of marine planktonic crustaceans among Canada's three oceans.

Chapter 2

Objective

The first objective of this research was to determine whether a biologically meaningful global pairwise sequence divergence (GPSD) threshold exists between intraspecific and interspecific DNA sequence divergences in planktonic marine crustaceans for the barcode region of the cytochrome *c* oxidase subunit I (COI-5P) gene.

Question

1. Does a COI-5P barcode gap, separating interspecific and intraspecific DNA sequence divergences, exist for marine planktonic Hexanauplia crustaceans (Copepoda and Thecostraca), which can reliably delimit specimens into species-like clusters largely concordant with known taxonomy?

Significance

This work contributes to our understanding of the behaviour of species delimitation methods in marine planktonic crustaceans with large population sizes and geographic ranges. My application of the easy-to-implement 'elbow' analysis approach is novel and significant in that it does not require an external criterion for threshold identification. This approach, not previously

used in published barcoding research, provides the optimum GPSD threshold for a research dataset. The use of concordance/discordance methods is also a significant contribution to the greater understanding of DNA barcode applicability in that it explicitly quantifies the fit between barcode-based MOTU and known external measures, such as morphological identifications.

Conclusion

The COI-5P barcode data for the marine planktonic crustaceans displayed an intraspecific/interspecific barcode gap between 2.1% and 2.6% GPSD, depending upon the MOTU clustering method employed. Using the obtained threshold, the resulting molecular clusters were largely concordant with morphological identifications. This range between 2.1-2.6%, obtained using all publically available data, provides a further refinement to the previously-suggested threshold of between 2-3%. Furthermore, I conclude that the two highlighted methods, obtaining molecular thresholds using an internal assessment (elbow analysis) and comparing cluster results between data sets using a formal concordance metric, should be considered for use in future barcoding studies.

Chapter 3

Objective

The second objective of my thesis was to assess the 18S-V4 molecular region in two separate aspects. First, I assessed the utility of 18S-V4 for delineating marine planktonic crustacean species that are largely concordant with established taxonomic groups. Secondly, I tested whether the 18S-V4 region is suitable for use as a DNA barcode, whereby the molecular region could be used to identify marine planktonic specimens to species. Finally, I discuss the applicability and limitations of the 18S-V4 region for metabarcoding studies employing high-throughput sequencing.

Questions

1. Can the 18S-V4 region be used as a molecular barcode to group specimens into established morphological species and MOTUs delineated using COI barcode data?
2. Do the MOTUs as determined through analysis with 18S-V4 sequence data show high levels of concordance to COI-5P molecular clusters?

Significance

This work contributes to the understanding of the 18S-V4 region and its applicability as a DNA barcode for metabarcoding marine planktonic crustaceans. Given the cost, time, and expertise constraints for traditional taxonomic research of marine planktonic crustaceans, metabarcoding studies have become increasingly popular. However, given the limitations of metabarcoding technologies, such as the requirement of a short sequence length (50-500 bp), an informative and easily amplifiable molecular barcode region is required. The 18S-V4 region is small (250-450 bp; Wu *et al.* 2015; Chapter 3) and readily amplifiable; however, it is unclear if it contains enough information to be effectively used as a DNA barcode to identify specimens to species and delimit suspected species groups. My work addresses the significant research gap in understanding the effectiveness of 18S-V4 data to accurately capture species-level information.

Conclusion

The 18S-V4 region was easily sequenced with a high (>80%) degree of success. However, the region did not contain enough information to be used as a molecular barcode and did not reliably delimit species, at least within the Multicrustacea studied here. However, with careful consideration of the research question and an understanding of the limitations of the information gained from the 18S-V4 molecular marker, the 18S-V4 may be a suitable choice for select studies, particularly when paired with amplification of other markers including COI.

Chapter 4

Objective

The third objective of my thesis was to investigate biogeographic patterns of marine planktonic crustaceans. This objective was approached by compiling two separate data sets. First, I assembled publically accessible COI-5P molecular sequence data of select marine calanoid (Multicrustacea: Copepoda: Calanoida) planktonic crustaceans and determined their Canadian biogeographic patterns. Using results obtained from a literature search for Canadian benthic marine invertebrates, I compared the patterns of biogeography present in the calanoid taxa with benthic invertebrates. Finally, I discuss the differences in the distributions across Canada's ocean regions with respect to benthic vs. planktonic species and discuss my findings in relation to historical habitat and evolutionary processes such as allopatric speciation.

Questions

1. Is there a shared distributional pattern of marine planktonic calanoids across Canada's three oceans, indicating a shared biogeographic history?
2. Is this distributional pattern similar to benthic marine species?

Significance

This work contributes to our understanding of planktonic distributional patterns across North Pacific, Arctic, and North Atlantic Oceans. Although taxonomically-focused studies have been conducted on benthic marine organisms (Hardy *et al.* 2011; Carr *et al.* 2011; Carr 2012; Layton *et al.* 2016), the distribution of marine planktonic calanoids across Canada's three ocean regions has received relatively little attention. Having a better understanding of the current distributional patterns of marine plankton may provide insights into potential future distributions as temperatures in the Arctic and Northern Canadian regions increase. Furthermore, this research contributes to our understanding of planktonic calanoid species boundaries and the role that allopatric divergence may have played in speciation.

Conclusion

Marine planktonic calanoid species with representation in Pacific, Atlantic, and Arctic waters displayed phylogeographic structure, with Pacific collections forming molecular groupings separate from Atlantic/Arctic collections. This closer relationship between Atlantic and Arctic waters is consistent with benthic marine invertebrate species and suggests that isolation across the Bering Strait similarly played a role in structuring planktonic copepod genetic diversity.

Tables and Figures

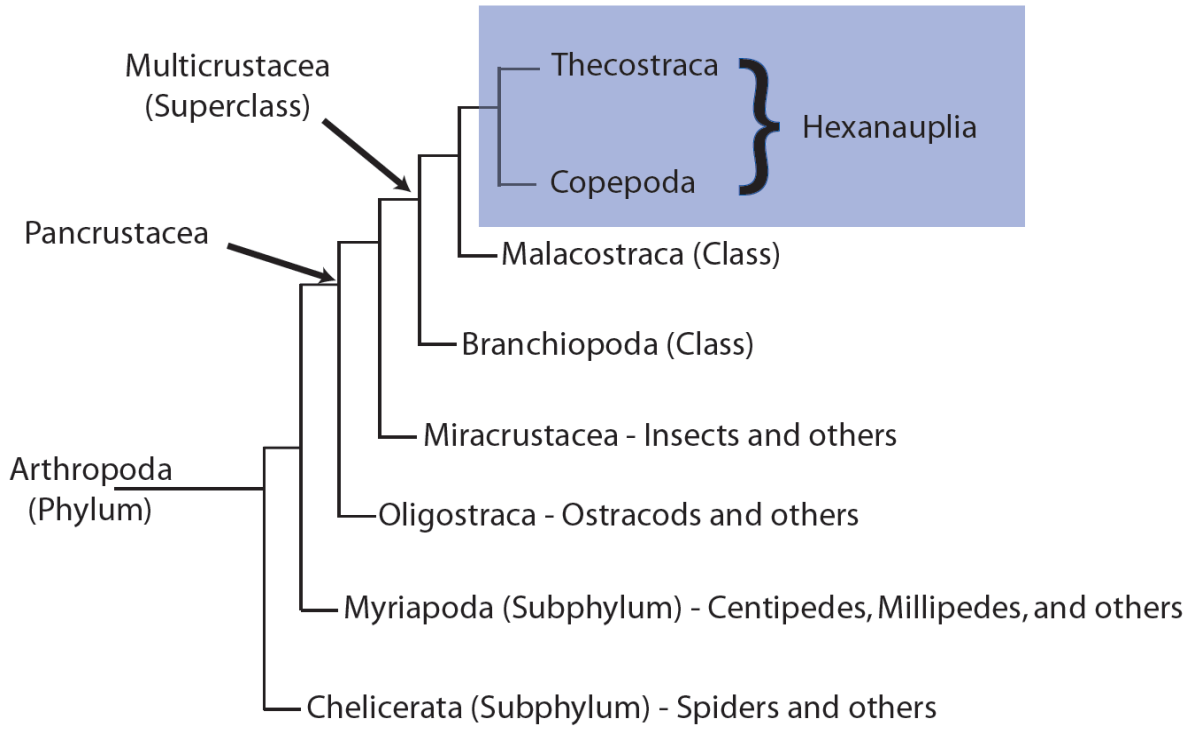


Figure 1.1. A phylogram showing the relationship of Thecostraca and Copepoda with respect to other major Arthropod lineages. Topology is based on Regier *et al.* (2010) and Oakley *et al.* (2013). The section of the tree shaded in blue indicates the group Hexanauplia (Oakley *et al.* 2013) and represents the taxonomic focus of this work.

Chapter 2

Barcode-based species delimitation in the marine realm: a test using Hexanauplia

Abstract

DNA barcoding has been used successfully for identifying specimens belonging to marine planktonic groups. However, the behaviour of methods for specimen identification and species delimitation remain untested for taxonomically diverse and widely-distributed marine groups, such as the Copepoda and Thecostraca. We investigate whether a cytochrome *c* oxidase subunit I (COI-5P) global pairwise sequence divergence threshold exists between intraspecific and interspecific divergences in the copepods plus the thecostracans (barnacles and allies). Using publicly accessible sequence data, we applied a graphical method to determine an optimal threshold value. With these thresholds, and using a newly-generated planktonic marine data set, we quantify the degree of concordance using a bi-directional analysis and discuss different analytical methods for sequence-based species delimitation (BIN, ABGD, jMOTU, UPARSE, Mothur, PTP, GMYC). Our results support a COI-5P threshold between 2.1 and 2.6% p-distance for these crustacean taxa and yielded molecular groupings largely concordant with traditional, morphologically-defined species. The adoption of internal methods for clustering verification enables rapid biodiversity studies and the exploration of unknown faunas using DNA barcoding. The approaches taken here for concordance assessment also provide a more quantitative comparison of clustering results (as contrasted with "success/failure" of barcoding), and we recommend their further consideration for barcoding studies.

Introduction

Quantifying biodiversity in marine ecosystems has become increasingly important due to a rise in anthropogenic disturbances and increases in local and global extinction rates (Singh 2002; Radulovici *et al.* 2010). Traditional morphological approaches to specimen identification are not realistic for completing comprehensive surveys of marine regions due to cost, expertise, and time required. Additionally, even when these are available, morphological identifications are challenging for many specimens due to variation in morphological characters across life stages, poorly-resolved taxonomy in some groups, and cryptic or undescribed species (Knowlton 1993; McManus & Katz 2009; Packer *et al.* 2009). These difficulties are further exacerbated when conducting broad geographical biodiversity surveys, as taxonomic expertise is often linked to specific taxonomic groups and/or geographic regions. As such, DNA sequence-based specimen identification systems, including DNA barcoding (Hebert *et al.* 2003a,b), are indispensable for conducting large-scale biodiversity surveys. Once established, molecular methodological pipelines, publicly accessible sequence databases, and tested analytical tools will not only facilitate biodiversity surveys but will also enable the rapid detection of introduced species through environmental sampling and high-throughput sequencing techniques (Cristescu 2014).

Identifying unknown specimens through DNA barcodes requires a reference library containing morphologically-identified, barcoded specimens against which unknowns can be compared (Collins & Cruickshank 2014). Currently, it is challenging to use DNA barcode databases for identification of many marine species; most species remain to be described, and even known species often have little to no DNA barcode coverage (McManus & Katz 2009; Bucklin *et al.* 2010a; Blanco-Bercial *et al.* 2014). Important first steps for enabling specimen identification using DNA barcodes are to investigate patterns of interspecific and intraspecific variation within target taxa and to determine the degree to which an integrative approach to species delimitation is necessary—as contrasted with more straightforward approaches using a single molecular marker (Collins & Cruickshank 2013, 2014). In addition to building an identification system for known species, it is important to be able to determine when a specimen is likely to represent a species that is novel to the database through some form of species delimitation.

The success of DNA barcoding has been documented across a range of marine taxa (e.g. da Silva *et al.* 2011; Blanco-Bercial *et al.* 2014; and references therein). Despite these successes, establishing a robust system is still challenging. With 31 phyla—and with an estimated million species remaining to be discovered—the genetic variation within marine metazoans is not fully appreciated (Bucklin *et al.* 2010a). Earlier molecular approaches to specimen identification relied on the presence of low intraspecific genetic variation and larger interspecific divergence between species, i.e. the presence of a "barcoding gap" (discussed in Meyer & Paulay 2005). Despite criticisms of strictly similarity-based approaches (Will *et al.* 2005; Hickerson *et al.* 2006; Collins & Cruickshank 2013), such delimitation methods have been shown to be useful among a wide range of taxa (Will *et al.* 2005; Ebach & Holdrege 2005; Radulovici *et al.* 2010; Huemer *et al.* 2014). Unfortunately, thresholds for species delimitation are often inappropriately chosen simply based on past literature (Collins & Cruickshank 2013). Although there has been work developing methodologies to justify the selection of a molecular threshold (Lefébure *et al.* 2006; Blanco-Bercial *et al.* 2014), these methods rely solely on external comparisons, often only to taxonomic species identifications, which can vary widely depending on the identifier (Shen *et al.* 2013). Nevertheless, once a well-defined and justified threshold is established, simple delimitation methods are advantageous to quickly and easily assess potential biodiversity.

The focus of this work is on a taxonomic group within the marine plankton which contains the Copepoda and Thecostraca. Past placement of these classes was under the superclass Maxillopoda (Newman 1992). However, more recent work has indicated that the Maxillopoda is not monophyletic (Regier *et al.* 2005), and the taxonomic placement of Copepoda and Thecostraca within the arthropods has been redefined (Newman 1992; Regier *et al.* 2010; Oakley *et al.* 2013) (Figure 1.1). Recent work using multiple lines of evidence (including transcriptome data, morphological data including fossil records, available expressed sequence tag data, mitochondrial DNA sequence data, and nuclear DNA sequence data) has placed the subclasses Copepoda and Thecostraca as sister lineages making up Hexanauplia (Regier *et al.* 2010; Oakley *et al.* 2013).

In this study, we investigate the prospects for using rapid species delimitation tools within a hyper-abundant and widely-distributed group of marine invertebrates, the subclasses Copepoda and Thecostraca. Using publicly available COI-5P barcode data, we first describe patterns of genetic divergence and explore the potential for a global pairwise sequence

divergence threshold to be used to delimit specimens into species-like units. Second, using a *novel* data set of morphologically-identified specimens, we also quantify the concordance among multiple sequence clustering methods as well as the concordance between molecular delimitations and current taxonomy. Our use of bidirectional concordances is a new approach for the barcoding literature. Our efforts provide insights into marine planktonic crustacean genetic divergence patterns and species boundaries under differing species definitions. In addition to contributing to the development of molecular identification systems for these taxa, the approaches employed here may be considered for other understudied marine invertebrates with large geographic ranges.

Methods

Specimen collection methodology

Plankton samples were collected from May 2011 to August 2012 at one proposed port location and eleven current port locations across all three of Canada's ocean regions (Arctic, Atlantic, and Pacific) (Figure 2.1). During off-ice periods, plankton samples were collected from small vessels using plankton nets of both 250 μm and 80 μm mesh sizes at 0 - 15 m depth. At most locations, two seasonal periods were sampled: July - September and November - December. Collections in northern regions were sometimes limited to one season due to logistical challenges. At each location, for each season collected, six separate plankton tows were made to provide representation across the entire port.

Samples were maintained in 95% ethanol and transferred to -20 °C cold storage within six months of collection. All samples were split into three fractions, one of which was used in this study. To reduce the number of samples for sorting and identification, all samples from a single mesh size from a single port were combined, and for the remainder of this manuscript these are referred to as "samples". Specimens within samples were sorted morphologically and were taxonomically identified to the lowest possible level (identification references: Nouvel 1950; Gardner *et al.* 1982; Roff *et al.* 1984; Kathman 1986; Todd *et al.* 1996; Johnson 1996; Gerber 2000; Johnson & Allen 2012). Four to six individuals of each morphologically-identified maxillopod taxon per sample were used for further molecular analysis.

Molecular laboratory methodology

DNA extraction consumed single whole individuals, as all individuals were less than approximately 1 mm^3. Batch vouchers were designated that consisted of individuals from the same site and assigned the same morphological identification as the sequenced specimen. The specimens will be archived at the Biodiversity Institute of Ontario, University of Guelph, and the digital specimen information is available through the Barcode of Life Data Systems (BOLD, http://www.boldsystems.org/) (Ratnasingham & Hebert 2007). DNA extraction followed Ivanova *et al.* (2006) with six variations: (1) a CTAB lysis solution (2% CTAB, 100mM Tris-HCL, pH 8.0, 20mM EDTA, pH8.0, 1.4M NaCl) was used in place of the indicated Vertebrate Lysis Buffer; (2) a 1 mm^3 piece of specimen was used; (3) a 1.2X dilution (using ddH_2O) of binding buffer was used; (4) after addition of binding buffer, the total solution was incubated on the bench top for 5 min; (5) a 2X dilution of protein wash buffer was used; and (6) final DNA extracts were eluted in 50 µl of ddH_2O.

Multiple PCR primer sets were used to amplify the animal barcode region. The predominant primer set used was LCO1490 / HCO2198 (Folmer *et al.* 1994). Additional primer sets used are listed here in order of amplification success: (1) Folmer *et al.* (1994) primers tailed with M13 (Messing 1983) (LCO1490_t1 / HCO2198_t1); (2) a zooplankton-specific primer pair (ZplankF1_t1 / ZplankR1_t1) (Prosser *et al.* 2013); (3) a degenerate primer set based on the Folmer primers (dgLCO1490 / dgHCO2198) (Meyer 2003); (4) a Lepidoptera and Folmer primer cocktail (C_LepFolF / C_LepFolR, 1:1) (Folmer *et al.* 1994; Hajibabaei *et al.* 2006); and (5) a crustacean primer set (CrustDF1 - GGTCWACAAAYCATAAAGAYATTGG, CrustDR1 - TAAACYTCAGGRTGACCRAARAAYCA) (Steinke unpublished). Initial PCRs were processed in 96-well plates following the Canadian Centre for DNA Barcoding (CCDB) protocols (Ivanova & Grainger 2007a). A second PCR protocol was tried when necessary and had a final chemistry of: 2 mM $MgCl_2$, 0.2 mM dNTP's, 0.4 µM forward and reverse primer, 1X PCR buffer, 10% DNA template by volume, and 0.0064 Units of Taq/µl in final volume. The thermocycling regime used for all reactions was: 95°C for 5min; 40 cycles of 95°C for 40sec, 50°C for 40sec, 72°C for 60sec; 72°C for 10min. PCR products were visualized using either a 1.5% agarose gel or a bufferless E-gel system (Life Technologies). PCR products were cleaned using EXOSap-IT. Bidirectional sequencing followed CCDB sequencing protocols using Big Dye 3.1 (Ivanova & Grainger 2007b).

CodonCode Aligner (CodonCode Corporation) was used to display sequence quality and assemble consensus sequences. All unassembled chromatograms and consensus sequences of less than half the expected length (~650 base pairs [bp]) were visually inspected and removed if the quality of the chromatograms and subsequent base calling was poor. Sequences were aligned using the default (FFT-NS-2) alignment strategy of the Multiple alignment program for nucleotide sequences (MAFFT Ver. 7) (Katoh & Standley 2013), and the multiple sequence alignment (MSA) was trimmed to a final length of 588bp (Appendix I). The MSA was then translated using the invertebrate mitochondrial code in MEGA6 (Tamura *et al.* 2013) for verification of the alignment. Single nucleotide insertions or deletions evident through frame shifts were further investigated and edited if they revealed an error in the nucleotide base reading. In cases of an unresolved frame shift or stop codon, the sequence was removed as this pattern suggests the presence of a nuclear pseudogene.

Screening data for potential contaminations and/or misidentifications

Neighbour-joining (NJ) phenograms (Saitou & Nei 1987) using Kimura-2-parameter (K2P) distances (Kimura 1980) were constructed through MEGA6 and tested using 10,000 bootstrap pseudoreplicates. The K2P model was chosen due to its prevalence in the barcoding literature, enabling comparisons across studies; p-distances, although advocated by some authors (Srivathsan & Meier 2012; Collins *et al.* 2012), vs. K2P are expected to behave similarly at small genetic distances (Srivathsan & Meier 2012; Collins *et al.* 2012). Phenograms were visually inspected for potential contaminants; non-target DNA could be sequenced as a result of DNA present in the shared ethanol storage medium, because of preferential binding of primers to a symbiont or prey item, or due to trace contamination during laboratory procedures. Two possible scenarios resulted in the removal of sequences: (1) a search of the COI-5P barcode data on BOLD's public data portal returned a result where there was 100% placement probability to a class other than Maxillopoda (mismatches were verified using the tree-based identification tool in BOLD to guard against matches to incorrectly identified sequences in the database); and (2) where three or more individuals were morphologically identified as being the same species, sequences that displayed more than 20% divergence (Blanco-Bercial *et al.* 2014) from others were deleted due to presumed contamination or misidentification (re-identifications for individual specimens, e.g. as in Renaud *et al.* 2012, were precluded by the consumptive DNA

analysis). Specimens morphologically identified as the same genus or species by the same identifier and from the same sample, but which grouped more closely with another genus, were flagged as potential contaminants and saved for further verification using additional information, as they may have indicated biologically significant variation, such as cryptic species or hybridization.

Data sets and molecular operational taxonomic units

Two data sets of COI-5P sequences were analyzed (Table 2.1). The *novel* data set consisted of marine collections sequenced here that were morphologically identified to the species level. The second set, *reference*, contained all Copepoda and Thecostraca specimens collected here which were not able to be identified to the species level together with all publicly accessible Maxillopoda data on the BOLD system (BOLD search for "Maxillopoda" in the public data portal, using the API search method conducted on December 1, 2014 http://www.boldsystems.org/index.php/API_Public/specimen?taxon=Maxillopoda). Although the taxonomic designation Maxillopoda is no longer accepted, it was necessary to use it when searching for public sequence data, as many databases have not been updated to the currently-accepted taxonomy. The two data sets were used in several different and several similar analyses, which are described below and visually displayed in Figure 2.2. The *reference* data set was then reduced by excluding those sequences not assigned a Barcode Index Number (BIN), to remove sequence data not meeting the minimum quality standards for BIN compliance (Ratnasingham & Hebert 2013).

Genetic distances were calculated and summarized using the 'Distance Summary' and 'Barcode Gap Analysis' tools on BOLD (Ratnasingham & Hebert 2007). All sequences longer than 500 base pairs were analyzed for the *reference* and *novel* data sets. Analyses were conducted using the BOLD sequence alignment, K2P (Kimura 1980) genetic distances, and pairwise deletion of missing data.

Five similarity-based and two coalescence-based analyses for generating molecular clusters or Molecular Operational Taxonomic Units (MOTU) (Blaxter 2004) were compared. Similarity-based methods included: Barcode Index Number (BIN) (Ratnasingham & Hebert 2013), Automated Barcode Gap Discovery (ABGD) (Puillandre *et al.* 2012), jMOTU (Jones *et al.* 2011), UPARSE (Edgar 2013), and Mothur (Schloss *et al.* 2009). The two coalescence-based

methods were Poisson Tree Processes (PTP) (Zhang *et al.* 2013) and Generalized Mixed Yule Coalescent (GMYC) (Fujisawa & Barraclough 2013). Coalescent-based phylogenetic methods were included in the comparison for the *novel* data set only, as the *reference* data set was too large and construction of an input tree too computationally expensive, while similarity-based methods were applied to both data sets. Prior to model testing and tree construction, all exact duplicate sequences were removed from the MSA using ElimDupes (https://hcv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html). Phylogenetic reconstruction for the *novel* data set used the best-fit model of nucleotide substitution as determined using the Bayesian Information Criterion (BIC) implemented in the jModelTest2 program (Guindon & Gascuel 2003; Darriba *et al.* 2012). Talavera *et al.* (2013) recently investigated the results from GMYC when using input trees variously constructed and largely found the same resulting clustering assignments. We also used three different tree-building methods (ultrametricized trees constructed using Bayesian, neighbour-joining, and maximum likelihood methods) for use by the coalescent-based MOTU delineation programs. As the clusters from GMYC, as well as PTP, across all three input trees were similar, detailed methods and results are presented for the Bayesian tree only, consistent with recommendations by Tang *et al.* (2014). PTP and GMYC clustering results using unique haplotypes were used to assign all sequences/specimens in the data set to a cluster.

The Bayesian tree was constructed with the Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond *et al.* 2012) program using the GTR+G+I model and a log-normal relaxed clock with rate estimation, as a strict clock is less likely to apply across the large and diverse multicrustacean groups studied. The tree prior was set to the Yule process, and the initial ucld.mean value was set between 0 and 10. Five independent runs of 100,000,000 Markov Chain Monte Carlo (MCMC) generations were conducted, with sampling every 1000 generations. Results for each of these runs were inspected using the Tracer program (Rambaut *et al.* 2014), with convergence visually verified. All five runs were combined using the BEAST program LogCombiner, where a 10% burn-in was applied (first 10% of trees eliminated), and 45,000 states for each MCMC run were subsampled from each individual run. The BEAST program TreeAnnotator was then used to summarize the LogCombined tree file into a single target tree by finding the best-fit tree using the specified maximum clade credibility type tree. The GMYC (Fujisawa & Barraclough 2013) analysis was conducted using python implementation of the

single-threshold model as downloaded from the Exelixis Lab webserver (http://species.h-its.org/gmyc/). The Poisson Tree Processes (PTP) model was also used to infer putative species boundaries (Zhang *et al.* 2013).

The BOLD-implemented refined single linkage (RESL) algorithm provided BIN assignments for each sequence (Ratnasingham & Hebert 2013). This method uses a 2.2% p-distance seed threshold, but then refines groupings for individual BINs and neighbouring clusters based on the level of continuity in the distribution of genetic divergences among sequences. The jMOTU program uses a similarity-based approach; a BLAST identity filter of 99 was employed, and clusters were arranged using the number of variable nucleotides, equivalent to p-distance (Jones *et al.* 2011). For the remaining similarity-based methods (ABGD, Mothur, and UPARSE), we set the programs to define molecular clusters using p-distance so that results could be compared across methods (Puillandre *et al.* 2012; Schloss 2009; Edgar 2013).

The ABGD method was implemented through the ABGD C source available on the ABGD website (Puillandre *et al.* 2012). Analysis settings were: Pmin value of 0.001, Pmax equal to 0.15, a minimum gap width X equal to 0.001, with 1000 steps using p-distance. The gap width used here was much smaller than the default width of 1.5, and the number of steps was much larger at 1000 compared to 10. This was done to obtain enough clustering results across a range of divergences to conduct an elbow analysis (explained below). UPARSE implementation with threshold values greater than 3% is not recommended, and a workaround for this problem was implemented in order to explore a broad range of possible thresholds, which can be seen in the supplementary files (Appendix II). The Mothur program had three clustering options available for MOTU assignment (nearest neighbour, furthest neighbour, and average neighbour), and the average neighbour method was used. Commands and scripts for the generation of the results from Mothur can be found in the supplementary files (Appendix II).

Determining optimal global thresholds

To calculate the optimal divergence threshold for each similarity-based clustering method, clusters were generated for the *reference* data set using a range between 0% and 15% p-distance pairwise divergence thresholds. This range was chosen as it is expected to encompass the general transition between intraspecific and interspecific variation within our COI-5P data set (Blanco-Bercial *et al.* 2014). Based upon prior studies of morphology and barcode variability

(e.g. Bucklin *et al.* 2003; Lefébure *et al.* 2006; Ratnasingham & Hebert 2013), treating every COI haplotype as a different species would result in oversplitting the data. Units would not resemble species defined by other criteria such as morphological characteristics. The opposite of this would also be a problem; if too many haplotypes were placed into a single species group, too few species would be represented using the data. Here, we hypothesis that the barcode data themselves can reveal the transition point between splitting and lumping.

For each program for which the user can specify the threshold (ABGD, jMOTU, Mothur, and UPARSE), the numbers of clusters generated at each p-distance were plotted, and the vertex point of the resulting curve was considered to represent the optimal threshold for clustering these data for that method (Handl *et al.* 2005) (Figure 2.3; Appendix II). For pairwise divergence thresholds below the proposed optimum, the sequences will be over-split into too many MOTUs, having lower correspondence to evolutionary species units. Conversely, at pairwise divergence thresholds above the optimum, sequences will be over-lumped into too few MOTUs. To determine this point, a graphical approach was employed where the Euclidian distance between the origin of the graph (0,0) and every point on the curve was obtained. The point on the curve with the smallest Euclidean distance to the origin was considered the hypothesized, ideal global threshold value (Figure 2.3). To determine this threshold using empirical data, thereby foregoing the need to approximate the curve, the y-axis (number of MOTUs at a given threshold) was scaled to be equal in length to the x-axis. Analyses conducted using divergence values between 0-10% and 0-20% and scaling the y-axis similarly yielded very similar results (not shown). Once obtained, this threshold could then be used to compare clustering results to external criteria (like taxonomic identification) for verification. If the obtained threshold was verified, then the resulting threshold could be considered appropriate for the taxonomic group when analyzing large, computationally challenging data sets of poorly described species.

Concordance among MOTUs and between MOTUs and morphological species

Concordance among MOTUs generated by different analytical methods, as well as concordance between MOTUs and morphological species assignments, was quantified using an Adjusted Wallace coefficient (Wallace 1983). This coefficient was selected because it takes into consideration potential chance events leading to cluster agreements and because it provides bi-directional results. Calculation of the coefficient required a data matrix containing all sequence

data with each row representing a unique specimen and each column a unique clustering analysis result. Specimens were removed from the analysis if data were missing in any one of the clustering columns being analyzed. Once a matrix was constructed, Adjusted Wallace coefficients were computed through the website Comparing Partitions (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Home) (Severiano *et al.* 2011). Coefficients were determined for the *reference* and *novel* data sets separately, and MOTUs were generated for each data set using the analysis-specific thresholds obtained from the larger (*reference*) data set only.

Comparing MOTUs and morphological identifications

Overall concordance between two clustering methods quantifies how well they agree; however, it does not inform us of the nature of the agreements/disagreements nor highlight those possible problematic sequences or taxa that are yielding the conflicting clustering results. To understand how clustering results obtained from the various molecular methods agreed with morphological groupings, we quantified molecular cluster agreement to all Linnaean species labels for both the *novel* and *reference* data sets using an R script (Appendix III). Agreement/disagreement comparison analyses resulted in four possible outcomes: complete 'match' where both clusters match exactly; an exact 'split' where the reference cluster was split into multiple clusters, with no members of the corresponding clusters being unaccounted; a complete 'lump' where the reference cluster was combined with one or more additional reference clusters in their entirety, with no members unaccounted; and a 'mixed' result where a reference cluster was both split and lumped (Ratnasingham & Hebert 2013). Species represented by a single specimen were removed from this analysis, as these specimens are only able to represent matches in the analysis and would thus bias the results towards concluding concordance. An agreement matrix was constructed, and total match, split, lump, and mixed numbers were tabulated for each data set and clustering analysis.

Results

Novel molecular data set of Canadian marine Hexanauplia

We successfully generated 404 new barcode sequences. After applying a 500 bp length criterion and removing sequences containing more than 1% Ns, our study yielded 366 (247 identified to species) novel barcode sequences for a newly-sampled and morphologically-identified planktonic crustacean collection (Table 2.1B). Amplification of the COI-5P region was challenging, with some specimens receiving 8 PCR attempts using up to 6 different primer sets. Once a protocol was established, successful amplifications across the Hexanauplia data set increased, and final sequencing success was 59% of individuals, representing 100% of the morphological species. There were several groups which remained more difficult to amplify, including: *Calanus*, *Microcalanus*, *Metridia* (specifically *Metridia longa*), *Oithona*, *Paracalanus*, and *Pseudocalanus*.

Patterns of genetic divergence

There was an overall separation between intraspecific vs. interspecific divergences for both the *reference* and *novel* data sets (Figure 2.4). For the *reference* data set, the genetic divergence analysis was limited to those sequences with associated Linnaean species names, which resulted in a data set of 2825 sequences comprised of 262 species. The average for the mean intraspecific K2P divergence values for all species in the *reference* data set was 1.84%. The mean of the maximum intraspecific pairwise distances was 3.03%, and the mean distance to the nearest neighbour divergence (the smallest pairwise distance to the closest individual of a different species) was 14.79%. The *novel* data set, with 247 sequences representing 27 species, had a mean maximum intraspecific distance of 4.35% and a mean distance to the nearest neighbour of 19.82%. The average for the mean intraspecific divergence values for all species in the *novel* data set was 1.81%.

Proposed threshold values through elbow analysis

Global pairwise sequence divergence (GPSD) thresholds, proposed to represent an optimum generally separating interspecific and intraspecific divergences, ranged from 2.1-2.6% across clustering analysis methods for the *reference* data set. Analysis with ABGD displayed the

lowest percent divergences with a 2.1% result. The jMOTU and UPARSE analyses had threshold values of 2.3% and 2.2%, respectively, while the Mothur result had the highest value at 2.6%.

Concordance among MOTU clustering methods

Molecular clustering results with 2825 specimens, representing 309 uniquely identified species, from the *reference* data set (Table 2.1A) were similar across the four similarity-based methods using GPSD threshold values (Table 2.2). This concordance index can take values between 0 and 1, with higher values indicating a strong ability of one clustering approach (row label) to explain the clusters generated by another method (column label). Concordance values between a pair of methods can differ in accordance with the direction of the comparison.

The adjusted Wallace concordance values (Table 2.3) for the *novel* data set (Table 2.1B) showed varied concordance across the eight clustering methods. Due to the UPARSE function of identifying suspected chimeric sequences, three sequences were removed from all concordance calculations, leaving 244 specimens for analysis. There were noticeably directional results in discriminatory power between molecular clustering results and morphological Linnaean species labels. For example, the Wallace coefficient value from BINs to Linnaean species labels was 0.925, meaning that two specimens in the same BIN have a 92.5% chance they would also have the same Linnaean species label. By contrast, two specimens with the same Linnaean species label only have 56.2% chance of falling in the same BIN. This example is similar in all comparisons between Linnaean species labels and molecular clusters, where the molecular clusters are more discriminant than the species designations. The PTP coalescence-based clustering of the specimens partitioned the data far more than all other methods, with 51 MOTUs, compared with 37-40 MOTUs for most of the molecular methods and 29 morphological species groups. Moreover, PTP had the lowest concordance in comparison to all other clustering results, including both molecular and morphological techniques. The morphological species were best able to explain the ABGD results, with a concordance value of 0.704, but generally exhibited more modest ability to explain the molecular results as compared to all other molecular clustering methods (0.664 - 0.819 for ABGD to explain other molecular clustering results, 0.994 - 1 for ABGD clusters to be explained by all other molecular clustering methods). Sets of clusters generated by BIN, jMOTU, Mothur, and UPARSE had the highest level of concordance with one another among all comparisons (0.988 - 1.00). Overall, molecular

clustering methods, including both similarity-based and coalescence-based, had a strong ability to explain the morphological species; by contrast, the morphological species had a generally low ability to explain the molecular clustering results. This suggests molecular data are more discriminant and generate more clusters than the morphological data, as examined for assigning specimens to Linnaean species, for the *novel* data set.

Results comparing molecular clusters to morphological groupings

The extent of agreement between molecular clusters and morphological species, containing more than one named specimen, varied between *reference* and *novel* data sets. For the *novel* data set, Linnaean names were assigned during this study and are based upon morphology; it is presumed this is primarily the case for the *reference* data set as well, but varied methods (including integrative consideration of molecular data) could have been used in the public data set. The molecular-based clusters displayed between 40-80% direct matches with species labels in the *novel* data set and 55-66% using the *reference* data set (Figure 2.5). Of the 29 morphologically-identified species in the *novel* data set, 20 were represented by more than one specimen and so were included in the agreement analysis. Ten morphologically-identified species shared a high degree of agreement between barcode-based clusters and morphological identifications. Two genera, *Acartia* and *Centropages*, showed varying degrees of agreement among clustering methods including mixed, split, or matched specimen clustering assignment. The species *Eurytemora herdmani* and *Tortanus discaudatus* showed predominately split agreement to molecular clusters. The remaining identified species (*Paracalanus parvus, Temora longicornis, Zaus abbreviatus*) predominately matched molecular clustering. The *reference* data set exhibited 56.0% (153/273) exact matches across all clustering methods. Of the remaining Linnaean names, approximately 2.2% were predominantly matched to most molecular clustering methods, 11% split, 2.9% lumped, and 25.6% mixed.

Discussion

Here, we empirically estimate a COI-5P threshold for Hexanauplia, which ranges between 2.1 and 2.6% p-distance among the analytical methods employed. We then used these thresholds to quantify the degree to which molecular clusters agree with species units according

to current taxonomy. We discovered that, for approximately 60% of Linnaean species labels, molecular species delineation methods assigned specimens into the same groupings as morphological identifications. We suggest that concordance of independent data sets can provide greater support for species boundaries, with more confidence that the groupings arising from these different character sets reflect evolutionarily independent species (see Mayden 1997 for treatment of species concepts). Below, we further consider the situations in which barcode-based species delineation methods may be appropriate and discuss the biological meaning of the discordances observed between molecular and morphological approaches.

Support for rapid species delimitation for Hexanauplia using global thresholds

Although the use of GPSD thresholds has been criticised, after suitable study in a given taxon it has been well established as a successful approach for delimiting specimens into clusters that are largely concordant to established taxonomic groups (Will *et al.* 2005; Ebach & Holdrege 2005; Huemer *et al.* 2014). Unfortunately, research to test the performance of COI barcode data in marine invertebrate taxa has been limited (but see Radulovici *et al.* 2010; Bucklin *et al.* 2010a, 2011; Blanco-Bercial *et al.* 2014). Furthermore, studies investigating barcode performance often do so by comparing the relative performance of barcode data against morphologically-identified specimens. The primary use of morphological identification as the gold standard for concluding "successes" and "failures" of barcode data, which is common but not universal across the barcoding literature, presents problems with marine taxa such as Hexanauplia, as some reproductively, evolutionarily, and even ecologically distinct taxa lack discernible and diagnostic morphological differences (Carrillo *et al.* 1974; Knowlton 1993; McManus & Katz 2009; Bucklin *et al.* 2011). Few studies have tested the utility of barcode data for providing consistent clustering through an internal, sequence-based measure of cluster validation (Handl *et al.* 2005).

Establishing a threshold for hexanauplian COI-5P barcode clusters, done here through elbow analysis, can provide an appropriate single threshold for the taxon as a whole. By basing this analysis upon similarity-based clustering methods, a threshold can be calculated much more efficiently than by applying coalescence-based methods, which are not presently feasible for extremely large data sets. This graphical method yields clusters similar to Linnaean taxonomy, but more finely sub-divided. This is evident in the high unidirectional Adjusted Wallace coefficient values, whereas Linnaean values are lower compared to molecular values. This

increased resolution by molecular clustering still allows for MOTUs to be linked back to current taxonomy, which is especially important for conservation studies and when screening for species presence/absence as part of invasive species monitoring efforts. Our molecular clustering approaches are partitioning the data into clusters smaller than current morphological species descriptions; for some study aims, a less stringent GPSD threshold may be needed if delimitation into current morphological species groups is desired as opposed to expected diversity of evolutionarily distinct species based on molecular data.

The similarity-based molecular methods yielded similar results to GMYC, in which a threshold is sought which explicitly divides nodes into those corresponding with Yule (inter-species) vs. coalescent (within-species) evolutionary processes. Therefore, both of these categories of methods may yield groupings that are closer to evolutionary species, a species concept which is increasingly favoured by several authors (e.g. Mayden 1997; de Queiroz 2007), than they are to morphological species. While MOTUs may represent evolutionarily distinct but in some cases morphologically similar species, further study based on additional molecular, morphological, and biological data is generally considered necessary before formal taxonomic revisions are supported (Collins & Cruickshank 2013). Researchers may also disagree about the biological meaning of divergent allopatric mitochondrial lineages within single Linnaean species, depending upon the preferred species concept. While a demonstration of reproductive isolation in sympatry may be required to meet species status under the biological species concept, separately evolving, genetically divergent populations that are currently on different evolutionary trajectories might be recognized as species under an evolutionary species concept (Mayden 1997). Nevertheless, our results support the use of a single molecular marker for rapid species delineation and for indicating likely cryptic species for further exploration. We sampled up to 34 individuals for those Linnaean species which were split into two or more MOTU here. As these exhibited up to 10% sequence divergence, we suggest that further study is warranted as some of these may meet the criterion for species status under biological or evolutionary species concepts.

We have presented results using limited specimen collections (Table 2.1) of Hexanauplia, and there could exist differing levels of molecular variation for sub groups within this focal taxon. However, this was not tested here as we were interested in a possible GPSD threshold which would partition the data into species-like groups across a large taxonomic data set.

Although there may be some taxonomic bias present in our data sets, there is a large phylogenetic breadth, particularly for Calanoida and Sessilia. Given the diverse data set, we expect that the global threshold will also work well across other poorly sampled groups across these classes, although this supposition would benefit from directed testing in future studies. Having evidence to support a single threshold for larger taxonomic breadth is important, as this is very desirable for analyses based upon high-throughput sequencing of mixed-species samples.

It is also important to note that there could be noise in our results potentially due to misidentified specimens. It is likely that some misidentifications are present in our data sets. This is especially true of the *reference* data set obtained from the publically accessible BOLD data base, originating from a large diversity of source studies. It is unlikely that such misidentifications would greatly influence the outcome of our elbow analysis, as this analysis only relied upon class-level identifications. Additionally, any major misidentifications or contamination events across distantly-related taxa would yield large divergence from other sequences, not impacting the inflection point in the elbow. However, the presence of misidentifications can more significantly impact our concordance and agreement analyses. This issue may explain why we observed a higher proportion of "lumps" in the *reference* dataset compared to our *novel* dataset. Misidentifications present in the *novel* data set would again add error to our agreement analysis comparing molecular clustering results with morphology-based identifications. Although not completed here due to limited resources, it would be beneficial to verify the accuracy of our morphological identifications. This could be accomplished by using the remaining batch vouchers as the basis for multiple sets of independent identifications by different investigators, followed by DNA barcoding of these specimens.

Current suggestions in the literature indicate the use of a Maxillopoda threshold ranging from 2-3% pairwise divergence (Radulovici *et al.* 2010; Blanco-Bercial *et al.* 2014). Here, the total range of determined thresholds for both data sets across similarity-based methods was fairly small, between 2.1 and 2.6% GPSD (p-distance) depending upon the method, and consistent with previous reports. This range, which was obtained using the same data set with each algorithm, shows how the choice of clustering analysis can impact the resulting clusters. Interestingly, the BIN 2.2% seed threshold—which was calibrated against morphological species using select groups of taxa: bees, butterflies and moths, fish, and birds (Ratnasingham & Hebert 2013)— is within the range of global thresholds calculated here. Results presented here show that GPSD

thresholds are relatively consistent across the four tested similarity-based analyses and that the following GSPD thresholds should be used for copepods and thecostracans (ABGD = 2.05%; jMOTU= 2.35%; Mothur= 2.6%; UPARSE= 2.2%).

Similarity-based vs. coalescent-based methods: Performance and feasibility

Similarity-based delimitation methods have certain advantages over coalescence-based methods: speed, simplicity to implement, and ability to accommodate large data sets. Among the taxa investigated here, the similarity-based delimitation methods (ABGD, jMOTU, Mothur, UPARSE, and BINs) displayed a higher concordance compared to morphology than did coalescence-based methods (PTP and GMYC). Specifically, BINs exhibited more direct matches to Linnaean names than PTP and GMYC results. This slightly better performance using the BIN method than the other clustering methods, when compared to current taxonomy, has also been reported in fish, birds, and two moth groups (Ratnasingham & Hebert 2013).

Ratnasingham & Hebert (2013) report an overall higher percentage of exact cluster matches between BINs and Linnaean species than we report (Ratnasingham & Hebert 2013: 83-97%, here: 40-80%). This difference is most likely indicating a lower correspondence in Hexanauplia between named species and evolutionary species—those currently on separate evolutionary trajectories as indicated by genetic separation, whether allopatric or not. This could also be due to the presence of more readily discernible diagnostic morphological characters in those taxonomic groups or a higher proportion of species pairs in Hexanauplia that are only distinguishable by chemical, ecological, and/or behavioural traits (e.g. see Knowlton 1993).

There were two pairs of taxa exhibiting mixing in our *novel* data set *(Acartia hudsonica* with *A.longiremis, Centropages abdominalis* with *C. hamatus*). Although there has been no study investigating barcode variation between *A. hudsonica* and *A. longiremis*, research using morphological and molecular evidence has shown close relationships among other *Acartia* species and high barcode variability within single *Acartia* species (da Costa *et al.* 2011; Blanco-Bercial *et al.* 2014). As with *Acartia*, the genus *Centropages* has also been noted as having discordant molecular clustering as compared to morphological identifications (Blanco-Bercial *et al.* 2014). In addition, species within the Centropagidae family have been noted as having a plastic response to differing environments, thereby making morphological identifications more

difficult (Blanco-Bercial *et al.* 2014). Although beyond the scope of this work, future efforts investigating the specific nature of these 'mixing' results are suggested.

Five species were split into two or more molecular groups as compared to taxonomic labels. *Paracalanus parvus* was split in two groups, with both MOTUs having representation in Pacific and Arctic regions. These MOTUs could represent subspecies, five of which are currently recognized (WoRMS Editorial Board 2016, accessed May 24, 2016), and additional sampling to elucidate the presence of subspecies or potential population structure among specimens of a single subspecies is required. *Centropages typicus* was represented by only two specimens, and so each MOTU for the split had only a single representative sequence, and further sampling is required to gain a better picture of the potential intraspecific variation. The remaining three (*Tortanus discaudatus, Eurytemora herdmani*, and *Temora longicornis*) species which displayed molecular splits have no known subspecies. As is the case for the *Centropages typicus* and *Paracalanus parvus*, additional collections are necessary to understand the geographic distributions and taxonomic status of the multiple MOTU detected within single Linnaean species.

Lower direct matches between Linnaean names and MOTU in copepods and thecostracans compared to several better-studied taxa may also be exacerbated by evolutionary processes that are largely unique to the marine realm, such as the accumulation of high levels of intraspecific diversity in extremely large populations with large geographic ranges. Due to this effect, combined with some cases of true recent speciation, Meyer & Paulay (2005) predicted greater overlap between intraspecific and interspecific divergences in marine environments as compared to terrestrial environments. Further investigation into this potential difference in patterns of sequence variation between marine and terrestrial systems is warranted and should include a broad variety of marine taxa. The optimal method for delimiting marine species may vary depending on the scientific questions and species concept. While in some instances the method of delimitation may be important, in others there may be little variation in the species counts across methods. This limited variation can be seen in our results where GMYC, BIN, jMOTU, Mothur, and UPARSE had very close concordance values, varying between approximately 0.88 and 1.0. This finding suggests that more rapid similarity-based methods are expected to yield groupings that largely correspond to the GMYC method, which is often favoured for its explicit evolutionary model. Consideration for the study question is also

important; for example, a more discriminant analysis option or threshold may be preferable for detecting invasive species or endangered species in a given habitat.

Quantifying concordance as an alternative to concluding failure/success of barcoding

Assessing "true" species boundaries—especially in a geographically widespread and taxonomically diverse group such as Hexanauplia —is difficult, and boundaries can differ according to the preferred primary species concept and the selected character system (Mayden 1997; de Queiroz 2007). By quantifying concordance, we can examine and compare the signal for various delineations emerging from different character types and analysis methods. This bi-directional concordance assessment, not previously used in the barcoding literature, provides more information than simply reporting failures when molecular clusters do not agree with morphological species. This extra information can support existing morphological species boundaries as evolutionary species, through a new character system, or provide new biological insights (e.g. into potential cryptic species prevalence) in cases of discordance.

Adjusted Wallace concordance values indicated that the molecular clusterings for the *novel* data set were more discriminant than current Linnaean species across all clustering analyses. This is not surprising as the low overall knowledge of the total biodiversity in Copepoda has been noted in the literature (Bucklin *et al.* 2011; Blanco-Bercial *et al.* 2014). Moreover, although morphological identifications for the *novel* data set were conducted by trained experts, some of the observed cases of discordance may be due to incorrect identification of specimens, due to the difficult taxonomy because juveniles and larvae are often damaged after preservation or have undeveloped diagnostic morphological characters. Using the *novel* data set, there was variation in clustering outcomes between coalescence-based and similarity-based analyses. Linnaean species better explain the similarity-based molecular clusters but are less concordant with the coalescent-based clusters. These results reflect some variation in the MOTUs generated using coalescence versus similarity-based analyses, with the latter generating groupings that can be somewhat better predicted by current species identifications. However, with the exception of PTP and ABGD, the differences among all molecular methods in their concordance to current taxonomy were modest. Thus, the molecular data appear to be revealing biological variation that was previously unrecognized in the current taxonomic hierarchy, which may be reflective of evolutionary species diversity.

