Molecular clocks and rates of evolution in marine invertebrates

by

Tzitziki del Rosario Loeza Quintana

A Thesis
Presented to
The Faculty of Graduate Studies
of
The University of Guelph

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

Guelph, Ontario, Canada

© Tzitziki del Rosario Loeza Quintana, August 2017
ABSTRACT

MOLECULAR CLOCKS AND RATES OF EVOLUTION IN MARINE INVERTEBRATES

Tzitziki del Rosario Loeza Quintana
University of Guelph, 2017

Advisor: Sarah J. Adamowicz

The molecular clock is an important tool in evolutionary biology for dating the tree of life, particularly for lineages with a poor fossil record such as many marine invertebrates. Historically, biogeographic calibrations assumed simultaneous divergence of allopatric pairs of sister lineages, which is problematic for long and complex vicariance events. This thesis presents a new methodology for calibrating the molecular clock using complex biogeographic events, the iterative calibration approach. This approach calibrates the rate of molecular evolution by finding the rate that achieves concordance between multiple lines of evidence: geographic distributions, genetic divergences, geological history, and the fossil record. Using the novel and successful method, this study explores two biogeographic events for clock dating, the opening/re-closure of the Bering Strait and the formation of the Isthmus of Panama, and investigates rates and patterns of molecular evolution in four major groups of marine invertebrates (echinoderms, molluscs, arthropods, and polychaetes) and between Arctic vs. tropical taxa. When investigating the divergence rate in 157 sister pairs of marine invertebrates based on publicly available sequences of the mitochondrial cytochrome c oxidase subunit I (COI) gene, the results show that, generally, echinoderms have slower rates of molecular evolution than molluscs and polychaetes. Interestingly, tropical arthropods have the slowest rate, while Arctic arthropods have the highest rate among the four groups. Absolute rates of molecular evolution were observed to be higher in Arctic lineages than in their tropical counterparts, which may be attributed, for example, to smaller effective population size in Arctic populations. Overall, the novel iterative calibration approach, the large geographic and taxonomic scales, and the unprecedented sample size of sister pairs used in this study provide significant advances in clock calibration research. Moreover, the rates of molecular evolution resulting from this study suggest rate heterogeneity across phyla and a difference in rates between tropical vs. Arctic lineages. This study represents a novel and valuable contribution to research in marine organisms, which are commonly understudied. Lastly, this thesis will
contribute to studying the tree of life and understanding the impact of prior climate change events upon the diversification of life.

**Key words:** molecular dating, clock calibration, Bering Strait, Isthmus of Panama, DNA barcoding, Echinodermata, Mollusca, Arthropoda, Polychaeta.
ACKNOWLEDGMENTS

This project would not have been possible without numerous sources of funding during my degree, and I am truthfully grateful! This research was supported by an International Graduate Studies Scholarship to TLQ (315757) from Consejo Nacional de Ciencia y Tecnología (CONACYT) in partnership with the University of Guelph, and by a research grant to Sarah J. Adamowicz from the Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery Grants 2010-386591 and 2016-06199). Additional travel support was awarded to TLQ by two Arthur D. Latornell Travel Scholarships (University of Guelph) and a Graduate Student Association Travel Grant (University of Guelph).

I am deeply grateful to my advisor Dr. Sarah Adamowicz for her continuous support, encouragement and enthusiasm through my degree. Sally, thank you for being an exceptional supervisor, for all the helpful discussions and advice, and for all the assistance you have provided, including your help with the R code. I feel very privileged for the opportunity to be part of your Lab, and I sincerely appreciate the freedom you gave me to pursue my research areas of interest.

Special thanks to the members of my advisory committee, Dr. Teresa J. Crease, Dr. Elizabeth G. Boulding, and Dr. Jonathan D.S. Witt, for their valuable comments, discussions, and suggestions through my degree. I am also very grateful to the members of my defence committee, Dr. Filipe Costa, Dr. Jinzhong Fu, Dr. Teresa J. Crease, Dr. Karl Cottenie, and Dr. Sarah Adamowicz for the thoughtful discussion and comments on my thesis. Also, I would like to acknowledge to all the researchers around the globe who have contributed and made available the data used in this research project, including those who have committed to the development of BOLD, which was the primary source of data for this project. Without their contribution, this project would not have been possible.

Infinite thanks to Christina M. Carr for providing genetic divergence data from Arctic polychaetes, and because her research on polychaetes was a significant influence to the foundation of this project. I am incredibly grateful to Tooba Khan and the volunteers Samantha Lyon and Yash A. Bhatt, for their interest, excellent work, and important contribution to this project. You guys are amazing! Many thanks to all my former and present lab-mates from Adamowicz and Hebert Labs, who have contributed in so many ways through discussions, comments, and advice at different stages of this project. It has been wonderful sharing grad school with all of you!
I have enjoyed every single step of pursuing my doctoral research at the University of Guelph, in a supportive, friendly, and beautiful environment. Thank you UofG and Canada. Additionally, I am sincerely grateful to my wonderful friends here in Canada, who have truly made this time more meaningful and less difficult to be away from home. I am extraordinarily lucky to find amazing people wherever I go.

I am deeply thankful to my husband Justin for his technical support, and for all those countless lunches and morning coffees that kept me running during school days; many thanks for the tremendous support, encouragement, love, and for being part of this journey. I also want to thank my mother-in-law Nancy, for her constant love and support through my degree.

Finally, I would like to acknowledge my parents José Luis Loeza and Socorro Quintana, and my sisters Catzingari & Xhangari, for their endless love, support, encouragement, and motivation. Thank you for understanding the physical absence of this curious mind, because no matter where my dreams would take me, you are always by my side.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................................................................................. IV
LIST OF TABLES .......................................................................................................................... IX
LIST OF FIGURES ........................................................................................................................ X
LIST OF SUPPLEMENTARY MATERIALS .................................................................................. XII

CHAPTER I ...................................................................................................................................... 1

AN INTRODUCTION TO MOLECULAR CLOCKS AND MOLECULAR EVOLUTIONARY RATES .......................................................................................................................... 1
THESIS OVERVIEW ...................................................................................................................... 1
THE MOLECULAR CLOCK AND MOLECULAR EVOLUTIONARY RATES .................................... 2
   Calibrating the Molecular Clock ............................................................................................... 3
   Time-dependency Hypothesis .................................................................................................. 4
   Rate Heterogeneity .................................................................................................................. 5
   Synthesis of Molecular Clock Research in Marine Invertebrates ........................................... 7
STUDY QUESTIONS, OBJECTIVES, HYPOTHESIS, AND PREDICTIONS ................................ 8
SIGNIFICANCE ............................................................................................................................. 9
TABLES ......................................................................................................................................... 10
FIGURES ....................................................................................................................................... 17

CHAPTER II ................................................................................................................................... 19

ITERATIVE CALIBRATION: A NOVEL APPROACH FOR CALIBRATING THE MOLECULAR CLOCK USING COMPLEX GEOLOGICAL EVENTS .................................................................................. 19
ABSTRACT ..................................................................................................................................... 19
INTRODUCTION ............................................................................................................................. 20
   Using the Bering Strait for Clock Calibration in Northern Marine Taxa .................................. 21
METHODS ....................................................................................................................................... 23
   Finding Potential Trans-Bering Sister Clades and Data Collection ....................................... 23
   Trans-Bering Sister Clades Confirmation through Maximum Likelihood Phylogenetic Analysis .................................................................................................................................................. 24
   Rate Constancy Test ................................................................................................................ 25
   Kimura 2-Parameter Distances between Trans-Bering Sister Clades ................................... 25
   Generation of Relative Node Ages through Bayesian Analysis .......................................... 26
   Iterative Calibration ............................................................................................................... 27
RESULTS ......................................................................................................................................... 29
   Trans-Bering Sister Clades ...................................................................................................... 29
   Relative Rate Test on Trans-Bering Lineages ........................................................................ 29
   Kimura 2-Parameter Distance Analysis and Clock Calibration ............................................ 29
   Bayesian Results and Clock Calibration ................................................................................. 31
   Considering Biological Traits of the Trans-Bering Sister Clades .......................................... 32
DISCUSSION .................................................................................................................................. 32
   Strengths of the Iterative Calibration Method ...................................................................... 32
   Divergence Times for Northern Echinoderms ........................................................................ 34
   Rate Heterogeneity in the Pseudostichopus Sister Pair ......................................................... 34
   Rate Estimation for Northern Echinoderms .......................................................................... 35
   Rate Estimation in Northern Echinoderms vs. other Rates in the Literature ....................... 36
   Phylogenetic Scope for the Substitution Rate Presented in this Study .................................. 38
LIST OF TABLES

**Table 1.1** Summary of molecular clock calibrations for the mitochondrial cytochrome c oxidase subunit I gene (COI) in marine invertebrates ..............................................................10

**Table 3.1** Summary of molecular clock calibrations for the mitochondrial cytochrome c oxidase subunit I gene (COI) in the four groups of Northern marine invertebrates used in this study ..............................................................................................................71

**Table 3.2** Summary of the statistical test results for the four groups of Northern marine invertebrates used in this study ..............................................................................................................71

**Table 4.1** Summary of the geological history of the formation of the Isthmus of Panama ........................................................................................................................................97

**Table 4.2** Summary of molecular rates for the mitochondrial cytochrome c oxidase subunit I gene (COI) in three groups of tropical marine invertebrates used in this study (chapter IV) and the four groups of Arctic marine invertebrates from chapters II and III ........................................................................100

**Table 4.3** Potential explanations for rate variability between Arctic and tropical taxa ........101
Figure 3.6 Divergence time estimates for molluscs and arthropods using K2P divergences based on the best calibration and mapping the biological zone ........................................77

Figure 3.7 Divergence time estimates for molluscs and arthropods based on the best calibration using K2P divergences and mapping the completeness of the sampling ..............................78

Figure 4.1 Kimura two-parameter (K2P) divergences retrieved from the literature for trans-isthmus sister clades of echinoderms, molluscs, and arthropods ...........................................106

Figure 4.2 Divergence time estimates for tropical marine echinoderms, molluscs, and arthropods using K2P distances and the best calibration according to the iterative calibration approach ........................................................................................................107

Figure 4.3 Depth habitat distribution of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III) ................................108

Figure 4.4 Comparison of the taxonomic composition at the class level of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III) ........................................................................................................109

Figure 4.5 Habitat zone of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III) ..................110

Figure 5.1 Divergence rates estimated in this study compared with the range of rates reported in the literature for marine invertebrates ..............................................................................116
LIST OF SUPPLEMENTARY MATERIALS

CHAPTER II

Part A: is provided in the Appendix in this thesis document
Supplementary Material 2.7 Bayesian analysis settings
Supplementary Material 2.8 List of trans-Bering sister clades of echinoderms

Part B: is available as an additional Excel document in the University of Guelph Research Data Repository (http://doi.org/10.5683/SP/B1IAGG)
Supplementary Material 2.1 Genus-level datasets and analysis for trans-Bering echinoderms
Supplementary Material 2.2 Example of a problematic sequence removed from the alignment
Supplementary Material 2.3 Sequence BOLD process IDs and GenBank accession numbers
Supplementary Material 2.4 Tajima’s Relative Rate test (Tajima 1993)
Supplementary Material 2.5 Results from the simple and traditional approach for clock calibration
Supplementary Material 2.9 Results from the iterative calibration using K2P divergences
Supplementary Material 2.10 Relative node ages from six independent runs in BEAST
Supplementary Material 2.11 Bayesian tree after eliminating Pseudostichopus pair
Supplementary Material 2.12 Results from the iterative calibration using relative node ages
Supplementary Material 2.13 Results from the iterative calibration using relative node ages after eliminating Pseudostichopus pair
Supplementary Material 2.14 Biological traits of trans-Bering echinoderms
Supplementary Material 2.15 Example of possible sources of error related to phylogenies with incomplete taxon sampling

Part C: is available as an additional text document in the University of Guelph Research Data Repository (http://doi.org/10.5683/SP/B1IAGG)
Supplementary Material 2.6 Sequence alignment

CHAPTER III

Part A: is provided in the Appendix in this thesis document
Supplementary Material 3.2 List of trans-Bering sister clades of molluscs and arthropods
Supplementary Material 3.6 Biological traits of trans-Bering molluscs, arthropods, and
polychaetes
Supplementary Material 3.7. List of trans-Bering sister clades of polychaetes retrieved from Carr (2010) and reanalyzed in Chapter III
Supplementary Material 3.8 Statistical analysis R commands

Part B: is available as an additional Excel document in the University of Guelph Research Data Repository (http://doi.org/10.5683/SP/B1IAGG)
Supplementary Material 3.1 Datasets and ML analysis for trans-Bering sisters of molluscs and arthropods
Supplementary Material 3.3 Tajima’s Relative Rate test
Supplementary Material 3.4 Saturation test (available in a separate Excel document)
Supplementary Material 3.5 Completeness of the sampling of trans-Bering molluscs and arthropods
Supplementary Material 3.9 Results from the iterative calibration in trans-Bering molluscs using K2P divergences
Supplementary Material 3.10 Results from the iterative calibration in trans-Bering molluscs using TN93+G divergences
Supplementary Material 3.11 Results from the iterative calibration in trans-Bering arthropods using K2P divergences
Supplementary Material 3.12 Results from the iterative calibration in trans-Bering arthropods using TN93+G divergences
Supplementary Material 3.13 Results from the iterative calibration applied to trans-Bering sisters of polychaetes from a previous study (Carr 2010)
Supplementary Material 3.14 Results from the iterative calibration in benthic trans-Bering molluscs using K2P divergences
Supplementary Material 3.15 Results from the iterative calibration in benthic trans-Bering arthropods using K2P divergences
Supplementary Material 3.16 Results from the iterative calibration in trans-Bering molluscs from phylogenies with good and fair taxon sampling
Supplementary Material 3.17 Results from the iterative calibration in trans-Bering arthropods from phylogenies with good and fair taxon sampling
CHAPTER IV

Part A: is provided in the Appendix in this thesis document
Supplementary Material 4.1 List of trans-isthmus sister clades of echinoderms, molluscs, and arthropods retrieved from the literature (Hart et al. 1997; Lessios 2008; ODea et al. 2016) and reanalyzed in Chapter IV

Part B: is available as an additional Excel document in the University of Guelph Research Data Repository (http://doi.org/10.5683/SP/B1IAGG)
Supplementary Material 4.2 K2P distances of trans-isthmus sister clades retrieved from previous literature
Supplementary Material 4.3 Completeness of the sampling of trans-isthmus echinoderms, molluscs, and arthropods
Supplementary Material 4.4 Results from the iterative calibration in trans-isthmus echinoderms using K2P divergences
Supplementary Material 4.5 Results from the iterative calibration in trans-isthmus molluscs using K2P divergences
Supplementary Material 4.6 Results from the iterative calibration in trans-isthmus arthropods using K2P divergences
Supplementary Material 4.7 Results from the iterative calibration in trans-isthmus molluscs from phylogenies with good and fair sampling
Supplementary Material 4.8 Results from the iterative calibration in trans-isthmus arthropods from phylogenies with good and fair sampling
Supplementary Material 4.9 Results from the iterative calibration in trans-isthmus molluscs from class Bivalvia using K2P divergences
CHAPTER I
An Introduction to Molecular Clocks and Molecular Evolutionary Rates

THESIS OVERVIEW

Establishing an evolutionary timescale for the tree of life is a central goal in evolutionary biology, where molecular clocks have become an important tool for dating biological and evolutionary processes. Molecular clocks are essential for dating evolutionary events in lineages with a patchy or absent fossil record, such as many marine invertebrates, which are the target of this study. In order to be useful, molecular clocks need to be calibrated using independent information, such as biogeographic events. In taxonomic groups with poor paleontological evidence, biogeographic events represent the main source for clock calibration. However, biogeographic events are usually complex processes that have nevertheless been primarily used in a simplistic manner in molecular clock research, thus leaving important research gaps. In addition, the evidence of rate variability across the tree of life evokes the need for further research to understand patterns of rate heterogeneity among lineages.

In this thesis, I examine the use of two complex biogeographic events for dating divergence events, the opening/re-closure of the Bering Strait and the formation of the Isthmus of Panama. Additionally, I explore rates and patterns of molecular evolution in four groups of marine invertebrates (echinoderms, molluscs, arthropods, and polychaetes), as well as between Arctic vs. tropical lineages. First, I develop a new method for dating divergence events using complex biogeographic calibrations, the iterative calibration approach (Chapter II). Then, using the novel iterative calibration and the geological history of the Bering Strait, I investigate and compare patterns of molecular divergence across different lineages of Northern marine invertebrates (Chapters II and III). Finally, I explore the use of the iterative calibration approach as applied to the Isthmus of Panama to calibrate molecular clocks in tropical marine invertebrates, with the final goal to compare absolute rates of molecular evolution between taxa inhabiting tropical vs. Arctic environments (Chapter IV). Overall, this thesis represents an advance in clock calibration research, especially for marine lineages, and contributes to our understanding of the impact of prior climate change events on the history of life.
The molecular clock postulates that molecules of life (i.e. nucleotides and proteins) evolve at a constant rate over time among lineages (Zuckerkandl & Pauling 1965a; b); therefore, genetic divergences between any two species would be proportional to the time since they last shared a common ancestor. The molecular clock is a prediction of the neutral theory of molecular evolution, in which the rate of nucleotide substitution is equivalent to the rate of neutral mutation per generation, and it should be constant over long periods of time (Ohta & Kimura 1971). However, an extension to the neutral theory was later proposed, the nearly neutral theory, recognizing the effects of effective population size ($N_e$) and genetic drift vs. natural selection upon the fate of slightly deleterious mutations (Ohta 1992). The nearly neutral theory highlights that the frequency of nearly neutral mutations with small selection coefficients, whether positive or negative, can increase to fixation by the action of genetic drift in small populations. However, as $N_e$ increases, genetic drift becomes weaker since single random events will have a smaller impact over the population (Lanfear et al. 2014). Thus, natural selection becomes more effective at eliminating slightly deleterious mutations from a population with large $N_e$ (Ohta 1992; Woolfit & Bromham 2005). According to the nearly neutral theory, large populations would have lower substitution rates per generation than smaller populations (Woolfit 2009; Lanfear et al. 2014). Originally, the molecular clock was based on the observation of rate constancy over time among lineages (Zuckerkandl & Pauling 1965a; b). However, further work has revealed rate heterogeneity across the tree of life (e.g. Martin et al. 1992; Martin & Palumbi 1993; Mooers & Harvey 1994; Bleiweiss 1998a; Bromham 2002). Thus, the molecular clock is considered a hypothesis that should be tested whenever possible. Despite this complexity, the molecular clock remains an important tool for estimating evolutionary time scales and for exploring the mechanisms and patterns of evolution.

Since its proposal half a century ago (Zuckerkandl & Pauling 1965a; b), the molecular clock has become exceptionally valuable in evolutionary biology and in many related areas, such as molecular ecology, systematics, and conservation genetics. Moreover, it is particularly significant for those organisms that have left few or no traces of their biological history in the fossil record, such as viruses, parasites, and many invertebrates.

Despite the significance of the molecular clock in the study of evolutionary biology, it has been subject to debate. One topic of controversy is the calibration of the molecular clock, which is one of the most basic and critical aspects that still need attention (Wilke et al. 2009; Ho & Duchêne 2014). In addition, the use of the molecular clock has been questioned due to the
evidence of rate variability across genes (e.g. Papadopoulou et al. 2010; Willett 2012; Lavinia et al. 2015) and among lineages (e.g. Jukes & Holmquist 1972; Martin & Palumbi 1993; Lanfear et al. 2010). Furthermore, there is also evidence of variability in the rate of evolution over time (Ho et al. 2005, 2007, 2011, 2015a). Evidence of rate heterogeneity might cause hesitation about the use of the molecular clock, but new methods that relax the assumption of rate constancy have been developed, allowing dating in the face of the variation in molecular rates across lineages (see review in Ho 2014; Ho & Duchêne 2014). Despite uncertainties, the molecular clock remains an irreplaceable tool in evolutionary biology, essential for elucidating the time and pace of evolutionary events in many taxonomic groups and thus providing a time scale for the history of life.

*Calibrating the Molecular Clock*

Molecular data alone can only provide relative divergence time estimations. In order to provide an absolute time scale for the history of life, molecular clocks need to be calibrated. When external information about the geological age of an evolutionary divergence event is available, based on fossils or geological events, the distance between sequences or the branch lengths on the tree can then be transformed into absolute geological times, thus calibrating the molecular clock (Rannala & Yang 2014). The fossil record remains the primary source for clock calibration (Hipsley & Müller 2014). However, fossil calibrations are often problematic because fossils can only provide minimum age estimates for divergence events. Additionally, fossils are often difficult to identify and to classify, and sometimes there are also uncertainties about the age of fossils (Wilke et al. 2009). Furthermore, fossil evidence is commonly absent for many lineages. For taxonomic groups with a poor or absent fossil record, geological events are the alternative source for calibration (Ho & Duchêne 2014; Ho et al. 2015b). Geological and climatic events can trigger the divergence of populations and speciation, and these can be used as biogeographic calibrations of the molecular clock.

Biogeographic calibrations can be established using a broad variety of geological events, from ancient tectonic events (e.g. Ericson et al. 2002; Phillips et al. 2013; Landis 2016) to recent drivers of population subdivision (e.g. Edwards & Beerli 2000; Marino et al. 2011; Crandall et al. 2012); thus, biogeographic calibrations can be informative across a wide range of timeframes. However, the nature of some geological events makes them difficult to use. For example, determining the date of a biogeographic calibration is particularly problematic when using complex geological events, for which assigning a single calibration date is difficult (e.g. the repetitive opening and closure of the Bering Strait). In other cases, the development of a major
barrier to dispersal would not have been immediate (e.g. the formation of the Isthmus of Panama). Unfortunately, complex geological events have frequently been used in a simplistic manner, generally assigning calibration dates to the most recent occurrence of the geological or climatic event instead of accounting for the prolonged nature of the event (Ho et al. 2015b). Furthermore, studies generally imply that a geological event simultaneously affected all organisms, ignoring taxon-specific differences in biological and ecological traits (e.g. Hickerson et al. 2003; Lessios 2008). Despite the ambiguities, biogeographic calibrations remain invaluable for clock dating. One of the advantages of biogeographic calibrations is that many independent sister lineages within multiple taxonomic groups can be used during the clock calibration process (Wilke et al. 2009). Additionally, biogeographic calibrations provide the only effective source of estimating divergence times independently of paleontological evidence (Wilke et al. 2009; De Baets et al. 2016).

**Time-dependency Hypothesis**

Several studies have reported that molecular evolutionary rates estimated based on recent calibrations are an order of magnitude or more greater than those estimated from older calibrations (e.g. Marko 2002; Henn et al. 2009; Papadopoulou et al. 2010; Ho & Lo 2013; Molak & Ho 2015). This phenomenon of non-constancy in the rate of evolution over time is the so-called “time-dependency hypothesis” (Ho et al. 2005, 2007, 2011, 2015a). There are several possible explanations for the phenomenon of time dependency. One explanation involves the balance of selection and genetic drift. Some of the mutations detected in studies using recent calibrations will eventually be fixed, but some will be lost by genetic drift and others will be eliminated by selection. Studies using older calibrations will only detect fixed mutations (Pulquério & Nichols 2007). In addition, mutational saturation can trigger patterns of time dependency (Ho et al. 2011). Saturation effects could lead to an underestimation of divergence times due the difference between the observed and expected number of mutations (Wilke et al. 2009). The saturation effect is likely to be less problematic over very short time frames, but it is an important factor to be considered in studies involving long time frames (Ho et al. 2011).

The time-dependency effect is evident when comparing rates based on recent calibrations (e.g. recent geological events) vs. those estimated using older calibrations (e.g. fossils) (Fig. 1.1). Fossil estimates usually represent the minimum age for a given clade; thus, the actual date of the divergence event can substantially predate the oldest known fossil for that clade and can lead to an underestimation of substitution rates (Ho et al. 2005). By contrast, many recent calibrations (e.g. recent vicariance events) are based on species-divergence events, which can
lead to an underestimate of the gene divergence time and therefore an overestimation of substitution rates. These factors relating to error in assigning the date of divergence, in addition to potential saturation effects, are particularly problematic when comparing rates based on calibrations from different time frames and types of calibration events (Fig. 1.1).

**Rate Heterogeneity**

Molecular evolution involves the interaction among mutation, selection, random genetic drift, and gene flow. Consequently, rate variation across lineages and the genome are a characteristic of the evolutionary process (Ho 2014). The causes of rate heterogeneity have been broadly divided into gene effects, lineage effects, and residual effects (Ho 2014). Gene effects lead to rate variability among loci, which can be explained by differences in the proportion of functionally constrained sites and by dissimilarities in mutation rates across the genome (Dickerson 1971; Gaut et al. 2011). Lineage effects, on the other hand, refer to factors affecting the entire genome, e.g. differences in life-history traits such as generation time, metabolic rate, body size, body temperature, and DNA repair mechanisms (Mooers & Harvey 1994; Bromham 2002, 2011; Gillooly et al. 2005). Residual effects are the interactions between gene and lineage effects (Ho 2014) and can be triggered by factors such as selection regime and variation in population size, which can explain the pattern and extent of rate variability among lineages (Takahata 1987; Cutler 2000; Ho 2014). A change in biological traits or ecological circumstances can result in such an effect of rate variability among lineages. For example, for (energy-related) mitochondrial genes, flightless insects have higher non-synonymous-to-synonymous substitution ratios than their flight-capable relatives (Mitterboeck & Adamowicz 2013).

Metabolic rate is one of the factors previously associated with rate variability. Elevated metabolic rates are associated with an increase in the production of oxygen radicals, which can increase the mutation rate (Martin & Palumbi 1993). This effect is emphasized in mitochondrial DNA (mtDNA) because the mitochondria consume most of the oxygen in cells; thus, oxidative damage is expected to be higher in mtDNA (Richter et al. 1988; Martin 1995; Bromham 2002; Lee & Wei 2007), explaining why rates of nucleotide substitution are greater in mtDNA than in nuclear DNA (Brown et al. 1979; Martin & Palumbi 1993; Lee & Wei 2007; Lynch et al. 2008; Welch et al. 2008). On the other hand, a comparative study across a wide range of animal taxa and 12 genes did not find evidence of any kind of association between basal metabolic rate and substitution rate (Lanfear et al. 2007). However, it has been suggested that active metabolic rate may be more strongly associated with molecular rates than the basal metabolic rate.
(Santos 2012); thus, additional research is needed to follow up upon those contrasting findings.

The effect of body size on mutation rates (Martin & Palumbi 1993; Mooers & Harvey 1994; Bromham et al. 1996; Bromham 2002) is associated with selection pressure for higher DNA repair efficacy and replication fidelity in organisms with large body size, due to pressure to maintain a large number of cells and cell generations (Bromham et al. 1996). Generation time, which is also correlated with body size, might also influence the rate of molecular evolution because the frequency of replication error in the germ-line genome increases with the frequency of reproduction (Gillman & Wright 2013). Thus, shorter generation time might increase the mutation rate per time unit (Rohde 1992; Gillooly et al. 2005; Thomas et al. 2010). Furthermore, \( N_e \) has been recognized as one of the key factors determining the substitution rate in a population (Ohta 1992; Woolfit & Bromham 2005; Woolfit 2009). \( N_e \) reflects the balance between the role of selection and drift in a population. According to the nearly neutral theory, in populations with small \( N_e \), slightly advantageous mutations are likely to be lost due to drift, decreasing the substitution rate of mutations of that type relative to larger populations; by contrast, slightly deleterious mutations might reach fixation more readily in small populations solely by chance, increasing their substitution rate in small relative to large populations (Woolfit 2009). On the other hand, populations with large effective population sizes have lower substitution rates per generation of slightly deleterious mutations but higher substitution rates of slightly advantageous mutations than populations with a small effective population size (Woolfit 2009; Lanfear et al. 2014). As \( N_e \) increases, genetic drift becomes weaker and natural selection becomes more effective at eliminating slightly deleterious mutations, thus reducing the substitution rate (Woolfit 2009; Lynch 2010, 2011; Akashi et al. 2012; Lanfear et al. 2014). Distinguishing among all these factors to explain rate heterogeneity may therefore be complicated. For example, free radical production and generation time are both correlated with metabolic rate (Martin & Palumbi 1993), which in turn varies according to body size and temperature (Gillooly et al. 2001). Likewise, generation time and metabolic rate are both correlated with body size (e.g. Mooers & Harvey 1994; Bromham et al. 1996).

Additionally, it has been proposed that evolutionary and speciation processes are faster in warmer climate regions (Rohde 1992; Gillman & Wright 2013). For example, in plants and some birds, molecular evolutionary rates are faster in tropical taxa than in related taxa from temperate latitudes or high altitudes (e.g. Bleiweiss 1998b; Gillooly et al. 2005, 2007; Wright et al. 2006). In fishes, there is evidence of slow evolutionary rates in response to the extremely low temperatures of the Southern Ocean; this is better known as the molecular slow-down.
hypothesis (Bargelloni et al. 1994). Furthermore, when comparing absolute rates of evolution between Arctic and tropical marine invertebrates from previous literature, there is an evident signal of disparity in molecular evolutionary rates (Table 1.1). However, rate variability between tropical and temperate taxa has not been supported for other taxa such as crustaceans (Held 2001), terrestrial turtles (Lourenço et al. 2013), lizards and snakes (Rolland et al. 2016), and birds (Bromham & Cardillo 2003). Moreover, comparing relative molecular evolutionary rates across studies is difficult due to discrepancies in methodological approaches. Thus, there is no clear directionality in the pattern of rates between Arctic and tropical taxa, and this remains a topic requiring research attention using consistent methods.

**Synthesis of Molecular Clock Research in Marine Invertebrates**

Marine invertebrates are an important target group in molecular clock studies (Fig. 1.2). For those groups of marine invertebrates with good fossil records (i.e., molluscs), fossil evidence has provided important independent evidence for clock dating (Table 1.1). By contrast, many marine invertebrates have poor fossil records; thus, there are few or no traces of their biological history. In those cases, molecular clocks are particularly important, and biogeographic calibrations are the only method for dating evolutionary events, providing a timescale for those taxonomic groups. Several biogeographic events, such as the closure of the Tethys Sea, the Messinian salinity crisis, the first opening of the Bering Strait, and the formation of the Isthmus of Panama, have been used for calibrating molecular clocks in marine invertebrates. The formation of the Isthmus of Panama is perhaps the most popular biogeographic calibration in the literature, while other geological events have received less attention (Table 1.1).

Molecular clock studies of marine invertebrates are numerous (Table 1.1), but important gaps in knowledge still need to be addressed, such as the proper use of calibration dates. The literature summary presented in Table 1.1 only includes a fraction of the available molecular clock studies, with a focus on those involving marine invertebrates, which are the target taxonomic groups in this thesis. Clock calibrations of marine invertebrates have mostly focused on tropical lineages (e.g. Knowlton & Weigt 1998; Schubart et al. 1998; Lessios et al. 1999; Marko & Moran 2002; Lessios 2008; summary in Table 1.1), while molecular clock research has been less frequent for Northern marine invertebrates (Foltz et al. 2008; Henzler & Ingólfsson 2008; Maggs et al. 2008; Carr 2010). Furthermore, divergence estimates for Northern lineages have commonly been established using calibrations from geographically distant taxa (e.g. Hart et al. 1997; Luttikhuizen et al. 2003; Blanco-Bercial et al. 2011; Boissin et al. 2011; Milligan et al. 2011) or from taxa with different ecological traits (e.g. Hart et al. 1997; Milligan et al. 2011;
Jung et al. 2013). This approach can be problematic, generating incorrect results if the evidence of rate variability among lineages with different traits and inhabiting different environments is not considered (e.g. Bromham 2002; Thomas et al. 2006; Wright et al. 2006; Mitterboeck & Adamowicz 2013). Hence, it remains important to pursue clock calibrations targeted to Northern marine lineages; and in that regard, the opening and re-closure events of the Bering Strait could provide an exceptional resource for clock calibration. In addition, previous clock calibrations in the literature were based on a single sister taxon pair or using the divergence average of a cluster of pairs (further details in Table 1.1), which could be improved by including several sister pairs during the calibration process. Moreover, as mentioned before, many biogeographic events are the result of lengthy or complex processes that require a better consideration when used for clock calibration. Lastly, the rapidly growing global sequence databases of standardized DNA regions (i.e. DNA barcoding; Hebert et al. 2003a) represent a great opportunity for cross-taxon and large-scale studies such as this thesis research.

STUDY QUESTIONS, OBJECTIVES, HYPOTHESIS, AND PREDICTIONS

Three main questions motivate this study. First, can complex geological events, such as the opening/closure of the Bering Strait and the formation of the Isthmus of Panama, be used to calibrate the molecular clock (Chapters II and IV)? Second, are rates of molecular evolution relatively constant among marine invertebrates (Chapters III and IV)? Lastly, do molecular evolutionary rates vary between marine invertebrates inhabiting tropical and Arctic environments (Chapters III and IV)? Thus, the objectives of this research are: i) to determine if complex geological events can be used effectively for clock calibration in marine lineages, ii) to examine rate variability among lineages of marine invertebrates and iii) among taxa inhabiting different latitudinal zones.

Chapter II explores the use of the geological history of the Bering Strait for clock calibration. The first opening of the Bering Strait occurred during the Late Miocene (Marincovich & Gladenkov 1999; Gladenkov et al. 2002; Gladenkov & Gladenkov 2004), followed by repeated opening and closure events due to glacial-interglacial periods (Maslin et al. 1996; Harris 2005). These opening and closure events of the Bering Strait have shaped the distribution and evolutionary history of Northern marine lineages, providing a potential source of absolute dates for clock calibration. However, assigning a single calibration date is problematic due to the complexity of the glacial history. Therefore, in Chapter II, I develop a new method for
biogeographic calibrations—the *iterative calibration approach*. The new methodological approach is successfully tested and applied to calibrating the molecular clock in Northern echinoderms. Rate variability is also explored across all echinoderms used in that chapter.

