12-2014

Phylogenomics and Historical Biogeography of the Gooseneck Barnacle Pollicipes elegans

Sergio Andres Marchant Rojas
Clemson University

PHYLOGENOMICS AND HISTORICAL BIOGEOGRAPHY OF THE GOOSENECK BARNACLE *POLLICIPES ELEGANS*

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biological Sciences

---

by
Sergio Andrés Marchant Rojas
December 2014

Accepted by:
Dr. Saara DeWalt, Committee Chair
Dr. Peter Marko
Dr. Amy Moran
Dr. Christina Wells
Dr. Amy Lawton-Rauh

ABSTRACT

This dissertation explores the systematics, biogeography, and genomics of the gooseneck barnacle *Pollicipes elegans*, a marine crustacean of the tropical Eastern Pacific. In Chapter 1, I provide a broad framework for my research by introducing and focusing on the long-standing debate of the mechanisms behind the latitudinal gradient in species diversity, which provided the initial motivation for using *Pollicipes elegans* as a model system to study the mechanisms leading to genetic differentiation and speciation in tropical regions. In Chapter 2, I examine the genetic structure, infer patterns of connectivity across the warm tropical waters of the eastern Pacific, and reconstruct the biogeographic history of *P. elegans* using a statistical phylogeographic framework. Using mitochondrial DNA sequences, I found strong evidence supporting an out-of-the tropics model of speciation in *P. elegans*, with a clear phylogeographical break between populations in Mexico and all populations to the south. In Chapter 3, I added sequence data from six nuclear genes to the analysis of genetic structure and found strong evidence for two cryptic species within the nominal *P. elegans* that likely originated by allopatric speciation across the Central American Gap. I estimated the divergence times between peripheral and central populations, and the effective population sizes of these populations, and found again support for an out-of-the-tropics model of diversification. In Chapter 4, I used RNA sequencing of individuals of *P. elegans* from each cryptic species to assemble the first transcriptome for this taxon. Data mining of the transcriptome allowed me to identify microsatellite and single nucleotide polymorphism (SNP) markers to be used in future research. Analyses using the SNP dataset revealed

evidence for 11 genes under natural selection between the two cryptic species; the genes that were identified may be influenced by spatial variation in sea surface temperature in the tropical eastern Pacific. Lastly, in Chapter 5, I provide guidelines for future studies that should be pursued to help elucidate patterns, mechanisms, and consequences of latitudinal gradients of temperature in the process of allopatric speciation. The phylogeographic and demographic reconstruction for *P. elegans* in this dissertation provide evidence of the role that temperature may play in population differentiation associated with speciation. The transcriptome analyses provided a large set of genetic markers and a list of candidate genes under selection, a crucial first step in the description of the genetic basis of local thermal adaptation in tropical regions. The information generated in this dissertation provides a novel empirical system that can help elucidate the evolution of tropical diversity and can be used to potentially predict the future impacts of climate change on tropical species.

DEDICATION

I dedicate my dissertation to my family in Chile, Brazil, and Colombia. Without your love, support and encouragement I would never had been able to accomplish this dissertation. I especially dedicate this dissertation to my wife Claudia: you have been a continuous source of encouragement and love at all times; also to my son Daniel, for inspiring me to pursue higher goals and do better things for him.

AKNOWLEDGMENTS

I am grateful to the people that helped me accomplish this professional goal. I thank my family and friends that kept me enthusiastic through my many years in graduate school. I would also like to thank Clemson University for being supportive of many student organizations that I was part of: Society for the Advancement of Chicanos/Latinos and Native-Americans in Science, Fulbright Fellows and Friends, and the Biological Sciences Graduate Student Association. Without the support of the friends that I made in those organizations, graduate life outside the laboratory would have felt empty.

I am indebted to all of the government officers that helped me to collecting information, samples, and permits to work on *Pollicipes elegans*. I would like to give special thanks to Enrique Barraza, Pancho Atoche and Carmen Yamashiro.

I am grateful for the advice from my committee: Drs. Peter Marko, Amy Moran, Saara DeWalt, Amy Lawton-Rauh, and Christina Wells. I owe special gratitude to Peter for being supportive of my research ideas and for helping me to become a better scientist (even after I booked non-refundable airplane tickets on the wrong dates!), and Saara for giving me advice and guidance in times of need. I am also grateful to the Fulbright Scholar Program, Comisión Nacional de Ciencia y Tecnología, and the National Science Foundation for funding my research and graduate studies at Clemson University.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION: *POLLICIPES ELEGANS* AND THE LATITUDINAL SPECIES DIVERSITY GRADIENT

The latitudinal species diversity gradient is the strongest and most important biogeographic pattern on Earth. Data accumulated over the last 200 years show that diversity peaks in the tropics and declines with increasing latitude, forming the most common distribution pattern in plants and animals (von Humbolt, 1808; Wallace, 1878; Fischer, 1960; Simpson, 1964). Considerable progress has been made towards understanding the causes of this global pattern (Pianka, 1966; Stevens, 1989, 1992; Roy *et al.*, 1994; Chown & Gaston, 2000; Willig *et al.*, 2003; Hillebrand, 2004b, a; Wiens *et al.*, 2006; Mittelbach *et al.*, 2007; Weir & Schluter, 2007), but there is still an ongoing debate among ecologists and evolutionary biologists about the mechanisms that control speciation, extinction, and migration along this gradient and the spatial scale at which these mechanisms operate (Wallace, 1878; Hutchinson, 1959; Terborgh, 1973; Stebbins, 1974; Rosenzweig, 1992; Jablonski *et al.*, 2006).

Previous studies have shown that tropics not only hold higher diversity but also that the tropics may be a major source of diversity for extra-tropical regions when species with tropical origin cross the tropical/extra-tropical boundaries (Rosenzweig, 1992, 1995; Hawkins & Porter, 2001; Wiens & Donoghue, 2004). Therefore, understanding the mechanisms by which tropical diversity is generated is fundamental to predict the potential effects of climate change in extra-tropical diversity. However, in the marine realm, little is known about patterns of gene flow, genetic drift, and natural

selection within the tropics and across the thermal gradients at the transitions between the tropics and extra-tropical regions

By studying the evolutionary history of species that are in the process of speciation at the interface between tropical and extra-tropical regions, it is possible to evaluate the roles that local adaptation, migration, and genetic drift play in the speciation process and, therefore, disentangle how the tropical diversity engine works. Understanding the effectiveness of the tropics as a barrier to gene flow as well as the role of natural selection across this thermal regime is essential for understanding the biogeographic mechanisms that lead to biotic interchange between the hemispheres. It is also pivotal for generating reliable models that can predict the effects of increasing temperatures, which are expected with human-induced climate change.

Pollicipes elegans (Lesson, 1831) as a model system for studies of biotic exchange across hemispheres

The four living species of *Pollicipes* are thought to be descendants from a common ancestor that lived in the Tethys Sea, an ocean that existed between North Africa and Eurasia, prior to 20 million years ago or MYA (Newman & Killingley, 1985; Newman, 1992a; Van Syoc *et al.*, 2010). The wide distribution of *Pollicipes* and its near absence in tropical regions suggest that temperature may limit survival and reproduction (Newman, 1992a). According to molecular clock estimates, *P. elegans*, *P. pollicipes* (=*P. cornucopia*) and *P. caboverdensis* diverged roughly 25-34 million years ago; *P. polymerus* is more distantly related, diverging 55-66 MYA from the other three species.