Consistency among molecular delimitation methods

Concordance values provide an overall description of the consistency of the clustering results among methods; however, this metric tells us little about the specifics of how individuals are partitioned into clusters. If we first consider the *novel* data, when the clustering method results in numerous MOTUs there are also fewer exact matches between molecular groupings and Linnaean species. Although the variation in the number of clusters did appear to influence the proportion of matches to current species, the number of matches for all analyses was relatively similar, with between 60-70% exact matches. ABGD, with the fewest total clusters, had the highest number of exact matches among all other methods for the *novel* data set, with no splits. The choice of the p-distance when conducting analyses using the ABGD program has been noted to provide differing results as opposed to other distance metrics (Kekkonen *et al.* 2015). The p-distance was necessary for this work to provide cross comparison among clustering methods. When I conducted an earlier ABGD analysis using K2P distances, there appeared to be little difference between results obtained using p-distance (results not shown). However, a more thorough evaluation of the variation in these analyses could be conducted in the future. Also, it is important to note that with a larger gap parameter setting than used here (such as the default) and recursive partitioning, the ABGD program performs an exploration of the variation in the data to create molecular groupings. As such, we found default parameters had a tendency to lump the data into larger groups (results not shown). This may have been because our marine focal taxon exhibits more continuity in the range of genetic divergences than many taxa previously tested. Instead, we used a small gap parameter to force a threshold upon ABGD and then obtain a specific global threshold using the elbow analysis approach. Therefore, ABGD may behave differently from here in future applications, depending upon the settings selected.

The remaining analyses (BIN, GMYC, jMOTU, Mothur, UPARSE) had very consistent agreement when considering both total matches and splits together (Figure 2.5). These consistent results indicate the clustering methods, whether resulting in a match or split, can accurately place specimens into Linnaean species using these methods in approximately 70-80% of cases. If we were to remove taxa showing highly unstable correspondence between molecular groupings and current species, which are likely in need of taxonomic revision (*Acartia hudsonica, A. longiremis, Centropages abdominalis, C. hamatus, Eurytemora herdmani, Temora longicornis*),

then the number of exact matches to current species for our methods would increase to approximately 78 and 91%, similar to barcode agreements reported in other taxonomic groups (Ratnasingham & Hebert 2013; Blanco-Bercial *et al.* 2014).

Conclusion

We conclude that when applying GPSD thresholds that the method selected for MOTU generation is important; our results indicate a need for careful selection of both the method of generating MOTU clusters and the threshold applied. Our study has also found a larger number of MOTUs generated as compared to morphological species labels. These data indicate either poor taxonomic identification in the databases, the presence of cryptic species, and/or evidence of substantial intraspecific diversity at the COI-5P gene region. Continued research is needed to quantify to what extent these MOTUs represent real biological entities under an explicit species concept of interest. Future work may also include the amplification of additional molecular markers (particularly from the nuclear genome) to verify the taxonomic placement of specimens and lend support toward the identified GPSD thresholds for the COI-5P molecular region. This work may be especially useful for specimens identified here as problematic taxa. In addition to further research on species boundaries in taxonomically problematic groups, further research is also needed towards protocol development for groups with low amplification and sequencing success. This may include primer design and investigation of the importance of specimen fixation in cold conditions immediately following field collection (Prosser *et al.* 2013). Finally, the adoption of internal methods for clustering verification, such as the analyses presented here, is encouraged in DNA barcode studies to enable rapid biodiversity study and exploration of unknown faunas.

Tables and Figures

Table 2.1 Taxonomic breakdown of the two data sets used during analysis. **A.** *reference* data set, **B.** *novel* data set. 'Sequences N' is the total number of sequences. The 'Species N' column provides the total number of named Linnaean species for each order for the two data sets. The 'BIN N' column provides the total number of BINs present for each order in the data set. The 'Clustering Outcome' column in part B indicates the results from the analyses of this work. Finally, the 'Subspecies' column indicates described, valid subspecies for the corresponding species (Boxshall *et al.* 2016).

A.

Sub class	Order	Sequences N	Species N	BIN N
Copepoda	Calanoida	1251	141	285
Copepoda	Cyclopoida	317	32	119
Copepoda	Harpacticoida	99	12	64
Copepoda	Poecilostomatoida	140	13	74
Copepoda	Siphonostomatoida	402	19	51
Thecostraca	Akentrogonida	5	0	2
Thecostraca	Ibliformes	2	1	2
Thecostraca	Kentrogonida	1	1	1
Thecostraca	Lepadiformes	24	11	18
Thecostraca	Lithoglyptida	3	0	1
Thecostraca	Scalpelliformes	38	4	5
Thecostraca	Sessilia	545	75	164
	Total	**2825**	**309**	**786**

B.

Order and Species	Seq N	BIN N	Clustering Outcome	Subspecies
Calanoida	211	30		
Acartia clausi	1	1	Single sequence, only match possible	*Acartia clausi gaboonensis* Scott T., 1894
Acartia hudsonica	15	4	Mix for all methods with *Acartia longiremis*	
Acartia longiremis	54	2	Mix for all methods with *Acartia hudsonica*	*Acartia longiremis spiniremis* Pinhey, 1926
Aetideus divergens	4	1	Match for all methods	
Calanus glacialis	9	1	Split for PTP, Match for all other methods	
Calanus hyperboreus	1	1	Single sequence, only match possible	
Calanus pacificus	10	1	Split for PTP, Match for all other methods	*Calanus pacificus californicus* Brodsky, 1965; *Calanus pacificus japonicus* Brodsky, 1959; *Calanus pacificus oceanicus* Brodsky, 1959
Centropages abdominalis	18	2	Mix for all methods with *Centropages hamatus*	
Centropages hamatus	10	3	Mix for all methods with *Centropages abdominalis*	
Centropages typicus	2	2	Either Match or Split outcome across methods	*Centropages typicus aucklandicus* Kramer, 1895
Epilabidocera amphitrites	1	1	Single sequence, only match possible	
Eurytemora herdmani	16	2	Either Match or Split outcome across methods	
Leptodiaptomus siciloides	4	1	Match for all methods	
Mesocalanus tenuicornis	3	1	Match for all methods	
Metridia pacifica	3	1	Match for all methods	
Paracalanus parvus	12	2	Either Match or Split outcome across methods	*Paracalanus parvus borealis* Wolfenden, 1906; *Paracalanus parvus indicus* Wolfenden, 1906; *Paracalanus parvus major* Tanaka, 1956; *Paracalanus parvus minor* Tanaka, 1956; *Paracalanus parvus parvus* Claus, 1863
Pseudocalanus minutus	1	1	Single sequence, only match possible	*Pseudocalanus minutus elongatus* Boeck, 1865; *Pseudocalanus minutus gracilis* Farran & Vervoort, 1951

Skistodiaptomus pallidus	2	1	Match for all methods	
Temora longicornis	11	2	Either Match or Split outcome across methods	
Tortanus discaudatus	34	3	ABGD Match, split for all other methods	
Cyclopoida	2	2		
Oithona atlantica	1	1	Single sequence, only match possible	
Oithona similis	1	1	Single sequence, only match possible	
Harpacticoida	17	3		
Microsetella norvegica	14	1	Split for PTP, Match for all other methods	
Tisbe furcata	1	1	Single sequence, only match possible	*Tisbe furcata furcata* Baird, 1837; *Tisbe furcata johnsoni* Monk, 1941; *Tisbe furcata tuberculata* Chislenko, 1971
Zaus abbreviatus	2	2	BINs split, match for all other methods	
Monstrilloida	1	1		
Monstrilla scotti	1	1	Single sequence, only match possible	
Poecilostomatoida	1	1		
Corycaeus anglicus	1	1	Single sequence, only match possible	
Sessilia	15	2		
Balanus balanus	12	1	Split for PTP, Match for all other methods	
Balanus glandula	3	1	Match for all methods	
Total	**247**	**43**		

Table 2.2 Adjusted Wallace coefficient (Wallace 1983) concordance values for the clustering results obtained from four programs for the *reference* data set. Thresholds applied to determine clusters are indicated in the row and column labels and were obtained for each analysis method via elbow analysis (Figure 2.3). Values in parentheses indicate the total number of MOTUs as determined by the corresponding analysis and threshold. Each value in the table indicates how well the clusters generated by the method indicated by the row label correspond to the clusters yielded by the method indicated in the column label. Each pair of methods is represented by two values in the table.

	ABGD 2.1% (759)	**jMOTU 2.3% (862)**	**Mothur 2.6% (856)**	**UPARSE 2.2% (878)**
ABGD 2.1%	--	0.73	0.72	0.65
jMOTU 2.3%	1	--	0.99	0.89
Mothur 2.6%	1	1	--	0.89
UPARSE 2.2%	1	1	1	--

Table 2.3 Bidirectional concordance among clustering methods for the *novel* data set using Adjusted Wallace's coefficients (Wallace 1983). Values in parentheses indicate the total number of clusters generated for each analysis. The global pairwise sequence divergence (GPSD) thresholds for ABGD, jMOTU, Mothur, and UPARSE are those obtained via elbow analysis using the *reference* data set. Each value in the table indicates how well the clusters generated by the method indicated by the row label correspond to the clusters yielded by the method indicated in the column label. Each pair of methods is represented by two values in the table.

	Morphology (29)	**BINs (40)**	**PTP-ML (51)**	**GMYC (37)**	**ABGD 2.1 (30)**	**jMOTU 2.3 (40)**	**Mothur 2.6 (39)**	**UPARSE 2.2 (39)**
Morphology (29)	--	0.562	0.46	0.52	0.704	0.563	0.569	0.559
BINs (40)	0.925	--	0.782	0.881	1	1	1	0.994
PTP-ML (51)	0.923	0.953	--	0.964	0.999	0.954	0.967	0.954
GMYC (37)	0.929	0.956	0.859	--	0.994	0.957	0.969	0.951
ABGD 2.1 (30)	0.938	0.81	0.664	0.742	--	0.81	0.819	0.806
jMOTU 2.3 (40)	0.926	0.999	0.783	0.881	1	--	1	0.994
Mothur 2.6 (39)	0.926	0.988	0.785	0.882	1	0.989	--	0.983
UPARSE 2.2 (39)	0.925	0.999	0.787	0.88	1	1	1	--

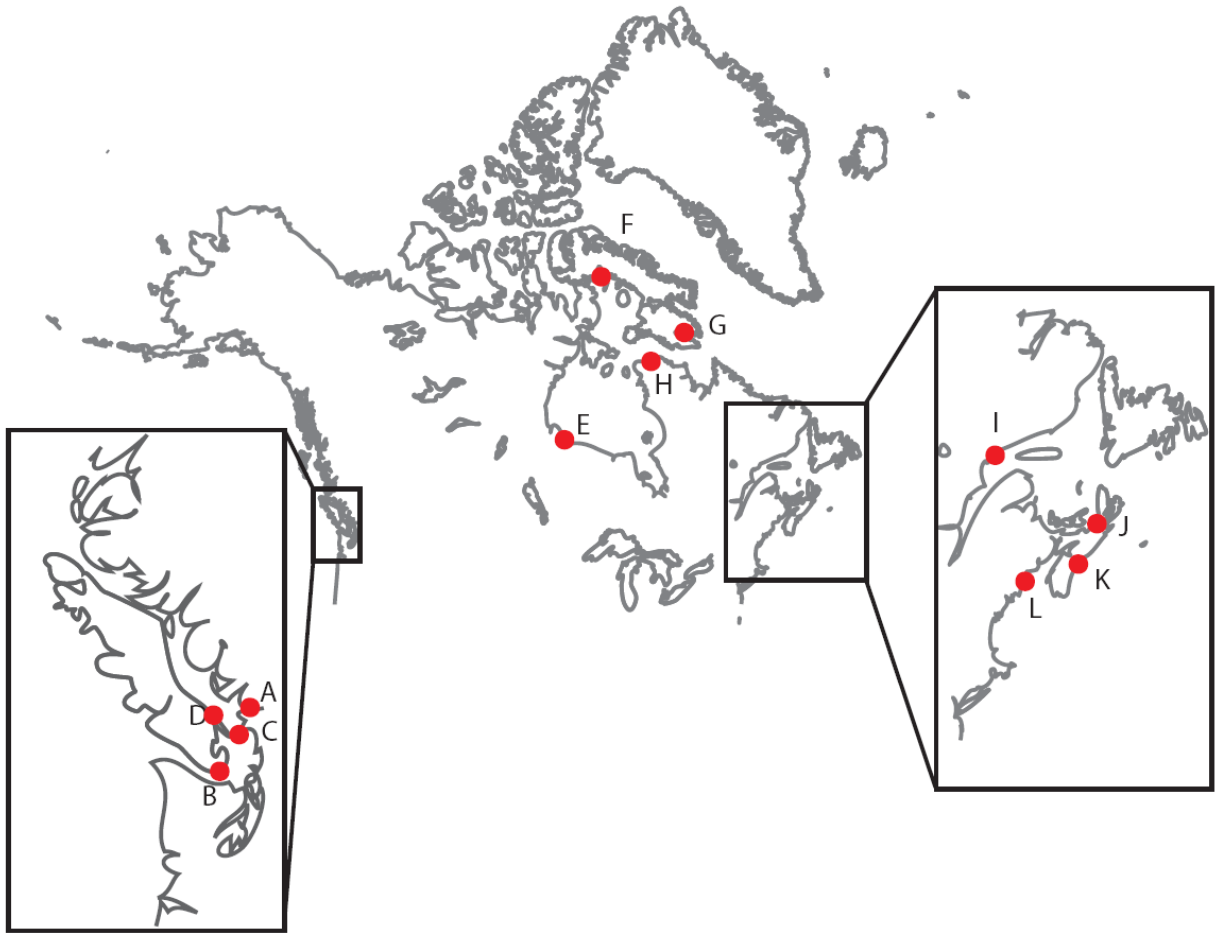


Figure 2.1 Map of Canada with plankton sampling sites indicated by red circles. Sites include: A - Vancouver, B - Victoria, C - Roberts Bank, D - Nanaimo, E - Churchill, F - Steensby Inlet, G - Iqaluit, H - Deception Bay, I - Baie de Sept-Iles, J - Port Hawksbury, K - Bedford Basin, L - Bayside.

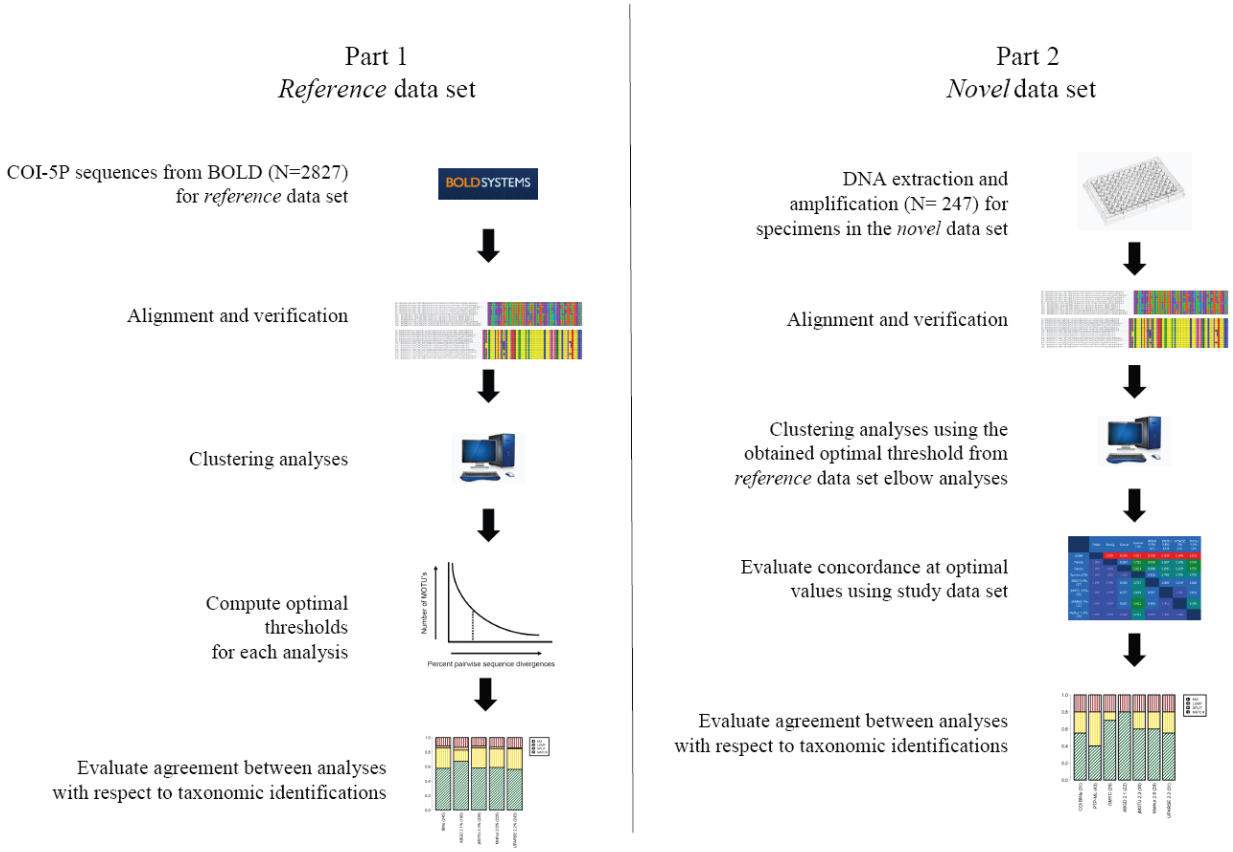
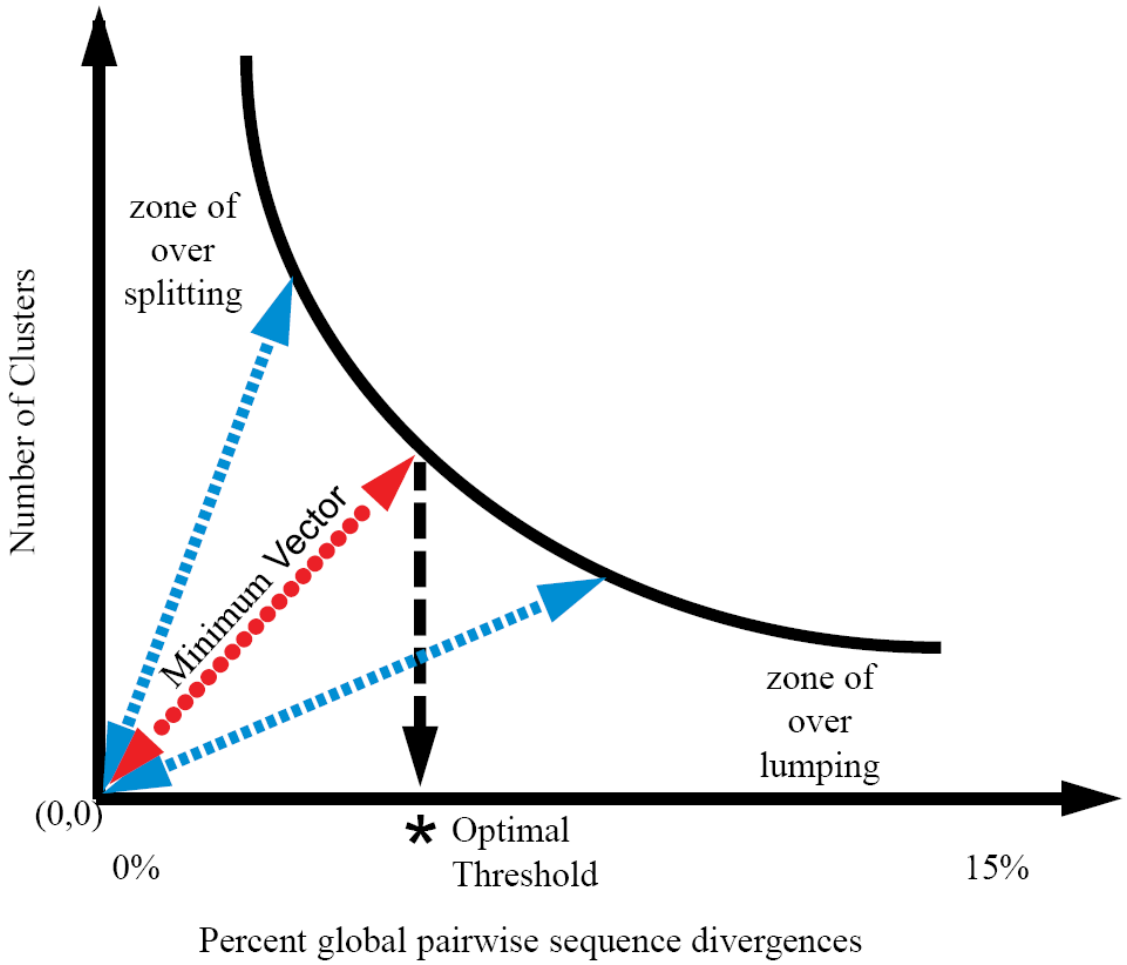


Figure 2.2 Flow chart for the analysis of the two study data sets used in this study. Part 1 shows the key steps used to analyze the *reference* data set. Clustering analyses included the use of 4 programs: ABGD, jMOTU, Mothur, and UPARSE. In addition to these four methods, BOLD BIN assignments were used to evaluate agreement to taxonomic identifications. Part 2 shows the steps in the analysis of the *novel* data set. The *novel* data set was clustered into MOTUs using the same four similarity-based analyses as for the *reference* data set (ABGD, jMOTU, Mothur, UPARSE). In addition to these, BINs and two coalescent (GMYC and PTP) clustering methods were used, and agreement to taxonomic identifications was quantified.



F

Figure 2.3 Conceptual diagram for the determination of the optimal molecular divergence threshold values. The vertical long-dashed black line indicates the point on the curve (elbow) representing a threshold value that does not over-split or over-lump the sequences into MOTUs. This point represents the closest distance to the origin (0,0) (red circular-dashed arrow), as contrasted with larger vectors (blue small-dotted arrows). The corresponding point on the x-axis indicates the value for the percent pairwise divergence representing the proposed optimal threshold for given data set using the graphed analysis method.

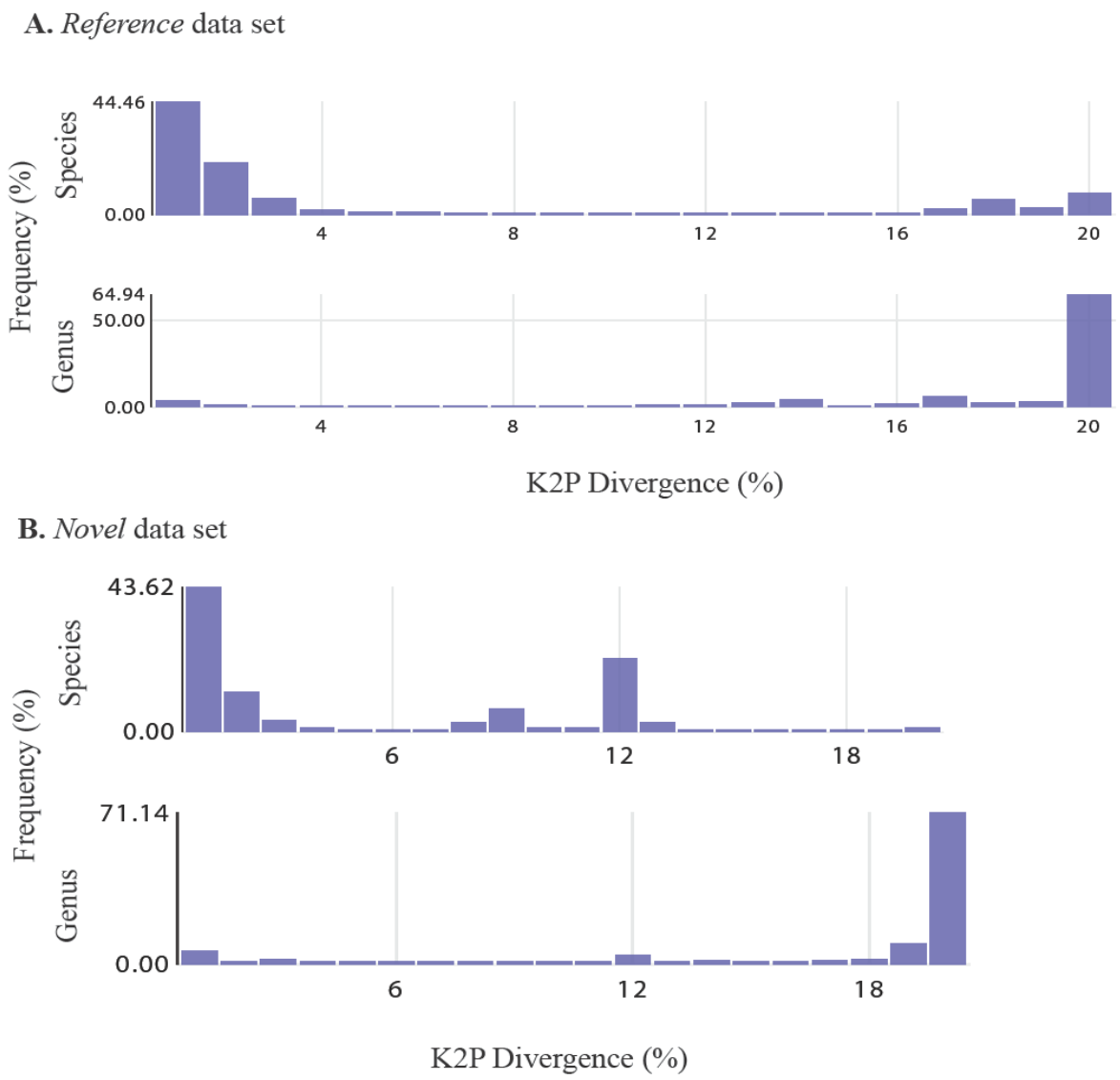


Figure 2.4 Histograms displaying the intraspecific and interspecific K2P pairwise sequence divergences for the A) *reference* and B) *novel* data sets (see Table 2.1 for information on the composition).

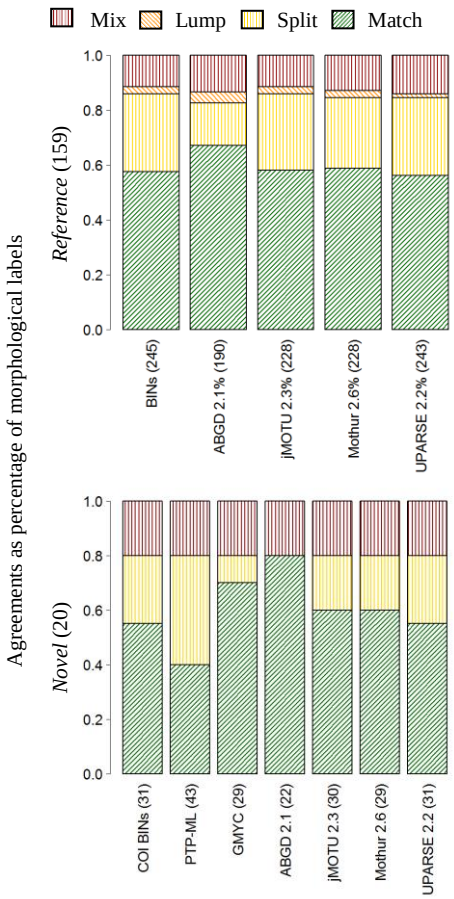


Figure 2.5 Agreement between morphologically-grouped specimens based on Linnaean species labels and clusters generated using molecular methods for the *reference* and *novel* data sets. The sample size of species included in each analysis is indicated in parentheses for each data set. The numbers of MOTUs generated are also indicated in parentheses after each analysis method.

Chapter 3

Molecular species delimitation in marine planktonic Multicrustacea: Testing the utility of the 18S-V4 molecular region using COI-5P barcodes and morphological identifications

Abstract

There remain significant gaps in our understanding of biodiversity, especially in the marine realm that is biogeographically complex, under-sampled relative to terrestrial systems, and taxonomically-confounded for many species groups. However, molecular tools are providing unique opportunities to better characterize and understand biological diversity. For example, the 18S gene has been used to investigate biological diversity across a wide range of taxa and is increasingly included in metabarcoding studies using high-throughput sequencing. However, little is known about its ability to delimit diverse collections of specimens into taxonomic groups. Here, using marine zooplankton samples, I test clusterings (i.e. Molecular Operational Taxonomic Units, MOTUs) generated using the 18S-V4 region against morphological identifications and COI-5P molecular clusters. Several methods often used in molecular biological research were used to cluster data (BIN, ABGD, jMOTU, UPARSE, Mothur, PTP, and GMYC) into hypothesized evolutionary genetic species. There was substantial variation in sequencing success of taxonomically-named specimens, with 81% of specimens yielding sequences for 18S-V4 but just 59% for COI-5P. Clustering results using the 18S-V4 region were found to underestimate biological diversity by 5-15% compared to COI-5P clusters. Results presented here indicate that the COI-5P is the best choice for use for species-level molecular delimitation in single-marker marine plankton research; however, the low COI-5P sequencing success remains a problem to be resolved. Until such a time, 18S-V4 may serve as an effective complementary molecular region, recognizing its limitations for species delimitation.

Introduction

The level of biodiversity in the marine realm is not well understood (Archambault *et al.* 2010; Bucklin *et al.* 2011). Increased anthropogenic pressures—including transportation of invasive species, climate change, pollution, and the physical disturbance of marine ecosystems—have the potential to increase the rate of local and global extinctions (Singh 2002). With rising rates of extinction, there is an urgent need for increased study of marine biodiversity before it is lost. Traditional morphological approaches to characterizing marine diversity are currently acting as a bottleneck due to costs and lack of taxonomic experts for many groups. Rare and difficult-to-identify biota is often missed when conducting traditional morphological marine surveys. This is particularly true in marine plankton where recent studies have highlighted the suspected high numbers of species present using high-throughput sequencing techniques (Zhan *et al.* 2013; Jensen *et al.* 2014; Zhan & MacIsaac 2015). A concern when assessing marine diversity using high-throughput sequencing methods is in separating the sequence data, which may include real, but rare species, from sequencing artefacts. One way to reduce the unknown number of sequences is to continue to populate DNA barcode reference libraries.

Molecular studies, using techniques such as DNA barcoding (Hebert *et al.* 2003a, b), are necessary, due to their speed compared to morphological approaches. These studies can result in a better understanding of marine biodiversity, thereby contributing to conservation and resource-management decisions (Dettai *et al.* 2011; Hubert & Hanner 2015). DNA barcodes can be applied to biodiversity studies for two distinct purposes (Collins & Cruickshank 2013). The first is to use DNA barcodes to identify a specimen to a species by matching the barcode to a sequence database with independent species-level identifications. The second is using DNA barcodes to cluster specimens into species-like groups, an approach that is especially useful for discovery in understudied biomes and taxa. Given challenges in obtaining COI-5P barcodes for some marine groups, and given increased attention and affordability of multi-marker metabarcoding approaches (Bucklin *et al.* 2011; Cristescu 2014), there remains a need to assess the utility of alternative or complementary molecular markers for the study of marine species diversity.

Marine DNA barcoding using the animal barcode region, a portion near the 5’ end of the cytochrome *c* oxidase subunit I gene (COI-5P), has been well supported across a large diversity

of marine animal life (Hebert *et al.* 2003a, b; Costa *et al.* 2007; Ivanova *et al.* 2007; Ward *et al.* 2009; Bucklin *et al.* 2010b; Radulovici *et al.* 2010; Carr *et al.* 2011; Blanco-Bercial *et al.* 2014; Layton *et al.* 2014; Raupach *et al.* 2015). DNA barcoding using the COI-5P molecular region works well because of its high interspecific and lower intraspecific divergences (Meyer & Paulay 2005; Hubert & Hanner 2015). In addition to molecular divergence patterns, several other elements contribute to the value of the COI-5P region as a molecular barcode, including its lack of introns, high number of mitochondria per cell, and its nearly universal uniparental maternal inheritance (Breton & Stewart 2015). These features facilitate sequence recovery as well as the ability to separate many recently-diverged species. Although these are favourable properties for a molecular barcode, the challenge for many marine groups is to effectively obtain the COI-5P molecular data. Difficult COI-5P amplifications are noted among marine planktonic crustacean taxa (Bucklin *et al.* 2010a, 2011; Lobo *et al.* 2013; Chapter 2), where numerous PCR attempts and primers are often needed before successful amplification is achieved, if at all. Another challenge is DNA-friendly preservation of specimens under difficult marine sampling conditions; rapid and high-quality fixation, such as using both ethanol and rapid cold storage, is needed to increase molecular sequencing success rates for aquatic invertebrates (Prosser *et al.* 2013).

Various ribosomal RNA genes have previously been used for molecular systematic analyses, including regions of both the 16S and 18S rRNA genes (Lefébure *et al.* 2006; Bucklin *et al.* 2011; Hubert & Hanner 2015; Kumar *et al.* 2015). Marine researchers have demonstrated that these molecular regions can place specimens into groups concordant with higher-level taxonomy (see Lefébure *et al.* 2006 - crustaceans using regions of 16S; Wu *et al.* 2015 - copepods using 18S; Zimmerman *et al.* 2011 - diatoms using regions of 18S). Tang *et al.* (2012) investigated the utility of the 18S gene for species delimitation of marine meiofauna, using all available data, with sequence lengths ranging from 319 to over 5000 nucleotides. They concluded that the 18S gene underestimated diversity by 60% compared to taxonomy. However, this work did not specifically focus in on the ability of the 18S-V4 region to delimit species of Hexanauplia, and they did not attempt to refine the threshold by which the data set was clustered.

The 18S-V4 region has been identified as having high levels of interspecific distance among congenerics and having a higher probability of being able to delimit species compared to other 18S regions (Wu *et al.* 2015). Some studies applying high-throughput sequencing techniques have used the 18S-V4 region toward large-scale biodiversity surveys for aquatic life

(Zhan *et al.* 2013; Brown *et al.* 2015; Flynn *et al.* 2015). However, there remains a research gap where complementary data sets—using multiple molecular markers combined with traditional morphological taxonomic identifications—are used to validate if this 18S region can be used effectively for species delimitation.

Here, I test the 18S-V4 region across a geographically diverse collection of marine planktonic crustaceans against morphological identifications and COI-5P sequence clusters. Past research has investigated the utility of the 18S gene in delimiting specimens into species-like groups in copepods and has indicated that the fourth variable region of the 18S gene (18S-V4) is the best region, as it provides the most information for separating species (Wu *et al.* 2015). Nevertheless, there remains a gap in understanding if the variability in the 18S-V4 molecular region is suitable for large-scale species delimitation; specifically, there is a need to better characterize intraspecific variability using geographically large-scale sampling for marine planktonic crustaceans. Moreover, given the prevalence of morphologically cryptic species (Appeltans *et al.* 2012), there is a need to compare delimitations based upon 18S-V4 with both morphology and the standard animal barcode region (COI-5P), which has a strong capacity to separate evolutionarily-isolated species (Bucklin *et al.* 2011). Here, I use morphological identifications and COI-5P clusters to identify likely species groupings and compare 18S-V4 data with respect to these groupings. Using this information, I evaluate whether 18S-V4 sequences can be clustered into species-like groups using various commonly-used software programs and draw conclusions regarding the potential utility and limitations of the 18S-V4 region for use in molecular biodiversity studies with planktonic crustaceans.

Methods

Field and molecular laboratory methods

Plankton samples were collected between May 15, 2011 and August 19, 2012 at eleven port locations as well as one proposed port location representing all three of Canada's ocean regions (Arctic, Atlantic, and Pacific) (Figure 2.1). Field collection, preservation and storage, and specimen selection followed protocols in Chapter 2. All specimen data and associated identifiers (BOLD and GenBank accession numbers) can be found as supplemental files in the appendix (Appendix IV). COI-5P data are from Chapter 2, while 18S sequences are newly

presented in this chapter. DNA extraction was completed on single whole individuals, as all individuals were no greater than approximately 1 mm^3. These individuals were consumed by the DNA extraction procedure and are referenced to batch vouchers. DNA lysis and extraction procedures followed the protocols from Chapter 2. Amplification of 18S-V4 used the primer set Uni18S and Uni18SR (Zhan *et al.* 2013). All sequence reactions contained: 2 mM $MgCl_2$; 0.2 mM dNTP's; 0.4 µM forward and reverse primer; 1X PCR buffer; 10% of DNA template by volume; 0.0064 Units of Taq/µl. PCR reaction conditions consisted of an initial denaturation step at 95 °C for 5 min, followed by ten cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 90 s; ten cycles of 95 °C for 30 s, 53 °C for 30 s, 72 °C for 90 s; twenty cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 90 s; and a final elongation step at 72 °C for 10 min. PCR amplifications were visualized, and clear single bands were deemed successes, cleaned using an EXOSap-IT clean-up protocol, and sequenced bi-directionally (Chapter 2).

The qualities of the resulting DNA sequence chromatograms were evaluated using CodonCode Aligner (CodonCode Corporation). Chromatograms which contained fewer than 25 called bases or had fewer than 50 called bases with a Phred value of 20 or less were removed from the data set. The end clipping function was applied at the start of the sequences, and sequences were trimmed until they contained fewer than 6 bases out of a 50-base window with a Phred score of less than 20. CodonCode was then used to combine the forward and reverse sequences into a single contiguous sequence for each specimen.

COI-5P sequences were aligned using the FFT-NS-2 alignment strategy in the multiple alignment program MAFFT (Ver. 7) (Katoh & Standley 2013). The multiple sequence alignment (MSA) was trimmed to a final length of 588 base pairs (bp). MEGA6 was used to translate the MSA using the invertebrate mitochondrial code to verify the alignment. Single nucleotide insertions or deletions evident through frame shifts were investigated and edited if base calling was incorrect. Cases of unresolved frame shifts or stop codons resulted in the removal of the sequence from the data set, as this pattern suggests that the sequence is of a nuclear pseudogene.

18S sequences were aligned using the SSU-ALIGN (Nawrocki 2009) suite of programs, which employs an alignment strategy using the secondary structure of eukaryote RNA. Different secondary structure models (archaeal, bacterial, or eukaryotic) were used in conjunction with a profile Hidden Markov Model to score submitted sequences and assign them to their appropriate domains. This first step ensured amplified sequences were of the target domain before continuing

further analyses. Primer sequences were removed from the resulting MSA, which was then further trimmed to a length of 980 bp. This further trimming was completed to provide reliable sequence blocks at the start and end regions with few unknown nucleotides. The resulting MSA was visually inspected using MEGA6 (Tamura *et al.* 2013). Sequences of less than half of the alignment length (including internal gap locations) following primer trimming were removed from the data set. 18S-V4 neighbour-joining (NJ) phenograms (Saitou & Nei 1987) using Kimura-2-parameter (K2P) distances (Kimura 1980) were constructed through MEGA6, and node support was estimated using 1,000 bootstrap pseudoreplicates.

The resulting phenograms were inspected for potential contaminants, with two possible scenarios resulting in the removal of sequence data: (1) a BLAST of the NCBI database resulted in a 100% query coverage and 0 e-value result for a named entry belonging to something other than the expected class-level identification based on morphology (this was verified by examining the second and third matches, to guard against cases of matching to a single erroneous sequence in the database); (2) the BLAST result was a closer match to specimens of a different family, as compared to morphology, indicative of 18S-V4 contamination. Specimens belonging to the same genus or species based on COI-5P clusters and morphological identification by the same identifier and from the same sample, but which grouped more closely with another genus on the NJ phenogram, were flagged as potential contaminants. However, these were saved for further comparisons using total evidence, as such cases could indicate hybridization or incomplete lineage sorting.

Data sets

Sequence data used to generate the *COI* (337 sequences representing 29 Linnaean species names, with two species having only a single representative) and *COI reference* (2825 sequences representing 262 Linnaean species names) data sets were obtained from Chapter 2. Two additional data sets were compiled for the 18S-V4 molecular region and are called the *18S reference* and the *18S* data sets. The *18S* data set contains 18S-V4 sequences which were obtained here and consists of the same 337 specimens as in the *COI* data set. The *18S reference* data set was populated with sequences obtained using the taxonomic browser and the publicly accessible database on the SILVA website (http://www.arb-silva.de/browser/ search for Maxillopoda on January 28, 2015). A total list of species present for each of the data sets can be

found in Appendix V. “Maxillopoda”, although now regarded as an obsolete name, was searched as the class in this database. This taxon was recently revised (Regier *et al.* 2010; Oakley *et al.* 2013) but previously contained all Copepoda and Thecostraca (Newman 1992). Sequences containing the terms "environmental", "uncultured", or "metagenome" in the fasta header were not used for further analysis. Sequences were screened for these three key terms to eliminate obvious signs of being obtained from an uncertain origin without morphological data and labelled as Maxillopoda based on molecular data alone. The SSU-ALIGN program was used to align the downloaded data. The *18S reference* MSA was also trimmed to a length of 980 nucleotides of the target V4 region in the same manner as the *18S* data set. This data set consisted of 1047 sequences containing 502 Linnaean species names.

Generating molecular operational taxonomic units

Sequences for all four data sets were clustered into Molecular Operational Taxonomic Units (MOTU) (Blaxter 2004) using four similarity-based methods: Automated Barcode Gap Discovery (ABGD) (Puillandre *et al.* 2012), jMOTU (Jones *et al.* 2011), UPARSE (Edgar 2013), and Mothur (Schloss *et al.* 2009). Commands for the execution of the clustering analyses are included in the appendix (Appendix II). In addition to the four similarity-based methods, the two COI-5P data sets were also clustered using BINs. Two coalescent-based methods were also used for the two smaller data sets (*18S* and *COI*): Poisson Tree Processes (PTP) (Zhang *et al.* 2013) and Generalized Mixed Yule Coalescent (GMYC) (Fujisawa & Barraclough 2013) (Table 3.1). Analyses for the *COI*, *18S*, and *18S reference* data sets are newly presented here, while results were used from Chapter 2 for the *COI reference* data set. All data sets were reduced to unique sequences for MOTU analysis to decrease the computational time of clustering analyses. Removed sequences were re-introduced and assigned to their appropriate MOTU for subsequent analysis. The best-fit model of nucleotide substitution for COI-5P was selected as per Chapter 2. The best-fit model for 18S-V4 was determined with the jModelTest program (Darriba *et al.* 2012) using the *18S reference* data set; the model with the lowest Bayesian information criteria (BIC) was selected. In both cases the best model was the General Time Reversible model with gamma distribution and invariant sites parameters, and so this model was used when generating trees for use in both coalescent-based analyses. All similarity-based MOTU-generating analyses were conducted with p-distance to enable comparisons.

Prior to GMYC and PTP analyses, ultrametric trees were constructed for the *COI* and *18S* data sets with the Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond *et al.* 2012) program using the selected models from model testing, with the Yule process selected and the initial ucld.mean value set between 0 and 10. Seven sequences from the crustacean order Decapoda generated during this study were used as the outgroup (Bron *et al.* 2011). Five independent runs of 200,000,000 Markov chain Monte Carlo (MCMC) generations were performed, sampling every 100,000 generations. Results were inspected using the Tracer program (Rambaut *et al.* 2014), and convergence of all five runs was verified based on visual inspection of the ln likelihood from the MCMC runs. BEAST results were combined using the sub program LogCombiner, with a 10% burn-in value. The combined results were then summarized using TreeAnnotator into a single target tree by finding the best-fit tree using the specified maximum clade credibility.

Metrics for testing for concordance

Two concordance analyses were conducted to assess the congruence between clusters from COI-5P, 18S-V4, and Linnaean species labels: An Adjusted Wallace concordance (Wallace 1983) and an Adjusted Rand (Pinto *et al.* 2007). Both concordance analyses were conducted using the Comparing Partitions online resource http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Home. A two-way Adjusted Wallace concordance was used to understand the directionality of the congruence of two data sets. In other words, how well is method A able to explain the clustering by B, and alternatively how well is B able to describe A? This two-way assessment can provide valuable information on how different character types and analysis methods are able to cluster specimens into species-like groups (Chapter 2). An Adjusted Rand concordance, which provides a single value for the global congruence between two clusterings (Hubert & Arabie 1985), was used to determine the sequence divergence threshold for 18S-V4 data (see below) that yielded the best concordance between clusters generated using different data sets (18S-V4, COI-5P, Linnaean species names).

Optimizing a global sequence divergence threshold through "elbow" and concordance analysis

Two different methods were employed to determine an optimal global pairwise sequence divergence (GPSD) threshold for each clustering analysis method for the *18S* data set. The first

approach, termed "elbow" analysis (see Figure 2.3), used a range (0-15%, stepping in increments of 0.1%) of pairwise divergence thresholds for grouping sequences into MOTUs. This method was used for the *COI reference* and the *18S reference* data sets. This range was chosen as it was expected to capture the threshold at which biologically significant (species-like) groupings would be created using both COI-5P (Blanco-Bercial *et al.* 2014) and 18S-V4 data (Tang *et al.* 2012; Wu *et al.* 2015).

The programs ABGD, jMOTU, Mothur, and UPARSE were used to test the natural vertex point of the pairwise sequence divergence data, as described in Chapter 2. This process involved a graphical approach, with the number of MOTUs plotted on the y-axis, transformed to be equal in length to that of the x-axis, and divergence thresholds, between 0-15%, plotted on the x-axis (Figure 3.1). The point on the curve closest to the origin (0,0) represents the natural vertex point, or elbow, in the data and the optimal threshold (Handl *et al.* 2005). The resulting GPSD on the x-axis, which corresponded to the elbow point, was determined to be the 'elbow-obtained' GPSD threshold.

Elbow-obtained GPSD thresholds were applied to the *18S* data set, and the resulting clusters were evaluated as to how well they represented likely species units, using the Adjusted Wallace concordance metric. *18S* data set clusters were compared against morphologically-identified species groups (morphospecies) as well as *COI* data set clusterings for all four similarity-based MOTU-generating methods (note BINs are only available for the COI gene). These congruence comparisons indicated that clusterings using 18S-V4 data at the elbow-obtained GPSD thresholds were not able to describe either the COI-5P or the morphospecies groupings (see Results and Discussion for further explanation). Therefore, a second method based upon concordance comparisons was used to obtain GPSD thresholds for the *18S* data set.

18S-V4 GPSD threshold identified through concordance to COI-5P and morphological data

GPSD thresholds for 18S-V4 data were obtained through concordance comparisons to COI-5P clusters and morphospecies. To obtain the optimal GPSD threshold for the 18S-V4 data based on the COI-5P data (hereafter called the 'COI-based' threshold), the clusters for the *COI* data set, using the COI-5P elbow-obtained optimal GPSD threshold (based on the full *COI reference* data set), were compared to *18S* clusters. Comparisons, using the Adjusted Rand concordance analysis, were completed using *18S* clusters generated using a range of GPSD

thresholds between 0-15%. Comparisons were only made across clusters generated using the same similarity-based clustering methods. The highest Adjusted Rand value was selected as the optimal threshold for the 18S-V4 data as determined by the COI-5P data. This method was repeated using morphospecies as the comparative data set in place of the COI-5P clusters (generating a 'morphology-based' 18S threshold).

Clustering agreement analysis

Morphospecies, as shown on the COI-5P Bayesian tree (Figure 3.2), were compared against 18S-V4 clusters—using both COI-based and morphology-based thresholds—through Adjusted Rand metric comparisons. Comparisons between *18S* clusters and morphospecies fell into four potential outcomes: match, split, lump, and mix (Ratnasingham & Hebert 2013, Chapter 2). Analysis to obtain these outcomes was completed using an R script (Appendix III). A complete 'match' was where the morphospecies (as shown on the COI-5P tree; Fig 3.2) exactly matched to an *18S* cluster; a 'split' was where the morphospecies was represented by two or more *18S* clusters, with no members of the corresponding clusters being unaccounted; a complete 'lump' was where two or more morphospecies were represented by a single *18S* cluster, with no members unaccounted; and a 'mixed' result was where a morphospecies was both split and lumped (Figure 3.2).

Results

Sequencing success: COI-5P vs. 18S-V4

The final sequence success for the COI-5P collection was 59% (Table 3.2). Amplification for the 18S-V4 region was more successful, most likely due to the universal nature of the Uni18S and Uni18SR primer set, enabled by evolutionary conservatism of that DNA region. Amplification of the 18S-V4 molecular region using Uni18S and Uni18SR primers yielded an 81% amplification and sequencing success after only a single attempt. No one taxon from the 18S-V4 data set had greater apparent amplification success compared to any other. Noted 18S-V4 sequencing failures were predominantly due to the amplification of non-target DNA.

Elbow-obtained GPSD thresholds for COI-5P and 18S-V4

The *COI reference* and *18S reference* data sets displayed distinctly different trends in the relationship between numbers of MOTUs generated and sequence threshold (Figure 3.1). The COI-5P data displayed more clearly-defined elbows, whereas the 18S-V4 exhibited a more linear trend. For the *COI reference* data set, elbow analyses yielded values for the similarity-based methods of ABGD = 2.1%, jMOTU = 2.3%, Mothur = 2.6%, and UPARSE = 2.2%; for the *18S reference* data set, these values were more variable: ABGD = 7.5%, jMOTU = 3.7%, Mothur = 6.7%, and UPARSE = 2.0%. Across all four similarity-based clustering methods, there was an overall low Adjusted Wallace concordance (Table 3.3A. species: 0.22 to 0.76, avg = 0.53; Table 3.3B, COI-5P: 0.18 to 0.48, avg = 0.33), indicating the poor ability of the 18S-V4 clusters to explain both COI-5P clusters and morphospecies.