Chapter III expands on the use of the novel iterative calibration approach (developed in Chapter II) applied to the Bering Strait, to calibrate the molecular clock in Northern marine molluscs, arthropods, and polychaetes. Absolute rates of molecular evolution among those three major groups of marine animals are then compared with echinoderms from Chapter II.

Chapter IV expands on the use of the iterative calibration approach by applying this method to the Panama system. I reanalyze public genetic divergence values to calibrate the molecular clock in tropical marine echinoderms, molluscs, and arthropods. This chapter first explores the efficacy of the iterative calibration approach when using the formation of the Isthmus of Panama to calibrate the molecular clock. Then, absolute rates of molecular evolution among the three phyla are compared with previous literature on tropical marine invertebrates (Table 1.1) and with the results from the Arctic taxa analyzed in Chapters II and III.

*Hypothesis and Prediction.*—I hypothesize that molecular rates in marine invertebrates vary along the latitudinal and temperature gradient and that substitution rates will be higher in tropical lineages than in their Arctic counterparts due to several biological and ecological traits.

**SIGNIFICANCE**

This thesis represents a novel and significant contribution to the field of molecular clock research, first by developing a new methodology for clock dating—the iterative calibration approach—and by filling important gaps in molecular clock research, especially for marine invertebrates. Chapters II–IV will provide a new context for dating the evolutionary history of marine lineages and will advance our understanding of the impact of prior climate change events upon the diversification of life. Chapter V will present the overall findings and their implications in the field of molecular evolution.
### Table 1.1

Summary of molecular clock calibrations for the mitochondrial cytochrome c oxidase subunit I gene (COI) in marine invertebrates. Abbreviations: Ma- million years ago; yr- years; Ka- thousand years ago; 5P- 5’ end of COI gene; 3P-3’ end of COI gene; W- whole COI gene; MY- million years. Citations are per article from ISI Web of Knowledge (last accessed July 2nd, 2017).

<table>
<thead>
<tr>
<th>Calibration Event</th>
<th>Assigned Age of Event</th>
<th>Number of Sister Pairs or Nodes Dated</th>
<th>Assumptions</th>
<th>Region of COI</th>
<th>Taxonomic Group</th>
<th>Divergence Rate Estimate (in %/MY)</th>
<th>Model of Molecular Evolution</th>
<th>Reference</th>
<th>No. of Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3 Ma</td>
<td>1/15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Mangrove species pair diverged at the end of the Isthmus closure</td>
<td>3P</td>
<td>Arthropoda</td>
<td>1.4%</td>
<td>K2P+G</td>
<td>(Knowlton &amp; Weigt 1998)</td>
<td>158/548&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3.1 Ma</td>
<td>1/8&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Mangrove species pair diverged at the end of the Isthmus closure</td>
<td>3P</td>
<td>Arthropoda</td>
<td>1.7%</td>
<td>K2P</td>
<td>(Williams &amp; Knowlton 2001)</td>
<td>126</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>4 Ma</td>
<td>1</td>
<td></td>
<td>3P</td>
<td>Arthropoda</td>
<td>3%</td>
<td>Only transversions</td>
<td>(Baldwin &lt;i&gt;et al.&lt;/i&gt; 1998)</td>
<td>112</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3 Ma</td>
<td>1</td>
<td></td>
<td>5P</td>
<td>Mollusca</td>
<td>2.4%</td>
<td>K2P</td>
<td>(Hellberg &amp; Vacquier 1999)</td>
<td>123</td>
</tr>
<tr>
<td>Event</td>
<td>Age (Ma)</td>
<td>Method</td>
<td>Taxon</td>
<td>Divergence</td>
<td>Model</td>
<td>Reference</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3.1</td>
<td>1/5d</td>
<td>Isolation of the pair with the smallest divergence at 3.1MY</td>
<td>5P</td>
<td>Mollusca</td>
<td>5.1%</td>
<td>TN93+G</td>
<td>(Marko &amp; Moran 2002)</td>
<td>32</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3.5</td>
<td>1</td>
<td></td>
<td>5P</td>
<td>Echinodermata</td>
<td>5%</td>
<td>K2P</td>
<td>(Hart et al. 1997)</td>
<td>135</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3.1</td>
<td>1</td>
<td></td>
<td>5P</td>
<td>Echinodermata</td>
<td>3.1%</td>
<td>K2P</td>
<td>(Lessios et al. 1999)</td>
<td>153</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3</td>
<td>1/3e</td>
<td>Isolation of the pair with the smallest divergence at 3MY</td>
<td>5P</td>
<td>Arthropoda</td>
<td>1.4% (for combined COI and 16S)</td>
<td>K2P</td>
<td>(Morrison et al. 2004)</td>
<td>74</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3</td>
<td>1/4'</td>
<td>The least-divergent pair diverged at 3MY</td>
<td>3P</td>
<td>Arthropoda</td>
<td>3.6%</td>
<td>GTR+G+I</td>
<td>(Mathews &amp; Anker 2009)</td>
<td>19</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>2.8–3.1 Ma</td>
<td>1/3e</td>
<td>The least-divergent pair diverged at 2.8-3.1MY</td>
<td>5P</td>
<td>Mollusca</td>
<td>2.1-3.1%</td>
<td>K2P</td>
<td>(Miura et al. 2010)</td>
<td>24</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3</td>
<td>8</td>
<td></td>
<td>3P</td>
<td>Echinodermata</td>
<td>1.95%</td>
<td>K2P</td>
<td>(Hickerson et al. 2003)</td>
<td>21</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3</td>
<td>15</td>
<td></td>
<td>3P</td>
<td>Arthropoda</td>
<td>1.8%</td>
<td>K2P</td>
<td>(Hickerson et al. 2003)</td>
<td>21</td>
</tr>
<tr>
<td>Event Description</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Isthmus of Panama</td>
<td>3.1 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Tethys Sea</td>
<td>19 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Isthmus of Panama and opening of the Bering Strait</td>
<td>3 and 3.5 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opening of the Bering Strait</td>
<td>3.5 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Isthmus of Panama</td>
<td>1.8 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Isthmus of Panama</td>
<td>3.1 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure of the Isthmus of Panama</td>
<td>2.8 Ma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Time</th>
<th>Methodology</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Schubart et al. 1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Williams &amp; Reid 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Luttikhuiizen et al. 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Foltz et al. 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Lessios et al. 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(McCartney et al. 2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Lessios 2008)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3P</td>
<td>Arthropoda</td>
</tr>
<tr>
<td>5P</td>
<td>Mollusca</td>
</tr>
<tr>
<td>5P</td>
<td>Mollusca</td>
</tr>
<tr>
<td>5P</td>
<td>Echinodermata</td>
</tr>
<tr>
<td>3P</td>
<td>Echinodermata</td>
</tr>
<tr>
<td>3P</td>
<td>Echinodermata</td>
</tr>
<tr>
<td>3P</td>
<td>Echinodermata</td>
</tr>
<tr>
<td>Event</td>
<td>Age</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>3 Ma</td>
</tr>
<tr>
<td>Opening of the Bering Strait</td>
<td>3.5 Ma</td>
</tr>
<tr>
<td>Mediterranean salinity crisis</td>
<td>5.64 Ma</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>2.75 Ma</td>
</tr>
<tr>
<td>Closure of Isthmus of Panama</td>
<td>2.75 Ma</td>
</tr>
</tbody>
</table>
## Closure of Isthmus of Panama

<table>
<thead>
<tr>
<th>Event</th>
<th>Age (Ma)</th>
<th>Distance (km)</th>
<th>Method</th>
<th>Calibration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure of Isthmus of</td>
<td>2.75</td>
<td>2</td>
<td>3P</td>
<td>Arthropoda</td>
<td>(Wilke et al. 2009)</td>
</tr>
<tr>
<td>Panama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0% (JC), 2.1% (K2P),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4% (F81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.2% (HKY), 3.2% (HKY+G+I)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0% (GTR), 3.2% (GTR+G+I)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Formation of Islands

<table>
<thead>
<tr>
<th>Island</th>
<th>Age (yr)</th>
<th>Method</th>
<th>Calibration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawaiian Islands</td>
<td>100,000</td>
<td>5P</td>
<td>Anchialine Arthropoda</td>
<td>(Craft et al. 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20%</td>
<td>P-distance</td>
</tr>
</tbody>
</table>

## Fossil Calibrations

<table>
<thead>
<tr>
<th>Event</th>
<th>Age (Ma)</th>
<th>Method</th>
<th>Calibration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 &amp; C2: Old</td>
<td>83-118</td>
<td>5P</td>
<td>Mollusca</td>
<td>(Marko 2002)</td>
</tr>
<tr>
<td>representative</td>
<td>Ma (C1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of subfamilies Noetiidae</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and Anadariae, C3:</td>
<td>79-89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First appearance of</td>
<td>Ma (C2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subgenera Anadara and</td>
<td>16-23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandiarca</td>
<td>Ma (C3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>calibrations (1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>calibration used at a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Fossil records from *Mytilus* and *Tellinacea*

<table>
<thead>
<tr>
<th>Age (Ma)</th>
<th>N</th>
<th>Model</th>
<th>Rate (K2P)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>30–251</td>
<td>5P</td>
<td>Mollusca</td>
<td>0.16–0.52%</td>
<td>(Luttikhuizen et al. 2003)</td>
</tr>
<tr>
<td>2–23</td>
<td>7</td>
<td>Mollusca</td>
<td>1.3%</td>
<td>(Marko et al. 2014)</td>
</tr>
<tr>
<td>41</td>
<td>1&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Mollusca</td>
<td>2.6%</td>
<td>(Williams &amp; Reid 2004)</td>
</tr>
</tbody>
</table>

### Coalescent model calibrations

<table>
<thead>
<tr>
<th>Event</th>
<th>Age (ka)</th>
<th>N</th>
<th>Model</th>
<th>Rate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea level rise into the Sunda Shelf</td>
<td>14.58–19.6</td>
<td>1&lt;sup&gt;k&lt;/sup&gt;</td>
<td>Unavailable</td>
<td>6.58% (A), 2.3% (M), 2.6% (E)</td>
<td>(Crandall et al. 2012)</td>
</tr>
<tr>
<td>Messinian Salinity Crisis</td>
<td>5.33–5.59</td>
<td>1</td>
<td>5P</td>
<td>Arthropoda</td>
<td>7.72%</td>
</tr>
<tr>
<td>Opening of the Bering Strait</td>
<td>3.5</td>
<td>1&lt;sup&gt;k&lt;/sup&gt;</td>
<td>5P</td>
<td>Echinodermata (E), Mollusca (M) and Arthropoda (A)</td>
<td>9.68% (E), 8.86% (M), 5.52% (A)</td>
</tr>
</tbody>
</table>

<sup>a</sup>One pair used for dating; 15 pairs considered in total, which diverged across a range of dates.<br>
<sup>b</sup>This study builds upon the concept and data used by Knowlton et al. (1993). Thus, 548 citations are the sum of citations for both studies (Knowlton et al. 1993; Knowlton & Weigt 1998), and 158 citations correspond to the study by Knowlton and Weigt (1998).<br>
<sup>c</sup>One pair used for dating; 8 pairs considered in total, which diverged across a range of dates. Sister pairs used in this study are a subset of the data used by Knowlton and Weigt (1998).<br>
<sup>d</sup>One pair used for dating; 5 pairs considered in total, which diverged across a range of dates.<br>
<sup>e</sup>One pair used for dating; 3 pairs considered in total, which diverged across a range of dates.<br>
<sup>f</sup>One pair used for dating; 4 pairs considered in total, which diverged across a range of dates.<br>
<sup>g</sup>Divergence rates were estimated for two clades, one with a rate of 1.66% and a second clade with a rate of 2.33%.
The first node in the *Echinolittorina* clade was dated to 41MY. Independently, one node was calibrated using the closure of the Tethys Sea approximately 19MY. One sister pair was also calibrated with the closure of the Isthmus of Panama; however, the biogeographic calibration was later discarded arguing that the rate based on the fossil calibration was closer to published rates from other molluscs with good fossil record (Williams & Reid 2004).

Four pairs used for dating; 20 pairs considered in total, which diverged across a range of dates.

From the 15 pairs suggested by Knowlton and Weigt (1998), only the seven most closely related pairs were used for rate estimation in Wilke *et al.* (2009).

One pair used for dating in each group included in the study (i.e. Echinodermata, Mollusca, and Arthropoda).
Figure 1.1 Divergence rates for marine invertebrates from the literature review presented in Table 1.1 and according to the date used for calibration.
Figure 1.2 Divergence rates for marine invertebrates estimated using calibration dates between 1–20 Ma from the literature review presented in Table 1.1.
CHAPTER II
Iterative Calibration: a Novel Approach for Calibrating the Molecular Clock Using Complex Geological Events

ABSTRACT

During the past 50 years, the molecular clock has become one of the main tools for providing a time scale for the history of life. In the era of robust molecular evolutionary analysis, clock calibration is still one of the most basic steps needing attention. When fossil records are limited, well-dated geological events are the main resource for calibration. However, biogeographic calibrations have often been used in a simplistic manner, for example assuming simultaneous divergence of sister lineages after a vicariance event. In this study, I propose a novel iterative calibration approach to define the most appropriate calibration date by seeking congruence between the dates assigned to multiple allopatric divergences and the complete geological history of the event used for calibration. Exploring patterns of molecular divergence in 16 trans-Bering sister clades of echinoderms, I demonstrate that the iterative calibration is predominantly advantageous when using complex geological or climatological events—such as the opening/reclosure events of the Bering Strait—providing a powerful tool for clock calibration that can be applied to other biogeographic calibration systems and further taxa. The results reveal a large range of genetic divergences, consistent with multiple pulses of trans-Bering migrations. Using Bayesian analysis, I observed that evolutionary rate variability in the mitochondrial cytochrome c oxidase subunit I gene is generally distributed in a clock-like fashion for Northern echinoderms. A resulting rate of 2.8% pairwise Kimura-2-parameter sequence divergence per million years is suggested for the COI DNA barcode region in Northern echinoderms. Given that molecular rates may vary across latitudes, this study provides a new context for dating the evolutionary history of Arctic marine life.
INTRODUCTION

The molecular clock is an important and valuable tool in evolutionary biology, although it has remained controversial since it was first proposed 50 years ago (Zuckerkandl & Pauling 1965a). In addition to concerns due to rate heterogeneity across taxa (e.g. Martin et al. 1992; Martin & Palumbi 1993; Mooers & Harvey 1994; Bleiweiss 1998a; Bromham 2002), calibrating the clock is also a challenge, especially when fossil evidence is poor or absent. For taxa without a fossil record, geological or climatological events represent the only calibration points to estimate rates of molecular divergence and to provide a time scale for the history of life. Yet, biogeographic calibrations are frequently used with simplistic methods and without considering the complexity of some geological events, as discussed below.

Well-dated geological events, such as island formation (e.g. Fleischer et al. 1998; Weir & Schluter 2008), paleotectonics (Landis 2016), or vicariance events—for example the closure of the Tethys Sea or the formation of the Isthmus of Panama (e.g. Held 2001; Williams & Reid 2004)—have been extensively used for clock calibrations. For tropical marine taxa, the most-used event has been the formation of the Isthmus of Panama (e.g. Knowlton & Weigt 1998; Lessios 2008; Hurt et al. 2009), with evolutionary rates calibrated using genetic divergences between sister species separated by the closure of the isthmus, which are commonly known as “geminate” species (Jordan 1908).

Molecular clocks for Northern marine taxa have been largely calibrated using fossils associated with the major trans-Arctic interchange 3.5 million years ago (Ma) (e.g. Palumbi & Kessing 1991; Collins et al. 1996b; Kober & Bernardi 2013). In addition, for Arctic taxa lacking a good fossil record, divergence times have been estimated using calibrations generated using tropical taxa (e.g. Luttikhuizen et al. 2003; Boissin et al. 2011). However, this approach has limitations. First, the formation of the Isthmus was a slow process (Coates & Obando 1996; Bacon et al. 2015; O’Dea et al. 2016), making divergence dating problematic. Additionally, the widely-used calibration from Panama ultimately relied on a single data point (Knowlton & Weigt 1998) and thus did not address possible rate heterogeneity among lineages. Moreover, using clock calibrations from other geographical regions could be problematic, especially as there is evidence that molecular evolutionary rates vary systematically in association with the environment (e.g. Bleiweiss 1998b; Gillooly et al. 2005; Wright et al. 2006). Therefore, further research is needed to investigate appropriate calibrations for Northern lineages.
Using the Bering Strait for Clock Calibration in Northern Marine Taxa

The geological history of the Bering Strait (Fig. 2.1) provides a unique, though complex, opportunity for calibrating molecular divergence rates for Northern taxa. North Pacific and Arctic-Atlantic marine faunas evolved separately since the middle Cretaceous (80-100 Ma) (Dunton 1992; Marincovich 1993) until the first opening of the Bering Strait. This opportunity for dispersal into a novel marine biogeographic region, followed by divergence in allopatry, is the first of a series of events that provide a strong opportunity for molecular clock calibration for Northern marine organisms.

Early fossil evidence of diatoms and of the mollusc Astarte suggest the first opening of the Bering Strait occurred between the Late Miocene and Early Pliocene approximately 4.8-7.4 Ma (Marincovich & Gladenkov 1999). However, with more recent fossil evidence of both diatom and bivalve fossils, the time of the first opening has been narrowed to 5.4-5.5 Ma, near the end of the Miocene (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004). However, according to diatom fossils from Northern Eurasia, temporary connections between Pacific and Arctic Oceans might have been existed across Eastern Siberia since the Early Miocene; thus, temporary openings of the Strait have been proposed to have occurred since approximately 17 Ma (Polyakova 2001). However, the inference of an open Strait since the Early Miocene is not fully accepted in the literature; instead, it has been strongly questioned using the argument that it is based on weak evidence and poorly preserved diatom fossils (see details in Marincovich & Gladenkov 2001 and Gladenkov & Gladenkov 2004), and thus should be considered only with great caution (Marincovich & Gladenkov 2001; Gladenkov & Gladenkov 2004). It is currently well accepted that the Bering Strait first opened 5.4-5.5 Ma (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004).

The first opening of the Bering Strait allowed marine migrations between the Pacific and Arctic-Atlantic Oceans and is one of the most important oceanographic (Shaffer & Bendtsen 1994; Hu et al. 2007) and biogeographical events shaping the Northern marine biota (Durham & MacNeil 1967; Vermeij 1991). While the Strait opened 5.4-5.5 Ma (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004), the major fossil evidence of the main trans-Arctic marine migration is dated at 3.5 Ma (Briggs 1970; Vermeij 1989a, 1991; Marincovich 2000; Dodson et al. 2007). The major trans-Arctic marine migration might have occurred in response to a major change in ocean circulation in the Northern Hemisphere associated with the formation of the Isthmus of Panama (Marincovich 2000; Matthiessen et al. 2009). Shortly after, approximately 3.4 Ma, a cooling transition period began in the Arctic (Shackleton & Opdyke 1977; Herman &
Hopkins 1980; Horikawa et al. 2015), culminating with extensive glaciation of the Northern hemisphere from 3 to 2.4 Ma (Einarsson et al. 1967; Schrader et al. 1976; Herman and Hopkins 1980; Maslin et al. 1996; Haug et al. 1999; Harris 2005; Horikawa et al. 2015). Glacial formations and the associated drop in sea level, of at least 50 meters, triggered the closure of the Bering Strait (Hopkins 1967), resulting in the re-isolation of the Pacific and Arctic-Atlantic marine biotas.

Glacial events were more frequent during the Quaternary period (Dunton 1992; Maslin et al. 1996; Haug et al. 1999), with ca. 9 successive glacial-interglacial cycles (Einarsson et al. 1967; Maslin et al. 1996; Haug et al. 1999; Harris 2005) causing the Bering Strait to open and close on several occasions. During interglacial periods, the sea level rose at least 20 metres and as much as 100 metres, suggesting an open Strait (Hopkins 1967). During glacial intervals, the reduction in sea level would have resulted in a continuous land connection between Alaska and Siberia, commonly known as the Bering Land Bridge (Hopkins 1967). Glacial-interglacial cycles during the Quaternary strongly influenced sea levels, habitat conditions, and the current distribution of Northern marine organisms. The most recent opening of the Bering Strait is dated to 15,000 years ago (Dodson et al. 2007; Hardy et al. 2011), allowing the contemporary exchange between Pacific and Arctic-Atlantic marine biotas.

The Bering Strait has been previously used for clock calibration in sea stars (Leptasterias) (Foltz et al. 2008) and polychaetes (Carr 2010). In polychaetes, divergence times for the barcode region (Hebert et al. 2003a) of the mitochondrial cytochrome c oxidase subunit I (COI-5P) gene were estimated in 20 trans-Arctic species pairs using Kimura-2-parameter (K2P; Kimura 1980) distances and assuming that the first wave of dispersal across the Bering Strait took place 3.5 Ma during the major marine migrations, resulting in a rate estimate of 4.4% pairwise sequence divergence per million years (MY). The results also suggested that sister lineages did not diverge simultaneously, but rather exhibited multiple migration episodes (Carr 2010). In sea stars, a rate was calibrated for a nuclear gene region (Elongation factor 1α subunit intron 4) and two mitochondrial regions (control region and COI-5P); a single trans-Arctic sister pair was used for calibration assuming divergence at 3.5 Ma, resulting in a rate estimate of 1.8% pairwise divergence per MY for COI-5P in Leptasterias (Foltz et al. 2008).

Here, I argue that the key to using the Bering Strait for calibration is to consider multiple potential episodes of migration followed by divergence rather than enforcing one single calibration point, which requires the tenuous assumption that all species migrated at the same time during the major trans-Arctic marine interchange. In this study I propose a novel iterative
calibration approach for clock dating and demonstrate its advantage when using complex geological events, such as the glacial history of the Bering Strait. I also explore patterns of molecular divergence in the COI-5P barcode region of trans-Bering sister clades of echinoderms inhabiting the North Pacific vs. Arctic-Atlantic ocean regions and generate a new molecular rate estimate for Northern echinoderms.

METHODS

I used publicly available DNA sequences of the COI-5P barcode region (Hebert et al. 2003b) from echinoderms, a taxonomic group that has been extensively used in tropical marine calibration studies (e.g. Lessios 1979, 2008; Hart et al. 1997; McCartney et al. 2000; Coppard et al. 2013). I explored the evolutionary rate of the COI-5P gene and focused explicitly on finding proper calibration dates. It is beyond the scope of this study to reconstruct phylogenies as species tree hypotheses. Despite the difficulties of using only one gene, I believe that COI represents the best choice across mitochondrial genes to answer the research questions based on its greater range of phylogenetic signal, compared with other mitochondrial genes (Hebert et al. 2003a). Also, this marker was favoured due to its patterns of divergence within and between species, rapid enough to allow the discrimination of phylogeographic groups (e.g. Wares & Cunningham 2001; Hebert et al. 2003b; Bastrop & Blank 2006; Bleidorn et al. 2006; Radulovici et al. 2010). Moreover, the extensive and fast-growing availability of public sequences and usage of this standardized marker across animal groups provides the ability to explore substitution patterns in the same genomic region for an exhaustive range of taxonomic groups, which represents a great opportunity for large-scale and cross-taxon research such as this study.

Finding Potential Trans-Bering Sister Clades and Data Collection

Using the phenogram-building tool available in the Barcode of Life Data Systems v3 (BOLD; http://www.boldsystems.org/; last accessed July 2015) (Ratnasingham & Hebert 2007), preliminary neighbour-joining (NJ) trees (Saitou & Nei 1987) were constructed for each of the five classes within the phylum Echinodermata using K2P distances and publicly accessible data, excluding sequences with stop codons, those flagged within BOLD as containing possible misidentifications or contaminants, and sequences shorter than 200 base pairs (bp). The trees were colour coded by Barcode Index Number (BIN) (Ratnasingham & Hebert 2013), and information about Country/Province of collection and sequence length were included in the tip label. BINs were used as taxonomic units since it has been demonstrated that barcodes perform
well to discriminate and delineate echinoderm species (e.g. Ward et al. 2008; Bribiesca-Contreras et al. 2013; Layton et al. 2016). Allopatric groups were recognized, and then I pinpointed potential trans-Bering sister clades, i.e. those having one BIN distributed only in the North Pacific and the closest related BIN distributed in the Arctic-Atlantic. Allopatric sister clades where one or both sisters consisted of multiple BINs were also considered. If one of the potential sister taxa was not yet identified to genus level, I still considered this as a target taxon to be included for further analysis. On the other hand, I also considered allopatric genetic clusters within a single Linnaean species name as target taxa. Distribution of the potential trans-Bering clades was verified by searching for the species name, when available, in the Encyclopedia of Life (EOL; http://www.eol.org last accessed August 2015) and the World Register of Marine Species (WoRMS; http://www.marinespecies.org last accessed August 2015). Trans-Bering pairs of echinoderms will be referred to as sister clades. The NJ trees for each class were used as preliminary trees to select the target genera for sequence data retrieval and further analysis, and to select the outgroup for each of the data sets based upon the next closest related lineage to each genus containing candidate sister clades.

**Trans-Bering Sister Clades Confirmation through Maximum Likelihood Phylogenetic Analysis**

When possible, previously published molecular phylogenies from the taxa of interest were used to confirm or refute sister clade relationships (Arndt et al. 1996; Lee 2003; Foltz et al. 2008; Mah & Foltz 2011; Kober & Bernardi 2013), giving preference to mitochondrial trees as I am most interested in accurate gene trees rather than species trees for calibration purposes. In the absence of published multi-gene trees, phylogenetic relationships were then reconstructed using a maximum likelihood (ML) tree for each genus containing a potential pair.

From April 18th to July 22nd, 2015, publicly available COI sequences from each genus containing potential trans-Bering sister clades were retrieved from BOLD v3, including all sequences within BINs that fell in the candidate sister clades, but excluding sequences with stop codons, those flagged as containing possible misidentifications, and sequences shorter than 200 bp. I then created genus-level datasets to be used in the following analysis. Each genus dataset, including its assigned outgroup sequence, was aligned in MEGA 6 (Tamura et al. 2013); then the alignment was visually checked for gaps and translated to amino acids to verify the reading frame and absence of stop codons. Even though the COI region used for DNA barcoding is standardized, COI sequences retrieved for echinoderms did not overlap completely due to different sequence length. As a result, each genus dataset was trimmed to a different alignment length (615-923 bp) (Supplementary Material 2.1). After inspecting each alignment I
eliminated problematic sequences when >1% of unknown bases were present, when potential misidentification cases were noticeable on the NJ tree, or when a base-calling error was suspected and trace files were not available (Supplementary Material 2.2). In addition, identical sequences were eliminated using the tool Duplicate Finder (http://bioinfotutlets.blogspot.ca/2009/09/duplicates-finder-java-standalone.html; last accessed September 2015). For each genus dataset, the best-fit molecular evolutionary substitution model was obtained according to the Bayesian Information Criterion (BIC) using MEGA 6. Phylogenetic relationships were then reconstructed in MEGA 6 using ML with 1000 bootstrap replicates. Trans-Bering sister clade relationships were supported for 16 pairs, displaying reciprocal monophyly on both the NJ and ML trees. BOLD process IDs and GenBank accession numbers for trans-Bering sister clades sequences can be found in Supplementary Material 2.3.

**Rate Constancy Test**

A Tajima’s relative rate test (Tajima 1993) was used to test for rate constancy within each of the 16 sister clades using MEGA 6. This test involved comparing relative rates between sequences of sister taxa using a third taxon as a reference point or outgroup. I selected the longest sequence per species; when multiple sequences with the same length were available, I randomly selected one sequence per species for the comparisons. The outgroups were designated based on the ML trees, selecting the closest related taxon to the sister clade of interest (Robinson et al. 1998).

A separate relative rate test was performed for each sister clade. The relative test was repeated for the two sister clades with poor bootstrap support (<70%) using a more distantly-related outgroup (further details in Supplementary Material 2.4).

**Kimura 2-Parameter Distances between Trans-Bering Sister Clades**

Using the full available sequence length for each genus-level dataset, average K2P genetic distances between each trans-Bering sister clade were calculated in MEGA 6, with pairwise deletion of missing nucleotides. In addition to enabling comparison with prior studies that used K2P distances (e.g. Lessios 2008; Carr 2010; Miura et al. 2010; Carr et al. 2011), a simple distance metric is expected to perform well at estimating evolutionary distances for shallow divergences (Hebert et al. 2003a). However, a drawback of this approach is that it would overlook systematic rate variability across the phylogeny, and taxa with anomalous evolutionary rates may have undue influence upon the final calibration. Another drawback is that evolutionary divergences for deeply divergent sister pairs might be underestimated due to
saturation. To address these limitations, I have complemented the use of K2P distances with a whole-tree approach for clock calibration (see below).

For clock calibration, I first mimic a traditional and simplistic practice where an average divergence of sister species pairs is calibrated using the most accepted single date for the vicariant event of interest (e.g. Knowlton & Weigt 1998; McCartney et al. 2000; Carr 2010; Hoareau et al. 2013). For the traditional clock calibration, only strict sister species pairs were used, and I calculated an average of divergences from the cluster of points with the most similar distances, which was visually selected. Using the average sequence divergence from the cluster of the most similar distances consisting of four sister species pairs, the divergence rate ($r$) was estimated as $r = D/T$, where $D$ is the average percentage of K2P distance and $T$ is the time of divergence assuming that the cluster of sister pairs with the most similar divergences migrated during the major marine trans-Arctic interchange 3.5 Ma (further details in Supplementary Material 2.5). K2P distances were also used for clock calibration using the iterative approach (see below).

**Generation of Relative Node Ages through Bayesian Analysis**

Relative node ages were also estimated using an ultrametric tree generated using a Bayesian approach (Drummond et al. 2012), applying the best-fit model of sequence evolution and smoothing rate heterogeneity across the tree. For computational efficiency, only unique sequences of each member of the 16 trans-Bering sister pairs were included. The dataset including 247 sequences was aligned and trimmed to a length of 459bp (Supplementary Material 2.6). The best-fit substitution model for this dataset was GTR+G+I (Generalised Time Reversible with gamma-distributed rate heterogeneity among sites and a proportion of invariant sites parameter) according to the lowest BIC score obtained in MEGA 6.

An ultrametricized Bayesian inference tree was constructed in BEAST v1.8.2 (Drummond et al. 2012) employing a Birth-Death speciation model (Yang & Rannala 1997) tree prior and an uncorrelated relaxed lognormal clock model. The BEAUTi v1.8.2 (Drummond et al. 2012) interface was used to generate the input XML files using the analysis settings that are detailed in the Supplementary Material 2.7. The only trans-Bering sister pair representing the echinoderm class that is most distantly related from all others (Crinoidea) (Littlewood et al. 1997; Perseke et al. 2010; Telford et al. 2014) was used as the outgroup; the remaining 15 pairs were constrained as the ingroup. Reciprocal monophyly of the outgroup, ingroup, and each of the 16 sister clades, 15 of which represented different genera, was assumed a priori for computational efficiency. Six independent Markov chain Monte Carlo (MCMC) analyses were
each run with a MCMC length of 100,000,000 generations, sampling every 10,000 with a final sample size of 10,000 trees. Adequate sampling (ESS>100) for all parameters and convergence were assessed in Tracer v1.6 (Rambaut et al. 2014). Due to low support for deeper nodes and independent MCMC analysis showing different tree topologies, it was not possible to combine two runs and their trees as suggested by the software authors, due to negative resulting branch lengths (Drummond et al. 2012). Within the phylum Echinodermata, relationships among classes, orders, and families remain controversial (e.g. Littlewood et al. 1997; Kamarudin et al. 2010; Perseke et al. 2010; Mah & Foltz 2011; Telford et al. 2014; Feuda & Smith 2015); thus, it was not feasible to constrain the full tree topology. Therefore, six independent runs and trees were analysed as a way to test for replicability of the calibrations. As results were similar across runs, I selected the run with the highest Effective Sample Size (ESS) values for detailed reporting of results. The final tree was visualized in FigTree v1.4.2 (Rambaut 2014), and all the branch tips were collapsed by allopatric taxa for visualization purposes. Branch lengths were given as relative time, and the posterior molecular evolutionary rate was mapped onto the tree. Relative node ages were then used for clock calibration using the iterative approach (see below).

*Iterative Calibration*

The iterative calibration approach involves a sequential process of clock calibration with the aim of finding congruence between the times assigned to allopatric divergences in relation to the geological/climatological event being used for calibration; this study uses the example of the Bering Strait to demonstrate this approach (Fig. 2.2). First, I applied the iterative approach for clock calibration using K2P distances and next using the relative node ages from the ultrametricized gene tree generated via Bayesian analysis.

To begin, a starting calibration date was assigned to the node that joins the most divergent pair of sister lineages (reference node), representing the most recent common ancestor preceding the dispersal event. I started with the earliest possible opening of the Bering Strait that has been suggested and well supported to date in the literature according to stratigraphic evidence and fossils from diatoms and molluscs (7 Ma; Marincovich and Gladenkov 1999) despite more recent evidence that the first opening has been narrowed to a date of 5.4-5.5 Ma (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004). Subsequently, I manually assigned the time of divergence for the remaining 15 pairs in relation to the most divergent sister node. Next, I graphed and compared the divergence time estimates from all 16 trans-Bering sister clades against the geological history of the Bering Strait. Geological
evidence and sea-level data available from the literature (Fig. 2.1) (e.g. Vermeij 1991; Dunton 1992; Haug et al. 1999; Gladenkov et al. 2002; Gladenkov & Gladenkov 2004; Harris 2005; Horikawa et al. 2015) were used to elucidate the timeframes when marine trans-Bering migrations were feasible vs. improbable. I gave special attention to the two periods of time when trans-Bering migrations were extremely unlikely: preceding the first opening of the Strait dated at 5.5 Ma and during the maximum glacial period (2.4 to 3 Ma). Calibration attempts that generated trans-Bering divergences within these time periods are considered highly improbable, as trans-Bering migration (followed by allopatric divergence) could not occur when the Strait was closed. Evidence from the major trans-Arctic marine interchange and the repeated glacial-interglacial cycles during the Quaternary provided additional information when looking for the most plausible calibration. If the genetic divergences and geological history of the Bering Strait were not concordant, then I proceeded to re-assign a new slightly lower date (stepping by 0.1 MY) to the most divergent sister clade and all other nodes proportionally; again, I compared the resulting date estimates with the geological history. The process was repeated until finding concordance between node divergences and the Bering Strait geological time scale. Using the same principles, 47 calibration dates were tested in the range between 2 to 7 Ma, using K2P distances and relative node ages from the ultrametricized gene tree independently. When two or more calibrations were concordant with the geological events, independent evidence (e.g. fossil evidence from the major trans-Arctic interchange) was used to select the best possible scenario and therefore the most probable calibration date. Lastly, the K2P divergence rate was estimated using the most probable calibration date. The divergence rate (r) for the most divergent trans-Bering sister clade was estimated as \( r = \frac{D}{T} \), where \( D \) is the percentage of K2P distance and \( T \) is the time of divergence according to the best calibration from the iterative calibration approach considering all pairs.