Pollicipes elegans, a marine gooseneck barnacle, is an ideal species to evaluate biotic exchange across warm tropical waters because of its disjunct distribution (Laguna, 1990; Van Syoc, 1994). This species inhabits wave-exposed rocks **(Fig. 1.1a)** from Punta Abreojos, Mexico to Southern Peru (Newman & Killingley, 1985; Kameya & Zeballos, 1988; Laguna, 1990), with several large distributional gaps between the equator and southern Mexico, with the exception of a few geographically isolated populations in El Salvador and Costa Rica (Newman & Foster, 1987; Laguna, 1990; Newman, 1992a). This distribution (**Fig. 2.1**) provides an unusual opportunity to compare recently diverged populations living under a wide range of temperatures. Because this species may be found in a wide range of temperatures (**Fig. 1.1b**), natural selection and local adaptation may be occurring in warm tropical populations or in colder habitats at the edge of the species' distribution.

My thesis is that the disjunct distribution of *P. elegans* is the result of vicariance of an ancestral population that used to be more abundant in the tropics rather than dispersal from one hemisphere to the other, which is often proposed for species with largely antitropical distributions (taxa that is present in both hemispheres and is absent from the tropics). I evaluate this hypothesis in Chapter 2 using model comparison of simulated and empirical data collected from one mitochondrial gene. In addition, peripheral populations of *P. elegans* are likely different species because the steep gradient combined with the large gaps in the distribution and oceanic circulation patterns seem to be an effective barrier to gene flow. I evaluate this using species delimitation methods based on the coalescent using data collected from six nuclear and

one mitochondrial gene. Finally, given that marine species often have large effective population sizes, I expect to observe high effects of natural selection and low effects of genetic drift contributing to the genetic differentiation between divergent groups. I address this in Chapter 4 by surveying transcriptome data from multiple individuals and by estimating the levels of genetic differentiation across thousands of single nucleotide polymorphisms.

REFERENCES

Chown, S.L. & Gaston, K.J. (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology & Evolution*, **15**, 311-315.

Fischer, A.G. (1960) Latitudinal variations in organic diversity. *Evolution*, **14**, 64-81.

Hawkins, B.A. & Porter, E.E. (2001) Area and the latitudinal diversity gradient for terrestrial birds. *Ecology Letters*, **4**, 595-601.

Hillebrand, H. (2004a) On the generality of the latitudinal diversity gradient. *American Naturalist*, **163**, 192-211.

Hillebrand, H. (2004b) Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series*, **273**, 251-267.

Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, **93**, 145-159.

Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102-106.

Kameya, A. & Zeballos, J. (1988) Distribucion y densidad de percebes Pollicipes elegans (Crustacea: Cirripedia) en el mediolitoral peruano (Yasila, Piura; Chilca, Lima). *Boletin-Instituto del Mar del Peru*, **12**, 5-22.

Laguna, J.E. (1990) Shore Barnacles (Cirripedia, Thoracica) and a Revision of Their Provincialism and Transition Zones in the Tropical Eastern Pacific. *Bulletin of Marine Science*, **46**, 406-424.

Lesson, R. (1831) Histoire naturelle des Mollusques, Annélides et Vers recueillis dans le voyage autour du monde de la corvette de sa majesté, la "Coquille," exécuté pendant les années 1822, 23, 24 et 1825, sous le commandement du capitaine Duperrey. *Paris, Zoologie*, **2**, 25-471.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M. & Turelli, M. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315-331.

Newman, W. (1992a) Biotic cognates of eastern boundary conditions in the Pacific and Atlantic: relicts of Tethys and climate change. *Proceedings of the San Diego Society of Natural History* (ed by, pp. 1-7.

Newman, W.A. & Killingley, J.S. (1985) The Northeast Pacific Intertidal Barnacle Pollicipes-Polymerus in India - a Biogeographical Enigma Elucidated by O-18 Fractionation in Barnacle Calcite. *Journal of Natural History*, **19**, 1191-1196.

Newman, W.A. & Foster, B.A. (1987) Southern-Hemisphere Endemism among the Barnacles - Explained in Part by Extinction of Northern Members of Amphitropical Taxa. *Bulletin of Marine Science*, **41**, 361-377.

Pianka, E.R. (1966) Latitudinal gradients in species diversity: a review of concepts. *American Naturalist*, **100**, 33-46.

Rosenzweig, M.L. (1992) Species diversity gradients: we know more and less than we thought. *Journal of mammalogy*, **73**, 715-730.

Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
Roy, K., Jablonski, D. & Valentine, J.W. (1994) Eastern Pacific molluscan provinces and latitudinal diversity gradient: no evidence for" Rapoport's rule". *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 8871.
Simpson, G.G. (1964) Species density of North American recent mammals. *Systematic zoology*, **13**, 57-73.
Stebbins, G.L. (1974) Flowering plants: evolution above the species level. *London: Arnold xviii, 399p..Illustrations.General (KR, 197500089)*,
Stevens, G.C. (1989) The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, **133**, 240-256.
Stevens, G.C. (1992) The elevational gradient in altitudinal range: an extension of Rapoport's latitudinal rule to altitude. *The American Naturalist*, **140**, 893-911.
Terborgh, J. (1973) On the notion of favorableness in plant ecology. *The American Naturalist*, **107**, 481-501.
Van Syoc, R.I. (1994) Genetic divergence between subpopulations of the eastern Pacific goose barnacle Pollicipes elegans: mitochondrial cytochrome c subunit 1 nucleotide sequences. *Molecular marine biology and biotechnology*, **3**, 338-346.
Van Syoc, R.J., Fernandes, J.N., Carrison, D.A. & Grosberg, R.K. (2010) Molecular phylogenetics and biogeography of Pollicipes (Crustacea: Cirripedia), a Tethyan relict. *Journal of experimental marine biology and ecology*, **392**, 193-199.
von Humbolt, A. (1808) *Ansichten der Natur mit wissenschaftlichen Erlauterungen Tubingen*. J. G. Cotta, Tubingen, Germany.
Wallace, A.R. (1878) *Tropical nature, and other essays*. Macmillan and co., London.
Weir, J.T. & Schluter, D. (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574.
Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639-644.
Wiens, J.J., Graham, C.H., Moen, D.S., Smith, S.A. & Reeder, T.W. (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *American Naturalist*, **168**, 579-596.
Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273-309.