Global threshold identified through concordance comparison

The morphology-based thresholds, identified through the highest Adjusted Rand concordance values between *18S* clusters and morphospecies, were: ABGD = 1.0%, jMOTU = 1.3%, Mothur = 2.8%, and UPARSE = 1.5%. The COI-based thresholds were ABGD = 4.2%, jMOTU = 1.3%, Mothur = 1.0%, and UPARSE = 0.6%. Both of these sets of concordance comparisons yielded a more constrained range of GPSD thresholds for each similarity-based analysis method than those obtained through elbow analysis for the 18S-V4 data set. There was a large degree of variation in the number of MOTUs generated for each similarity-based clustering method. This large variation was present when using both the morphology-based and COI-based thresholds applied to the *18S* data as well as when comparing the *18S* and *COI* data sets. Using the morphology-based 18S-V4 threshold, between 54 and 64 clusters were generated across the four similarity-based methods; when the COI-based thresholds were applied to the *18S* data set, there were 50 to 75 clusters. Both of these results are in comparison to 65 to 90 clusters generated for the same specimens using the *COI* data set. Furthermore, each of the four methods, ABGD, jMOTU, Mothur, and UPARSE, generated a larger number of clusters using the *COI* data set compared to the same method using the *18S* data set, with between 8 and 34 more clusters across all methods.

The 18S-V4 generated between 20-64% more clusters using similarity analyses as compared to morphological species. However, the 18S-V4 clusters only matched or split the

morphological species half of the time (9 of 18 morphologically-identified species with more than one representative sequence) (Figure 3.2). COI-5P yielded 16-56% more clusters than morphologically-identified species. Although, the 18S-V4 data yielded a greater estimate of diversity compared to morphological identifications, this diversity matched neither morphology nor COI-5P groups. Meanwhile, COI-5P yielded more diversity than morphology and was able to explain both the 18S-V4 groupings and the Linnaean species most of the time.

The adjusted Wallace concordance analysis revealed lower discriminatory power for 18S-V4 to describe the morphospecies compared to COI-5P (Table 3.3C, D, and E). Adjusted Wallace concordance values for 18S-V4 clusterings, using morphology-based GPSD thresholds, indicated that 18S-V4 was only able to correctly place a specimen into a morphospecies approximately 75% of the time (Table 3.3C). Additionally, morphospecies were only able to correctly place specimens into 18S-V4 groupings approximately 50% of the time. Adjusted Wallace concordance values for 18S-V4 clusters, using COI-based GPSD thresholds, were nearly identical to concordance values using the morphology-based GPSD thresholds (Table 3.3D). Finally, comparing the 18S-V4 and COI-5P performance at predicting morphospecies, the COI-5P molecular data exhibited values that were 0.15-0.18 higher (Table 3.3E). This suggests that, in general, clustering analyses using COI-5P data has more discriminatory power than 18S-V4 data.

Bayesian phylogenetic analysis and clustering agreements

All genera on the COI-5P Bayesian tree were recovered as monophyletic with two exceptions: the first where *Mesocalanus* grouped within the genus *Calanus*, and second where *Eurytemora* grouped within *Centropages*. Species-level monophyly was also observed within most genera, with again only two exceptions, as *Acartia* and *Calanus* each contained polyphyletic Linnaean species. While morphospecies within genera were generally well supported as clades (75% of the clusters displayed in Figure 3.2 had posterior support values of 0.80 or greater), genus-level support varied, with posterior probability support values ranging from 0.03 to 1. There were several clades with very low posterior support including two *Acartia longiremis* clades with 0.10 and 0.01 values, two of the three *Calanus* groupings having 0.14 and 0.24 posterior probabilities, and finally two Pacific *Pseudocalanus* groupings with 0.18 and 0.24

posterior probability support values. Although low support values were recorded for ten of the morphospecies on the COI-5P tree, geographic clustering of COI haplotypes was still evident.

Morphospecies showed mixed agreement when compared to the *18S* clusters (Figure 3.2). Of the 15 genera with matches for any of the MOTU-generating methods, only four contained species having exact matches across all 18S-V4 clustering analyses; *Eurytemora, Mesocalanus, Eucalanus,* and *Paracalanus.* One of these four genera was only comprised of two sequences (*Eucalanus*), *Mesocalanus* was comprised of three sequences, and the remaining two genera had larger numbers, *Eurytemora* (n=16) and *Paracalanus* (n=6). Furthermore, for each of these four clades there was either only a single species identified (*Eurytemora hermandi, Mesocalanus tenuicornis,* and *Paracalanus parvus*), or no specimens were able to be identified to species (*Eucalanus*). The 18S-V4 clusters for three genera (*Acartia, Pseudocalanus, Scolecithrecella)* were exclusively mixed with respect to COI-5P MOTUs, and *Aetideus* was predominantly mixed. The results for the MOTU-generating methods for *Tisbe furcata* and *Zaus abbreviates* were mostly split, and the remaining genera analysed had varied results between mixed, lumped, split, and matched across analysis methods.

Discussion

Sequencing success: COI-5P vs. 18S-V4

Given the importance of characterizing biodiversity using DNA barcoding and high-throughput sequencing technologies, my results provide a clearer understanding for methodological choices and molecular marker selection, at least for marine planktonic crustaceans. The 18S-V4 molecular region was not able to delimit specimens into species groups congruent with Copepoda and Thecostraca morphological species. Also, using the 18S-V4 region for metabarcoding underestimates the species-level biological diversity by 5-15% in comparison with COI-5P. In light of lower discriminatory power of 18S-V4 compared to COI-5P and morphology, careful consideration of the data required to address study questions is necessary. In situations where 18S-V4 data would provide enough information to address research questions it may be recommended over the COI-5P pending further protocol development, due to greater sequencing success. For example, the 18S-V4 region may be applicable in situations where researchers have *a priori* knowledge that they will not require

species-level information to address their research question. As well, a very low divergence value for matching could be selected if the study goal was not to delimit species or estimate species richness but rather to screen for invasive species.

The first (and only) amplification and sequencing attempt for 18S-V4 data resulted in high sequencing success; by contrast, after several attempts, the sequencing success of COI-5P remained low by comparison. This low COI-5P amplification success could be due, in part, to the taxonomic composition of the data set, as the variable amplification success for some planktonic crustacean taxonomic groups has been previously documented (Bucklin *et al.* 1999; Roe & Sperling 2007; Bucklin *et al.* 2010b). Another possible reason for the lower COI-5P sequencing success could be that the specimens used in this study, due to logistics of sampling, were not immediately and consistently stored in cold (-20^{o}C) storage medium. Initial fixation of specimens in cold storage medium has been documented to increase successful sequencing of the COI-5P molecular region (Prosser *et al.* 2013).

It is important to note that the main reason for the removal of sequences from the 18S-V4 data set was suspected contamination, possibly from environmental DNA or specimen gut content. My laboratory procedures were careful, and I minimized the potential of contaminant DNA. The gut contents, epibionts, and/or the bulk collection and storage of specimens are therefore assumed to have introduced the majority of incidences of contamination, rather than cross-contamination among targets; this is supported by the finding that most cases of non-target sequences were organisms other than the classes targeted. However, given the relatively small potential of contamination due to the small amount of non-target DNA present, compared to the larger target specimen, concerns of preferential amplification should be addressed where one-by-one barcoding is not completed or where large amounts of non-target DNA are present. Another possibility is that the high incidence of contamination may be a result of a primer annealing bias for 18S-V4, and if this were the case metabarcoding may not resolve these instances of contamination and instead may miss valuable information, which may be partially overcome through the great sequencing depth of modern platforms. While some research has been performed on artificially-constructed communities for the 18S-V4 region (Brown *et al.* 2015), amplification bias in diverse potential source communities remains an area for further exploration.

Optimal threshold determination

Delimiting species using molecular data and a barcoding approach may not have been the primary intention of DNA barcoding, but since Hebert *et al.* (2003a) introduced the concept of barcode-based specimen identification, barcoding has been increasingly used for this purpose (Hubert & Hanner 2015). Single-marker species delimitation has drawn substantial research attention regarding methods development (Hubert & Hanner 2015) as well as criticism (Meyer & Paulay 2005; Will *et al.* 2005; Krishnamurthy & Francis 2012). However, in order to effectively and rapidly analyze large barcode sequence datasets, a similarity-based approach to species delimitation is most appropriate. Before analyzing a data set in this manner, one must first choose a global pairwise sequence divergence threshold to cluster the data. Many past studies which apply a threshold-based approach to species delimitation may not effectively support their threshold selection (Collins & Cruickshank 2013). There are exceptions such as Lefébure *et al.* (2006), where they not only supported a threshold for COI-5P data but also recommended their methodology be used to obtain thresholds for other markers and taxa. Here, I have taken a similar approach and have provided a much less complex methodology to achieve optimal thresholds. Whereas Lefébure *et al.*'s (2006) methodology required the calculation of the divergences since the hypothesized last common ancestor to compare two clades, my method estimates the most common transition point in the dataset between interspecific and intraspecific divergences. My method assumes that heterogeneity in rates of molecular evolution are modest, an assumption that GMYC does not make by insisting upon an ultrametric input tree, which in turn increases the difficulty and computation time for analysis.

I took several approaches to obtain the 18S-V4 global pairwise sequence divergence threshold (GPSD) necessary to cluster specimens into species-like groups. The first method, the internal elbow analysis method, obtained the optimal threshold necessary to group sequences into clusters, where the data set was neither over-split nor over-lumped. The COI-5P data had a clear 'elbow' point; however, this does not appear to be the case with the 18S-V4 data. This difference is not surprising as the COI-5P region is expected to coalesce more quickly than the 18S-V4 gene region due to the smaller effective population size of mitochondrial compared with nuclear markers (Hubert & Hanner 2015). Additionally, the absence of recombination—as well as positive selection, selective sweeps, and purifying selection likely constraining COI-5P intraspecific variability—contribute to the clear elbow present in the COI-5P data (Hebert *et al.*

2003b; Stoeckle & Thaler 2014; Hubert & Hanner 2015; Raupach *et al.* 2015). The 18S-V4 curve is more linear with no clear elbow, and therefore no clear optimal GPSD threshold is obtainable with this approach. These more linear clustering results across divergence thresholds for 18S could also be due to concerted evolution, whereby repetitive genetic regions experience unequal crossing over and tend to evolve in concert, resulting in more similar 18S-V4 sequences within species and between closely related species (Hillis *et al.* 1991; Liao 1999). This constraint of variability among the copies within a genome and subsequent recombination between individuals within the species can lead to lower variability. Evidence of likely concerted evolution has been suggested for calanoid species (Laakmann *et al.* 2013), in which COI displays high intraspecific sequence variation as compared to little to no variation between specimens of the same species in 18S (Bucklin *et al.* 2003). In addition to the lower variability of the 18S region, the 18S-V4 elbow results differed more than the COI-5P results across clustering methods; this finding is similar to that of Tang *et al.* (2012), where it was noted that there was variation in 18S-V4 clusters among the methods applied.

In light of the uncertainty of the threshold value obtained through 18S-V4 elbow analysis, and the low concordance to morphological species identifications using this threshold, two additional thresholds were obtained and further explored. The COI-based thresholds and the morphology-based thresholds showed very little variation in clusterings when applied to 18S-V4 data. The jMOTU clustering was the most consistent across the COI-based and morphological-based *18S* GPSD thresholds, with the same 1.3% optimal threshold value. All other similarity-based MOTU-generating methods yielded optimal thresholds that varied in magnitude by 23-40% between COI-based and morphological-based thresholds. Although there was quite a large difference in the obtained thresholds across the four similarity methods using the COI-based and morphology-based approaches for determining the optimal threshold, there was very low variation in the Adjusted Wallace concordance values, which indicate the ability of the 18S-V4 data to correctly place specimens into morphospecies. This result is consistent with the shape of the MOTU curves for the 18S-V4 data (Figure 3.1), where the MOTU/divergence curve was more linear and, in the region of the candidate thresholds, displayed shallow slopes. Interestingly, Wu *et al.* (2015) suggested that a GPSD threshold near 0, allowing for all nucleotide differences to reflect placement into a cluster, was needed to enable species-level placement using 18S-V4 data. My results, through the use of the Adjusted Wallace concordance,

indicate that a slightly higher value provides the best species-level delimitation, and at lower levels the delimitation of specimens into the morphological groupings would decline. If a very well-sampled database of 18S-V4 sequences were created, a threshold near 0 may possibly permit specimen identification, even if species delimitation is not optimal at that level.

Agreement analysis

The 18S-V4 clusters were relatively consistent across clustering methods as compared to morphospecies as shown beside the Bayesian COI-5P tree as a reference (Figure 3.2), in spite of the overall low number of total matches. The most variation in agreement as compared to the morphospecies was between GMYC, PTP, and Mothur. There are two main elements which separate these methods from the other clustering analyses. PTP and GMYC methods use a coalescence-based approach, and Mothur differs from the other similarity-based methods in how it treats gaps during analysis (Table 3.1). The Mothur-generated clusters exhibited more variable levels of agreementto the morphospecies than the other clustering methods. This was evident when comparing the results between the two obtained thresholds (COI-based = 2.8%; morphology-based = 1%). As the implementation of Mothur treated multiple gaps as a single gap during analysis, careful consideration of the use of this program should be taken when implementing it with molecular regions known to exhibit indels, such as 18S-V4. It is possible to change how Mothur treats gaps in the settings for the program, and this may be appropriate if using Mothur as the clustering method of choice for gene alignments with many gaps.

The PTP and GMYC analyses appear to have opposite general outcomes when compared to other clustering methods. GMYC appears to split the data when not in agreement with the other clusterings, while PTP appears to lump. This discrepancy between PTP and GMYC methods was more pronounced for 18S-V4 (this study) than for COI-5P (Chapter 2). When a coalescence analysis is used, a similarity-based analysis may also be conducted for comparison, giving consideration to the study aims. Given that the similarity-based methods are much easier and faster compared to coalescent-based methods, this additional work should be weighed considering the study aims. Completing both types of analyses can provide a basis of comparison for the coalescence method and may indicate potential instances of over-splitting or over-lumping.

The ABGD results, which are based upon a recursive clustering method, do not appear to have a larger proportion of matches to morphospecies than other similarity-based clustering methods. The four similarity-based clustering methods did not show substantial differences in their proportion of matches of 18S-V4 clusters to morphospecies. There were 6 morphospecies which had high posterior probability support values on the COI tree (>0.95) and matched MOTUs generated using the similarity-based methods (*Balanus glandula, Eurytemora hermandi, Centropages typicus, Mesocalanus* sp., *Eucalanus* sp., *Paracalanus parvus*). For these species, my results suggest that the 18S-V4 data can provide clusters congruent with morphology. However, these 6 examples appear to be the exception. Of the total 22 morphospecies that formed clades and exhibited high (>0.95) posterior probability, 12 resulted in predominantly lumped or mixed agreements across clustering methods. These results further indicate that clusters obtained using the 18S-V4 data provided low agreement to species groups established using traditional methods as well as DNA barcoding using COI-5P.

Analyses show *Acartia* and *Pseudocalanus* species mixing across genera. Species within these genera are known to have problematic morphological identifications (Blanco-Bercial *et al.* 2014; Bucklin *et al.* 2015). Uncertain species-level discrimination among *Acartia* using 18S-V4 is similar to COI-5P results, where inconsistent delimitation is known to occur (Blanco-Bercial *et al.* 2014). While *Aetideus and Scolecithrecella* are not known to be taxonomically problematic, with few representative sequences in my study, further sampling is necessary before extensive discussion could occur with regards to delimitation using 18S-V4 for these genera. This may be especially necessary for the Aetideidae family as there are closely-related species among pelagic members, which can often be found in sympatry, potentially leading to difficult morphological identifications (Laakmann *et al.* 2012). *B. balanus* had predominantly mixed results across similarity-based methods. The sequences used for this work were obtained from nauplii, which display very few morphological characteristics, and so uncertain morphological identifications could contribute toward the variations in the results seen with *B. balanus*. Finally, two genera contained species which had lumped clustering results (*Centropages* and *Calanus*), where a single molecular cluster contained all representatives of more than one morphological species. This outcome is not surprising for the recently-radiated *Calanus* species, as hybridization among member species has been reported (Parent *et al.* 2012).

Six morphologically-identified species showed predominantly matching between 18S clusters and morphospecies, indicating that the 18S-V4 region can effectively delineate these morphospecies. Although these six species were predominantly matched to the 18S clusters, it is important to note that species were sparsely sampled in these genera. Although results presented here were obtained through expert identifications, independent verification of the identifications may provide more insight into the reasons for the discordant results observed here. Cases of discordance between morphological and molecular results using both COI and 18S data may more strongly indicate morphological misidentification. If morphological identifications remain unchanged and the agreement results stay the same, then 18S-V4 is not suitable for delimiting congeneric species. With the small data set presented here, it is uncertain if a true barcode gap exists for some groups even though there was correspondence between morphological species and 18S-V4 MOTU in some cases; with increased sampling, the overlap between interspecific and intraspecific divergence increases (Meyer & Paulay 2005), and the 18S-V4 region will most likely not be effective.

Given the low sampling density for this study, it is likely that the average genetic distance to the closest neighbour would significantly decrease with increased sampling of species within the target genera as well as increased sampling of localities within species (Bergsten *et al.* 2012). With increased sampling, the main results presented here—substantially lower discriminatory power of 18S-V4 compared to COI-5P at the species level—may become even more prominent. Therefore, the use of 18S-V4 does not appear to be ideal for species delimitation, an important application in the greatly-understudied marine realm. However, further work exploring its capacity for enabling specimen identification in the case of matching to well-populated databases is warranted. This may be of interest for high-priority taxonomic groups, such as those containing invasive species.

Conclusions

The efforts detailed here have identified a slightly larger GPSD threshold compared to Wu *et al.* (2015): between 0.6% and 4.3%, depending upon the clustering method used as well as the criterion for optimizing the threshold. Using these thresholds, results support the findings of Tang *et al.* (2012) and indicate that the 18S-V4 region underestimates species-level diversity

when compared to morphological species groups. Furthermore, a slightly higher GPSD threshold is less prone to inflating the number of molecular clusters because of sequencing errors than when using a 0% threshold such as suggested by Wu *et al.* (2015). Eliminating or trimming sequences with potential errors is an integral step in all DNA barcoding efforts, and verification methods such as translating protein-coding gene sequences are standard in large sequence data base generation efforts (Ratnasingham & Hebert 2007), an approach not possible for 18S. Although a more sophisticated data validation procedure based upon secondary structure may be possible, this could be challenging across all of life to develop a robust and computationally efficient system. A 0% GPSD threshold would therefore be expected to establish a new cluster for every single nucleotide difference; with increasing use of error-prone sequencing methods, an inflation of true diversity is certain to occur (Kunin *et al.* 2010).

Without the ability to verify ribosomal gene sequences through translation, detecting errors is difficult and relies upon complex secondary structure models. Methods to eliminate rare sequences are standard steps in molecular studies and are available in most metabarcoding pipelines (including UPARSE and Mothur used to cluster sequence data sets here). Current metabarcoding technologies produce sequence reads of lengths between approximately 50 and 500 bp (Shokralla *et al.* 2012). With the possibility of hundreds of thousands or millions of sequence reads when using a metabarcoding approach, compared to smaller numbers in one-by-one Sanger sequencing work, the difficulty is in identifying reads containing sequencing errors. When larger numbers of short sequences are generated, using the same platform and experiencing the same potential error, the chance to have identical sequences due to the same sequence error increases. If one were to use a 0% threshold, the chance of multiple instances of the same error increases, thereby increasing the number of sequences in an erroneous cluster, which may provide incorrect support for the cluster based on incorrect sequence data (Kunin *et al.* 2010).

With these difficulties using the 18S-V4 region, caution should be taken when assessing marine biological diversity through a metabarcoding approach. In addressing new methodologies with a metabarcoding approach and using the 18S-V4 molecular region, Zhan *et al.* (2013) described a much higher-than-expected level of diversity present in fresh water collections in Hamilton Harbour, Ontario, Canada. These results were largely driven by high number of singletons, sequences in the data set represented by only a single sequence (Zhan *et al.* 2013).

Flynn *et al.* (2015) investigated the informatics approaches and their influence on the diversity estimates using the same pyrosequencing platform. One concern which can be noted from these studies is the difficulty in determining what represents true biological data and what represents errors during the sequencing process. The pyrosequencing process has been reported to have a much higher rate of sequencing errors compared to traditional Sanger sequencing technologies (Balzer *et al.* 2011). This is especially concerning when using the 18S-V4 region, which naturally contains many insertion and deletions, and so distinguishing between errors and biologically meaningful data may be challenging, confounding conclusions. Flynn *et al.* (2015) showed that depending on how gaps were treated, there could be up to a two-fold difference in the resulting diversity estimates using a MOTU approach with pyrosequencing data. Applying methodologies in an attempt to eliminate these singletons, such as applying a filter threshold of 0.5% (Brown *et al.* 2015), may reduce the number of singletons but may also be eliminating data from closely-related sister taxa for gene regions with low variability or rare species in the samples such as introduced species.

With the very small threshold needed to delimit marine planktonic crustacean species using the 18S-V4 gene region, as shown both here and by Wu *et al.* (2015), the elimination or correction of sequencing errors may not be possible and, when attempted, may be eliminating true biological diversity. Although it is unlikely that the diversity present in aquatic systems is as high as reported in Zhan *et al.* (2013), further work is required to gain an understanding of intraspecific vs. interspecific variability in 18S-V4 that would allow one to separate what is biological diversity and what is error induced via metabarcoding through high-throughput sequencing.

The 18S-V4 region may still be beneficial for specific study questions. For example, multiple-marker phylogenetic and taxonomic studies may still benefit from the 18S-V4 region, as COI-5P has been noted to provide resolution among closely-related taxa but is less informative of higher taxonomic relationships (Bucklin *et al.* 2011). If possible, additional carefully-selected molecular markers to those studied here can further support taxonomic revision studies (Dupuis *et al.* 2012; Chesters *et al.* 2015). Based on results, the use of COI-5P remains the best option when conducting one-by-one traditional DNA barcoding work for marine planktonic crustaceans (Bucklin *et al.* 1999). While it is clear that COI-5P would be the 'gold' standard for use in single-marker high-throughput marine planktonic crustacean research, given

the amount of information it provides, the poor success of amplifications across a broad range of taxonomic groups needs to be addressed. Until such a time, the use of 18S-V4 for environmental DNA screening may be effective when applied to certain studies, for example, screening samples for the presence of a target species, where the target was divergent enough from the other specimens to be able to identify the specimen using the 18S-V4 region. This could be a possible solution when screening for the presence of invasive species in an environmental sample where the native species pool differs sufficiently that the 18S-V4 region may be used (Comtet *et al.* 2015). Another application could be the exploration of biological turnover, or for ongoing assessment. Given the limitations of the 18S-V4 region to correctly group specimens into known species, it seems clear that if the COI-5P is obtainable, then the COI-5P marker is the optimal choice.

Tables and Figures

Table 3.1 Clustering algorithms used to generate Molecular Operational Taxonomic Units (MOTU) from 18S sequences. The underlying mechanisms indicate the type of clustering performed by each program. The recursive clustering indicates that once a barcode gap is determined, the entire data set is partitioned. Each member cluster of this primary partition is then subjected to the same partitioning method, yielding secondary partitions. This process is repeated until no further partitioning occurs. A greedy algorithm selects the best locally optimal choice to provide globally optimal clusterings. Treatment of gaps varies across methods, as specified in the final column.

	Underlying Mechanism	**Basis of clustering**	**Treatment of alignment gaps**
ABGD	Recursive partitioning	Pairwise distance matrix based on pre-aligned sequence input	Gaps are accounted for using pairwise deletion during the distance matrix construction
jMOTU	Greedy algorithm	MegaBLAST pre-clustering followed by Needleman-Wunch algorithm to calculate distances which are used to cluster sequences	The jMOTU program aligns the sequence data and gaps are ignored on a pairwise basis when calculating pairwise distances
Mothur	Greedy algorithm	Pairwise distance matrix based on pre-aligned sequence input	Single unknown nucleotides are treated as a gap and multiple gaps are treated as a single gap
UPARSE	Greedy algorithm	Clustering occurs on a pairwise basis from input of pre-aligned sequences	Gaps are coded as unknown nucleotides and are treated as a difference in nucleotides on a pairwise basis
GMYC	Optimum maximum likelihood solution between Yule and neutral coalescent models	Ultrametric and bifurcating input tree	Missing data are treated as equally probable for the 4 base nucleotides
PTP	Poisson Tree Processes model	Phylogenetic input tree	Missing data are treated as equally probable for the 4 base nucleotides

Table 3.2 Number of specimens morphologically identified to each taxonomic level. Success within each taxonomically-identified group is based on final cleaned, edited, and aligned nucleotide sequences. Instances of non-target amplification are not included.

	Class	Order	Family	Genus	Species	Total
Number of specimens identified to a given taxonomic rank (lowest possible identification)	111	34	1	121	416	683
18S sequencing success	91	24	0	110	325	550
COI sequencing success	57	18	1	75	253	404
Successful for both COI-5P and 18S-V4	50	12	0	68	214	337

Table 3.3 A comparison between *18S* and *COI novel* data set clustering results using the Adjusted Wallace bidirectional concordance metric. Each sub-table reports Adjusted Wallace bidirectional concordance values between data sets for all four similarity-based analysis methods (ABGD, jMOTU, Mothur, UPARSE). The global pairwise sequence divergence (GPSD) threshold used for each analysis is indicated in brackets beside the analysis method. The rows in the sub-tables indicate the direction of the concordance value results.

A. Adjusted Wallace concordance values comparing 18S-V4 clustering analyses using the elbow-obtained GPSD threshold against morphological identifications.

Adjusted Wallace	ABGD (7.5%)	jMOTU (3.7%)	Mothur (6.7%)	UPARSE (2.0%)
Species explain 18S	0.501	0.44	0.499	0.503
18S explain Species	0.509	0.215	0.757	0.64

B. Adjusted Wallace concordance values comparing 18S-V4 clusterings using elbow-obtained GPSD thresholds against COI-5P data (using elbow-obtained COI GPSD thresholds).

Adjusted Wallace	ABGD (COI - 2.0%, 18S - 7.5%)	jMOTU (COI - 2.3%, 18S - 3.7%)	Mothur (COI - 2.6%, 18S - 6.7%)	UPARSE (COI - 2.2%, 18S - 2.0%)
COI explain 18S	0.84	0.813	0.7	0.727
18S explain COI	0.313	0.183	0.484	0.358

C. Adjusted Wallace concordance values comparing 18S-V4 clustering analyses using the species-obtained GPSD threshold against morphological identifications.

Adjusted Wallace	ABGD (1.0%)	jMOTU (1.3 %)	Mothur (2.8%)	UPARSE (1.5%)
Species explain 18S	0.51	0.519	0.496	0.505
18S explain Species	0.767	0.758	0.764	0.755

D. Adjusted Wallace concordance values comparing 18S-V4 clustering analyses using the COI-5P-obtained GPSD threshold against morphological identifications.

Adjusted Wallace	ABGD (4.3%)	jMOTU (1.3 %)	Mothur (1.0%)	UPARSE (0.6%)
Species explain 18S	0.513	0.519	0.488	0.487
18S explain Species	0.757	0.758	0.761	0.76

E. Adjusted Wallace concordance values comparing COI-5P clusters using elbow-obtained GPSD threshold against species morphological identifications.

Adjusted Wallace	ABGD (2.0%)	jMOTU (2.3%)	Mothur (2.6%)	UPARSE (2.2%)
Species explain COI	0.612	0.476	0.484	0.471
COI explain Species	0.921	0.912	0.938	0.937

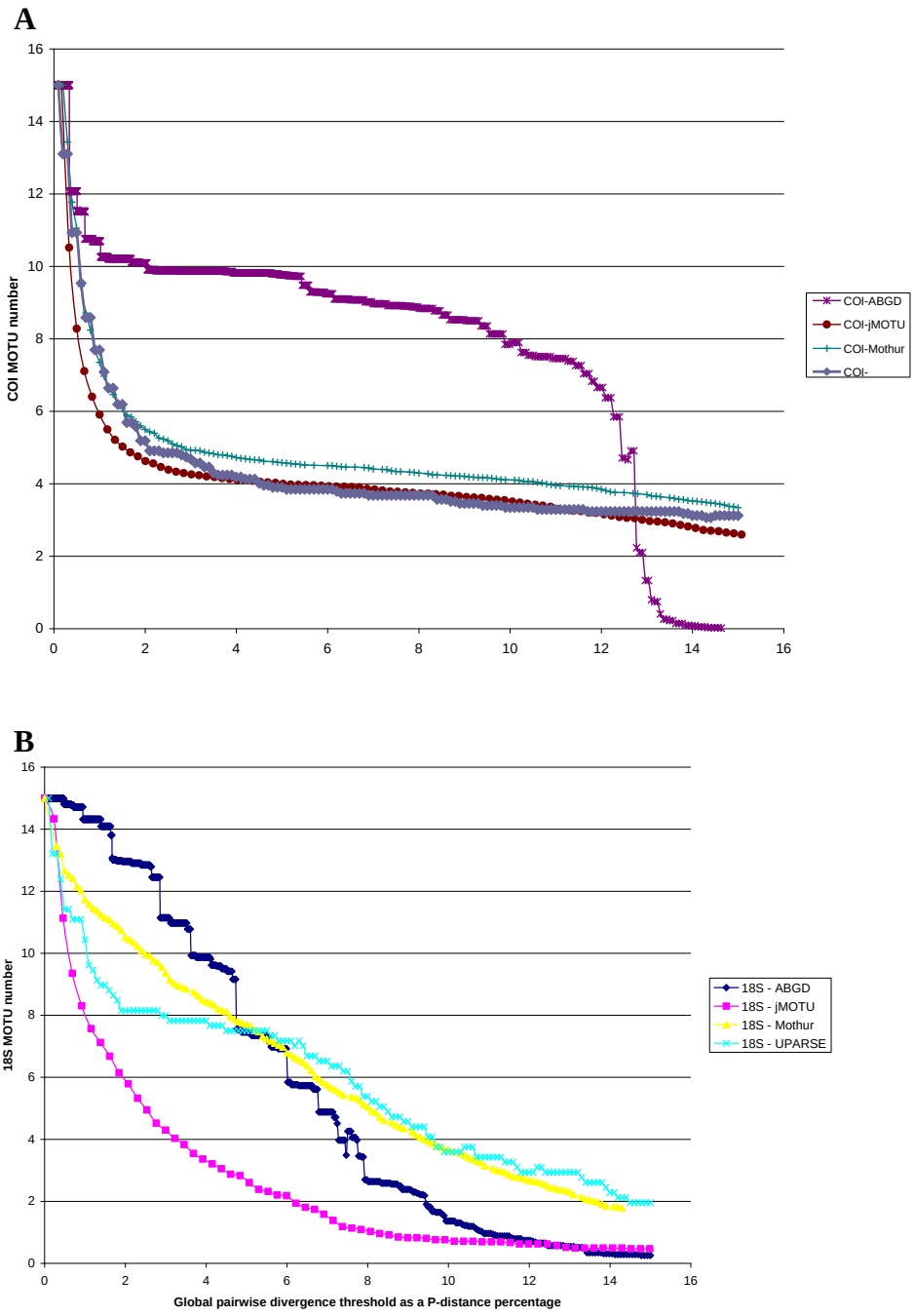


Figure 3.1 Graphed total number of molecular clusters (MOTU) yielded for each similarity-based clustering method for all global pairwise sequence divergence thresholds (0-15%). **A.** *COI reference* data set, **B.** *18S reference* data set.

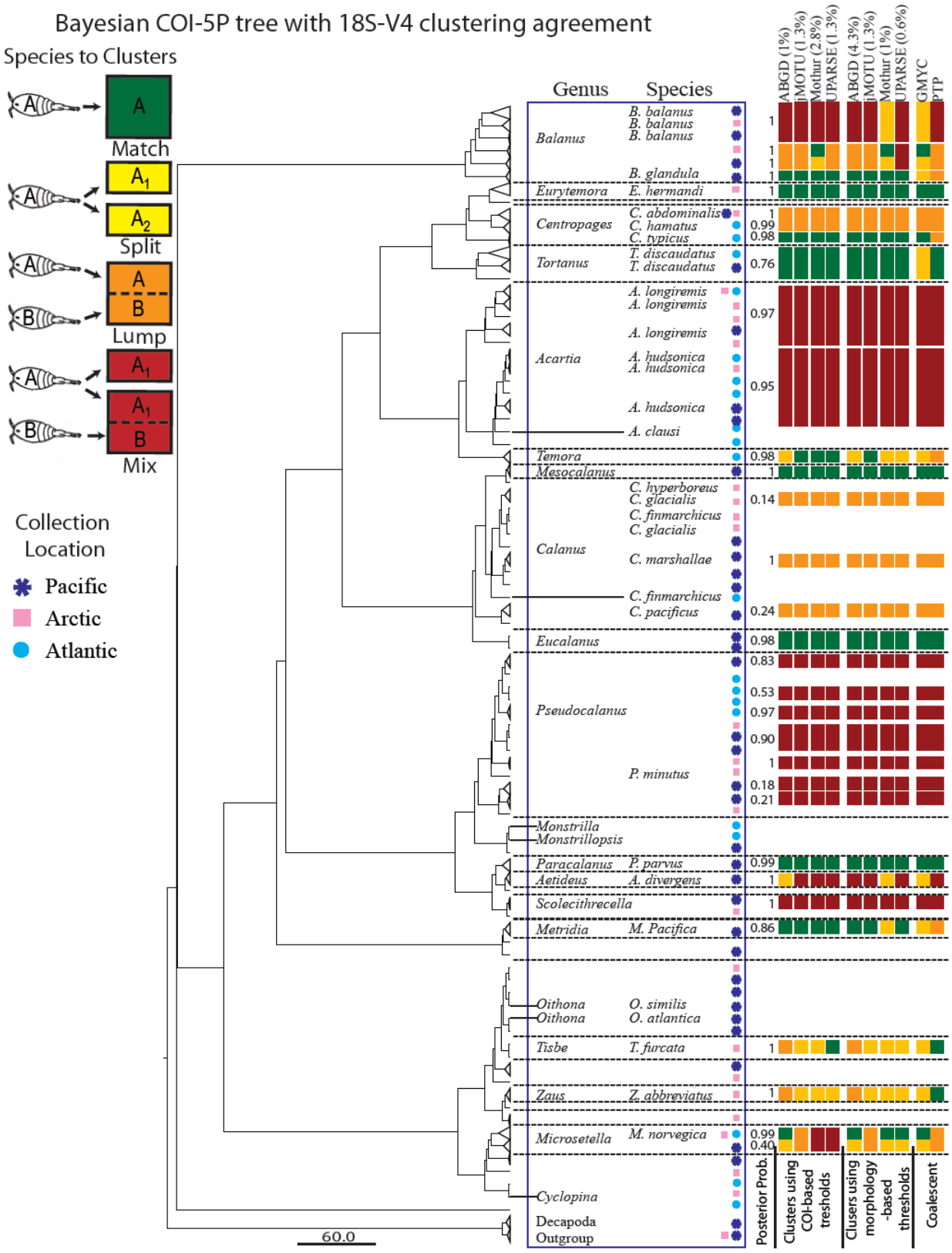

Bayesian COI-5P tree with 18S-V4 clustering agreement
Species to Clusters
Match
Split
Lump
Mix
Collection Location
Pacific
Arctic
Atlantic
Genus
Species
ABGD (1%)
iMOTU (1.3%)
Mothur (2.8%)
UPARSE (1.3%)
ABGD (4.3%)
iMOTU (1.3%)
Mothur (1%)
UPARSE (0.6%)
GMYC
PTP
Balanus
B. balanus
B. balanus
B. balanus
B. glandula
Eurytemora
E. hermandi
Centropages
C. abdominalis
C. hamatus
C. typicus
Tortanus
T. discaudatus
T. discaudatus
Acartia
A. longiremis
A. longiremis
A. longiremis
A. hudsonica
A. hudsonica
A. hudsonica
A. clausi
Temora
Mesocalanus
Calanus
C. hyperboreus
C. glacialis
C. finmarchicus
C. glacialis
C. marshallae
C. finmarchicus
C. pacificus
Eucalanus
Pseudocalanus
P. minutus
Monstrilla
Monstrillopsis
Paracalanus
P. parvus
Aetideus
A. divergens
Scolecithrecella
Metridia
M. Pacifica
Oithona
O. similis
Oithona
O. atlantica
Tisbe
T. furcata
Zaus
Z. abbreviatus
Microsetella
M. norvegica
Cyclopina
Decapoda Outgroup
60.0
Posterior Prob.
Clusters using COI-based tresholds
Clusers using morphology -based thresholds
Coalescent

Figure 3.2 Bayesian COI-5P tree with collapsed terminal groups and branches concordant with COI-5P BINs as obtained from the BOLD database. Genus and species identifications are mapped onto the tree. Geographic collection locations are indicated and correspond to each of the terminal groups or individual branches; dark purple star = Pacific region, pink square = Arctic region, blue circle = Atlantic region. Posterior probability values are shown for the associated collapsed terminal groups. Agreement analysis results are presented in three columns. The first, indicated on the bottom of the figure as "Clustering using COI-based thresholds", displays results from 4 similarity-based analyses (ABGD, jMOTU, Mothur, UPARSE), using 18S-V4 data and thresholds obtained from comparison to COI-5P clusterings. The second, "Clustering using morphology-based thresholds", again has results from the four similarity-based clustering methods but using thresholds obtained from comparison to morphospecies. And finally, the last set of two results indicates the clusterings from coalescent-based analyses using the 18S-V4 molecular data. Results obtained from each clustering analysis of 18S sequences, as compared to morphospecies on the COI-5P Bayesian tree, fall into four potential categories. A complete 'match' was where the morphospecies on the COI-5P tree exactly matched to a cluster from *18S* data; a 'split' was where the morphospecies on the COI-5P tree was represented by two or more clusters from the *18S* data analysis, with no members of the corresponding clusters being unaccounted; a complete 'lump' was where two or more morphospecies on the COI-5P tree were represented by a single clustering outcome from the *18S* data, with no members unaccounted; and a 'mixed' result was where a morphospecies on the COI-5P tree was both split and lumped.

Chapter 4

The impact of the Bering Strait on the biogeography of marine zooplankton (Multicrustacea: Copepoda: Calanoida)

Abstract

Arctic glaciations and the closing and opening of the Bering Strait have influenced the structure of the biota across North Pacific, Arctic, and North Atlantic marine waters. The aim of this work was to determine if there is a shared distributional pattern of marine planktonic calanoid species across Canada's three ocean regions, based on current sampling, and to further our understanding of the impact of the Bering Strait on the biogeography of Northern plankton. A subsequent aim was to compare the planktonic patterns to known benthic phylogeographic patterns. Cytochrome *c* oxidase subunit I (COI-5P) sequence data were generated and mined from public data sources across the three ocean regions that border northern North America (Arctic, Atlantic, and Pacific) for eleven Calanoida (Copepoda) genera (*Acartia, Aetideus, Calanus, Centropages, Eurytemora, Mesocalanus, Metridia, Paracalanus, Pseudocalanus, Temora, Tortanus*). Using maximum likelihood phylogenetic methods, I reconstructed the topology for 39 taxonomically-identified species from the Pacific, Arctic, or Atlantic Oceans, comprising 79 molecular groups (BINs, Barcode Index Numbers), to discover if there are shared biogeographical patterns across species and genera. My results indicate the calanoid genera display varying levels of phylogeographic structure both between species and within species using the COI-5P sequence data. Biogeographical patterns of benthic marine organisms, which generally display recent dispersal between north Atlantic and Arctic waters, showed similarity to calanoid copepod biogeographical patterns; both groups also display cases of apparent allopatric divergence across the Bering Strait. Finally, I suggest a need for more extensive sampling of marine calanoid copepods across Canadian ocean regions with emphasis on populating publicly accessible sequence databases with taxonomically-identified sequences.

Introduction

Marine copepods comprise a large component of marine biodiversity, and marine planktonic copepods are one of the most abundant groups of multicellular organisms on earth (Humes 1994; Bucklin *et al.* 2010b; Blanco-Bercial *et al.* 2014). Marine calanoid copepods (Multicrustacea: Copepoda: Calanoida), having vast population sizes and large geographic distributions across diverse habitats (Bron *et al.* 2011; Blanco-Bercial *et al.* 2014), are excellent candidates for studying the dispersal patterns and connectivity of holoplankton from Northern Pacific to Northern Atlantic Ocean regions (Bucklin *et al.* 2003). Calanoid species distributions and biogeographical patterns are related to their biological attributes as well as oceanic history. With increasing global temperatures in polar regions, gaining a better understanding of current Arctic biogeography can provide valuable information to better predict future impacts on species distributions in the region. Gaining a better understanding of contemporary marine copepod species distributions will provide further understanding of past impacts on the distribution of marine biota.

In particular, a major influence on current marine species distributional patterns is the Bering Strait, while the past Bering Land Bridge acted as a barrier to gene flow. The Bering Land Bridge is a physical land bridge that closed off the Bering Strait approximately 80 mya, during the late Cretaceous Period, resulting in the isolation of the Pacific and Arctic Oceans (Briggs 2003). At that time, the connection between Arctic and Atlantic Ocean regions was also narrowed but remained open (Dayton *et al.* 1994). Approximately 5.3 million years ago, a significant opening of the Bering Strait allowed the free flow of water between North Pacific and Arctic Oceans (Dunton 1992; Gladenkov *et al.* 2002).

Since this opening, the Arctic region has experienced intermittent glacial advances and retreats until the last glacial maximum approximately 18,000 years ago, when the ice began its recession to what we see today (Dunton 1992; Gladenkov *et al.* 2002). Periods of glaciation caused dramatic drops in sea levels, exposing sea floor habitat as deep as 85m and severely damaging benthic communities (Dayton *et al.* 1994; Piepenburg 2005). By contrast, glacial melts allowed species to colonize the relatively unpopulated regions of the North Pacific and Arctic (Adey & Steneck 2001; Gladenkov *et al.* 2002). The opening of the Bering Strait also resulted in

ocean current changes, which assisted species moving into the Arctic region due to northward-flowing waters (Dayton *et al.* 1994; Adey & Steneck 2001; Gladenkov *et al.* 2002). These changes and subsequent species migrations from the Pacific into the Arctic Ocean have been supported, with fossil evidence indicating a prominent trans-Arctic movement of species approximately 3.5 MYA (Vermeij 1991). These glacial and colonization events are prominent influences upon current Arctic marine biogeographical patterns (Dayton *et al.* 1994).

Numerous studies covering a wide diversity of marine life have provided contradictory evidence for the geographic centre of origin of the biota of the Arctic Ocean. There are three main groups that have been targeted in prior biogeographic studies: near-shore marine macro algae, benthic invertebrates with planktonic larvae, and holoplanktonic species. Using a survey of morphologically-identified species, Dunton (1992) suggested that there is an Arctic paradox in the geographic origin of the species that inhabit the region, whereby benthic near-shore algae originate from Atlantic waters, while invertebrate species generally display Pacific origins. With increased sampling and more molecular evidence, more recent studies have shown that the origin of the Arctic flora is more complex and includes probable occurrences of Pacific and Atlantic origin as well as endemic species that emerged from glacial refugia, which have occupied the Arctic since before the most recent glaciations (Saunders & McDevit 2013; Laughinghouse *et al.* 2015).

There is no consensus on the origins of Arctic marine invertebrate species, with some taxa displaying stronger Pacific affinities and others Atlantic. Polychaete diversity and distribution patterns in Canadian marine waters showed greater similarity between Arctic and Atlantic species composition compared to Pacific and Arctic and even less between Pacific and Atlantic Oceans (Bodil *et al.* 2011; Carr *et al.* 2011; Carr 2012). Arctic species composition of Echinodermata is more similar to the North Atlantic than the North Pacific (Hardy *et al.* 2011). The bryozoan species present in Arctic waters have been noted to represent relatively recent movements of species into the region, most likely from North Atlantic into Arctic waters (Kuklinski *et al.* 2013). Arctic marine molluscs include species displaying Arctic-Atlantic affinities (Kuklinski *et al.* 2013), ones with Arctic-Pacific affinities (Dyke *et al.* 1996), as well as species which appear to have Arctic origins dating back to before the last glaciation (Layton *et al.* 2014). Mollusc species with longstanding Arctic origins most likely occupied past areas of glacial refugia in the Northern Arctic and Atlantic regions (Layton *et al.* 2014). Interestingly, in

contrast to the biogeographic patterns among some benthic marine invertebrates, water flow patterns and molecular evidence indicate that Arctic planktonic protist species likely have Pacific origins (Darling *et al.* 2007; Lovejoy 2014).

Larval dispersal range and the life history stages associated with dispersal in marine organisms directly contribute to population dynamics and connectivity in marine environments (Cowen & Sponaugle 2009). Of the approximately 40 marine metazoan phyla, only 6 are completely lacking species with a free-living larval stage (Pechenik 1999). Furthermore, over 55% of the known species in the remaining 34 phyla produce planktonic larvae that remain in the water column for days to weeks (Pechenik 1999). However, examinations of genetic structure for holoplanktonic species between North Pacific and North Atlantic populations are relatively few compared to studies of benthic species.

One such study by Blanco-Bercial *et al.* (2011a) looked at two species of the calanoid genus *Clausocalanus* (*C. arcuicornis* and *C. lividus*), and different biogeographical patterns were observed for these two species across North Pacific and North Atlantic waters. Interestingly, *C. lividus* displayed distinct phylogeographical structure, with clustering of sequences from each region using cytochrome *c* oxidase subunit I (COI-5P) data, while *C. arcuicornis* appeared to show a mixture of haplotypes across the sampled range (Blanco-Bercial *et al.* 2011a). Goetze (2005) researched two sister species, *Eucalanus hyalinus* and *E. spinifer*. Although *E. hyalinus* exhibited genetic structure across North Atlantic and North Pacific regions, there was no apparent partitioning of genetic variation for *E. spinifer* across the Bering Strait. These two studies highlight how closely-related taxa with seemingly similar biology and dispersal capability can exhibit differing phylogeographic patterning. As holoplankton species, calanoid copepods remain in the water column for their entire life, likely contributing to their large distributions (Goetze 2005; Hardy *et al.* 2011). Despite large distributions of calanoid copepods, geographic population genetic structure is often found across large distances (Blanco-Bercial *et al.* 2011a; Chen & Hare 2011; Goetze 2011).

Elucidating the biogeographical patterns of marine calanoid species is problematic due to the conservation of morphological characteristics in the order (Bucklin *et al.* 2003). With few and difficult-to-discern defining characters present between sibling species, identifications among related species with shared distributional ranges are difficult (Bucklin *et al.* 2003). Utilizing a DNA barcoding approach, specimens can be more easily identified to a species. In

addition, past studies have shown that the COI-5P animal barcode region can effectively delineate specimens into molecular operational taxonomic units (MOTUs) that are largely congruent to taxonomic identifications (Bucklin *et al.* 2010a, 2011; Chapter 2). With few studies investigating the distribution of calanoid copepods between North Pacific and North Atlantic Oceans through the Arctic Ocean region—and conflicting accounts of Pacific-Arctic-Atlantic connectivity—this work addresses an important research gap in our knowledge of Northern biogeography.

Here, I compile barcode data for multiple co-distributed Calanoida species to gain a better understanding of the overall biogeographic patterns across these three ocean regions. With predicted migrations of southern taxa northward and increasing northern temperatures, gaining an understanding of the geographic distribution of zooplankton species present in these waters, and testing historical hypotheses about their present distributions, is essential before other species are introduced (Doney *et al.* 2012). I investigate the biogeographical patterns using distributions of species and MOTU for 11 Calanoida copepod genera (*Acartia, Aetideus, Calanus, Centropages, Eurytemora, Mesocalanus, Metridia, Paracalanus, Pseudocalanus, Temora*, and *Tortanus*). Despite large distributions, I predict that calanoid plankton species will still display genetic divergence between Pacific and Arctic-Atlantic regions because of historical vicariance and low contemporary gene flow across the Bering Strait.

Methods

Molecular data sets and assignment of geographic collection locations

COI-5P sequences were obtained from Chapter 2 for eleven Calanoida genera. This data set was used as a core data set because the specimens were identified by expert identifiers using consistent taxonomic keys; moreover, the collection locations and collecting methods were standardized, representing 12 marine sites across Canadian shores (Chapter 2 Figure 2.1). All morphologically-identified Calanoida genera for which novel sequences were obtained were analyzed here, and additional data were mined through taxonomic searches for all sequences representing the eleven focal genera using the Barcode of Life Data Systems (BOLD, accessed January 13, 2016) public database (Ratnasingham & Hebert 2007). The total number of sequences used for the analyses was 1944, with 39 taxonomically-identified species analyzed

(Table 4.1). (An additional species from the Antarctic was included in the ML tree to increase phylogenetic breadth but was not analyzed for geographic patterns; a list of all species included in the trees can be found in Appendix V.) Sequences obtained in Chapter 2 that received a Barcode Index Number (BIN) designation from BOLD (i.e. sequences > 500 nucleotides in length from a specific portion of the COI gene and with <1% unknown nucleotides) were assigned a taxonomic name, when available and when just one name was associated with a specific BIN (Ratnasingham & Hebert 2013). A BIN is a MOTU designation which is achieved through implementation of a refined single linkage (RESL) analysis (Ratnasingham & Hebert 2013). The RESL algorithm is implemented in BOLD and provides BIN assignments using a 2.2% p-distance sequence divergence seed threshold, and then through refining the resulting groups with neighbouring clusters based on the level of continuity in the distribution of genetic divergences among sequences (Ratnasingham & Hebert 2013).