Additionally, biological traits from all trans-Bering sisters were investigated when available. Deep (>50 m) vs. shallow distribution (≤50 m), habitat zone during adult stage (benthic vs. pelagic), thermal affinity (polar, temperate, or tropical), and developmental model (planktonic larvae vs. direct development) were the target of attention. Specific traits might have allowed alternative dispersal routes, and the phylogenetic events may not be truly associated with the geological history of the Bering Strait. For example, taxa inhabiting deep waters are less likely to disperse during a shallow open Strait than taxa from the intertidal, especially as the Bering strait is approximately 50 meters deep (Hopkins 1967). Likewise, taxa with a wide range of temperature tolerance are more likely to have alternative migration routes through tropical pathways (e.g. before the closure of the Isthmus of Panama) than taxa that exclusively inhabit
polar waters. After integrating all information sources (e.g. biological traits), the resulting date estimates were compared to the geological history, with the traits graphically visualized. Information about biological traits was primarily obtained from online databases such as WoRMS and SeaLifeBase (http://www.sealifebase.org; last accessed May 4th, 2017).

RESULTS

Trans-Bering Sister Clades

After examining the preliminary NJ COI gene trees generated in BOLD for each of the five classes of echinoderms, 16 potential trans-Bering sister clades were identified (Supplementary Material 2.8). Previous molecular phylogenies were not available in most cases; therefore, sister clade relationships were confirmed based on the genus-level ML phylogenetic trees. Fourteen of the candidate sister clades were supported with high bootstrap support (>76%), while support for one pair (Asterias) was only 47%, although this relationship was also supported in previous studies (Mah & Foltz 2011). Support for the other sister clade (Ophiopholis) was 69%. The sixteen sister clades were phylogenetically diverse, falling within 15 genera from four classes of echinoderms: seven sister clades from class Asteroidea, four sister clades from each of Holothuroidea and Ophiuroidea, and one more pair from Crinoidea. However, trans-Bering pairs from the class Echinoidea were not detected in the available data. Thirteen trans-Bering sister clades were comprised of genetically divergent clusters and are likely to be separate species according to the BIN system (Ratnasingham & Hebert 2013). Sequence divergence between the remaining three sister clades was low (<2% average pairwise K2P distance), and they are likely to be populations within species. Detailed information about genus-level datasets and sequence lengths can be found in Supplementary Material 2.1.

Relative Rate Test on Trans-Bering Lineages

The molecular clock hypothesis was not rejected for 15 of the 16 pairs (Supplementary Material 2.4). Although relative rates significantly differed for Pseudostichopus tuberosus and P. aemulatus ($\chi^2 = 8.53; p < 0.005$), this pair was first included in the following analysis after accounting for rate variability. The calibration was then repeated after omitting this one pair.

Kimura 2-Parameter Distance Analysis and Clock Calibration

Mean COI K2P distances between trans-Bering sister clades ranged from 0.5% to 15.5% divergence (Fig. 2.3). Both small and large K2P distances were observed within multiple echinoderm classes. A cluster of four sister pairs with the most similar divergences (8.4-11.4%...
K2P) was observed. The remaining pairs showed a much lower or higher genetic divergence or consisted of a sister lineage made up of two or more BINs (Supplementary Material 2.5).

After considering only strict sister species pairs (9 pairs; Fig. 2.3 and Supplementary Material 2.5) and using a traditional clock calibration approach assuming that the cluster of similar divergences (8.4-11.4% K2P) migrated during the major marine trans-Arctic interchange 3.5 Ma, an average sequence divergence of 9.7% K2P (10% after removing Pseudostichopus) indicates an average COI divergence rate of 2.8%/MY (2.9%/MY after removing Pseudostichopus).

Additionally, given the complexity of the Bering Strait glacial history, I used my new iterative calibration approach instead of assuming simultaneous divergence during the major trans-Arctic interchange. The most divergent trans-Bering sister pair was Ophiopholis aculeata and O. japonica, with 15.5% K2P divergence and representing the earliest migration event through the Bering Strait; this genetic distance was then assigned to the 47 candidate calibration dates in sequence and with divergence ages for all other pairs assigned relative to this (for complete details see Supplementary Material 2.9). After comparing the distribution of divergence ages using all calibrations against the geological events, calibrations in the range of 6.7 to 5.5 Ma for the reference node were concordant with the possible glacial history of the Bering Strait. However, calibrations in this range are not equally likely when considering previous independent fossil and geological evidence (Vermeij 1991; Marincovich 2000) and the fact that the date for the first opening of the Bering Strait has been narrowed to 5.4-5.5 Ma according to fossil evidence from diatoms and molluscs and stratigraphic evidence (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004). After considering the best possible scenario, it was found that dating the most divergent sister clade at 5.5 Ma is likely to be the most feasible calibration (Fig. 2.4). Dating the most divergent trans-Bering sister clade at 5.5 Ma would lead to divergence estimations for five sister clades in the range of 3-4 Ma, consistent with the well-documented major marine migration through the Bering Strait (Briggs 1970; Vermeij 1991; Marincovich 2000; Dodson et al. 2007). In addition, stratigraphic evidence suggests that at around 5 Ma the sea-level rose by >75m around the Bering Strait (Haq et al. 1987; Marincovich 2000), enabling migration through the strait and supporting the best calibration date. According to the best calibration (5.5 Ma), the divergence time for the 16 trans-Bering sister clades ranged from 0.2 to 5.5 Ma (Fig. 2.4), while the inferred divergence rate was 2.8% pairwise K2P sequence divergence per million years for the COI barcode region in echinoderms.
Bayesian Results and Clock Calibration

Similar results were found across all six independent runs (Supplementary Material 2.10), one of which is presented in detail. The following results correspond to the run with the highest ESS for all the parameters. The estimated coefficient of variation (CV) across the whole tree was 0.407, indicating that the COI-5P gene evolves nearly in a clock-like fashion in Northern echinoderms. Rate heterogeneity among lineages can be determined by the CV, where values closer to zero (<0.1) are consistent with a strict clock, larger values (>0.1) suggest the use of a relaxed molecular clock would be appropriate, and values greater than 1 suggest substantial rate heterogeneity among lineages such that those data should not be used for divergence time estimations (Drummond et al. 2006; Drummond & Bouckaert 2015). The results suggest that sequences of echinoderms in this analysis typically varied by 40.7% of the absolute clock rate. Therefore, the choice to impose a relaxed clock is largely supported.

Before clock calibration, results from Bayesian analysis are only relative estimates as neither a lineage rate nor a calibration date were pre-assigned. The relative mean rate of evolution was 6.9 substitutions per site (95% highest posterior density regions (HPD): 2.05-10.91). The estimated relative rate for the 16 trans-Bering sister clade nodes ranged from 5.0 to 12.6 substitutions per site, with the fastest branch 2.5 times faster than the slowest branch (Fig. 2.5). It is interesting that the second-slowest branch corresponds to P. tuberosus, while the second-fastest branch is from P. aemulatus (Fig. 2.5), comprising the sole sister pair for which the clock hypothesis was rejected via the relative rate test. Aside from Pseudostichopus, rate variability was distributed along the entire tree without major deviations.

The Bayesian analysis was repeated without sequences from the sister clade with the highest rate variability (Pseudostichopus). The best substitution model describing this dataset was GTR+G+I. In this case, the relative mean rate of evolution was 6.8 substitutions per site (95% HPD: 1.89-10.25). The rate of the fastest branch (7.9 substitutions per site) was now just 1.3 times faster than that of the slowest-evolving branch (6.0 substitutions per site) (Supplementary Material 2.11). The CV was 0.2, supporting the use of a relaxed molecular clock but approaching a strict clock and indicating generally constrained rate variability across the tree.

The relative node ages were used to calibrate the molecular clock using the iterative calibration approach applied to an ultrametricized gene tree (Fig. 2.5), meaning that terminal nodes are aligned with each other and internal branch lengths are scaled to display the divergence among sister clades rather than among individual species (Gregory 2008). Imitating
the iterative calibration using K2P distances, I used the relative node age from the most divergent trans-Bering sister clade (O. aculeata vs. O. japonica) assigned to the 47 calibration dates; relative to this node, I then assigned node dates to the remaining sister clades (Fig. 2.6; for complete details see Supplementary Material 2.12). After comparing the distribution of node ages using all calibrations against the geological events, a single calibration was supported, as with the K2P analysis above. Dating the most divergent trans-Bering sister clade at 5-4.7 Ma yielded the preferred calibration according to the narrowed date for the first opening of the Bering Strait and the earliest possible migration from the Arctic to the Pacific Ocean at 5.4-5.5 Ma (Gladenkov et al. 2002). According to the preferred calibration, the divergence time of trans-Bering sister clades ranged from 0.2 to 5 Ma (Fig. 2.6). Very similar results were found when repeating the calibration process using the relative node ages from the analysis without the Pseudostichopus pair. However, the range of dates for the best calibration is narrower (5-4.9 Ma instead of 5-4.7 Ma) (Supplementary Material 2.13).

Considering Biological Traits of the Trans-Bering Sister Clades

Biological traits from trans-Bering sister clades were not always accessible (See Supplementary Material 2.14). From those trans-Bering pairs with available information, it was noticed that the majority of pairs have planktonic larval development, and only two pairs have benthic direct development (Supplementary Material 2.14). Similarly, the majority of pairs have shallow distribution, and only three sisters are distributed in deep waters (Fig. 2.7 and Supplementary Material 2.14). Interestingly, the results suggest that trans-Bering pairs of echinoderms with deep distributions (>50 m) have divergences between 3–5.5 Ma after the first opening of the Bering Strait and before the subsequent closure of the Strait due to the maximum glacial (Fig 2.7). Moreover, all trans-Bering sister pairs of echinoderms are benthic as adults and have affinity for temperate to polar waters (Supplementary Material 2.14).

DISCUSSION

Strengths of the Iterative Calibration Method

An important element in molecular-dating research is the choice of calibrations. It is recognized that uncertainty in divergence estimates is mainly driven by the calibrations (e.g. Yang & Rannala 2006; Ho & Duchêne 2014). In the absence of fossil evidence, biogeographical events are the key for clock dating. However, accurate biogeographic calibrations are determined by the availability of reliable geological dates (Ho & Phillips 2009). This study
revealed that the iterative calibration approach can be effectively used for clock dating when using complex geological events such as the repeated opening and reclosing of the Bering Strait.

A reason for the successful use of the Bering Strait is the constrained periods of time associated with the geological event. While migration through the Strait could have occurred at any time while it was open, there are strong constraints about time periods when migrations did not occur. The level of uncertainty as to the timing of the closures generally ranges from 10 to 100s of thousands of years, rather than millions. By contrast, the closure of the Isthmus of Panama was a process spanning approximately 12 MY (Duque-Caro 1990; Coates & Obando 1996; Collins et al. 1996a; Lessios 2008; Leigh et al. 2014).

Additional strengths of the iterative calibration method are that it not only recognizes the complexity of the geological history; it also incorporates further sources of evidence such as a comprehensive geological signal. Additionally, the iterative calibration approach searches for the most accurate calibration date, guarding against using a single or few sister pairs, which may possess anomalous rates of molecular evolution within their larger taxonomic group. The usage of many sister lineages is possible because it is not necessary to assume that all pairs are perfect sister species, as was required in previous calibration attempts using the Isthmus of Panama and Bering Strait. Instead, all geographically separated sister lineages are considered, recognizing that trans-Bering migration may have occurred at varying times among them. Subsequent diversification within one or both sister lineages is compatible with the iterative calibration approach. Moreover, allopatric groups below the species level are also suitable for contributing to the total evidence. The iterative calibration would benefit from but does not require globally comprehensive sampling of species within genera. One important assumption of this methodology is that organisms dispersed through the water, rather than aerially or via any other means. Therefore, this approach is best for taxa for which this assumption is likely to be true as well as for aquatic taxa represented by many allopatric sister lineage pairs, such that the total signal is apparent despite anomalous data points. In that regards, echinoderms are an exceptional taxonomic group for study due to their ecology (i.e. exclusively marine dispersal; there are no terrestrial or freshwater members of the phylum). On the other hand, the fact that the Bayesian analysis allows for rate variability across lineages, through using a relaxed clock, should minimize the potential impact of errors associated with molecular rate variation. Despite the uncertainty associated with the use of the molecular clock for divergence dating, here I used more data and investigated the influence of anomalous data points to integrate all available
sources of evidence. Consequently, the times of trans-Bering divergence in echinoderms estimated during this study are all in agreement with the glacial history of the Bering Strait.

**Divergence Times for Northern Echinoderms**

The initial finding of a wide range of K2P distances between trans-Bering sister clades (0.5% to 15.5%) could have been due to rate heterogeneity across taxonomic groups or to non-simultaneous divergence of the sister clades. However, I found that rate heterogeneity was not the dominant explanation since the molecular clock hypothesis was rejected for only one trans-Bering pair. In addition, the Bayesian results suggested that the COI-5P gene evolves in a nearly clock-like fashion for Northern echinoderms, although significant rate variation is observed for a single taxon (*Pseudostichopus*) (Fig. 2.5). Thus, the wide range of K2P divergences observed in this study (Fig. 2.3), with a similar pattern shown by the relative node ages from the Bayesian gene tree (Fig. 2.5), strongly suggests multiple pulses of trans-Bering migrations for Northern echinoderms. Multiple trans-Bering migrations are concordant with previously suggested scenarios. It has been proposed that the marine trans-Arctic interchange occurred in two stages. The first marine interchange occurred from the Atlantic to the Pacific Ocean, immediately after the initial opening of the Bering Strait (5.5 Ma) (Marincovich & Gladenkov 1999; Marincovich 2000). Then, a second interchange took place around 3.5 Ma from Pacific to Atlantic, co-occurring with a reorganization of Northern Hemisphere ocean circulation after the formation of the Isthmus of Panama (Marincovich & Gladenkov 1999; Marincovich 2000). The Pacific-Arctic/Atlantic marine interchanges during the Pleistocene are also consistent with ca. 9 cycles of glaciation in the last 2.4 MY (Harris 2005). Multiple marine migrations through the Bering Strait have also been suggested in polychaetes according to molecular evidence (Carr 2010). In addition to migratory events during the major trans-Arctic interchange 3.5 Ma, several polychaete migrations also occurred between 1.9 and 0.4 Ma and even during the last 120,000 years (Carr 2010).

**Rate Heterogeneity in the Pseudostichopus Sister Pair**

The only trans-Bering pair for which the rate constancy hypothesis was rejected was *Pseudostichopus*. Further examination of *Pseudostichopus* sequences and alignment allowed us to detect high sequence divergence and probable base-calling errors in this taxon (Supplementary Material 2.2). After eliminating the sequence with possible editing errors, high rate variability in this pair was still observed, particularly in *P. aemulatus* in relation to other rates across the tree (Fig. 2.5). The observed high COI-5P rate in *P. aemulatus* could be explained by the expected rate heterogeneity across taxa (e.g. Martin and Palumbi 1993). However, further,
more subtle, sequence editing errors are still a possibility, and investigation of the sequence chromatograms, which are not presently available, would be needed. Having access to the original trace files, which is feasible for many barcode sequences in BOLD database, is highly valued and would have permitted in this case a re-examination of the original sequences to discard or correct errors. Despite the rate heterogeneity in *Pseudostichopus*, the inclusion of this pair is still valuable to allow comparison with the rest of the trans-Bering sister clades. Moreover, after accounting for rate variability, it was possible to assign the time of divergence (3.1 Ma) for the *Pseudostichopus* pair.

Rate Estimation for Northern Echinoderms

Despite that clock calibration using K2P divergences was very similar when using a simplistic and traditional approach (COI divergence rate of 2.77%/MY) versus using the iterative calibration (2.81%/MY), the second method provides a more reliable outcome by including further sources of information such as a comprehensive geological history and the use of more sister lineages. A problem when using an average of divergences for clock dating is that if one or more sister species pairs were missing, the results would greatly fluctuate. For example, the calibrated rate estimate here varies between 2.6% and 2.9% per MY upon dropping each of the 4 data points in sequence and recalculating the rate based upon 3 points; the rate varies from 2.5% to 3.0% per MY when systematically dropping the largest two divergences, followed by the lowest two divergences. Moreover, assuming simultaneous divergence of trans-Bering pairs through the trans-Arctic interchange can produce misleading estimates of divergence times (see further discussion below).

A single preferred calibration date assigned to the most-divergent sister pair is suggested by the K2P results using the iterative approach (5.5 Ma) and supported by the Bayesian results (5-4.7 Ma). It is unlikely that echinoderms from this study would have migrated before the first opening of the Bering Strait at 5.5 Ma (Gladenkov *et al.* 2002; Gladenkov & Gladenkov 2004). Although this scenario cannot be completely ruled out, it is less probable; therefore, it was not used for clock calibration. It is unlikely that trans-Bering taxa would have colonized both Pacific and Arctic-Atlantic Oceans well before the first opening of the Bering Strait as it is well established that North Pacific and Arctic-Atlantic marine faunas evolved separately since the mid-Cretaceous (80-100 Ma) until the formation of the Bering Strait marine corridor (Dunton 1992; Marincovich 1993). Although migration of the trans-Bering pairs through the Panama seaway could have been possible before the first opening of the Bering Strait, that is also unlikely given that all trans-Bering sisters included in this study have affinity for temperate to
Polar waters (Supplementary Material 2.14) and also because echinoderms tend to exhibit limited ranges in their thermal tolerance, and entire genera and even families tend to be endemic to the Arctic or tropics (e.g. Lessios et al. 2012; Mah & Blake 2012). Another implausible scenario is if the most divergent pair would have migrated between 3 and 3.4 Ma. This would imply that only one of the 16 pairs migrated through the Bering Strait shortly after the major trans-Arctic interchange (3.5 Ma; Vermeij 1991). Dating the beginning of the second group of invasions right after the maximum glacial, around 1.9-2.4 Ma, would suggest ongoing trans-Arctic migrations within the last 2.4 MY.

The difference in the best calibration date proposed by my iterative approach when using K2P distances (5.5 Ma) vs. Bayesian results (5-4.7 Ma) can be understood as an effect of rate variability across taxa and due to differences in sequence length between K2P distances and Bayesian analysis (>651 vs. 459bp). It has been demonstrated that low-frequency variants, indicative of sequencing or base-calling errors, are located near the ends in the COI-5 barcode region in the first and last 50bp of the sequences (Athey 2013). Thus, trimming the alignment to 459bp for the Bayesian analysis could have reduced miscalculations due to low-frequency variants.

*Rate Estimation in Northern Echinoderms vs. other Rates in the Literature*

The results suggest a rate of 2.8% pairwise K2P sequence divergence per million years for a >651-bp section of the COI barcode region in trans-Bering echinoderms. Here, I report a faster COI divergence rate for Arctic echinoderms than a prior report for *Leptasterias* sea stars (0.009 nucleotide substitutions per site per MY, equating to a pairwise divergence rate of approximately 1.8%/MY; Foltz et al. 2008). Substitution rates in *Leptasterias* were estimated using one trans-Bering sister pair (*L. muelleri* and *L. stolacantha*) assuming divergence at 3.5 ± 0.25 Ma during the major trans-Arctic interchange (Foltz et al. 2008). Contrary to the previous *Leptasterias* study, the results from this study suggest that *Leptasterias* migrated through the Bering Strait more recently, approximately 2 Ma (Figs. 2.4 and 2.6). Assigning a time of divergence of 3.5 Ma to the *Leptasterias* trans-Bering sister clade in this study would imply a pairwise divergence of approximately 1.6%/MY and that the most divergent pair migrated as early as 9.7 Ma, prior to the first opening of the Bering Strait, which is very unlikely.

The divergence rate reported here for Arctic echinoderms is faster than a previous report for the tropical *Diadema* sea urchin, but slower than previously stated for other tropical echinoderms. However, comparison is challenging because the taxonomic range, sequence lengths, and the included region of the COI gene are not identical among studies. Low genetic
divergences for the COI-3P region were observed between a geminate pair of *Diadema* (4.57% K2P), which the authors interpreted as being consistent with isolation at the last point of the closure of the Isthmus of Panama (Lessios *et al.* 2001); therefore, the molecular clock for *Diadema* was calibrated using a date of 1.8 Ma, resulting in a K2P divergence rate of 2.6%/MY (Lessios *et al.* 2001). By contrast, higher rates of K2P divergence in COI-5P have been reported for one pair of sister species in the genus *Oreaster* (5%/MY; Hart *et al.* 1997) and for the COI-3P region for the sea urchins *Echinometra* (3.49%/MY; McCartney *et al.* 2000). Furthermore, a per-lineage substitution rate of 3.12%/MY, equating to a pairwise divergence rate of 6.23%/MY, was reported for the COI 5P and 3P regions in the sand dollar *Melitina* (Coppard *et al.* 2013).

Previous studies based their calibration on the assumption that geminate species diverged at the completion of the Panamanian Isthmus (Hart *et al.* 1997; McCartney *et al.* 2000; Coppard *et al.* 2013). Additionally, a range of 4.2-13.5% K2P divergences has been reported for eight geminate species of tropical sea urchins (Lessios 2008). The author selected which genetic distances were “close enough to each other to be considered as a cluster” (Lessios 2008, p.79), suggesting similar times of isolation at the time of the Isthmus of Panama closure (Lessios 2008). The resulting average divergence rate of 3.7%/MY (calculated by later authors from geminate species divergence values presented in Lessios 2008) has been then used for dating in other echinoderms (Vogler *et al.* 2008; Boissin *et al.* 2011; Hoareau *et al.* 2013). The results presented here revealed that nine of the 16 trans-Bering clades exhibited K2P divergences (5.6-13.7%, Fig. 2.3) similar to those reported for tropical sea urchins (4.2-13.5%; Lessios 2008). However, after discounting rate heterogeneity for Northern echinoderms as the primary cause of this large range, the results suggested different times of divergence instead of simultaneous isolation, contrary to what has been generally anticipated or assumed for tropical sea urchins. It is possible that a temporal spread of divergence times has occurred among tropical echinoderm sister pairs, as noted as likely in the case of shrimps inhabiting various habitats (Knowlton & Weigt 1998).

Interestingly, the above studies on tropical echinoderms have used different dates for the closure of the Panamanian Isthmus in their respective clock calibrations: 3.5 Ma (Hart *et al.* 1997), 3.1 Ma (McCartney *et al.* 2000), 3 Ma (Lessios 2008; Coppard *et al.* 2013), and as recently as 1.8 Ma (Lessios *et al.* 2001). The use of different calibration dates for the same geological event has affected molecular rate estimates for tropical echinoderms. In addition, the long period of time involved in the Isthmus formation and a probable secondary oceanic connection through the land bridge are other factors needing further consideration when estimating molecular rates in tropical lineages. Further research for tropical taxa could benefit
from using the iterative calibration approach in order to unify molecular and geological evidence and provide more consistent results.

In contrast with the well-cited divergence rate of 1.4%/MY in *Alpheus* shrimps estimated using the Isthmus of Panama (Knowlton & Weigt 1998), this study revealed a higher K2P divergence rate for Northern echinoderms (2.8%/MY). Results from trans-Bering echinoderms also differ from the high rate in the COI barcode region previously reported for Arctic polychaetes (4.4%/MY; Carr 2010) and tropical marine molluscs from the genus *Arcopsis* (5.1%/MY; Marko & Moran 2002). Further work is necessary to expand the novel calibration approach to other marine groups, in order to investigate rate estimates among taxa using consistent methods and genetic regions.

*Phylogenetic Scope for the Substitution Rate Presented in this Study*

Difference in evolutionary rates could be taxon specific and fluctuate across genes (Ho & Duchêne 2014); therefore, the substitution rate for the COI-5P gene presented in this study should not be generalized to all taxa, nor to all genes. Furthermore, the resulting substitution rate might not be appropriate for dating recent demographic events, nor for dating extremely deep nodes in the tree of life. In the context of the time-dependency hypothesis, evolutionary rates are not constant over time, with an exponential decline as a function of the calibration age (Ho *et al.* 2005, 2011). Thus, the rate based on interspecific divergences between sister lineages would not be an accurate generalization for divergence estimations far outside of the calibration time frame (2-5 MY) from this study, especially if the transition from high, short-term (<1-2 MY) mutation rates to lower, long-term substitution rates (Ho *et al.* 2005; Ho & Larson 2006) applies to echinoderms as previously observed in other lineages (Ho *et al.* 2015a).

This study used a relaxed uncorrelated clock to account for rate variability across the phylum as suggested for the dataset, which slightly deviates from the strict clock model. The observed rate variability across the phylum was not excessively large (with the only exception of *Pseudostichopus*), indicating that the calibration for COI-5P presented here might be carefully used in molecular clock research for Arctic echinoderms.

It has been proposed that evolutionary rates and speciation processes are faster in warmer climate regions (e.g. Rohde 1992; Bargelloni *et al.* 1994; Bleiweiss 1998b; Gillooly *et al.* 2005; Wright *et al.* 2006). Therefore, my results are most appropriate for research involving Northern echinoderms and might only be used for comparative purposes when estimating rates and dates for tropical echinoderms, considering that rates might vary along the latitudinal
gradient. While the resulting rate for Northern echinoderms is somewhat lower than prior estimates for tropical lineages, further research is needed to evaluate tropical vs. Arctic rates using phylogenetically paired taxa as well as consistent gene regions and analysis methods. Furthermore, it is important to consider that these results might not represent the appropriate evolutionary rate for Northern members of the class Echinoidea. Sea urchins were not included here as I was not able to detect any trans-Bering sister clade, given the limited COI-5P data available for that class. The absence of sea urchins in this analysis contrasts with the fact that most tropical research is based on the assessment of trans-Isthmus sea urchin divergences (e.g. Lessios 1979, 2008; McCartney et al. 2000; Lessios et al. 2001).

Sources of Error of this Method

Extremely high rate variability would make this method difficult to use for obtaining a single calibration for a large taxon. In those cases, evidence for more phylogenetically localized clocks may be explored. However, a strength of the iterative calibration approach is explicitly considering potential rate variability. One more source of error is related to currently partial phylogenies. For highly incomplete phylogenies, it is possible that the geographical shift might have been assigned to the incorrect node (Supplementary Material 2.15). Another source of error is that phylogenetic events may not be associated with the biogeographical event. For example, divergence could have occurred earlier, and then one lineage later migrated through the Bering Strait; as well, migrations may not be recorded due to extinctions. These uncertainties may be partially mitigated through seeking large sample sizes of sister pairs and generalities in the divergence trends. Moreover, uncertainties about the exact age of the geological events persist as a limitation when calibrating the clock. Whereas some major events like the closure of the Isthmus of Panama in Central America are subject to ongoing controversy, other events like the opening/closing of the Bering Strait are dated with more accuracy and with strongly constrained time periods; nevertheless, some error associated with the calibration itself likely remains. On the other hand, calibration failure using the iterative approach can occur when there is no resolution between geological and genetic divergence data. This issue could arise due to excessive molecular rate variability or due to different migratory routes (e.g. aerial/terrestrial dispersal); in those cases, highly variable and uncertain node age estimations might occur.

Future Directions

This research would benefit from estimating confidence intervals on the calibration itself. In addition, it would be valuable to estimate divergence rates using other substitution models in
order to allow further comparisons across taxonomic groups and to provide a range of calibrations for future use. Complementing COI-5P with additional genes would be advantageous for assessing whether patterns in molecular rates are consistent beyond the mitochondrial genome. Ideally, more trans-Bering sister clades should be used to estimate the most accurate rate of evolution for Arctic echinoderms. However, this improvement is associated with the general need to increase the taxonomic and geographic breadth of molecular data for marine fauna. Molecular clock research can benefit from the iterative calibration approach by integrating all available sources of evidence when dating phylogenies. A promising future direction would be to explore further the iterative calibration method using taxa with an excellent fossil record, such as many molluscs. Clock calibrations for terrestrial groups can also take advantage of the iterative calibration approach, for example considering that land migration events would have occurred during the existence of the Bering Land Bridge, when the Bering Strait was closed. Well-dated phylogenies may, in turn, represent new opportunities for dating other taxonomic groups with neither fossil nor geological data associated with divergence events, always guarding against anomalous rates of evolution or an exceptional dispersal history.

CONCLUSIONS

This chapter has presented a novel approach for molecular clock calibration integrating phylogenetic, distributional, and geological/climatological history data. Incorporating several sources of evidence through the iterative calibration anticipates more accurate results than when using simplistic methods for clock dating (e.g. using a divergence average). Therefore, the iterative calibration represents the most appropriate approach when using complex biogeographic events for clock dating. This study demonstrates how the opening/closing of the Bering Strait, together with my new iterative calibration approach, can successfully be used in clock dating research to shed light upon the evolutionary history of Northern taxa and the influence of major climatic upheaval upon biogeography and diversity. The results suggest that the COI-5P divergence rate in trans-Bering echinoderms is slower than rates previously reported for tropical sea urchins (Hart et al. 1997; McCartney et al. 2000; Lessios 2008; Coppard et al. 2013), but faster than earlier-reported rates for Arctic sea stars (Foltz et al. 2008). Additionally, the results strongly support several migrations across the Bering Strait as previously suggested for polychaetes (Carr 2010). This study represents a significant advantage over prior calibrations after using a substantial number of sister clades, together with the iterative calibration approach. Lastly, I recognize the potential to expand upon molecular clock
research using the iterative calibration approach in other taxa and with additional biogeographical events.
Figure 2.1 Geological history and map of the Bering Strait (modified from Marincovich 2000; Gladenkov and Gladenkov 2004). a) Shows the chronostratigraphic scale (Ogg et al. 2016) in relation to eustatic sea level changes (Haq et al. 1987) and major Bering Strait geological events during the last 10 MY in the Northern Hemisphere; the black dot represents the first opening of the Bering Strait at the end of the Miocene as inferred from diatom and bivalve fossils (Gladenkov et al. 2002; Gladenkov & Gladenkov 2004). b) Map showing the Bering Strait.
Figure 2.2 Flowchart describing the iterative calibration approach. The iterative calibration approach is a consecutive process of clock calibration that involves comparing node dates with the geological time scale and any other available information, such as the fossil record, in order to reach congruence between molecular divergences and biogeographic evidence. For the purpose of this study, the oldest possible date of dispersal through the Bering Strait was assigned to the most-divergent node as the starting calibration date. 1All reasonable calibrations according with the geological event of interest. 2If the iterative calibration approach results in calibration failure, alternative sources of calibration should be used (e.g. fossil calibration, different geological event). In extreme cases, clock calibration could be unsuccessful.
**Figure 2.3** Mean COI K2P % divergences between 16 trans-Bering sister clades of echinoderms. The 16 trans-Bering comparisons are assigned to one of three categories: sister clades (blue), sister pairs within BINs (light pink), and strict sister species pairs (white). Two points with the same divergence are offset vertically. Only strict sister species pairs (white) were considered for clock calibration with the traditional methodology using the average sequence divergence of the visually-determined cluster of points marked with arrows and assuming a divergence time of 3.5 Ma (see materials and methods section for further details). The triangle represents the divergence of the trans-Bering sister clade of sea stars in the genus *Leptasterias*. A trans-Bering sister species pair of *Leptasterias* was previously used in the literature (Foltz et al. 2008) for clock calibration, assuming divergence during the major trans-Arctic interchange 3.5 Ma.
Figure 2.4 Divergence time estimation of 16 trans-Bering sister clades of echinoderms using K2P divergences and the iterative calibration approach. Dots represent divergence date estimations in Ma (X axis). Each horizontal row of points represents a different calibration attempt (Y axis, right) using a different calibration date and the divergence rate generated in K2P%/MY (Y axis, left). For visualization purposes, 8 of the total 47 calibration attempts are presented in this figure including the first calibration date (7 Ma; calibration attempt 1) and the last calibration date (2 Ma; calibration attempt 8) tested (for further details see Supplementary Material 2.9). Overlapping points representing two identical sister divergences are highlighted with a thicker outline. All calibration attempts were compared with the geological history of the Bering Strait to find congruence between the genetic and geologic evidence. Shaded areas on the timescale indicate times when trans-Bering migration was extremely unlikely. Calibrations generating divergence times that fall in these regions are considered improbable, whereas calibrations where these areas contain no points are favoured. Calibration attempt 3 (star), to which a divergence time of 5.5 Ma was assigned to the most-divergent clade, represents the most feasible calibration that is concordant with the geological evidence.
**Figure 2.5** Maximum credibility ultrametricized tree for the COI-5P gene in 16 trans-Bering sister clades of echinoderms generated in BEAST v1.8.2. The 16 trans-Bering sister clade nodes are marked by stars. Rates are colour mapped for each branch where the brightest red branch represents the fastest rate, while the lightest blue represents the lowest rate. The scale bar represents branch length in nucleotide substitutions per site. Tip nodes were collapsed by allopatric sister taxa.
Figure 2.6 Divergence time estimation of 16 trans-Bering sister clades of echinoderms using relative node ages from the ultrametricized gene tree generated in BEAST v1.8.2 and the iterative calibration approach. Dots represent divergence date estimations in MY (X axis). Each horizontal row of points represents a different calibration attempt (Y axis) using a different calibration date. For visualization purposes, 8 of the total 47 calibrations attempts are presented in this figure including the first calibration date (7 Ma; calibration attempt 1) and the last calibration date (2 Ma; calibration attempt 8) tested (for further details see Supplementary Material 2.12). All calibrations were compared with the geological history of the Bering Strait to find congruence between the genetic and geologic evidence. Shaded areas on the timescale indicate times when trans-Bering migration was extremely unlikely. Calibrations generating node dates that fall in these regions are considered improbable, whereas calibrations where these areas contain no points are favoured. Calibration attempt 4 (star), to which a divergence time of 5 Ma was assigned to the most-divergent node, represents the most feasible calibration that is concordant with the geological evidence. Divergence rates are not presented because results from Bayesian analysis are only relative estimates since neither a lineage rate nor a calibration date were pre-assigned during the Bayesian analysis (see further details in Results).
Figure 2.7 Divergence time estimates for trans-Bering echinoderms using K2P divergences based on the best calibration and mapping depth distribution. The trans-Bering sister clades that were used for calibration and the sister clade that was omitted during the calibration process after rejecting the clock hypothesis are shown. The shaded boxes show the two periods of time when trans-Bering migrations were extremely unlikely, during the maximum glacial (2.4–3 Ma) and before the earliest possible opening of the Bering Strait at 5.5 Ma.
CHAPTER III
A New Molecular Clock for Arctic Marine Invertebrates

ABSTRACT

Divergence times for Arctic marine lineages have commonly been estimated based on calibrations from geographically distant taxa. However, due to evidence of rate heterogeneity among taxa and environments, it is essential to pursue clock calibrations targeted to Northern lineages. The opening and re-closure events of the Bering Strait provide an exceptional resource for calibrating the molecular clock in Northern marine taxa. Here, I expanded the use of the novel *iterative calibration approach*, which incorporates the complete glacial history of the Bering Strait for clock dating. Following the success of the iterative calibration for clock dating in trans-Bering sisters of echinoderms (Chapter II), patterns of molecular divergence across Northern marine molluscs and arthropods were investigated using publicly available sequences of the cytochrome c oxidase subunit I (COI) gene. Additionally, previous genetic distances from trans-Bering polychaetes (Carr 2010) were re-examined using the iterative calibration approach. The novel results from molluscs, arthropods, and polychaetes were then compared with previous research in Arctic echinoderms (Chapter II). The wide range of Kimura two-parameter (K2P) divergences between 91 total trans-Bering sister taxa (0.12–26.37%) strongly supports multiple pulses of trans-Bering migrations. The results indicate a rate of K2P divergence of 2.8%/MY in echinoderms, 3.2%/MY in molluscs, 3.5–4.7%/MY in polychaetes, and 5–5.2%/MY in arthropods. These rates contrast with a highly-cited low divergence rate reported for tropical lineages (1.4%/MY), but agree with several other published calibrations in the range of 3–5%/MY. By integrating genetic, biogeographic, and fossil evidence, and by using a substantial number of sister clades, this study anticipates more accurate calibrations than those based on simplistic assumptions. The new rates presented here provide an advance for dating evolutionary events in the marine realm and for understanding the influence of prior climatic changes upon the history of life.
INTRODUCTION

The molecular clock is a fundamental tool for establishing an absolute time scale for the tree of life. Traditionally, fossil evidence has provided the main source of dates for calibrating the clock (De Baets et al. 2016), although for many lineages the fossil record is highly incomplete or absent. Biogeographic events are an effective resource for clock calibration especially for soft-bodied taxa, such as diverse marine lineages, for which divergence times cannot be estimated using good fossil evidence. Biogeographic calibrations of the molecular clock are common for tropical marine taxa (e.g. Knowlton & Weigt 1998; Lessios 2008; Coppard et al. 2013; summary in Table 1.1) but limited for Northern marine groups (Foltz et al. 2008; Henzler & Ingólfsson 2008; Carr 2010).