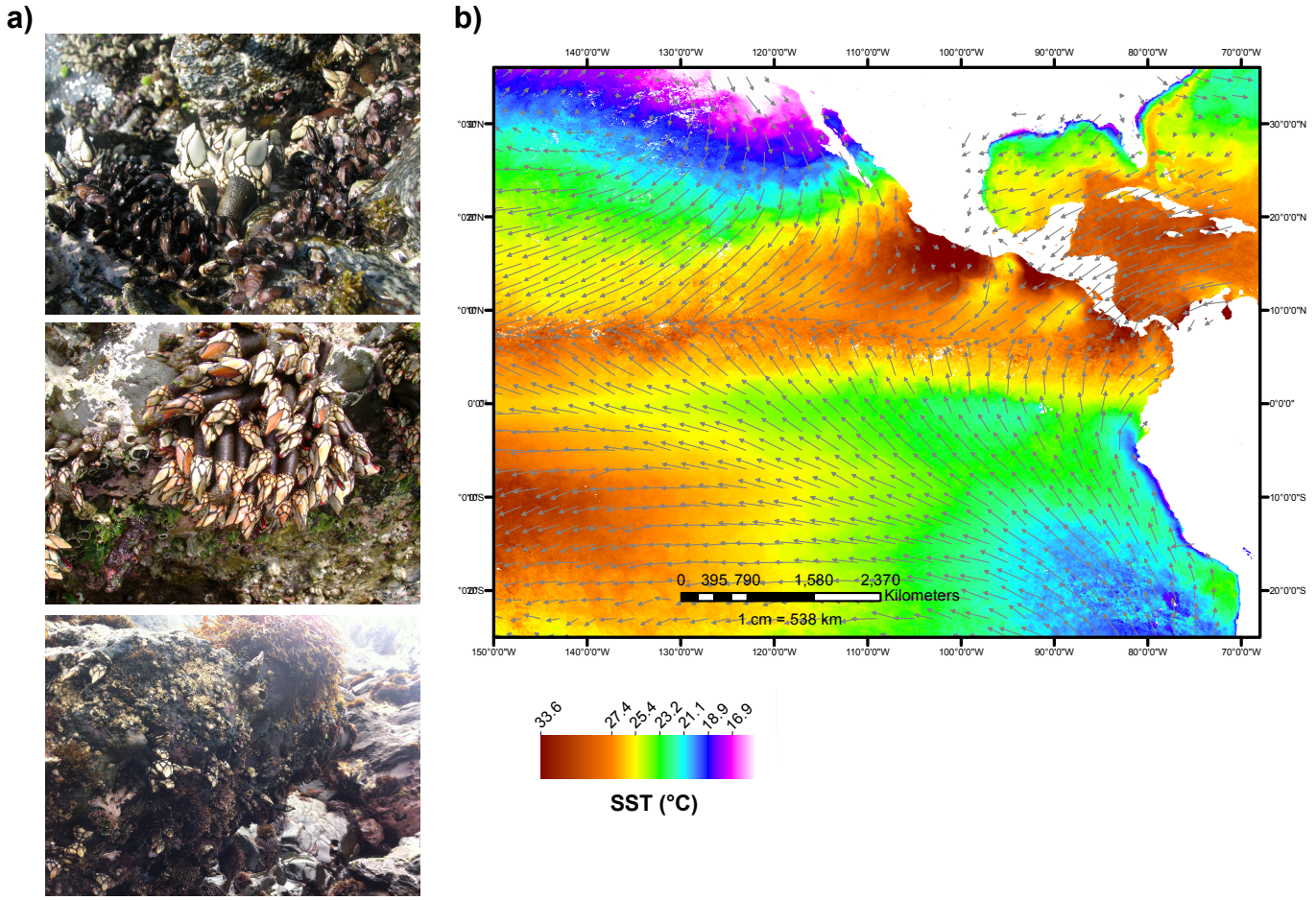

Figure 1.1: a. *Pollicipes elegans* collected in the rocky intertidal shore of Mexico, El Salvador and Peru. b. Map of the eastern tropical Pacific region showing average sea surface temperature and wind stress. Grey arrows represent wind stress climatology data after (Pennington, *et al*. 2006). Sea surface temperature (SST) corresponds to average values of November 2013. Satellite data obtained from MODIS-Aqua mission, 4km resolution (GHRSSTL41sst).

CHAPTER TWO

TROPICAL EXTINCTION OR TRANS-PACIFIC DISPERSAL? THE ORIGINS OF THE DISJUNCT DISTRIBUTION OF THE GOOSENECK BARNACLE *POLLICIPES ELEGANS*

ABSTRACT

By studying species with disjunct distributions, biogeographers can evaluate factors controlling species ranges, limits on gene-flow, and the process of allopatric speciation. Here, I use phylogeographic and population genetic studies of the barnacle *Pollicipes elegans* to discriminate between two primary hypotheses about the origin of disjunct distributions: trans-tropical stepping-stone dispersal versus vicariance through tropical extinction. Mitochondrial cytochrome c oxidase-1 sequences were gathered from individuals collected in Mexico, El Salvador, and Peru to reconstruct the demographic history of *Pollicipes elegans.* Sequence diversity statistics were estimated and phylogenetic analyses were conducted for eight sites within the species' range. Demographic parameters were inferred with coalescent-based methods. Four demographic models of disrupted distribution patterns in *P. elegans* were compared with Approximate Bayesian Computational (ABC) methods. Nucleotide diversity peaked in the centre of the species' range in El Salvador and was lower at higher latitudes in Mexico and Peru. Haplotypes from Salvadoran sites also showed a deeper coalescence or time to a most recent common ancestor. The deepest phylogeographical break occurred between Mexico and all sites to the south; isolation-with-migration analyses showed no evidence of significant gene flow between any of the three regions. ABC testing found strong support for an out-of-the tropics model of near antitropicality for *P.*

elegans. Moreover, I found little evidence consistent with a stepping-stone history of trans-tropical colonization, but instead found strong evidence for a tropical origin model for the largely antitropical distribution of *P. elegans*. Sea surface temperature and habitat suitability are likely mechanisms driving vicariance by tropical extinction.

INTRODUCTION

Species with disjunct or geographically discontinuous ranges are important systems for understanding the factors controlling species' distributions, population connectivity, and the process of allopatric speciation (Darwin, 1859; Ekman, 1953; Briggs, 1987; Wiley, 1988; Lindberg, 1991; Palumbi, 1994; Parmesan, 2006; Cowen & Sponaugle, 2009). Two main biogeographic mechanisms that have been proposed to explain range disjunctions: dispersal across a preexisting barrier or uninhabited region; or vicariance, the fragmentation of a species' ancestral range through extinction in the uninhabited region. In both cases, allopatric speciation can take place if the gap in the species' range sufficiently limits gene flow (Crisp *et al.*, 2010) and either natural selection or genetic drift allows the development of phylogenetically distinct species.

Distinguishing dispersal from vicariance remains a major challenge of biogeography. Dispersal and vicariance may be differentiated when the ages of both the barrier and the population separations are known (de Queiroz, 2005a), thus understanding the timing of divergent events is crucial when distinguishing such hypotheses (Hunn & Upchurch, 2001). However, for many marine taxa, resolution of the mechanisms leading to the formation of disjunct distributions is particularly difficult given that many of the barriers to dispersal in the sea are likely geologically short-lived

(Hellberg, 1998; Marko, 1998). Genetic data have been used to test hypotheses about range disjunctions in the sea (Bowen & Grant, 1997; Burridge, 2002; Keever *et al.*, 2009), but characterizing how species' distributions evolve with genetic data is often difficult: as time goes by, the telltale genetic diversity signatures of major demographic events will be erased by the evolutionary forces of mutation, gene flow, genetic drift, and natural selection.

However, the fragmented geographic distribution of the gooseneck barnacle, *Pollicipes elegans,* (Lesson, 1830) provides a potentially useful model to investigate the origin of a disjunct distribution in the tropical eastern Pacific (TEP). Extra-tropical populations of this single nominal species are abundant in Mexico and Peru, but records of this species are rare between central Mexico and northern Peru (**Fig. 2.1**). Within the warmest waters of the tropical eastern convergence - located predominantly to the north of the equator - *P. elegans* is known to be abundant in El Salvador, but only a handful of observations of the species are known from the two intervening areas to the north and south of El Salvador (Newman & Foster, 1987; Laguna, 1990; Newman, 1992b; Van Syoc, 1994). Although the unusual distribution of *P. elegans* in the TEP has been referred to as antitropical (Laguna, 1990; Van Syoc, 1994) or absent from the tropics, a more recent surveys of the region and the literature indicate (**Fig. 2.1**) that the fragmented distribution of *P. elegans* is best described as parantitropical (Ekman, 1953), a transtropical distribution in which a species is more abundant towards the range periphery than towards the center of its latitudinal range (Newman & Foster, 1987; Newman, 1992b).

The disjunctions in the geographic range of *P. elegans* could have been formed by either dispersal or vicariance. First, populations of *P. elegans* in the warmest regions of TEP could be intermediate populations colonized by extra-tropical populations from one side of the TEP. Such a trans-tropical dispersal model posits that extra-tropical taxa dispersed across the tropics in a steeping-stone fashion, from one hemisphere to the other, likely during periods of cooler climate (Darwin, 1859; Ekman, 1953; Carlquist, 1981; Lindberg, 1991). Alternatively, populations of *P. elegans* at the center of the species' range could be relicts of a larger and continuous ancestral distribution. Under this vicariance scenario, the disjunct distribution is the result of extinction of populations in tropical regions and persistence and/or expansion of peripheral populations at the edge of the distribution.