Downloaded sequence data were identified as originating from the Pacific Ocean, Atlantic Ocean, or Arctic Ocean region (Figure 4.1). For the definition of the Arctic Ocean region as distinct from the Pacific and Atlantic Ocean regions, I primarily followed the marine ecoregions of the world as described by Spalding *et al.* (2007). Spalding *et al.* (2007) ecoregions were used as there were clear descriptions of their methodology for establishing boundaries. In addition, the ecoregions have been widely used for other research, including studying invasive species distributions and for conservation planning and analysis (Molnar *et al.* 2008).

To contrast the distribution patterns between Pacific and Arctic+Atlantic Ocean regions, the division between Pacific and Arctic Ocean regions was slightly adjusted from that described in Spalding *et al.* (2007) and placed to span the Bering Strait (Figure 4.1). The division line separating Arctic and Atlantic Oceans from Spalding *et al.* (2007) varied slightly in comparison to geographical distribution data from the Marine Planktonic Copepods website (http://copepodes.obs-banyuls.fr/en/) (Razouls *et al.* 2005), which was also consulted for this study (Figure 4.1). The zones from the Marine Planktonic Copepod website were used for expected ranges of the study species (Table 4.1) as well as for analysis of the overlap in species composition among ocean regions based upon known distributions of marine copepod species (see below).

Obtained sequences where the closest associated target ocean region was ambiguous (e.g. Indian Ocean collections) were not assigned to one of the three study regions. Nevertheless,

these sequences were still included in the phylogenetic analysis, as denser sampling of species within genera is expected to break up long branches and contribute to phylogenetic accuracy. When geographic collection location data were not available on the BOLD system and where the sequence had an associated GenBank record, the GenBank record was searched to populate this information. Geographic collection data could include GPS coordinates or stated region of collection in an online database or associated peer-reviewed reference. When geographic location data were not available using these methods, the sequences were kept and used in phylogenetic analysis with no geographic collection data. In addition to geographic information for the molecular data, previously-recorded geographic distributions were also obtained for all marine planktonic copepod species using the World Registry of Marine Species online portal and database (www.marinespecies.org, Boxshall *et al.* 2016) as well as the Marine Planktonic Copepods website (Razouls *et al.* 2005).

Maximum Likelihood phylogenetic analysis

Monophyly has not been well established for many copepod genera. Therefore, two *Eucalanus* sequences (ProcessIDs: CAISN632-13 and CAISN633-13), obtained from Chapter 2, were used together for the outgroup; this genus is within the order Calanoida but in a different superfamily (Figure 4.2) from the 11 genera studied here. This provides a high likelihood that the *Eucalanus* outgroup is more distantly related to the ingroup sequences than they are to one another, enabling reliable rooting of the phylogenetic trees. This genetic divergence in relation to other study genera was also present in the tree seen in Chapter 3 (Figure 3.2). Sequence data for each genus were independently aligned using the default (FFT-NS-2) alignment strategy of the MAFFT multiple alignment program (MAFFT Ver. 7) (Katoh & Standley 2013). Sequences were translated using the invertebrate mitochondrial translation matrix in MEGA6 (Tamura *et al.* 2013) as a data validation step to ensure accurate base calling and alignment.

Sequences exhibiting clear alignment errors—such as gaps of 1-2 bp or a shift in reading frame, possibly arising from a base-calling error in the source sequence or amplification of a pseudogene—were removed from the multiple sequence alignment (MSA). Each genus-level MSA was trimmed to a fixed length greater than 500 nucleotides (with the exception of *Pseudocalanus*, which had a final MSA of 471 nucleotides) to provide reliable sequences at 5' and 3' sequence ends; all MSA sequence lengths are provided in Appendix VI. Sequences with

fewer base pairs (bp) than the MSA total alignment lengths were removed from the data set. Model testing was conducted on each aligned genus data set including the two outgroup sequences using MEGA6. The model which had the lowest BIC score was used for further analysis. The selected model for four of the genera (*Acartia, Calanus, Centropages*, and *Paracalanus*) was the Tamura 3-parameter model plus a gamma distribution parameter. The preferred model for the remaining seven genera (*Aetideus, Eurytemora, Mesocalanus, Metridia, Pseudocalanus, Temora, Tortanus*) was the Tamura 3-parameter model plus both gamma and invariant sites parameters (G+I). Using the selected model and 95% pairwise deletion of missing data, maximum likelihood (ML) tree construction was completed for each genus independently using MEGA6 with 1,000 bootstrap pseudoreplicates.

Patterns of copepod species distributions across the Pacific, Arctic, and Atlantic

For the geographic phylogenetic analysis, 80 tip clades or lineages were defined, representing a total of 79 BINs and 39 Linnaean species. Each BIN represented by a single sequence was considered its own terminal, and each BIN that formed a single clade was collapsed into a single terminal clade. Paraphyletic BINs were divided into their constituent monophyletic clusters, each of which was analyzed as a terminal clade. There were two terminal clades defined that each contained more than one BIN. Specifically, one of the two clades within *Pseudocalanus newmani* and the species *Metridia pacifica* were each collapsed into a single group with multiple BINs, due to the close and alternating relationship of the BINs within these groups. Finally, there were two terminal branches each with a single sequence which was clearly separated (> 2% average divergence from the nearest MOTU) from other sequences but which did not have an assigned BIN. Although the data set did contain other sequences without a BIN assignment (due to the sequence having fewer than <500bp), all other instances grouped with sequences having an assigned BIN.

The collection locations using the three geographic ocean regions were then displayed on these defined tips of the ML trees for each genus separately. Potential transitions between Pacific and Arctic+Atlantic Oceans were mapped onto interior nodes of each phylogeny, excluding the node connecting the ingroup and outgroup, according to the maximum parsimony criterion using Mesquite V 3.04 (Maddison & Maddison 2015). Mesquite assigned each branch a hypothesized geographic region based on the fewest changes that would result in the geographic regions for

the terminal lineages. Nodes which split into two lineages occupying separate geographic regions were pinpointed as likely cases of allopatric cladogenesis. Nodes which split into two identified species or BINs within the same ocean region were indicated as proposed cases of intra-ocean cladogenesis. These cases were tabulated in relation to the total nodes, 69 ingroup nodes across the 11 trees, for investigation of historical dispersal and colonization events between the Pacific and Arctic+Atlantic regions. Recent dispersal among defined terminal lineages are considered and visualized by displaying the geographic assignments at the tips of the trees. In addition, these recent dispersals are also visualized as a whole using Venn diagrams to show the overlap in BIN composition and species composition using the study dataset as well as species composition using the Marine Planktonic Copepod database.

Results

Data set, missing data, and quality control

There were numerous instances of potential misidentifications and/or uncertain phylogenetic placement of morphologically-identified specimens on the COI-5P ML tree. Of the 39 calanoid species investigated, 18 did not have COI-5P sequence data for one or more geographic regions where that species was expected to be found based on available data from the Marine Planktonic Copepod database (Table 4.1). The *Paracalanus* tree (Figure 4.3A) exhibited two instances where different named species were placed within the same genetic group (BIN): *Paracalanus indicus* (3 sequences) and *P. quasimodo* (4 sequences). The sister clade to this *P. indicus*/*P. quasimodo* clade contained specimens bearing both *P. parvus* and *P. indicus* identifications, and collection locations for the sequences in this clade were from various Northern Atlantic European areas. *Tortanus dextrilobatus* and *T. derjugnini* also clustered together in a single BIN; however, a sister lineage represented by a single sequence identified as *T. derjugnini* was also present (Figure 4.3B). *Acartia* (Figure 4.3C) had some instances of suspected specimen misidentifications; there was also a single BIN of *A. hudsonica* which was more closely related to *A. clausi* clades. Among all of the ML trees, there were three recorded instances, in addition to the above-noted suspected misidentifications, of species which were not monophyletic on the COI-5P tree: *Acartia californiensis, Paracalanus parvus,* and *P. indicus.*

These species may be examples of cryptic species, and further investigation is warranted. Because this evidence was not available, geographic collection locations for this work used the current taxonomic names. However, given these issues, patterns were also tabulated using the molecular data alone.

Geographic transitions on the phylogenetic trees

On the 11 phylogenetic trees containing a total of 80 tip lineages, there were 27 interior nodes where an oceanic transition was observed (Figure 4.3; Appendix VII). Of these 27 nodes, there were 16 transitions from the Pacific region to the Arctic+Atlantic region, and there were 5 from the Atlantic and/or Arctic Ocean into the Pacific Ocean region. The direction of the transition was inferred based on the reconstructed region of the ancestral node via parsimony mapping. The direction of the transition was not able to be estimated for 6 of the nodes because of the equal possibility of occurring in more than one ocean region. In addition to the nodes which displayed geographic transitions, there were 43 nodes representing lineage splits within regions, including 24 Pacific lineage splits and 19 cases of lineage splits within the Atlantic region. There were no lineage splits occurring within the Arctic Ocean region.

Twenty four of the 69 total ingroup nodes within these 11 trees were below the Linnaean species level, where ten species were split into two or more BINs. These splits occurred within species belonging to 6 genera: *Paracalanus parvus* (Figure 4.3A), *Tortanus discaudatus* (Figure 4.3B), *Acartia californiensis, A. hudsonica, A. longiremis* (Figure 4.3C), *Centropages abdominalis, C. hamatus, C. typicus* (Figure 4.3D), *Mesocalanus tenuicornis* (Figure 4.3F), *Pseudocalanus newmani* (Figure 4.3I), and *Metridia lucens* (Figure 4.3K). There were eight lineage splits within Linnaean species in the Atlantic, two occurring in *Acartia,* two in *Pseudocalanaus,* one in *Metridia*, and three in *Centropages*. There were five splits within Linnaean species in the Pacific Ocean region, one in each of the genera: *Paracalanus, Acartia, Mesocalanus, Metridia,* and *Centropages.* There were no splits within Linnaean species in the Arctic Ocean region.

Species diversity and complementarity in Pacific, Arctic, and Atlantic Oceans

There were just over twice as many BINs (79) in the data set as there were taxonomically (39) named species (Figure 4.4). Four Linnaean species were distributed in all three ocean

regions, but just a single BIN was so broadly distributed (Table 4.1). The morphological species *Acartia hudsonica*, containing the BIN BOLD:AAJ3150, spanned all three ocean regions (Table 4.1). There were 2222 recorded copepod planktonic species, of which 215 (9.67%) are present in all three ocean regions according to the Marine Planktonic Copepoda data base (Figure 4.5). This percentage of the whole is similar to the 10.5% present in all three ocean regions among those species represented in the molecular data set. By contrast, just 1.5% of taxa were shared among all three oceans considered when examining the BIN distributions for the calanoid molecular data set. For all three ocean regions, BINs revealed more instances of potential endemism in each of the three ocean regions as compared to Linnaean names. In all cases, there was a larger percent connectivity, as indicated by both shared Linnaean species and shared BINs, between the Atlantic and Arctic Oceans than between the Pacific and Arctic.

Discussion

This study has used collections of specimens for 11 calanoid genera from three ocean regions, Pacific, Arctic, and Atlantic, to better understand current distributional patterns and possible shared patterns of dispersal and vicariance resulting from past climatic and geological events. Comparisons were made among genera within the same order and among congeneric species to elucidate the consequences of the opening and reclosure of the Bering Strait on the biogeography and cladogenesis of marine copepods. The number of suspected allopatric cladogenesis events which span the Bering Strait was 21, representing 30% of the total ingroup nodes. The phylogenetic evidence indicated that dispersal was most prevalent in a Pacific eastward direction (16 nodes) as contrasted with the reverse scenario (6 nodes). This indicates that the calanoid species analyzed in this study displayed a general movement from Pacific into Arctic and Northern Atlantic waters, in accordance with the direction of movement for other groups noted from fossil evidence (Vermeij 1991). Results indicate that once this movement across the Bering Strait occurred, the Bering Bridge/Strait acted as a barrier to gene flow between the two ocean populations. In addition to a greater migration from the Pacific eastward, the Pacific Ocean had overall higher species diversity as compared to the Arctic and Atlantic Ocean regions. Both the Pacific and Atlantic Oceans had higher instances of cladogenesis as compared to the Arctic Ocean; indeed, there were no apparent cases of endemic cladogenesis in

the Arctic Ocean. The lack of cladogenesis in the Arctic is likely reflecting the relatively recent colonization or could be due, in part, to undersampling in the region. These findings indicate the important role that the Bering Strait and North Pacific organisms have played in the biological structuring of northern North American marine biodiversity.

Sampling efforts and impact on analysis

The sampling strategies used in my analysis, although not exhaustive for the geographic range studied, were important for two reasons. First, they provide information for 12 specific sites across the investigated range with expertly-identified specimens. Secondly, supplementing these collections with other sequence data from the three ocean regions, I was able to obtain a more complete picture for the eleven evaluated genera, while in some cases linking taxonomy to unlabeled sequence data. However, there were several limitations of my dataset, such as the close physical clustering of the targeted Pacific sampling sites (Chapter 2, Figure 2.1), which would have limited coverage for Pacific North America in relation to the true total diversity. Additionally, Arctic collections were focused in eastern regions of northern North America, and further western Arctic collections would thus be beneficial for examining genetic divergence across the Bering Strait. Finally, increased coverage of species within sampled genera will also assist in our understanding of relationships between species. For example, *Acartia* samples analyzed here represented just over 5% of the total *Acartia* species known globally, and increasing the taxonomic breadth for this and all other sampled genera would increase our understanding of the interrelationships within genera, thereby providing a clearer picture of potential geographic lineage shifts (see Table 4.1 for sampling totals compared to global species totals per genus).

Phylogenetic geographic transitions and the role of the Bering Strait

My results indicate that the Bering Strait has played an important role in structuring the current North Pacific, Arctic, and North Atlantic biodiversity of marine calanoid copepods. Predominantly Pacific to Arctic and/or Atlantic shifts indicate that movement occurred through the Bering Strait from the North Pacific and then east but that gene flow was not maintained, leading to allopatric cladogenesis. For instances where there was a direct transition mapped from Atlantic into Pacific or from Pacific into Atlantic, it is presumed that the transition occurred via

the Arctic Ocean and that this Arctic population was either not sampled or went extinct. Transitions present on the ML trees from Pacific to Atlantic Ocean bypassing the Arctic could also be due to introduction via a different vector. This scenario seems less likely given that the genetic divergences between lineages after these transitions are pronounced and not suggestive of a recent anthropogenic introduction. Following trans-Bering dispersal, Arctic+Atlantic populations diverged in isolation of the Pacific lineages. This pattern has also been found in planktonic protists (Darling *et al.* 2007; Lovejoy 2014).

Results from past studies have shown that there is allopatric divergence of calanoid taxa between Atlantic and Pacific Ocean regions. Carrillo *et al.* (1974) was unable to interbreed Pacific populations of *Acartia clausi* with Atlantic Ocean populations, providing support for the differentiation of these two populations into biological species. Blanco-Bercial *et al.* (2011a) studied two sister species of *Clausocalanus* (Calanoida) using COI-5P data and haplotype networks and showed *Clausocalanus lividus* species had two distinct clusters, one comprised of Pacific haplotypes and the second Atlantic. In addition to the phylogenetic results, Linnaean species compositions for the three ocean regions show more overlap in species composition between Arctic and Atlantic regions than Pacific and Arctic or Atlantic. This higher similarity in species composition is consistent when looking at MOTU results using BINs. And finally, looking at the distribution of all known planktonic copepod species across the region, there is a similar pattern detected as when using the calanoid species and molecular data sets analyzed here. These results further support the conclusion that the Bering Strait has had an impact on holoplankton distributions and has served as a barrier to postglacial recolonization of the Arctic Ocean.

Complex biogeographic patterns such as those displayed by northern calanoid species across the Bering Strait are not uncommon in the Northern Pacific, Arctic, and Northern Atlantic regions. Benthic studies showing higher Atlantic and Arctic connections suggest that areas of glacial refuge have contributed to these patterns for some marine invertebrates, including echinoderms, polychaetes, and molluscs (Hardy *et al.* 2011; Bodil *et al.* 2011; Carr *et al.* 2011; Carr 2012; Layton *et al.* 2014). Geographic distribution patterns among species within a single genus have also been similarly reported for freshwater planktonic crustaceans. Adamowicz *et al.* (2009) investigated *Daphnia* species across large (inter-continental) spatial scales and indicated

that the highly dispersive genus also exhibited a large number of allopatric divergences, leading to a global scale for many cladogenesis events.

The large-scale biogeographical patterning for marine zooplanktonic groups is influenced not only by the relatively shallow water access of the Bering Strait but also movement through the Arctic Ocean. Movement of species through the Arctic Ocean is disrupted by other geographic barriers such as the underwater Lomonosov Ridge, which divides the two underwater Arctic basins, Eurasian and Canadian (Kosobokova *et al.* 2011).

Along with water currents and Arctic Ocean access, either through the Bering Strait from the Pacific or from the Atlantic, biological traits are also contributors to the success of calanoid copepods in the Arctic region. Biological traits are especially important in light of challenging environmental conditions in northern and Arctic regions. Copepod vertical migration in the water column has been linked to success in colder climates with surface ice cover during winter months (McLaren 1974; Kaartvedt 1996). Vertical structure of zooplankton diversity has been documented in the Arctic Ocean (Grainger 1965; Kosobokova *et al.* 2011), which may additionally contribute to variability among species in their exposure to passive dispersal routes via oceanic currents. With variation in temperature and the melting of seasonal ice, salinity variation is an environmental stressor that may also serve as a barrier to colonization of the Arctic by some more southerly species (Gradinger 2001; Rochet & Grainger 1988).

The successful colonization of calanoid species in the Arctic region, with its unique environmental factors, has been linked to a species's egg production and juvenile survivorship (McLaren *et al.* 1969). Halvorsen (2015) revealed a positive relationship between egg initial lipid content and total egg production and survivorship for the calanoid species *Calanus hyperboreus*, suggesting that a strong ability to store lipids could be required for successful reproduction in cold waters. Another strong factor in reproductive success is the timing of the development of the eggs into the adult form. Hirche (2013) indicated that the timing of reproductive activity is important in calanoid survivorship in colder conditions, and species with delayed development of eggs in response to environmental conditions are more successful in the quickly changing Arctic environment. The overall complex patterns, both vertically in the water column as well as biogeographically, seen in the Arctic region are influenced by many factors including physical connection to Pacific and Atlantic Oceans, water currents through the region, and physiological challenges (such as varying salinity, temperature, and food availability). With

the increasing temperatures in the northern region, the traits involved in responding to environmental cues, which enable species to be successful Arctic inhabitants today, may not be as effective in the future. Further investigation into the influence of trait variability on the observed biogeographic patterns and extinction risk in the face of climate change is beyond the scope of this work but is suggested for future study.

The complex Arctic biogeographic patterns detected here and in other studies of northern marine fauna are also present in the Arctic marine flora. Dunton (1992) had suggested an Arctic paradox, where the fauna had Pacific origins, while the Arctic flora was more closely related to that of Atlantic waters. My results do not support this hypothesis and instead indicate an important role for historical migration from Pacific waters through the Bering Strait, followed by allopatric speciation, with more recent dispersal events displaying greater Arctic+Atlantic ties. Interestingly, recent work looking at a near-shore macro alga, *Fucus distichus*, has indicated very similar patterns to those of marine benthic organisms, with stronger Arctic+Atlantic relationships, with suspected regions of glacial refugia within the Northern Atlantic and Arctic regions (Laughinghouse *et al.* 2015).

While the patterns for all of these groups are influenced by many factors, including past glacial impacts, a general pattern does exist for holoplankton as well as benthic organisms with planktonic larvae. Arctic and North Atlantic holoplankton tend to have Pacific origins (Darling *et al.* 2007; Blanco-Bercial *et al.* 2011a; Lovejoy 2014; results presented here); as well, fossil evidence for some marine molluscs indicates migration from the Pacific into the Arctic during interglacial periods (Vermeij 1991; Dyke *et al.* 1996). Many marine invertebrate taxa, including several examples of molluscs (Layton *et al.* 2014, 2016), have a stronger contemporary connection between Atlantic and Arctic Oceans as indicated by patterns of shared species or MOTUs (polychaetes, Carr *et al.* 2011; Carr 2012; echinoderms, Hardy *et al.* 2011; bryozoans, Kuklinski *et al.* 2013; cnidarians, Hotke 2015). The main difference between these two groups is that the benthic organisms are only present in the plankton for a portion of their life, while holoplankton spend their entire life in the water column. Although the time spent in the water column is different between benthic and holoplankton species, there was still substantial similarity in the overall biogeographic patterns for these two groups in the Arctic region.

Regional diversity and taxonomic trends among calanoid copepods

There was no apparent overall relationship between phylogeographic patterns and specific higher Calanoida taxa. When the geographic transitions are mapped onto the Calanoida phylogenetic tree, there do not appear to be particular higher taxa in which transitions from Pacific to Arctic-Atlantic Ocean regions are more prevalent than in other lineages. Differing distributional patterns among closely-related taxa have been recorded for marine copepods before, showing high gene flow across geographic regions in one species and phylogeographic clustering for the other (Blanco-Bercial *et al.* 2011a). Evidence of phylogeographic variation between two closely-related species has also been noted in benthic invertebrates. Layton *et al.* (2016) studied two marine bivalve species, *Hiatella arctica* and *Macoma balthica*, with similar larval dispersal and noted that *H. arctica* had less geographic structure based on COI-5P data than did the *M. balthica*. Layton *et al.* (2016) used mollusc collections from across Pacific, Arctic, and Atlantic waters and hypothesized that some of the structure was most likely due to regions of glacial refugia during past glacial cycles. During the last glacial retreat, variation in glacial melting and water flow in the North Pacific, Arctic, and North Atlantic could have resulted in differing distributional patterns for sister species strictly by chance. More recently, random events facilitating gene flow from one ocean region to another may also be occurring via human-mediated means.

Molecular operational taxonomic units and morphological identifications

Aside from the general trend that there are more BINs than there are morphological species, indicating potentially cryptic species, there were several taxa with problematic BIN assignments as compared to the phylogenetic trees. There were three identified groups displaying such problems, *Paracalanus indicus/quasimodo/parvus* complex, *Pseudocalanus newmani* complex, and the *Tortanus dextrilobatus/derjugnini* complex. In addition to these specific examples, there were two genera with morphologically-named species displaying polyphyly on the ML trees (*Paracalanus parvus, Acartia californiensis, A. clausi,* and *A. hudsonica*). The phylogenetic clustering of the *Pseudocalanus newmani* specimens was conflicting between morphological species names and molecular groupings. Phylogenetic analysis for two of these three groups using the COI-5P molecular barcode region has been previously noted to provide mixed placement on phylogenetic trees with respect to morphological identifications (*Acartia:*

Blanco-Bercial *et al.* 2014; *Paracalanus*: Blanco-Bercial *et al.* 2014). *Pseudocalanus* has been reported to have phylogenetic placement consistent with current taxonomy using several different molecular markers including COI-5P, CytB, and ITS-1 (Aarbakke *et al.* 2014). The results of my analysis using COI-5P show poor consistency of placement for *Pseudocalanus* into groups congruent with current taxonomy, suggesting the possibility of morphological species misidentifications in my data set.

Due to the difficult taxonomy for these taxa, further sampling across the geographic region, amplification of additional molecular markers, and reconsideration of morphological identifications are necessary to provide further evidence of species diversity and evolutionary species boundaries in this group. COI-5P sequence data have been shown to provide reliable species identifications for many copepod species (Bucklin *et al.* 2011). However, additional information may be required when conducting a biogeographic investigation of calanoid species. Additional molecular markers could include CytB and ITS-1, previously used in comparative phylogenetic analyses of *Pseudocalanus* species (Aarbakke *et al.* 2014). The combination of both mitochondrial and nuclear evidence can also elucidate important relationships among species such as potential introgression of mitochondrial genomes (Ballard & Whitlock 2004).

Additional data in the form of alternative molecular evidence may be required to provide further information about population genetic structure and dispersal history not evident from the more conserved molecular data (such as the COI-5P region). Other molecular techniques such as single nucleotide polymorphism (SNP) and Restriction Fragment Length Polymorphism (RFLP) analyses may provide more information between populations of a single species. SNP's have been successfully used for studies of calanoid copepods (Unal & Bucklin 2010). RFLP analyses have been successfully used to distinguish between species within the copepod genera *Clausocalanus* (Blanco-Bercial & Álvarez-Marqués 2007) and *Paracalanus* (Jagadeesan *et al.* 2009). More recent analyses using high-throughput sequencing technologies, like restriction site-associated DNA (RAD) sequencing, have also proven valuable in discerning population genetic structure for calanoid species (Blanco-Bercial & Bucklin 2016). With increased data, additional biogeographic analysis methods can also be implemented using programs like the Reconstruct Ancestral State in Phylogenies (RASP) program (Yu *et al.* 2015). In this program, a Bayesian approach to Statistical Dispersal-Vicariance Analyses can be implemented in which biogeographic reconstructions are averaged over highly probable Bayesian tree constructions.

Conclusions

Planktonic species across the Pacific, Arctic, and Atlantic regions have similar biogeographical patterns when compared to benthic and near-shore marine species. Variations in the biogeographical patterns are likely due to the increased time in the water column for holoplankton species versus benthic species, facilitating colonization of new areas via ocean currents. However, based on my results, simply being a holoplankton species is not enough to predict the biogeographic patterns present across Northern Pacific, Arctic, and Northern Atlantic Ocean regions.

With warming temperatures leading to less northern ice cover and increased accessibility of the region, there is a need for increased studies of the North Pacific, Arctic, and North Atlantic Ocean regions. Understanding current regional biodiversity patterns and past biogeographic influences on the region can further our understanding and potentially provide valuable evidence to predict and understand future biotic changes. It is clear that future research looking into the patterns of marine planktonic copepods should include molecular evidence. My results indicated that the COI-5P data estimated a much larger number of potential species as compared to morphology, with a corresponding higher level of genetic isolation as seen by higher endemism in the three ocean regions. Future research to increase taxonomic breadth and sampling effort for Calanoida and other planktonic groups may perhaps resolve some of the taxonomic difficulties and paint a better picture of the biogeography of marine plankton.

Tables and Figures

Table 4.1 Collected COI-5P sequence data for 11 genera of Calanoida (Copepoda). Table of presences and absences for sequence data of each morphologically-identified species used. The numbers in brackets following the genus names indicate the number of species used in this study followed by the total number of marine species present in the genus. Sequence data were grouped into three separate categories, Arctic (AR), Pacific (PA), and Atlantic (AL). A checkmark indicates that a sequence for that geographic region is in the data set. An x indicates that the species is expected for that geographic region (based on data from the Marine Planktonic Copepod database, http://copepodes.obs-banyuls.fr/en) but is not yet represented by sequence data in public databases or in my Chapter 2.

Superfamily	Family	Genus	Species	Sequence Data		
				PA	AR	AL
Augaptilidae	Metridinidae	*Metridia* (3/25)	*M. effusa*			✓
			M. lucens	✓	✖	✓
			M. pacifica	✓	✖	
Centropagidae	Acartiidae	*Acartia* (4/65)	*A. californiensis*	✓		
			A. clausi	✓		✓
			A. hudsonica	✓	✓	✓
			A. longiremis	✓	✓	✓
	Centropagidae	*Centropages* (3/34)	*C. abdominalis*	✓	✓	
			C. hamatus	✖	✖	✓
			C. typicus	✖		✓
	Temoridae	*Eurytemora* (4/26)	*E. affinis*	✖	✖	✓
			E. carolleeae			✓
			E. herdmani		✓	✓
			E. lacustris		✖	✓
		Temora (2/5)	*T. discaudata*	✓		✖
			T. longicornis		✖	✓
	Tortanidae	*Tortanus* (6/40)	*T. derjugini*	✓		
			T. dextrilobatus	✓		
			T. discaudatus	✓	✓	✓
			T. gracilis	✓		✖
			T. komachi	✓		
			T. vermiculus	✓		
Calanidae	Calanidae	*Calanus* (3/17)	*C. glacialis*	✓	✓	✖
			C. hyperboreus		✓	✓
			C. pacificus	✓		
		Mesocalanus (1/2)	*M. tenuicornis*	✓		✓
	Paracalanidae	*Paracalanus* (5/15)	*P. aculeatus*	✓		✓
			P. indicus	✓		✓
			P. parvus	✓	✓	✓
			P. quasimodo	✖		✓
			P. denudatus	✖		✓
Clausocalanidae	Aetideidae	*Aetideus* (3/12)	*A. armatus*	✖		✓
			A. bradyi	✖		✓
			A. divergens	✓		
	Clausocalanidae	*Pseudocalanus* (5/7)	*P. acuspes*	✖	✓	✓
			P. elongatus	✖	✖	✓
			P. minutus	✖	✓	✓
			P. moultoni	✖	✓	✓
			P. newmani	✓	✖	✓

Figure 4.1 Region boundaries used for assigning specimens to one of three ocean regions. The solid red line provides the border between Pacific and Arctic as well as Atlantic and Arctic ecoregions as per Spalding *et al.* (2007) but with the line adjusted to traverse the Bering Strait. The red line was used to place molecular data into ocean regions. The broken yellow line provides the boundaries for the regions as searched on the Marine Planktonic Copepod database (http://copepodes.obs-banyuls.fr/en). These regions were used for the expected sites in Table 4.1 and the copepod overlapping ranges in Figure 4.5.

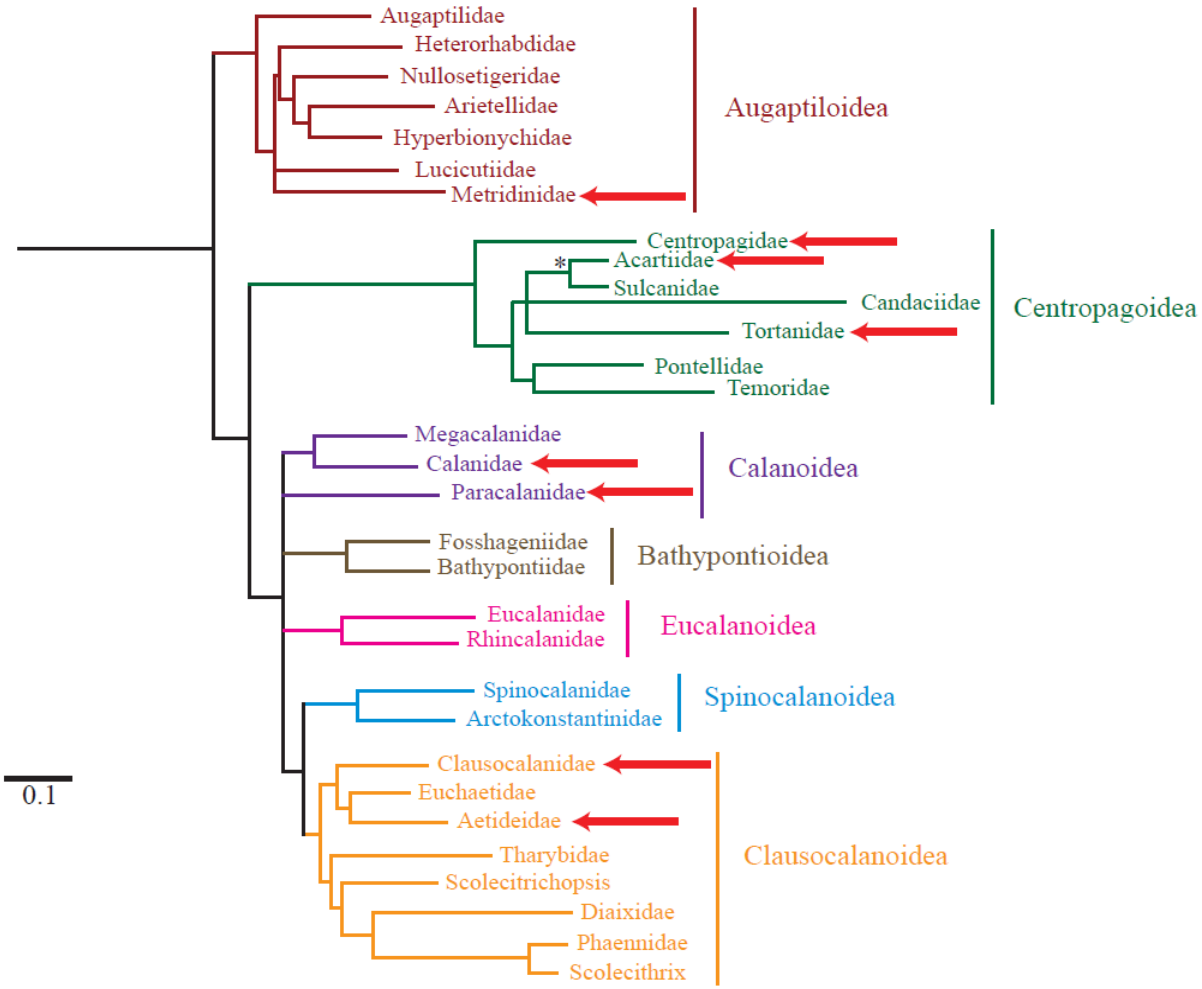
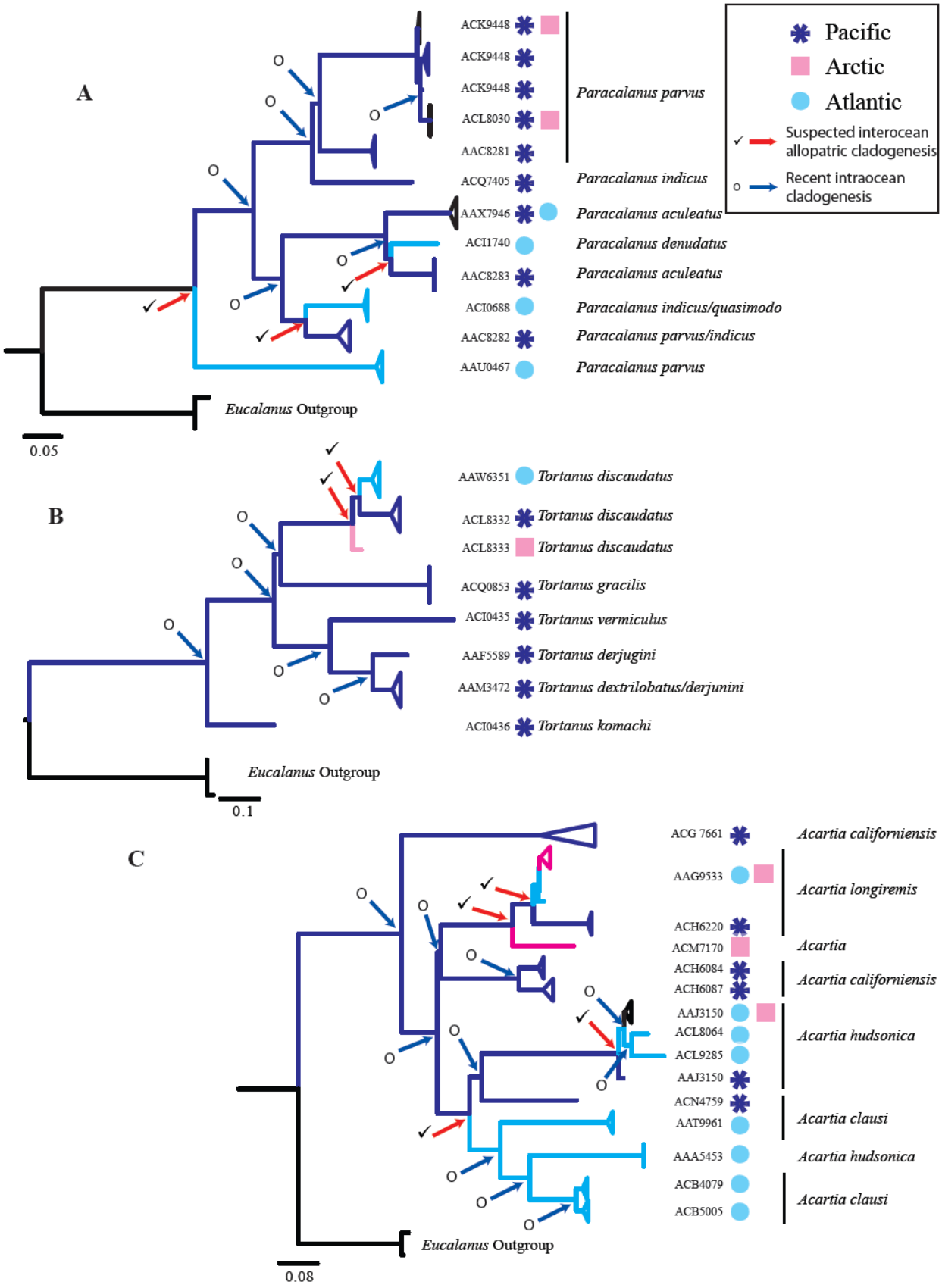
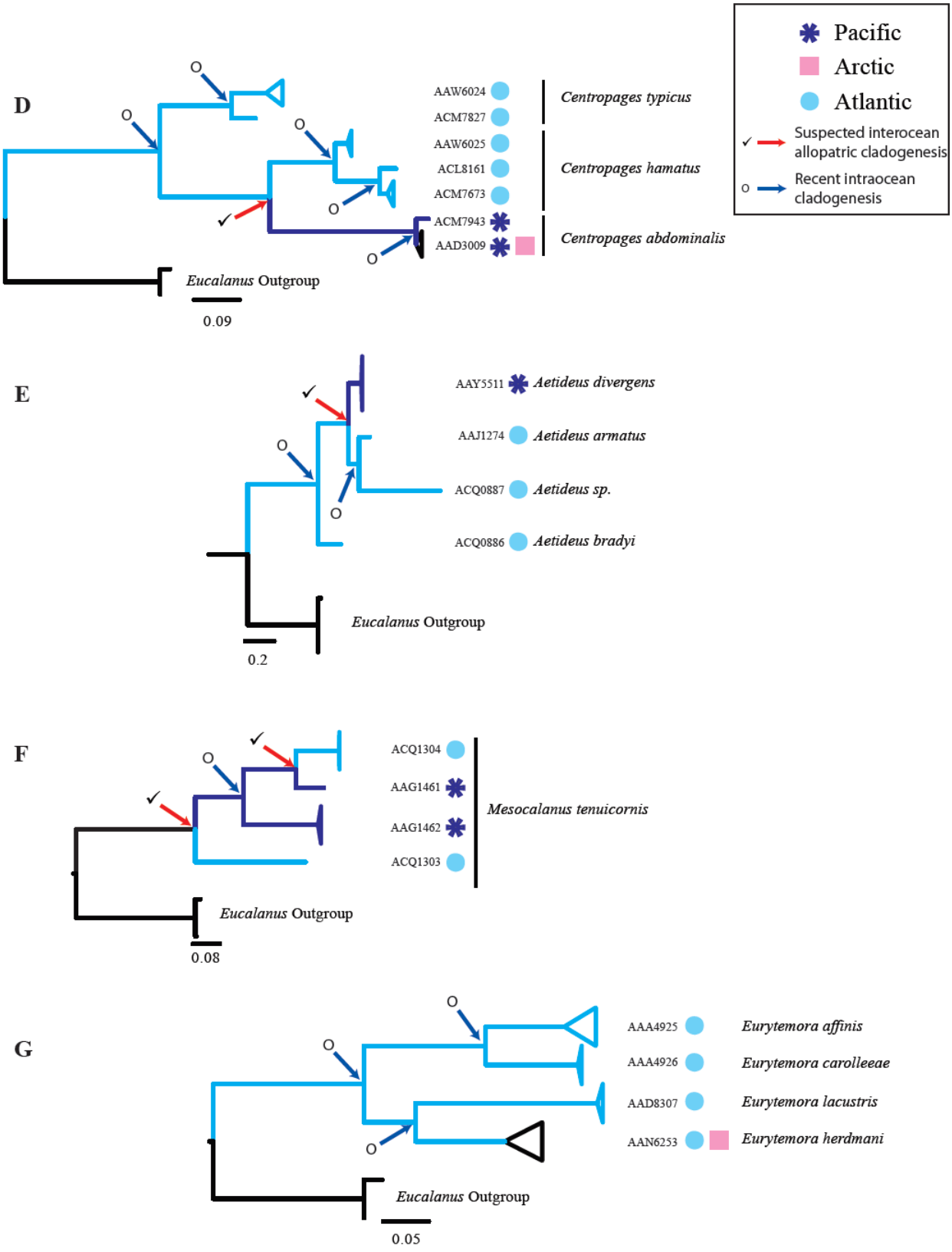


Figure 4.2 Calanoida phylogenetic tree. Tree showing families within the order Calanoida based on maximum likelihood phylogenetic analysis using nuclear large (28S) and small (18S) subunits ribosomal RNA (rRNA) gene sequences and the mitochondrial protein-coding genes cytochrome c oxidase subunit I (COI) and cytochrome b (Cyt b) (re-drawn from Blanco-Bercial *et al.* 2011b). Red arrows indicate families containing species included in this study. *Acartiidae was not included in the Blanco-Bercial *et al.* (2011b) analysis, and the Acartiidae family was placed here based on cladistic analysis of morphological data (Bradford-Grieve *et al.* 2010). Acartiidae was placed as a sister to the Sulcanidae (included in both Blanco-Bercial *et al.* (2011b) and Bradford-Grieve *et al.* (2010)), and the branch lengths of both families were kept equal in length to the branch length of Sulcanidae (Blanco-Bercial *et al.* 2011b).


A
ACK9448
ACK9448
ACK9448
ACL8030
AAC8281
ACQ7405
AAX7946
ACI1740
AAC8283
ACI0688
AAC8282
AAU0467
Paracalanus parvus
Paracalanus indicus
Paracalanus aculeatus
Paracalanus denudatus
Paracalanus aculeatus
Paracalanus indicus/quasimodo
Paracalanus parvus/indicus
Paracalanus parvus
Eucalanus Outgroup
0.05
Pacific
Arctic
Atlantic
Suspected interocean allopatric cladogenesis
Recent intraocean cladogenesis
B
AAW6351
ACL8332
ACL8333
ACQ0853
ACI0435
AAF5589
AAM3472
ACI0436
Tortanus discaudatus
Tortanus discaudatus
Tortanus discaudatus
Tortanus gracilis
Tortanus vermiculus
Tortanus derjugini
Tortanus dextrilobatus/derjunini
Tortanus komachi
Eucalanus Outgroup
0.1
C
ACG 7661
AAG9533
ACH6220
ACM7170
ACH6084
ACH6087
AAJ3150
ACL8064
ACL9285
AAJ3150
ACN4759
AAT9961
AAA5453
ACB4079
ACB5005
Acartia californiensis
Acartia longiremis
Acartia
Acartia californiensis
Acartia hudsonica
Acartia clausi
Acartia hudsonica
Acartia clausi
Eucalanus Outgroup
0.08


D
AAW6024
ACM7827
AAW6025
ACL8161
ACM7673
ACM7943
AAD3009
Centropages typicus
Centropages hamatus
Centropages abdominalis
Eucalanus Outgroup
0.09
Pacific
Arctic
Atlantic
Suspected interocean allopatric cladogenesis
Recent intraocean cladogenesis
E
AAY5511
AAJ1274
ACQ0887
ACQ0886
Aetideus divergens
Aetideus armatus
Aetideus sp.
Aetideus bradyi
Eucalanus Outgroup
0.2
F
ACQ1304
AAG1461
AAG1462
ACQ1303
Mesocalanus tenuicornis
Eucalanus Outgroup
0.08
G
AAA4925
AAA4926
AAD8307
AAN6253
Eurytemora affinis
Eurytemora carolleeae
Eurytemora lacustris
Eurytemora herdmani
Eucalanus Outgroup
0.05

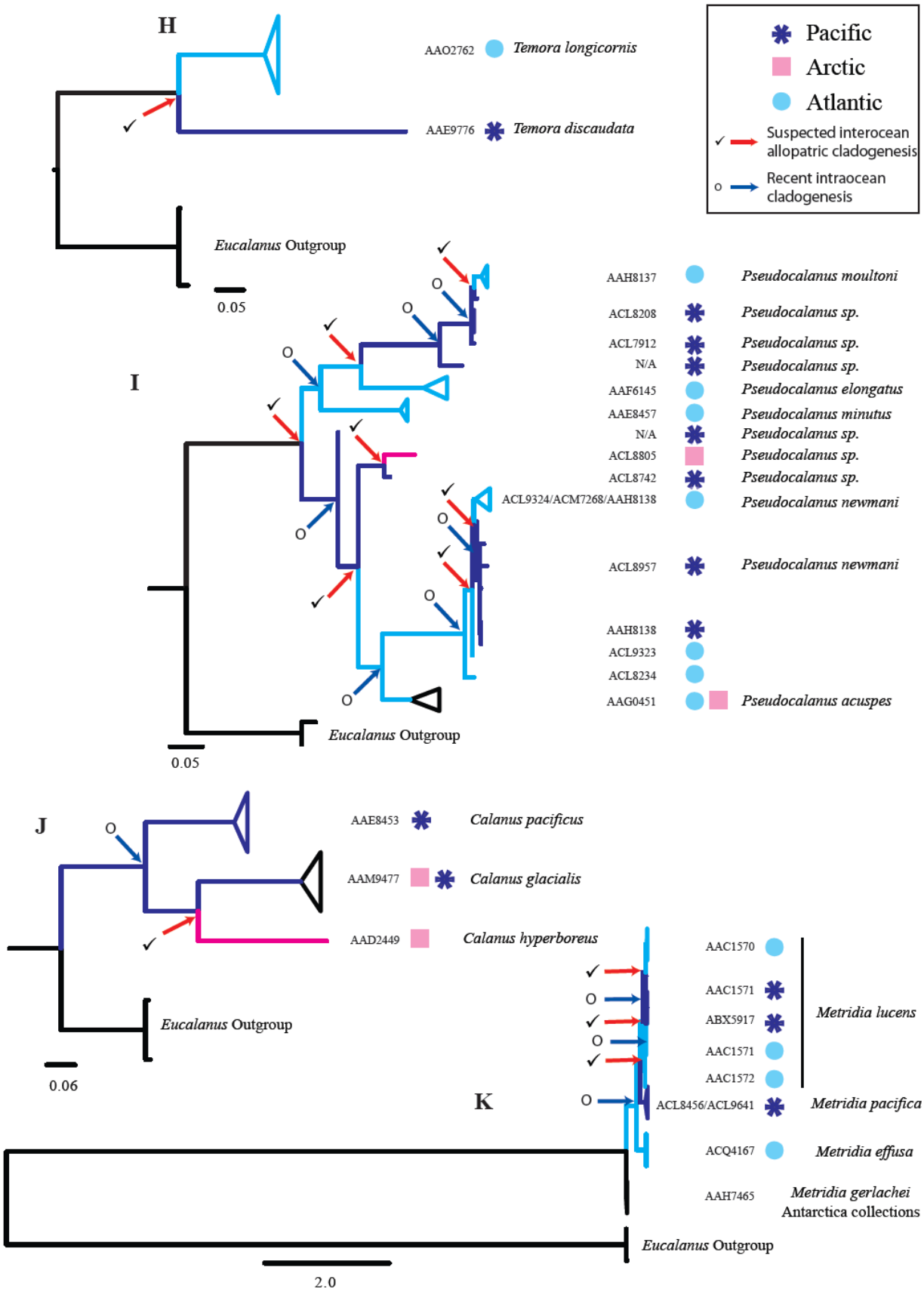
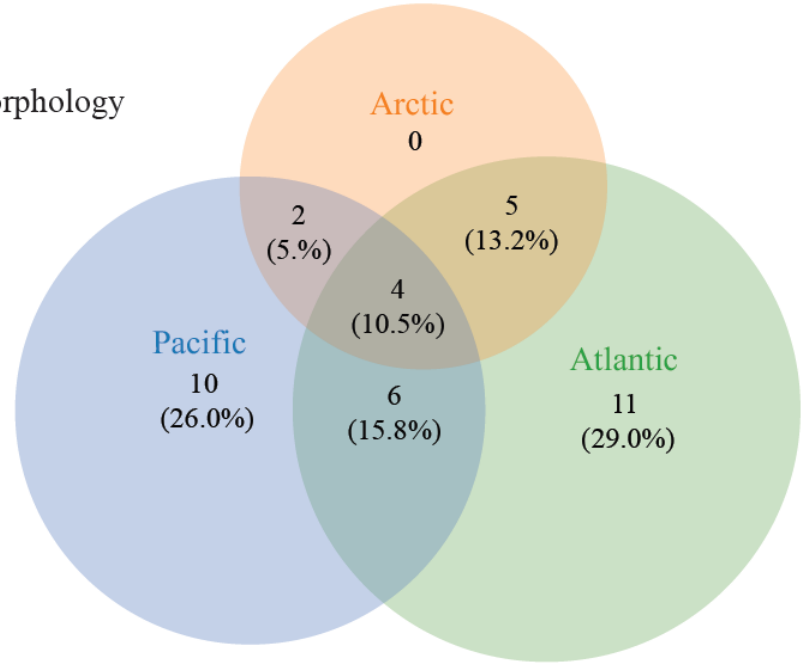

H
AAO2762
Temora longicornis
AAE9776
Temora discaudata
Eucalanus Outgroup
0.05
Pacific
Arctic
Atlantic
Suspected interocean allopatric cladogenesis
Recent intraocean cladogenesis
I
AAH8137
Pseudocalanus moultoni
ACL8208
Pseudocalanus sp.
ACL7912
Pseudocalanus sp.
N/A
Pseudocalanus sp.
AAF6145
Pseudocalanus elongatus
AAE8457
Pseudocalanus minutus
N/A
Pseudocalanus sp.
ACL8805
Pseudocalanus sp.
ACL8742
Pseudocalanus sp.
ACL9324/ACM7268/AAH8138
Pseudocalanus newmani
ACL8957
Pseudocalanus newmani
AAH8138
ACL9323
ACL8234
AAG0451
Pseudocalanus acuspes
Eucalanus Outgroup
0.05
J
AAE8453
Calanus pacificus
AAM9477
Calanus glacialis
AAD2449
Calanus hyperboreus
Eucalanus Outgroup
0.06
K
AAC1570
AAC1571
ABX5917
AAC1571
Metridia lucens
AAC1572
ACL8456/ACL9641
Metridia pacifica
ACQ4167
Metridia effusa
AAH7465
Metridia gerlachei
Antarctica collections
Eucalanus Outgroup
2.0

Figure 4.3 Maximum likelihood phylogenetic trees for each of 11 Calanoida genera collected from Pacific, Atlantic, and Arctic Ocean regions. Specimens were assigned to three collection regions: Pacific (purple star), Atlantic (light blue circle), and Arctic (pink square). Branches on the trees are also colourized based on each lineage's mapped geographic range based on maximum parsimony. Branches were left black where the geographic region was uncertain. The 7-character codes present in front of the collection location symbols at the tips of the trees are unique identifier Barcode Index Numbers; for searching for the BINs in BOLD, these need to be prefaced by "BOLD:". Each genus is shown in a separate panel: A) *Paracalanus,* B) *Tortanus,* C) *Acartia,* D) *Centropages,* E) *Aetideus,* F) *Mesocalanus,* G) *Eurytemora,* H) *Temora,* I) *Pseudocalanus,* J) *Calanus,* K) *Metridia.* Red arrows with a ✓ represent suspected allopatric cladogenesis between ocean regions; blue arrows with an O indicate mapped intra-ocean cladogenesis. Scale bars indicate nucleotide substitutions per site according to the selected model of nucleotide evolution and are located below each of the trees.