Divergence estimations for Arctic marine lineages have commonly been established based on calibrations from geographically distant taxa (e.g. Hart et al. 1997; Luttikhuizen et al. 2003; Blanco-Bercial et al. 2011; Boissin et al. 2011; Milligan et al. 2011) or from taxa with different ecological traits (e.g. Hart et al. 1997; Milligan et al. 2011; Jung et al. 2013). However, that approach might generate inaccurate time estimations if one does not account for variability in the rate of molecular evolution among taxa exhibiting different traits (e.g. Bromham 2002; Bromham & Penny 2003; Thomas et al. 2006; Mitterboeck & Adamowicz 2013) and inhabiting varied environments (e.g. Bleiweiss 1998b; Gillooly et al. 2005; Wright et al. 2006). Thus, it is necessary to pursue clock calibrations targeted to Northern groups. In that regard, the opening and re-closure events of the Bering Strait provide an exceptional resource for clock calibration in Northern taxa.

The Bering Strait has a complex geological history (reviewed in Chapter II, Fig. 2.1). However, it can successfully be used for clock calibration when implementing the iterative calibration approach, a novel method that I proposed and developed in Chapter II. The iterative calibration approach incorporates several sources of evidence (e.g. phylogenetic, distributional, and geological/climatological information) and seeks congruence between the dates assigned to multiple divergence events and geological history. With a more realistic consideration of biogeographic events, I expect more accurate results using the novel methodology than when using simplistic methods (e.g. using a divergence average), as demonstrated in Chapter II.

In Chapter II, I investigated the divergence rate of Northern echinoderms using the Bering Strait and applying the iterative calibration approach. The results showed a broad range of allopatric divergences consistent with multiple pulses of trans-Bering migrations during the
repeated opening and closing of the Bering Strait. The results also suggested a divergence rate of 2.8% per million years for the COI DNA barcode region, using Kimura-2-parameter (K2P; Kimura 1980) genetic distances. The breakthrough in calibration for Northern echinoderms detailed in Chapter II can be pursued for other marine invertebrates, which is the purpose of this study. Moreover, many previous studies have used only one or a few pairs of sister species separated by a barrier when calibrating the molecular clock (e.g. Knowlton & Weigt 1998; Lessios et al. 1999, 2012), and further research with a larger number of pairs is still needed.

In Chapter III, I expand the application of the iterative calibration approach in order to provide novel divergences and estimate new rates for other major groups of marine invertebrates: molluscs and arthropods. I also reanalyzed genetic distance data from polychaetes (Carr 2010) using the iterative calibration, and compared the results with those obtained for echinoderms. Overall, this chapter will advance our ability to use molecular clocks to date evolutionary events in the marine realm, thus contributing to the field of molecular evolution.

METHODS

Selecting Orders for Targeted Analysis Using Public Data

I used publicly available sequences of the mitochondrial cytochrome c oxidase subunit I (COI-5P) barcode region (Hebert et al. 2003a) from marine molluscs and arthropods. Within the two groups of interest and following the taxonomic classification listed in the Barcode of Life Data Systems v4 (BOLD; http://v4.boldsystems.org; last accessed March 9th, 2017) (Ratnasingham & Hebert 2007), I first discarded orders without marine taxa, those orders that consist entirely of parasitic species, and those containing solely extinct species according to the World Register of Marine Species (WoRMS; http://www.marinespecies.org last accessed March 9th, 2017). Therefore, I selected only those orders containing extant, free-living marine taxa for further analysis.

To search for candidate trans-Bering sister clades within the two target phyla, I constructed preliminary neighbour-joining (NJ) trees (Saitou & Nei 1987) using K2P distances for each order, using the phenogram-building tool available in BOLD v4 and following the same settings outlined in Chapter II. Allopatric Pacific vs. Arctic/Atlantic pairs were then recognized as potential trans-Bering sister clades, following the same criteria used in Chapter II. Distribution of the potential trans-Bering clades was verified by searching for the species name, when available, in the Encyclopedia of Life (EOL; http://www.eol.org last accessed March 9th, 2017)
and WoRMS. These NJ trees were used to select the target orders for sequence data retrieval and further analysis.

Identifying Trans-Bering Sister Clades through Maximum Likelihood Phylogenetic Analysis

To identify sister-clade relationships, I used previously published molecular phylogenies from the taxa of interest, when available (e.g. Reid et al. 1996), giving preference to mitochondrial trees to avoid bias relating to incomplete lineage sorting (Mendes & Hahn 2016). In the absence of published trees, phylogenetic relationships were reconstructed using a maximum likelihood (ML) analysis for each target taxon, i.e. those that contained a potential trans-Bering sister pair in the NJ phenogram.

When possible, I constructed ML trees including all available sequences for each target order. The order level was selected for analysis, whenever feasible, due to missing lower-level taxonomy for some specimens. However, for orders that contained >6,000 sequences, I used family-level datasets for computational efficiency and visualization purposes. From July 18th to February 17th 2017, publicly available COI sequences from all taxa containing potential trans-Bering sister clades were retrieved within the workbench in BOLD v4 using a taxonomy-based search. Available data from each target order and family were downloaded excluding sequences with stop codons, those flagged as containing possible misidentifications, and sequences shorter than 400 base pairs (bp). The geographic distributions of target taxa were verified at this point. In addition, the genera Hyperoche (Amphipoda) and Odostomia (Heterostropha) were excluded because they contain parasitic species. Parasites were excluded due to previous evidence of an unusual pattern of high divergences in parasitic taxa compared with other crustaceans (Costa et al. 2007).

Each order or family sequence dataset was aligned using the ClustalW algorithm (gap opening penalty = 20; gap extension penalty = 0.88, all other settings as default), and trimmed in MEGA 7 (Kumar et al. 2016). Each order or family dataset was trimmed independently, retaining the full length of the barcode region of COI to the maximum extent possible (i.e. the region at the 5' end of the COI gene amplified by Folmer et al. primers (Folmer et al. 1994) or similar primers, with length of 658 bp for most taxa). For some datasets containing a large fraction of shorter sequences, it was necessary to trim the ends until at least approximately 60% of the sequences contained bases at both ends (Supplementary Material 3.1). All alignments were inspected for gaps verifying that, if present, they were in multiples of three nucleotides, reflecting an amino acid insertion or deletion; if not, gapped sequences were deleted. The alignments were also translated to amino acids to verify the reading frame and absence of stop
Problematic sequences (>1% of unknown bases aside from the flanking regions, potential misidentification cases, and possible base-calling errors (see Chapter II), with no trace files available for verification) were then eliminated from the datasets. Duplicate sequences were eliminated using the online tool ElimDupes (https://hcv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html; last accessed February 23rd 2017), considering subsequences as duplicates. For the majority of the datasets, I eliminated only sequences that were 100% identical. However, when the datasets contained >1,500 sequences, I eliminated sequences that were more than 98% similar.

For each of the datasets, a ML tree was built in MEGA 7 using the best-fit substitution model according to the Bayesian Information Criterion (BIC), with nodal support estimated using 1000 bootstrap replicates. Details about trans-Bering sister clades and ML analysis can be found in Supplementary Material 3.1. The list of trans-Bering sister clades of molluscs and arthropods can be found in Supplementary Material 3.2. As defined in Chapter II, trans-Bering sister clades are reciprocally monophyletic groups having one Barcode Index Number (BIN) (Ratnasingham & Hebert 2013) distributed only in the Pacific and the closest related BIN distributed in the Arctic-Atlantic. Allopatric sister clades where one or both sisters consisted of multiple BINs were also considered. Sister-clade relationships were considered well supported if the bootstrap value was ≥70%.

Rate Constancy Test

To test for rate constancy within each of the trans-Bering sister clades, I used a Tajima’s relative rate test (Tajima 1993) in MEGA 7. This test compares relative rates of molecular evolution between sequences of sister taxa against an outgroup. The outgroups were chosen based on the ML trees, selecting the most closely related taxon to the sister clade of interest (Robinson et al. 1998) (further details in Supplementary Material 3.3).

Sequences corresponding to each trans-Bering sister pair and their target outgroup were selected from the order or family sequence datasets used for ML analysis after eliminating duplicates. Subsequently, for each rate constancy test, I selected the longest sequence per species; when multiple sequences with the same length were available, I randomly selected one sequence per species for the comparisons. Taxon pairs for which the null hypothesis of constancy rate was rejected were not used for clock calibration, but were retained for plotting on the time scale and were subject of further discussion.
**Genetic Distances between Trans-Bering Sister Clades**

Average pairwise genetic distances between members of each trans-Bering sister clade were estimated using only unique sequences. Genetic distances were calculated in MEGA 7 using two metrics. First, I used Kimura-2-parameter (K2P; Kimura 1980) distances with pairwise deletion of missing nucleotides. I selected this distance metric to compare my results with prior studies (e.g. Lessios 2008; Wilke et al. 2009; Carr 2010; Miura et al. 2010) and with results on echinoderms from Chapter II. Second, I complemented the use of K2P distances with a more complex model of substitution: Tamura-Nei (TN93; Tamura & Nei 1993) distances with gamma-distributed rate heterogeneity among sites (model TN93+G) and pairwise deletion of missing nucleotides. To be consistent across taxonomic groups, the gamma parameter was set to 0.3 for all pairs. This value was the result of calculating the average (mean) of the median values for each of the two phyla, based upon the estimated gamma parameters from the TN93+G model for each pair.

The more complex model selected here, TN93+G, takes into account different substitution rates between nucleotides, different nucleotide frequencies, and different substitution rates between purines and pyrimidines (Tamura & Nei 1993), together with variability in substitution rates among nucleotide positions (gamma parameter). Although TN93+G was not the best-fit model for all trans-Bering sister pairs, it was typically among the top five models (among 24 tested) for 35 of 55 pairs, according to the BIC. Selecting a single distance model for all pairs allows for consistent comparison of results across pairs and taxonomic groups.

**Saturation Test**

To test whether the association between number of substitutions and accumulation of genetic divergence deviated from linearity, a saturation test was performed using DAMBE v.6 (Xia & Xie 2001; Xia 2013). Two corrected genetic distances (K2P and GTR) were each plotted against the number of transitions and transversions, including all nucleotide positions. To find a representative estimate of substitution saturation for all the target groups, a saturation test was performed in the two largest order-level datasets within each of the phyla, molluscs and arthropods, for a total of four order-level datasets: Sacoglossa, Nudibranchia, Decapoda, and Isopoda (Supplementary Material 3.4).

A saturation threshold was determined for molluscs and arthropods based on the linearity of the saturation plots. For each group, plots of divergence vs. transitions were generated. Data subsets were created with a different maximum pairwise divergence value, stepping in intervals...
of 1%. A least-squares linear regression model was then fit for each, forced through the origin. Afterwards, the threshold was determined by choosing the interval with the largest drop in the $R^2$ value of the linear model. Trans-Bering sister clades exceeding the divergence of the saturation threshold were not used as primary sources for clock calibration, but they were not entirely excluded. Their divergences were also plotted on the time scale and were the subject of further discussion.

*Iterative Calibration of Molecular Rates Using Opening/Closure Events of the Bering Strait*

The molecular clock was independently calibrated using the iterative calibration approach (Fig. 2.2) for molluscs and arthropods, first using K2P distances and next using TN93+G distances. Considering the results from the relative rate test and saturation test, trans-Bering clades rejecting the clock hypothesis or with divergences larger than 17% K2P were avoided as primary sources of information during the calibration process. A reference node for time scaling was then selected. The clock hypothesis must be accepted for the reference node, which must also have a divergence estimate below the saturation threshold, be from a well-sampled phylogeny (ratio of described species diversity and the number of total available BINs >66%) and, in the case of molluscs, have a fossil record associated with the trans-Arctic interchange for comparison.

Following the process described in Chapter II, I assigned a starting calibration date to the reference node. To assign a calibration date, I started with the earliest possible opening of the Bering Strait suggested to date in the literature according to stratigraphic evidence and fossils from diatoms and molluscs (7 Ma; Marincovich & Gladenkov 1999) (see chapter II for further details), although more recent evidence has narrowed the first opening of the Bering Strait to a date of 5.4–5.5 Ma (Gladenkov *et al.* 2002; Gladenkov & Gladenkov 2004), which is the most commonly accepted date for the initial opening of the Bering Strait (e.g. Hardy *et al.* 2011; Vautravers 2014; De Schepper *et al.* 2015). Subsequently, the times of divergence for the remaining pairs were manually assigned in relation to the genetic distance of the reference node. The divergence time estimates from all trans-Bering sister clades were graphed and then compared against the geological history of the Bering Strait. Using the same principle, I tested 121 calibration dates in the range between 1 to 7 Ma (stepping by 0.05 MY). There are two periods of time when trans-Bering migrations were extremely unlikely: preceding the first opening of the Strait dated at 5.5 Ma according to recent evidence (Gladenkov *et al.* 2002; Gladenkov & Gladenkov 2004) and during the maximum glacial period (2.4 to 3 Ma) (Einarsson *et al.* 1967; Schrader *et al.* 1976; Herman & Hopkins 1980; Maslin *et al.* 1996; Haug *et al.* 1999;
Calibration attempts generating trans-Bering divergences within these time periods are considered highly improbable, as trans-Bering migration (followed by allopatric divergence) could not occur when the Strait was closed. However, specific biological traits (e.g. intertidal species, tolerance to freezing temperatures) might have permitted the dispersal through the Bering Strait through rare water channels during glacial episodes. Migrations between Pacific and Atlantic Oceans were also possible through the Panama seaway until the end of the Pliocene. However, after the isthmus formed, the only likely migratory route for Northern marine taxa has been through the Bering Strait. Distribution and divergences should therefore be closely related to the opening/closing events of the Bering Strait, and additional dispersal paths should be rare; therefore, the periods when the Strait was closed are used as time constraints for the calibration.

Lastly, the most likely calibration date was chosen once I found concordance between node divergences and the Bering Strait geological time scale (see details in Chapter II). Then, using the most probable calibration date, the K2P divergence rate and the TN93+G divergence rates were estimated. The divergence rate (r) for the reference node was estimated as \[ r = \frac{D}{T} \], where D is the percentage of genetic distance and T is the time of divergence according to the best calibration from the iterative calibration approach, considering all trans-Bering pairs passing the inclusion criteria.

**Sensitivity Analyses: Fossil Record, Taxon Sampling, and Biological Traits**

The trans-Bering marine interchange has been primarily recognized from the fossil record in molluscs (Vermeij 1991). Thus, the fossil record from the Pliocene associated with the major trans-Arctic interchange (Vermeij 1991; Reid et al. 1996) was compared against the divergence times resulting from biogeographic calibrations using the iterative calibration approach. Unfortunately, a fossil record of marine arthropods associated with the trans-Bering interchange is not available. Fossil evidence from the major trans-Arctic marine interchange is particularly valuable because it provides target points to compare with biogeography-based calibrations.

The sensitivity of the results to taxon sampling and ecological traits was also examined. Highly incomplete phylogenies, due to poor taxon sampling, can influence the identification of trans-Bering sisters (Chapter II, Supplementary Material 2.15). In those cases, it is possible that allopatric divergences might have been assigned to the incorrect node, and in such cases the age estimate for the node would not be associated with the trans-Bering migration event. The completeness of sampling was examined using the ratio of the described species diversity according to WoRMs and the number of total BINs available in BOLD for a specific taxon.
(Supplementary Material 3.5). The calibration process was repeated after excluding the trans-Bering sisters from a taxonomic group with poor sampling (<33%; Supplementary Material 3.5).

Biological traits from all trans-Bering sisters were considered during the calibration process when available. Deep (>50 m) vs. shallow distribution (≤50 m), habitat zone during adult stage (benthic vs. pelagic), thermal affinity (polar, temperate, or tropical), and developmental model (planktonic larvae vs. direct development) were the targets of attention. Specific traits might have allowed different dispersal routes, and the phylogenetic events may not be truly associated with the geological history of the Bering Strait. For example, taxa inhabiting deep waters are less likely to disperse during a shallow open Strait than taxa from the intertidal, especially as the Bering Strait is approximately 50 meters deep (Hopkins 1967). Likewise, taxa with a wide range of temperature tolerance are more likely to have alternative migration routes through tropical pathways (e.g. before the closure of the Isthmus of Panama) than taxa that exclusively inhabit polar waters. After considering the available information, clades that are more likely to be highly dispersive, or that could have used a different migratory route, were dropped from a secondary calibration process and compared with the main results (Supplementary Material 3.6). After integrating all information sources (e.g. fossil evidence and biological traits), the resulting date estimates were compared to the geological history.

Comparison with Trans-Bering Polychaetes and Echinoderms

Divergence time estimates from molluscs and arthropods were plotted and compared with prior trans-Bering sister taxa from echinoderms (Chapter II) and polychaetes (Carr 2010; Supplementary Material 3.7). Previous K2P divergences reported from trans-Bering polychaetes (Carr 2010) were re-examined using the iterative calibration approach, and the resulting times of divergence were mapped onto the timescale from this study.

Statistical Test

In accordance with the geological history of the Bering Strait, trans-Bering migrations are not expected to have occurred when the Strait was closed during the glacial maximum (2.4–3 Ma). After iterative calibration, I observed an absence of estimated dates in this range for three phyla, but for Mollusca one pair fell in this range, with a value of 2.97 Ma. To assess whether this observed number of divergences between 2.4–3 Ma would be expected solely by chance, a statistical test was created using R (R v3.3.3, R Core Team 2017) to compare the observed data with simulated data.

Simulated sets of divergence times between 0 and 7 Ma were randomly generated under
the uniform distribution. The upper date limit was selected to reflect the earliest possible date of
the first opening of the Bering Strait (7 Ma). The sample size of points generated for each data
set was selected to match the number of trans-Bering sisters used for the calibration in each
phylum, after omitting the pairs rejecting the clock hypothesis and/or that were beyond the
saturation threshold.

After generating 100K random sets of divergence times, the proportion of data sets
containing the number of observed values between 2.4–3 Ma was determined. This proportion
was interpreted as a p-value; i.e., how commonly did the simulated data contain a value as
extreme as the observed number of divergences between 2.4–3 Ma? The test was repeated
independently for all four phyla and for the entire dataset of all 73 trans-Bering sister taxa
across phyla (Supplementary Material 3.8).

RESULTS

Trans-Bering Sister Clades for Marine Molluscs and Arthropods

Ninety-three order-level NJ trees were constructed in BOLD, of which 44 orders were
retained for further analysis using ML. Trans-Bering sister clades were then identified after
examining the ML COI gene trees.

Within molluscs, 25 trans-Bering sisters were supported with high bootstrap support
(≥70%), while one node had low bootstrap support (60%) but was also supported from prior
literature (Reid et al. 1996), for a total of 26 pairs considered for calibration. The pairs were
phylogenetically diverse, falling within four classes, 15 orders, and 20 families (Supplementary
Material 3.1). Two trans-Bering sisters had low divergences (<2% average pairwise K2P
distance) and are most likely to be populations within species. The remaining 24 sister clades
were comprised of genetically divergent clusters and have been previously assigned to separate
species according to the BIN system (Ratnasingham & Hebert 2013).

Within arthropods, 29 trans-Bering sisters were supported with high bootstrap support
(≥70%). The pairs were phylogenetically diverse, falling within two classes, nine orders, and 23
families (Supplementary Material 3.1). Twenty-eight sister clades where comprised of
genetically divergent clusters and have been previously assigned as separate species
according to the BIN system. Only one trans-Bering sister is considered to be two populations
within a species according to the low divergence (<2% average pairwise K2P distance). Details
about order-level and family-level datasets and sequence lengths for molluscs and arthropods
Relative Rate Test on Trans-Bering Lineages

The molecular clock hypothesis was not rejected in most cases (p>0.05). Relative rates significantly differed (p<0.05) two times within molluscs and two times within arthropod pairs (Supplementary Material 3.3). Trans-Bering pairs showing evidence of rate heterogeneity were not included in the calibration process but were still plotted. P-values were not corrected for multiple testing in order to maintain higher sensitivity to detect rate variability, given the strict boundaries between the times when the Bering Strait was open vs. closed. All pairs would accept the clock hypothesis if a sequential Bonferroni correction is used.

Kimura 2-Parameter and Tamura-Nei (+Gamma) Genetic Distances for Trans-Bering Lineages

Mean COI K2P distances between trans-Bering sister clades ranged from 1.57% to 21.25% in molluscs and from 0.91% to 26.37% in arthropods (Fig. 3.1). On the other hand, the mean COI TN93+G distances between trans-Bering sisters ranged from 1.66% to 39.91% in molluscs and from 0.95% to 57.22% in arthropods. The wide range of K2P and TN93+G distances were observed within the multiple classes of each phylum.

Saturation Analysis

Within all four orders (Sacoglossa, Nudibranchia, Decapoda, and Isopoda), saturation of transversional substitution remained linear with the increase of K2P and GTR distances, but evidence of saturation of transitional substitutions began at approximately 16–18% K2P and GTR divergences (Supplementary Material 3.4). The threshold was based on the degree of linearity of the divergences vs. transitions plots, where the saturation of transitional substitutions was observed approximately at 16% K2P divergence. However, the evidence of saturation was pronounced at 17% K2P divergence (Supplementary Material 3.4); therefore, the threshold was considered at this point. Within molluscs only two trans-Bering sisters exceeded the 17% K2P divergence threshold while 10 arthropod trans-Bering sisters exceeded the divergence threshold. These 12 trans-Bering sisters were not used for the calibration, but were still plotted on the timescale.

Clock Calibration and Times of Divergence

Using K2P and TN93+G divergences separately, the target trans-Bering reference node for molluscs and arthropods was dated using the iterative calibration approach considering 121 calibration dates in sequence, and assigning divergence ages for all other pairs relative to the
genetic divergence of the reference node. The graphical representation of the iterative calibration approach shows evident gaps in divergence times for molluscs (Fig. 3.2) and arthropods (Fig. 3.3).

**Molluscs**— The reference node used during the calibration process had a divergence of 11.49% K2P and 17.86% TN93+G. According to K2P divergences, dating the reference node at 3.55 Ma provided the most concordant calibration after comparing the distribution of divergence ages from all trans-Bering pairs against the geological history of the Bering Strait (Supplementary Material 3.9). Divergence times for all trans-Bering sister clades ranged from 0.48 to 6.56 Ma (Fig. 3.4), while the inferred divergence rate was 3.2 % pairwise K2P sequence divergence per million years for the COI barcode region in northern marine molluscs (Supplementary Material 3.8). On the other hand, according to TN93+G divergences, calibrations in a range of 3.9 to 4.2 Ma for the reference node were concordant with the possible glacial history of the Bering Strait. Using the range of best calibration dates (3.9–4.2 Ma), divergence times for trans-Bering sisters ranged from 0.36 to 9.38 Ma (Fig. 3.4), while the inferred divergence rate was 4.3–4.6% pairwise TN93+G sequence divergence per million years for the COI barcode region in Northern molluscs (Supplementary Material 3.10).

**Arthropods**— The reference node used during the calibration process had a divergence of 11.81% K2P and 17.56% TN93+G. According to K2P divergences, dating the reference node to a range of 2.25 to 2.35 Ma yielded the most concordant calibration after comparing the distribution of divergences from all trans-Bering pairs against the geological history of the Bering Strait (Supplementary Material 3.11). According to this range of dates for the most concordant calibration, divergence times for trans-Bering sisters ranged from 0.17 to 5.24 Ma (Fig. 3.4), while the inferred divergence rate was 5.0–5.2% pairwise K2P sequence divergence per million years for the COI barcode region in northern marine arthropods (Supplementary Material 3.11). Similarly, when looking at TN93+G results and using the best range of calibration dates (2.05–2.4 Ma), divergence times for the trans-Bering sister clades ranged from 0.11 to 7.82 Ma (Fig. 3.4), while the inferred divergence rate was 7.3–8.6 % pairwise TN93+G sequence divergence per million years for the COI barcode region in Northern marine arthropods (Supplementary Material 3.12).

**Polychaetes**— Applying the iterative calibration approach to the K2P divergences previously reported by Carr (2010) showed that dating the reference node (15.97% pairwise K2P) to 3.4–4.5 Ma generated the most concordant calibration after comparing the distribution of divergences from all trans-Bering pairs against the geological history of the Bering Strait. The
results from the most concordant calibration suggested divergence times from 0.025 to 4.5 Ma, with an inferred divergence rate of 3.5–4.7% K2P sequence divergence per million years for trans-Bering polychaetes (Fig. 3.5 and Supplementary Material 3.13).

**Sensitivity Analysis: Considering Taxon Sampling and Biological Traits**

Information from the fossil record from the trans-Bering interchange was common for molluscs but absent for arthropods. Divergence times based on the iterative calibration were concordant with the fossil record (Vermeij 1991; Reid et al. 1996) for six trans-Bering sisters (Axinopsida, Ennucula, Ariadnaria, Neptunea, Margarites, and the most-divergent pair of Littorina) (Fig. 3.4) considering that fossils provide only minimum ages of clades (De Baets et al. 2016), thus supporting the accuracy of the results from the biogeography-based iterative calibration. However, for three other sister clades (Modiolus, Velutina, and the least-divergent pair of Littorina), the estimated divergence times from the iterative calibration were not concordant with the fossil record from the trans-Arctic interchange (Vermeij 1991; Reid et al. 1996).

Biological traits from trans-Bering sister clades were not always accessible (see Supplementary Material 3.6). The most commonly available information was about depth distribution (mapped in Fig. 3.4). From those trans-Bering pairs with available information, it was noticed that most pairs have planktonic larval development and a shallow distribution in both molluscs and arthropods. Although trans-Bering taxa have a temperature affinity from temperate to polar waters, only one taxon has been shown to tolerance freezing temperatures (Dendronotus frondosus) (Gionet & Aiken 1992). The biological zone distribution of trans-Bering molluscs and arthropods was mapped in Figure 3.5. Only four sister taxa of molluscs are pelagic, and the remaining 22 are benthic. Within arthropods, seven sisters are pelagic and 15 are benthic; information was unavailable for the other seven sister taxa. All trans-Bering sisters of echinoderms and polychaetes have benthic adults, so the calibration process for molluscs and arthropods was then repeated after excluding pelagic trans-Bering sisters. For molluscs, after excluding pelagic sisters in the calibration process, the resulting divergence rate was very similar to the result including all 26 pairs (3.28%/MY vs. 3.23%/MY) (Supplementary Material 3.14). However, for arthropods the resulting divergence rate after excluding pelagic sisters was lower than the rate including all 29 pairs (4.3–4.9%/MY vs. 5.0–5.2%/MY) (Supplementary Material 3.15).

Very poor taxon sampling (<33%) was evident in 20 of the 43 families in which trans-Bering clades were analyzed. The calibration process was therefore repeated after excluding
those trans-Bering sisters (Fig. 3.7). Within molluscs, seven trans-Bering sisters were from families with good (>66%) sampling, another seven from families with fair sampling (33–66%), and the remaining 12 sisters were from families with sparse taxon sampling (<33%). Pairs from poorly-sampled phylogenies were distributed across all ranges of genetic divergences, and results from the calibration process were similar to the results before eliminating trans-Bering sisters from sparsely-sampled phylogenies (Supplementary Material 3.16). Within arthropods, 11 trans-Bering sisters were from families with good (>66%) sampling, eight sisters from families with fair sampling (33–66%), and ten sisters were from families with highly incomplete sampling (<33%). Most of the pairs from poorly-sampled phylogenies have high genetic divergences, and results from the calibration process were similar to the results before eliminating trans-Bering sisters from sparsely-sampled phylogenies (Supplementary Material 3.17).

Statistical Test

In three phyla, trans-Bering divergences did not occur during the glacial maximum, between 2.4 and 3 Ma using the best calibration. Thus, the null distribution was compared against zero (i.e. the observed number of divergences between 2.4 and 3 Ma) for Echinodermata, Arthropoda and Polychaeta. In Mollusca, one trans-Bering divergence occurred at 2.97 Ma. Thus, the null distribution was compared against one (i.e. the observed number of divergences between 2.4 and 3 Ma) for Mollusca and for the entire dataset of all 73 sisters across phyla. For echinoderms, the sample size for each generated dataset was 15 hypothetical trans-Bering sisters, as 15 observed sisters were used for the calibration process after omitting the pairs rejecting the clock hypothesis and/or that were beyond the saturation threshold. The sample sizes for the other phyla were 22 pairs in molluscs, 17 pairs in arthropods, and 19 pairs in polychaetes.

The results of the statistical test indicate that divergence times between 2.4–3 Ma occur frequently within the randomly-generated datasets; i.e., it is not very likely to find an absence of divergence values during the glacial maximum (2.4–3 Ma) solely by chance. When the test was performed separately for each phylum, p-values were not significant. However, the test including the total sample size from four phyla yielded a highly significant p-value (p = 0.0097), indicating we would expect to observe a gap in estimated divergence times between 2.4–3 Ma only ca. 0.9% of the time by chance alone (Table 3.2).
DISCUSSION

Rate Estimation for Northern Marine Invertebrates

This chapter supports the efficacy of the iterative calibration approach for clock dating in marine invertebrates when using complex geological events. After integrating additional sources of evidence (e.g. fossil, depth) and considering the constrained periods of time when migrations through the Bering Strait did not occur, divergence rate estimation was feasible for molluscs and arthropods.

The results suggest a rate of 3.2% pairwise K2P and 4.3–4.6% pairwise TN93+G sequence divergence per million years for the COI barcode region in Northern marine molluscs (Table 3.1). On the other hand, for Northern marine arthropods a higher rate was suggested, 5–5.2% pairwise K2P and 7.3–8.6% pairwise TN93+G sequence divergence per million years for the COI barcode region (Table 3.1).

The higher divergence rate for TN93+G model, in comparison with the rate for K2P, is not surprising. Models with gamma distribution and invariant sites parameters have previously shown an elevated rate (e.g. Wilke et al. 2009; Papadopoulou et al. 2010). Results based on K2P distances indicate a faster rate in molluscs and arthropods than the rate reported in Chapter II for Northern echinoderms (2.8%/MY). Furthermore, the results suggest that trans-Bering arthropods have a slightly faster rate than Arctic polychaetes (5–5.2%/MY vs. 3.5–4.7%/MY in polychaetes), although the rate in polychaetes is faster than in molluscs (4.4%/MY vs. 3.2%/MY). Moreover, if the iterative calibration approach is applied to trans-Bering polychaetes using the K2P divergences reported by Carr (2010), it can be concluded that the first wave of dispersal across the Bering Strait occurred between 3.4 and 4.5 Ma, suggesting a rate of divergence of 3.5–4.7% K2P (Fig. 3.5 and Supplementary Material 3.13). The results from the iterative calibration applied to polychaetes (Table 3.1) are very similar to previous results assuming that the first wave of trans-Bering migrations in polychaetes occurred 3.5 Ma, which suggested the rate of 4.4% K2P divergences per million years (Carr 2010).