Although *P. elegans* has an extended planktonic larval period (25-35 days) consisting of six naupliar stages and a cypris (Barnes, 1996; S. Crickenberger, personal communication, July 2013), mtDNA sequences (cytochrome oxidase I) gathered from extra-tropical populations (Mexico and Peru) showed significant sequence divergence (1.2%) (Van Syoc, 1994), suggesting limited genetic exchange across the TEP. Van Syoc (1994) proposed that temperature was the primary factor limiting the colonization and persistence of populations in the warmest water of the TEP (and therefore gene flow across the TEP). Partially consistent with this idea, larval thermal tolerance of extra-tropical populations (Mexico and Peru) of *P. elegans* is lower than for larvae from tropical populations (El Salvador), providing some evidence of a potential physiological barrier to gene flow into the tropics, but not for larvae migrating out of the tropics

(Walther *et al.*, 2013). In addition to temperature, habitat availability may also be an important factor influencing the distribution of *P. elegans*, given that the species only lives on rocky intertidal substrate with moderate to extreme wave action (Newman & Foster, 1987; Laguna, 1990). The lack of such habitat over large stretches of the TEP coastline is well known (Springer, 1959; Walker, 1960; Rosenblatt & Walker, 1962; Stephens Jr, 1963; Rosenblatt, 1967), such as across the Sinaloan Gap (SG), the Central American gap (CAG), and the mainland Mexican coast of Baja California GAP (BCG) (**Fig. 2.1**), areas dominated by sandy beaches and estuaries interrupted by only short stretches of rocky coastline (Hastings, 2000; Robertson & Cramer, 2009).

As a first step to characterize the origins of the disjunct distribution of *P. elegans*, I have gathered mtDNA sequence data to describe the genetic structure and historical demography of this species as a way to infer its recent range-wide geographic history. Using a combination of phylogenetic analyses, coalescent population genetic methods, and Approximate Bayesian Computation (ABC), I compared alternative biogeographical histories that varied in population separation order, isolation time, and ancestral N_e, based on expectations of either trans-tropical dispersal or tropical extinction models. If the disjunct distribution is a consequence of recent trans-tropical, stepping-stone dispersal, I expect an overall genetic signature of declining genetic diversity (e.g. Marko, 2004) from one hemisphere to the other. Alternatively, if the fragmented population is the product of a relatively recent tropical extinction, genetic diversity should either be uniform across all regions or greatest in the tropics.

MATERIAL AND METHODS

Geographical sampling, DNA isolation and gene sequencing

Adult individuals of *Pollicipes elegans* were collected between 2009 and 2011 from sites chosen to capture variation in populations at the north, south, and centre of the species' range (**Fig. 2.1**). A total of 178 adult individuals were collected across three different sites in Mexico and El Salvador, and from two places in Peru (**Fig. 2.1**, **Table 2.1**). Tissue samples from the peduncle (muscle) of live barnacles were preserved in 70% ethanol. Genomic DNA was extracted by overnight proteinase K incubation in 2X cetyl trimethyl ammonium bromide (CTAB) followed by two chloroform extractions and recovered with ethanol precipitation. LCO1490 and HCO2198 universal primers (Folmer *et al.*, 1994) were used to amplify a portion of the mitochondrial cytochrome c oxidase subunit I gene (COI). Polymerase chain reaction (PCR) was carried out in 20 µL volumes containing 1x PCR buffer, 3 mM $MgCl_2$, 0.4 mM each dNTPs, 0.5 U of GoTaq DNA polymerase (Promega), 0.3 µM each primer, and approximately 50 ng of genomic DNA. Amplification conditions consisted of 94ºC for 2 min followed by 5 cycles of 94ºC for 30 s, 42ºC for 1:30 min, and 72ºC for 1 min. These first 5 cycles were followed by 35 cycles of 94ºC for 30 sec, 49ºC for 1 min, 72ºC for 1 min, with a final incubation step at 72ºC for 5 min. PCR products were purified using QIAquick spin filter columns (QIAgen) and sequenced in both directions on an ABI-3730 sequencer (Applied Biosystems). The sequences were inspected for basecalling errors with the forward and reverse strands aligned using SEQUENCHER 4.7 (Gene Codes Corporation). Sequences were edited using BIOEDIT (Hall, 1999) and easily aligned with CLUSTALW (Thompson *et al.*, 2002).

Sequence diversity statistics and demographic history reconstruction

Frequency based statistics were calculated to evaluate patterns of genetic diversity and demographic history across the distribution range of *P. elegans*. Nucleotide diversity (π), haplotype diversity (H), pairwise differences (k), segregating sites (S), Tajima's D (Tajima, 1989), and Fu's F_S (Fu, 1997) were calculated with ARLEQUIN version 3.5.1.2 (Excoffier & Lischer, 2010) with 10,000 permutations to test significance. R_2 (Ramos-Onsins & Rozas, 2002) was calculated with DNASP version 5 (Librado & Rozas, 2009) with 10,000 coalescent simulations. To assess the demographic history at each sampling site, I calculated nucleotide mismatch distribution plots, the sum of squared deviations (SSD), Harpending's r, and the parameter tau (τ) in ARLEQUIN to evaluate the fit of a sudden demographic expansion. The historical demography at each site was reconstructed by using the GMRF skyride plot (Minin *et al.*, 2008) implemented in BEAST version 1.7.2 (Drummond & Rambaut, 2007). Each GMRF skyride analysis was conducted twice for 10,000,000 generations each, sampling every 1,000 generations. The output files were checked in TRACER version 1.5 to ensure that all effective sampling sizes (ESS) values were greater than 200. I considered population size changes as significant when the upper and lower 95% confidence intervals at the root and the tips did not overlap.

Phylogenetic analyses and population structure

Phylogenetic relationships were inferred using Bayesian Inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP). For the model-based methods, the best substitution model was identified with MRMODELTEST version 2.3 (Nylander, 2004) using the Akaike Information Criterion (AIC). BI analyses were performed using MRBAYES

version 3.1.2 (Ronquist & Huelsenbeck, 2003), with five independent runs of four Markov chains for 10 million generations and default heating values, sampling every 100 generations with 2500 samples discarded as burn-in. ML analyses were conducted using RAxML version 7.04 (Stamatakis *et al.*, 2008). Nodal support was estimated from 1,000 fast bootstrap replicates in five independent runs. MP was implemented using PAUP* 4.0b10a (Swofford, 2003), with heuristic searches, tree bisection reconnection, and with 1,000 random addition replicates with all characters treated as unordered and equally weighted. Node reliabilities in the MP tree were assessed with 100 bootstrap replicates and averaged over five independent runs. Additionally, an unrooted gene network using the median-joining algorithm was built using SplitsTree 4.12.6 (Huson & Bryant, 2006). Partial COI sequences from *P. pollicipes* (GenBank accession #HM563669) and *P. caboverdensis* (GenBank accession #HM563667) were retrieved from GenBank and used as outgroups.

Genetic structure was characterized with Φ_{ST}, a measure of the proportion of nucleotide diversity within sub-populations relative to the total which takes into account the molecular distance between alleles. I used the substitution model in Arlequin most similar to the model identified with MrModeltest. Significance was assessed with 10,000 permutations of the data. A Spatial Analysis of Molecular Variance (SAMOVA) was then used to define groups of populations by maximizing the proportion of total genetic variance due to differences between groups of populations in SAMOVA version 1.0 (Dupanloup *et al.*, 2002). Additionally, partial Mantel tests (Smouse *et al.*, 1986) were performed using the Isolation by Distance Web Service (Jensen *et al.*, 2005) to

assess the significance of correlations between genetic and geographical distances among populations, while controlling for the effects of the zones delineated by SAMOVA.