A. Morphology



B. BINs

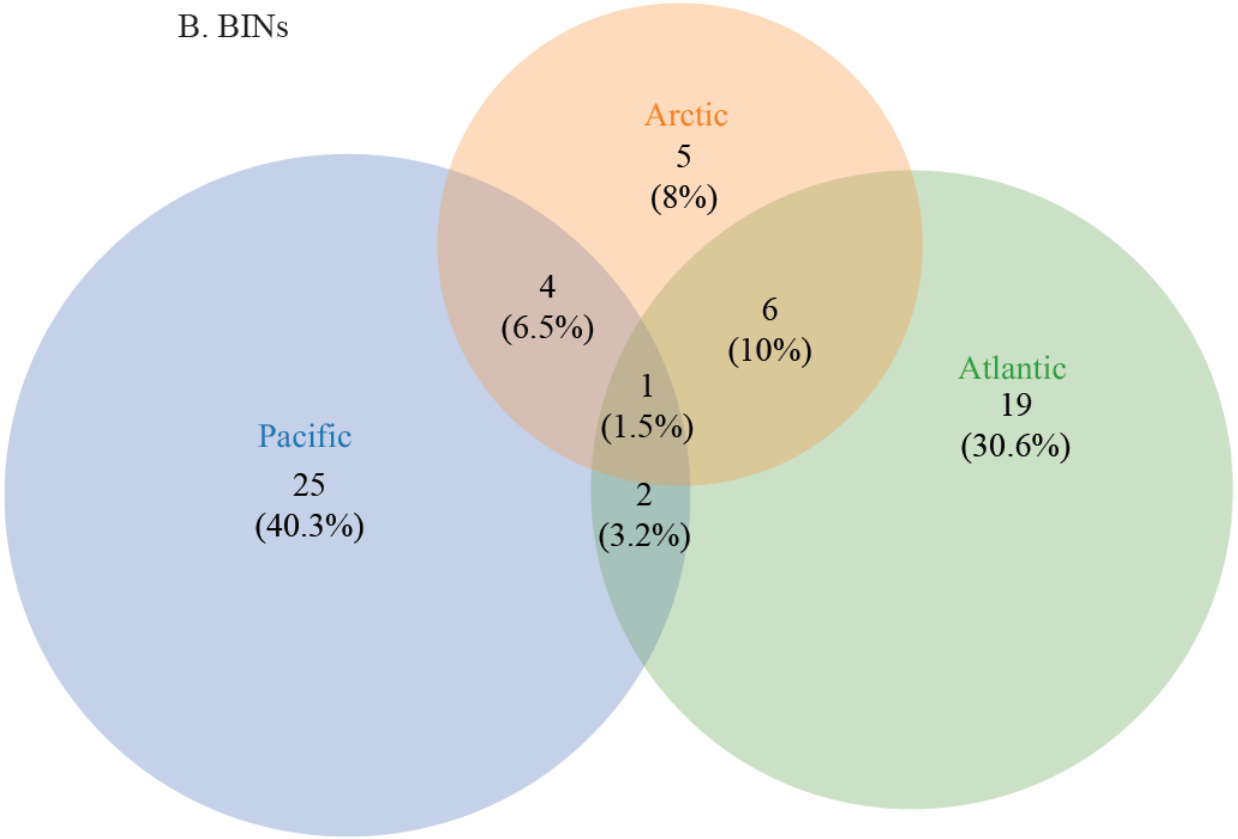


Figure 4.4 Calanoida species and BIN presence across the three ocean regions: Pacific, Arctic, and Atlantic. The two Venn diagrams represent the amount of overlap of species composition in the three ocean regions, using the boundaries indicated by the red line in Figure 4.1. **A.** Displays Linnaean species identifications of 11 genera of marine Calanoida copepods. **B.** Uses the same data set as panel A; however, the numbers represent the number of BINs for each geographic collection region.

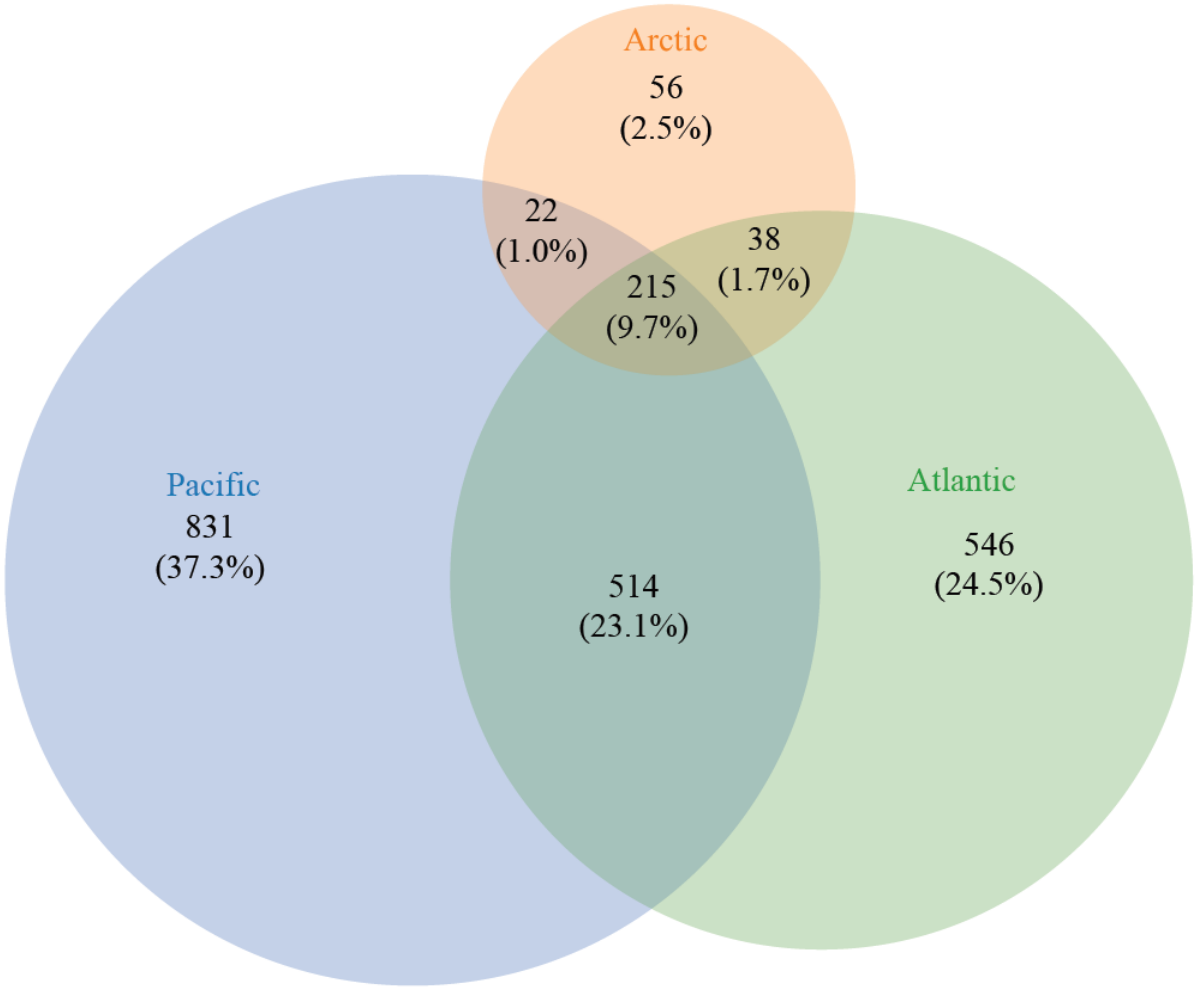


Figure 4.5 Planktonic copepod species presence across the three ocean regions: Pacific, Arctic, and Atlantic. Venn diagram that represents the amount of overlap of planktonic Copepoda species composition in the three ocean regions using data collected from the Marine Planktonic Copepod website (http://copepodes.obs-banyuls.fr/en/). The yellow dotted line on Figure 4.1 portrays the boundaries used for the ocean regions in that database.

Chapter 5
General Conclusions

My thesis research provides valuable information on the biological diversity present in Canadian marine waters by sampling a variety of taxa from 12 locations in all three Canadian ocean regions: Pacific, Arctic, and Atlantic. This work is important as it contributes information toward answering the question of what diversity is present in Canadian marine plankton. My work contributes over 250 expertly-identified specimens and includes molecular data for two genes widely used for DNA barcoding or metabarcoding: one mitochondrial gene fragment (COI-5P) and one nuclear gene fragment (18S-V4). These molecular data, which are stored in publicly available databases, provide accessible information that is essential for conducting future species-level identifications using molecular evidence and for applications such as biomonitoring via metabarcoding. Among the collected data, I provide two sites not previously sampled for molecular evidence of marine zooplankton. My work provides new knowledge of species distributions and connectivity between collection locations. Specifically, my work provides crucial knowledge about the species composition in 11 ports and one proposed port across Canada. With the increase in shipping traffic and human-mediated species movements of small planktonic species, these records are essential to understand the current level of diversity and species distributions in Canadian waters. With this knowledge, we will have a better understanding of the species present across the region, which can contribute to our knowledge of ecosystem structure as a basis of comparison to monitor future changes. With this body of work, I have presented several key, unique contributions to three distinct areas within marine plankton research.

First, I have made a contribution to methodological development for studying marine planktonic crustaceans. Past research conducted on marine zooplankton in North Pacific, Arctic, and North Atlantic regions has been either geographically or taxonomically constrained (for example, Carrillo *et al.* 1974; Bucklin *et al.* 2003; Blanco-Bercial *et al.* 2011a; Laakmann *et al.* 2012). In my Chapters 2-4, I address this gap by looking at two subclasses of planktonic multicrustaceans across all three of Canada's ocean regions. I first described the patterns of genetic divergence for the marine multicrustacean groups Copepoda and Thecostraca. I then

identified a COI-5P global pairwise sequence divergence threshold between 2.1 and 2.6% for use in species delimitation for marine planktonic multicrustaceans. The obtained range of this threshold is interesting in that it is very similar to ranges previously studied for marine species (Bucklin *et al.* 2011; Carr et al. 2011; Layton 2013; Layton *et al.* 2014). This narrow range of divergence across broad marine life is promising for using standardized methods in estimating species diversity across a wide range of taxa.

The obtained range between 2.1 and 2.6% GPSD among MOTU-generating methods used in this work was often referred to as the 'optimal' threshold for the data set. This term optimal may be subjective; however, here it refers to an internal validation of the data where the sequence data are neither over split nor over lumped (Handl *et al.* 2005), as discussed in Chapter 2. Although my choice of the 'elbow' method provided what I referred to as the 'optimal' point to cluster the data, there are other methods to evaluate the effectiveness of clustering data. Other internal methods are described in depth in Handl *et al.* (2005) and the references therein. Optimization of clustering parameters in comparison with external references has also been conducted using various methods. For example, Renaud *et al.* (2012) determined the optimal threshold for a group of flies by comparing MOTUs to taxonomic identifications using cumulative error rates per threshold divergence, weighting false positives and false negatives equally; Lefébure *et al.* (2006) used a Bayesian phylogenetic approach where family, genus, and species taxonomic levels were used determine the distribution of the pairwise distances, and the selected threshold represented that expected to provide the 'best compromise' for molecular data to match to taxonomic identifications; and Handl *et al.* (2005) summarized numerous different methods to validate clustering results using both internal analyses and external validations.

Additionally, I have provided explicit analyses using concordance metrics to assess the relative success of a molecular marker to groups specimens into known taxonomic groups (Wallace 1983; Pinto *et al.* 2007). Using Adjusted Wallace and Adjusted Rand concordance analyses, I demonstrate the importance of having quantitative values to evaluate how well DNA barcode-based clusters match to current taxonomy. Using these concordance metrics, I provide evidence supporting the use of barcodes within Copepoda and Thecostraca; my results indicated that, depending on the analysis method used, COI-5P clusters yielded 40-80% exact placement of specimens into a single Linnaean species, with the majority of the remaining cases involving splits (rather than lumps or mixes). These features bode well for rapid, barcode-based

identification of multicrustacean specimens. I highlighted the variation in richness between the molecular operational taxonomic units (MOTU) using the COI-5P data and morphologically-identified Linnaean species. There was a higher number of MOTUs than morphologically-identified species.

The second research area to which my work contributes is that of the rapid assessment of zooplankton biodiversity. In this work I have provided an evidence-based evaluation of the 18S-V4 molecular marker using a verified molecular dataset of taxonomically-identified material. The 18S-V4 marker is an increasingly-used marker for high-throughput sequencing projects of marine plankton (Wu *et al.* 2015). I found that the 18S-V4 molecular region was not able to accurately group specimens into morphological species for Copepoda and Thecostraca. I also found that using the 18S-V4 region for metabarcoding would underestimate the species-level biological diversity by 5-15% in comparison with COI-5P.

The final research area to which my work has contributed has been in the greater understanding of the biogeographical patterns present among marine plankton across Canadian marine waters. Previous studies of this nature have used only single or small numbers of taxa (Blanco-Bercial *et al.* 2011a) or have been confined to relatively small geographic areas (Bucklin *et al.* 2010a). In contrast, the analyses conducted here were completed across large geographic scales and on a broad range of taxonomic diversity. Using 11 genera and 39 Linnaean species of Calanoida (Copepoda), I assessed the degree to which the phylogenies of the genera displayed shared biogeographic patterns. The calanoid genera displayed varying levels of biogeographic structure both between species and within species using the COI-5P sequence data. However, the preponderance of the data indicates the historical dispersal of calanoid copepods from the Pacific into the Arctic and Atlantic Oceans. This initial dispersal was followed by the apparent allopatric divergence of species on either side of the Bering Strait. This work contributes to our understanding of the historical events that have led to the current distribution of marine copepods, as well as a better understanding of the degree to which copepod communities in Canada's three ocean regions are connected.

The work accumulated in this thesis is intellectually novel through the approaches taken to addressing my study questions and hypotheses, as well as the outcomes obtained. My use of an elbow analysis to determine if there was sufficient information for species delimitation in the COI-5P molecular region was the first explicit use of an internal test (Handl *et al.* 2005) for

clustering DNA barcode data. By contrast, most other barcoding studies apply a method for generating MOTUs and then compare the MOTUs to Linnaean species *a posteriori*; my development of this approach can therefore contribute to the further study of poorly-known faunas for which such external tests are not possible. My use of this methodology is unique in that it provides a simple method that can be used with very large data sets, unlike previous methods, which were more specific to the molecular region or were more complex or computationally intensive (Lefébure *et al.* 2006; Wu *et al.* 2015).

The elbow analysis used a range of pairwise sequence divergence thresholds to isolate the optimum threshold to delimit species. The elbow analysis provided a succinct pairwise sequence divergence at which the specimens could be best placed into species-like groups using a similarity-based method of delimitation and COI-5P sequence data. This elbow analysis can be applied to see if there is sufficient information in any molecular marker. The outcome of applying this methodology contributed two sets of information toward marine copepod research. Firstly, it provided verification that the COI-5P contained enough data to delineate specimens into clusters. Secondly, it provided a more specific similarity threshold (2.1-2.6% pairwise sequence divergence, depending upon the analysis method) than previously reported in the literature (2-3% pairwise sequence divergence) (Radulovici *et al.* 2010; Blanco-Bercial *et al.* 2014).

Conducting a check on the performance of molecular data to accurately place specimens into species-like groups contributes to our overall understanding of species boundaries. My analyses provide an example for future work to gain a better understanding of the genetic variation among species using multiple lines of evidence and differing species concepts. Assessing "true" species boundaries is difficult, and boundaries can differ according to the preferred primary or secondary species concept, as well as the selected character system (Mayden 1997; de Queiroz 2007). This difficulty is increased when studying geographically widespread and taxonomically diverse groups such as the Multicrustacea. By quantifying concordance, we can examine and compare the signal for various delimitations emerging from different character types and analysis methods. The concordance assessments provide more information than simply reporting failures when molecular clusters do not agree with morphological species. This extra information can support existing morphological species

boundaries as evolutionary species through a new character system or provide new biological insights (e.g. into potential cryptic species prevalence) in cases of discordance.

My thesis work contained several possible sources of error. The most likely possible source of error for my project is in the potential misidentification of specimens. This is of particular concern when mining publicly available databases, where misidentifications are common, and the source of the identifications can oftentimes be difficult to ascertain, let alone verify (Vilgalys 2003). The chance that possible misidentifications have significantly impacted the results of my larger analyses of genetic divergence patterns seems less likely, primarily because I obtained all available sequences at the class level. Misidentifications of specimens at this level are less likely because the morphological characters are more pronounced. Furthermore, if there were a few errors in these larger datasets, it is probable that the noise of these few sequences did not alter the outcomes from my analyses using the very large data sets.

The use of material misidentified at lower taxonomic levels in my analysis and for my smaller datasets poses greater concerns in influencing the outcomes of my project. All possible efforts were made to reduce this possibility, including sequence verification steps such as checking identifications using translated alignments and modeled secondary structure alignments. For future work I would suggest and encourage additional steps to verify identifications such as using additional molecular markers. As well, "voucher recovery" protocols (e.g. Porco *et al.* 2010) could possibly be further developed for copepods, or very small specimen fragments used for DNA barcoding, such that post-sequencing taxonomic re-identification could be conducted. Misidentifications for many of my analyses for Chapters 2-4 present minimal impacts on my conclusions, because these conclusions were not exclusively dependent on morphological identifications. Nevertheless, my concordance values involving Linnaean species may be reduced due to such misidentifications.

Aside from misidentifications from both public databases and expert identifiers, the second area of potential weakness for the results included in this work is the sampling effort. My sampling provided only 392 sequences of 31 morphologically identified species representing only one of the 11 phyla expected in the Canadian marine waters sampled (Bucklin *et al.* 2010b). Although all possible means of reducing potential sampling error were taken, resources to increase sampling were not available. The 12 Canadian port locations sampled by the Canadian Aquatic Invasive Species Network were carefully selected for their importance in establishing

baseline information and enabling future monitoring. This is due to the presence of human influence, specifically the presence of shipping ports or the expected future presence of a port. Unfortunately, due to the geographic scale, these collections were only able to be conducted twice for two different seasons at each site. Increasing the number of samples at each collection site would most likely not have affected the outcome of the majority of my conclusions for mapping geographical patterns across the Bering Strait. However, additional sampling would increase the number of close relatives, resulting in more sequence data with smaller nearest neighbour interspecific distances, potentially refining my results obtained in the elbow analyses (Chapters 2 and 3).

In addition, I hypothesize that the storage of the specimens for this project added an additional layer of uncertainty for obtaining quality sequence data. Due to the geographic expanse of the collections, multiple teams of individuals were, at various times, obtaining plankton tows and sending them to the lab for further analysis. Due to the logistics of the collection and shipping of these samples, the storage of these samples was not always ideal, and the specimens were not immediately fixed in cold temperatures and kept this way until DNA extraction. Not keeping planktonic crustaceans suspended in a cold storage solution has been noted to cause poor-quality DNA, which is expected to be a major contributor to my low amplification and sequencing success for the COI-5P molecular region (Prosser *et al.* 2013). With only a 59% COI-5P sequencing success at the individual level, the storage of the specimens before molecular extraction most likely contributed to this low value. This appeared to be less of a problem for 18S-V4 (80% success with one attempt), likely due to the more universal nature of the primers as well as the shorter fragment size. Finally, another problem faced during my research was that of amplification of non-target sequences from identified specimens. These non-target amplifications were most likely due to amplification of gut contents, or surface DNA on the specimens from the shared bulk storage medium. All steps were taken to minimize these situations; however, due to the nature of the collections and storage, no immediate recommendation for a solution to this problem is forthcoming.

Future work should focus on gaining better and more standardized collections of Canada's planktonic organisms. Choosing numerous sites evenly spaced across the Canadian and Alaskan marine coastline is recommended to provide better coverage across the three ocean regions. In addition, specific protocols whereby the collected specimens are immediately placed

and maintained in a cold storage medium is necessary to optimize DNA extraction and sequencing success. Multiple sampling events should occur throughout the year to provide a better understanding of the total diversity present, as well as potential turnover occurring at each site. Finally, in addition to increased sampling, conducting exhaustive sorting of a select number of collected samples can provide valuable information about the methodology used to obtain specimen identifications and the expected missing data, such as revealing to what degree standard sorting protocols overlook rare and cryptic species. Optimizing the collection protocols for future research will provide greater availability of data to investigate the biogeographical patterns.

To conclude, my thesis has shown that marine planktonic copepod diversity across the North Pacific, Arctic, and North Atlantic is underestimated in the current literature, and COI-5P molecular results indicate that species diversity is greater than indicated by morphological investigation alone. Furthermore, our current understanding of the diversity in the marine waters of Canada's coastlines is poorly understood. In addition, I have shown that our current knowledge of the geographic distribution of these species may not be an accurate explanation of the true diversity and endemism present across Canadian marine waters. To better understand the diversity present across Canadian waters, further efforts, particularly in the poorly-sampled Arctic region, need to occur. With increased sampling and further study, the true diversity and biogeography of Canadian marine plankton can be obtained. This knowledge will assist in future monitoring of these waters to fully appreciate human impacts, such as species introductions and climate change, and can greatly assist management and protection efforts for Canada's natural resources.

Literature Cited

Aarbakke, O.N.S., Bucklin, A., Halsband, C., & Norrbin, F. 2014. Comparative phylogeography and demographic history of five sibling species of *Pseudocalanus* (Copepoda: Calanoida) in the North Atlantic Ocean. *Journal of Experimental Marine Biology and Ecology*, **461**: 479-488.

Adamowicz, S.J., Petrusek, A., Colbourne, J.K., Hebert, P.D.N., & Witt, J.D.S. 2009. The scale of divergence: a phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, **50**(3): 423-436.

Adey, W.H., & Steneck, R.S. 2001. Thermogeography over time creates biogeographic regions: a temperature/space/time-integrated model and an abundance-weighted test for benthic marine algae. *Journal of Phycology*, **37**(5): 677-698.

Agapow, P.-M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., & Purvis, A. 2004. The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*, **79**(2): 161-179.

Appeltans, W., Ahyong, S.T., Anderson, G., Angel, M.V., Artois, T., Bailly, N., ... & Błażewicz-Paszkowycz, M. 2012. The magnitude of global marine species diversity. *Current Biology*, **22**(23): 2189-2202.

Archambault, P., Snelgrove, P.V.R., Fisher, J.A.D., Gagnon, J.M., Garbary, D.J., Harvey, M., ... & Michel, P. 2010. From sea to sea: Canada's three oceans of biodiversity. *PLoS ONE*, **5**(8): e12182.

Arnot, D.E., Roper, C., & Bayoumi, R.A.L. 1993. Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology*, **61**(1): 15-24.

Balzer, S., Malde, K., & Jonassen, I. 2011. Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics*, **27**(13): i304-i309.

Ballard, J.W.O., & Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, **13**(4): 729-744.

Barr, C.M., Neiman, M., & Taylor, D.R. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist*, **168**(1): 39-50.

Bensasson, D., Zhang, D.X., Hartl, D.L., & Hewitt, G.M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, **16**(6): 314-321.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., ... & Vogler, A.P. 2012. The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**(5): 851-869.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., ... & Das, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**(3): 148-155.

Blanco-Bercial, L., & Álvarez-Marqués, F. 2007. RFLP procedure to discriminate between *Clausocalanus* Giesbrecht, 1888 (Copepoda, Calanoida) species in the Central Cantabrian Sea. *Journal of Experimental Marine Biology and Ecology*, **344**(1): 73-77.

Blanco-Bercial, L., & Bucklin, A. 2016. New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*. *Molecular Ecology*, **25**(7): 1566-1580.

Blanco-Bercial, L., Álvarez-Marqués, F., & Bucklin, A. 2011a. Comparative phylogeography and connectivity of sibling species of the marine copepod *Clausocalanus* (Calanoida). *Journal of Experimental Marine Biology and Ecology*, **404**(1): 108-115.

Blanco-Bercial, L., Bradford-Grieve, J., & Bucklin, A. 2011b. Molecular phylogeny of the Calanoida (Crustacea: Copepoda). *Molecular Phylogenetics and Evolution*, **59**(1): 103-113.

Blanco-Bercial, L., Cornils, A., Copley, N., & Bucklin, A. 2014. DNA barcoding of marine copepods: assessment of analytical approaches to species identification. *PLoS Currents*, **6**(1): 1-21.

Blaxter, M.L. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**(1444): 669-679.

Bodil, B.A., Ambrose Jr., W.G., Bergmann, M., Clough, L.M., Gebruk, A.V., Hasemann, C., ... & Włodarska-Kowalczuk, M. 2011. Diversity of the arctic deep-sea benthos. *Marine Biodiversity*, **41**(1): 87-107.

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *AMBIO: A Journal of the Human Environment*, **37**(2): 114-118.

Boxshall, G.A., Mees, J., Costello, M.J., Hernandez, F., Bailly, N., Boury-Esnault, N., ... & Zeidler, W. 2016. World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. [accessed on 16 May 2016].

Bradford-Grieve, J.M., Boxshall, G.A., Ahyong, S.T., & Ohtsuka, S. (2010). Cladistic analysis of the calanoid Copepoda. *Invertebrate Systematics*, **24**(3): 291-321.

Breton, S., & Stewart, D.T. (2015). Atypical mitochondrial inheritance patterns in eukaryotes. *Genome*, **58**(10): 423-431.

Briggs, J.C. 2003. Marine centres of origin as evolutionary engines. *Journal of Biogeography*, **30**(1); 1-18.

Bron, J.E., Frisch, D., Goetze, E., Johnson, S.C., Lee, C.E., & Wyngaard, G.A. 2011. Observing copepods through a genomic lens. *Frontiers in Zoology*, **8**(1): 22-37.

Brown, E.A., Chain, F.J.J., Crease, T.J., MacIsaac, H.J., & Cristescu, M.E. 2015. Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecology & Evolution*, **5**(11): 2234-2251.

Bucklin, A., Guarnieri, M., Hill, R.S., Bentley, A.M., & Kaartvedt, S. 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *Hydrobiologia*, **401**: 239-254.

Bucklin, A., Frost, B.W., Bradford-Grieve, J., Allen, L.D., & Copley, N.J. 2003. Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Marine Biology*, **142**(2): 333-343.

Bucklin, A., Wiebe, P.H., Smolenack, S.B., Copley, N.J., Beaudet, J.G., Bonner, K.G., ... & Pierson, J.J. 2007. DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). *Journal of Plankton Research*, **29**(6): 483-493.

Bucklin, A., Hopcroft, R.R., Kosobokova, K.N., Nigro, L.M., Ortman, B.D., Jennings, R.M., & Sweetman, C.J. 2010a. DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**(1): 40-48.

Bucklin, A., Ortman, B.D., Jennings, R.M., Nigro, L.M., Sweetman, C.J., Copley, N.J., … & Wiebe, P.H. 2010b. A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**(24-26): 2234-2247.

Bucklin, A., Steinke, D., & Blanco-Bercial, L. 2011. DNA Barcoding of Marine Metazoa. *Annual Review of Marine Science*, **3**(1): 471-508.

Bucklin, A., McGillicuddy Jr., D.J., Wiebe, P.H., & Davis, C.S. 2015. Habitat usage by the cryptic copepods *Pseudocalanus moultoni* and *P. newmani* on Georges Bank (Northwest Atlantic). *Continental Shelf Research*, **111**(2015): 83-94.

Carr, C.M. 2012. Polychaete diversity and distribution patterns in Canadian marine waters. *Marine Biodiversity*, **42**(2): 93-107.

Carr, C.M., Hardy, S.M., Brown, T.M., Macdonald, T.A., & Hebert, P.D.N. 2011. A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. *PLoS ONE*, **6**(7): e22232.

Carrillo, E.B.G., Miller, C.B., & Wiebe, P.H. 1974. Failure of interbreeding between Atlantic and Pacific populations of the marine calanoid copepod *Acartia clausi* Giesbrecht. *Limnology and Oceanography*, **19**(3): 452-458.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A., & Kembel, S.W. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**(7): 693-715.

Chen, G., & Hare, M.P. 2011. Cryptic diversity and comparative phylogeography of the estuarine copepod *Acartia tonsa* on the US Atlantic coast. *Molecular Ecology*, **20**(11): 2425-2441.

Chesters, D., Zheng, W.M., & Zhu, C.D. 2015. A DNA Barcoding system integrating multigene sequence data. *Methods in Ecology & Evolution*, **6**(8): 930-937.

Collins, R.A., & Cruickshank, R.H. 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, **13**(6): 969-975.

Collins, R.A., & Cruickshank, R.H. 2014. Known knowns, known unknowns, unknown unknowns and unknown knowns in DNA barcoding: a comment on Dowton *et al.* *Systematic Biology*, **63**(6): 1005-1009.

Collins, R.A., Boykin, L.M., Cruickshank, R.H., & Armstrong, K.F. 2012. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology & Evolution*, **3**(3): 457-465.

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. 2015. DNA (meta) barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, **17**(3): 905-922.

Costa, F.O., deWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M., & Hebert, P.D.N. 2007. Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, **64**(2): 272-295.

Cowen, R.K., & Sponaugle, S. 2009. Larval dispersal and marine population connectivity. *Marine Science*, **1**(2009):443-466.

Cracraft, J. 1987. Species concepts and the ontology of evolution. *Biology and Philosophy*, **2**(3): 329-346.

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 & Evolution*, **29**(10): 566-571.

da Costa, K.G., Vallinoto, M., & da Costa, R.M. 2011. Molecular identification of a new cryptic species of *Acartia tonsa* (Copepoda, Acartiidae) from the Northern coast of Brazil, based on mitochondrial COI gene sequences. *Journal of Coastal Research*, **S1**(64): 359-363.

da Silva, J.M., Creer, S., Dos Santos, A., Costa, A.C., Cunha, M.R., Costa, F.O., & Carvalho, G.R. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS ONE*, **6**(5): e19449.

Darling, K.F., Kucera, M., & Wade, C.M. 2007. Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of Sciences*, **104**(12): 5002-5007.

Darriba, D., Taboada, G.L., Doallo, R., & Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**(8): 772-772.

Dayton, P.K., Mordida, B.J., & Bacon, F. 1994. Polar marine communities. *American Zoologist*, **34**(1): 90-99.

de Carvalho, M.R., Bockmann, F.A., Amorim, D.S., Brandão, C.R.F., de Vivo, M., de Figueiredo, J.L., ... & Nelson, G.J. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, **34**(3-4): 140-143.

de Queiroz, K. 1998. The General Lineage Concept of Species, Species Criteria, and the Process of Speciation. *Edited by* Howard B.D. *Endless forms: species and speciation.* Oxford University Press, New York. pp.57-75

de Queiroz, K. 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, **102**(1): 6600-6607.

de Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology*, **56**(6): 879-886.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., ... & Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**(6237): 1261605.

Dettai, A., Adamowicz, S.J., Allcock, L., Arango, C.P., Barnes, D.K.A., Barratt, I., … & Ameziane, N. 2011. DNA barcoding and molecular systematics of the benthic and demersal organisms of the CEAMARC survey. *Polar Science*, **5**(2): 298-312.

Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., ... & Talley, L.D. 2012. Climate change impacts on marine ecosystems. *Marine Science*, **4**: 11-37.

Doyle, J.J., & Gaut, B.S. 2000. Evolution of genes and taxa: a primer. *Plant Molecular Evolution* **42**(1): 1-23.

Drummond A.J., Suchard M.A., Xie D., & Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molecular Biology and Evolution*, **29**(8): 1969-1973.

Dunton, K. 1992. Arctic biogeography: the paradox of the marine benthic fauna and flora. *Trends in Ecology & Evolution*, **7**(6): 183-189.

Dupuis, J.R., Roe, A.D., & Sperling, F.A.H. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**(18): 4422-4436.

Dyke, A.S., Dale, J.E., & McNeely, R.N. 1996. Marine molluscs as indicators of environmental change in glaciated North America and Greenland during the last 18 000 years. *Géographie physique et Quaternaire*, **50**(2): 125-184.

Ebach, M.C., & Holdrege, C. 2005. DNA barcoding is no substitute for taxonomy. *Nature*, **434**(7034): 697-697.

Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**(10): 996-998.

Floyd, R., Abebe, E., Papert, A., & Blaxter, M. 2002. Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**(4): 839-850.

Flynn, J.M., Brown, E.A., Chain, F.J.J., MacIsaac, H.J., & Cristescu, M.E. 2015. Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecology & Evolution*, **5**(11): 2252-2266.

Folmer, O., Black M., Hoeh, W., Lutz, R., & Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*. **3**(5): 294-299.

Freeman, S., & Herron, J.C. 2007. Evolutionary analysis. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Frézal, L., & Leblois, R. 2008. Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution*, **8**(5): 727-736.

Fujisawa, T., & Barraclough, T.G. 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology*, **62**: 707-724.

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G.D.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**(22): 4541-4550.

Gamfeldt, L., Lefcheck, J.S., Byrnes, J.E.K., Cardinale, B.J., Duffy, J.E., & Griffin, J.N. 2015. Marine biodiversity and ecosystem functioning: what's known and what's next? *Oikos*, **124**(3): 252-265.

Gardner, G.A., & Szabo, I. 1982. British Columbia pelagic marine Copepoda: an identification manual and annotated bibliography. Department of Fisheries and Oceans. Vol. 62.

Garnett, S.T., & Christidis, L. 2007. Implications of changing species definitions for conservation purposes. *Bird Conservation International*, **17**(3):187-195.

Gerber, R.P. 2000. An identification manual to the coastal and estuarine zooplankton of the Gulf of Maine region from Passamaquoddy Bay to Long Island Sound. Freeport Village Press. Parts 1 and 2.

Giangrande, A. 2003. Biodiversity, conservation, and the 'Taxonomic impediment'. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **13**(5): 451-459.

Gladenkov, A.Y., Oleinik, A.E., Marincovich Jr., L., & Barinov, K.B. 2002. A refined age for the earliest opening of Bering Strait. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **183**(3): 321-328.

Godfray, H.C.J. 2007. Linnaeus in the information age. *Nature*, **446**(7133): 259-260.

Goetze, E. 2005. Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution*, **59**(11): 2378-2398.

Goetze, E. 2011. Population differentiation in the open sea: insights from the pelagic copepod *Pleuromamma xiphias*. *Integrative and Comparative Biology*, **51**(4): 580-597.

Gradinger, R.R. 2001. Adaptation of Arctic and Antarctic ice metazoa to their habitat. *Zoology*, **104**(3): 339-345.

Grainger, E.H. 1965. Zooplankton from the Arctic Ocean and adjacent Canadian waters. *Journal of the Fisheries Board of Canada*, **22**(2): 543-564.

Grassle, J.F., & Maciolek, N.J. 1992. Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *American Naturalist*, **139**(2): 313-341.

Guindon, S., & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**(5): 696-704.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., & Hebert, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(4): 968-971.

Halvorsen, E. 2015. Significance of lipid storage levels for reproductive output in the Arctic copepod *Calanus hyperboreus*. *Marine Ecology Progress Series*, **540**: 259-265.

Handl, J., Knowles, J., & Kell, D.B. 2005. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, **21**(15): 3201-3212.

Hardy, S.M., Carr, C.M., Hardman, M., Steinke, D., Corstorphine, E., & Mah, C. 2011. Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. *Marine Biodiversity*, **41**(1): 195-210.

Hays, G.C., Richardson, A.J., & Robinson, C. 2005. Climate change and marine plankton. *Trends in Ecology & Evolution*, **20**(6): 337-344.

Hebert, P.D.N., Cywinska, A., Ball, S.L., & deWaard, J.R. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**(1512): 313-21.

Hebert, P.D.N., Ratnasingham, S., & deWaard, J.R. 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**(Suppl 1): S96-S99.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., & Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(41): 14812-14817.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, **405**(6789): 907-913.

Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**(1442): 183-195.

Hickerson, M.J., Meyer, C.P., & Moritz, C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**(5): 729-739.

Hillis, D.M. 1987. Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics*, **18**(1):23-42.

Hillis, D.M., Moritz, C., Porter, C.A., & Baker, R.J. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science*, **251**(4991): 308-311.

Hirche, H.J. 2013. Long-term experiments on lifespan, reproductive activity and timing of reproduction in the Arctic copepod *Calanus hyperboreus. Marine Biology*, **160**(9): 2469-2481.

Hoagland, K.E. 1996. The taxonomic impediment and the convention on biodiversity. *Association of Systematics Collections Newsletter*, **24**(5): 61-62.

Hotke, K.M. 2015. DNA Barcode Variability in Canadian Cnidaria. M.Sc. thesis. Integrative Biology, University of Guelph, Guelph, ON.

Hubert, L., & Arabie, P. 1985. Comparing partitions. *Journal of Classification*, **2**(1): 193-218.

Hubert, N., & Hanner, R. 2015. DNA Barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes*, **3**(1): 44-58.

Huemer, P., Mutanen, M., Sefc, K.M., & Hebert, P.D.N. 2014. Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts. *PLoS ONE*, **9**(12): e115774.

Humes, A.G. 1994. How many copepods?. *Hydrobiologia*, **292**(1): 1-7.

Ivanova, N.V., & Grainger, C.M. 2007a. CCDB Protocols, COI amplification. Available from http://ccdb.ca//CCDB_DOCS/CCDB_Amplification.pdf [accessed May 20, 2014].

Ivanova, N.V., & Grainger, C.M. 2007b. CCDB Protocols, COI sequencing. Available from http://ccdb.ca//CCDB_DOCS/CCDB_Sequencing.pdf [accessed May 20, 2014].

Ivanova, N.V., deWaard, J., & Hebert, P.D.N. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**(4): 998-1002.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H., & Hebert, P.D.N. 2007. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**(4): 544-548.

Jagadeesan, L., Perumal, P., & Thangaraj, M. 2009. Molecular identification of marine calanoid copepod *Paracalanus parvus* (Claus 1863) using RFLP. *World Journal of Fish and Marine Sciences*, **1**(3): 239-242.

Jeffery, N.W., Elías-Gutiérrez, M., & Adamowicz, S.J. 2011. Species diversity and phylogeographical affinities of the Branchiopoda (Crustacea) of Churchill, Manitoba, Canada. *PLoS ONE*, **6**(5): e18364.

Jennings, R.M., Bucklin, A., & Pierrot-Bults, A. 2010. Barcoding of arrow worms (Phylum Chaetognatha) from three oceans: genetic diversity and evolution within an enigmatic phylum. *PLoS ONE*, **5**(4): e9949.

Jensen, S., Lynch, M.D.J., Ray, J.L., Neufeld, J.D., & Hovland, M. 2014. Norwegian deep-water coral reefs: cultivation and molecular analysis of planktonic microbial communities. *Environmental Microbiology*, **17**(10): 3597-3609.

Johnson, K.B. 1996. A guide to marine coastal plankton and marine invertebrate larvae. Kendall Hunt, U.S.A.

Johnson, W.S., & Allen, D.M. 2012. Zooplankton of the Atlantic and Gulf coasts: a guide to their identification and ecology. The Johns Hopkins University Press, U.S.A.

Jones, M., Ghoorah, A., & Blaxter, M. 2011. jMOTU and Taxonerator: turning DNA Barcode sequences into annotated operational taxonomic units. *PLoS ONE*, **6**(4): e19259.

Kaartvedt, S. 1996. Habitat preference during overwintering and timing of seasonal vertical migration of *Calanus finmarchicus*. *Ophelia*, **44**(1-3): 145-156.

Kathman, R.D. 1986. Identification manual to the Mysidacea and Euphausiacea of the northeast Pacific. Fisheries and Oceans, Information and Publications Branch, Canada. Vol. 93.

Katoh, K., & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4): 772-780.

Kekkonen, M., Mutanen, M., Kaila, L., Nieminen, M., & Hebert, P.D.N. 2015. Delineating species with DNA barcodes: a case of taxon dependent method performance in moths. *PLoS ONE*, **10**(4): e0122481.

Kerr, K.C., Stoeckle, M.Y., Dove, C.J., Weigt, L.A., Francis, C.M., & Hebert, P.D.N. 2007. Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, **7**(4): 535-543.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**(2): 111-120.

Knowlton, N. 1993. Sibling species in the sea. *Annual Review of Ecology and Systematics*, **24**(1): 189-216.

Kosobokova, K.N., Hopcroft, R.R., & Hirche, H.J. 2011. Patterns of zooplankton diversity through the depths of the Arctic's central basins. *Marine Biodiversity*, **41**(1): 29-50.

Krishnamurthy, P.K., & Francis, R.A. 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*, **21**(8): 1901-1919.

Kuklinski, P., Taylor, P.D., Denisenko, N.V., & Berning, B. 2013. Atlantic origin of the Arctic biota? Evidence from phylogenetic and biogeographical analysis of the cheilostome bryozoan genus *Pseudoflustra*. *PLoS ONE*, **8**(3): e59152.

Kumar, C.S., Prabu, V.A., & Kumar, C.P. 2015. DNA Barcode Genes (rbcL, 18s rRNA and ITS Phylogeny) in *Skeletonema costatum* Grevelli (Cleve, 1873). *International Journal of Current Microbiology and Applied Sciences*, **4**(9): 195-203.

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, **12**(1): 118-123.

Laakmann, S., Auel, H., & Kochzius, M. 2012. Evolution in the deep sea: Biological traits, ecology and phylogenetics of pelagic copepods. *Molecular Phylogenetics and Evolution*, **65**(2): 535-546.

Laakmann, S., Gerdts, G., Erler, R., Knebelsberger, T., Martínez Arbizu, P., & Raupach, M.J. 2013. Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences. *Molecular Ecology Resources*, **13**(5): 862-876.

Laughinghouse IV, H.D., Müller, K.M., Adey, W.H., Lara, Y., Young, R., & Johnson, G. 2015. Evolution of the Northern Rockweed, *Fucus distichus*, in a Regime of Glacial Cycling: Implications for Benthic Algal Phylogenetics. *PLoS ONE*, **10**(12): e0143795.

Layton, K. 2013. Examining patterns of genetic variation in Canadian marine molluscs through DNA barcodes. M.Sc. thesis, Integrative Biology, University of Guelph, Guelph ON.

Layton, K.K., Martel, A.L., & Hebert, P.D.N. 2014. Patterns of DNA barcode variation in Canadian marine molluscs. *PLoS ONE*, **9**(4): e95003.

Layton, K.K., Martel, A.L., & Hebert, P.D.N. 2016. Geographic patterns of genetic diversity in two species complexes of Canadian marine bivalves. *Journal of Molluscan Studies*, **82**: 282-291.

Lefébure, T., Douady, C.J., Gouy, M., & Gibert, J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**(2): 435-447.

Liao, D. 1999. Concerted evolution: molecular mechanism and biological implications. *The American Journal of Human Genetics*, **64**(1): 24-30.

Lobo, J., Costa, P.M., Teixeira, M.A.L., Ferreira, M.S.G., Costa, M.H., & Costa, F.O. 2013. Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC Ecology*, **13**(34): 1-8.

Longhurst, A.R. 1985. The structure and evolution of plankton communities. *Progress in Oceanography*, **15**(1): 1-35.

Lovejoy, C. 2014. Changing views of Arctic protists (marine microbial eukaryotes) in a changing Arctic. *Acta Protozoologica*, **53**(1): 91-100.

Maddison, W.P., & D.R., Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04 http://mesquiteproject.org.

May, R.M. 1988. How many species are there on earth? *Science*, **241**(4872): 1441-1449.

May, R.M. 2010. Tropical arthropod species, more or less? *Science*, **329**(5987): 41-42.

Mayden, R.L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. *In* Species: The Units of Biodiversity. *Edited by* Claridge, M.F., Dawah, H.A. & Wilson, M.R. Chapman and Hall, London, 381-324.

Mayr, E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, U.S.A.

Mayr, E. 1996. What is a species, and what is not? *Philosophy of Science*, **63**(1): 262-277.

McLaren, I.A. 1974. Demographic strategy of vertical migration by a marine copepod. *American Naturalist*, **108**(959):91-102.

McLaren, I.A., Corkett, C.J., & Zillioux, E.J. 1969. Temperature adaptations of copepod eggs from the arctic to the tropics. *The Biological Bulletin*, **137**(3): 486-493.

McManus, G.B., & Katz, L.A. 2009. Molecular and morphological methods for identifying plankton: what makes a successful marriage? *Journal of Plankton Research*, **31**(10): 1119-1129.

Messing, J., 1983. New M13 vectors for cloning. *Methods in Enzymology*, **101**: 20-78.

Meyer, C.P. 2003. Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnaean Society*, **79**(3): 401-459.

Meyer, C.P., & Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**(12): e422.

Mitchell, A. 2008. DNA barcoding demystified. *Australian Journal of Entomology*, **47**(3): 169-173.

Molnar, J.L., Gamboa, R.L., Revenga, C., & Spalding, M.D. 2008. Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, **6**(9): 485-492.

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., & Worm B. 2011. How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, **9**(8): e1001127.

Moritz, C., & Cicero, C. 2004. DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**(10): 1529-1531.

Nadolny, A.A., Omelko, M.M., Marusik, Y.M., & Blagoev, G. 2016. A new species of spider belonging to the *Pardosa lugubris*-group (Araneae: Lycosidae) from Far East Asia. *Zootaxa*, **4072**(2): 263-281.

Nawrocki, E.P. 2009. Structural RNA Homology Search and Alignment using Covariance Models. Ph.D. thesis, Washington University in Saint Louis, School of Medicine, Saint Louis, MO.

Newman, W.A. 1992. Origin of Maxillopoda. *Acta Zoologica*, **73**(5): 319-322.

Nixon, K.C., & Wheeler, Q.D. 1990. An amplification of the phylogenetic species concept. *Cladistics*, **6**(3): 211-223.

Nouvel, H. 1950. Mysidacea. *In* Zooplankton. Conseil International Pour L'exploration De la Mer. No. 18-27.

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., & Zaharoff, A.K. 2013. Phylotranscriptomics to bring the understudied into the fold: monophyletic Ostracoda, fossil placement, and pancrustacean phylogeny. *Molecular Biology and Evolution*, **30**(1): 215-233.

Packer, L., Gibbs, J., Sheffield, C., & Hanner, R. 2009. DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, **9**(s1): 42-50.

Parent, G.J., Plourde, S., & Turgeon, J. 2012. Natural hybridization between *Calanus finmarchicus* and *C. glacialis* (Copepoda) in the Arctic and Northwest Atlantic. *Limnology and Oceanography*, **57**(4): 1057-1066.

Passamonti, M., Ricci, A., Milani, L., & Ghiselli, F. 2011. Mitochondrial genomes and Doubly Uniparental Inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia Mytilidae). *BMC Genomics*, **12**(1): 1-19.

Pechenik, J.A. 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series*, **177**(1): 269-297.

Piepenburg, D. 2005. Recent research on Arctic benthos: common notions need to be revised. *Polar Biology*, **28**(10): 733-755.

Pinto, F.R., Carriço, J.A., Ramirez, M., & Almeida, J.S. 2007. Ranked Adjusted Rand: integrating distance and partition information in a measure of clustering agreement. *BMC Bioinformatics*, **8**(44): 1-13.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., ... & Vogler, A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**(4): 595-609.

Porco, D., Rougerie, R., Deharveng, L., & Hebert, P.D.N. 2010. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high-throughput context: the example of Collembola. *Molecular Ecology Resources*, **10**(6): 942-945.

Prosser, S., Martínez-Arce, A., and Elías-Gutiérrez, M. 2013. A new set of primers for COI amplification from freshwater microcrustaceans. *Molecular Ecology Resources*, **13**(6): 1151-1155.