The COI divergence rates reported here for Northern marine molluscs (3.2% K2P/MY and 4.3–4.6% TN93+G/MY) and arthropods (5–5.2% K2P/MY and 7.3–8.6% TN93+G/MY) are similar to prior reports in other marine invertebrates, but contrast with a highly-cited divergence rate of 1.4%/MY (Knowlton & Weigt 1998) reported for tropical shrimps (further details about previous studies are available in Chapter I, Table 1.1). Rate variability between Northern taxa and tropical shrimps could be due to methodological problems, such as the variability in
taxonomic composition between datasets. In addition, biological differences would have an impact on rate estimations; thus, further investigation would be required.

Comparable to the results in this study, divergence rates between 3–6.58%/MY have been reported for marine molluscs (Wares & Cunningham 2001; Marko & Moran 2002; Luttikhuizen et al. 2003; Williams & Reid 2004; Miura et al. 2010), marine arthropods (Baldwin et al. 1998; Mathews & Anker 2009; Crandall et al. 2012), and echinoderms (Hart et al. 1997; Lessios et al. 1999; McCartney et al. 2000). By contrast, slower rates (0.16–2.6%/MY) have also been reported in the literature for marine invertebrates, including molluscs (Hellberg & Vacquier 1999; Marko 2002; Luttikhuizen et al. 2003; Miura et al. 2010; Crandall et al. 2012; Marko et al. 2014), echinoderms (Lessios et al. 2001; Foltz et al. 2008; Crandall et al. 2012), and arthropods (Schubart et al. 1998). Additionally, low mean divergence rates (2.36–2.48%/MY) were reported for five groups of marine invertebrates using various models of sequence evolution (Wilke et al. 2009). Interestingly, when the same data were analyzed using models with gamma distribution and invariant sites, the mean rates were higher, in a range of 3.14–3.52%/MY (Wilke et al. 2009) consistent with the higher rates observed when using TN93+G divergences in this study.

Commonly, slow rates in marine lineages have been described in studies using fossils for calibrating the molecular clock (e.g. Marko 2002; Luttikhuizen et al. 2003; Williams & Reid 2004; Marko et al. 2014), whereas most of the higher rates are from studies where the molecular clock was calibrated using coalescent methods or biogeographic calibrations (e.g. Wares & Cunningham 2001; Marko & Moran 2002; Mathews & Anker 2009; Marino et al. 2011; Hart 2012). The reliability of the “high” rates (>3%/MY) has been previously questioned, arguing that divergence events may predate geological events (e.g. Marko 2002; Luttikhuizen et al. 2003). However, here I argue that the difference in rates could be due to three main factors. First, the range of dates used for calibration (fossil vs. recent geological events) might be influencing the rate estimate according to the time-dependency hypothesis—i.e. a negative correlation between evolutionary rate estimates and calibration times (Ho et al. 2007, 2011, 2015a). Additionally, difference in rates might be due to the effect of saturation on COI, slowing the rate of accumulation of genetic variations. Lastly, assuming simultaneous divergence for all organisms associated with a biogeographic barrier would also have an impact on rate estimates.

The majority of previous studies assume simultaneous divergence of sister taxa, ignoring differences in the environmental tolerances, ecological requirements, dispersal ability, and that biogeographic events will impact different ecological groups at different times (De Baets et al. 2016). Most studies that used the Bering Strait for clock dating assumed divergence of trans-
Bering sister taxa around the first opening of the Bering Strait (e.g. Hrincevich et al. 2000; Wares & Cunningham 2001; Foltz et al. 2008; Gérard et al. 2008; Hallas et al. 2016) without considering the repeated opening and closure events and that multiple pulses of trans-Bering migrations have already been suggested (Carr 2010). If the calibration process is repeated in molluscs using a simplistic approach assuming simultaneous divergence of a cluster of points during the major trans-Arctic interchange, the calibrated rate estimate would vary between 2.7% and 3.7% per MY. Moreover, dating only one or two sister clades at 3.5 MY, a range of rates between 0.5%/MY and 6.1%/MY would be obtained. The majority of the rates obtained when using a simplistic approach are not concordant with the available fossil evidence from the major trans-Arctic interchange. In addition, applying those rates to estimate divergence dates for the rest of the sister clades would generate divergence times inconsistent with the geological history of the Bering Strait. Thus, using simplistic assumptions of simultaneous divergence can produce misleading estimates of divergence times. Therefore, I consider that using the iterative calibration approach provides more accurate results than when using prior simplistic methods and smaller sample sizes of pairs for clock dating. Here, a substantial number of pairs was sampled across a wide taxonomic range and without a priori assumption of simultaneous divergence during the major trans-Arctic interchange. Once a major gap in genetic divergences consistent with the geological history of the Bering Strait was found, only those trans-Bering clades passing the inclusion criteria were retained for the calibration process using the iterative calibration approach. According to the results from the statistical test (Table 3.2), the gap in divergence observed between 2.4–3 Ma is not very likely to occur solely by chance. Divergences between 2.4–3 Ma occur frequently in randomly-generated datasets; by chance, we would expect to observe a major gap, such as the one observed in the calibrations results, only ca. 0.9% of the time. Hence, the results illustrate temporal congruence with the Bering Strait glacial history.

Exploring biological traits from trans-Bering sister clades was essential to elucidate the effect of the geological events upon organisms with different traits and possible alternative dispersal routes. In the case of echinoderms and polychaetes, all trans-Bering sisters have benthic adults. However, within molluscs and arthropods, several sisters are pelagic. Evolutionary rates of pelagic taxa might be affected by the mutagenic effect of UV radiation, which is expected to be higher in pelagic taxa compared to benthic taxa (DeWaard 2004). When the calibration process was repeated after excluding trans-Bering sisters with pelagic adults (Fig. 3.6), the results were very similar for molluscs but not for arthropods. In molluscs, the divergence rate based on benthic sisters was 3.3%/MY (vs. 3.2%/MY when including all 26
pairs) (Supplementary Material 3.14). In contrast, when pelagic sisters were excluded from the arthropods calibration, the resulting divergence rate was lower than the rate including all 29 pairs (4.3–4.9%/MY vs. 5.0–5.2%/MY) (Supplementary Material 3.15); suggesting that the inclusion of pelagic taxa might overestimate the molecular rate for benthic arthropods. Pelagic arthropods are often transparent or translucent and inhabit shallower waters; thus, UV exposure might have a greater effect on the mutation rate. Both results should be considered especially when comparing rates across studies in order to achieve more accurate conclusions.

Rate Estimation in Northern Marine Invertebrates vs. Rates in Tropical Marine Invertebrates

Faster rates of molecular evolution and speciation processes have been suggested for lineages inhabiting warmer climate regions (e.g. Rohde 1992; Bargelloni et al. 1994; Bleiweiss 1998b; Gillooly et al. 2005; Wright et al. 2006). When comparing molecular rates from previous studies in tropical taxa (Table 1.1) vs. Arctic rates from this study, there is no clear pattern. The divergence rate for Arctic molluscs is comparable with rates reported previously for tropical taxa, and it was also congruent with the fossil record of the Bering region. However, the divergence rate from Northern echinoderms is lower than most rates for tropical echinoderms reported in the literature. By contrast, the rate reported here for Northern arthropods is higher than all prior rates from tropical marine arthropods. Furthermore, comparison of divergence rates across studies becomes challenging due to the disparities in methodology, the use of different models of nucleotide substitution, and due to the disagreement in the dates of the biogeographic events used for calibration (see Table 1.1). In an effort to evaluate tropical vs. Arctic relative rates, Orton et al. (unpublished) have used a consistent gene region and methods to analyze more than eight thousand phylogenetic pairs of latitudinally-separated BINs spanning six animal phyla. Overall, there was only a weak trend of higher rates at lower latitudes, but the strongest latitudinal pattern was found in echinoderms (Orton et al. unpublished), mirroring the results reported here for absolute date estimates. Therefore, the geographic scope of the calibrations presented here requires further investigation, but a correction for latitude may be necessary for echinoderms.

Phylogenetic Scope for the Substitution Rates Presented in this Study

Substitution rates for the COI-5P gene presented in this study should not be generalized due to rate variability across taxa and genes (Ho & Duchêne 2014). It has been revealed that COI evolves more slowly than other mitochondrial genes (e.g. Pons & Vogler 2005; Lavinia et al. 2015); therefore, it would be inappropriate to assume that the rates suggested here for COI apply to other mitochondrial genes. As previously mentioned, evolutionary rates are not
constant over time (Ho et al. 2005, 2011). Therefore, the substitution rates presented here might not be appropriate for dating events far outside the calibration time frame (2-7 MY) used here (i.e. recent demographic events and extremely deep nodes in the tree of life). The present results are most appropriate for research involving Northern marine lineages. However, it is important to consider that they might not represent the appropriate evolutionary rate for other marine groups.

Divergence Times for Northern Marine Invertebrates

The observed wide range of genetic divergences between trans-Bering sisters in molluscs and arthropods agrees with multiple pulses of trans-Bering migrations suggested for other Northern marine invertebrates such as polychaetes (Carr 2010) and echinoderms (see Chapter II). Although rate heterogeneity across taxonomic groups cannot be fully rejected, it is not the main explanation for the broad genetic divergence range in trans-Bering marine invertebrates (see Chapter II). Thus, multiple pulses of trans-Bering migrations are evident in all four phyla (Fig. 3.5).

After exploring alternative calibration scenarios derived from previous literature, it is clear that using the iterative calibration approach yielded an outcome more concordant with the geological history of the Bering Strait than former studies. For example, the trans-Bering pair Acanthodoris was previously used for clock calibration to estimate divergence times for the genus (Hallas et al. 2016). Using Bayesian analysis, Hallas and collaborators (2016) dated the divergence of Acanthodoris trans-Bering sisters at 5.32 Ma with the first opening of the Bering Strait. The authors also included an additional node calibration based on the formation of the Baja Peninsula. The results suggested a divergence time of approximately 2.4 Ma for the trans-Bering pair Acanthodoris (Hallas et al. 2016), which differs from the time of divergence based on the results presented here (1.2–1.4 Ma). If the trans-Bering pairs from this chapter were dated using the divergence time previously described for Acanthodoris (2.4 Ma; Hallas et al. 2016), the results would suggest a range of divergences of 0.8–10.86 Ma for the 26 trans-Bering pairs of molluscs in this study, implying that divergence of most trans-Bering pairs pre-dated the first opening of the Bering Strait. Despite temporary connections between Pacific and Arctic Oceans across Eastern Siberia might have occurred since the Early Miocene (approx. 17 Ma; Polyakova 2001), that inference has been strongly questioned in the literature arguing that it is based on weak evidence and poorly preserved diatom fossils (see details in Marincovich & Gladenkov 2001 and Gladenkov & Gladenkov 2004); thus, I consider that trans-Bering migrations before 5.5 Ma are highly improbable.
There is some disagreement between the time scale in this study and the fossil record associated with the major trans-Arctic interchange. That is the case of one trans-Bering pair from the genus *Littorina*. The divergence of both *Littorina* trans-Bering clades has previously been linked to the major trans-Arctic interchange according to their fossil record from the Late Pliocene in the Atlantic (Reid *et al.* 1996). While the divergence time estimate from the iterative calibration of one *Littorina* trans-Bering pair (3.55–3.9 Ma) is concordant with its fossil record, the genetic divergence of the other pair would have occurred long after the major trans-Arctic interchange. Disagreement between genetic divergences and the fossil record in the less-divergent trans-Bering pair of *Littorina* could be due to an anomalous rate related to clade-specific biological factors. For example, taxa from the less-divergent *Littorina* sister clade are nonplanktotrophic or ovoviviparous, while the rest of the species in the genus have a planktotrophic developmental mode (Reid *et al.* 1996). Another possible explanation for the disagreement between the fossil evidence and genetic divergences is that the calibration presented here does accurately date the timing of the split of the sampled lineages, but that the mitochondrial lineage arising from the trans-Bering interchange has not be sampled.

Additionally, there is evidence of a high rate of extinction in Northern Pacific and Atlantic molluscs (Stanley 1986a; b, Vermeij 1989a; b), and in some cases, local extinctions were followed by subsequent colonization events (e.g. Palumbi & Kessin; Azuma *et al.* 2017). Therefore, it is possible that lineages may have gone extinct, and the extant populations of this morphospecies may be descended from a more recent migration event. These complexities also emphasize the value of considering numerous sister pairs for calibration, such that a general signal can emerge from the data.

**Alternative Scenarios for Trans-Bering Migration**

For both Arctic marine molluscs and arthropods, a single preferred calibration date assigned to the reference node was suggested when using the iterative calibration and K2P divergences. In molluscs, dating the reference node (11.49% K2P) at 3.55 Ma provided the most concordant calibration. On the other hand, the most concordant calibration in arthropods was achieved by dating the reference node (11.81% K2P) in a range of 2.25 to 2.35 Ma. Similarly, for both groups, a single preferred calibration was supported when looking at TN93+G results. In molluscs, the most concordant calibration resulted when dating the reference node (17.86% TN93+G) in a range of 3.9 to 4.2 Ma. In arthropods, dating the reference node (17.56% TN93+G) in a range of 2.05–2.4 Ma yielded the most concordant calibration. However, alternative scenarios could be explored, though they are less likely, as explained below.
For molluscs, two alternate calibrations assigned to the reference node (5.6–5.85 MY and 3.9 MY) could also be inferred from the K2P results. Assigning a calibration date of 3.9 MY to the reference node would imply that one trans-Bering pair migrated across the Strait at approximately 2.66 Ma. Although this scenario is less probable, it cannot be fully ruled out since *Dendronotus frondosus* is known to tolerate freezing temperatures (Gionet & Aiken 1992), and as an intertidal species it might have migrated through rare water channels. A less-likely scenario is assigning a calibration date in the range of 5.6–5.85 MY to the reference node. In that case, the divergence of 10 sister pairs would have predated the first opening of the Bering Strait, with divergences as early as 10.8 Ma.

For arthropods, one alternate calibration date assigned to the reference node could be suggested by the K2P results (6.05 MY) and another one by the TN93+G results (4.5 MY). Both scenarios are highly unlikely since assigning a calibration date of 6.05 MY to the reference node would imply the divergence of 15 sister pairs before the first opening of the Bering Strait with divergences as early as 13.5 Ma. Similarly, assigning a calibration date of 4.5 MY would imply the divergence of 13 sister pairs before the first opening of the Bering Strait, with divergences as early as 14.66 Ma.

**Future Directions**

A future direction would be to explore further the rate variability among lineages in this study, first by testing for rate heterogeneity across entire trees (e.g. using Bayesian analysis) and, second, by examining rate heterogeneity in relation to biological traits. This research would also benefit from estimating confidence intervals on the calibration itself. In addition, it would be valuable to estimate divergence rates using other substitution models in order to correct possible underestimations of genetic distances and to allow further comparisons across taxonomic groups. Furthermore, it would be beneficial to complement COI-5P with additional genes to assess the consistency of rates across genes. Another promising future direction would be to expand the research to other taxonomic groups, including those lineages without a fossil record and that would benefit from clock calibrations based on biogeographic events.

**CONCLUSION**

This study provides strong evidence of multiple pulses of trans-Bering migrations in all four groups of marine invertebrates investigated (Fig. 3.4). The results presented here suggest a divergence rate for COI-5P of 2.8% K2P/MY in echinoderms, 3.2% K2P/MY in molluscs, 3.5–4.7% K2P/MY in polychaetes, and 5.0–5.2% MY in arthropods. Using a substantial number of
sister clades—91 in total—together with my iterative calibration approach, this chapter presents important advantages over previous calibrations. Divergence rates reported here are higher than highly-cited molecular rates (e.g. Knowlton & Weigt 1998; see Table 1.1). The prevalent use of the low divergence rate reported for *Alpheid* shrimps (1.4%/MY) could have led to a systematic underestimation of rate for dating phylogenies in the marine realm. Lastly, dating evolutionary events accurately is essential for understanding the impacts of prior climatic changes upon the history of life.
### Table 3.1
Summary of molecular clock calibrations for the mitochondrial cytochrome c oxidase subunit I gene (COI) in the four groups of Northern marine invertebrates used in this study.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Total number of pairs plotted</th>
<th>Total number of pairs passing the inclusion criteria</th>
<th>Range of divergences among pairs used for calibration (% K2P)</th>
<th>Divergence rate estimate (% K2P per MY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echinodermata</td>
<td>16</td>
<td>15</td>
<td>0.45–15.45%</td>
<td>2.8%</td>
</tr>
<tr>
<td>Mollusca</td>
<td>26</td>
<td>22</td>
<td>1.57–16.87%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>29</td>
<td>17</td>
<td>0.91–16.67%</td>
<td>5.0–5.2%</td>
</tr>
<tr>
<td>Polychaeta</td>
<td>20</td>
<td>19</td>
<td>0.12–15.97%</td>
<td>3.5–4.7%</td>
</tr>
</tbody>
</table>

### Table 3.2
Summary of the statistical test results for the four groups of Northern marine invertebrates used in this study. The test assesses how frequently randomly-generated data sets contain more than the observed number of divergence times during the glacial maximum period, 2.4–3 Ma.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number of pairs included in the calibration &amp; sample size per data set in the randomization test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echinodermata</td>
<td>15</td>
<td>0.260</td>
</tr>
<tr>
<td>Mollusca</td>
<td>22</td>
<td>0.286</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>17</td>
<td>0.216</td>
</tr>
<tr>
<td>Polychaeta</td>
<td>19</td>
<td>0.182</td>
</tr>
<tr>
<td>All taxa</td>
<td>73</td>
<td>0.00972</td>
</tr>
</tbody>
</table>
Figure 3.1 Kimura two-parameter (K2P) divergences for trans-Bering sister clades of echinoderms, molluscs, arthropods, and polychaetes. Genetic divergences are colour-coded by class in each group.
Figure 3.2 Iterative calibration approach for molluscs using K2P divergences. Dots represent divergence date estimations in Ma (X axis). Each horizontal row of points represents a different calibration attempt (Y axis, right) using a different calibration date and the divergence rate generated in K2P%/MY (Y axis, left). For visualization purposes, 62 of the total 121 calibration attempts are presented in this figure (for further details see Supplementary Material 3.9). All calibration attempts were compared with the geological history of the Bering Strait to find congruence between the genetic and geologic evidence. The boxes with darker red shading show the two periods of time when trans-Bering migrations were extremely unlikely; the light red shading (5.5-7 Ma) shows the period of time where the Bering Strait is considered to be closed, after the earliest possible opening was narrowed to 5.5 Ma. The solid black arrow indicates the most feasible calibration (calibration attempt 36) that is concordant with the geological evidence. The points falling within the shading areas in the most feasible calibration failed the relative rate test or are above the saturation threshold.
Figure 3.3 Iterative calibration approach for arthropods using K2P divergences. Dots represent divergence date estimations in Ma (X axis). Each horizontal row of points represents a different calibration attempt (Y axis, right) using a different calibration date and the divergence rate generated in K2P%/MY (Y axis, left). For visualization purposes, 62 of the total 121 calibration attempts are presented in this figure (for further details see Supplementary Material 3.11). All calibration attempts were compared with the geological history of the Bering Strait to find congruence between the genetic and geologic evidence. The boxes with darker red shading show the two periods of time when trans-Bering migrations were extremely unlikely; the light red shading (5.5-7 Ma) shows the period of time where the Bering Strait is considered to be closed, after the earliest possible opening was narrowed to 5.5 Ma. The solid black arrow indicates the most feasible calibration (calibration attempt 49) that is concordant with the geological evidence. The dashed grey arrow marks a calibration that would yield a divergence rate of 2.46%/MY (as in Wilke et al. 2009); the dashed black arrow marks the calibration using the rate of 1.4%/MY (Knowlton & Weigt 1998). Both of those rates were previously reported for Alpheus shrimps.
Figure 3.4 Divergence time estimates for trans-Bering molluscs and arthropods using K2P and TN93+G distances and the best calibration according to the iterative calibration approach. The figure shows the reference node used for time scaling, which accepted the clock hypothesis, had a divergence estimate below the saturation threshold, was from a well-sampled phylogeny and, in the case of molluscs, had a fossil record associated with the trans-Arctic interchange. Trans-Bering sister clades that were used for calibration and the sister clades that were omitted during the calibration process are also indicated. Available information from the fossil record associated with the major trans-Arctic interchange is also indicated. The boxes with dark grey shading show the two periods of time when trans-Bering migrations were extremely unlikely; the light grey shading (5.5-7 Ma) shows the period of time where the Bering Strait is considered to be closed, after the earliest possible opening was narrowed to 5.5 Ma.
Figure 3.5 Divergence time estimates for Northern marine echinoderms, molluscs, polychaetes, and arthropods using K2P divergences and the best calibration according to the iterative calibration approach. The shaded boxes show the two periods of time when trans-Bering migrations were extremely unlikely, during the maximum glacial (2.4–3 Ma) and before the earliest possible opening of the Bering Strait at 5.5 Ma.
Figure 3.6 Divergence time estimates for molluscs and arthropods using K2P divergences based on the best calibration and mapping the biological zone. The figure shows the trans-Bering sister clades with benthic and pelagic adults. The reference node used for time scaling is marked with a star. Trans-Bering sister clades that were used for calibration and the sister clades that were omitted during the calibration process are also indicated. The shaded boxes show the two periods of time when trans-Bering migrations were extremely unlikely, during the maximum glacial (2.4–3 Ma) and before the earliest possible opening of the Bering Strait at 5.5 Ma.
Figure 3.7 Divergence time estimates for molluscs and arthropods based on the best calibration using K2P divergences and mapping the completeness of the sampling. The reference node used for time scaling is marked with a star. The figure shows the trans-Bering sister clades from families with fair to good sampling (>33%) that were used for calibration and the sister clades that were omitted during the calibration process due to poor sampling (<33%). The shaded boxes show the two periods of time when trans-Bering migrations were extremely unlikely, during the maximum glacial (2.4–3 Ma) and before the earliest possible opening of the Bering Strait at 5.5 Ma.
CHAPTER IV
New Molecular Clock Calibrations from the Isthmus of Panama using the Iterative Calibration Approach

ABSTRACT

The formation of the Isthmus of Panama caused vicariance between Pacific and Atlantic marine populations. Sister species on either side of the isthmus have been widely used for calibrating the molecular clock in marine lineages, and molecular rates based on trans-isthmus sisters have been extensively applied in the literature. Nevertheless, important uncertainties with calibrations derived from the isthmus still persist. Here, I expand the use of the novel iterative calibration approach to the Panama system, incorporating the full timeline of the isthmus formation process for clock dating. Divergence rates in tropical marine echinoderms, molluscs, and arthropods were examined after re-analyzing publicly available Kimura two-parameter (K2P) divergences of the cytochrome c oxidase subunit I (COI) gene. By integrating genetic and biogeographic evidence, and using a substantial number of sister clades (66 total trans-isthmus sisters) during the calibration process, I anticipate more accurate results than those generated during previous work using simplistic assumptions. Additionally, absolute rates of divergence were compared between tropical and Arctic marine invertebrates. The results indicate a rate of K2P divergence of 2–2.1%/MY in echinoderms, 2.3–2.5%/MY in molluscs, and 1.8–1.9%/MY in arthropods. These rates contrast with the rates reported for Arctic marine invertebrates and suggest that molecular rates from tropical taxa are generally slower than rates from Arctic taxa, perhaps due to differences in effective population size between these regions. The results presented here provide an advance for dating evolutionary events in the marine realm and support previous evidence of rate variability across the tree of life. Additionally, the comparison between tropical and Arctic lineages provides novel and improved calibrations, which are important for revealing how prior climatic events have shaped the diversity and distribution of life.
INTRODUCTION

The formation of the Isthmus of Panama through the late Pliocene (2.8 Ma; Coates et al. 2005; Lessios 2008; O’Dea et al. 2016) fragmented the distribution of many marine species into Atlantic and Pacific populations. Sister species separated by the closure of the isthmus are commonly known as geminate species (Jordan 1908) and have become an important tool for estimating divergence rates and patterns of molecular evolution in marine taxa. Genetic divergences between trans-isthmus sisters have been widely used in molecular clock research based on the observation that rates of molecular evolution are relatively constant over time (molecular clock hypothesis; Zuckerkandl & Pauling 1965a; b) and that divergence of geminate species occurred during the final closure of the isthmus (e.g. Bermingham & Lessios 1993; Knowlton et al. 1993; Bermingham et al. 1997; Knowlton & Weigt 1998; Schubart et al. 1998; Miura et al. 2010).

Although there has been considerable research, several important uncertainties remain with calibrations derived from the isthmus. Using the iterative calibration approach, the novel method for clock dating developed in Chapter II, here I will approach some of those uncertainties aiming for more accurate results than those presented in previous studies. The balance of this introduction reviews the geological history of the Isthmus of Panama and previous molecular clock studies based on the formation of the isthmus, prior to introducing how I will conduct revised calibrations using the new approach for clock dating.

A Brief Geological Time Scale of the Isthmus of Panama Formation

The formation of the Isthmus of Panama was a slow process that lasted approximately 12 MY, affecting the habitat and distribution of marine organisms through the entire process. In the following paragraphs I briefly describe the geological history of the formation of the Isthmus of Panama according with current available knowledge.

Previous to the formation of the isthmus, between 23–25 Ma, the subduction of the Pacific-Farallon Plate under the South American and Caribbean plates yielded the development of a volcanic arc on the edge of the Caribbean Plate (O’Dea et al. 2016). The volcanic arc impacted on South America, generating a moderate but constant uplift that has been ongoing ever since and that was essential for the formation of the Isthmus of Panama (Farris et al. 2011; Coates 2013). The continuous impact of the volcanic arc over South America led to a transition from deep waters to sedimentation in both Pacific and Atlantic regions at about 15 Ma, initiating the formation of the isthmus (Coates et al. 2004).
The seaway between the Americas started narrowing and shallowing by the middle Miocene (12–13 Ma), with an ultimate interruption of the deep-water connections by 9.2–12 Ma (Coates et al. 2005; Leigh et al. 2014; Osborne et al. 2014). The continuous shoaling began to split deeper-water populations around 7 Ma, when the water temperature, salinity, and carbonate contents of the seabed sediments began to increase in the southern Caribbean (Leigh et al. 2014). The rising of the Panama Arc began again at about 6 Ma, and it has continued to do so through the present (O’Dea et al. 2016).

Major ecological changes between the Pacific Ocean and Caribbean Sea started at about 4.6 Ma (Keigwin 1982; Haug et al. 2001). At that point, surface water salinity began to increase drastically until reaching current values at approximately 4.2 Ma (Keigwin 1982; Haug et al. 2001). There is also evidence of coastal upwelling, shoaling of the Caribbean-Pacific sill, and the increase of the eastern Pacific thermocline temperature at around 4.2 Ma, suggesting that the Panama Arc was mostly emergent at that time (Keigwin 1982; Keller et al. 1989; Billups et al. 1999; Chaisson & Ravelo 2000; Haug et al. 2001; Schneier & Schmittner 2006; Kamikuri et al. 2009; Groeneveld et al. 2014). Until that time, there was little taxonomic or ecological difference in shelf benthic and nektonic communities between the Pacific and Caribbean, indicating the movement of water carrying larvae or adults between oceans (Jackson et al. 1993; Landau et al. 2009; Aguilera Socorro et al. 2011).

The continuous uplift of the Panamanian Arc, followed by the drop of the sea level due to the Pleistocene glaciations, triggered the establishment of the land bridge connecting North and South America at about 3 Ma (Coates et al. 2004; Bartoli et al. 2005; Coates 2013), with the closure of the Isthmus of Panama sensu stricto at 2.8 Ma (Coates et al. 2005; Lessios 2008; O’Dea et al. 2016). Some researchers, however, have pointed out that the final closure of the isthmus did not occur until 2.7 Ma (Leigh et al. 2014). A subsequent water connection between the Pacific and Caribbean might have occurred at around 2–2.45 Ma (Cronin & Dowsett 1996; Groeneveld et al. 2014). The evidence of similar salinity and sea surface temperature between both oceans has suggested the possible temporary breaching of the isthmus, possibly due to the rise in sea level during the interglacial periods of the Pleistocene (Cronin & Dowsett 1996; Groeneveld et al. 2014).

The changes in oceanic productivity during the formation of the isthmus had a great effect on marine biota. As the water flow from the Pacific to the Caribbean became restricted, there was a decline in nutrient concentrations in the Caribbean. Thus, organisms suited to high planktonic productivity were subjected to strong negative selection pressure (Allmon 2001;
O’Dea et al. 2007; O’Dea & Jackson 2009), resulting in a regional extinction event in the Caribbean between 2 to 4 Ma (Jackson & O’Dea 2013). On the other hand, the formation of the isthmus allowed the exchange of terrestrial organisms across the forming land bridge. The major exchange of mammals between North and South America has been dated at 2.6–2.7 Ma, although some mammal migrations occurred before 3 Ma (Molnar 2008).

*Calibration of the Molecular Clock Using the Isthmus of Panama*

Molecular clock calibrations based on the divergence of trans-isthmus marine species are common in fishes (Gorman & Kim 1977; Vawter et al. 1980; Bermingham et al. 1997; Donaldson & Wilson 1999; Eytan & Hellberg 2010; Lessios & Robertson 2013), but predominate for marine invertebrates such as echinoderms (Hart et al. 1997; Lessios et al. 1999, 2001; McCartney et al. 2000; Lessios 2008; Coppard et al. 2013), molluscs (Hellberg & Vacquier 1999; Marko & Moran 2002; Miura et al. 2010), and arthropods (Baldwin et al. 1998; Knowlton & Weigt 1998; Schubart et al. 1998; Williams & Knowlton 2001; Morrison et al. 2004; Mathews & Anker 2009). Using the isthmus for calibration, molecular rates for marine invertebrates have been reported that range from 1.4–6.23% divergence per MY (Table 1.1).

From the wide range of divergence rates reported for tropical marine invertebrates (Table 1.1), the rate most frequently cited in the literature is from alpheid shrimps (Knowlton & Weigt 1998). These results revealed high variability in genetic divergences across trans-isthmus sisters, where the most divergent sister pairs were those restricted to offshore islands or deeper habitats. Moreover, the only two pairs from mangrove environments, which were likely to be the last marine habitats separated by the formation of the isthmus, were in fact the least-divergent pairs (Knowlton and Weigt 1998). Therefore, according to the authors, the lowest genetic divergence provides a better calibration point using the final closure of the isthmus (3 Ma) (Knowlton and Weigt 1998). The oft-cited rate of 1.4% COI divergence per MY is thus calibrated from a single pair of trans-isthmus arthropods, after considering habitat and discarding deeper divergence values (Knowlton and Weigt 1998).

Although molecular clock research using trans-isthmus sisters is vast, ambiguities and problems still persist. Potential sources of error are: i) the use of different dates for the closure of the isthmus, ranging from 1.8 (e.g. Lessios et al. 2001) to 3.1 Ma (e.g. Hart et al. 1997); ii) most studies assumed the simultaneous divergence of trans-isthmus species when calibrating the molecular clock (e.g. Hickerson et al. 2003; Lessios 2008); iii) the use of different genes for rate estimation (e.g. Tringali et al. 1999; Morrison et al. 2004; Robles et al. 2007); iv) only a few studies (e.g. Knowlton & Weigt 1998; Williams & Knowlton 2001; Miura et al. 2010) considered
biological and ecological differences among taxa and that they might have been affected in a different way by the vicariance event; v) most studies did not test for rate heterogeneity; vi) or for saturation effects (e.g. Hart et al. 1997; Lessios et al. 1999; Lessios 2008). Furthermore, there is a lack of homogeneity in the available data, as different regions have been used within the studied gene. Additionally, not all sequences used for divergence calculations (e.g. Lessios 2008) are available in public sequence databases. Thus, it is difficult to compare results across studies, and a complete re-analysis of the entire data on trans-isthmus species, including testing for rate constancy and saturation, is not feasible. Here, I will address concerns i to iv to generate more realistic calibrations using the Panama system.

In this chapter, I apply the iterative calibration approach for the re-analysis of trans-isthmian taxa (Supplementary Material 4.1), expanding from research using the Bering Strait (Chapters II and III). The use of the iterative calibration will allow for the complete geological history of the isthmus formation to be taken into account through the calibration process. Furthermore, this approach considers divergence values from all available sister pairs. Therefore, with a more realistic consideration of the geological event, I anticipate more accurate estimates of divergence rates for trans-isthmian sister clades than presented in previous studies. Furthermore, in this chapter I will compare absolute rates of divergence between marine invertebrates from Arctic (Chapters II and III) and tropical (this chapter) environments. Overall, this chapter will advance our ability to use molecular clocks to date evolutionary events for marine lineages.

METHODOLOGY

Data Collection

Genetic divergences from the mitochondrial cytochrome c oxidase gene (COI), including both the 5P and 3P regions, previously reported from trans-isthmus sister clades of marine invertebrates were retrieved from the literature and re-examined using the iterative calibration approach. Specifically, Kimura two-parameter (K2P) divergences from trans-isthmus echinoderms, molluscs, and arthropods were re-analyzed (Supplementary Material 4.2). The main sources of information were two studies synthesizing data from earlier research (Lessios 2008; O’Dea et al. 2016), but additional data were obtained from a further source (Hart et al. 1997).
Establishing the Geological Time Scale of the Isthmus of Panama Formation

The completion of the Isthmus of Panama is usually considered to have occurred at 2.7–3 Ma (e.g. Knowlton & Weigt 1998; Lessios 2008; Wilke et al. 2009; Miura et al. 2010; Coppard et al. 2013; Lessios & Robertson 2013). However, it was a slow process that lasted approximately 12 MY (Coates et al. 2004; Lessios 2008). The geological history the isthmus formation is summarized in Table 4.1, and it was incorporated during the calibration process using the iterative calibration approach. Then, in relation to the series of events (i.e. the transition between deep to shallow waters, ecological differences between Caribbean and Pacific, and the formation of the land bridge), the probability of dispersal between the Pacific and Caribbean was evaluated and classified as high dispersal, low dispersal, limited, and null dispersal. Periods of time where there is high uncertainty of the events were also mapped (Table 4.1).