Demographic parameter estimates

I estimated demographic parameters by sampling the coalescent in a Bayesian framework with the program IMA2 (Hey, 2010), which is based on an isolation-with-migration (IM) model (Nielsen & Wakeley, 2001; Hey & Nielsen, 2007; Hey, 2010) that uses Metropolis-coupled Markov-chain techniques to estimate the posterior densities of the time of divergence (t), theta (θ) and migration (m). I used the mean of trans-isthmian estimates of COI divergence rates (2%) from the most recently separated species among crustacean lineages (Knowlton & Weigt, 1998; Schubart *et al.*, 1998) and a generation time of one year, previously described for the congeneric species *P. pollicipes* (Cruz & Araujo, 1999; Cruz *et al.*, 2010), to obtain parameter estimates in demographic units (i.e. time in years, effective population size and number of migrants per generation as number of individuals). Initial 8 hour IMA2 runs (with a burn-in of 8 hours) of forty independent heated chains (Geyer, 1991) were performed to assess if the priors were suitable and if the heating conditions were appropriate. I then conducted two longer independent runs (100,005 saved genealogies) for each analysis. Stationarity was assessed by comparisons between parameter estimates generated from genealogies from the first and the second half of the run, and by visual inspection of the splitting time trend plots (Hey, 2010).

Likelihood ratio tests (LRTs) were computed to test the significance of migration (m=0) between each pair of populations (Nielsen & Wakeley, 2001; Hey, 2010). Since the LRT statistic is not a good fit to the theoretical expectation until higher values of LRT are achieved, I decided to use the same approach as Diaz-Perez *et al.* (2012), by considering LRT values less than 1 as a failure to reject the hypothesis of m=0 ; values greater than 1 but lower than 2.74 viewed as inconclusive (Hey, 2010); values greater than 2.74 were considered significant. A Bonferroni correction was used to reduce the chances of obtaining false positives (type I errors) with multiple pairwise comparisons.

Approximate Bayesian computation

To comprehensively evaluate the likelihood of different demographic scenarios, I used an Approximate Bayesian Computation (ABC) approach (Beaumont *et al.*, 2002). BAYESIAN SERIAL SIMCOAL (BSSC) (Anderson *et al.*, 2005) was used to determine if the observed nucleotide and haplotype diversity, number of segregating sites, genetic differentiation F_{ST} and Tajima's D could be explained as a result of one of four specific demographic models (**Fig. 2.2**). Given that there was no significant genetic differentiation within subtropical and tropical sites within Mexico, or within any other regions I pooled samples to reduce the number of parameters to be analyzed. I evaluated four basic biogeographic models of historical divergence: 1) Tropical Origin, Early Mexican Split (TOEMS) (**Fig. 2.2a**), a scenario that proposes Mexican populations split from an ancestral central (tropical) population in El Salvador, followed by a split between El Salvador and Peru; 2) Tropical Origin, Early Peruvian Split (TOEPS) (**Fig. 2.2b**), a history in which Peru diverged from a tropical Salvadoran ancestral population before

Mexican populations diverged from El Salvador; 3) Northern Origin Stepping-Stone (NOSS), and 4) Southern Origin Stepping-Stone (SOSS), models consistent with migration during periods of cooler climate (**Fig. 2.2c-d**). Each scenario was tested by modeling 1,000,000 coalescent simulations. In all models, I assumed that the split (disruption of connectivity) between Mexican and Salvadoran, and Mexican and Peruvian populations could be produced either by dispersal toward un-colonized regions in time t_0 and t_1, or by extinction of intermediate populations (**Fig. 2.2**). However, since models 3 and 4 assume the older populations are those present at one end of the modern distribution of *P. elegans*, high posterior probability support for models 3 or 4 suggests that trans-tropical dispersal is the most likely scenario that could explain divergence in *P. elegans*. The sizes of ancestral populations at the moment of splitting were represented by parameters Anc_0 and Anc_1. This parameter is obtained by modeling the ancestral population size at t_0 and t_1 to be any number between one and twenty times the size of the modern N_e. Large Anc suggests that environmental conditions were favorable at the time of the split with large effective population size (N_e) of the common ancestor populations, whereas small Anc values suggests adverse conditions at the time of split.

I conducted model fitting in R using ABC package (Csillery *et al.*, 2012) with feed-forward neural networks (nonlinear regression). Neural networks reduce a large number of summary statistics into a smaller number of dimensions (Blum & Francois, 2010), a more robust and accurate approach when the number of summary statistics is large (Csillery *et al.*, 2010). Cross validation was performed with four tolerance levels (δ = 0.001, 0.002, 0.005, and 0.01) and 100 steps. Competing models were compared based

on their posterior probability and Bayes Factor scores. To ensure that the model selection procedure was reliable, model selection and cross-validation steps were performed twice, first by comparing all four models against each other simultaneously and then by pairwise comparisons of all models. Parameter estimates for splitting times (parameters t_0 and t_1) and ancestral N_e (parameters Anc_0 and Anc_1) were estimated for the best-fitting model and validated using four tolerance levels.

RESULTS

Sequence diversity statistics

After alignment, visual inspection and trimming, a single 590 bp partial COI sequence was obtained for all individuals and used in this study. Nucleotide diversity was greatest in the central (tropical) portion of the species range, where populations of *P. elegans* are currently most scarce; diversity peaked in El Salvador, with lower values in Mexico and in Peru (**Table 2.1**). Haplotype diversity was lowest in Peru but similarly high in Mexico and El Salvador. Sequences are available in GenBank under accession numbers KF958514-KF958701.

Given high haplotype diversity but very low nucleotide diversity, each of Tajima's D, Fu's F_S and R_2 were significant for all Mexican populations, providing evidence of either a recent demographic expansion or a selective sweep in this region (**Table 2.1**). Fu's F_S was significant in El Salvador and Peru, but Tajima's D and R_2 were not. The mismatch distribution reflected the summary statistics, in that Mexican populations had one large peak relatively close to zero (mean τ=2.67) created by many sequences with few pairwise differences, plus a smaller peak due to the presence of a small number of

more divergent haplotypes (**Fig. 2.3a-c**). On average, τ was larger for Peru (τ=5.50, Fig. **2.3g-h**) and El Salvador (τ=6.09, **Fig. 2.3d-f**) and the mismatch distributions for populations in both of these regions was more ragged. Despite these differences, the sum of squared deviations and Harpending's raggedness statistic were not significant for any population, indicating that a sudden demographic expansion could not be rejected for any sample (**Fig. 2.3a-h**). Bayesian skyride plots showed little evidence of changes in N_e across time (**Fig. 2.3i-p**). However, the time to the most recent common mtDNA ancestor increased consistently towards the centre of the range of *P. elegans* (**Fig. 2.3i-p**).

Phylogenetic analyses and population structure

The unrooted median joining haplotype network for COI showed a strong association of haplotypes with geography, with three main haplogroups (**Fig. 2.4**). Haplogroup 1 consisted primarily of individuals collected in Mexico plus nine individuals from El Salvador. In contrast, haplogroups 2 and 3 primarily consisted of individuals from El Salvador and Peru, with only four individuals from Mexico (**Fig. 2.4**). The rooted phylogenetic analyses using MP, ML and BI revealed a consistent pattern in which haplogroup 1 (primarily Mexican individuals) formed a monophyletic clade with high node support from bootstrap percentages and posterior probabilities (see **Appendix A Fig. A-1**).