Puillandre, N., Strong, E.E., Bouchet, P., Boisselier, M.C., Couloux, A., & Samadi, S. 2009. Identifying gastropod spawn from DNA barcodes: possible but not yet practicable. *Molecular Ecology Resources*, **9**(5): 1311-1321.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, **21**(8): 1864-77.

Radulovici, A.E., Archambault, P., & Dufresne, F. (2010). DNA Barcodes for Marine Biodiversity: Moving Fast Forward? *Diversity*, **2**(4): 450-472.

Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. 2014. Tracer v1.6 [online]. Available from http://beast.bio.ed.ac.uk/Tracer [accessed 12 July 2014].

Ratnasingham, S., & Hebert, P.D.N. 2007. Barcoding BOLD: The Barcode of Life Data System. *Molecular Ecology Notes*, **7**(3): 355-364.

Ratnasingham, S., & Hebert, P.D.N. 2013. A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoS ONE*, **8**(7): e66213.

Raupach, M.J., & Radulovici, A.E. 2015. Looking back on a decade of barcoding crustaceans. *ZooKeys*, **539**(2015): 53-81.

Raupach, M.J., Barco, A., Steinke, D., Beermann, J., Laakmann, S., Mohrbeck, I., ... & Knebelsberger, T. 2015. The application of DNA barcodes for the identification of marine crustaceans from the North Sea and adjacent regions. *PLoS ONE*, **10**(9): e0139421.

Razouls, C., De Bovée, F., Kouwenberg, J., & Desreumaux, N. 2005. Diversity and geographic distribution of marine planktonic copepods. Available from http://copepodes.obs-banyuls.fr/en/ [accessed 11 Jan 2015].

Regier, J.C., Shultz, J.W., & Kambic, R.E. 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**(1561): 395-401.

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., ... & Cunningham, C.W. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, **463**(7284): 1079-1083.

Renaud, A.K., Savage, J., & Adamowicz, S.J. 2012. DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. *BMC Ecology*, **12**(1): 24.

Renaud, P.E., Sejr, M.K., Bluhm, B.A., Sirenko, B., & Ellingsen, I.H. 2015. The future of Arctic benthos: expansion, invasion, and biodiversity. *Progress in Oceanography*, **139**(2015): 244-257.

Richly, E., & Leister, D. 2004. NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution*, **21**(6): 1081-1084.

Rochet, M., & Grainger, E.H. 1988. Community structure of zooplankton in eastern Hudson Bay. *Canadian Journal of Zoology*, **66**(7): 1626-1630.

Roe, A.D., & Sperling, F.A. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, **44**(1): 325-345.

Roff, J.C., Davidson, K.G., Pohle, G., & Dadswell, M.J. 1984. *A guide to the marine flora and fauna of the Bay of Fundy and Scotian Shelf: larval Decapoda: Brachyura*. Canadian Technical Report of Fisheries and Aquatic Sciences **1322**: iv-57.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., & Reyes, A. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238**(1): 195-209.

Saitou, N., & Nei, M. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**(4): 406-425.

Saunders, G.W., & McDevit, D.C. 2013. DNA barcoding unmasks overlooked diversity improving knowledge on the composition and origins of the Churchill algal flora. *BMC Ecology*, **13**(1): 1-23.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., … & Weber, C.F. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**(23): 7537-7541.

Severiano, A., Pinto, F.R., Ramirez, M., & Carriço, J.A. 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. *Journal of Clinical Microbiology*, **49**(11): 3997-4000.

Shen, Y.Y., Chen, X., & Murphy, R.W. 2013. Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE*, **8**(2): e57125.

Shokralla, S., Spall, J.L., Gibson, J.F., & Hajibabaei, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**(8): 1794-1805.

Singh, J.S. 2002. The biodiversity crisis: a multifaceted review. *Current Science*, **82**(6): 638-647.

Song, H., Buhay, J.E., Whiting, M.F., & Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, **105**(36): 13486-13491.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Ferdaña, Z.A., ... & Robertson, J. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**(7): 573-583.

Spangenberg, J.H. 2007. Biodiversity pressure and the driving forces behind. *Ecological Economics*, **61**(1): 146-158.

Srivathsan, A., & Meier, R. 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, **28**(2): 190-194.

Stoeckle, M.Y., & Thaler, D.S. 2014. DNA barcoding works in practice but not in (neutral) theory. *PLoS ONE*, **9**(7): e100755.

Talavera, G., Dincă, V., & Vila, R. 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology & Evolution*, **4**(12): 1101-1110.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**(12): 2725-2729.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., & Fontaneto, D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences*, **109**(40): 16208-16212.

Tang, C.Q., Humphreys, A.M., Fontaneto, D., & Barraclough, T.G. 2014. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology & Evolution*, **5**(10): 1086-1094.

Tang, R.W., Yau, C., & Ng, W.C. 2010. Identification of stomatopod larvae (Crustacea: Stomatopoda) from Hong Kong waters using DNA barcodes. *Molecular Ecology Resources*, **10**(3): 439-448.

Teske, P.R., Papadopoulos, I., Barker, N.P., & McQuaid, C.D. 2012. Mitochondrial DNA paradox: sex-specific genetic structure in a marine mussel–despite maternal inheritance and passive dispersal. *BMC Genetics*, **13**(1): 45-51.

Todd, C.D., Laverack, M.S., & Boxshall, G. 1996. Coastal marine zooplankton: a practical manual for students. Cambridge University Press, Cambridge UK.

Unal, E., & Bucklin, A. 2010. Basin-scale population genetic structure of the planktonic copepod *Calanus finmarchicus* in the North Atlantic Ocean. *Progress in Oceanography*, **87**(1): 175-185.

Vermeij, G.J. 1991. Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology*, **17**(3): 281-307.

Vermeij, G.J., & Roopnarine, P.D. 2008. The coming Arctic invasion. *Science*, **321**(5890): 780-781.

Vilgalys, R. 2003. Taxonomic misidentification in public DNA databases. *New Phytologist,* **160**(1): 4-5.

Vincent, W.F., Mueller, D., Van Hove, P., & Howard-Williams, C. 2005. Glacial periods on early Earth and implications for the evolution of life. *In* Origins: genesis, evolution and diversity of life. Edited by Seckbach, J. Springer Netherlands. Netherlands, pp. 483-501.

Wallace, D.L. 1983. A method for comparing two hierarchical clusterings: comment. *Journal of the American Statistical Association,* **78**(383): 569-576.

Ward, R.D., Hanner, R., & Hebert, P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**(2): 329-356.

Wares, J.P. 2002. Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology*, **11**(7): 1131-1144.

Wares, J.P., & Cunningham, C.W. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, **55**(12): 2455-2469.

Wassmann, P., Duarte, C.M., Agusti, S., & Sejr, M.K. 2011. Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology*, **17**(2): 1235-1249.

White, D.J., Wolff, J.N., Pierson, M., & Gemmell, N.J. 2008. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**(23): 4925-4942.

Wiley, E.O., & Mayden, R.L. 2000. The evolutionary species concept. *In* Species concepts and phylogenetic theory: a debate. Edited by Wheeler, Q.D., & Meier, R. Columbia University Press, USA, pp. 70-89.

Will, K.W., Mishler, B.D., & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**(5): 844-51.

WoRMS Editorial Board 2016. World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. [accessed on 24 May 2016].

Wu, S., Xiong, J., & Yu, Y. 2015. Taxonomic Resolutions Based on 18S rRNA Genes: A Case Study of Subclass Copepoda. *PLoS ONE*, **10**(6): e0131498.

Yu, Y., Harris, A.J., Blair, C., & He, X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution,* **87**(2015) 46-49.

Zhan, A., & MacIsaac, H.J. 2015. Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. *Conservation Genetics*, **16**(3): 513-522.

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A. A., Abbott, C. L., … & MacIsaac, H. J. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology & Evolution*, **4**(6): 558-565.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, **29**(22): 2869-2876.

Zhou, X., Jacobus, L.M., DeWalt, R.E., Adamowicz, S.J., & Hebert, P.D.N. 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. *Journal of the North American Benthological Society*, **29**(3): 814-837.

Zimmermann, J., Jahn, R., & Gemeinholzer, B. 2011. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, **11**(3): 173-192.

Zouros, E. 2013. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evolutionary Biology*, **40**(1): 1-31.

Appendix I: List of sequences used for Chapter 2 analyses

List of all process identifiers (BOLD database sequence unique identifiers) for the *COI reference* and *COI* data sets used for Chapter 2.

COI reference

BIOZO004-14	GBA10677-13	GBA14308-13	GBCX1950-14	GBMIN41480-14	SACRU006-08
BIOZO008-14	GBA10678-13	GBA14309-13	GBCX1951-14	GBMIN41481-14	SACRU008-08
BIOZO010-14	GBA10679-13	GBA14310-13	GBCX1952-14	GBMIN41482-14	SACRU009-08
BIOZO049-14	GBA10686-13	GBA14311-13	GBCX1953-14	GBMIN41483-14	SACRU014-08
BIOZO052-14	GBA10687-13	GBA14312-13	GBCX1954-14	GBMIN41484-14	SACRU078-10
BIOZO093-14	GBA10688-13	GBA14313-13	GBCX1955-14	GBMIN41485-14	SACRU127-10
BNSC085-10	GBA10689-13	GBA14315-13	GBCX1956-14	GBMIN41490-14	SACRU130-10
BNSC296-11	GBA10690-13	GBA14320-13	GBCX1957-14	GBMIN41495-14	SACRU140-10
BNSC303-11	GBA10691-13	GBA14322-13	GBCX1958-14	GBMIN41496-14	SACRU141-10
BNSC304-11	GBA10692-13	GBA14326-13	GBCX1965-14	GBMIN41498-14	SACRU153-10
BNSC306-11	GBA10693-13	GBA14327-13	GBCX1966-14	GBMIN41500-14	SBDSC190-11
BNSC307-11	GBA10694-13	GBA14329-13	GBCX1967-14	GBMIN41501-14	SBDSC365-11
BNSC326-11	GBA10695-13	GBA14331-13	GBCX1968-14	GBMIN41502-14	SBDSC368-11
BNSC327-11	GBA10696-13	GBA14332-13	GBCX1969-14	GBMIN41505-14	SGRZC019-09
BNSC337-11	GBA10697-13	GBA14336-13	GBCX1970-14	GBMIN41506-14	SGRZC032-09
BNSC353-11	GBA10698-13	GBA14337-13	GBCX1971-14	GBMIN41509-14	SGRZC037-09
BNSC354-11	GBA10699-13	GBA14339-13	GBCX1974-14	GBMIN41510-14	SLAVA006-11
BNSC355-11	GBA10701-13	GBA14340-13	GBCX1975-14	GBMIN41513-14	SLAVA008-11
BNSC356-11	GBA10702-13	GBA14343-13	GBCX1976-14	GBMIN41515-14	SLAVA014-11
BNSC357-11	GBA10703-13	GBA14345-13	GBCX1977-14	GBMIN41516-14	SLAVA016-11
BNSC358-11	GBA10704-13	GBA14347-13	GBCX1979-14	GBMIN41521-14	SLAVA017-11
BNSC376-11	GBA10706-13	GBA14349-13	GBCX1980-14	GBMIN41527-14	SLAVA018-11
BNSC379-11	GBA10707-13	GBA14351-13	GBCX1984-14	GBMIN41529-14	SLAVA020-11
BNSC380-11	GBA10708-13	GBA14353-13	GBCX1985-14	GBMIN41531-14	SLAVA021-11
BNSC424-12	GBA10709-13	GBA14355-13	GBCX1987-14	GBMIN41532-14	SLAVA022-11
BNSC425-12	GBA10711-13	GBA14356-13	GBCX1988-14	GBMIN41534-14	SLAVA025-11
BNSC426-12	GBA10712-13	GBA14358-13	GBCX1999-14	GBMIN41535-14	SLAVA031-11
BNSC427-12	GBA10713-13	GBA14359-13	GBCX2001-14	GBMIN41538-14	SLAVA032-11
BNSC428-12	GBA10714-13	GBA14360-13	GBCX2002-14	GBMIN41539-14	SLAVA033-11
BNSC429-12	GBA10715-13	GBA14361-13	GBCX2003-14	GBMIN41540-14	SLAVA038-11
BNSC430-12	GBA10716-13	GBA14364-13	GBCX2004-14	GBMIN41541-14	SLAVA039-11
BNSC431-12	GBA10717-13	GBA14365-13	GBCX2005-14	GBMIN41542-14	SLAVA044-11
BNSC432-12	GBA10718-13	GBA14366-13	GBCX2006-14	GBNM0031-06	SLAVA048-11
BNSC438-12	GBA10719-13	GBA14370-13	GBCX2007-14	HVDBC064-11	SLAVA051-11
BNSC440-12	GBA10720-13	GBA14377-13	GBCX2008-14	HVDBC065-11	SLAVA052-11
BNSC441-12	GBA10721-13	GBA14380-13	GBCX2009-14	HVDBC066-11	SLAVA055-11
BNSC445-12	GBA10722-13	GBA14381-13	GBCX2010-14	HVDBC067-11	SLAVA057-11

BNSC447-12
BNSC455-12
BNSC457-12
BNSC458-12
BNSC459-12
BNSC462-12
BNSC463-12
BNSC464-12
BNSC471-12
BNSC473-12
BNSCP031-11
BNSCP095-11
BNSDE078-11
GBA0012-06
GBA11364-13
GBA1931-07
GBA4776-09
GBCX0008-06
GBCX0279-06
GBCX0466-06
GBCX1769-14
GBFCD053-14
GBFCM0001-06
GBFCM077-07
ACSD003-08
ACSD005-08
ACSD006-08
ACSD007-08
ACSD011-08
ACSD012-08
ACSD013-08
ACSD016-08
ACSD017-08
ACSD018-08
ACSD019-08
ACSD025-08
ACSD026-08
ACSD028-08
ACSD032-08
ACSD033-08
ACSD039-08
ACSD042-08
GBA10723-13
GBA10724-13
GBA10725-13
GBA10727-13
GBA10728-13
GBA10729-13
GBA10730-13
GBA10731-13
GBA10732-13
GBA10734-13
GBA10735-13
GBA10736-13
GBA10737-13
GBA10738-13
GBA10740-13
GBA10741-13
GBA10742-13
GBA10743-13
GBA10745-13
GBA10746-13
GBA10747-13
GBA10748-13
GBA10749-13
GBA10750-13
GBA10751-13
GBA10752-13
GBA10753-13
GBA10754-13
GBA10756-13
GBA10767-13
GBA10768-13
GBA10769-13
GBA10776-13
GBA10777-13
GBA10778-13
GBA10779-13
GBA10780-13
GBA10781-13
GBA10782-13
GBA10783-13
GBA10784-13
GBA10785-13
GBA14382-13
GBA14385-13
GBA14391-13
GBA14392-13
GBA14393-13
GBA14394-13
GBA14399-13
GBA14400-13
GBA14404-13
GBA14405-13
GBA14406-13
GBA14407-13
GBA14412-13
GBA14413-13
GBA14416-13
GBA14421-13
GBA14425-13
GBA14426-13
GBA14427-13
GBA14428-13
GBA14429-13
GBA14430-13
GBA14431-13
GBA14432-13
GBA14433-13
GBA14434-13
GBA14435-13
GBA14436-13
GBA14437-13
GBA14438-13
GBA14439-13
GBA14440-13
GBA14442-13
GBA14443-13
GBA14444-13
GBA14445-13
GBA14446-13
GBA14447-13
GBA14448-13
GBA14449-13
GBA14450-13
GBA14452-13
GBCX2011-14
GBCX2012-14
GBCX2013-14
GBCX2014-14
GBCX2015-14
GBCX2016-14
GBCX2017-14
GBCX2018-14
GBCX2019-14
GBCX2020-14
GBCX2021-14
GBCX2027-14
GBCX2028-14
GBCX2029-14
GBCX2030-14
GBCX2031-14
GBCX2032-14
GBCX2033-14
GBCX2034-14
GBCX2035-14
GBCX2036-14
GBCX2037-14
GBCX2038-14
GBCX2040-14
GBCX2041-14
GBCX2042-14
GBCX2043-14
GBCX2044-14
GBCX2045-14
GBCX2046-14
GBCX2047-14
GBCX2048-14
GBCX2049-14
GBCX2052-14
GBCX2054-14
GBCX2055-14
GBCX2056-14
GBCX2057-14
GBCX2063-14
GBCX2064-14
GBCX2070-14
GBCX2071-14
HVDBC068-11
HVDBC071-11
HVDBC072-11
HVDBC081-11
HVDBC100-12
HVDBC102-12
HVDBC104-12
HVDBM517-11
HVDBM519-11
HZPLY1108-13
INVMP016-13
INVMP017-13
INVMP018-13
JMCRU053-09
JMCRU055-09
JMCRU068-09
JMCRU075-09
JMCRU087-09
JMCRU090-09
JMCRU098-09
JMCRU101-09
JMCRU106-09
JMCRU109-09
JMCRU116-09
JMCRU121-09
JMCRU127-09
JMCRU133-09
JMCRU134-09
JMCRU135-09
JMCRU139-09
JMCRU144-09
JMCRU147-09
JMCRU149-09
JMCRU160-09
JMCRU161-09
JMCRU170-09
JMCRU183-09
JMCRU193-09
JMCRU205-09
JMCRU211-09
JMCRU224-09
JMCRU226-09
SLAVA058-11
SLAVA059-11
SLAVA061-11
SLAVA065-11
SLAVA066-11
SLAVA067-11
SLAVA068-11
SLAVA071-11
SLAVA072-11
SLAVA074-11
SLAVA076-11
SLAVA079-11
SLAVA080-11
SLAVA082-11
SLAVA083-11
SLAVA084-11
SLAVA092-11
SLAVA093-11
SLAVA094-11
SLAVA097-11
SLAVA105-11
SLAVA107-11
SLAVA108-11
SLAVA109-11
SLAVA116-11
SLAVA122-11
SLAVA124-11
SLAVA125-11
SLAVA126-11
SLAVA127-11
SLAVA129-11
SLAVA134-11
SLAVA135-11
SLAVA136-11
SLAVA138-11
SLAVA140-11
SLAVA141-11
SLAVA142-11
SLAVA143-11
SLAVA144-11
SLAVA145-11
SLAVA146-11

ACSD043-08	GBA10786-13	GBA14453-13	GBCX2072-14	JMCRU229-09	SLAVA147-11
ACSD044-08	GBA10787-13	GBA14467-13	GBCX2073-14	JMCRU234-09	SLAVA148-11
ACSD048-08	GBA10788-13	GBA14468-13	GBCX2074-14	JMCRU235-09	SLAVA152-11
ACSD050-08	GBA10789-13	GBA477-04	GBCX2075-14	JMCRU241-09	SLAVA154-11
ACSD054-08	GBA10790-13	GBA4817-09	GBCX2077-14	JMCRU243-09	SLAVA158-11
ACSD056-08	GBA10791-13	GBA7206-11	GBCX2078-14	JMCRU249-09	SLAVA159-11
ACSD060-08	GBA10792-13	GBA7207-11	GBCX2079-14	KBINV032-11	SLAVA161-11
ACSD062-08	GBA10793-13	GBA7208-11	GBCX2080-14	KBINV033-11	SLAVA162-11
ACSD063-08	GBA10794-13	GBA8990-13	GBCX2081-14	KBINV035-11	SLAVA165-11
ACSD067-08	GBA10795-13	GBA9013-13	GBCX2112-14	KBINV056-11	SLAVA166-11
ACSD068-08	GBA10796-13	GBA9014-13	GBCX2113-14	KLMPC002-09	SLAVA172-11
ACSD069-08	GBA10797-13	GBA9016-13	GBCX2115-14	KLMPC003-09	SLAVA179-11
ACSD070-08	GBA10798-13	GBA9017-13	GBCX2117-14	KLMPC010-09	SLAVA180-11
ACSD084-08	GBA10799-13	GBA9018-13	GBCX2118-14	KLMPC036-09	SLAVA182-11
ACSD089-08	GBA10800-13	GBA9019-13	GBCX2119-14	KLMPC055-09	SLAVA185-11
ACSD097-11	GBA10801-13	GBA9020-13	GBCX2120-14	KLMPC070-09	SLAVA186-11
ACSD100-11	GBA10802-13	GBA9021-13	GBCX2121-14	KLMPC089-09	SLAVA187-11
ACSD101-11	GBA10803-13	GBA9053-13	GBCX2122-14	KLMPC092-09	SLAVA188-11
ACSD102-11	GBA10804-13	GBA9054-13	GBCX2125-14	LIMC607-08	SLOBC001-06
ACSD104-11	GBA10805-13	GBA9056-13	GBCX2127-14	LIMC609-08	SLOBC007-06
ACSD105-11	GBA10806-13	GBA9059-13	GBCX2155-14	LIMC610-08	SLOBC011-06
ACSD115-11	GBA10807-13	GBA9061-13	GBCX2156-14	LIMC611-08	SLOBC013-06
ACSD118-11	GBA10808-13	GBA9062-13	GBCX2157-14	LIMC693-10	SLOBC014-06
ACSD124-11	GBA10809-13	GBA9063-13	GBCX2158-14	MBMIA433-06	SLOBC015-06
ACSD126-11	GBA10810-13	GBA9064-13	GBCX2159-14	MBMIA434-06	SLOBC017-06
ACSD129-11	GBA10811-13	GBA9065-13	GBCX2160-14	MMAST002-12	SLOBC018-06
ACSD136-11	GBA10812-13	GBA9066-13	GBCX2161-14	MMAST008-12	SLOBC024-06
ACSD142-11	GBA10813-13	GBA9067-13	GBCX2162-14	MSCP001-09	SLOBC025-06
ACSD143-11	GBA10814-13	GBA9068-13	GBCX2163-14	MSCP003-09	SLOBC027-06
ACSD145-11	GBA10815-13	GBA9069-13	GBCX2164-14	MSCP010-09	SLOBC029-06
ACSD147-11	GBA10816-13	GBA9070-13	GBCX2165-14	MSCP015-09	SLOBC030-06
ACSD148-11	GBA10817-13	GBA9071-13	GBCX2166-14	MSCP020-09	SLOBC032-06
ACSD152-11	GBA10818-13	GBA9072-13	GBCX2167-14	MSCP025-09	SLOBC033-06
ACSD155-11	GBA10819-13	GBA9073-13	GBCX2168-14	MSCP047-09	SLOBC036-06
ACSD156-11	GBA1097-04	GBA9075-13	GBCX2170-14	MSCP055-09	SLOBC038-06
ACSD170-11	GBA11253-13	GBA9076-13	GBCX2172-14	MSCP059-09	SLOBC041-06
ACSD171-11	GBA11254-13	GBA9077-13	GBCX2174-14	MSCP063-09	SLOBC048-06
ACSD173-11	GBA11255-13	GBA9078-13	GBCX2608-14	MSCP083-09	SLOBC049-06
ACSD174-11	GBA11256-13	GBA9079-13	GBCX2609-14	MSCP086-09	SLOBC052-06
ACSD184-11	GBA11257-13	GBA9080-13	GBCX2622-14	MSCP092-09	SLOBC055-06
ACSD185-11	GBA11258-13	GBA9081-13	GBCX2623-14	MSCP126-09	SLOBC056-06
ACSD188-11	GBA11259-13	GBA9082-13	GBCX2754-14	MSCP135-09	SLOBC062-06

ARCMI434-14	GBA11260-13	GBA9083-13	GBCX2755-14	MSCP137-09	SLOBC063-06
ARCMI435-14	GBA11261-13	GBA9084-13	GBCX2756-14	MSCP146-09	SLOBC066-06
ARCMI437-14	GBA11262-13	GBA9085-13	GBCX2758-14	MSCP163-09	SLOBC078-06
ARCMI439-14	GBA11263-13	GBA9086-13	GBCX2759-14	MSCP178-09	SLOBC079-06
ARCMI452-14	GBA11264-13	GBA9087-13	GBCX2760-14	MSCP179-09	SLOBC080-06
ARCMI503-14	GBA11265-13	GBA9088-13	GBCX2761-14	MSCP183-09	SLOBC086-06
ARCMI505-14	GBA11266-13	GBA9089-13	GBCX2766-14	MSCP185-09	SLOBC088-06
ARCMI507-14	GBA11267-13	GBA9090-13	GBCX2767-14	MSCP188-09	SLOBC093-06
ARCMI510-14	GBA11268-13	GBA9091-13	GBCX2768-14	MSCP223-09	SLOBC100-07
ARCMI511-14	GBA11269-13	GBA9092-13	GBCX2769-14	MSCP233-09	SLOBC105-07
ARCMI512-14	GBA11270-13	GBA9093-13	GBCX2770-14	MSCP234-09	SLOBC108-07
ARCMI513-14	GBA11271-13	GBA9094-13	GBCX2771-14	MSCP239-09	SLOBC109-07
ARCMI514-14	GBA11272-13	GBA9095-13	GBCX2772-14	MSCP248-09	SLOBC110-07
ARCMI519-14	GBA11273-13	GBA9096-13	GBCX2773-14	MSCP250-09	SLOBC111-07
ARCMI520-14	GBA11274-13	GBA9097-13	GBCX2774-14	MSCP253-09	SLOBC114-07
ARCMI530-14	GBA11275-13	GBA9098-13	GBCX2775-14	MSCP254-09	SLOBC117-07
BALAP007-11	GBA11277-13	GBA9099-13	GBCX2778-14	MSCP292-09	SLOBC118-07
BALAP008-11	GBA11279-13	GBA9100-13	GBCX2779-14	MSCP293-09	SLOBC121-07
BALAP019-11	GBA11281-13	GBA9101-13	GBCX2780-14	MSCP295-09	SLOBC124-07
BALAP053-11	GBA11284-13	GBA9102-13	GBCX2781-14	MSCP302-09	SLOBC127-07
BALAP054-11	GBA11286-13	GBA9103-13	GBCX2782-14	MSCP321-09	SLOBC128-07
BAPAZ012-10	GBA11287-13	GBA9104-13	GBCX2783-14	MSCP326-09	SLOBC129-07
BAPAZ024-10	GBA11288-13	GBA9105-13	GBCX2784-14	MSCP330-09	SLOBC130-07
BAPAZ045-10	GBA11293-13	GBA9106-13	GBCX2785-14	MSCP345-09	SLOBC133-07
BAPAZ056-10	GBA11295-13	GBA9107-13	GBCX2786-14	MSCP346-09	SLOBC134-07
BAPAZ058-10	GBA11296-13	GBA9108-13	GBCX2787-14	MSCP348-09	SLOBC135-07
BAPAZ059-10	GBA11297-13	GBA9109-13	GBCX2802-14	MSCP349-09	SLOBC143-07
BAPAZ060-10	GBA11298-13	GBA9111-13	GBCX2803-14	MSCP369-09	SLOBC145-07
BAPAZ116-11	GBA11300-13	GBA9112-13	GBCX2804-14	MSCP378-09	SLOBC148-07
BAPAZ118-11	GBA11303-13	GBA9113-13	GBCX2805-14	MSCP409-11	SLOBC149-07
BAPAZ122-11	GBA11305-13	GBA9114-13	GBCX2806-14	MSCP459-11	SLOBC150-07
BAPAZ142-11	GBA11306-13	GBA9115-13	GBCX2807-14	NBCRU053-12	SLOBC152-07
BAPAZ143-11	GBA11308-13	GBA9116-13	GBCX2808-14	NBCRU054-12	SLOBC156-07
BAPAZ151-11	GBA11309-13	GBA9117-13	GBCX2815-14	NBCRU086-12	SLOBC157-07
BAPAZ161-11	GBA11333-13	GBA9118-13	GBCX2816-14	NBCRU087-12	SLOBC160-07
BAPAZ164-11	GBA11334-13	GBA9119-13	GBCX2818-14	NBCRU088-12	SLOBC162-07
BAPAZ168-11	GBA11335-13	GBA9120-13	GBCX2820-14	NBCRU089-12	SLOBC164-07
BAPAZ175-11	GBA11336-13	GBA9121-13	GBCX2821-14	NJCGS086-09	SLOBC165-07
BAPAZ177-11	GBA11338-13	GBA9122-13	GBCX2824-14	NJCGS1130-11	SLOBC166-07
BAPAZ188-11	GBA11339-13	GBA9123-13	GBCX2825-14	NJCGS1145-11	SLOBC169-07
BASKL079-09	GBA11342-13	GBA9124-13	GBCX2831-14	NJCGS181-10	SLOBC171-07
BBCRU019-10	GBA11343-13	GBA9125-13	GBCX2834-14	NJCGS182-10	SLOBC172-07

BBCRU020-10
BBCRU021-10
BBCRU022-10
BBCRU024-10
BBCRU069-10
BBCRU112-10
BBCRU113-10
BBCRU114-10
BBCRU115-10
BBCRU123-10
BBCRU148-10
BBCRU149-10
BBCRU151-10
BBCRU153-10
BBLZI216-14
BBLZI234-14
BBLZI237-14
BCRUS025-10
BCRUS031-10
BCRUS032-10
BCRUS040-10
BCRUS043-10
BCRUS045-10
BCRUS054-10
BCRUS069-10
BCRUS070-10
BCRUS073-10
BCRUS079-10
BCRUS089-10
BCRUS090-10
BCRUS091-10
BCRUS098-10
BCRUS133-10
BCRUS141-10
BENTH277-08
CMARA222-10
CMARA233-10
COAPP150-12
COAPP154-12
COAPP160-12
COAPP178-12
COAPP179-12
GBA11344-13
GBA11345-13
GBA11347-13
GBA11348-13
GBA11349-13
GBA11350-13
GBA11351-13
GBA11353-13
GBA11366-13
GBA11367-13
GBA11368-13
GBA11370-13
GBA11371-13
GBA11372-13
GBA11373-13
GBA11374-13
GBA11375-13
GBA11376-13
GBA11377-13
GBA11378-13
GBA11381-13
GBA11382-13
GBA11383-13
GBA11385-13
GBA11391-13
GBA11410-13
GBA11421-13
GBA11431-13
GBA11446-13
GBA11498-13
GBA11501-13
GBA11505-13
GBA11517-13
GBA11519-13
GBA11520-13
GBA11521-13
GBA11522-13
GBA11523-13
GBA11524-13
GBA11525-13
GBA11526-13
GBA11527-13
GBA9126-13
GBA9127-13
GBA9128-13
GBA9129-13
GBA9130-13
GBA9131-13
GBA9132-13
GBA9133-13
GBA9134-13
GBA9135-13
GBA9136-13
GBA9137-13
GBA9138-13
GBA9139-13
GBA9140-13
GBA9141-13
GBA9142-13
GBA9143-13
GBA9144-13
GBA9145-13
GBA9146-13
GBA9147-13
GBA9148-13
GBA9149-13
GBA9150-13
GBA9151-13
GBA9152-13
GBA9153-13
GBA9154-13
GBA9156-13
GBA9157-13
GBA9158-13
GBA9159-13
GBA9160-13
GBA9162-13
GBA9163-13
GBA9164-13
GBA9165-13
GBA9166-13
GBA9167-13
GBA9168-13
GBA9169-13
GBCX2836-14
GBCX2841-14
GBCX2842-14
GBCX2845-14
GBCX2851-14
GBCX2858-14
GBCX2862-14
GBCX2874-14
GBCX2875-14
GBCX2876-14
GBCX2877-14
GBCX2879-14
GBCX2882-14
GBCX2883-14
GBCX2884-14
GBCX2887-14
GBCX2890-14
GBCX2891-14
GBCX2892-14
GBCX2894-14
GBCX2895-14
GBCX2897-14
GBCX2898-14
GBCX2905-14
GBCX2906-14
GBCX2909-14
GBCX2910-14
GBCX2911-14
GBCX2912-14
GBCX2913-14
GBCX2914-14
GBCX2916-14
GBCX2919-14
GBCX2921-14
GBCX2923-14
GBCX2924-14
GBCX2925-14
GBCX2926-14
GBCX2927-14
GBCX2928-14
GBCX2929-14
GBCX2930-14
NJCGS184-10
NJCGS185-10
NJCGS312-10
NJCGS316-10
NJCGS350-10
NJCGS353-10
NJCGS355-10
NJCGS359-10
NJCGS364-10
NJCGS365-10
NJCGS366-10
NJCGS367-10
NJCGS372-10
NJCGS375-10
NJCGS376-10
NJCGS377-10
NJCGS403-10
NJCGS404-10
NJCGS407-10
NJCGS409-10
NJCGS410-10
NJCGS411-10
NJCGS466-10
NJCGS469-10
NJCGS630-10
NJCGS898-11
NJCGS899-11
NJCGS900-11
NJNAH001-14
NJSAH094-14
NNMC079-08
NNMC110-08
NNMC111-08
NNMC119-08
NNMC120-08
NNMC123-08
NNMC193-08
NNMC199-08
NNMC200-08
NNMC252-08
NNMC253-08
NNMC255-08
SLOBC173-07
SLOBC176-07
SLOBC192-07
TREAR100-12
TREAR103-12
TREAR117-12
TREAR119-12
TREAR126-12
TREAR127-12
TREAR128-12
TREAR136-12
TREAR145-12
TREAR147-12
TREAR174-12
WSCRU008-09
WSCRU009-09
WSCRU010-09
WSCRU011-09
WW820-08
WW832-08
WW839-08
WW937-08
WWAMP1227-13
WWAMP1265-13
WWAMP1274-13
ZMIII1011-12
ZMIII1031-12
ZOOP417-13
ZOOP419-13
ZOOP421-13
ZOOP422-13
ZOOP763-13
ZOOP799-13
ZPAU002-13
ZPAU081-13
ZPC005-13
ZPC009-13
ZPC035-13
ZPC045-13
ZPC047-13
ZPC050-13
ZPC055-13

COAPP181-12	GBA11528-13	GBA9170-13	GBCX2932-14	NNMC257-08	ZPC061-13
COAPP185-12	GBA11529-13	GBA9171-13	GBCX2933-14	NNMC263-08	ZPC063-13
COAPP187-12	GBA11531-13	GBA9172-13	GBCX2935-14	NNMC295-08	ZPC065-13
COAPP188-12	GBA11532-13	GBA9173-13	GBCX2936-14	NNMC296-08	ZPC080-13
COAPP314-13	GBA11801-13	GBA9174-13	GBCX2937-14	NNMC349-08	ZPC083-13
COAPP315-13	GBA11802-13	GBA9175-13	GBCX2938-14	NNMC350-08	ZPC089-13
COAPP327-13	GBA11803-13	GBA9177-13	GBCX2939-14	NNMC351-08	ZPC092-13
COAPP337-13	GBA11804-13	GBA9178-13	GBCX2940-14	NNMC354-08	ZPC103-14
COAPP365-13	GBA11807-13	GBA9179-13	GBCX2941-14	NNMC362-08	ZPC104-14
COAPP376-13	GBA11808-13	GBA9180-13	GBCX2948-14	NNMC374-08	ZPC105-14
COAPP393-13	GBA11809-13	GBA9181-13	GBCX2950-14	NNMC379-08	ZPC110-14
COAPP415-13	GBA11811-13	GBA9182-13	GBCX2951-14	NNMC381-08	ZPC118-14
COAPP430-13	GBA11812-13	GBA9184-13	GBCX2952-14	NNMC422-08	ZPC119-14
COAPP435-13	GBA11813-13	GBA9186-13	GBCX2955-14	NNMC423-08	ZPC122-14
COAPP498-13	GBA11814-13	GBA9188-13	GBCX2956-14	NNMC453-08	ZPC123-14
CRCH032-09	GBA11815-13	GBA9198-13	GBCX2958-14	NNMC463-08	ZPC126-14
CRCH099-09	GBA11816-13	GBA9199-13	GBCX2960-14	NRMMC129-10	ZPC139-14
CRCH122-09	GBA11817-13	GBA9201-13	GBCX2963-14	NZCYC010-13	ZPC146-14
CRCN025-09	GBA11818-13	GBA9203-13	GBCX2964-14	NZCYC026-13	ZPC150-14
CRCN046-09	GBA11820-13	GBA9206-13	GBCX2966-14	NZCYC032-13	ZPC151-14
CRCN047-09	GBA11821-13	GBA9210-13	GBCX2967-14	NZCYC033-13	ZPC152-14
CRCN053-09	GBA11822-13	GBA9958-13	GBCX2968-14	NZCYC035-13	ZPC167-14
CRM090-10	GBA11823-13	GBA9959-13	GBCX2969-14	NZCYC039-13	ZPC181-14
CTM057-10	GBA11825-13	GBA9972-13	GBCX2970-14	NZCYC040-13	ZPC184-14
CTM059-10	GBA11826-13	GBA9973-13	GBCX2971-14	NZCYC045-14	ZPC185-14
CTM082-10	GBA11827-13	GBA9975-13	GBCX2972-14	NZCYC051-14	ZPC188-14
CTM089-10	GBA11828-13	GBA9976-13	GBCX2973-14	NZCYC066-14	ZPC202-14
CTM090-10	GBA11829-13	GBA9977-13	GBCX2977-14	NZCYC082-14	ZPC204-14
CTM094-10	GBA11830-13	GBCX0005-06	GBCX2980-14	NZCYC086-14	ZPC206-14
CTM133-10	GBA11831-13	GBCX0006-06	GBCX2981-14	NZCYC091-14	ZPC215-14
CTM138-10	GBA11832-13	GBCX0007-06	GBCX2982-14	NZCYC095-14	ZPC219-14
CTM215-13	GBA11833-13	GBCX0042-06	GBCX2986-14	NZCYC098-14	ZPC221-14
CTM216-13	GBA11834-13	GBCX0043-06	GBCX2993-14	NZCYC102-14	ZPC241-14
CTM218-13	GBA11835-13	GBCX0280-06	GBCX2994-14	NZCYC105-14	ZPC242-14
CTM221-13	GBA11836-13	GBCX0281-06	GBCX2998-14	NZCYC106-14	ZPC243-14
CTM224-13	GBA11837-13	GBCX0282-06	GBCX3002-14	NZCYC107-14	ZPC244-14
CTM227-13	GBA11838-13	GBCX0283-06	GBCX3003-14	NZCYC108-14	ZPC245-14
CTM228-13	GBA11839-13	GBCX0284-06	GBCX3004-14	NZCYC109-14	ZPC265-14
CTM230-13	GBA11840-13	GBCX0286-06	GBCX3005-14	NZCYC110-14	ZPC270-14
CTM232-13	GBA11841-13	GBCX0289-06	GBCX3007-14	NZCYC111-14	ZPII1002-11
CTM234-13	GBA11843-13	GBCX0290-06	GBCX3009-14	NZCYC112-14	ZPII1175-11
CTM245-13	GBA11844-13	GBCX0292-06	GBCX3010-14	NZCYC113-14	ZPII1189-11

DFO059-11	GBA11845-13	GBCX0293-06	GBCX3011-14	NZCYC114-14	ZPII1196-11
DFO067-11	GBA11846-13	GBCX0294-06	GBCX3012-14	NZCYC119-14	ZPII1198-11
DFO263-11	GBA11847-13	GBCX0295-06	GBCX3013-14	NZCYC127-14	ZPII1219-11
DFO264-11	GBA11848-13	GBCX0297-06	GBCX3014-14	NZCYC128-14	ZPII1223-11
DFO265-11	GBA11850-13	GBCX0298-06	GBCX3015-14	NZPL154-10	ZPII1272-11
DFO266-11	GBA11851-13	GBCX0299-06	GBCX3016-14	NZPL156-10	ZPII1335-11
DFO267-11	GBA11852-13	GBCX0301-06	GBCX3017-14	NZPL182-10	ZPII1340-11
DQCS089-08	GBA11854-13	GBCX0302-06	GBCX3018-14	NZPL198-10	ZPII1344-11
DQCS094-08	GBA11855-13	GBCX0303-06	GBCX3019-14	NZPL199-10	ZPII1345-11
DQCS096-08	GBA11856-13	GBCX0304-06	GBCX3020-14	NZPL200-10	ZPII1356-11
DQCS099-08	GBA11857-13	GBCX0305-06	GBCX3024-14	NZPL201-10	ZPII1359-11
DQCS103-10	GBA11858-13	GBCX0306-06	GBCX3025-14	NZPL202-10	ZPII1360-11
DQCS105-10	GBA11859-13	GBCX0307-06	GBCX3027-14	NZPL203-10	ZPII1361-11
DQCS106-10	GBA11860-13	GBCX0308-06	GBCX3029-14	NZPL205-10	ZPII1374-11
DQCS107-10	GBA11861-13	GBCX0309-06	GBCX3031-14	NZPL214-10	ZPII1388-11
DQCS108-10	GBA11862-13	GBCX0312-06	GBCX3033-14	NZPL222-10	ZPII1410-11
DQCS111-10	GBA11863-13	GBCX0313-06	GBCX3034-14	NZPL226-10	ZPII1414-11
DQCS165-10	GBA11864-13	GBCX0314-06	GBCX3036-14	NZPL227-10	ZPII1486-11
ECCRU034-10	GBA11865-13	GBCX0315-06	GBCX3037-14	NZPL228-10	ZPII1502-11
ECCRU041-10	GBA11867-13	GBCX0316-06	GBCX3038-14	NZPL229-10	ZPII1546-11
ECTCR001-14	GBA11868-13	GBCX0319-06	GBCX3039-14	NZPL230-10	ZPII1555-11
ECTCR002-14	GBA11869-13	GBCX0320-06	GBCX3040-14	NZPL234-10	ZPII635-07
ECTCR003-14	GBA11872-13	GBCX0321-06	GBCX3041-14	NZPL241-10	ZPLCA011-06
ECTCR004-14	GBA11873-13	GBCX0322-06	GBCX3042-14	NZPL244-10	ZPLIV401-11
ECTCR005-14	GBA11880-13	GBCX0323-06	GBCX3043-14	OGL374-11	ZPLIV403-11
ECTCR006-14	GBA11881-13	GBCX0325-06	GBCX3044-14	OGL377-11	ZPLIV414-11
ECTCR007-14	GBA11882-13	GBCX0326-06	GBCX3045-14	OGL378-11	ZPLIV432-11
ECTCR008-14	GBA11883-13	GBCX0328-06	GBCX3046-14	OGL379-11	ZPLIV438-11
ECTCR009-14	GBA11884-13	GBCX0329-06	GBCX3049-14	OGL380-11	ZPLIV457-11
ECTCR010-14	GBA11885-13	GBCX0330-06	GBCX3050-14	OGL381-11	ZPLIV481-11
ECTCR011-14	GBA11890-13	GBCX0331-06	GBCX3051-14	OGL382-11	ZPLIV483-11
ECTCR012-14	GBA11891-13	GBCX0332-06	GBCX3052-14	OGL384-11	ZPLIV485-11
ECTCR013-14	GBA11892-13	GBCX0333-06	GBCX3053-14	OGL385-11	ZPLIV486-11
ECTCR014-14	GBA11893-13	GBCX0334-06	GBCX3054-14	OGL387-11	ZPLIV487-11
ECTCR015-14	GBA11894-13	GBCX0335-06	GBCX3057-14	OGL388-11	ZPLIV570-11
ECTCR018-14	GBA11896-13	GBCX0336-06	GBCX3058-14	OGL389-11	ZPLIV613-11
ECTCR021-14	GBA11897-13	GBCX0338-06	GBCX3063-14	OGL390-11	ZPLIV616-11
ECTCR022-14	GBA11898-13	GBCX0340-06	GBCX3064-14	OGL393-11	ZPLIV627-11
ECTCR023-14	GBA11899-13	GBCX0341-06	GBCX3065-14	OGL394-11	ZPLIV633-11
ECTCR024-14	GBA11900-13	GBCX0342-06	GBCX3066-14	OGL396-11	ZPLIV635-11
ECTCR029-14	GBA1190-04	GBCX0343-06	GBCX3067-14	OGL397-11	ZPLIV680-11
ECTCR032-14	GBA11901-13	GBCX0344-06	GBCX3068-14	OGL398-11	ZPLIV695-11

ECTCR034-14	GBA11902-13	GBCX0345-06	GBCX3069-14	OGL399-11	ZPLIV701-11
ECTCR035-14	GBA11903-13	GBCX0346-06	GBCX3070-14	OGL400-11	ZPLIV705-11
ECTCR037-14	GBA11904-13	GBCX0347-06	GBCX3071-14	OGLA1733-11	ZPLIV708-11
ECTCR038-14	GBA11905-13	GBCX0348-06	GBCX3073-14	OGLA1750-11	ZPLIV713-11
ECTCR039-14	GBA11906-13	GBCX0349-06	GBCX3074-14	OGLA1751-11	ZPLIV732-11
ECTCR041-14	GBA11907-13	GBCX0353-06	GBCX3075-14	OGLA1766-11	ZPLIV744-11
ECTCR051-14	GBA11908-13	GBCX0354-06	GBCX3076-14	OGLA1768-11	ZPLMX165-06
ECTCR052-14	GBA11911-13	GBCX0358-06	GBCX3077-14	OGLA1769-11	ZPLMX166-06
ECTCR053-14	GBA11912-13	GBCX0359-06	GBCX3078-14	OGLA1788-11	ZPLMX170-06
ECTCR054-14	GBA11913-13	GBCX0361-06	GBCX3079-14	OGLA1789-11	ZPLMX172-06
ECTCR055-14	GBA11916-13	GBCX0363-06	GBCX3080-14	OGLA3125-13	ZPLMX175-06
ECTCR058-14	GBA11949-13	GBCX0364-06	GBCX3081-14	OGLA3126-13	ZPLMX182-06
ECTCR061-14	GBA11950-13	GBCX0365-06	GBCX3082-14	OGLA3127-13	ZPLMX194-06
ECTCR063-14	GBA11951-13	GBCX0370-06	GBCX3084-14	OGLA3128-13	ZPLMX212-06
ECTCR064-14	GBA11952-13	GBCX0372-06	GBCX3086-14	OGLA3129-13	ZPLMX214-06
ECTCR065-14	GBA11953-13	GBCX0373-06	GBCX3087-14	OGLA3130-13	ZPLMX215-06
ECTCR067-14	GBA11954-13	GBCX0374-06	GBCX3088-14	OGLA3133-13	ZPLMX218-06
ECTCR070-14	GBA11955-13	GBCX0375-06	GBCX3089-14	OGLA3135-13	ZPLMX219-06
ECTCR073-14	GBA11956-13	GBCX0379-06	GBCX3090-14	OGLA3137-13	ZPLMX220-06
ECTCR078-14	GBA11957-13	GBCX0380-06	GBCX3093-14	OGLA3139-13	ZPLMX221-06
ECTCR079-14	GBA11958-13	GBCX0381-06	GBCX3104-14	OGLA3141-13	ZPLMX226-06
ECTCR080-14	GBA11959-13	GBCX0383-06	GBCX3105-14	OGLA3145-13	ZPLMX233-06
ECTCR086-14	GBA11960-13	GBCX0384-06	GBCX3113-14	OGLA3146-13	ZPLMX247-06
ECTCR090-14	GBA11961-13	GBCX0385-06	GBCX3116-14	OGLA3147-13	ZPLMX512-06
ECTCR091-14	GBA11966-13	GBCX0386-06	GBCX3117-14	OGLA3148-13	ZPLMX513-06
ECTCR094-14	GBA1250-04	GBCX0387-06	GBCX3118-14	OGLA3149-13	ZPLMX514-06
ECTCR095-14	GBA12637-13	GBCX0388-06	GBCX3119-14	OGLA3150-13	ZPLMX517-06
ECTCR106-14	GBA12638-13	GBCX0392-06	GBCX3120-14	OGLA3151-13	ZPLMX520-06
ECTCR108-14	GBA12639-13	GBCX0394-06	GBCX618-12	OGLA3152-13	ZPLMX521-06
ECTCR109-14	GBA12642-13	GBCX0399-06	GBCX634-12	OGLA3154-13	ZPLMX523-06
ECTCR130-14	GBA12643-13	GBCX0401-06	GBCX636-12	OGLA3155-13	ZPLMX526-06
ECTCR153-14	GBA12644-13	GBCX0403-06	GBCX637-12	OGLA3161-13	ZPLMX529-06
ECTCR154-14	GBA12647-13	GBCX0404-06	GBCX638-12	OGLA3170-13	ZPLMX530-06
ECTCR156-14	GBA12648-13	GBCX0405-06	GBCX640-12	OGLA3171-13	ZPLMX533-06
ECTCR157-14	GBA12649-13	GBCX0406-06	GBCX646-12	OGLA3172-13	ZPLMX536-06
ECTCR185-14	GBA12650-13	GBCX0408-06	GBCX647-12	OGLA3174-13	ZPLMX540-06
EES002-12	GBA12652-13	GBCX0410-06	GBCX649-12	OGLA3175-13	ZPLMX542-06
EES007-12	GBA12653-13	GBCX0412-06	GBCX654-12	OGLA3178-13	ZPLMX544-06
EES013-12	GBA12654-13	GBCX0413-06	GBCX669-12	OGLA3180-13	ZPLMX546-06
EES015-12	GBA12657-13	GBCX0414-06	GBCX671-12	OGLA3181-13	ZPLMX547-06
EES016-12	GBA12658-13	GBCX0417-06	GBCX673-12	OGLA3182-13	ZPLMX549-06
EES017-12	GBA12660-13	GBCX0418-06	GBCX674-12	OGLA3184-13	ZPLMX551-06