Iterative Calibration of Molecular Rates Using the Closure of the Isthmus of Panama

The iterative calibration approach was applied to estimate divergence times for echinoderms, molluscs, and arthropods separately using K2P divergences. First, the pair with the smallest divergence was selected as the reference node for time scaling. Then, a starting calibration date (1.5 Ma) was assigned to the reference node following the process described in Chapter II. Subsequently, times of divergence for the remaining pairs were manually assigned in relation to the reference node. The divergence time estimates from all trans-isthmus sister clades were graphed and then compared against the geological history of the Isthmus of Panama closure. Using the same principle, I tested 136 calibration dates for each phylum in the range between 1.5 to 15 Ma (stepping by 0.1 Ma).

According to the literature, the completion of the Isthmus of Panama occurred sensu stricto around 2.8 Ma (Coates et al. 2005; Lessios 2008; O’Dea et al. 2016), with possible temporary breaching of the isthmus between 2–2.45 Ma (Cronin & Dowsett 1996; Groeneveld et al. 2014). Limited marine dispersal through water channels may have been possible between the Pacific and Caribbean around 2–2.45 Ma and is therefore permitted during the iterative calibration. However, calibration attempts generating divergence times between 2.45–2.8 Ma and <2 Ma are considered highly improbable, as marine migrations would have not occurred when the isthmus was fully completed and assuming that terrestrial or aerial dispersal did not occur.

The most likely calibration date was chosen after finding concordance between divergence date estimates for trans-isthmus species pairs and the Isthmus of Panama.
geological time scale. Then, using the most probable calibration date, the K2P divergence rate for each major group was estimated. The divergence rate \( r \) for the reference node was estimated as \( r = \frac{D}{T} \), where \( D \) is the percentage of genetic distance and \( T \) is the time of divergence according to the best calibration obtained from the iterative calibration approach considering all pairs.

**Sensitivity Analyses: Taxon Sampling and Biological and Ecological Traits**

The sensitivity of the results to taxon sampling and variability in biological traits was also examined. Highly incomplete phylogenies, due to poor taxon sampling, can influence the results. It is possible that allopatric divergence events might have been assigned to the incorrect node, and in such cases the age estimate for the node would not be associated with the trans-isthmus vicariance event (Chapter II, Supplementary Material 2.15). The completeness of the sampling was examined at the family level using the ratio of the described species diversity according to the World Register of Marine Species (WoRMs; http://www.marinespecies.org; last accessed May 4th, 2017) and the number of total Barcode Index Numbers (BINs) (Ratnasingham & Hebert 2013) available in the Barcode of Life Data Systems v4 (BOLD; http://v4.boldsystems.org; last accessed May 4th, 2017) (Ratnasingham & Hebert 2007) for a specific taxon (Supplementary Material 4.3). BOLD was used rather than GenBank because BOLD mines sequence data from GenBank and is thus more inclusive specifically for barcode data. The calibration process was repeated after excluding the trans-Bering sisters from taxonomic groups with poor sampling (<33%; Supplementary Material 4.3).

Biological traits from all trans-isthmus sisters were also considered in the analysis when available. Deep (>50 m) vs. shallow distribution (≤50 m), thermal affinity (tropical, subtropical, or temperate), developmental mode (planktonic larvae vs. direct development), and habitat zone (pelagic vs. benthic) were the target of attention. Information about biological traits was primarily obtained from online databases such as WoRMS and SeaLifeBase (http://www.sealifebase.org; last accessed May 4th, 2017). Specific traits might have allowed different dispersal patterns or routes, and the phylogenetic splitting events may not be truly associated with the final closure of the isthmus. For example, taxa inhabiting shallow waters, intertidal pools, or mangroves are more likely to be the last groups to migrate during the final closure of the isthmus, while populations of taxa from deep waters would have been separated earlier. Likewise, taxa with a wide range of temperature tolerance are more likely to have had alternative migration routes available to them after the completion of the Isthmus of Panama than taxa that exclusively inhabit warm tropical waters. After considering the available information, clades that are more
likely to be highly dispersive, or that could have used a different migratory route, were dropped from a secondary calibration process and compared with the main results. After integrating all information sources (e.g. biological and ecological traits), the resulting date estimates were compared to the geological history.

Comparison of Absolute Rates between Tropical and Arctic Lineages

After applying the iterative calibration, the resulting absolute rates of molecular evolution for COI in tropical marine invertebrates were compared with prior research on Arctic marine invertebrates (Chapters II and III). The comparison is feasible because the biogeographic events used for calibration in both systems occurred at a similar time, which is expected largely to cancel the effect of time dependency (Ho et al. 2011, 2015a) on the rate estimates. Additionally, the taxonomic composition at the class level and habitat of trans-barrier pairs were compared between the tropical and Arctic sisters used for calibration.

RESULTS

Trans-Isthmus Sister Clades and Kimura 2-Parameter Genetic Distances for Tropical Marine Invertebrates

A total of sixty-six trans-isthmian sister clades have been reported in the literature from which K2P divergences for the COI gene were available and therefore included here. Within echinoderms, 11 trans-isthmus sister clades were included in the analysis, with mean COI K2P distances ranging from 4.5% to 18.7%. Two trans-isthmus sisters had the lowest divergences (4.5 and 4.8%), while the remaining nine sisters had larger divergences of ≥8.7% K2P (Fig. 4.1). For the majority of echinoderms sisters, the available genetic distances correspond to the 3’ end of the COI gene. The echinoderm pairs were phylogenetically diverse, failing within two classes, seven orders, and eight families. Within molluscs, 28 trans-isthmus sister clades were included, with mean COI K2P distances between 5.3–40.1%. One trans-isthmus sister had the lowest divergence (5.3% K2P), while the remaining 27 sisters had divergences ≥7.4% K2P (Fig. 4.1). All the genetic distances available from trans-isthmus molluscs are from the 5’ end of the COI gene. Mollusc pairs were also phylogenetically diverse, failing within two classes, seven orders, and 12 families. Finally, 27 trans-isthmus sister clades of arthropods were available, with mean COI K2P distances between 4.1–22.6%. Two sisters had the lowest divergences (4.1% and 5.4% K2P), while the remaining 25 sisters had divergences ≥6.2% K2P (Fig. 4.1). For the majority of trans-Bering arthropods, the available genetic distances correspond to the 3’ end of the COI gene. All of the trans-isthmus sisters from arthropods are from the order Decapoda and
fall within five families.

Clock Calibration and Times of Divergence

Using the iterative calibration approach, it was possible to achieve concordance between time estimates based upon the genetic divergences of trans-isthmus sisters and the geological history of the Isthmus of Panama. The reference node for each major group was dated considering 136 calibration dates in sequence and assigning divergence ages for all other pairs relative to this clade according to the iterative calibration approach.

Echinoderms—The reference node used during the calibration process had a divergence of 4.5% K2P. Dating the reference node at 2.1–2.2 Ma provided the most concordant calibration considering the geological history of the Isthmus of Panama formation. Divergence times for all trans-isthmus sisters ranged from 2.1 to 19.4 Ma (Fig. 4.2), while the inferred divergence rate was 2.0–2.1 % pairwise K2P sequence divergence per million years for the COI gene in tropical marine echinoderms (further details in Supplementary Material 4.4).

Molluscs—The reference node had a divergence of 5.3% K2P. Dating this node at 2.1–2.3 Ma provided the most concordant calibration. Divergence times for all trans-isthmus sisters ranged from 2.1 to 17.4 Ma (Fig. 4.2), with an estimated divergence rate of 2.3–2.5% pairwise K2P sequence divergence per million years for the COI gene in tropical marine molluscs (further details in Supplementary Material 4.5).

Arthropods—The reference node exhibited a divergence of 4.1% K2P. The most concordant calibration was obtained by dating this node to 2.2–2.3 Ma. Across trans-isthmus sisters, divergence time estimates were 2.2–12.7 Ma (Fig. 4.2), while the inferred divergence rate was 1.8–1.9% pairwise K2P sequence divergence per million years for the COI gene in tropical marine arthropods of the order Decapoda (Supplementary Material 4.6).

Sensitivity Analysis: Considering Taxon Sampling and Biological Traits

Information about the biological traits of trans-isthmus sister clades was not always accessible (see Supplementary Material 4.2). Information about developmental mode was not available for most molluscs and arthropods, but all trans-isthmus echinoderms have planktonic larval development. In addition, from those trans-isthmus pairs with available information, it was noticed that all pairs of echinoderms, molluscs, and arthropods are distributed in shallow waters (Fig. 4.3). Most trans-isthmus sister pairs within the three phyla have a temperature affinity from tropical waters and are benthic as adults (Supplementary Material 4.2).
Very low sampling (<33%) was evident in 12 of the total 25 families where the trans-isthmus clades are included (Supplementary Material 4.3). The calibration process for the isthmus pairs was repeated after excluding those sisters with low sampling. Within echinoderms, two trans-isthmus sisters were from a family with good (>66%) sampling, and only one pair was from a family with fair sampling (33–66%). The remaining eight sisters were from families with sparse sampling. Pairs from poorly-sampled phylogenies were distributed across a wide range of genetic divergences. If sisters from sparsely-sampled phylogenies had been removed, the calibration process would have only included three pairs, and the results would not be informative; thus, the calibration was not repeated after eliminating sisters from sparsely-sampled phylogenies for the trans-isthmus echinoderms.

Ten trans-isthmus sister pairs of molluscs were from families with good (>66%) sampling, and 11 were from families with fair sampling (33–66%). The remaining seven sisters were from families with poor sampling. Pairs from poorly-sampled and fairly-sampled phylogenies were distributed across the full range of genetic divergences, and results from the calibration process were very similar to those obtained before eliminating sisters from sparsely-sampled phylogenies (Supplementary Material 4.7). Within arthropods, 22 trans-isthmus sisters were from the single family with good (>66%) sampling, one sister was from a family with fair sampling (33–66%), and four sisters were from families with highly incomplete sampling. Pairs from poorly-sampled phylogenies were distributed across the range of genetic divergences, and results from the calibration process were similar to the findings before eliminating sisters from sparsely-sampled phylogenies (Supplementary Material 4.8).

Comparison of Absolute Rates between Tropical and Arctic Lineages

After applying the iterative calibration approach to trans-Isthmus sisters, the resulting absolute rates of molecular evolution for the COI gene in tropical marine invertebrates were compared with prior rates estimated for Arctic marine invertebrates (Chapter II and Chapter III). Higher rates of molecular evolution were found for Arctic taxa in comparison with tropical taxa (Table 4.2). For echinoderms and molluscs, a difference of approximately 40% between tropical vs. Arctic rates was observed, but in arthropods the difference between rates was greater, 1.8–1.9%/MY vs. 5.0–5.2%/MY (Table 4.2).

When comparing the biological traits and taxonomic composition between tropical and Arctic taxa used for calibrations, interesting differences were noticed. Most taxa inhabit shallow waters in both systems, but within Arctic taxa 2 to 5 pairs within each phylum inhabit deeper waters (Fig. 4.3). Furthermore, the taxonomic composition between both systems is similar for
molluscs and arthropods but not for echinoderms. For Arctic echinoderms, the clock calibration is based on sister clades from four of the five classes within the phylum, with Echinoidea not represented; by contrast, results from tropical echinoderms are based mostly on the single class Echinoidea (Fig. 4.4). Lastly, most taxa have benthic adults in both systems, but four and seven pelagic sister pairs are present within Arctic molluscs and Arctic arthropods, respectively (Fig. 4.5).

DISCUSSION

Rate Estimation for Trans-Isthmus Marine Invertebrates

My iterative calibration approach was effective when using the Isthmus of Panama to calibrate the molecular clock in tropical marine invertebrates. After considering the complete geological history of the isthmus formation and incorporating additional sources of evidence (e.g. depth, habitat zone), it was feasible to estimate divergence rates for trans-isthmian sister clades, thus supporting the success of the new method for clock dating when using complex geological events. As suggested in Chapter III, divergence rates estimated using the iterative calibration approach are expected to be more accurate than results from previous literature that used various dates for isthmus closure and, often, simplistic assumptions of simultaneous divergence of sibling species. The methodological approach and results presented here provide an advance in clock dating and will enable future research on tropical marine lineages based upon the presented calibrations.

My results suggest a rate of 2.0–2.1% pairwise K2P sequence divergence per million years for the COI gene in tropical echinoderms from the class Echinoidea. This rate in echinoderms is slightly lower than my resulting divergence rate for tropical molluscs (2.3–2.5% per million years). On the other hand, my results indicate a rate of 1.8–1.9% per million years for tropical arthropods from the order Decapoda, which is the lowest divergence rate among the three phyla in this study (Table 4.2).

The divergence rates reported in this chapter, estimated using the iterative calibration approach are within the range of rates for trans-isthmus sisters in the literature (Table 1.1). For echinoderms, previous rates have been reported in a range of 1.95–6.23% divergence per MY (Hickerson et al. 2003; Coppard et al. 2013). Here, I presented a divergence rate for tropical echinoderms in a range of 2.0–2.1%/MY, which is narrower than the previous range (1.95–6.23%/MY) reported in the literature. Results from previous studies were based on calibrations using a single pair or using the divergence average of a cluster; there was also variability in the
calibration date applied for the isthmus closure (further details in Table 1.1). By contrast, the divergence rate presented here (2.0–2.1%/MY) incorporates all 11 available sister pairs and the complete geological history of the isthmus formation in the calibration process, and thus is also expected to be more accurate than previous rates. On the other hand, divergence rates in the literature for tropical molluscs range widely, from 0.16–5.1%/MY (Marko & Moran 2002; Luttikhuizen et al. 2003). Generally, the lower rates (0.16–2.6%/MY) are based on older fossil calibrations, whereas higher rates (2.1–5.1%/MY) were calibrated using the Isthmus of Panama (Table 1.1). The results from this study using the iterative calibration approach suggest a more precise rate for tropical molluscs (2.3–2.5%/MY) than the range of rates reported in the literature (0.16–5.1%/MY). Applying prior calibrations (e.g. Hallas et al. 2016) to the data analyzed here would yield divergence time estimates incompatible with the geological history. In addition, divergence times based on the iterative calibration were largely concordant with the fossil record from the major trans-Arctic interchange available for molluscs (Vermeij 1991; Reid et al. 1996). Thus, evidence from fossil record supports the accuracy of the results when using the iterative calibration approach and including all 28 trans-isthmus sisters together with the complete geological history of the isthmus formation through the calibration process. Interestingly, the divergence rate reported here for trans-isthmus arthropods (1.8–1.9%/MY) is slightly higher than the highly-cited rate for Alpheus shrimps (1.4%/MY; Knowlton & Weigt 1998). Besides the low divergence rate reported for Alpheus, other rates for trans-isthmus arthropods in the literature range from 1.66–3.88%/MY (Schubart et al. 1998; Wilke et al. 2009). Contrary to most previous studies on trans-isthmus arthropods that use a single pair to calibrate the molecular clock (Table 1.1), the divergence rate presented in this study incorporates a total of 27 sister pairs, thus providing more general rates for decapod arthropods than prior research.

Although molecular clock research using trans-isthmus sisters is vast and tropical rate estimates are common in the literature, the results presented here are proposed to be more accurate. In this study, I was able to approach some problems present in previous works. First, using the novel iterative calibration approach I was able to incorporate the complete geological history of the isthmus formation during the calibration process instead of using a single date. As pointed out before, different dates for the closure of the isthmus have been used for clock dating, from 1.8 (e.g. Lessios et al. 2001) to 3.1 (e.g. Hart et al. 1997) million years. This previous variability in dates makes it difficult to compare studies. Furthermore, instead of assuming simultaneous divergence such as in many previous studies (e.g. Hickerson et al. 2003; Lessios 2008), I used a substantial number of trans-isthmus pairs for calibration because sister species might have diverged at different times through the entire duration of isthmus
formation due to biological and ecological differences among taxa, as well as stochastic effects. Some problems raised in the introduction still remain to be addressed (e.g. testing for rate heterogeneity and saturation effects and including more genes), but the results presented here represent an advance in molecular clock research for tropical marine lineages.

**Rates in Tropical vs. Arctic Marine Invertebrates**

The rate variability among phyla observed in this study agrees with previous evidence that rates vary across lineages (e.g. Schubart *et al.* 1998; Bromham 2002; Bromham & Cardillo 2003; Thomas *et al.* 2006; Mitterboeck & Adamowicz 2013). Likewise, it is not surprising that absolute rates for tropical taxa (this Chapter) differ from the absolute rates estimated for Arctic taxa (Chapters II and III). However, the pattern of divergence rates between Arctic and tropical taxa observed in this study was unexpected.

In general, the COI divergence rates for tropical marine invertebrates reported in this chapter are lower than the rates reported for Arctic marine invertebrates (Chapters II and III). Assuming both timescales are correct, the rate for Arctic echinoderms is approximately 40% higher than in tropical lineages (2.8%/MY vs. 2.0–2.1%/MY). Similarly, the rate for Arctic molluscs is approximately 40% higher than for tropical taxa (3.2%/MY vs. 2.3–2.5%/MY). On the other hand, the rate for Arctic arthropods is almost 2.8 times higher than for tropical arthropods (5.0–5.2%/MY vs. 1.8–1.9%/MY). According to these results, rate variability between tropical and Arctic taxa is general for the taxonomic groups studied here but it is more evident in arthropods than in echinoderms and molluscs (Table 4.2).

Contrary to the results of this study, faster rates of molecular evolution and speciation have been suggested for taxa from warmer climate regions (e.g. Rohde 1992; Bargelloni *et al.* 1994; Bleiweiss 1998b; Gillooly *et al.* 2005; Wright *et al.* 2006) as a response to temperature and higher metabolic rates, which are thought to increase the mutation rate (Martin & Palumbi 1993; Mooers & Harvey 1994; Bleiweiss 1998a; Gillooly *et al.* 2005). In addition, the mutagenic effect of UV exposure is greater towards the equator (Bromham *et al.* 2015), and considering that tropical taxa have also been exposed to higher salinities (hypersaline environments can increase the metabolic rate; Hebert *et al.* 2002), these two factors might also suggest higher mutation rates in tropical taxa. However, the results from this study do not support the initial hypothesis of faster rates of molecular evolution in tropical lineages, given the lineages available for analysis here.

Higher molecular rates observed for Arctic taxa might have several explanations.
Methodological problems (e.g. variability in taxonomic composition and inclusion of undersampled phylogenies) might have had a minor effect on the results (Table 4.3). However, several biological factors—effective population size, extinction rate, ancestral polymorphism, UV, and salinity exposure (Table 4.3)—would likely have had a stronger impact on the estimated rates. After considering various potential explanations and their influence on molecular rates (summarized in Table 4.3), I cannot fully attribute the observed pattern to a single cause. Instead, a few factors might contribute to the finding of higher Arctic rates in this thesis. For example, differences between the taxonomic composition of Arctic and tropical datasets (Fig. 4.4) might explain some of the rate variability observed, at least in echinoderms and arthropods (Table 4.3). The effect of taxonomic composition disparities was also investigated in molluscs after observing that all trans-isthmus sisters from class Bivalvia had high K2P divergences. If the calibration would include only bivalves (seven pairs), the resulting rate would be up to three times faster (7%/MY) than the rate based on the total 28 pairs from tropical molluscs (2.3–2.5%/MY) (Supplementary Material 4.9). Thus, further investigation of rate variability among classes would be necessary. Another explanation for the rate variability observed could be that divergence estimations from Arctic taxa are based only on the 5P region of the COI gene, but in tropical taxa the genetic divergences are based mostly on the 3P region of the COI gene (Supplementary Material 4.2). This could be an explanation if the rate of nucleotide change is not homogeneous across both regions of the COI gene (Roe & Sperling 2007) as initially suggested (Hebert et al. 2003b). Thus, it would be essential to homogenize the divergence estimations and comparison of results to a single region of the COI gene and further evaluate the potential rate heterogeneity across both regions of the COI gene. Also, the mutagenic effect of UV in pelagic taxa could explain the high rate estimated for Arctic arthropods in comparison with tropical arthropods. All sisters from the tropics have benthic adults, but several Arctic sisters of arthropods and molluscs are pelagic during their adult phase (Fig. 4.5). When the calibration was repeated after excluding pelagic taxa, the divergence rate was very similar to the initial results for molluscs (3.3%/MY; Supplementary Material 3.14), but a lower rate was suggested for benthic arthropods (4.3–4.9%/MY; Supplementary Material 3.15). Thus, the effect of UV on pelagic taxa could be another factor that partially explains the higher rates observed for Arctic arthropods. Prior research that compared phylogenetically-related planktonic vs. benthic lineages has also suggested higher rates in planktonic aquatic organisms (DeWaard 2004).

Another possible explanation for the higher rates observed in Arctic lineages could be the disparities between the Panama and the Bering Strait systems, especially in relation to
ancestral polymorphism. Before the formation of the Isthmus of Panama, which separated Pacific and Caribbean populations, ancestral tropical populations might have been continuously distributed populations with large effective population size \((N_e)\). After vicariance, the presence of shared ancestral polymorphisms on each side of the isthmus would increase the time for coalescence and the evolution of reciprocal monophyly, causing an overestimation of divergence times (Edwards & Beerli 2000) and hence underestimation of rates, relative to the Arctic system. Thus, lower divergence rates in tropical sisters might be explained by shared polymorphism. On the other hand, for Arctic groups, a single ancestral population might have existed in the Pacific or Atlantic Oceans, but only a subset of individuals dispersed through the open strait. Founder populations might have had low genetic variability, a condition that was then intensified due to population declines and bottlenecks due to glaciation, resulting in smaller \(N_e\) in ancestral populations. These conditions would be expected to result in the faster evolution of trans-Bering reciprocal monophyly, which may contribute to higher estimated rates of molecular evolution in Arctic taxa.

An additional element that might have a substantial effect on rate estimates explaining the observed rate variability between Arctic and tropical lineages, is effective population size \((N_e)\). Genetic drift has a greater effect in populations with small \(N_e\), where nearly neutral (slightly deleterious) mutations could become fixed rather than eliminated by purifying natural selection, thus increasing the divergence rate (Ohta 1992; Woolfit & Bromham 2005; Lanfear et al. 2014). Smaller \(N_e\) is expected for Arctic populations, which were drastically disturbed by environmental changes during glacial-interglacial periods (Palumbi & Kessing 1991; Maggs et al. 2008). Northern populations were also subject to high rates of extinction (Stanley 1986a; b; Vermeij 1989a; Palumbi & Kessing 1991) in some areas, resulting in geographic structure (Maggs et al. 2008). After several episodes of trans-Bering migrations and colonization of new habitats, Arctic populations had to recover from severe bottlenecks (Palumbi & Kessing 1991). Extinctions, population structure, and bottlenecks together might have contributed to a reduced \(N_e\) in ancestral Arctic populations, leading to higher evolutionary rates. However, this hypothesis should be further assessed and examination of \(N_e\) in Northern populations would be highly valuable.

Although Caribbean populations were also subject to extinction, thus decreasing \(N_e\), the decline in \(N_e\) in Arctic populations due to extinction and demographic events was potentially exacerbated by additional factors. Northern lineages were also affected by significant fluctuations in salinity during glacial-interglacial periods (Maslin et al. 1996). Tolerance to salinity
concentration might have affected the survival of Arctic populations, which could have further reduced population size, reinforcing the explanation for higher rates in Arctic taxa. Furthermore, adaptations to new environments might have influenced the rate of molecular evolution in Arctic lineages. Higher substitution rates could be due to the effect of relaxed or positive selection linked to adaptations for novel environments (Mitterboeck et al. 2016). As mentioned before, Arctic populations were subject to drastic environmental changes and habitat fluctuations. However, this hypothesis would need further examination before conclusions could be made about the underlying mechanisms. For example, future research could estimate ratios of non-synonymous-to-synonymous substitutions, which are informative about the role of selection vs. drift in molecular evolution in different settings (e.g. Witt et al. 2008; Corstorphine 2010; Halanych et al. 2013; Mitterboeck et al. 2016).

**Phylogenetic and Geographic Scope for the Substitution Rates Presented in this Study**

Substitution rates presented in this study calibrated using the isthmus system are most appropriate for research involving tropical marine lineages. They might not represent the appropriate evolutionary rate for marine groups inhabiting higher latitudes due to the pattern of rate variability. It is also important to consider that substitution rates from this study are only representative of the COI gene. It would be inappropriate to assume that they could be applied to other mitochondrial genes as there is evidence that COI evolves more slowly than other mitochondrial genes (e.g. Pons & Vogler 2005; Lavinia et al. 2015). Furthermore, substitution rates presented here might not be appropriate for dating events far outside the calibration time frame (2-7 MY) of this study (i.e. recent demographic events and extremely deep nodes in the tree of life), due to the negative relationship between evolutionary rate estimates and calibration times (Ho et al. 2005, 2011, 2015a) as observed in Figure 1.1.

**Divergence Times for Trans-Isthmus Marine Invertebrates**

A wide range of divergence times was observed for the three phyla, agreeing with previous observations of divergence events prior to the final closure of the Isthmus of Panama (e.g. Knowlton & Weigt 1998; Lessios 2008; O’Dea et al. 2016). In the literature, the split of trans-isthmus sisters has been dated to be as recent as 1.8 Ma (Lessios et al. 2001) and as long ago as 27.5 Ma (O’Dea et al. 2016). Divergence times for trans-isthmus sisters in this study range from 2.1–19.4 Ma in echinoderms, from 2.1–17.4 Ma in molluscs, and from 2.2–12.7 Ma in arthropods. Similar divergence patterns were observed for all three phyla. The divergence of one (in molluscs and arthropods) or two pairs (in echinoderms) is dated after the final closure of the Isthmus of Panama, in agreement with possible temporary gaps in the isthmus between 2–
2.45 Ma. The divergence of the remaining 62 pairs would have occurred between 2.9–19.4 Ma. Of those 62 pairs, 58 pairs (88%) diverged more recently than 12 Ma, after deep-water connections between the Pacific and Caribbean started to become severed (Osborne et al. 2014). The results suggest that trans-isthmus sister taxa of marine invertebrates split at different times during the long-lasting period of the isthmus formation. Thus, assuming simultaneous divergence at the final closure of the isthmus, as in many previous studies (Table 1.1), is inappropriate.

**Future Directions**

An essential improvement of this study would be the further reanalysis of trans-isthmus pairs directly from the DNA sequences, in order to test for saturation effects and rate variability among trans-isthmus sisters. It would be important to explore rate variability not only between members of each sister pair, but also among lineages across entire trees and in relation to biological traits (e.g. using Bayesian analysis). Sequence analysis would be valuable to estimate divergence rates using additional substitution models in order to correct possible underestimations of genetic distances and to allow direct comparisons with previous rates based on substitution models other than K2P (see Table 1.1). Currently, such a reanalysis is only possible for a subset of pairs, because some sequence data associated with published divergence values (e.g. Lessios 2008) is not available in public datasets. Thus, while sequence databases continue to grow, the analysis of new sequence data from these same pairs may be possible in the future. This study would also benefit from including more trans-isthmus pairs, which is feasible in light of growing global sequence datasets. Increasing sequence availability will also enable the examination of phylogenetic hypotheses at broader taxonomic and geographic scales, thus improving inferences of sister relationships. Another important improvement would be to estimate confidence intervals on the calibration itself. Furthermore, it would be important to complement this research with the inclusion of additional genes to assess the consistency of phylogenetic relationships and to estimate rates across genes to facilitate multi-marker dating studies in the future. Also, it would be interesting to investigate the time-dependency hypothesis of molecular rates in more detail by comparing rates from different calibration events; this question could be extended to different genes, which may have different levels of time dependency. Additionally, it would be essential to explore non-synonymous/synonymous ratios across latitude to understand the role of selection vs. drift in relation to the latitudinal gradient. Finally, given the potential contribution of various biological traits and environmental features to rate variability, future work may productively involve
multivariate analysis conducted in a phylogenetic context.

CONCLUSION

The results presented here for tropical taxa suggest a divergence rate for COI of 2.0–2.1%/MY in echinoderms, 2.3–2.5%/MY in molluscs, and 1.8–1.9%/MY in arthropods. These results are in a narrower range compared with previous rates for tropical marine invertebrates in the literature. Using a generous number of sister clades—66 in total—together with the iterative calibration approach, this study suggests more accurate divergence rates for tropical marine invertebrates than in previous studies using simplistic methods. Thus, I anticipate an important advantage over previous calibrations generated using the Isthmus of Panama. Moreover, the results suggest that divergence rates for tropical taxa in this study are lower than for Arctic marine invertebrates. The widespread use of the divergence rate in Alpheid shrimps (1.4%/MY) could have led to a systematic application of an underestimated COI rate for dating evolutionary events in the marine realm, especially for lineages from higher latitudes. Lastly, dating evolutionary events accurately and exploring patterns of rate variability across the tree of life is important for understanding the impacts of prior climatic events upon diversity.
### Table 4.1. Summary of the geological history of the formation of the Isthmus of Panama according with the literature.

<table>
<thead>
<tr>
<th>Date</th>
<th>Dispersal probability between Pacific and Atlantic</th>
<th>Event</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–23 Ma</td>
<td>High dispersal</td>
<td>The geologic collision of Central America with South America took place.</td>
<td>(Coates 2013)</td>
</tr>
<tr>
<td>15 Ma</td>
<td>High dispersal</td>
<td>Transition from generally deep water to sedimentation and increasing of emergence.</td>
<td>(Coates et al. 2004)</td>
</tr>
<tr>
<td>13–12 Ma</td>
<td>Low dispersal</td>
<td>The seaway between the Americas was narrowing and shallowing by the middle Miocene 13 to 12 Ma.</td>
<td>(Coates et al. 2005; Leigh et al. 2014)</td>
</tr>
<tr>
<td>12–9.2 Ma</td>
<td>Low dispersal</td>
<td>Strong evidence that deep-water connections were shut off. The deepest part of the Panama Arc was as shallow as 1200m.</td>
<td>(Osborne et al. 2014)</td>
</tr>
<tr>
<td>9–8 Ma</td>
<td>Low dispersal</td>
<td>Deepening of the Panama Arc possible due to the rise of the sea level.</td>
<td>(Coates et al. 2004)</td>
</tr>
<tr>
<td>7 Ma</td>
<td>Low dispersal</td>
<td>In the southern Caribbean, the temperature and salinity of sea water and carbonate contents of sea-bottom sediments began to increase.</td>
<td>(Leigh et al. 2014)</td>
</tr>
<tr>
<td>7 Ma</td>
<td>Low dispersal</td>
<td>Shoaling began to split deeper-water populations.</td>
<td>(Leigh et al. 2014)</td>
</tr>
<tr>
<td>Time</td>
<td>Dispersal Type</td>
<td>Event Description</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6 Ma</td>
<td>Low dispersal</td>
<td>Panama Arc began rising again and has continued to do until the present day.</td>
<td>(O’Dea et al. 2016)</td>
</tr>
<tr>
<td>5–4 Ma</td>
<td>Low dispersal</td>
<td>Circulation of deep water began to be affected.</td>
<td>(Keigwin 1982; Haug &amp; Tiedemann 1998)</td>
</tr>
<tr>
<td>4.6 Ma</td>
<td>Low dispersal</td>
<td>Surface water salinity began to diverge between Pacific and Atlantic.</td>
<td>(Keigwin 1982; Haug et al. 2001)</td>
</tr>
<tr>
<td>4.2 Ma</td>
<td>Limited dispersal</td>
<td>The Caribbean reached modern salinity values. This is a strong indication that the Panama Arc was mostly emergent at this point. Presence of foraminifers in Western Caribbean are distinct indicators of increased upwelling starting about 4.2 Ma, interpreted as shoaling of the Caribbean-Pacific sill.</td>
<td>(Keigwin 1982; Keller et al. 1989; Haug et al. 2001; Schneier &amp; Schmittner 2006; Kamikuri et al. 2009; Groeneveld et al. 2014)</td>
</tr>
<tr>
<td>4.2–3.8 Ma</td>
<td>Limited dispersal</td>
<td>Increase of the temperature gradient in the eastern Pacific thermocline, consistent with increased coastal upwelling.</td>
<td>(Billups et al. 1999; Chaisson &amp; Ravelo 2000)</td>
</tr>
<tr>
<td>4 Ma</td>
<td>Limited dispersal</td>
<td>Until this time there was little taxonomic or ecological difference in shelf benthic and nektonic communities between Pacific and Caribbean, demonstrating easy movement of water carrying larvae or adults between oceans.</td>
<td>(Jackson et al. 1993; Landau et al. 2009; Aguilera Socorro et al. 2011)</td>
</tr>
<tr>
<td>4 Ma</td>
<td>Limited dispersal</td>
<td>Caribbean’s primary productivity began to decline due to the disappearance of upwelling.</td>
<td>(Leigh et al. 2014)</td>
</tr>
<tr>
<td>4–2 Ma</td>
<td>Limited dispersal</td>
<td>Regional extinction across the Caribbean highly selective against animals suited to high planktonic productivity, implicating declining nutrients due to the restriction of Pacific waters entering the Caribbean.</td>
<td>(Jackson et al. 1993; Allmon 2001; O’Dea et al. 2007; O’Dea &amp; Jackson 2009)</td>
</tr>
<tr>
<td>Time</td>
<td>Dispersal Type</td>
<td>Event Description</td>
<td>Reference</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>3.5–3 Ma</td>
<td>Limited dispersal</td>
<td>Foraminiferan species disappeared from the Caribbean, indicative of shoaling of the Panama Arc.</td>
<td>(Keigwin 1978)</td>
</tr>
<tr>
<td>3 Ma</td>
<td>Possible null dispersal</td>
<td>Arc uplift &amp; sea level falls due to the Pleistocene glaciations resulted in the land bridge connecting North and South America.</td>
<td>(Coates et al. 2004; Bartoli et al. 2005; Coates 2013)</td>
</tr>
<tr>
<td>2.8 Ma</td>
<td>Null dispersal</td>
<td><strong>Closure of the Isthmus of Panama sensu stricto.</strong></td>
<td>(Coates et al. 2005; Lessios 2008; O’Dea et al. 2016)</td>
</tr>
<tr>
<td>2.7 Ma</td>
<td>Null dispersal</td>
<td>The final closure of the isthmus might not have occurred until this point as the consequence of falling sea levels brought about by growing northern glaciers.</td>
<td>(Leigh et al. 2014)</td>
</tr>
<tr>
<td>2.7–2.6 Ma</td>
<td>Null dispersal</td>
<td>The major exchange of mammals between North and South America took place at this time, although a few mammals from South America crossed into North America and vice versa before 3 Ma.</td>
<td>(Molnar 2008)</td>
</tr>
<tr>
<td>2.45 Ma</td>
<td>Possible limited dispersal</td>
<td>Similar salinity and sea surface temperature between the Tropical Eastern Pacific and Caribbean suggest the possible temporary breaching of the isthmus at this time due to rise in the sea level during interglacial periods.</td>
<td>(Cronin &amp; Dowsett 1996; Groeneveld et al. 2014)</td>
</tr>
<tr>
<td>2 Ma</td>
<td>Possible limited dispersal</td>
<td>Salinity decreased in the Caribbean, suggesting a possible subsequent gap in the isthmus.</td>
<td>(Cronin &amp; Dowsett 1996)</td>
</tr>
<tr>
<td>1.8 Ma</td>
<td>Null dispersal</td>
<td>Salinity in both oceans reached modern values.</td>
<td>(Cronin &amp; Dowsett 1996)</td>
</tr>
</tbody>
</table>
Table 4.2 Summary of molecular rates for the mitochondrial cytochrome c oxidase subunit I gene (COI) in three groups of tropical marine invertebrates used in this study (chapter IV) and the four groups of Arctic marine invertebrates from chapters II and III.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Tropical calibration</th>
<th>Arctic calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echinodermata</td>
<td>2.0–2.1%</td>
<td>2.8%</td>
</tr>
<tr>
<td>Mollusca</td>
<td>2.3–2.5%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>1.8–1.9%</td>
<td>5.0–5.2%</td>
</tr>
<tr>
<td>Polychaeta</td>
<td>N/A</td>
<td>3.5–4.7%</td>
</tr>
</tbody>
</table>
Table 4.3 Potential explanations for rate variability between Arctic and tropical taxa.