The general time reversible model (Tavaré, 1986) with a gamma rate distribution and invariant sites parameter (GTR+I+G) was the best-fit model for the dataset. I used (Tamura & Nei, 1993) substitution model, the most parameterized model

implemented in Arlequin, with a gamma correction to compare genetic diversity levels within sampling locations. The geographical patterns in the haplotype trees corresponded closely to very strong and significant genetic differentiation (**Fig. 2.5**) between Mexico and each of Peru (Φ_{ST} from 0.566 to 0.655) and El Salvador (Φ_{ST} from 0.358 to 0.507). In contrast, all pairwise Φ_{ST} estimates between El Salvador and Peru were low and not significant, with the exception of the SAL2-PER2 sampling site pair (**Fig. 2.5**); estimates of Φ_{ST} between Peru and El Salvador were similar to values between populations within each region (**Fig. 2.5**). The SAMOVA analysis reflected these patterns maximizing the variance among groups relative to the total variance (F_{CT}), when Mexico populations were considered as a single group and all El Salvador and Peru populations as a second group (**Table 2.2**).

Despite the sharp phylogeographical break between Mexico and all populations to the south, a partial Mantel test showed highly significant isolation by distance across all populations (r=0.8320, P=0.0001), even after controlling for the effects of the large genetic break between Mexico and all other southern samples (**Table 2.3**).

Coalescent-based demographic parameter estimates

Likelihood ratio tests from the IM analysis showed evidence for migration between *P. elegans* populations in only three of 28 pairwise comparisons. These three comparisons were not significant after a Bonferroni correction for multiple tests (see **Appendix A Table A-1**). In all 28 comparisons, the joint posterior density for m increased asymptotically as m approached zero, indicating either no gene flow across populations or a lack of migration signal in the data (see **Appendix A Fig A-2**).

Divergence times varied among pairwise comparisons, with more old divergences among populations within Mexico than compared to populations within El Salvador and Peru (see **Appendix A Fig. A-3**). The posterior density distribution for the divergence times between Mexico and Peru (~150 to 300 ka) overlapped with the divergence times between Mexico and El Salvador (~150 to 250 ka). However, the divergence time between El Salvador and Peru showed consistently more recent and narrower joint posterior density distributions (~50 to 150 ka) (see **Appendix A Fig. A-4**).

Only a few posteriors for N_e showed a well-defined peak with tails that returned to zero (Appendix S3b). However, estimates of N_e towards the centre of the species' range were potentially very large, especially when compared to Peruvian populations. Considering only the posterior density plots with clear peaks, El Salvador showed a slightly higher mean historical N_e (2×10^6 individuals) than either Mexico (1.5×10^6 individuals) or Peru (7.5×10^5 individuals) (see **Appendix A Fig. A-5**).

Approximate Bayesian computation phylogeography

Simultaneous evaluation of all four competing models indicated that Tropical Origin, Early Mexican Split (TOEMS, **Fig. 2.2a**) was the best-fitting model for all tolerance levels. When comparing all models simultaneously, cross validation averaged 83% (**Table 2.4**). However, in the pairwise model comparisons, cross validation support improved substantially (mean=93%), and as in the simultaneous evaluation, TOEMS was consistently selected as the best fit model, with the highest posterior probability (PP) value when comparing all models simultaneously (average PP=0.86), or in a pairwise fashion (average PP=0.89) (pairwise results not shown).

Estimates of ancestral effective population size (Anc_0 and Anc_1) and time of divergence (t_0 and t_1) for the best-fitting TOEMS model showed contrasting patterns between the northern and southern portions of the species' range. In the north, Mexico and El Salvador separated 260 ka, with ancestral Salvadoran populations about half the size (50%) of the averaged historical N_e over the coalescent of Salvadoran populations (**Table 2.5**, δ=0.001). In the south, Peru separated from El Salvador only 55 ka, but with an ancestral Salvadoran population that was much smaller, only 3% of the averaged historical N_e over the coalescent of Salvadoran populations (**Table 2.5**). Confidence intervals of the Anc parameters suggest that there was a significant reduction in N_e moving forward in time between t_1 and t_0 (**Table 2.5**). Diagnostic plots of parameters Anc and t suggested little influence of the tolerance level and priors in the posterior density plots (**Table 2.5**). In addition, cross validation of the estimated parameters showed a minimum effect of the tolerance level in the error rate.

DISCUSSION

Although taxa with disjunct distributions have been the subject of intensive research (Barber & Bellwood, 2005; Floeter *et al.*, 2008; Lessios, 2008; Cowman & Bellwood, 2013), rigorous comparisons of alternative hypotheses of historical divergence based on modeling demographic history have been rare. Using a combination of coalescent population genetic analyses and ABC modeling, this study found very little support for a recent history of trans-tropical dispersal. Instead, the results support the idea that the extra-tropical populations of *P. elegans* likely developed through a process of extinction of tropical populations, best described by the

Tropical Origin, Early Mexican Split, or TOEMS model. Further, the data and analyses indicate that for *P. elegans*, the separation of peripheral populations from central populations in the warmest tropical water was achieved in two temporally distinct steps: the isolation of Mexican populations followed by a more recent isolation of Peruvian populations.

From every perspective I considered, the genetic data are most consistent with the tropical extinction hypothesis and lack any evidence of recent trans-tropical dispersal. First, a peak in nucleotide diversity in El Salvador (tropical region) is not consistent with recent dispersal across the tropics (**Table 2.1**); in a trans-tropical dispersal scenario, sequential colonization events should leave behind a signature of declining genetic diversity from north to south or from south to north, similar to what has been observed in many other taxa that have recently expanded their geographical range in one direction over time (e.g. Hickerson & Cunningham, 2005; Marko *et al.*, 2010). Although haplotype diversity did not peak in the tropics, high haplotype diversity in Mexico appears to be the result of a relatively recent demographic expansion (**Fig. 2.3a-c**, **Table 2.1**), and not the result of recent separation from tropics (**Table 2.5**, **Fig. S3a**). Viewed from the perspective of the coalescent, Bayesian skyride plots showed that haplotypes in tropical populations had a consistently deeper history compared to haplotypes in all other populations to the north or south (**Fig. 2.3l-n**), indicating either a relatively larger historical N_e or older population ages in the tropical region. Second, the pattern of significant isolation by distance (IBD) observed in *P. elegans* is not expected in a linear dispersal-driven colonization model (e.g. Sanford *et al.*, 2003; Marko, 2004).

Significant IBD indicates that populations are now, or recently have been, in a gene flow/genetic drift equilibrium (Hutchison & Templeton, 1999; Wares, 2002). Given that I could find no evidence of recent gene flow, significant IBD suggests that substantial connectivity existed across all populations of *P. elegans* in the past, a pattern consistent with the idea that tropical populations were more abundant in the past. Finally, given evidence of more diverse and possibly older and/or larger Salvadoran populations, I used ABC model testing to comprehensively evaluate four likely biogeographical hypotheses potentially explaining the origins of near antitropicality in *P. elegans*. These analyses found that a tropical origin model was a better fit for the data than a trans-tropical dispersal model. Under the best-fitting tropical origin scenario, *P. elegans* was abundant and widely distributed throughout the tropics, but through the loss of most populations in the tropical region (perhaps coupled with colonization at higher latitudes), the species' range was sequentially divided into three isolated populations.