EES019-12	GBA12661-13	GBCX0419-06	GBCX682-12	OGLA3185-13	ZPLMX553-06
EES026-12	GBA12662-13	GBCX0420-06	GBCX684-12	OGLA3192-13	ZPLMX555-06
EES032-12	GBA12663-13	GBCX0421-06	GBCX688-12	OGLA3194-13	ZPLMX558-06
EES041-12	GBA12666-13	GBCX0423-06	GBCX720-12	OGLA3205-13	ZPLMX561-06
EES043-12	GBA12667-13	GBCX0424-06	GBCX721-12	OGLA3206-13	ZPLMX562-06
EES050-12	GBA12668-13	GBCX0426-06	GBCX722-12	OGLA3209-13	ZPLMX564-06
EES051-12	GBA12669-13	GBCX0428-06	GBCX723-12	OGLA3210-13	ZPLMX566-06
EES052-12	GBA12670-13	GBCX0429-06	GBCX724-12	OGLA3212-13	ZPLMX568-06
EES053-12	GBA12671-13	GBCX0431-06	GBCX725-12	OGLA3213-13	ZPLMX716-06
ELPPC010-09	GBA12672-13	GBCX0436-06	GBCX727-12	OZFWC119-11	ZPLMX722-06
ELPPC045-09	GBA12673-13	GBCX0437-06	GBCX728-12	OZFWC140-11	ZPLMX759-06
ELPPC061-09	GBA12674-13	GBCX0444-06	GBCX730-12	OZFWC147-11	ZPLMX772-06
ELPPC063-09	GBA12675-13	GBCX0445-06	GBCX732-12	OZFWC191-11	ZPLMX795-06
ELPPC071-09	GBA12676-13	GBCX0446-06	GBCX733-12	OZFWC212-11	ZPLMX803-06
ELPPC073-09	GBA12679-13	GBCX0448-06	GBCX734-12	OZFWC216-11	ZPLMX815-06
ELPPC092-09	GBA12680-13	GBCX0449-06	GBCX737-12	OZFWC221-11	ZPLMX816-06
ELPPC093-09	GBA12681-13	GBCX0455-06	GBCX745-12	OZFWC261-11	ZPLMX863-06
ELPPC094-09	GBA12682-13	GBCX0458-06	GBCX780-12	OZFWC288-11	ZPLMX865-06
ELPPC105-09	GBA12689-13	GBCX1182-14	GBCX781-12	OZFWC355-11	ZPLMX866-06
ELPPC116-09	GBA1269-04	GBCX1183-14	GBCX782-12	OZFWC363-11	ZPLMX908-06
FCCOM028-09	GBA12693-13	GBCX1185-14	GBCX784-12	OZFWC370-11	ZPLMX911-06
FCCOM029-09	GBA12694-13	GBCX1208-14	GBCX785-12	OZFWC387-11	ZPLMX912-06
FCCOM030-09	GBA12700-13	GBCX1251-14	GBFCC0004-06	OZFWC388-11	ZPLMX924-06
FCCOM068-09	GBA12704-13	GBCX1271-14	GBFCC0005-06	OZFWC409-11	ZPLMX933-06
FCFC001-04	GBA12705-13	GBCX1272-14	GBFCC664-13	OZFWC467-11	ZPLMX994-06
FCFC035-04	GBA12706-13	GBCX1287-14	GBFCC665-13	OZFWC508-12	CAISN1053-13
FCFC058-04	GBA12709-13	GBCX1288-14	GBFCC687-13	OZFWC550-12	CAISN1058-13
FCFC059-04	GBA12712-13	GBCX1289-14	GBFCC688-13	OZFWC662-12	CAISN1059-13
FCFC083-04	GBA12715-13	GBCX1291-14	GBFCC689-13	OZFWC688-12	CAISN1061-13
FCFC084-04	GBA12716-13	GBCX1340-14	GBFCC690-13	OZFWC689-12	CAISN1062-13
FCFC085-04	GBA12717-13	GBCX1341-14	GBFCC691-13	OZFWC701-12	CAISN1063-13
FCFC087-04	GBA12719-13	GBCX1342-14	GBFCC692-13	OZFWC756-12	CAISN1065-13
FPMAR125-08	GBA12720-13	GBCX1343-14	GBFCC693-13	OZFWC788-12	CAISN1069-13
FPMAR126-08	GBA12721-13	GBCX1344-14	GBFCC694-13	OZFWC793-12	CAISN1080-13
FPMAR128-08	GBA12722-13	GBCX1345-14	GBFCC695-13	OZFWC864-12	CAISN1090-13
FPMAR136-08	GBA12723-13	GBCX1346-14	GBFCC697-13	OZFWC938-12	CAISN1107-13
FPMAR137-08	GBA12724-13	GBCX1439-14	GBFCC702-13	OZFWC952-12	CAISN1109-13
FPMAR143-08	GBA12728-13	GBCX1566-14	GBFCC703-13	OZFWZ330-11	CAISN1118-13
FPMAR208-08	GBA12729-13	GBCX1567-14	GBFCC704-13	OZFWZ341-11	CAISN1132-13
FPMAR209-08	GBA12732-13	GBCX1569-14	GBFCC706-13	OZFWZ357-11	CAISN1139-13
GBA10011-13	GBA12740-13	GBCX1570-14	GBFCC708-13	OZFWZ401-11	CAISN1335-13
GBA10099-13	GBA12743-13	GBCX1572-14	GBFCC710-13	OZFWZ421-11	CAISN1336-13

GBA10100-13
GBA10172-13
GBA10177-13
GBA10179-13
GBA10180-13
GBA10181-13
GBA10182-13
GBA10184-13
GBA10185-13
GBA10186-13
GBA10187-13
GBA10368-13
GBA10369-13
GBA10372-13
GBA10376-13
GBA10377-13
GBA10380-13
GBA10381-13
GBA10386-13
GBA10400-13
GBA10401-13
GBA10402-13
GBA10403-13
GBA10404-13
GBA10405-13
GBA10406-13
GBA10407-13
GBA10408-13
GBA10409-13
GBA10410-13
GBA10411-13
GBA10412-13
GBA10413-13
GBA10414-13
GBA10415-13
GBA10416-13
GBA10417-13
GBA10418-13
GBA10420-13
GBA10421-13
GBA10422-13
GBA10423-13
GBA12744-13
GBA12748-13
GBA12759-13
GBA12763-13
GBA12764-13
GBA12767-13
GBA12780-13
GBA12783-13
GBA12787-13
GBA12789-13
GBA12793-13
GBA12794-13
GBA12810-13
GBA13684-13
GBA13690-13
GBA13692-13
GBA14007-13
GBA14011-13
GBA14027-13
GBA14032-13
GBA14070-13
GBA14071-13
GBA14086-13
GBA14088-13
GBA14096-13
GBA14098-13
GBA14108-13
GBA14118-13
GBA14123-13
GBA14124-13
GBA14125-13
GBA14128-13
GBA14134-13
GBA14135-13
GBA14144-13
GBA14147-13
GBA14151-13
GBA14157-13
GBA14165-13
GBA14167-13
GBA14173-13
GBA14175-13
GBCX1573-14
GBCX1576-14
GBCX1577-14
GBCX1578-14
GBCX1583-14
GBCX1584-14
GBCX1585-14
GBCX1586-14
GBCX1587-14
GBCX1588-14
GBCX1589-14
GBCX1590-14
GBCX1591-14
GBCX1592-14
GBCX1593-14
GBCX1594-14
GBCX1595-14
GBCX1596-14
GBCX1597-14
GBCX1598-14
GBCX1599-14
GBCX1600-14
GBCX1601-14
GBCX1604-14
GBCX1605-14
GBCX1606-14
GBCX1608-14
GBCX1609-14
GBCX1613-14
GBCX1614-14
GBCX1615-14
GBCX1618-14
GBCX1619-14
GBCX1620-14
GBCX1621-14
GBCX1623-14
GBCX1625-14
GBCX1626-14
GBCX1627-14
GBCX1628-14
GBCX1654-14
GBCX1659-14
GBFCC714-13
GBFCC718-13
GBFCC720-13
GBFCC721-13
GBFCC723-13
GBFCC724-13
GBFCC732-13
GBFCC735-13
GBFCC737-13
GBFCC742-13
GBFCC745-13
GBFCC750-13
GBFCC752-13
GBFCC755-13
GBFCC758-13
GBFCC759-13
GBFCC760-13
GBFCC762-13
GBFCC767-13
GBFCC769-13
GBFCC770-13
GBFCC771-13
GBFCC775-13
GBFCC781-13
GBFCC783-13
GBFCC784-13
GBFCC787-13
GBFCD001-14
GBFCD005-14
GBFCD006-14
GBFCD008-14
GBFCD055-14
GBFCM0061-06
GBFCM075-07
GBFCM076-07
GBFCM103-13
GBFCM104-13
GBFCM106-13
GBFCM108-13
GBFCM110-13
GBFCM111-13
GBFCM112-13
OZFWZ449-11
OZFWZ458-11
OZFWZ459-11
OZFWZ466-11
PCQC023-08
PCQC025-08
PCQC027-08
PCQC029-08
PCQC031-08
PDMX028-11
PDMX034-11
PDMX035-11
PDMX036-11
PDMX039-11
PDMX059-11
PDMX063-11
PDMX065-11
PDMX066-11
PDMX068-11
PDMX075-11
PDMX076-11
PDMX077-11
PDMX078-11
PDMX079-11
PDMX080-11
PDMX081-11
PDMX092-11
RBGC099-03
RBGC100-03
RBGC101-03
RBGC102-03
RBGC103-03
RBGC104-03
RBGC105-03
RBGC106-03
RBGC107-03
RBGC109-03
RBGC110-03
RBGC112-03
RDCNA001-06
RDCNA002-06
RDCNA003-06
CAISN1337-13
CAISN1339-13
CAISN1343-13
CAISN1347-13
CAISN1349-13
CAISN1356-13
CAISN1357-13
CAISN1358-13
CAISN1359-13
CAISN1366-13
CAISN1369-13
CAISN1377-13
CAISN1382-13
CAISN152-12
CAISN177-12
CAISN189-12
CAISN190-12
CAISN197-12
CAISN218-12
CAISN229-12
CAISN231-12
CAISN235-12
CAISN239-12
CAISN240-12
CAISN245-12
CAISN248-12
CAISN249-12
CAISN256-12
CAISN258-12
CAISN263-12
CAISN268-12
CAISN282-12
CAISN284-12
CAISN286-12
CAISN288-12
CAISN294-12
CAISN299-12
CAISN303-12
CAISN308-12
CAISN317-12
CAISN318-12
CAISN321-12

GBA10424-13	GBA14176-13	GBCX1661-14	GBFCM113-13	RDCNA004-06	CAISN322-12
GBA10425-13	GBA14178-13	GBCX1667-14	GBFCM115-13	RDCNA007-06	CAISN325-12
GBA10426-13	GBA14182-13	GBCX1668-14	GBFCM116-13	RDCNA008-06	CAISN326-12
GBA10427-13	GBA14185-13	GBCX1680-14	GBFCM118-13	RDCNA009-06	CAISN331-12
GBA10428-13	GBA14188-13	GBCX1681-14	GBFCM122-13	RDCNA011-06	CAISN332-12
GBA10429-13	GBA14189-13	GBCX1682-14	GBMIN41407-14	RDCNA015-06	CAISN339-12
GBA10430-13	GBA14192-13	GBCX1683-14	GBMIN41408-14	RDCNA017-06	CAISN340-12
GBA10431-13	GBA14193-13	GBCX1772-14	GBMIN41409-14	RDCNA018-06	CAISN352-12
GBA10434-13	GBA14196-13	GBCX1777-14	GBMIN41411-14	RDCNA031-06	CAISN354-12
GBA10435-13	GBA14197-13	GBCX1778-14	GBMIN41414-14	RDCNA048-06	CAISN355-12
GBA10436-13	GBA14201-13	GBCX1779-14	GBMIN41418-14	RDCNA054-06	CAISN356-12
GBA10439-13	GBA14202-13	GBCX1780-14	GBMIN41420-14	RDCNA056-06	CAISN357-12
GBA10445-13	GBA14209-13	GBCX1792-14	GBMIN41421-14	RDCNA057-06	CAISN362-12
GBA10449-13	GBA14211-13	GBCX1795-14	GBMIN41422-14	RDCNA058-06	CAISN364-12
GBA10453-13	GBA14215-13	GBCX1800-14	GBMIN41423-14	RDCNA059-06	CAISN365-12
GBA10462-13	GBA14216-13	GBCX1801-14	GBMIN41425-14	RDCNA065-06	CAISN366-12
GBA10463-13	GBA14217-13	GBCX1816-14	GBMIN41426-14	RDCNA068-06	CAISN369-12
GBA10466-13	GBA14220-13	GBCX1817-14	GBMIN41427-14	RDCNA073-06	CAISN370-12
GBA10467-13	GBA1423-04	GBCX1830-14	GBMIN41428-14	RDCNA074-06	CAISN371-12
GBA10468-13	GBA14238-13	GBCX1831-14	GBMIN41429-14	RDCNA075-06	CAISN372-12
GBA10470-13	GBA14249-13	GBCX1900-14	GBMIN41430-14	RDCNA076-06	CAISN374-12
GBA10471-13	GBA14251-13	GBCX1901-14	GBMIN41431-14	RDCNA077-06	CAISN376-12
GBA10472-13	GBA14258-13	GBCX1902-14	GBMIN41432-14	RDCNA078-06	CAISN377-12
GBA10473-13	GBA14265-13	GBCX1903-14	GBMIN41433-14	RDCNA079-06	CAISN378-12
GBA10474-13	GBA14266-13	GBCX1904-14	GBMIN41434-14	RDCNA080-06	CAISN385-13
GBA10477-13	GBA14267-13	GBCX1905-14	GBMIN41435-14	RDCNA084-06	CAISN394-13
GBA10573-13	GBA14268-13	GBCX1908-14	GBMIN41436-14	RDCNA085-06	CAISN395-13
GBA10574-13	GBA14269-13	GBCX1909-14	GBMIN41443-14	RDCNA087-06	CAISN399-13
GBA10581-13	GBA14270-13	GBCX1910-14	GBMIN41444-14	RDCNA089-06	CAISN425-13
GBA10583-13	GBA14271-13	GBCX1911-14	GBMIN41445-14	RDCNA092-06	CAISN447-13
GBA10652-13	GBA14273-13	GBCX1920-14	GBMIN41446-14	RDCNA094-06	CAISN475-13
GBA10653-13	GBA14274-13	GBCX1921-14	GBMIN41447-14	RDCNA097-06	CAISN571-13
GBA10654-13	GBA14275-13	GBCX1925-14	GBMIN41449-14	RDCNA101-06	CAISN573-13
GBA10655-13	GBA14278-13	GBCX1926-14	GBMIN41454-14	RDCNA102-06	CAISN578-13
GBA10656-13	GBA14280-13	GBCX1927-14	GBMIN41455-14	RDCNA105-06	CAISN579-13
GBA10657-13	GBA14281-13	GBCX1928-14	GBMIN41456-14	RDCNA106-06	CAISN586-13
GBA10658-13	GBA14284-13	GBCX1929-14	GBMIN41457-14	RDCNA111-06	CAISN588-13
GBA10659-13	GBA14285-13	GBCX1930-14	GBMIN41458-14	RDCNA113-06	CAISN593-13
GBA10660-13	GBA14286-13	GBCX1931-14	GBMIN41459-14	RDCNA116-06	CAISN612-13
GBA10661-13	GBA14287-13	GBCX1934-14	GBMIN41460-14	RDCNA118-06	CAISN625-13
GBA10662-13	GBA14289-13	GBCX1935-14	GBMIN41462-14	RDCNA119-06	CAISN626-13
GBA10663-13	GBA14290-13	GBCX1936-14	GBMIN41463-14	RDMPC004-10	CAISN628-13

GBA10664-13	GBA14291-13	GBCX1937-14	GBMIN41464-14	RDMPC046-10	CAISN630-13
GBA10665-13	GBA14292-13	GBCX1938-14	GBMIN41465-14	RDMPC056-10	CAISN632-13
GBA10666-13	GBA14293-13	GBCX1939-14	GBMIN41466-14	RDMPC072-10	CAISN633-13
GBA10667-13	GBA14295-13	GBCX1940-14	GBMIN41467-14	RDMPC113-10	CAISN658-13
GBA10668-13	GBA14296-13	GBCX1941-14	GBMIN41468-14	RDMPC114-10	CAISN660-13
GBA10669-13	GBA14299-13	GBCX1942-14	GBMIN41470-14	RDMPC137-10	CAISN672-13
GBA10670-13	GBA14300-13	GBCX1943-14	GBMIN41471-14	RDMPC164-10	CAISN673-13
GBA10671-13	GBA14301-13	GBCX1944-14	GBMIN41472-14	RDMPC185-10	CAISN710-13
GBA10672-13	GBA14302-13	GBCX1945-14	GBMIN41473-14	SACOP037-08	CAISN716-13
GBA10673-13	GBA14303-13	GBCX1946-14	GBMIN41474-14	SACOP044-08	CAISN717-13
GBA10674-13	GBA14304-13	GBCX1947-14	GBMIN41475-14	SACOP069-08	CAISN720-13
GBA10675-13	GBA14305-13	GBCX1948-14	GBMIN41476-14	SACOP079-08	CAISN746-13
GBA10676-13	GBA14307-13	GBCX1949-14	GBMIN41479-14	SACOP091-08	CAISN749-13
CAISN1233-13	CAISN1218-13	CAISN1214-13	CAISN861-13	CAISN751-13	

COI

CAISN1051-13	CAISN315-12	CAISN714-13	CAISN1445-14	CAISN1226-13	CAISN546-13
CAISN1052-13	CAISN316-12	CAISN726-13	CAISN1454-14	CAISN1227-13	CAISN614-13
CAISN1054-13	CAISN323-12	CAISN727-13	CAISN1455-14	CAISN1228-13	CAISN617-13
CAISN1055-13	CAISN324-12	CAISN729-13	CAISN1456-14	CAISN1229-13	CAISN619-13
CAISN1056-13	CAISN327-12	CAISN732-13	CAISN1457-14	CAISN1230-13	CAISN623-13
CAISN1057-13	CAISN350-12	CAISN735-13	CAISN1467-14	CAISN1232-13	CAISN693-13
CAISN1060-13	CAISN360-12	CAISN739-13	CAISN1469-14	CAISN1322-13	CAISN696-13
CAISN1085-13	CAISN361-12	CAISN761-13	CAISN1483-14	CAISN1067-13	CAISN700-13
CAISN1094-13	CAISN382-13	CAISN762-13	CAISN1486-14	CAISN1102-13	CAISN709-13
CAISN1105-13	CAISN383-13	CAISN763-13	CAISN1487-14	CAISN1103-13	CAISN713-13
CAISN1131-13	CAISN389-13	CAISN764-13	CAISN1491-14	CAISN1104-13	CAISN715-13
CAISN1133-13	CAISN392-13	CAISN765-13	CAISN1508-14	CAISN1137-13	CAISN725-13
CAISN1136-13	CAISN397-13	CAISN766-13	CAISN1509-14	CAISN1376-13	CAISN728-13
CAISN1140-13	CAISN401-13	CAISN767-13	CAISN1510-14	CAISN1379-13	CAISN864-13
CAISN1352-13	CAISN402-13	CAISN768-13	CAISN1511-14	CAISN208-12	CAISN869-13
CAISN1353-13	CAISN428-13	CAISN773-13	CAISN1513-14	CAISN254-12	CAISN872-13
CAISN1368-13	CAISN435-13	CAISN777-13	CAISN1519-14	CAISN297-12	CAISN875-13
CAISN1378-13	CAISN436-13	CAISN780-13	CAISN1146-13	CAISN301-12	CAISN1438-14
CAISN1393-13	CAISN444-13	CAISN781-13	CAISN1148-13	CAISN302-12	CAISN1471-14
CAISN1394-13	CAISN473-13	CAISN783-13	CAISN1149-13	CAISN304-12	CAISN1472-14
CAISN1413-13	CAISN545-13	CAISN794-13	CAISN1150-13	CAISN310-12	CAISN1473-14
CAISN1417-13	CAISN572-13	CAISN835-13	CAISN1152-13	CAISN312-12	CAISN1488-14
CAISN980-13	CAISN587-13	CAISN856-13	CAISN1161-13	CAISN359-12	CAISN1489-14
CAISN1040-13	CAISN590-13	CAISN857-13	CAISN1163-13	CAISN373-12	CAISN1490-14
CAISN205-12	CAISN594-13	CAISN858-13	CAISN1164-13	CAISN379-12	CAISN1515-14

CAISN212-12	CAISN595-13	CAISN859-13	CAISN1169-13	CAISN381-13	CAISN1520-14
CAISN246-12	CAISN596-13	CAISN860-13	CAISN1170-13	CAISN384-13	CAISN1145-13
CAISN271-12	CAISN604-13	CAISN863-13	CAISN1176-13	CAISN390-13	CAISN1147-13
CAISN273-12	CAISN613-13	CAISN865-13	CAISN1179-13	CAISN391-13	CAISN1151-13
CAISN274-12	CAISN616-13	CAISN866-13	CAISN1184-13	CAISN398-13	CAISN1188-13
CAISN277-12	CAISN618-13	CAISN867-13	CAISN1185-13	CAISN404-13	CAISN1194-13
CAISN278-12	CAISN620-13	CAISN868-13	CAISN1186-13	CAISN434-13	CAISN1195-13
CAISN287-12	CAISN622-13	CAISN873-13	CAISN1187-13	CAISN437-13	CAISN1196-13
CAISN291-12	CAISN648-13	CAISN874-13	CAISN1193-13	CAISN439-13	CAISN1204-13
CAISN293-12	CAISN686-13	CAISN1427-14	CAISN1202-13	CAISN440-13	CAISN1205-13
CAISN298-12	CAISN687-13	CAISN1431-14	CAISN1203-13	CAISN442-13	CAISN1206-13
CAISN300-12	CAISN688-13	CAISN1432-14	CAISN1210-13	CAISN443-13	CAISN1207-13
CAISN305-12	CAISN692-13	CAISN1433-14	CAISN1221-13	CAISN445-13	CAISN1208-13
CAISN309-12	CAISN704-13	CAISN1441-14	CAISN1222-13	CAISN472-13	CAISN1211-13
CAISN311-12	CAISN705-13	CAISN1443-14	CAISN1223-13	CAISN543-13	CAISN1225-13
CAISN314-12	CAISN712-13	CAISN1444-14	CAISN1224-13	CAISN544-13	CAISN1231-13
CAISN1323-13					

Appendix II: Clustering commands

UPARSE commands

```
usearch -cluster_otus A.fas -otu_radius_pct 0.1 -otus out_0.1per.fas -uparseout out0.1per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.2 -otus out_0.2per.fas -uparseout out0.2per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.3 -otus out_0.3per.fas -uparseout out0.3per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.4 -otus out_0.4per.fas -uparseout out0.4per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.5 -otus out_0.5per.fas -uparseout out0.5per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.6 -otus out_0.6per.fas -uparseout out0.6per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.7 -otus out_0.7per.fas -uparseout out0.7per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.8 -otus out_0.8per.fas -uparseout out0.8per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 0.9 -otus out_0.9per.fas -uparseout out0.9per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1 -otus out_1per.fas -uparseout out1per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.1 -otus out_1.1per.fas -uparseout out1.1per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.2 -otus out_1.2per.fas -uparseout out1.2per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.3 -otus out_1.3per.fas -uparseout out1.3per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.4 -otus out_1.4per.fas -uparseout out1.4per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.5 -otus out_1.5per.fas -uparseout out1.5per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.6 -otus out_1.6per.fas -uparseout out1.6per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.7 -otus out_1.7per.fas -uparseout out1.7per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.8 -otus out_1.8per.fas -uparseout out1.8per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 1.9 -otus out_1.9per.fas -uparseout out1.9per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2 -otus out_2per.fas -uparseout out2per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.1 -otus out_2.1per.fas -uparseout out2.1per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.2 -otus out_2.2per.fas -uparseout out2.2per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.3 -otus out_2.3per.fas -uparseout out2.3per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.4 -otus out_2.4per.fas -uparseout out2.4per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.5 -otus out_2.5per.fas -uparseout out2.5per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.6 -otus out_2.6per.fas -uparseout out2.6per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.7 -otus out_2.7per.fas -uparseout out2.7per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.8 -otus out_2.8per.fas -uparseout out2.8per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 2.9 -otus out_2.9per.fas -uparseout out2.9per.log -sizein -sizeout
usearch -cluster_otus A.fas -otu_radius_pct 3 -otus out_3per.fas -uparseout out3per.log -sizein -sizeout
usearch -cluster_smallmem out_2per.fas -id 0.965 -centroids out_3.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.960 -centroids out_4per.fas
usearch -cluster_smallmem out_2per.fas -id 0.955 -centroids out_4.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.950 -centroids out_5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.945 -centroids out_5.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.940 -centroids out_6per.fas
usearch -cluster_smallmem out_2per.fas -id 0.935 -centroids out_6.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.930 -centroids out_7per.fas
usearch -cluster_smallmem out_2per.fas -id 0.925 -centroids out_7.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.920 -centroids out_8per.fas
usearch -cluster_smallmem out_2per.fas -id 0.915 -centroids out_8.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.910 -centroids out_9per.fas
usearch -cluster_smallmem out_2per.fas -id 0.905 -centroids out_9.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.900 -centroids out_10per.fas
usearch -cluster_smallmem out_2per.fas -id 0.895 -centroids out_10.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.890 -centroids out_11per.fas
usearch -cluster_smallmem out_2per.fas -id 0.885 -centroids out_11.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.880 -centroids out_12per.fas
usearch -cluster_smallmem out_2per.fas -id 0.875 -centroids out_12.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.870 -centroids out_13per.fas
usearch -cluster_smallmem out_2per.fas -id 0.865 -centroids out_13.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.860 -centroids out_14per.fas
usearch -cluster_smallmem out_2per.fas -id 0.855 -centroids out_14.5per.fas
```

usearch -cluster_smallmem out_2per.fas -id 0.850 -centroids out_15per.fas
usearch -cluster_smallmem out_2per.fas -id 0.845 -centroids out_15.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.840 -centroids out_16per.fas
usearch -cluster_smallmem out_2per.fas -id 0.835 -centroids out_16.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.830 -centroids out_17per.fas
usearch -cluster_smallmem out_2per.fas -id 0.825 -centroids out_17.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.820 -centroids out_18per.fas
usearch -cluster_smallmem out_2per.fas -id 0.815 -centroids out_18.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.810 -centroids out_19per.fas
usearch -cluster_smallmem out_2per.fas -id 0.805 -centroids out_19.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.800 -centroids out_20per.fas
usearch -cluster_smallmem out_2per.fas -id 0.795 -centroids out_20.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.790 -centroids out_21per.fas
usearch -cluster_smallmem out_2per.fas -id 0.785 -centroids out_21.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.780 -centroids out_22per.fas
usearch -cluster_smallmem out_2per.fas -id 0.775 -centroids out_22.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.770 -centroids out_23per.fas
usearch -cluster_smallmem out_2per.fas -id 0.765 -centroids out_23.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.760 -centroids out_24per.fas
usearch -cluster_smallmem out_2per.fas -id 0.755 -centroids out_24.5per.fas
usearch -cluster_smallmem out_2per.fas -id 0.750 -centroids out_25per.fas

Mothur Commands

dist.seqs(fasta=A.fas,output=lt)
cluster(phylip=A.phylip.dist,method=nearest,cutoff=0.25,hard=t,precision=1000)
cluster(phylip=A.phylip.dist,method=furthest,cutoff=0.25,hard=t,precision=1000)
cluster(phylip=A.phylip.dist,method=average,cutoff=0.25,hard=t,precision=1000)

ABGD

./abgd -p0.001 -P0.15 -n1000 -a -d3 -X 0.001 A.fas | tee A.LOG

Appendix III: Agreement analysis R commands

```
rm(list=ls())

# prompting to choose the file of interest
Cluster_file<-file.choose()

# load in the file data for the lines
cluster_data<-read.table(Cluster_file,header=T,skip=1,sep="\t",dec=".")

#Pulling all unique entries in the first column into the variable count to use as counter for the loop

count<- unique(cluster_data[,"C1"])

#Initalizing the two vectors to store the results
x<-c()
y<-c()

#Initalizing the table to eventually be printed to file

data <- data.frame(first=count)
data$first<- NULL

for (i in 1:length(count)) {

        # using subset function to get all column 1 clusters of the same value
        newdata <- subset(cluster_data, C1 == count[i], select=c(C1,C2))

        #get all unique variables in column 2 and store in array(Col2_check)
        Col2_check<- unique(newdata[,"C2"])

        # This if statement entered if there is only a single value in column 2 for the element in column 1

        if(length(Col2_check) ==1){

                # using subset get all column 1 clusters with the same value in column 2 from all items

                lumpcheck <- subset(cluster_data, C2 == Col2_check[1], select=c(C1,C2))

                #get all unique variables in column 1 and store in Col1_check
                Col1_check<- unique(lumpcheck[,"C1"])

                #if the number of values is equal to 1 for all possible column 2 values then exact match

                if(length(Col1_check) <2){
                        x<-append(x,count[i])
                        y<-append(y,"MATCH")
                }

                #initalize the lumpcount for the beginning of the loop
                lump_count = 0
                #if there was more than one value in column 1 matching in column 2
                if(length(Col1_check)>1){

                        #initalize the j for the beginning of the loop
                        j=0
```

```
                # Start looping through all column 1 values to check column 2
                for (j in 1:length(Col1_check)){

                        #Look at the values in column 2 for Col1_check[j] if there is only a single value
                        #then this value can be dismissed as a possible MIX and move on to the next

                        # using subset function to get all cloumn 1 clusters of the same value
                        newlumpdata <- subset(cluster_data, C1 == Col1_check[j], select=c(C1,C2))

                        #get all unique variables in column 2 and store in array(Col2_lumpcheck)
                        Col2_lumpcheck<- unique(newlumpdata[,"C2"])

                        if(length(Col2_lumpcheck) > 1){
                                lump_count = lump_count + 1
                        }
                        rm(Col2_lumpcheck)
                }

                if(lump_count <1){
                        x<-append(x,count[i])
                        y<-append(y,"LUMP")
                }
                if(lump_count > 0){
                        x<-append(x,count[i])
                        y<-append(y,"MIX")
                }
        }

rm(Col1_check, lumpcheck)

}

#initalize the split_count for the beginning of the loop
split_count = 0

if(length(Col2_check) > 1){

        #Check to see if things are split first. So loop through all possible values in column 2

        #if each value in column two only has values in column 1 equal to i then I have a SPLIT

        #if not then I have a MIX
        j<-0
        for (j in 1:length(Col2_check)){

                #Look at the values in column 1 for Col2_check[j] and if there is only a single value

                #then this value can be dismissed as a possible MIX and move on to the next

                # using subset function to get all column 1 clusters of the same value
                newsplitdata <- subset(cluster_data, C2 == Col2_check[j], select=c(C1,C2))

                #get all unique variables in column 1 and store in array(Col2_splitcheck)
                Col2_splitcheck<- unique(newsplitdata[,"C1"])
```

```
                        if(length(Col2_splitcheck) > 1){
                                split_count = split_count + 1
                        }
                                rm(Col2_splitcheck)
                }
                if(split_count < 1){
                        x<-append(x,count[i])
                        y<-append(y,"SPLIT")
                }
                if(split_count > 0){
                        x<-append(x,count[i])
                        y<-append(y,"MIX")
                }
        }
        rm(Col2_check,newdata)
}

data$Morph<- c(x)
data$C2<-c(y)
```

Appendix IV: Sequence alignment files

List of all process identifiers (BOLD database sequence unique identifiers) for the Chapter 3 *COI* and *18S* data sets and National Center for Biotechnology Information (NCBI) accession numbers (NCBI sequence unique identifiers) for the *18S reference* data set.

***18S* and *COI* data sets**					
CAISN1051-13	CAISN208-12	CAISN362-12	CAISN596-13	CAISN765-13	CAISN1513-14
CAISN1054-13	CAISN212-12	CAISN363-12	CAISN604-13	CAISN767-13	CAISN1515-14
CAISN1055-13	CAISN218-12	CAISN364-12	CAISN612-13	CAISN768-13	CAISN1516-14
CAISN1056-13	CAISN229-12	CAISN365-12	CAISN613-13	CAISN776-13	CAISN1519-14
CAISN1057-13	CAISN230-12	CAISN371-12	CAISN614-13	CAISN777-13	CAISN1520-14
CAISN1060-13	CAISN231-12	CAISN372-12	CAISN616-13	CAISN780-13	CAISN1145-13
CAISN1061-13	CAISN235-12	CAISN376-12	CAISN617-13	CAISN781-13	CAISN1147-13
CAISN1062-13	CAISN239-12	CAISN377-12	CAISN618-13	CAISN783-13	CAISN1148-13
CAISN1063-13	CAISN240-12	CAISN378-12	CAISN619-13	CAISN835-13	CAISN1149-13
CAISN1064-13	CAISN243-12	CAISN379-12	CAISN620-13	CAISN856-13	CAISN1151-13
CAISN1069-13	CAISN245-12	CAISN383-13	CAISN622-13	CAISN857-13	CAISN1152-13
CAISN1080-13	CAISN246-12	CAISN384-13	CAISN623-13	CAISN858-13	CAISN1161-13
CAISN1090-13	CAISN248-12	CAISN385-13	CAISN625-13	CAISN860-13	CAISN1163-13
CAISN1094-13	CAISN249-12	CAISN386-13	CAISN626-13	CAISN864-13	CAISN1164-13
CAISN1102-13	CAISN254-12	CAISN387-13	CAISN628-13	CAISN865-13	CAISN1169-13
CAISN1105-13	CAISN256-12	CAISN389-13	CAISN630-13	CAISN866-13	CAISN1170-13
CAISN1107-13	CAISN258-12	CAISN390-13	CAISN632-13	CAISN867-13	CAISN1176-13
CAISN1108-13	CAISN263-12	CAISN391-13	CAISN633-13	CAISN868-13	CAISN1178-13
CAISN1109-13	CAISN271-12	CAISN394-13	CAISN640-13	CAISN873-13	CAISN1179-13
CAISN1131-13	CAISN273-12	CAISN395-13	CAISN641-13	CAISN874-13	CAISN1180-13
CAISN1132-13	CAISN277-12	CAISN397-13	CAISN642-13	CAISN894-13	CAISN1184-13
CAISN1133-13	CAISN278-12	CAISN398-13	CAISN643-13	CAISN1427-14	CAISN1185-13
CAISN1137-13	CAISN282-12	CAISN399-13	CAISN648-13	CAISN1429-14	CAISN1186-13
CAISN1139-13	CAISN284-12	CAISN401-13	CAISN657-13	CAISN1431-14	CAISN1187-13
CAISN1140-13	CAISN286-12	CAISN402-13	CAISN658-13	CAISN1432-14	CAISN1188-13
CAISN1335-13	CAISN290-12	CAISN425-13	CAISN660-13	CAISN1433-14	CAISN1193-13
CAISN1337-13	CAISN291-12	CAISN428-13	CAISN672-13	CAISN1435-14	CAISN1194-13
CAISN1343-13	CAISN293-12	CAISN434-13	CAISN673-13	CAISN1436-14	CAISN1195-13
CAISN1347-13	CAISN294-12	CAISN435-13	CAISN678-13	CAISN1437-14	CAISN1196-13
CAISN1349-13	CAISN298-12	CAISN436-13	CAISN679-13	CAISN1438-14	CAISN1201-13
CAISN1352-13	CAISN299-12	CAISN437-13	CAISN680-13	CAISN1441-14	CAISN1202-13
CAISN1356-13	CAISN300-12	CAISN439-13	CAISN686-13	CAISN1443-14	CAISN1203-13
CAISN1357-13	CAISN301-12	CAISN440-13	CAISN687-13	CAISN1444-14	CAISN1204-13
CAISN1358-13	CAISN303-12	CAISN443-13	CAISN688-13	CAISN1445-14	CAISN1205-13
CAISN1359-13	CAISN304-12	CAISN445-13	CAISN692-13	CAISN1454-14	CAISN1206-13
CAISN1366-13	CAISN308-12	CAISN447-13	CAISN693-13	CAISN1455-14	CAISN1207-13
CAISN1368-13	CAISN314-12	CAISN471-13	CAISN696-13	CAISN1456-14	CAISN1208-13
CAISN1369-13	CAISN315-12	CAISN472-13	CAISN700-13	CAISN1457-14	CAISN1210-13
CAISN1374-13	CAISN316-12	CAISN473-13	CAISN704-13	CAISN1460-14	CAISN1211-13
CAISN1376-13	CAISN317-12	CAISN475-13	CAISN709-13	CAISN1467-14	CAISN1214-13
CAISN1377-13	CAISN318-12	CAISN543-13	CAISN710-13	CAISN1469-14	CAISN1215-13
CAISN1378-13	CAISN319-12	CAISN544-13	CAISN712-13	CAISN1471-14	CAISN1218-13

CAISN1379-13	CAISN320-12	CAISN545-13	CAISN714-13	CAISN1472-14	CAISN1221-13
CAISN1382-13	CAISN321-12	CAISN546-13	CAISN716-13	CAISN1473-14	CAISN1222-13
CAISN1384-13	CAISN322-12	CAISN555-13	CAISN717-13	CAISN1479-14	CAISN1223-13
CAISN1393-13	CAISN325-12	CAISN556-13	CAISN720-13	CAISN1481-14	CAISN1225-13
CAISN1394-13	CAISN326-12	CAISN566-13	CAISN725-13	CAISN1482-14	CAISN1226-13
CAISN1413-13	CAISN327-12	CAISN571-13	CAISN726-13	CAISN1483-14	CAISN1227-13
CAISN1417-13	CAISN331-12	CAISN572-13	CAISN727-13	CAISN1486-14	CAISN1228-13
CAISN152-12	CAISN332-12	CAISN573-13	CAISN728-13	CAISN1487-14	CAISN1229-13
CAISN159-12	CAISN339-12	CAISN578-13	CAISN729-13	CAISN1488-14	CAISN1230-13
CAISN177-12	CAISN340-12	CAISN579-13	CAISN732-13	CAISN1489-14	CAISN1231-13
CAISN189-12	CAISN352-12	CAISN586-13	CAISN735-13	CAISN1490-14	CAISN1232-13
CAISN190-12	CAISN354-12	CAISN587-13	CAISN746-13	CAISN1491-14	CAISN1233-13
CAISN980-13	CAISN355-12	CAISN588-13	CAISN749-13	CAISN1508-14	
CAISN1040-13	CAISN356-12	CAISN589-13	CAISN751-13	CAISN1509-14	
CAISN197-12	CAISN359-12	CAISN593-13	CAISN763-13	CAISN1510-14	
CAISN205-12	CAISN360-12	CAISN595-13	CAISN764-13	CAISN1511-14	

18S* and *COI data sets

CAISN1051-13	CAISN208-12	CAISN362-12	CAISN596-13	CAISN765-13	CAISN1513-14
CAISN1054-13	CAISN212-12	CAISN363-12	CAISN604-13	CAISN767-13	CAISN1515-14
CAISN1055-13	CAISN218-12	CAISN364-12	CAISN612-13	CAISN768-13	CAISN1516-14
CAISN1056-13	CAISN229-12	CAISN365-12	CAISN613-13	CAISN776-13	CAISN1519-14
CAISN1057-13	CAISN230-12	CAISN371-12	CAISN614-13	CAISN777-13	CAISN1520-14
CAISN1060-13	CAISN231-12	CAISN372-12	CAISN616-13	CAISN780-13	CAISN1145-13
CAISN1061-13	CAISN235-12	CAISN376-12	CAISN617-13	CAISN781-13	CAISN1147-13
CAISN1062-13	CAISN239-12	CAISN377-12	CAISN618-13	CAISN783-13	CAISN1148-13
CAISN1063-13	CAISN240-12	CAISN378-12	CAISN619-13	CAISN835-13	CAISN1149-13
CAISN1064-13	CAISN243-12	CAISN379-12	CAISN620-13	CAISN856-13	CAISN1151-13
CAISN1069-13	CAISN245-12	CAISN383-13	CAISN622-13	CAISN857-13	CAISN1152-13
CAISN1080-13	CAISN246-12	CAISN384-13	CAISN623-13	CAISN858-13	CAISN1161-13
CAISN1090-13	CAISN248-12	CAISN385-13	CAISN625-13	CAISN860-13	CAISN1163-13
CAISN1094-13	CAISN249-12	CAISN386-13	CAISN626-13	CAISN864-13	CAISN1164-13
CAISN1102-13	CAISN254-12	CAISN387-13	CAISN628-13	CAISN865-13	CAISN1169-13
CAISN1105-13	CAISN256-12	CAISN389-13	CAISN630-13	CAISN866-13	CAISN1170-13
CAISN1107-13	CAISN258-12	CAISN390-13	CAISN632-13	CAISN867-13	CAISN1176-13
CAISN1108-13	CAISN263-12	CAISN391-13	CAISN633-13	CAISN868-13	CAISN1178-13
CAISN1109-13	CAISN271-12	CAISN394-13	CAISN640-13	CAISN873-13	CAISN1179-13
CAISN1131-13	CAISN273-12	CAISN395-13	CAISN641-13	CAISN874-13	CAISN1180-13
CAISN1132-13	CAISN277-12	CAISN397-13	CAISN642-13	CAISN894-13	CAISN1184-13
CAISN1133-13	CAISN278-12	CAISN398-13	CAISN643-13	CAISN1427-14	CAISN1185-13
CAISN1137-13	CAISN282-12	CAISN399-13	CAISN648-13	CAISN1429-14	CAISN1186-13
CAISN1139-13	CAISN284-12	CAISN401-13	CAISN657-13	CAISN1431-14	CAISN1187-13
CAISN1140-13	CAISN286-12	CAISN402-13	CAISN658-13	CAISN1432-14	CAISN1188-13
CAISN1335-13	CAISN290-12	CAISN425-13	CAISN660-13	CAISN1433-14	CAISN1193-13
CAISN1337-13	CAISN291-12	CAISN428-13	CAISN672-13	CAISN1435-14	CAISN1194-13
CAISN1343-13	CAISN293-12	CAISN434-13	CAISN673-13	CAISN1436-14	CAISN1195-13
CAISN1347-13	CAISN294-12	CAISN435-13	CAISN678-13	CAISN1437-14	CAISN1196-13
CAISN1349-13	CAISN298-12	CAISN436-13	CAISN679-13	CAISN1438-14	CAISN1201-13
CAISN1352-13	CAISN299-12	CAISN437-13	CAISN680-13	CAISN1441-14	CAISN1202-13

CAISN1356-13	CAISN300-12	CAISN439-13	CAISN686-13	CAISN1443-14	CAISN1203-13
CAISN1357-13	CAISN301-12	CAISN440-13	CAISN687-13	CAISN1444-14	CAISN1204-13
CAISN1358-13	CAISN303-12	CAISN443-13	CAISN688-13	CAISN1445-14	CAISN1205-13
CAISN1359-13	CAISN304-12	CAISN445-13	CAISN692-13	CAISN1454-14	CAISN1206-13
CAISN1366-13	CAISN308-12	CAISN447-13	CAISN693-13	CAISN1455-14	CAISN1207-13
CAISN1368-13	CAISN314-12	CAISN471-13	CAISN696-13	CAISN1456-14	CAISN1208-13
CAISN1369-13	CAISN315-12	CAISN472-13	CAISN700-13	CAISN1457-14	CAISN1210-13
CAISN1374-13	CAISN316-12	CAISN473-13	CAISN704-13	CAISN1460-14	CAISN1211-13
CAISN1376-13	CAISN317-12	CAISN475-13	CAISN709-13	CAISN1467-14	CAISN1214-13
CAISN1377-13	CAISN318-12	CAISN543-13	CAISN710-13	CAISN1469-14	CAISN1215-13
CAISN1378-13	CAISN319-12	CAISN544-13	CAISN712-13	CAISN1471-14	CAISN1218-13
CAISN1379-13	CAISN320-12	CAISN545-13	CAISN714-13	CAISN1472-14	CAISN1221-13
CAISN1382-13	CAISN321-12	CAISN546-13	CAISN716-13	CAISN1473-14	CAISN1222-13
CAISN1384-13	CAISN322-12	CAISN555-13	CAISN717-13	CAISN1479-14	CAISN1223-13
CAISN1393-13	CAISN325-12	CAISN556-13	CAISN720-13	CAISN1481-14	CAISN1225-13
CAISN1394-13	CAISN326-12	CAISN566-13	CAISN725-13	CAISN1482-14	CAISN1226-13
CAISN1413-13	CAISN327-12	CAISN571-13	CAISN726-13	CAISN1483-14	CAISN1227-13
CAISN1417-13	CAISN331-12	CAISN572-13	CAISN727-13	CAISN1486-14	CAISN1228-13
CAISN152-12	CAISN332-12	CAISN573-13	CAISN728-13	CAISN1487-14	CAISN1229-13
CAISN159-12	CAISN339-12	CAISN578-13	CAISN729-13	CAISN1488-14	CAISN1230-13
CAISN177-12	CAISN340-12	CAISN579-13	CAISN732-13	CAISN1489-14	CAISN1231-13
CAISN189-12	CAISN352-12	CAISN586-13	CAISN735-13	CAISN1490-14	CAISN1232-13
CAISN190-12	CAISN354-12	CAISN587-13	CAISN746-13	CAISN1491-14	CAISN1233-13
CAISN980-13	CAISN355-12	CAISN588-13	CAISN749-13	CAISN1508-14	
CAISN1040-13	CAISN356-12	CAISN589-13	CAISN751-13	CAISN1509-14	
CAISN197-12	CAISN359-12	CAISN593-13	CAISN763-13	CAISN1510-14	
CAISN205-12	CAISN360-12	CAISN595-13	CAISN764-13	CAISN1511-14	

Appendix V: Total table of species used in thesis

Table of the species used in this thesis. The chapter(s) in which each species was analyzed is also indicated.