<table>
<thead>
<tr>
<th>Potential explanation</th>
<th>Consideration of factors that may influence rate estimates and current evidence</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Taxonomic composition</strong></td>
<td>Difference in the taxonomic composition of the pairs used for calibration might be introducing a bias in the results. For example, there is an evident difference in the taxonomic composition between the datasets for Arctic and tropical echinoderms (Figure 4.4 and Supplementary Materials 3.1 and 4.2) and between Arctic and tropical arthropods (Supplementary Materials 3.1 and 4.2). The rate for tropical echinoderms is based mostly on a single class (Echinoidea), which was absent from the Arctic dataset, while the rate for tropical arthropods is based on a single order (Decapoda). If the Arctic calibration is repeated including only Decapoda (10 sister pairs), the resulting rate would be in a range of 4.3–4.9%/MY, which is lower than the rate based on the total 29 pairs from Northern arthropods (5.0–5.2%/MY). The effect of taxonomic composition disparities was also investigated in molluscs. In tropical molluscs, all sisters from class Bivalvia have high K2P divergences; if the calibration is repeated including only bivalves (seven pairs), the resulting rate would be up to three times faster (7%/MY) than the rate based on the total 28 pairs from tropical molluscs (2.3–2.5%/MY) (Supplementary Material 4.9). However, if the Arctic calibration is repeated with only bivalves, the resulting rate is very similar to the full mollusc data set rate. When the calibration was repeated with only gastropods, the results do not change for Northern or tropical taxa. Thus, rate variability between tropical and Arctic lineages could be somewhat explained by taxonomic composition bias.</td>
<td></td>
</tr>
<tr>
<td><strong>Undersampling</strong></td>
<td>After excluding pairs from poorly-sampled phylogenies, the rate estimates were not affected for either the Arctic or the tropical calibration. (Nevertheless, undersampling could influence time estimates for specific nodes.)</td>
<td>Negligible effect</td>
</tr>
<tr>
<td><strong>Alternative dispersal mechanisms</strong></td>
<td>Only two trans-isthmus sister pairs within arthropods (genus Sesarma) are known to be terrestrial. They present two of the three lowest genetic divergences within tropical arthropods. If those two pairs are omitted, the results would not be affected. There is no current evidence of terrestrial/aerial dispersal for the rest of the sister pairs.</td>
<td>Negligible effect</td>
</tr>
</tbody>
</table>
Alternative dispersal routes

An alternative dispersal route for Arctic taxa would be through the Panama seaway before the isthmus formation. On the other hand, an alternative dispersal route for tropical taxa would be through the Bering Strait or around the Cape Horn. However, the thermal tolerance for all the pairs with available information is restricted to either cold or warm waters (Supplementary Materials 3.6 and 4.2).

Species extinction rate

There is evidence of high rates of extinction in the Arctic due to glaciations (Stanley 1986a; b; Vermeij 1989a; Palumbi & Kessing 1991). There is also evidence of extinction in the Caribbean in association with the closure of the isthmus (Jackson et al. 1993). Knowing whether the rate of extinction was higher in the Arctic or in the Caribbean region would be necessary to predict the directionality of this effect upon rates estimates. Nevertheless, it is possible that trans-barrier sisters would have been assigned to the incorrect node if the true sister(s) went extinct, and therefore divergence rates might be overestimated with a high rate of extinction. Given the nature of the iterative calibration approach, this would have to affect all or most pairs for the rates to be biased upwards due to extinction, and thus the effect on overall rate estimates may be minor. However, the rate of extinction would also have a direct effect on the effective population size and consequently on rate estimates (see below).

Effective population size (N_e)

The effect of genetic drift is greater in populations with small N_e; thus, nearly neutral yet slightly deleterious mutations can become fixed more easily, increasing the divergence rate (Ohta 1992; Woolfit & Bromham 2005; Lanfear et al. 2014). Despite that Caribbean populations were subject to extinction, decreasing N_e, the effect of extinction over N_e in Arctic populations was potentially exacerbated by additional demographic factors affecting Artic lineages. Northern populations were drastically affected by environmental changes during glacial periods (Palumbi & Kessing 1991; Maggs et al. 2008) and were subject to high rates of extinction (Stanley 1986a; b; Vermeij 1989a; Palumbi & Kessing 1991) and extirpation in some areas, resulting in geographic population genetic structure (Maggs et al. 2008). Following trans-Bering migrations, colonizing populations in Northern oceans were recovering from severe bottlenecks (Palumbi & Kessing 1991). Extinctions, population structure, and bottlenecks might have contributed to reduced N_e in ancestral populations in the Arctic. Accordingly, evolutionary rates are expected to be higher in Arctic taxa.
Difference between systems: vicariance vs. dispersal

Panama system.— The ancestral population might have been a continuously-distributed population with larger \( N_e \) that was then divided by a vicariance event. It would have taken a longer time to observe reciprocal monophyly than for the trans-Bering pairs. This factor, in combination with ancestral polymorphism (see below), could contribute to explaining the lower divergence rates observed for tropical lineages.

Bering Strait system.— For many groups, a single ancestral population might have existed in the Pacific or Atlantic Oceans, but only a subset of individuals dispersed through the newly open strait. The founding trans-Bering population is expected generally to have had low genetic diversity, a condition that was intensified during population bottlenecks and population declines associated with glacial episodes. Thus, reciprocal monophyly may have occurred faster in Arctic populations, and divergence rates then might be overestimated in comparison with the tropics.

Ancestral polymorphism

Polymorphism of the ancestral population prior to the divergence of sister pairs might be introducing bias to the rate estimation (Wilke et al. 2009). If the size of the ancestral population was large, speciation dates can be greatly overestimated as a result of the potentially deep coalescence of ancestral polymorphisms (Edwards & Beerli 2000). Variance in coalescence time due to polymorphism in ancestral populations could explain the variation in genetic distance among sister species. Coalescence time for two random alleles from a structured species will increase as the number of populations increases and the migration rate decreases. Thus, if the alleles that gave rise to those in descendant species derive from different subpopulations, it could drive an overestimation of the divergence times (Edwards & Beerli 2000), which would be associated with a decrease in rate estimates in tropical lineages.

On the other hand, reciprocal monophyly and coalescence times might have occurred faster in Arctic lineages, which may contribute to higher evolutionary rates in Arctic taxa. This could be explained by the effect of severe population declines and bottlenecks, presumably triggering low genetic variability and smaller \( N_e \) in ancestral populations. Hence, Arctic divergence rates might be overestimated relative to tropical rates due to lower shared ancestral polymorphism. However, the effect of ancestral polymorphism in rate estimates is less problematic for deeper nodes than for population-level estimations (Ho et al. 2005, 2008). Therefore, it is difficult to know the precise effect of ancestral polymorphism over rate estimation in this study. While ancestral polymorphism could explain some of the variation in evolutionary rates between Arctic and tropical lineages, further investigation is needed to evaluate this hypothesis fully.

Minor effect

Moderate effect
Developmental mode

Direct and non-planktonic development is common in many cold-water echinoderms (Fetzer & Arntz 2008; Pearse et al. 2009). Non-planktonic larvae in echinoderms have been associated with higher rates of evolution due to smaller $N_e$ (Foltz 2003; Foltz & Mah 2009, 2010; Corstorphine 2010). For echinoderms, non-planktonic developmental mode is present in two trans-Bering pairs from this study, while all sister pairs from the tropics have a planktonic development. However, there is no evidence of difference in the genetic divergences related to developmental mode in echinoderms, given the limited trait variability among the pairs available here. Developmental mode was mostly unavailable for the other groups; therefore, it was not possible to conclude if this factor might contribute to explaining the results.

Unknown

UV damage

UV radiation has a direct mutagenic effect on DNA; thus, exposure may lead to an increase in rate of molecular evolution (Smith et al. 1992; Pawlowski et al. 1997; Hebert et al. 2002; DeWaard 2004). UV exposure increases towards the equator (Bromham et al. 2015); thus, UV damage would be expected to be higher in tropical taxa. However, the results suggest higher rates for Arctic taxa. In addition, UV damage is expected to be higher for pelagic taxa (Smith et al. 1992; Pawlowski et al. 1997; DeWaard 2004), which are often transparent or translucent and inhabit shallower waters, and this could be an explanation for the higher rates observed in Arctic molluscs and arthropods. All sisters from the tropics have benthic adults, but several sister pairs of Arctic arthropods and molluscs are pelagic during their adult phase. When the calibration was repeated after excluding pelagic taxa, the divergence rate was very similar for molluscs as when calibrated using the full dataset, but a lower rate was suggested for benthic arthropods (4.3–4.9%/MY). In conclusion, the effect of UV could partially explain the higher rates observed for Arctic arthropods.

Moderate effect (in arthropods)

Salinity

It has been suggested that higher intracellular salinity reduces the fidelity of DNA replication, increasing mutagenic effects (Favre & Rudin 1996). Therefore, it is expected that organisms inhabiting hypersaline waters would present higher rates of molecular evolution (Hebert et al. 2002). Accordingly, increased salinity in the Caribbean in association with the formation of the Isthmus of Panama (Molnar 2008) would suggest higher mutation rates in tropical taxa (i.e. Caribbean lineages), but this is contrary to the observed results. On the other hand, with the first opening of the Bering Strait, low-salinity waters were transported from the Pacific into the Arctic and North Atlantic

Minor effect
Oceans (Matthiessen et al. 2009). The subsequent closure of the Bering Strait during glacial periods led to an increase in surface salinities in the Arctic and North Atlantic Oceans (Matthiessen et al. 2009). Salinity fluctuations during glacial-interglacial periods (Maslin et al. 1996) might have subjected individuals to sub-optimal ionic conditions, potentially increasing the mutation rate, and also affected the survival of Arctic populations, intensifying fluctuations in population size. If population size were reduced due to salinity fluctuations, subsequent increased molecular rates would then be expected for Arctic populations. A broad-scale study of relative rates between freshwater and marine taxa revealed only a weak and taxon-specific relationship (Mitterboeck et al. 2016).

Relaxed or positive selection on genes associated with adaptations to new environments might influence patterns of molecular evolution by increasing the substitution rate (Mitterboeck et al. 2016). Arctic populations were subject to drastic environmental changes and habitat fluctuations, so under new ecological settings molecular rates might have been affected. Therefore, higher rates would be expected for Arctic taxa. In addition, low \( N_e \) during bottlenecks and colonization events would strength the expectation of higher rates in Arctic taxa. This possibility may be of minor effect in a conserved gene such as COI.

<table>
<thead>
<tr>
<th>Early phases of divergence</th>
<th>Minor effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1 Kimura two-parameter (K2P) divergences retrieved from the literature for trans-isthmus sister clades of echinoderms, molluscs, and arthropods. Genetic divergences are colour-coded by class in each phylum.
Figure 4.2 Divergence time estimates for tropical marine echinoderms, molluscs, and arthropods using K2P distances and the best calibration according to the iterative calibration approach. Using the geological history of the Isthmus of Panama formation (Table 4.1), the time scale was divided into episodes of high, low, limited, and null dispersal in accordance with the probability of marine dispersal between the Pacific and Caribbean. The final closure of the Isthmus of Panama has been dated *sensu stricto* at 2.8 Ma (Coates *et al.* 2005; Lessios 2008; O’Dea *et al.* 2016); however, possible shallow and narrow marine connections might have occurred around 2.0–2.45 Ma (light grey), allowing limited dispersal between oceans (Cronin & Dowsett 1996; Groeneveld *et al.* 2014). Dark shaded boxes show the periods of time when the Isthmus of Panama is considered to be entirely formed; therefore, migrations of marine organisms between the oceans were extremely unlikely.
**Figure 4.3** Depth habitat distribution of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III). Sister clades were classified in two categories according to their habitat distribution: deep-waters (>50 m) and shallow-waters (≤50 m).
### Figure 4.4

Comparison of the taxonomic composition at the class level of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III).

<table>
<thead>
<tr>
<th>Class</th>
<th>Trans-isthmus sisters per class</th>
<th>Trans-Bering sisters per class</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A) Echinodermata</strong></td>
<td><img src="chart1.png" alt="Pie Chart" /></td>
<td><img src="chart2.png" alt="Pie Chart" /></td>
</tr>
<tr>
<td>Asteroidea</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Crinoidea</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Echinoidea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holothuroidea</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Ophiuroidea</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B) Mollusca</strong></td>
<td><img src="chart3.png" alt="Pie Chart" /></td>
<td><img src="chart4.png" alt="Pie Chart" /></td>
</tr>
<tr>
<td>Bivalvia</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>Cephalopoda</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Gastropoda</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyplacophora</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C) Arthropoda</strong></td>
<td><img src="chart5.png" alt="Pie Chart" /></td>
<td><img src="chart6.png" alt="Pie Chart" /></td>
</tr>
<tr>
<td>Hexanauplia</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>Malacostraca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pycnogonida</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.5 Habitat zone of the trans-barrier sister clades used for clock calibration using the Isthmus of Panama (Chapter IV) and using the Bering Strait (Chapters II and III).
Chapter V

General Conclusions

Summary of Key Findings

This thesis has presented three main advances in molecular clock research: first, the development of the iterative calibration approach; second, the new rates of molecular evolution for Arctic and tropical marine lineages; lastly, the suggestion that rates of molecular evolution tend to be faster in Arctic marine invertebrates than in their tropical counterparts.

Chapter II presented a novel and successful method—the iterative calibration approach—for calibrating the molecular clock using complex biogeographic events. By integrating several sources of information (e.g. molecular and geological information, fossil evidence, and a large number of sister pairs), the iterative calibration seeks congruence between the dates assigned to multiple divergence events and geological history, thus anticipating more accurate calibrations than previous studies using simplistic methods. In this chapter, the opening and re-closure of the Bering Strait was successfully used to calibrate the molecular clock in echinoderms using the iterative calibration approach. The results suggested multiple pulses of trans-Bering migration and a divergence rate of 2.8%/MY in Northern echinoderms.

Chapter III extended the use of my iterative calibration approach to calibrate molecular clocks in additional groups of marine invertebrates. This chapter presented novel clock calibrations for Northern molluscs and arthropods, as well as a re-analysis of previously-generated pairs of Northern polychaetes (from Carr 2010). The results support multiple pulses of trans-Bering migrations in all three taxa and suggested a divergence rate of 3.2%/MY in molluscs, 5.0–5.2%/MY in arthropods, and 3.5–4.7%/MY in polychaetes. Divergence times based on the iterative calibration were largely concordant with available fossil record from the major trans-Arctic marine interchange (Vermeij 1991; Reid et al. 1996), thus supporting the accuracy of the results when using the iterative calibration approach. The divergence rates from Northern marine invertebrates presented in chapters II and III contrast with previously highly-cited low divergence rates for tropical marine lineages (e.g. Knowlton & Weigt 1998; Schubart et al. 1998; Marko 2002).

Chapter IV expanded the use of the iterative calibration approach to calibrate molecular clocks for tropical marine invertebrates. This chapter revealed that my iterative calibration approach is successful when using the formation of the Isthmus of Panama for clock dating.
Also, in this chapter I presented new clock calibrations and rates for tropical marine invertebrates, suggesting a divergence rate of 2.0–2.1%/MY in tropical echinoderms, 2.3–2.5%/MY in tropical molluscs, and 1.8–1.9%/MY in tropical arthropods. The results presented here are more precise, falling within the range from previous rates in the literature (Table 1.1 and Fig. 1.2), but are in a narrower range (Fig. 5.1). More accurate results are expected by incorporating more data and additional information (e.g. the complete geological history and multiple time boundaries) than prior studies during the calibration. The results presented here agree with prior evidence of trans-isthmus divergences previous to the final closure of the Isthmus of Panama (e.g. Knowlton & Weigt 1998; Lessios 2008; Selin et al. 2016). Furthermore, the results from tropical marine invertebrates contrast with the rates reported in chapters II and III for Arctic lineages, suggesting that molecular rates from tropical taxa are generally slower than rates from Arctic taxa over the time range investigated here. Higher divergence rates observed in Arctic lineages may be attributed, for example, to smaller effective population size and lower shared ancestral polymorphism in Arctic populations.

**Novel Contributions to the Field of Molecular Clock Research**

This thesis represents an original contribution to the field of molecular clock research by introducing an innovative methodological approach for calibrating molecular clocks using biogeographic events—the iterative calibration approach. As discussed in this study, more accurate calibrations are expected by incorporating several sources of information through the use of the novel calibration approach. Assuming simultaneous divergence or using a single sister pair for clock calibration would have resulted in different divergence rates and would have yielded divergence date estimates for other nodes that lacked congruence with the geological history. In addition, problems arising from an incomplete fossil record could be overcome by including several sister pairs during the calibration process, thus improving our ability to date the tree of life. This research also fills gaps in knowledge by presenting new rates of molecular evolution for Northern marine invertebrates and proposes more accurate rates for tropical marine invertebrates. Furthermore, this thesis suggests a pattern of higher evolutionary rates in Arctic marine invertebrates compared with tropical lineages. Although the mutation rate is expected to be higher in the tropics due to several factors (i.e. UV exposure, metabolic rate, salinity, shorter generation times) (Jukes & Holmquist 1972; Rohde 1992; Martin & Palumbi 1993; Thomas et al. 2010; Bromham et al. 2015), this research showed the importance of the fixation rate in governing observed molecular divergence rates. Finally, the observed pattern of higher rates in the Arctic cautions researchers to avoid the use of tropical rates to date
evolutionary events in lineages from high latitudes, and *vice versa*. Prior to applying rates across major latitudinal zones, it would be necessary to investigate further the underlying causes of the difference and to assess whether a correction, such as for effective population size, could help to mitigate the difference and enable rates to be applied across broader latitudinal regions. The results from this study also provide a new context for dating and understanding the evolutionary history in the marine realm.

**Implications**

Findings from this study are broadly relevant in the field of molecular evolution. First, the iterative calibration can successfully be used to calibrate molecular clocks in additional Arctic marine lineages using the Bering Strait. Second, it can effectively be applied to the Isthmus of Panama system to calibrate molecular clocks in other tropical marine groups. Third, the new method can be adjusted to calibrate molecular clocks in terrestrial lineages using both events, the glacial history of Bering Strait and the formation of the Isthmus of Panama. Fourth, the iterative calibration approach could also be adjusted and applied to calibrate molecular clocks using other biogeographic events.

The divergence rates presented in this study can be cautiously used to date evolutionary events in other related taxa and be used in comparison with future research. Overall, this work represents an advance for dating evolutionary events, which is important for elucidating patterns of diversification and for revealing factors that might have influenced historical speciation and extinction events. Dating evolutionary events is an important piece of evidence towards our understanding of how and why these events occurred.

The evidence of rate variability between Arctic and tropical lineages observed in this study reveals that it can be inappropriate to use calibrations derived from taxonomic groups from different latitudinal regions. This subject should be considered in future work and when comparing or using molecular rates from previous literature.

**Limitations**

One limitation of this work is the use of a single genetic marker. Substitution rates presented in this study only represent the evolutionary history of the COI gene and should not be generalized to other genes; it would be also inappropriate to assume that they apply to the entire mitochondrial genome (Lavinia *et al.* 2015). Another limitation is that the rates presented in this study would not be appropriate for dating events far outside the calibration time frames (2-7 MY) used here (i.e. recent demographic events and extremely deep nodes in the tree of
life), due to the evidence of rate variability over time (time-dependency hypothesis) (Ho et al. 2007, 2011, 2015a). Lastly, it is important to consider that the substitution rates presented here might not represent the appropriate evolutionary rate for specific marine groups with extreme rate variability.

**Future Work and Recommendations**

An important future direction to this research would be to include additional markers. Complementing COI with other genes is essential to assess the consistency of rates across genes and to facilitate multi-marker dating studies in the future. Also, a future improvement involves the inclusion of both regions (5P and 3P) of the COI gene. First, it would be important to evaluate rate variability between the two different regions of the COI gene. If rate heterogeneity is present between the different regions of the gene, it would be critical to homogenize the information available for rate estimations using COI sequences, particularly because many available sequences from tropical taxa correspond to the 3P end of the COI gene.

Previously, I pointed out that the rates presented here would be more appropriate for dating events within the time frames from the calibrations used. In that regard, future work could further explore the time-dependency hypothesis of molecular rates in COI and in marine invertebrates. This question could be then extended to different genes, which may have different levels of time dependency (Lavinia et al. 2015). This research would also benefit from estimating confidence intervals on the calibration itself. In addition, it would be important to investigate rate variability among and within each major group included in this study. Here, I estimated general rates for large taxonomic groups and examined relative rates in sister pairs, but I only quantified rate variability across a phylum tree for a subset of echinoderms. Thus, rate variability among lineages across entire trees and in relation to biological traits should be further investigated in order to better understand evolutionary patterns in marine invertebrates. Furthermore, it would be important to explore non-synonymous/synonymous substitution ratios across latitude in order to understand the role of selection vs. drift in relation to the latitudinal pattern in molecular rates observed in this study.

The opportunity to include additional sister pairs and/or pairs from more complete phylogenies would be valuable to strengthen the results for the current taxonomic groups. Additionally, a promising future direction would be to calibrate the molecular clock for other marine lineages, including those without fossil records and understudied groups that would benefit from molecular clock research using biogeographic calibrations. This advance will
become increasingly feasible in light of growing global sequence datasets.

Lastly, I warn about the possible systematic underestimation of molecular rates in the marine realm due to the widespread use of the divergence rate in *Alpheus* shrimps (1.4%/MY from Knowlton & Weigt 1998) in many previous studies (e.g. Blanco-Bercial *et al.* 2011; Milligan *et al.* 2011; Jung *et al.* 2013; Alam *et al.* 2016; Anker *et al.* 2017; Roussel & Van Wormhoudt 2017). Hence, it is worth reconsidering published estimates of divergence time and conclusions that were based upon that rate.

**FINAL REMARKS**

This thesis represents a novel and important contribution to the field of molecular evolution in the following aspects:

i) It presents a new methodological approach for clock calibration, which suggests more accurate results than when using few data points and simplistic assumptions.

ii) It focuses on marine invertebrates, which are usually understudied but represent an important amount of diversity.

iii) It is characterized by the use of a large geographic and taxonomic scale to explore molecular divergence rates.

iv) It is characterized by the largest number of sister pairs (157 sister pairs) ever used to calibrate the molecular clock in marine invertebrates.

v) It supports previous findings of rate variability across the tree of life and between lineages with different biological and ecological traits.

vi) It includes work on a poorly-explored subject, rate variability between tropical vs. arctic marine lineages.

vii) It discovered a new trend, higher rates in Arctic lineages compared with tropical taxa, which is contrary to previous expectations. It points to the importance of fixation rate differences—as contrasted with mutation rate differences as a primary driver of observed rate differences—and sets up future research to investigate the mechanisms underlying this pattern in more detail.

viii) Lastly, this thesis represents an important contribution to the study of the tree of life and to our understanding of the impact of prior climate change events upon the diversification of life.
Figure 5.1 Divergence rates estimated in this study compared with the range of rates reported in the literature for marine invertebrates. Stars represent the divergence rates estimated in this study for Arctic (white) and tropical (grey) sister pairs using the iterative calibration approach. Divergence rates from the literature (boxes) were estimated using a calibration date between 1–20 Ma (see Table 1.1).
REFERENCES


Durham J, MacNeil F (1967) Cenozoic migrations of marine invertebrates through the Bering


Foltz DW (2003) Invertebrate species with nonpelagic larvae have elevated levels of nonsynonymous substitutions and reduced nucleotide diversities. *Journal of Molecular Evolution*, 57, 607–612.


Harris SA (2005) Thermal history of the Arctic Ocean environs adjacent to North America during the last 3.5 Ma and a possible mechanism for the cause of the cold events (major glaciations and permafrost events). *Progress in Physical Geography*, **29**, 218–238.


Harvey HR, Pleuthner RL, Lessard EJ, Bernhardt MJ, Tracy Shaw C (2012) Physical and biochemical properties of the euphausiids *Thysanoessa inermis*, *Thysanoessa raschii*, and


Marko PB, Moran AL, Kolotuchina NK, Zaslavskaya NI (2014) Phylogenetics of the gastropod genus *Nucella* (Neogastropoda: Muricidae): species identities, timing of diversification and


R Core Team (2017) R: A language and environment for statistical computing.


Witt JDS, Threlfall DL, Hebert PDN (2008) Genetic zoogeography of the


Supplementary Material 2.7. Bayesian analysis settings

Relative node ages were estimated using an ultrametricized Bayesian inference tree generated in BEAST v1.8.2 (Drummond et al. 2012) using the following analysis settings. The best-fit substitution model for the dataset was used, the GTR+G+I model with empirical base frequencies, four gamma categories, all codon positions partitioned with unlinked base frequencies and substitution rates. The Birth-Death speciation model (Yang & Rannala 1997) tree prior was selected, and an uncorrelated relaxed lognormal clock model was used with the rate estimated for the data. The mean (ucld.mean) parameter prior distribution was set to a uniform distribution with an initial value of 1, a lower boundary of 0, and 10 as an upper boundary. The relative rate parameters for codon positions 1, 2, and 3 were set to an exponential distribution. For positions 1 and 2 (CP1mu and CP2mu), the mean was set to 0.05, while the mean for position 3 (CP3mu) was set to 5 (Oaks & Callahan 2014). All other settings were left as default according to the BEAST v1.8.2 manual (Drummond et al. 2012). Six independent Markov chain Monte Carlo (MCMC) analyses were each run with a MCMC length of 100,000,000 generations, sampling every 10,000 with a final sample size of 10,000 trees. Effective Sample Size (ESS) values and trace files of runs were evaluated in Tracer v1.6 (Rambaut et al. 2014) resulting in adequate sampling (ESS>100) for all parameters and convergence in the stationary distribution (Drummond & Bouckaert 2015). In a first attempt, the best two runs (with higher ESS values) were combined using the BEAST utility program LogCombiner v1.8.2 with 10% burn-in to obtain an estimate of the posterior distribution. Maximum clade credibility trees with mean node heights of the target tree were constructed with 10% burn-in in the BEAST utility program TreeAnnotator v1.8.2. Due to low support for deeper nodes and independent MCMC analysis showing different tree topologies, it was not possible to combine two runs and their trees as suggested by the software authors (Drummond et al. 2012). Within the phylum Echinodermata, relationships among classes (Littlewood et al. 1997; Perseke et al. 2010; Telford et al. 2014; Feuda & Smith 2015), orders, and families remain controversial (e.g. Kamarudin et al. 2010; Mah & Foltz 2011; Feuda & Smith 2015), and so imposing constraints on the deep topology was not feasible. In order to overcome this difficulty in combining runs, I analysed six independent runs and trees as a way to test for replicability of the calibration; then, I selected the run with the highest ESS values for detailed reporting of results and to construct the final phylogenetic tree. For the selected MCMC run, a total of 9001 phylogenetic trees were recovered after discarding 10% as burn-in using TreeAnnotator v1.8.2. The Maximum clade credibility tree with mean node heights was then visualized in FigTree v1.4.2 (Rambaut 2014).
## Supplementary Material 2.8. List of Trans-Bering sister clades of echinoderms

<table>
<thead>
<tr>
<th>Class</th>
<th>Genus</th>
<th>North Pacific</th>
<th>Arctic/North Atlantic Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Specific epithet</td>
<td>BIN</td>
</tr>
<tr>
<td>Asteroidea</td>
<td>Pteraster</td>
<td>militaris</td>
<td>BOLD:ABZ2331</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteraster</td>
<td></td>
<td>BOLD:AAH7926</td>
</tr>
<tr>
<td></td>
<td>Henricia</td>
<td>leviuscula spiculifera</td>
<td>BOLD:AEE3892</td>
</tr>
<tr>
<td></td>
<td>Leptasterias</td>
<td>stolocanta/fisheri</td>
<td>BOLD:AAE4892</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Solaster</td>
<td></td>
<td>BOLD:AAB5053</td>
</tr>
<tr>
<td></td>
<td>Crossaster</td>
<td>papposus</td>
<td>BOLD:AAB5053</td>
</tr>
<tr>
<td></td>
<td>Asterias</td>
<td>amurensis</td>
<td>BOLD:AAD2155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holothuroidea</td>
<td>Cucumaria</td>
<td>pallida/miniata</td>
<td>BOLD:AEE3243</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BOLD:AAC2853</td>
</tr>
<tr>
<td></td>
<td>Pentamera</td>
<td>lissoplaca</td>
<td>BOLD:AAX1529</td>
</tr>
<tr>
<td></td>
<td>Psolus</td>
<td>chitonoides</td>
<td>BOLD:AEE0864</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudostichopus</td>
<td>tuberosus</td>
<td>BOLD:AAC8222</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ophiuroidea</td>
<td>Asteroschema</td>
<td>sublaveve</td>
<td>BOLD:AAJ9638</td>
</tr>
<tr>
<td></td>
<td>Ophiopholis</td>
<td>japonica</td>
<td>BOLD:AAI1097</td>
</tr>
<tr>
<td></td>
<td>Amphiophiura</td>
<td>superba</td>
<td>BOLD:AAE6089</td>
</tr>
<tr>
<td></td>
<td>Ophiura</td>
<td>sarsii</td>
<td>BOLD:AAD3481</td>
</tr>
<tr>
<td>Crinoidea</td>
<td>Florometra</td>
<td>serratissima</td>
<td>BOLD:AAC3353</td>
</tr>
</tbody>
</table>
### Supplementary Material 3.2 List of trans-Bering sister clades of molluscs and arthropods