Modern and historical factors shaping Pollicipes elegans distribution

Even though I cannot determine how specific environmental and biotic factors shaped patterns of genetic differentiation in *P. elegans* in the eastern Pacific, there are several historical and contemporary factors that may explain the formation of the disjunct distribution and the persistence of apparently tropical relict populations in the warmest region of the TEP. First, the current range disjunction and restricted connectivity between tropical and extra-tropical populations could be maintained by a lack of suitable habitat for *P. elegans*. Salvadoran populations at the center of the range of *P. elegans* lie in the Central American Gap or CAG (Springer 1959, Hastings 2000), a

transition zone between the Mexican and Panamanian Zoogeographic Provinces (Briggs, 1974; Laguna, 1990) that is dominated by sandy or muddy coastline interspersed with mangrove-lined lagoons (Springer, 1959; Walker, 1960; Rosenblatt & Walker, 1962; Stephens Jr, 1963; Rosenblatt, 1967). The absence of suitable habitats for *P. elegans* across large stretches of this region may substantially limit the production, spread, and settlement of larvae, especially along the northern CAG coastline between existing Salvadoran and Mexican populations (**Fig. 3.1**). However, considerable rocky shore habitat can be found south of the CAG, and yet, to the best of my knowledge, *P. elegans* is absent from over 2,000 km of shore from El Salvador to northern Ecuador (**Fig. 3.1**), suggesting that other factors may prevent the colonization and persistence of *P. elegans*, at least to south of El Salvador. For example, patterns of ocean circulation and thermal tolerance may provide a mechanism of geographic isolation for *P. elegans*. Ocean currents, thermocline depth and nutrient supply are all governed by wind stress in the TEP (for a review see Pennington *et al.*, 2006), with winds driving the north Pacific subtropical gyre clockwise and the south Pacific gyre counterclockwise, causing a convergence of surface waters in the TEP (**Fig. 3.1**). This circulation pattern, that pushes water outside the coast in the extra-tropical regions and towards the coast in the tropics, could prevent the establishment of *P. elegans* at higher latitudes and, although currents will carry larvae from extra-tropical regions into the warm tropical pool, the region is unsuitable for cold adapted larvae (Walther *et al.*, 2013, **Fig. 3.1**). The temperature tolerance of larvae of *P. elegans* has been proposed and tested as a barrier to transtropical dispersal, preventing the spread of larvae from extra-tropical regions

into the warmest tropical water of the TEP (Walther *et al.*, 2013). Larvae of *P. elegans* collected at extra-tropical sites (near MEX1 and PER1 from this study, respectively) are unable to survive at temperatures commonly found in the region of the warmest tropical pool (Fig. 2.1, Walther *et al.*, 2013). Even though I found no evidence of gene flow between extra-tropical sites (i.e. Mexico and Peru), gene-flow was also absent among sites with similar sea surface temperature, suggesting that either the single mtDNA locus that I used provides too little information about gene flow in the past, or other factors such as biological interactions, and food availability may be involved. Nevertheless, persistence and isolation of *P. elegans* in the tropics may be consequence of local recruitment. Crustacean larvae are able to regulate depth and may show retention or return to coastal habitats after initial transport offshore.

Although there are modern environmental factors (e.g. sea surface temperature and habitat availability) that may currently regulate the distribution of *P. elegans*, historical environmental changes may have caused the initial range disruption. Large fluctuations in sea level (Cronin & Forward, 1986; Phillips & Mcwilliam, 1986), patterns of ocean circulation, and sea surface temperature have been invoked as important factors contributing to diversification of marine species in the TEP. Acknowledging the uncertainty of the molecular clock (Paulay, 1990; but see Williams & Reid, 2004) in the parameter estimates, the ABC-based estimates of t0 and t1 (53 kyr and 292 kyr, respectively) indicate that isolation of both extra-tropical populations occurred during cold glacial periods, suggesting that Mexican and Peruvian populations became peripheral isolates when the main range of the species likely contracted towards the

tropics. I can only speculate, but strong advection of cold water from the Peruvian Current system during glacial cycles may have generated a latitudinal shift of the Intertropical Convergence Zone northward, disrupting gene flow in the equatorial region by sudden contraction of populations north of the equator (Arbogast *et al.*, 2002; Pulquerio & Nichols, 2007). The intensification of upwelling in the eastern equatorial ocean during the mid-Pleistocene climate transition, followed by an increase in temperature in the equatorial Pacific Ocean, might also have strongly affected connectivity across the TEP (Rincon-Martinez *et al.*, 2010).

The effects of decadal variation in temperature on connectivity and resilience of modern marine populations may also have left an imprint on the population structure of *P. elegans*. El Niño Southern Oscillation and acute seasonal variation profoundly impacted population and community dynamics through repeated regional die-offs and peripheral colonization events, especially in South America (Marlow *et al.*, 2000; Liu & Herbert, 2004; de Garidel-Thoron *et al.*, 2005). For example, during the El Niño event of 1997-1998, high sea surface temperature and run-off of terrestrial guano during intense rainfall seasons caused mass mortalities of *P. elegans* in Lobos de Afuera island, a location 200 km away from the PER3 site on this study (Richmond, 1990; Defeo *et al.*, 2013); during the 1982-1983 El Niño, *P. elegans* colonized rocky shores of central Peru, establishing a local short-lived local fishery (Carbajal *et al.*, 1998). Data analyses shows that Peruvian historical N_e was considerably smaller than either Salvadoran or Mexican populations (see **Appendix A Fig. A-4**), which may reflect repeated bottleneck events

caused by extinction and colonization events in Peru during strong ENSO years, rather than consistently small N_e or a single strong founder event in Peru.

Evolutionary significant units within Pollicipes elegans distribution range

Speciation in marine taxa like *P. elegans* with high dispersal capabilities remains incompletely understood (Kameya & Zeballos, 1988; Arntz *et al.*, 2006). The very strong phylogeographic break (i.e., near reciprocal monophyly of mtDNA) between Mexican and Peruvian haplogroups suggest that Mexican *P. elegans* populations are on an independent evolutionary trajectory with respect to all populations to the south. The strong phylogeographical differentiation (which the isolation-with-migration model attributes to a relatively ancient isolation time) that has persisted across the last two glacial-interglacial cycles (150-300 ka) indicates at the very least, the existence of at least two evolutionarily significant units. Given incomplete lineage sorting of mtDNA haplotypes, Peruvian and Salvadoran populations may provide an earlier snapshot of climate-driven differentiation between tropical and extra-tropical populations. Nonetheless, my findings from a single mtDNA locus must to be tested with analyses that include more unlinked nuclear genes to overcome the stochasticity of the coalescent.

In the absence of gene flow between regions with different environmental conditions, local adaptation through niche divergence is likely. However, it is difficult to conclusively determine if the larval phenotypic changes (Walther *et al.,* 2013) are genetically based or the result of phenotypic plasticity or population acclimatization. For example, a study of the brooding activities of the gooseneck barnacle *Pollicipes*

polymerus in California indicated the presence of two distinct phenotypic "races", one that broods maximally in colder seawater temperatures (14 °C) and other that broods maximally at warmer temperatures (20 °C) (Cimberg, 1981). However, no significant genetic differences between the two races were found, leading the authors to conclude that the reproductive differences observed in *P. polymerus* may the result of phenotypic plasticity in brooding activity (Knowlton, 1993; Palumbi, 1994; Lessios *et al.*, 2001; Swanson & Vacquier, 2002). Similarly, although the thermal tolerance of larvae of *P. elegans* showed population-dependent physiological phenotypes (Miner, 2002), only Mexican and Salvadoran populations show a strong phylogenetic break that is consistent with the physiological study, unless the separation of Peruvian and Salvadoran populations has been too recent for consistent mtDNA differentiation to develop. Therefore, the evolutionary status of Peruvian populations also warrants future investigation. The use of additional genetic markers for species delimitation coupled with niche overlap analyses, should shed light on the relative contributions of genetic drift or selection to speciation in *P. elegans.*

Out of the tropics?