Class	Order	Family	Species	Chapter 2		Chapter 3		
				COI Ref	COI Novel	18S Ref	18S & COI Novel	Chap 4
Copepoda	Calanoida	Acartiidae	*Acartia bifilosa*			✓		
Copepoda	Calanoida	Acartiidae	*Acartia californiensis*	✓				✓
Copepoda	Calanoida	Acartiidae	*Acartia clausi*	✓	✓	✓	✓	✓
Copepoda	Calanoida	Acartiidae	*Acartia discaudata*					
Copepoda	Calanoida	Acartiidae	*Acartia hudsonica*		✓		✓	✓
Copepoda	Calanoida	Acartiidae	*Acartia levequei*	✓				
Copepoda	Calanoida	Acartiidae	*Acartia longiremis*		✓	✓	✓	✓
Copepoda	Calanoida	Acartiidae	*Acartia pacifica*	✓		✓		
Copepoda	Calanoida	Acartiidae	*Acartia tonsa*	✓		✓		
Copepoda	Calanoida	Acartiidae	*Paracartia grani*	✓				
Copepoda	Calanoida	Acartiidae	*Paralabidocera grandispina*					
Copepoda	Calanoida	Aetideidae	*Aetideus armatus*	✓		✓		✓
Copepoda	Calanoida	Aetideidae	*Aetideus bradyi*	✓				✓
Copepoda	Calanoida	Aetideidae	*Aetideus divergens*		✓		✓	✓
Copepoda	Calanoida	Aetideidae	*Euchirella amoena*			✓		
Copepoda	Calanoida	Aetideidae	*Gaetanus variabilis*			✓		
Copepoda	Calanoida	Aetideidae	*Undeuchaeta major*			✓		
Copepoda	Calanoida	Arietellidae	*Paraugaptilus buchani*			✓		
Copepoda	Calanoida	Augaptilidae	*Euaugaptilus laticeps*					
Copepoda	Calanoida	Augaptilidae	*Haloptilus longicornis*			✓		
Copepoda	Calanoida	Augaptilidae	*Haloptilus ocellatus*	✓				
Copepoda	Calanoida	Bathypontiidae	*Temorites brevis*	✓		✓		
Copepoda	Calanoida	Calanidae	*Calanoides acutus*					
Copepoda	Calanoida	Calanidae	*Calanoides carinatus*	✓		✓		
Copepoda	Calanoida	Calanidae	*Calanus finmarchicus*	✓		✓	✓	
Copepoda	Calanoida	Calanidae	*Calanus glacialis*	✓	✓		✓	✓
Copepoda	Calanoida	Calanidae	*Calanus helgolandicus*	✓		✓		
Copepoda	Calanoida	Calanidae	*Calanus hyperboreus*	✓	✓		✓	✓
Copepoda	Calanoida	Calanidae	*Calanus marshallae*				✓	
Copepoda	Calanoida	Calanidae	*Calanus pacificus*	✓	✓	✓	✓	✓
Copepoda	Calanoida	Calanidae	*Calanus propinquus*					
Copepoda	Calanoida	Calanidae	*Calanus simillimus*					
Copepoda	Calanoida	Calanidae	*Calanus sinicus*			✓		
Copepoda	Calanoida	Calanidae	*Canthocalanus pauper*			✓		
Copepoda	Calanoida	Calanidae	*Cosmocalanus darwinii*	✓		✓		
Copepoda	Calanoida	Calanidae	*Mesocalanus tenuicornis*		✓		✓	✓

Copepoda	Calanoida	Calanidae	*Neocalanus cristatus*			✓		
Copepoda	Calanoida	Calanidae	*Neocalanus flemingeri*			✓		
Copepoda	Calanoida	Calanidae	*Neocalanus gracilis*			✓		
Copepoda	Calanoida	Calanidae	*Neocalanus plumchrus*			✓		
Copepoda	Calanoida	Calanidae	*Neocalanus robustior*			✓		
Copepoda	Calanoida	Calanidae	*Neocalanus tonsus*					
Copepoda	Calanoida	Calanidae	*Undinula vulgaris*			✓		
Copepoda	Calanoida	Candaciidae	*Candacia bispinosa*	✓		✓		
Copepoda	Calanoida	Candaciidae	*Candacia bradyi*					
Copepoda	Calanoida	Candaciidae	*Candacia columbiae*			✓		
Copepoda	Calanoida	Candaciidae	*Candacia discaudata*	✓		✓		
Copepoda	Calanoida	Candaciidae	*Candacia pachydactyla*	✓		✓		
Copepoda	Calanoida	Candaciidae	*Candacia simplex*	✓		✓		
Copepoda	Calanoida	Candaciidae	*Candacia truncata*			✓		
Copepoda	Calanoida	Centropagidae	*Boeckella gracilipes*					
Copepoda	Calanoida	Centropagidae	*Boeckella gracilis*	✓				
Copepoda	Calanoida	Centropagidae	*Boeckella titicacae*					
Copepoda	Calanoida	Centropagidae	*Boeckella triarticulata*					
Copepoda	Calanoida	Centropagidae	*Calamoecia ampulla*	✓				
Copepoda	Calanoida	Centropagidae	*Centropages abdominalis*	✓	✓	✓	✓	✓
Copepoda	Calanoida	Centropagidae	*Centropages furcatus*			✓		
Copepoda	Calanoida	Centropagidae	*Centropages hamatus*		✓	✓	✓	✓
Copepoda	Calanoida	Centropagidae	*Centropages tenuiremis*			✓		
Copepoda	Calanoida	Centropagidae	*Centropages typicus*		✓	✓	✓	✓
Copepoda	Calanoida	Centropagidae	*Centropages violaceus*			✓		
Copepoda	Calanoida	Centropagidae	*Isias clavipes*			✓		
Copepoda	Calanoida	Centropagidae	*Limnocalanus johanseni*					
Copepoda	Calanoida	Centropagidae	*Limnocalanus macrurus*			✓		
Copepoda	Calanoida	Centropagidae	*Osphranticum labronectum*					
Copepoda	Calanoida	Centropagidae	*Parabroteas sarsi*					
Copepoda	Calanoida	Centropagidae	*Sinocalanus sinensis*			✓		
Copepoda	Calanoida	Centropagidae	*Sinocalanus tenellus*			✓		
Copepoda	Calanoida	Clausocalanidae	*Clausocalanus arcuicornis*			✓		
Copepoda	Calanoida	Clausocalanidae	*Clausocalanus furcatus*			✓		
Copepoda	Calanoida	Clausocalanidae	*Clausocalanus laticeps*	✓				
Copepoda	Calanoida	Clausocalanidae	*Clausocalanus lividus*					
Copepoda	Calanoida	Clausocalanidae	*Clausocalanus pergens*	✓				
Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus acuspes*					✓
Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus elongatus*	✓		✓		✓
Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus mimus*	✓				
Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus minutus*	✓	✓		✓	✓
Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus moultoni*	✓		✓		✓

Copepoda	Calanoida	Clausocalanidae	*Pseudocalanus newmani*	✓			✓
Copepoda	Calanoida	Diaixidae	*Diaixis hibernica*			✓	
Copepoda	Calanoida	Diaptomidae	*Aglaodiaptomus leptopus*			✓	
Copepoda	Calanoida	Diaptomidae	*Aglaodiaptomus spatulocrenatus*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus dorsalis*			✓	
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus dorsalis1*	✓			
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus dorsalis2*				
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus dorsalis3*	✓			
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus dorsalis4*				
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus salinus*			✓	
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus stephanidesi*			✓	
Copepoda	Calanoida	Diaptomidae	*Arctodiaptomus wierzejskii*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Diaptomus castor*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Diaptomus cyaneus*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Diaptomus kenitraensis*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Diaptomus mirus*			✓	
Copepoda	Calanoida	Diaptomidae	*Eudiaptomus gracilis*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Eudiaptomus graciloides*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Eudiaptomus vulgaris*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus amblyodon*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus gurneyi*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus gurneyi canaanita*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus hungaricus*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus ignatovi*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus ingens*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus maroccanus*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus roubaui*			✓	
Copepoda	Calanoida	Diaptomidae	*Hemidiaptomus superbus*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Hesperodiaptomus shoshone*			✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus ashlandi*			✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus coloradensis*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus cuauhtemoci*				
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus garciai*				
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus minutus*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus moorei*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus novamexicanus*	✓			
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus sicilis*	✓		✓	
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus siciloides*	✓	✓		
Copepoda	Calanoida	Diaptomidae	*Leptodiaptomus tyrrelli*				
Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus albuquerquensis*				
Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus montezumae*				
Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus nesus*	✓		✓	

Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus patzcuarensis*			
Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus reidae*			
Copepoda	Calanoida	Diaptomidae	*Mastigodiaptomus texensis*			
Copepoda	Calanoida	Diaptomidae	*Metadiaptomus chevreuxi*			
Copepoda	Calanoida	Diaptomidae	*Neodiaptomus meggitti*			
Copepoda	Calanoida	Diaptomidae	*Neodiaptomus schmackeri*	✓		
Copepoda	Calanoida	Diaptomidae	*Prionodiaptomus colombiensis*			
Copepoda	Calanoida	Diaptomidae	*Skistodiaptomus mississippiensis*	✓		✓
Copepoda	Calanoida	Diaptomidae	*Skistodiaptomus oregonensis*			✓
Copepoda	Calanoida	Diaptomidae	*Skistodiaptomus pallidus*		✓	✓
Copepoda	Calanoida	Diaptomidae	*Skistodiaptomus pygmaeus*	✓		✓
Copepoda	Calanoida	Diaptomidae	*Copidodiaptomus numidicus*			✓
Copepoda	Calanoida	Diaptomidae	*Onychodiaptomus sanguineus*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Eucalanus spinifer*			✓
Copepoda	Calanoida	Eucalanidae	*Pareucalanus attenuatus*			✓
Copepoda	Calanoida	Eucalanidae	*Pareucalanus langae*			✓
Copepoda	Calanoida	Eucalanidae	*Rhincalanus cornutus*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Rhincalanus gigas*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Rhincalanus nasutus*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Rhincalanus rostrifrons*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus crassus*			✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus longiceps*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus mucronatus*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus pileatus*			✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus subcrassus*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Subeucalanus subtenuis*	✓		✓
Copepoda	Calanoida	Eucalanidae	*Eucalanus bungii*			
Copepoda	Calanoida	Eucalanidae	*Eucalanus elongatus*			✓
Copepoda	Calanoida	Eucalanidae	*Eucalanus inermis*	✓		✓
Copepoda	Calanoida	Euchaetidae	*Euchaeta acuta*	✓		✓
Copepoda	Calanoida	Euchaetidae	*Euchaeta concinna*			✓
Copepoda	Calanoida	Euchaetidae	*Euchaeta indica*			✓
Copepoda	Calanoida	Euchaetidae	*Euchaeta media*			✓
Copepoda	Calanoida	Euchaetidae	*Euchaeta rimana*	✓		✓
Copepoda	Calanoida	Euchaetidae	*Paraeuchaeta antarctica*	✓		
Copepoda	Calanoida	Fosshageniidae	*Temoropia mayumbaensis*			✓
Copepoda	Calanoida	Heterorhabdidae	*Heterorhabdus tanneri*			✓
Copepoda	Calanoida	Heterorhabdidae	*Heterostylites major*			✓
Copepoda	Calanoida	Heterorhabdidae	*Paraheterorhabdus compactus*			✓
Copepoda	Calanoida	Hyperbionycidae	*Hyperbionyx athesphatos*			✓
Copepoda	Calanoida	Lucicutiidae	*Lucicutia flavicornis*			✓
Copepoda	Calanoida	Lucicutiidae	*Lucicutia ovaliformis*			✓

Copepoda	Calanoida	Metridinidae	*Gaussia princeps*	✓		✓		
Copepoda	Calanoida	Metridinidae	*Metridia asymmetrica*			✓		
Copepoda	Calanoida	Metridinidae	*Metridia curticauda*	✓		✓		
Copepoda	Calanoida	Metridinidae	*Metridia effusa*			✓		✓
Copepoda	Calanoida	Metridinidae	*Metridia gerlachei*	✓		✓		✓
Copepoda	Calanoida	Metridinidae	*Metridia longa*			✓		
Copepoda	Calanoida	Metridinidae	*Metridia lucens*	✓		✓		✓
Copepoda	Calanoida	Metridinidae	*Metridia okhotensis*	✓		✓		
Copepoda	Calanoida	Metridinidae	*Metridia pacifica*	✓	✓	✓	✓	✓
Copepoda	Calanoida	Metridinidae	*Pleuromamma abdominalis*			✓		
Copepoda	Calanoida	Metridinidae	*Pleuromamma antarctica*			✓		
Copepoda	Calanoida	Metridinidae	*Pleuromamma borealis*	✓		✓		
Copepoda	Calanoida	Metridinidae	*Pleuromamma piseki*					
Copepoda	Calanoida	Metridinidae	*Pleuromamma scutullata*			✓		
Copepoda	Calanoida	Metridinidae	*Pleuromamma xiphias*			✓		
Copepoda	Calanoida	Nullosetigeridae	*Nullosetigera auctiseta*			✓		
Copepoda	Calanoida	Paracalanidae	*Acrocalanus andersoni*	✓		✓		
Copepoda	Calanoida	Paracalanidae	*Acrocalanus gibber*	✓		✓		
Copepoda	Calanoida	Paracalanidae	*Acrocalanus gracilis*	✓		✓		
Copepoda	Calanoida	Paracalanidae	*Acrocalanus longicornis*			✓		
Copepoda	Calanoida	Paracalanidae	*Acrocalanus monachus*			✓		
Copepoda	Calanoida	Paracalanidae	*Bestiolina similis*			✓		
Copepoda	Calanoida	Paracalanidae	*Calocalanus minutus*	✓				
Copepoda	Calanoida	Paracalanidae	*Calocalanus pavo*			✓		
Copepoda	Calanoida	Paracalanidae	*Calocalanus plumulosus*			✓		
Copepoda	Calanoida	Paracalanidae	*Calocalanus styliremis*			✓		
Copepoda	Calanoida	Paracalanidae	*Calocalanus tenuis*					
Copepoda	Calanoida	Paracalanidae	*Delibus nudus*			✓		
Copepoda	Calanoida	Paracalanidae	*Mecynocera clausi*	✓		✓		
Copepoda	Calanoida	Paracalanidae	*Paracalanus aculeatus*	✓		✓		✓
Copepoda	Calanoida	Paracalanidae	*Paracalanus denudatus*			✓		✓
Copepoda	Calanoida	Paracalanidae	*Paracalanus indicus*	✓		✓		✓
Copepoda	Calanoida	Paracalanidae	*Paracalanus parvus*		✓	✓	✓	✓
Copepoda	Calanoida	Paracalanidae	*Paracalanus quasimodo*			✓		✓
Copepoda	Calanoida	Paracalanidae	*Paracalanus tropicus*			✓		
Copepoda	Calanoida	Paracalanidae	*Parvocalanus crassirostris*			✓		
Copepoda	Calanoida	Phaennidae	*Phaenna spinifera*	✓		✓		
Copepoda	Calanoida	Pontellidae	*Anomalocera patersoni*	✓		✓		
Copepoda	Calanoida	Pontellidae	*Calanopia thompsoni*					
Copepoda	Calanoida	Pontellidae	*Epilabidocera amphitrites*		✓			
Copepoda	Calanoida	Pontellidae	*Labidocera acuta*			✓		
Copepoda	Calanoida	Pontellidae	*Labidocera euchaeta*	✓		✓		

Copepoda	Calanoida	Pontellidae	*Labidocera japonica*					
Copepoda	Calanoida	Pontellidae	*Labidocera madurae*					
Copepoda	Calanoida	Pontellidae	*Labidocera rotunda*					
Copepoda	Calanoida	Pontellidae	*Pontella chierchiae*					
Copepoda	Calanoida	Pontellidae	*Pontella fera*			✓		
Copepoda	Calanoida	Pontellidae	*Pontellina plumata*			✓		
Copepoda	Calanoida	Pontellidae	*Pontellopsis yamadae*					
Copepoda	Calanoida	Pseudocyclopidae	*Exumella mediterranea*			✓		
Copepoda	Calanoida	Pseudocyclopidae	*Pseudocyclops juanibali*			✓		
Copepoda	Calanoida	Pseudocyclopidae	*Pseudocyclops schminkei*			✓		
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus annandalei*			✓		
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus aurivillii*			✓		
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus euryhalinus*			✓		
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus inopinus*			✓		
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus ishigakiensis*					
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus koreanus*					
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus nansei*					
Copepoda	Calanoida	Pseudodiaptomidae	*Pseudodiaptomus poplesia*			✓		
Copepoda	Calanoida	Scolecitrichidae	*Scaphocalanus magnus*			✓		
Copepoda	Calanoida	Scolecitrichidae	*Scolecithricella longispinosa*			✓		
Copepoda	Calanoida	Scolecitrichidae	*Scolecithrix bradyi*			✓		
Copepoda	Calanoida	Scolecitrichidae	*Scolecithrix danae*			✓		
Copepoda	Calanoida	Spinocalanidae	*Foxtonia barbatula*			✓		
Copepoda	Calanoida	Spinocalanidae	*Spinocalanus abyssalis*			✓		
Copepoda	Calanoida	Stephidae	*Stephos longipes*					
Copepoda	Calanoida	Sulcanidae	*Sulcanus conflictus*			✓		
Copepoda	Calanoida	Temoridae	*Eurytemora affinis*	✓		✓		✓
Copepoda	Calanoida	Temoridae	*Eurytemora carolleeae*	✓				✓
Copepoda	Calanoida	Temoridae	*Eurytemora herdmani*		✓		✓	✓
Copepoda	Calanoida	Temoridae	*Eurytemora lacustris*					✓
Copepoda	Calanoida	Temoridae	*Temora discaudata*			✓		✓
Copepoda	Calanoida	Temoridae	*Temora longicornis*		✓	✓	✓	✓
Copepoda	Calanoida	Temoridae	*Temora stylifera*	✓		✓		
Copepoda	Calanoida	Temoridae	*Temora turbinata*			✓		
Copepoda	Calanoida	Tharybidae	*Tharybis groenlandica*			✓		
Copepoda	Calanoida	Tortanidae	*Tortanus derjugini*					✓
Copepoda	Calanoida	Tortanidae	*Tortanus dextrilobatus*					✓
Copepoda	Calanoida	Tortanidae	*Tortanus discaudatus*		✓		✓	✓
Copepoda	Calanoida	Tortanidae	*Tortanus gracilis*			✓		✓
Copepoda	Calanoida	Tortanidae	*Tortanus komachi*					✓
Copepoda	Calanoida	Tortanidae	*Tortanus vermiculus*					✓
Copepoda	Calanoida	Candaciidae	*Candacia bipinnata*	✓		✓		

Copepoda	Calanoida	Diaptomidae	*Aglaodiaptomus clavipoides*	✓	✓
Copepoda	Calanoida	Megacalanidae	*Bathycalanus princeps*		✓
Copepoda	Calanoida	Paracalanidae	*Calocalanus curtus*		✓
Copepoda	Calanoida	Pseudodiaptomidae	*Calanipeda aquaedulcis*	✓	✓
Copepoda	Cladocera	Daphniidae	*Daphnia pulex*		✓
Copepoda	Cyclopoida	Incertae sedis	*Pachos punctatum*		✓
Copepoda	Cyclopoida	Cyclopettidae	*Paracyclopina nana*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops americanus*		
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops bicuspidatus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops galbinus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops robustus*		
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops vernalis*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Acanthocyclops viridis*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Cyclops abyssorum*		
Copepoda	Cyclopoida	Cyclopidae	*Cyclops insignis*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Cyclops kolensis*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Cyclops strenuus*		
Copepoda	Cyclopoida	Cyclopidae	*Diacyclops bicuspidatus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Diacyclops galbinus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Diacyclops improcerus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Diacyclops incolotaenia*		✓
Copepoda	Cyclopoida	Cyclopidae	*Diacyclops jasnitskii*		✓
Copepoda	Cyclopoida	Cyclopidae	*Ectocyclops polyspinosus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops arcanus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops dumonti*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops macruroides*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops macrurus*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops serrulatus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops serrulatus baicalocorrepus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops serrulatus serrulatus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Eucyclops speratus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Macrocyclops albidus*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Macrocyclops distinctus*		
Copepoda	Cyclopoida	Cyclopidae	*Megacyclops viridis*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Mesocyclops edax*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Mesocyclops leuckarti*	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Mesocyclops pehpeiensis*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Paracyclops fimbriatus*	✓	
Copepoda	Cyclopoida	Cyclopidae	*Thermocyclops crassus*		✓
Copepoda	Cyclopoida	Cyclopidae	*Thermocyclops inversus*		
Copepoda	Cyclopoida	Cyclopidae	*Thermocyclops oithonoides*		
Copepoda	Cyclopoida	Cyclopidae	*Tropocyclops aztequei*		

Copepoda	Cyclopoida	Cyclopidae	*Tropocyclops prasinus*				
Copepoda	Cyclopoida	Cyclopinidae	*Cyclopina gracilis*			✓	
Copepoda	Cyclopoida	Lernaeidae	*Lamproglena chinensis*	✓		✓	
Copepoda	Cyclopoida	Lernaeidae	*Lamproglena orientalis*			✓	
Copepoda	Cyclopoida	Lernaeidae	*Lernaea cyprinacea*			✓	
Copepoda	Cyclopoida	Notodelphyidae	*Notodelphys prasina*	✓		✓	
Copepoda	Cyclopoida	Oithonidae	*Oithona atlantica*		✓		✓
Copepoda	Cyclopoida	Oithonidae	*Oithona brevicornis*	✓		✓	
Copepoda	Cyclopoida	Oithonidae	*Oithona davisae*			✓	
Copepoda	Cyclopoida	Oithonidae	*Oithona dissimilis*				
Copepoda	Cyclopoida	Oithonidae	*Oithona oculata*				
Copepoda	Cyclopoida	Oithonidae	*Oithona similis*		✓	✓	✓
Copepoda	Cyclopoida	Cyclopidae	*Apocyclops royi*			✓	
Copepoda	Harpacticoida	Ameiridae	*Nitokra hibernica*			✓	
Copepoda	Harpacticoida	Ameiridae	*Nitokra spinipes*	✓		✓	
Copepoda	Harpacticoida	Argestidae	*Eurycletodes laticauda*	✓		✓	
Copepoda	Harpacticoida	Cancrincolidae	*Cancrincola plumipes*			✓	
Copepoda	Harpacticoida	Canthocamptidae	*Attheyella crassa*			✓	
Copepoda	Harpacticoida	Canthocamptidae	*Itunella muelleri*			✓	
Copepoda	Harpacticoida	Canthocamptidae	*Mesochra rapiens*			✓	
Copepoda	Harpacticoida	Canuellidae	*Canuella perplexa*			✓	
Copepoda	Harpacticoida	Cletodidae	*Cletocamptus deitersi*				
Copepoda	Harpacticoida	Dactylopusiidae	*Sewellia tropica*	✓		✓	
Copepoda	Harpacticoida	Ectinosomatidae	*Microsetella norvegica*		✓		✓
Copepoda	Harpacticoida	Harpacticidae	*Tigriopus californicus*			✓	
Copepoda	Harpacticoida	Harpacticidae	*Tigriopus fulvus*			✓	
Copepoda	Harpacticoida	Harpacticidae	*Tigriopus japonicus*			✓	
Copepoda	Harpacticoida	Harpacticidae	*Zaus abbreviatus*		✓		✓
Copepoda	Harpacticoida	Harpacticidae	*Zaus caeruleus*			✓	
Copepoda	Harpacticoida	Harpacticidae	*Tigriopus brevicornis*			✓	
Copepoda	Harpacticoida	Laophontidae	*Onychocamptus bengalensis*			✓	
Copepoda	Harpacticoida	Miraciidae	*Miracia efferata*			✓	
Copepoda	Harpacticoida	Miraciidae	*Paramphiascella fulvofasciata*			✓	
Copepoda	Harpacticoida	Miraciidae	*Schizopera akation*	✓			
Copepoda	Harpacticoida	Miraciidae	*Schizopera knabeni*	✓			
Copepoda	Harpacticoida	Miraciidae	*Schizopera uranusi*	✓			
Copepoda	Harpacticoida	Miraciidae	*Stenhelia pubescens*	✓			
Copepoda	Harpacticoida	Miraciidae	*Typhlamphiascus typhlops*	✓		✓	
Copepoda	Harpacticoida	Paramesochridae	*Remanea naksanensis*				
Copepoda	Harpacticoida	Phyllognathopodidae	*Phyllognathopus viguieri*				
Copepoda	Harpacticoida	Tachidiidae	*Euterpina acutifrons*	✓		✓	
Copepoda	Harpacticoida	Tachidiidae	*Tachidius triangularis*	✓		✓	

Copepoda	Harpacticoida	Tachidiidae	*Tanais tinhauae*			✓	
Copepoda	Harpacticoida	Tisbidae	*Tisbe furcata*	✓	✓		✓
Copepoda	Harpacticoida	Tisbidae	*Tisbe tenera*	✓		✓	
Copepoda	Harpacticoida	Ameiridae	*Ameira scotti*			✓	
Copepoda	Harpacticoida	Canthocamptidae	*Bryocamptus pygmaeus*	✓		✓	
Copepoda	Harpacticoida	Miraciidae	*Amphiascoides atopus*			✓	
Copepoda	Misophrioida	Misophriidae	*Misophriopsis okinawensis*			✓	
Copepoda	Monstrilloida	Monstrillidae	*Monstrilla scotti*		✓		✓
Copepoda	Monstrilloida	Monstrillidae	*Monstrilla clavata*			✓	
Copepoda	Poecilostomatoida	Catiniidae	*Catinia plana*			✓	
Copepoda	Poecilostomatoida	Chondracanthidae	*Lernentoma asellina*			✓	
Copepoda	Poecilostomatoida	Clausidiidae	*Clausidium vancouverense*			✓	
Copepoda	Poecilostomatoida	Clausidiidae	*Hemicyclops thalassius*	✓		✓	
Copepoda	Poecilostomatoida	Corycaeidae	*Corycaeus affinis*				
Copepoda	Poecilostomatoida	Corycaeidae	*Corycaeus anglicus*		✓		
Copepoda	Poecilostomatoida	Corycaeidae	*Corycaeus speciosus*			✓	
Copepoda	Poecilostomatoida	Corycaeidae	*Farranula gibbula*				
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus anchoratus*			✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus briani*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus hypomesi*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus peregrinus*			✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus scalaris*			✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus tumidus*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Ergasilus yaluzangbus*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Paraergasilus brevidigitus*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Paraergasilus medius*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Sinergasilus polycolpus*	✓		✓	
Copepoda	Poecilostomatoida	Ergasilidae	*Sinergasilus undulatus*			✓	
Copepoda	Poecilostomatoida	Lichomolgidae	*Astericola clausii*			✓	
Copepoda	Poecilostomatoida	Lichomolgidae	*Lichomolgus canui*	✓		✓	
Copepoda	Poecilostomatoida	Lichomolgidae	*Lichomolgus marginatus*	✓		✓	
Copepoda	Poecilostomatoida	Myicolidae	*Pseudomyicola spinosus*				
Copepoda	Poecilostomatoida	Mytilicolidae	*Mytilicola intestinalis*			✓	
Copepoda	Poecilostomatoida	Mytilicolidae	*Mytilicola orientalis*			✓	
Copepoda	Poecilostomatoida	Mytilicolidae	*Pectenophilus ornatus*			✓	
Copepoda	Poecilostomatoida	Oncaeidae	*Oncaea shmelevi*	✓			
Copepoda	Poecilostomatoida	Oncaeidae	*Oncaea waldemari*	✓			
Copepoda	Poecilostomatoida	Oncaeidae	*Triconia minuta*				
Copepoda	Poecilostomatoida	Oncaeidae	*Triconia umerus*				
Copepoda	Poecilostomatoida	Pionodesmotidae	*Pionodesmotes domhainfharraigeanus*				
Copepoda	Poecilostomatoida	Rhynchomolgidae	*Doridicola agilis*			✓	
Copepoda	Poecilostomatoida	Sapphirinidae	*Copilia mirabilis*			✓	

Copepoda	Poecilostomatoida	Sapphirinidae	*Sapphirina metallina*		
Copepoda	Poecilostomatoida	Sapphirinidae	*Sapphirina opalina*		
Copepoda	Poecilostomatoida	Sapphirinidae	*Sapphirina scarlata*		✓
Copepoda	Poecilostomatoida	Taeniacanthidae	*Clavisodalis abbreviatus*		✓
Copepoda	Poecilostomatoida	Taeniacanthidae	*Irodes sauridi*		✓
Copepoda	Poecilostomatoida	Ergasilidae	*Pseudergasilus parasiluri*	✓	✓
Copepoda	Poecilostomatoida	Ergasilidae	*Sinergasilus major*	✓	✓
Copepoda	Poecilostomatoida	Mytilicolidae	*Trochicola entericus*	✓	✓
Copepoda	Poecilostomatoida	Sabelliphilidae	*Sabelliphilus elongatus*	✓	✓
Copepoda	Poecilostomatoida	Sapphirinidae	*Sapphirina darwinii*	✓	✓
Copepoda	Poecilostomatoida	Taeniacanthidae	*Taeniacanthus kitamakura*	✓	✓
Copepoda	Poecilostomatoida	Taeniacanthidae	*Taeniacanthus zeugopteri*	✓	✓
Copepoda	Poecilostomatoida	Taeniacanthidae	*Umazuracola elongatus*	✓	✓
Copepoda	Pseudanthessiidae	Pseudanthessiidae	*Mecomerinx heterocentroti*		✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus belones*		✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus brevipedis*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus centrodonti*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus clemensi*		✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus curtus*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus elongatus*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus fugu*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus gurnardi*		✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus longirostris*		
Copepoda	Siphonostomatoida	Caligidae	*Caligus pelamydis*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus quadratus*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus rogercresseyi*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Caligus uniartus*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Gloiopotes watsoni*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus cuneifer*	✓	
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus hippoglossi*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus hospitalis*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus mugiloidis*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus natalensis*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus parvicruris*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus pectoralis*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus pollachius*	✓	✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus salmonis*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus thompsoni*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus yanezi*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus zbigniewi*		✓
Copepoda	Siphonostomatoida	Caligidae	*Paralebion elongatus*	✓	✓
Copepoda	Siphonostomatoida	Dichelesthiidae	*Anthosoma crassum*		✓

Copepoda	Siphonostomatoida	Dirivultidae	*Aphotopontius mammillatus*	✓	✓
Copepoda	Siphonostomatoida	Dirivultidae	*Stygiopontius brevispina*		
Copepoda	Siphonostomatoida	Dirivultidae	*Stygiopontius hispidulus*		
Copepoda	Siphonostomatoida	Dirivultidae	*Stygiopontius lauensis*		
Copepoda	Siphonostomatoida	Dirivultidae	*Rhogobius contractus*		✓
Copepoda	Siphonostomatoida	Dissonidae	*Dissonus manteri*		✓
Copepoda	Siphonostomatoida	Ecbathyriontidae	*Ecbathyrion prolixicauda*		✓
Copepoda	Siphonostomatoida	Eudactylinidae	*Eudactylina pusilla*	✓	✓
Copepoda	Siphonostomatoida	Eudactylinidae	*Nemesis lamna*		✓
Copepoda	Siphonostomatoida	Eudactylinidae	*Eudactylinodes niger*		✓
Copepoda	Siphonostomatoida	Hatschekiidae	*Hatschekia cadenati*		
Copepoda	Siphonostomatoida	Hatschekiidae	*Hatschekia cyanopodus*	✓	
Copepoda	Siphonostomatoida	Hatschekiidae	*Hatschekia iridescens*	✓	
Copepoda	Siphonostomatoida	Hatschekiidae	*Hatschekia maculatus*		
Copepoda	Siphonostomatoida	Hatschekiidae	*Hatschekia pagrosomi*	✓	✓
Copepoda	Siphonostomatoida	Kroyeriidae	*Kroyeria dispar*	✓	✓
Copepoda	Siphonostomatoida	Kroyeriidae	*Kroyeria longicauda*	✓	✓
Copepoda	Siphonostomatoida	Kroyeriidae	*Kroyeria papillipes*	✓	✓
Copepoda	Siphonostomatoida	Kroyeriidae	*Kroyeria sphyrnae*	✓	✓
Copepoda	Siphonostomatoida	Lernaeopodidae	*Clavella addunca*	✓	✓
Copepoda	Siphonostomatoida	Lernaeopodidae	*Parabrachiella bispinosa*	✓	✓
Copepoda	Siphonostomatoida	Lernanthropidae	*Lernanthropus callionymicola*	✓	
Copepoda	Siphonostomatoida	Lernanthropidae	*Lernanthropus mugilii*	✓	
Copepoda	Siphonostomatoida	Megapontiidae	*Hyalopontius typicus*		✓
Copepoda	Siphonostomatoida	Nanaspididae	*Nanaspis tonsa*		✓
Copepoda	Siphonostomatoida	Nicothoidae	*Choniosphaera maenadis*		✓
Copepoda	Siphonostomatoida	Nicothoidae	*Nicothoe astaci*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Pandarus smithi*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Perissopus dentatus*	✓	✓
Copepoda	Siphonostomatoida	Pandaridae	*Phyllothyreus cornutus*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Pseudopandarus longus*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Achtheinus oblongus*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Dinemoura latifolia*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Nesippus crypturus*	✓	✓
Copepoda	Siphonostomatoida	Pandaridae	*Nesippus orientalis*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Nesippus vespa*		✓
Copepoda	Siphonostomatoida	Pandaridae	*Pannosus japonicus*		✓
Copepoda	Siphonostomatoida	Pennellidae	*Lernaeocera branchialis*	✓	✓
Copepoda	Siphonostomatoida	Pontoeciellidae	*Pontoeciella abyssicola*	✓	✓
Copepoda	Siphonostomatoida	Sphyriidae	*Paeon elongatus*		✓
Copepoda	Siphonostomatoida	Caligidae	*Lepeophtheirus chilensis*		✓
Thecostraca	Akentrogonida	Clistosaccidae	*Clistosaccus paguri*		✓

Thecostraca	Akentrogonida	Clistosaccidae	*Sylon hippolytes*	✓	✓
Thecostraca	Akentrogonida	Polysaccidae	*Polysaccus japonicus*	✓	✓
Thecostraca	Akentrogonida	Thompsoniidae	*Pottsia serenei*	✓	✓
Thecostraca	Akentrogonida	Thompsoniidae	*Diplothylacus sinensis*	✓	✓
Thecostraca	Akentrogonida	Chthamalophilidae	*Chthamalophilus delagei*	✓	✓
Thecostraca	Dendrogastrida	Dendrogastridae	*Dendrogaster asterinae*		✓
Thecostraca	Dendrogastrida	Dendrogastridae	*Dendrogaster ludwigi*		✓
Thecostraca	Dendrogastrida	Dendrogastridae	*Ulophysema oeresundense*		✓
Thecostraca	Ibliformes	Iblidae	*Ibla cumingi*	✓	✓
Thecostraca	Ibliformes	Iblidae	*Ibla quadrivalvis*	✓	✓
Thecostraca	Kentrogonida	Lernaeodiscidae	*Lernaeodiscus porcellanae*	✓	✓
Thecostraca	Kentrogonida	Peltogastridae	*Peltogasterella sulcata*		✓
Thecostraca	Kentrogonida	Peltogastridae	*Peltogaster paguri*		✓
Thecostraca	Kentrogonida	Peltogastridae	*Septosaccus rodriguezii*		✓
Thecostraca	Kentrogonida	Sacculinidae	*Heterosaccus californicus*		✓
Thecostraca	Kentrogonida	Sacculinidae	*Heterosaccus dollfusi*		✓
Thecostraca	Kentrogonida	Sacculinidae	*Heterosaccus lunatus*		✓
Thecostraca	Kentrogonida	Sacculinidae	*Polyascus gregaria*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Polyascus plana*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Polyascus polygenea*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Sacculina carcini*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Sacculina confragosa*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Sacculina leptodiae*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Sacculina oblonga*		✓
Thecostraca	Kentrogonida	Sacculinidae	*Sacculina sinensis*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Bosmaella japonica*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Loxothylacus panopaei*	✓	✓
Thecostraca	Kentrogonida	Sacculinidae	*Loxothylacus texanus*		✓
Thecostraca	Laurida	Lauridae	*Baccalaureus maldivensis*		✓
Thecostraca	Laurida	Petrarcidae	*Zibrowia auriculata*	✓	✓
Thecostraca	Lepadiformes	Heteralepadidae	*Heteralepas japonica*		✓
Thecostraca	Lepadiformes	Heteralepadidae	*Heteralepas quadrata*		✓
Thecostraca	Lepadiformes	Heteralepadidae	*Koleolepas avis*	✓	✓
Thecostraca	Lepadiformes	Lepadidae	*Conchoderma hunteri*		✓
Thecostraca	Lepadiformes	Lepadidae	*Conchoderma virgatum*		✓
Thecostraca	Lepadiformes	Lepadidae	*Lepas anatifera*		✓
Thecostraca	Lepadiformes	Lepadidae	*Lepas anserifera*		✓
Thecostraca	Lepadiformes	Lepadidae	*Lepas australis*		✓
Thecostraca	Lepadiformes	Lepadidae	*Lepas pectinata*		✓
Thecostraca	Lepadiformes	Lepadidae	*Lepas testudinata*		✓
Thecostraca	Lepadiformes	Lepadomorpha	*Conchoderma auritum*		✓
Thecostraca	Lepadiformes	Oxynaspididae	*Oxynaspis celata*		✓

Thecostraca	Lepadiformes	Oxynaspididae	*Oxynaspis ryukyuensis*	✓	
Thecostraca	Lepadiformes	Poecilasmatidae	*Megalasma striatum*		✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Octolasmis angulata*		✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Octolasmis cor*		✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Octolasmis warwickii*		✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Temnaspis amygdalum*		✓
Thecostraca	Lepadiformes	Heteralepadidae	*Paralepas dannevigi*	✓	✓
Thecostraca	Lepadiformes	Heteralepadidae	*Paralepas palinuri*		✓
Thecostraca	Lepadiformes	Heteralepadidae	*Paralepas xenophorae*	✓	✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Poecilasma inaequilaterale*		✓
Thecostraca	Lepadiformes	Poecilasmatidae	*Poecilasma kaempferi*		✓
Thecostraca	Lithoglyptida	Lithoglyptidae	*Berndtia purpurea*		✓
Thecostraca	Lithoglyptida	Lithoglyptidae	*Auritoglyptes bicornis*	✓	✓
Thecostraca	Lithoglyptida	Trypetesidae	*Trypetesa lampas*	✓	✓
Thecostraca	Parthenopeidae	Parthenopeidae	*Parthenopea subterranea*	✓	✓
Thecostraca	Scalpelliformes	Calanticidae	*Calantica spinosa*		✓
Thecostraca	Scalpelliformes	Calanticidae	*Calantica villosa*	✓	✓
Thecostraca	Scalpelliformes	Calanticidae	*Smilium peronii*		✓
Thecostraca	Scalpelliformes	Calanticidae	*Smilium scorpio*		✓
Thecostraca	Scalpelliformes	Eolepadidae	*Ashinkailepas seepiophila*	✓	✓
Thecostraca	Scalpelliformes	Eolepadidae	*Neolepas rapanuii*	✓	✓
Thecostraca	Scalpelliformes	Eolepadidae	*Neolepas zevinae*	✓	✓
Thecostraca	Scalpelliformes	Eolepadidae	*Vulcanolepas scotiaensis*	✓	✓
Thecostraca	Scalpelliformes	Eolepadidae	*Vulcanolepas osheai*		✓
Thecostraca	Scalpelliformes	Lithotryidae	*Lithotrya valentiana*		✓
Thecostraca	Scalpelliformes	Pollicipedidae	*Capitulum mitella*		✓
Thecostraca	Scalpelliformes	Pollicipedidae	*Pollicipes mitella*	✓	
Thecostraca	Scalpelliformes	Pollicipedidae	*Pollicipes pollicipes*		✓
Thecostraca	Scalpelliformes	Pollicipedidae	*Pollicipes polymerus*		✓
Thecostraca	Scalpelliformes	Scalpellidae	*Arcoscalpellum africanum*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Arcoscalpellum beuveti*		✓
Thecostraca	Scalpelliformes	Scalpellidae	*Arcoscalpellum sociabile*		✓
Thecostraca	Scalpelliformes	Scalpellidae	*Scalpellum scalpellum*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Scalpellum stearnsii*		✓
Thecostraca	Scalpelliformes	Scalpellidae	*Ornatoscalpellum stroemii*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Trianguloscalpellum balanoides*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Trianguloscalpellum regium*	✓	✓
Thecostraca	Scalpelliformes	Eolepadidae	*Leucolepas longa*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Litoscalpellum discoveryi*	✓	✓
Thecostraca	Scalpelliformes	Scalpellidae	*Litoscalpellum regina*	✓	✓
Thecostraca	Sessilia	Archaeobalanidae	*Armatobalanus allium*	✓	✓
Thecostraca	Sessilia	Archaeobalanidae	*Armatobalanus cepa*		✓

Thecostraca	Sessilia	Archaeobalanidae	*Conopea fidelis*				
Thecostraca	Sessilia	Archaeobalanidae	*Conopea galeata*	✓			
Thecostraca	Sessilia	Archaeobalanidae	*Conopea saotomensis*				
Thecostraca	Sessilia	Archaeobalanidae	*Membranobalanus longirostrum*				
Thecostraca	Sessilia	Archaeobalanidae	*Semibalanus balanoides*	✓		✓	
Thecostraca	Sessilia	Archaeobalanidae	*Semibalanus cariosus*	✓		✓	
Thecostraca	Sessilia	Balanidae	*Amphibalanus amphitrite*	✓		✓	
Thecostraca	Sessilia	Balanidae	*Amphibalanus eburneus*	✓		✓	
Thecostraca	Sessilia	Balanidae	*Amphibalanus reticulatus*	✓			
Thecostraca	Sessilia	Balanidae	*Amphibalanus rhizophorae*	✓			
Thecostraca	Sessilia	Balanidae	*Amphibalanus variegatus*				
Thecostraca	Sessilia	Balanidae	*Amphibalanus zhujiangensis*				
Thecostraca	Sessilia	Balanidae	*Balanus balanus*	✓	✓	✓	✓
Thecostraca	Sessilia	Balanidae	*Balanus crenatus*			✓	
Thecostraca	Sessilia	Balanidae	*Balanus glandula*		✓	✓	✓
Thecostraca	Sessilia	Balanidae	*Balanus nubilus*			✓	
Thecostraca	Sessilia	Balanidae	*Balanus perforatus*	✓		✓	
Thecostraca	Sessilia	Balanidae	*Balanus trigonus*	✓			
Thecostraca	Sessilia	Balanidae	*Fistulobalanus albicostatus*	✓			
Thecostraca	Sessilia	Balanidae	*Megabalanus ajax*				
Thecostraca	Sessilia	Balanidae	*Megabalanus californicus*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus coccopoma*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus occator*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus rosa*				
Thecostraca	Sessilia	Balanidae	*Megabalanus spinosus*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus stultus*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus tintinnabulum*			✓	
Thecostraca	Sessilia	Balanidae	*Megabalanus volcano*				
Thecostraca	Sessilia	Balanidae	*Megabalanus zebra*				
Thecostraca	Sessilia	Balanidae	*Wanella milleporae*			✓	
Thecostraca	Sessilia	Bathylasmatidae	*Bathylasma corolliforme*	✓			
Thecostraca	Sessilia	Catophragmidae	*Catomerus polymerus*	✓		✓	
Thecostraca	Sessilia	Catophragmidae	*Catophragmus imbricatus*			✓	
Thecostraca	Sessilia	Chthamalidae	*Caudoeuraphia caudata*	✓		✓	
Thecostraca	Sessilia	Chthamalidae	*Chamaesipho brunnea*			✓	
Thecostraca	Sessilia	Chthamalidae	*Chamaesipho columna*			✓	
Thecostraca	Sessilia	Chthamalidae	*Chamaesipho tasmanica*			✓	
Thecostraca	Sessilia	Chthamalidae	*Chthamalus anisopoma*	✓		✓	
Thecostraca	Sessilia	Chthamalidae	*Chthamalus antennatus*	✓		✓	
Thecostraca	Sessilia	Chthamalidae	*Chthamalus bisinuatus*	✓		✓	
Thecostraca	Sessilia	Chthamalidae	*Chthamalus challengeri*			✓	
Thecostraca	Sessilia	Chthamalidae	*Chthamalus dentatus*			✓	

Thecostraca	Sessilia	Chthamalidae	*Chthamalus fissus*		✓
Thecostraca	Sessilia	Chthamalidae	*Chthamalus fragilis*	✓	✓
Thecostraca	Sessilia	Chthamalidae	*Chthamalus malayensis*		✓
Thecostraca	Sessilia	Chthamalidae	*Chthamalus montagui*		✓
Thecostraca	Sessilia	Chthamalidae	*Chthamalus panamensis*		
Thecostraca	Sessilia	Chthamalidae	*Chthamalus proteus*		✓
Thecostraca	Sessilia	Chthamalidae	*Chthamalus stellatus*		✓
Thecostraca	Sessilia	Chthamalidae	*Euraphia rhizophorae*		
Thecostraca	Sessilia	Chthamalidae	*Hexechamaesipho pilsbryi*		✓
Thecostraca	Sessilia	Chthamalidae	*Microeuraphia depressa*		✓
Thecostraca	Sessilia	Chthamalidae	*Microeuraphia rhizophorae*		✓
Thecostraca	Sessilia	Chthamalidae	*Microeuraphia withersi*	✓	✓
Thecostraca	Sessilia	Chthamalidae	*Nesochthamalus intertextus*		✓
Thecostraca	Sessilia	Chthamalidae	*Octomeris angulosa*		✓
Thecostraca	Sessilia	Chthamalidae	*Octomeris brunnea*		✓
Thecostraca	Sessilia	Chthamalidae	*Pseudoctomeris sulcata*		✓
Thecostraca	Sessilia	Coronulidae	*Chelonibia patula*		✓
Thecostraca	Sessilia	Coronulidae	*Chelonibia testudinaria*	✓	✓
Thecostraca	Sessilia	Platylepadidae	*Stomatolepas elegans*		
Thecostraca	Sessilia	Platylepadidae	*Stomatolepas praegustator*	✓	✓
Thecostraca	Sessilia	Platylepadidae	*Stomatolepas transversa*	✓	✓
Thecostraca	Sessilia	Pyrgomatidae	*Adna anglica*	✓	✓
Thecostraca	Sessilia	Pyrgomatidae	*Cantellius hoegi*		
Thecostraca	Sessilia	Pyrgomatidae	*Cantellius pallidus*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Cantellius sextus*		
Thecostraca	Sessilia	Pyrgomatidae	*Ceratoconcha domingensis*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Ceratoconcha paucicostata*	✓	✓
Thecostraca	Sessilia	Pyrgomatidae	*Darwiniella angularis*	✓	
Thecostraca	Sessilia	Pyrgomatidae	*Darwiniella conjugatum*	✓	✓
Thecostraca	Sessilia	Pyrgomatidae	*Galkinia adamanteus*		
Thecostraca	Sessilia	Pyrgomatidae	*Galkinia altiapiculus*		
Thecostraca	Sessilia	Pyrgomatidae	*Galkinia equus*		
Thecostraca	Sessilia	Pyrgomatidae	*Galkinia indica*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Galkinia tabulatus*	✓	
Thecostraca	Sessilia	Pyrgomatidae	*Hiroa stubbingsi*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Neotrevathana elongatum*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Nobia grandis*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Nobia orbicellae*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Pyrgopsella annandalei*		
Thecostraca	Sessilia	Pyrgomatidae	*Pyrgopsella youngi*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Savignium crenatum*		✓
Thecostraca	Sessilia	Pyrgomatidae	*Trevathana jensi*		✓

Thecostraca	Sessilia	Pyrgomatidae	*Trevathana paulayi*		✓
Thecostraca	Sessilia	Tetraclitidae	*Tesseropora rosea*	✓	
Thecostraca	Sessilia	Tetraclitidae	*Tetraclita ehsani*	✓	
Thecostraca	Sessilia	Tetraclitidae	*Tetraclita japonica*	✓	✓
Thecostraca	Sessilia	Tetraclitidae	*Tetraclita kuroshioensis*		
Thecostraca	Sessilia	Tetraclitidae	*Tetraclita rubescens*		✓
Thecostraca	Sessilia	Tetraclitidae	*Tetraclita squamosa*	✓	✓
Thecostraca	Sessilia	Tetraclitidae	*Tetraclitella chinensis*		
Thecostraca	Sessilia	Tetraclitidae	*Tetraclitella divisa*	✓	✓
Thecostraca	Sessilia	Tetraclitidae	*Tetraclitella karandei*	✓	
Thecostraca	Sessilia	Tetraclitidae	*Tetraclitella purpurascens*	✓	✓
Thecostraca	Sessilia	Verrucidae	*Metaverruca recta*		✓
Thecostraca	Sessilia	Verrucidae	*Verruca laevigata*		✓
Thecostraca	Sessilia	Verrucidae	*Verruca spengleri*	✓	✓
Thecostraca	Sessilia	Verrucidae	*Verruca stroemia*	✓	✓
Thecostraca	Sessilia	Austrobalanidae	*Elminius kingii*	✓	✓
Thecostraca	Sessilia	Austrobalanidae	*Elminius modestus*		✓
Thecostraca	Sessilia	Austrobalanidae	*Austrobalanus imperator*		✓
Thecostraca	Sessilia	Balanidae	*Austromegabalanus psittacus*	✓	✓
Thecostraca	Sessilia	Balanidae	*Menesiniella aquila*	✓	✓
Thecostraca	Sessilia	Chelonibiidae	*Chelonibia caretta*	✓	✓
Thecostraca	Sessilia	Chelonibiidae	*Chelonibia manati*		✓
Thecostraca	Sessilia	Chthamalidae	*Notochthamalus scabrosus*		✓
Thecostraca	Sessilia	Chthamalidae	*Tetrachthamalus oblitteratus*		✓
Thecostraca	Sessilia	Chthamalidae	*Jehlius cirratus*		✓
Thecostraca	Sessilia	Coronulidae	*Coronula diadema*	✓	✓
Thecostraca	Sessilia	Coronulidae	*Cryptolepas rhachianecti*		✓
Thecostraca	Sessilia	Coronulidae	*Tubicinella cheloniae*	✓	✓
Thecostraca	Sessilia	Coronulidae	*Xenobalanus globicipitis*		✓
Thecostraca	Sessilia	Neoverrucidae	*Neoverruca brachylepadoformis*		✓
Thecostraca	Sessilia	Pachylasmatidae	*Pachylasma japonicum*	✓	✓
Thecostraca	Sessilia	Platylepadidae	*Cylindrolepas darwiniana*		✓
Thecostraca	Sessilia	Platylepadidae	*Cylindrolepas sinica*		✓
Thecostraca	Sessilia	Platylepadidae	*Platylepas decorata*	✓	✓
Thecostraca	Sessilia	Platylepadidae	*Platylepas hexastylos*		✓
Thecostraca	Sessilia	Platylepadidae	*Stephanolepas muricata*	✓	✓
Thecostraca	Sessilia	Pyrgomatidae	*Pyrgoma cancellatum*		✓
Thecostraca	Sessilia	Verrucidae	*Rostratoverruca krugeri*	✓	✓

Appendix VI: Sequence alignment data for Chapter 4

Multiple sequence alignment lengths for the 11 studied genera.

Genus	# of sequences in data set	Sequence Length (in bp)
Acartia	138	554
Aetideus	7	534
Calanus	180	585
Centropages	149	557
Eurytemora	352	546
Mesocalanus	8	557
Metridia	65	558
Paracalanus	60	501
Pseudocalanus	120	471
Temora	28	543
Tortanus	48	576

Appendix VII: Tabulation of node transitions for Chapter 4

Values do not include outgroup terminal branches or clades nor the node connecting the ingroup and outgroup. The Atlantic+Arctic Oceans are considered as one oceanic region here.

	Terminal Clades	BINs	Nodes	Within-ocean lineage split	Pacific to Arctic or Atlantic	Arctic or Atlantic to Pacific	Transition at node with uncertain direction	Notes
Acartia	15	15	14	10	4			
Aetideus	4	4	3	2		1		
Calanus	3	3	2	1	1			
Centropages	7	7	6	5		1		
Eurytemora	4	4	3	3				
Mesocalanus	4	4	3	1	1		1	
Metridia	8	8	7	3	2	1	1	One node contains an Antarctic species and was considered of uncertain direction; one paraphyletic BIN was treated as two terminal clades
Paracalanus	10	10	9	6	2		1	
Pseudocalanus	15	14	14	7	4	2	1	Two terminal clades lacked BIN labels; one terminal clade contained three intermixed BINs; one paraphyletic BIN was treated as two terminal clades
Temora	2	2	1				1	
Tortanus	8	8	7	5	2			
Total	**80**	**79**	**69**	**43**	**16**	**5**	**6**	