<table>
<thead>
<tr>
<th>Class</th>
<th>Genus</th>
<th>Specific epithet</th>
<th>North Pacific BIN</th>
<th>Arctic/North Atlantic Ocean Specific epithet</th>
<th>Arctic/North Atlantic Ocean BIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOLLUSCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivalvia</td>
<td><em>Thyasira</em></td>
<td><em>flexuosa</em></td>
<td>BOLD:AAV6708</td>
<td><em>sarsi</em></td>
<td>BOLD:AAI0391</td>
</tr>
<tr>
<td></td>
<td><em>Axinopsida</em></td>
<td><em>serricata</em></td>
<td>BOLD:ABA6564</td>
<td><em>orbiculata</em></td>
<td>BOLD:AAW3037</td>
</tr>
<tr>
<td></td>
<td><em>Ennucula</em></td>
<td><em>cardara</em></td>
<td>BOLD:ACH6511</td>
<td><em>teanis</em></td>
<td>BOLD:AAE3708</td>
</tr>
<tr>
<td></td>
<td><em>Yoldiella</em></td>
<td><em>orcia</em></td>
<td>BOLD:ACI2405</td>
<td><em>frigida</em> / <em>nana</em> / <em>inconspicua_inconspicua</em></td>
<td>BOLD:AOA3808 / BOLD:AOA3807 / BOLD:ACI2404</td>
</tr>
<tr>
<td></td>
<td><em>Yoldia</em></td>
<td><em>scissurata</em></td>
<td>BOLD:ACH5248</td>
<td>sp.</td>
<td>BOLD:ACR1874</td>
</tr>
<tr>
<td></td>
<td><em>Pecten</em></td>
<td><em>albicans</em></td>
<td>BOLD:AAX1745</td>
<td><em>jacobaeus</em> / <em>maximus</em> / <em>maximus</em></td>
<td>BOLD:ABZ0150 / BOLD:AAE7060 / BOLD:ACV7290</td>
</tr>
<tr>
<td></td>
<td><em>Modiolus</em></td>
<td><em>modiolus</em></td>
<td>BOLD:ACM2550</td>
<td><em>modiolus</em></td>
<td>BOLD:AAG5349</td>
</tr>
<tr>
<td>Cephalopoda</td>
<td><em>Onychoteuthis</em></td>
<td><em>compacta</em></td>
<td>BOLD:AAX0494</td>
<td><em>banksii</em></td>
<td>BOLD:AAX0487</td>
</tr>
<tr>
<td></td>
<td><em>Helicocranchia</em></td>
<td><em>pfefferi</em></td>
<td>BOLD:AAY2019</td>
<td>sp.</td>
<td>BOLD:ACQ6631</td>
</tr>
<tr>
<td></td>
<td><em>Taonius</em></td>
<td><em>borealis</em></td>
<td>BOLD:AAK0251</td>
<td><em>pavo</em></td>
<td>BOLD:AAM9951</td>
</tr>
<tr>
<td></td>
<td><em>Rossia</em></td>
<td>sp.</td>
<td>[GBCPH0134-06]</td>
<td><em>palpebrosa</em></td>
<td>BOLD:AA6541</td>
</tr>
<tr>
<td>Polyplacophora</td>
<td><em>Boreochiton</em></td>
<td><em>beringensis</em></td>
<td>BOLD:ABV1920 / BOLD:ACQ6736</td>
<td><em>Tonicella rubra</em></td>
<td>BOLD:ABW0554</td>
</tr>
<tr>
<td></td>
<td><em>Polyschides</em></td>
<td><em>quadrifissatus</em></td>
<td>BOLD:AAU2327</td>
<td><em>carolinensis</em></td>
<td>BOLD:AAY3381</td>
</tr>
<tr>
<td>Gastropoda</td>
<td><em>Velutina</em></td>
<td>sp.</td>
<td>BOLD:ABV1790</td>
<td><em>velutina</em></td>
<td>BOLD:ABV1790</td>
</tr>
<tr>
<td></td>
<td><em>Littorina</em></td>
<td><em>aleutica / natica</em></td>
<td>BOLD:ACB8366</td>
<td>*obtusata / fabalis / saxatilis / compresa / arcana</td>
<td>BOLD:AAA7469 / BOLD:AAG1552</td>
</tr>
<tr>
<td>Gastropoda</td>
<td>Littorina</td>
<td>squalida</td>
<td>BOLD:ACB8238</td>
<td>littorea</td>
<td>BOLD:AAA1269</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ariadnaria</td>
<td>insignis</td>
<td>BOLD:ABV3540</td>
<td></td>
<td></td>
<td>BOLD:ACM0991</td>
</tr>
<tr>
<td>Trichotropis</td>
<td>cancellata</td>
<td>BOLD:ABA6567</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neptune</td>
<td>phoenicea</td>
<td>BOLD:AAW7290</td>
<td></td>
<td></td>
<td>BOLD:ABV9665</td>
</tr>
<tr>
<td>Neptune</td>
<td>amianta</td>
<td>BOLD:AAW7290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neptune</td>
<td>lyrata</td>
<td>BOLD:AAW7290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dendronotus</td>
<td>venustus</td>
<td>BOLD:AAO4633</td>
<td>frondosus</td>
<td>BOLD:AAC7990</td>
<td></td>
</tr>
<tr>
<td>Dendronotus</td>
<td>patricki</td>
<td>BOLD:ACB7882</td>
<td>robustus</td>
<td>BOLD:AAL1485</td>
<td></td>
</tr>
<tr>
<td>Dendronotus</td>
<td>kalikal</td>
<td>BOLD:ACP9856</td>
<td>frondosus</td>
<td>BOLD:ACR7570</td>
<td></td>
</tr>
<tr>
<td>Flabellina</td>
<td>trophina / fusca</td>
<td>BOLD:ACB8064</td>
<td>gracilis</td>
<td>BOLD:ACR6252</td>
<td></td>
</tr>
<tr>
<td>Acanthodoris</td>
<td>atrogriseata / pilosa</td>
<td>BOLD:AAX0711</td>
<td>pilosa</td>
<td>BOLD:AAE9648</td>
<td></td>
</tr>
<tr>
<td>Doris</td>
<td>montereyensis</td>
<td>BOLD:ABU5771</td>
<td>pseudoargus</td>
<td>BOLD:AAE0792</td>
<td></td>
</tr>
<tr>
<td>Onchidoris</td>
<td>bilamellata</td>
<td>BOLD:ABU5771</td>
<td>pseudoargus</td>
<td>BOLD:AAE0792</td>
<td></td>
</tr>
<tr>
<td>Alderia</td>
<td>willowi</td>
<td>BOLD:AAW1549</td>
<td>modesta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margarites</td>
<td>helicinus</td>
<td>BOLD:AAK7703</td>
<td>helicinus</td>
<td>BOLD:AAG1563</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ARTHROPODA</th>
<th>Calocaris</th>
<th>investigatoris</th>
<th>BOLD:AAI6879</th>
<th>macandreae</th>
<th>BOLD:ACF6499</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eusergestes</td>
<td>similis</td>
<td>BOLD:AAD3509</td>
<td>articus</td>
<td>BOLD:AAE1889</td>
<td></td>
</tr>
<tr>
<td>Pandalus</td>
<td>goniurus</td>
<td>BOLD:AAX1996</td>
<td>montagui</td>
<td>BOLD:AAB2199</td>
<td></td>
</tr>
<tr>
<td>Pandalus</td>
<td>borealis</td>
<td>BOLD:AAA9256</td>
<td>borealis</td>
<td>BOLD:AAA9256</td>
<td></td>
</tr>
<tr>
<td>Eualus</td>
<td>suckleyi</td>
<td>BOLD:AAC6386</td>
<td>gaimardii</td>
<td>BOLD:AAJ0881</td>
<td></td>
</tr>
<tr>
<td>Spirontocaris</td>
<td>spinus</td>
<td>BOLD:AAE9107</td>
<td>spinus sp.</td>
<td>BOLD:AAB6145</td>
<td>BOLD:ACR1068</td>
</tr>
<tr>
<td>Eualus</td>
<td>sp.</td>
<td>BOLD:ACL8098</td>
<td>pusiolus</td>
<td>BOLD:AAC7990</td>
<td></td>
</tr>
</tbody>
</table>

136
<table>
<thead>
<tr>
<th>Malacostraca</th>
<th>BOLD: AAC3832</th>
<th>avinus</th>
<th>BOLD: AAF3776</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eualus macilentus</td>
<td>sp.</td>
<td>pubescens</td>
<td>sp.</td>
</tr>
<tr>
<td>Pagurus sp.</td>
<td>BOLD: AAF1438</td>
<td>avinus</td>
<td>BOLD: AAD6867</td>
</tr>
<tr>
<td>Pagurus pectinatus</td>
<td>BOLD: ACH8698</td>
<td>arcuratus</td>
<td>BOLD: AAE6975</td>
</tr>
<tr>
<td>Thysanoessa longipes</td>
<td>BOLD: AAI0439</td>
<td>sp. AB-2009</td>
<td>BOLD: AAO1754</td>
</tr>
<tr>
<td>Metopa sp.</td>
<td>BOLD: AAO8058</td>
<td>alderi</td>
<td>BOLD: ACP7788</td>
</tr>
<tr>
<td>Stenothoe frecanda</td>
<td>BOLD: ACH0349</td>
<td>marina</td>
<td>BOLD: ABW3806</td>
</tr>
<tr>
<td>Eritctonius brasiliensis</td>
<td>BOLD: ACH6329</td>
<td>punctatus</td>
<td>BOLD: ABW2100</td>
</tr>
<tr>
<td>Eritctonius sp.</td>
<td>BOLD: AAV1493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gnathopleustes pachychaetus</td>
<td>BOLD: AAG7139</td>
<td>Paraclestes bicuspis</td>
<td>BOLD: AAX1074</td>
</tr>
<tr>
<td>Rhachotropis sp. A</td>
<td>BOLD: ACH0271</td>
<td>aculeata</td>
<td>BOLD: AAB3310</td>
</tr>
<tr>
<td>Ampitheoe valida</td>
<td>BOLD: AAG3718</td>
<td></td>
<td>BOLD: ACM2468</td>
</tr>
<tr>
<td>Ampitheoe valida</td>
<td>BOLD: ACH6369</td>
<td></td>
<td>BOLD: ACM2468</td>
</tr>
<tr>
<td>Ampitheoe valida</td>
<td>BOLD: ACH6369</td>
<td></td>
<td>BOLD: ACM2468</td>
</tr>
<tr>
<td>Ampitheoe valida</td>
<td>BOLD: ACH6369</td>
<td></td>
<td>BOLD: ACM2468</td>
</tr>
<tr>
<td>Peramphitoe humeralis</td>
<td>BOLD: AAG7127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byblis veleronis millsi</td>
<td>BOLD: AAO4641</td>
<td>frigidus</td>
<td>BOLD: AAV5048</td>
</tr>
<tr>
<td>Edotia sublittoralis</td>
<td>BOLD: AAR0618</td>
<td>triloba</td>
<td>BOLD: AAF6319</td>
</tr>
<tr>
<td>Nebalia pseudotroncosoi</td>
<td>BOLD: ACI0616</td>
<td>Nebaliacea</td>
<td></td>
</tr>
<tr>
<td>Hexanauplia</td>
<td>BOLD: ACH6220</td>
<td>sp./longiremis</td>
<td>BOLD: AAG9533</td>
</tr>
<tr>
<td>Acartia longiremis / californiensis</td>
<td>BOLD: ACL8332</td>
<td>discaudatus</td>
<td>BOLD: AAW6351</td>
</tr>
<tr>
<td>Tortanus discaudatus</td>
<td>BOLD: ACL8333</td>
<td></td>
<td>BOLD: ACL8333</td>
</tr>
<tr>
<td>Hexanauplia</td>
<td>BOLD:ACL8742</td>
<td>BOLD:ACL8805</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td><em>Clausocalanus</em></td>
<td><em>lividus</em> N/A</td>
<td><em>lividus</em> BOLD:AAR9479</td>
<td></td>
</tr>
<tr>
<td><em>Scolecithricella</em></td>
<td>sp.</td>
<td><em>minor</em> BOLD:AAE4125</td>
<td></td>
</tr>
<tr>
<td><em>Oithona</em></td>
<td>sp.</td>
<td><em>similis</em> BOLD:AAG5172</td>
<td></td>
</tr>
<tr>
<td><em>Harpacticella</em></td>
<td><em>jejuensis</em> BOLD:ACS6380</td>
<td><em>Harpacticus flexus</em> BOLD:ACS7639</td>
<td></td>
</tr>
<tr>
<td>Pycnogonida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Nymphon</em></td>
<td><em>pixellae</em> BOLD:AAO2714</td>
<td><em>stroemi</em> BOLD:AAG3655</td>
<td></td>
</tr>
<tr>
<td><em>Nymphon</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Material 3.6 Biological traits of trans-Bering molluscs, arthropods, and polychaetes**

**Table S3.6.** Summary of the fossil record (when available) and biological traits for trans-Bering sister lineages of molluscs, arthropods, and polychaetes. Abbreviations: P Pacific; ArAt Arctic-Atlantic, At Atlantic.

<table>
<thead>
<tr>
<th>Trans-Bering sister clade</th>
<th>Stratigraphic distribution</th>
<th>Direction of the trans-Bering migration according to the fossil record</th>
<th>Depth in metres</th>
<th>Developmental mode</th>
<th>Thermal affinity</th>
<th>Habitat zone during adult stage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MOLLUSCS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Yoldiella orcia</em> vs. <em>Y. frigida, Y. nana, Y. inconspicua</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pecten albicans</em> vs. <em>P. jacobaeus, P. maximus</em></td>
<td></td>
<td>0–500 (Bergh &amp; Strand 2001; Palomares &amp; Pauly 2017)</td>
<td></td>
<td></td>
<td></td>
<td>Planktonic (Morzezen et al. 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Temperate (Palomares &amp; Pauly 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Benthic (Palomares &amp; Pauly 2017)</td>
</tr>
<tr>
<td>Species Pair</td>
<td>Time Period</td>
<td>Distribution</td>
<td>Ecological Zone</td>
<td>References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>------------------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Modiolus modiolus vs. M. modiolus</strong></td>
<td>Early Pliocene and Recent (Vermeij 1991)</td>
<td>Intertidal up to 100m (Halanych et al. 2013)</td>
<td>Benthic (Halanych et al. 2013)</td>
<td>Boreal (Palomares &amp; Pauly 2017) (Halanych et al. 2013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Onychoteuthis compacta vs. O. banksii</strong></td>
<td>Early to ArAt (Vermeij 1991)</td>
<td>0–800 (Palomares &amp; Pauly 2017)</td>
<td>Pelagic (Palomares &amp; Pauly 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Taonius borealis vs. T. pavo</strong></td>
<td>300–4000m (Quetglas et al. 2013)</td>
<td>Temperate (Palomares &amp; Pauly 2017)</td>
<td>Pelagic (Quetglas et al. 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rossia sp. vs. R. palpebrosa</strong></td>
<td>50–600 (Sweeney et al. 1992; Gotshall 1994)</td>
<td>Benthic (Gardiner &amp; Dick 2010)</td>
<td>Pelagic (Sweeney et al. 1992)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species Pair</td>
<td>Time Period</td>
<td>Distribution</td>
<td>Life History</td>
<td>Temperature</td>
<td>Depth Range</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td><em>Dendronotus venustus</em> vs. <em>D. frondosus</em></td>
<td>Intertidal up to 20m (Ekimova et al. 2015)</td>
<td>2–2161 (Stout et al. 2011; Ekimova et al. 2015)</td>
<td>Planktotrophic (Sisson 2005a; b)</td>
<td>Boreal-Polar (Gionet &amp; Aiken 1992)</td>
<td>Benthic (Palomares &amp; Pauly 2017)</td>
<td></td>
</tr>
<tr>
<td><em>Dendronotus patricki</em> vs. <em>D. robustus</em></td>
<td>0–33 (Palomares &amp; Pauly 2017)</td>
<td>24–60 (Ekimova et al. 2015)</td>
<td>Planktotrophic (Sisson 2005a; b)</td>
<td>Benthic (Palomares &amp; Pauly 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species comparison</td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ARTHROPODS**

<table>
<thead>
<tr>
<th>Species comparison</th>
<th>Ecology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pandalus borealis vs. P. borealis</td>
<td>60–100 (Nygård et al. 2007) 50–100 (Birkely &amp; Guilliksen 2003)</td>
</tr>
<tr>
<td>Eualus suckleyi vs. E. gaimardii</td>
<td>Planktonic (Grabe 2003)</td>
</tr>
<tr>
<td>Spirontocaris spinus vs. S. spinus, S. sp.</td>
<td></td>
</tr>
<tr>
<td>Hippolytidae vs. Eualus pusiolus</td>
<td></td>
</tr>
<tr>
<td>Eualus macilentus vs. E. avinus</td>
<td></td>
</tr>
<tr>
<td>Species Comparison</td>
<td>Lower Temperature Range</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Pagurus sp. vs. P. pubescens, P. sp.</td>
<td>0–30 (Kim &amp; Hong 2005; Selin et al. 2016)</td>
</tr>
<tr>
<td>Pagurus pectinatus vs. P. arcuatus</td>
<td>Planktonic (Kim &amp; Hong 2005)</td>
</tr>
<tr>
<td>Thysanoessa longipes vs. T. sp.</td>
<td>&gt;200 (Harvey et al. 2012)</td>
</tr>
<tr>
<td>Metopa sp. vs. M. alderi</td>
<td></td>
</tr>
<tr>
<td>Stenothoe frecanda vs. S. marina, Stenula sp.</td>
<td></td>
</tr>
<tr>
<td>Ericthonius brasiliensis, Ischyroceridae vs. E. punctatus</td>
<td></td>
</tr>
<tr>
<td>Gnathopleustes pachychaetus vs. Parapleustes bicuspis</td>
<td></td>
</tr>
<tr>
<td>Ampithoe valida, A. sp. vs. A. valida</td>
<td></td>
</tr>
<tr>
<td>Species Comparisons</td>
<td>Habitat</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------</td>
</tr>
<tr>
<td><em>Byblis veleronis</em>, <em>B. millsi</em> vs. <em>B. frigidus</em>, <em>B. gaimardii</em></td>
<td>Temperate-Polar</td>
</tr>
<tr>
<td><em>Edotia sublittoralis</em> vs. <em>E. triloba</em></td>
<td>Benthic</td>
</tr>
<tr>
<td><em>Nebalia pseudotroncosoi</em> vs. <em>Nebaliacea</em></td>
<td>&lt;5</td>
</tr>
<tr>
<td><em>Acartia longiremis</em>, <em>A. californiensis</em> vs. <em>A. longiremis</em>, <em>A. sp.</em></td>
<td>Estuaries and coastal lagoons</td>
</tr>
<tr>
<td><em>Tortanus discaudatus</em> vs. <em>T. discaudatus</em></td>
<td>Coastal and estuarine</td>
</tr>
<tr>
<td>Calanoida BOLD:ACL8742 vs. Calanoida BOLD:ACL8805</td>
<td>Pelagic</td>
</tr>
<tr>
<td><em>Clausocalanus lividus</em> vs. <em>C. lividus</em></td>
<td>0–200</td>
</tr>
<tr>
<td><em>Scolecithricella</em> sp. vs. <em>S. minor</em></td>
<td>Boreal</td>
</tr>
<tr>
<td><em>Harpacticella jejuensis</em>, Harpacticoida vs. <em>Harpacticus flexus</em></td>
<td>Pelagic</td>
</tr>
</tbody>
</table>
## Nymphon pixellae vs. Pantopoda
*(BOLD:AAG3655), Nymphon stroemi, Nymphonidae (BOLD:AAG4695), Phoxichilidiidae (BOLD:AGG6271), Anoplactylus lentus*

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>100–400</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

## POLYCHAETES

### Harmothoe rarispina vs. H. rarispina

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–824</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Eunoe nodosa vs. E. nodosa

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–824</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Harmothoe imbricata vs. H. imbricata

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4000</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Lepidonotus squamatus vs. L. squamatus

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–824</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Pholoe baltica vs. P. baltica

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4000</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Pholoe minuta vs. P. minuta

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4000</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Nephtys punctata vs. N. punctata

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–824</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
</tbody>
</table>

### Alitta sp. vs. A. virens

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–824</td>
<td>Temperate</td>
<td>Benthic</td>
</tr>
<tr>
<td>Species Comparison</td>
<td>Range</td>
<td>Habitat</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Nereis pelagica vs. N. pelagica</td>
<td>12–18</td>
<td>Planktonic</td>
</tr>
<tr>
<td>Glycera capitata vs. G. capitata</td>
<td>12–18</td>
<td>Planktonic</td>
</tr>
<tr>
<td>Glycera capitata vs. G. capitata, G. capitata</td>
<td>12–18</td>
<td>Planktonic</td>
</tr>
<tr>
<td>Phyllopoce sp vs P. sp.</td>
<td>6–38</td>
<td></td>
</tr>
<tr>
<td>Phyllopoce groelandica vs. P. groelandica</td>
<td>0–180</td>
<td>Polar</td>
</tr>
<tr>
<td>Myxicola infundibulum vs. M. infundibulum</td>
<td>200–1500</td>
<td>Benthic</td>
</tr>
<tr>
<td>Nothria conchylega vs. N. conchylega</td>
<td>0–180</td>
<td>Benthic</td>
</tr>
<tr>
<td>Pectinaria granulata vs. P. granulata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Praxillella praetermissa vs. P. praetermissa</td>
<td>0–50</td>
<td>Temperate</td>
</tr>
<tr>
<td>Cirratulus cirratus vs. C. cirratus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalibregma inflatum vs. S. inflatum</td>
<td>37–177</td>
<td>Polar</td>
</tr>
</tbody>
</table>
Supplementary Material 3.7. List of trans-Bering sister clades of polychaetes retrieved from Carr (2010) and reanalyzed in Chapter III.

<table>
<thead>
<tr>
<th>Class</th>
<th>Trans-Bering sister clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polychaeta</td>
<td>Pectinaria granulata vs. P. granulata</td>
</tr>
<tr>
<td></td>
<td>Cirratulus cirratus vs. C. cirratus</td>
</tr>
<tr>
<td></td>
<td>Myxicola infundibulum vs. M. infundibulum</td>
</tr>
<tr>
<td></td>
<td>Lepidonotus squamatus vs. L. squamatus</td>
</tr>
<tr>
<td></td>
<td>Alitta sp. vs. A. virens</td>
</tr>
<tr>
<td></td>
<td>Glycera capitata vs. G. capitata, G. capitata</td>
</tr>
<tr>
<td></td>
<td>Nephtys punctata vs. N. punctata</td>
</tr>
<tr>
<td></td>
<td>Harmothoe rarispina vs. H. rarispina</td>
</tr>
<tr>
<td></td>
<td>Nothria conchylega vs. N. conchylega</td>
</tr>
<tr>
<td></td>
<td>Scalibregma inflatum vs. S. inflatum</td>
</tr>
<tr>
<td></td>
<td>Nereis pelagica vs. N. pelagica</td>
</tr>
<tr>
<td></td>
<td>Harmothoe imbricata vs. H. imbricata, H. imbricata</td>
</tr>
<tr>
<td></td>
<td>Pholoe baltica vs. P. baltica</td>
</tr>
<tr>
<td></td>
<td>Harmothoe imbricata vs. H. imbricata</td>
</tr>
<tr>
<td></td>
<td>Phyllodoce sp. vs. P. sp.</td>
</tr>
<tr>
<td></td>
<td>Pholoe minuta vs. P. minuta</td>
</tr>
<tr>
<td></td>
<td>Eunoe nodosa vs. E. nodosa</td>
</tr>
<tr>
<td></td>
<td>Praxillella praetermissa vs. P. praetermissa</td>
</tr>
<tr>
<td></td>
<td>Phyllodoce groelandica vs. P. groelandica</td>
</tr>
<tr>
<td></td>
<td>Glycera capitata vs. G. capitata</td>
</tr>
</tbody>
</table>
Supplementary Material 3.8 Statistical analysis R commands

This script provides a simple statistical test of whether we expect to have no divergences between 2.4 and 3 MY by chance alone. According to the geological history of the Bering Strait, trans-Bering migrations are not expected to have occurred when the Strait was closed during the glacial maximum (2.4–3 Ma). After iterative calibration, an absence of estimated dates in this range was observed for three phyla, but for Mollusca one pair fell in this range with a value of 2.97 Ma. That result would be predicted if there were no trans-Bering migrations during the period 2.4-3 MY. However, could this finding arise by chance alone, unrelated to the geological history of the Bering Strait?

First, many artificial sets of divergence time values will be generated, to create a null distribution, against which the observed number of divergences in the range 2.4-3 Ma will be compared. The sample size of points generated will be selected according to the number of trans-Bering sisters used for the calibration in each phylum, after omitting the pairs rejecting the clock hypothesis and/or that were beyond the saturation threshold. Values between 0 and 7 will be randomly selected from a uniform distribution. The maximum value of 7 MY is based upon the oldest possible date for the first opening of the Bering Strait according with the literature, i.e. when the Bering Strait was closed due to glaciation.

###ECHINODERMATA

#Generating a vector with 15 values between 0 and 7, using a uniform distribution.
x<-c(runif(15, 0, 7))

#Create a plot to verify the number of points and their distribution.
x
plot(x)

#Sorting and printing the vector to the screen.
x1<-c(sort(x))
x1

#Counting the values falling inside of the range 2.4-3 MY.
length(which(2.4 < x1 & x1 <3))

###Performing the test by first generating 100K data sets randomly under the uniform distribution and counting the number with values between 2.4-3 Ma.
glacialmax <- function(){
x<-c(runif(15, 0, 7))
length(which(2.4 < x & x <3 ))
}
distribution <- replicate(100000,glacialmax())
distribution
hist(distribution)
length(which(distribution == 0))

#This next line will provide the p-value.
#i.e. How often does the null distribution contain a value as extreme as our observed value?
#the observed value was 0 for echinoderms.
length(which(distribution == 0))/100000

###MOLLUSCA

#Generating a vector with 22 values between 0 and 7, using a uniform distribution.
x<-c(runif(22, 0, 7))

#Create a plot to verify the number of points and their distribution.
x
plot(x)

#Sorting and printing the vector to the screen.
x1<-c(sort(x))
x1

#Counting the values falling inside of the range 2.4-3 MY.
length(which(2.4 < x1 & x1 <3))

###Performing the test by first generating 100K data sets randomly under the uniform distribution and counting the number with values between 2.4-3 Ma.
glacialmax <- function(){
x<-c(runif(22, 0, 7))
length(which(2.4 < x & x <3 ))
}
distribution <- replicate(100000,glacialmax())
distribution
hist(distribution)
length(which(distribution == 1))

#This next line of code will provide our p-value.
#i.e. How often does the null distribution contain a value as extreme as our observed value?
#The observed value for molluscs was 1
length(which(distribution == 1))/100000

###ARTHROPODA

#Generating a vector with 17 values between 0 and 7, using a uniform distribution.
x<-c(runif(17, 0, 7))

#Create a plot to verify the number of points and their distribution.
x
plot(x)

# Sorting and printing the vector to the screen.
x1<-c(sort(x))
x1

# Counting the values falling inside of the range 2.4-3 MY.
length(which(2.4 < x1 & x1 <3))

### Performing the test by first generating 100K data sets randomly under the uniform distribution and counting the number with values between 2.4-3 Ma.
glacialmax <- function(){
x<-c(runif(17, 0, 7))
  length(which(2.4 < x & x <3))
}
distribution <- replicate(100000,glacialmax())
distribution
hist(distribution)
length(which(distribution == 0))

# This next line of code will provide our p-value.
# i.e. How often does the null distribution contain a value as extreme as our observed value?
# The observed value for arthropods was 0.
length(which(distribution == 0))/100000

### POLYCHAETA
# Generating a vector with 19 values between 0 and 7, using a uniform distribution.
x<-c(runif(19, 0, 7))

# Create a plot to verify the number of points and their distribution.
x
plot(x)

# Sorting and printing the vector to the screen.
x1<-c(sort(x))
x1

# Counting the values falling inside of the range 2.4-3 MY.
length(which(2.4 < x1 & x1 <3))

### Performing the test by first generating 100K data sets randomly under the uniform distribution and counting the number with values between 2.4-3 Ma.
glacialmax <- function(){
x<-c(runif(19, 0, 7))
  length(which(2.4 < x & x <3))
}
distribution <- replicate(100000,glacialmax())
distribution
hist(distribution)
length(which(distribution == 0))

# This next line of code will provide our p-value.
# i.e. How often does the null distribution contain a value as extreme as our observed value?
# The observed value for polychaetes was 0.
length(which(distribution == 0))/100000

### OVERALL TEST INCLUDING ECHINODERMATA, MOLLUSCA, ARTHROPODA, AND POLYCHAETA
# Generating a vector with 73 values between 0 and 7, using a uniform distribution.
x <- c(runif(73, 0, 7))

# Create a plot to verify the number of points and their distribution.
x
plot(x)

# Sorting and printing the vector to the screen.
x1 <- c(sort(x))
x1

# Counting the values falling inside of the range 2.4-3 MY.
length(which(2.4 < x1 & x1 < 3))

### Performing the test by first generating 100K data sets randomly under the uniform distribution and counting the number with values between 2.4-3 Ma.
glacialmax <- function(){
x <- c(runif(73, 0, 7))
length(which(2.4 < x & x < 3))
}
distribution <- replicate(100000, glacialmax())
distribution
hist(distribution)
length(which(distribution == 1))

# This next line of code will provide our p-value.
# i.e. How often does the null distribution contain a value as extreme as our observed value?
# The observed value for the dataset including all 4 taxa was 1.
length(which(distribution == 1))/100000
Supplementary Material 4.1 List of trans-isthmus sister clades of echinoderms, molluscs, and arthropods retrieved from the literature (Hart et al. 1997; Lessios 2008; O'Dea et al. 2016) and reanalyzed in Chapter IV

<table>
<thead>
<tr>
<th>Class</th>
<th>Trans-isthmus sister clade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECHINODERMATA</strong></td>
<td></td>
</tr>
<tr>
<td>Echinoidea</td>
<td><em>Meoma ventricosa grandis</em> vs. <em>M. ventricosa</em></td>
</tr>
<tr>
<td></td>
<td><em>Diadema mexicanum</em> vs. <em>D. antillarum</em></td>
</tr>
<tr>
<td></td>
<td><em>Tripneustes depressus</em>, <em>T. gratilla</em> vs. <em>T. ventricosus</em></td>
</tr>
<tr>
<td></td>
<td><em>Eucidaris thouarsi</em> vs. <em>E. tribuloides</em></td>
</tr>
<tr>
<td></td>
<td><em>Astropyga pulvinata</em> vs. <em>A. magnifica</em></td>
</tr>
<tr>
<td></td>
<td><em>Echinometra vanbrunti</em> vs. <em>E. lucunter</em></td>
</tr>
<tr>
<td></td>
<td><em>Arbacia spatuligera</em> vs. <em>A. lixula</em></td>
</tr>
<tr>
<td></td>
<td><em>Lytechinus semituberculatus</em> vs. <em>L. variegatus</em>, <em>L. williamsi</em></td>
</tr>
<tr>
<td></td>
<td><em>Mellita quinquesperforata</em> vs. <em>M. notabilis</em>, <em>M. kanakoffi</em></td>
</tr>
<tr>
<td></td>
<td><em>Mellita longifissa</em>, <em>M. grantii</em> vs. <em>M. isometra</em>, <em>M. tenuis</em>, <em>M. quinquesperforata</em></td>
</tr>
<tr>
<td>Asteroidea</td>
<td><em>Oreaster reticulatus</em> vs. <em>O. occidentalis</em></td>
</tr>
<tr>
<td><strong>MOLLUSCA</strong></td>
<td></td>
</tr>
<tr>
<td>Gastropoda</td>
<td><em>Conus ermineus</em> vs. <em>C. purpurascens</em></td>
</tr>
<tr>
<td></td>
<td><em>Strombus peruvianus</em> vs. <em>S. raninus</em></td>
</tr>
<tr>
<td></td>
<td><em>Tegula verrucosa</em> vs. <em>T. viridula</em></td>
</tr>
<tr>
<td></td>
<td><em>Crepidula incurva</em> &quot;Peru&quot; vs. <em>C. convexa&quot;</em> &quot;Bocas&quot;</td>
</tr>
<tr>
<td></td>
<td><em>Littoraria irrorata</em> vs. <em>L. variegata</em></td>
</tr>
</tbody>
</table>
Gastropoda

Conus gladiator vs. E. mus, E. tabidus
Nerita funiculata, vs. N. tessellata, N. fulgarans, N. senegalensis
Macrocyprea cervinetta vs. M. zebra, M. cervus
Echinolittorina aspera, E. dubiosa, E. tenuistriata vs. E. interrupta
Echinolittorina modesta, E. compresa vs. E. ziczac, E. ziczac E. soroziczac
Strombus gracilior vs. S. alatus, S. pugilis
Conus perplexus vs. puncticulatus
Bulla gouldiana, B. punctulata vs. B. mabillei
Littoraria nebulosa vs. L. varia, L. zebra, L. irrorara, L. variegate
Nerita scabricosta vs. N. peloronta, N. versicolor
Echinolittorina galapagiensis vs. E. tuberculata, E. vermeiji, E. granosa, E. miliaris, E. helenae
Littoraria rosewateri vs. L. tessellata
Elysia diomedea vs. E. crispa, E. clarki
Conus regius vs. C. bartschii, C. brunneus
Echinolittorina apicina, E. paytensis vs. E. riisei
Tegula corteziana vs. T. fasciata

Bivalvia

Barbatia candida vs. B. reeveana
Barbatia illota vs. B. tenera
Brachidontes semilaevis vs. E. exustus-Gulf- Atlantic
Brachidontes adamsianus vs. E. exustus-W. Caribbean
Anadara chemnitzi vs. A. nux
Arca mutabilis vs. A. imbricata
Arcopsis adamsi vs. A. solida
ARTHROPODA

Sesarma rhizophorae vs. S. reticulatum, S. sp nr. reticulatum, S. curacaoense

Alpheus antepenultimus A vs. A. chacei

(Sesarma sulcatum, S. aequatoriale) vs. S. crassipes

Alpheus rostratus vs. A. paracrinitus spot

Alpheus colombiensis vs. A. estuarensis

Alpheus websteri vs. A. websteri

Alpheus cylindricus vs. A. cylindricus

Synalpheus brevicarpus vs. S. brevicarpus

Alpheus canalis- sp.B (blue) = millsae vs. A. nuttingi

Alpheus floridanus B’ vs. A. floridanus B

Alpheus panamensis vs. A. formosus-sp.A

Alpheus paracrinitus no spot vs. A. paracrinitus no spot

Synalpheus digueti vs. S. minus

Alpheus bouvieri vs. A. bouvieri

Xiphopenaeus riveti vs. X. sp. 2

Alpheus malleator vs. A. malleator

Penaeus vannamei vs. P. duorarum, P. paulensis, P. setiferus

Synalpheus fritzmuelleri vs. S. fritzmuelleri

Alpheus umbo vs. A. schmitti

Alpheus saxidomus vs. A. simus

Alpheus floridanus A’ vs. A. floridanus A

Alpheus normanni vs. A. normanni Brazil
<table>
<thead>
<tr>
<th>Malacostraca</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Emerita rathbunae</em> vs. <em>E. brasiliensis, E. benedicti, E. talpoida</em></td>
<td></td>
</tr>
<tr>
<td><em>(Alpheus naos, A. aequus)</em> vs. <em>A. christofferseni</em></td>
<td></td>
</tr>
<tr>
<td><em>Alpheus cristulifrons</em> vs. <em>A. cristulifrons</em></td>
<td></td>
</tr>
<tr>
<td><em>Austinixa felipensis</em> vs. <em>A. cristata clade</em></td>
<td></td>
</tr>
<tr>
<td><em>Synalpheus bannerorum</em> vs. <em>S. dominicensis</em></td>
<td></td>
</tr>
</tbody>
</table>