Population genetic patterns of mtDNA diversity combined with phylogeographical inference for *P. elegans* provide a novel example of diversification in the tropics likely mediated by glacial-interglacial climate change during the Pleistocene, a scenario in which extra-tropical species arise from a tropical ancestor. The overall pattern of differentiation among populations of *P. elegans* fits some but not all of the expectations of the out-of-the tropics model of diversification (OTT), a model that

explains the formation and maintenance of the latitudinal diversity gradient (LDG) (Walther *et al.*, 2013). In the OTT model, tropical diversity is the result of high speciation and low extinction rates in the tropics ("cradle" and "museum" effects, respectively) combined with latitudinal shifts of "bridge species" towards extra-tropical regions where speciation rates are lower and extinction rates are higher. This study showed that at least one highly divergent extra-tropical population (likely a bridge/cryptic species) originated through this process, crossing the tropical/extra-tropical boundaries of *Pollicipes* in the TEP. However, the OTT model depends fundamentally on relatively high speciation and low extinction rates in the tropics, and at present, the relative scarcity of tropical populations suggests that the future for *Pollicipes* in the warmest waters of the TEP is uncertain.

CONCLUSION

The formation of ephemeral dispersal barriers that arose during sea level fluctuations in the Pleistocene (Jablonski *et al.*, 2006; Jablonski *et al.*, 2013) combined with cycles of local extinction and recolonization, have been proposed as opportunities for rapid speciation of marine species in the tropics (Valentine & Jablonski, 1983; Mcmillan & Palumbi, 1995). Using frequency-based statistics, coalescent methods, and ABC modeling, I consistently support vicariance as the main mechanism that produced the largely antitropical distribution of *P. elegans* in the eastern Pacific. Although is not clear how much habitat suitability, ocean circulation patterns and physiological response to thermal stress can explain the current disjunct geographical distribution and isolation of *P. elegans*, future studies should address if the distinct evolutionary

significant units found in Mexico and Southern Mexico may correspond to a novel example of cryptic species in a latitudinal thermal gradient.

REFERENCES

Anderson, C.N.K., Ramakrishnan, U., Chan, Y.L. & Hadly, E.A. (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733-1734.

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

Arntz, W.E., Gallardo, V.A., GutiÃ©rrez, D., Isla, E., Levin, L.A., Mendo, J., Neira, C., Rowe, G.T., Tarazona, J. & Wolff, M. (2006) El Niño and similar perturbation effects on the benthos of the Humboldt, California, and Benguela Current upwelling ecosystems. *Advances in geosciences*, **6**, 243-265.

Barber, P.H. & Bellwood, D.R. (2005) Biodiversity hotspots: evolutionary origins of biodiversity in wrasses (Halichoeres : Labridae) in the Indo-Pacific and new world tropics. *Molecular Phylogenetics and Evolution*, **35**, 235-253.

Barnes, M. (1996) Pedunculate Cirripedes of the genus Pollicipes. *Oceanography and Marine Biology, Vol 34*, **34**, 303-394.

Beaumont, M.A., Zhang, W.Y. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025-2035.

Blum, M.G.B. & Francois, O. (2010) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, **20**, 63-73.

Bowen, B.W. & Grant, W.S. (1997) Phylogeography of the sardines (Sardinops spp): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution*, **51**, 1601-1610.

Briggs, J.C. (1974) *Marine zoogeography*. McGraw-Hill, New York.

Briggs, J.C. (1987) Antitropicality and vicariance. *Systematic Biology*, **36**, 206-207.

Burridge, C.P. (2002) Antitropicality of Pacific fishes: molecular insights. *Environmental Biology of Fishes*, **65**, 151-164.

Carbajal, W., Castañeda, J., Castro, J., De la Cruz, J., Ramírez, P., Quevedo, M. & Zeballos, C. (1998) Informe del Seguimiento de las Pesquerías Artesanales. In: *Informe Progressivo*, p. 28. Instituto del Mar de Perú (IMARPE), Chiclayo, Perú.

Carlquist, S. (1981) Chance Dispersal. *American Scientist*, **69**, 509-516.

Cowen, R.K. & Sponaugle, S. (2009) Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science*, **1**, 443-466.

Cowman, P.F. & Bellwood, D.R. (2013) Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. *Proceedings of the Royal Society B-Biological Sciences*, **280**

Crisp, M.D., Isagi, Y., Kato, Y., Cook, L.G. & Bowman, D.M.J.S. (2010) Livistona palms in Australia: Ancient relics or opportunistic immigrants? *Molecular Phylogenetics and Evolution*, **54**, 512-523.

Cronin, T.W. & Forward, R.B. (1986) Vertical Migration Cycles of Crab Larvae and Their Role in Larval Dispersal. *Bulletin of Marine Science*, **39**, 192-201.
Cruz, T. & Araujo, J. (1999) Reproductive patterns of Pollicipes pollicipes (Cirripedia : Scalpellomorpha) on the southwestern coast of Portugal. *Journal of Crustacean Biology*, **19**, 260-267.
Cruz, T., Castro, J.J. & Hawkins, S.J. (2010) Recruitment, growth and population size structure of Pollicipes pollicipes in SW Portugal. *Journal of experimental marine biology and ecology*, **392**, 200-209.
Csillery, K., Francois, O. & Blum, M.G.B. (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475-479.
Csillery, K., Blum, M.G.B., Gaggiotti, O.E. & Francois, O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410-418.
Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life: Murray*. London.
de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F. & Beaufort, L. (2005) Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature*, **433**, 294-298.
de Queiroz, A. (2005a) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution*, **20**, 68-73.
Defeo, O., Castrejón, M., Ortega, L., Kuhn, A.M., Gutiérrez, N.L. & Castilla, J.C. (2013) Impacts of Climate Variability on Latin American Small-scale Fisheries. *Ecology and Society*, **18**, 30.
Diaz-Perez, A.J., Sequeira, M., Santos-Guerra, A. & Catalan, P. (2012) Divergence and biogeography of the recently evolved Macaronesian red Festuca (Gramineae) species inferred from coalescence-based analyses. *Molecular ecology*, **21**, 1702-1726.
Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology*, **7**, 214.
Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular ecology*, **11**, 2571-2581.
Ekman, S. (1953) *Zoogeography of the sea*. Sidgwick and Jackson, London,.
Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.
Floeter, S., Rocha, L., Robertson, D., Joyeux, J., Smith-Vaniz, W., Wirtz, P., Edwards, A., Barreiros, J.P., Ferreira, C. & Gasparini, J.L. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22-47.
Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, **3**, 294-299.
Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.
Geyer, C.J. (1991) Markov-Chain Monte-Carlo Maximum-Likelihood. *Computing Science and Statistics*, 156-163.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* (ed by, pp. 95-98.
Hastings, P.A. (2000) Biogeography of the Tropical Eastern Pacific: distribution and phylogeny of chaenopsid fishes. *Zoological Journal of the Linnean Society*, **128**, 319-335.
Hellberg, M.E. (1998) Sympatric sea shells along the sea's shore: The geography of speciation in the marine gastropod Tegula. *Evolution*, **52**, 1311-1324.
Hey, J. (2010) Isolation with Migration Models for More Than Two Populations. *Molecular biology and evolution*, **27**, 905-920.
Hey, J. & Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785-2790.
Hickerson, M.J. & Cunningham, C.W. (2005) Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (Xiphister) revealed by multilocus DNA analysis. *Evolution*, **59**, 344-360.
Hunn, C.A. & Upchurch, P. (2001) The importance of time/space in diagnosing the causality of phylogenetic events: Towards a "chronobiogeographical" paradigm? *Systematic Biology*, **50**, 391-407.
Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, **23**, 254-267.
Hutchison, D.W. & Templeton, A.R. (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898-1914.
Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102-106.
Jablonski, D., Belanger, C.L., Berke, S.K., Huang, S., Krug, A.Z., Roy, K., Tomasovych, A. & Valentine, J.W. (2013) Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10487-10494.
Jensen, J.L., Bohonak, A.J. & Kelley, S.T. (2005) Isolation by distance, web service. *Bmc Genetics*, **6**, 13.
Kameya, A. & Zeballos, J. (1988) Distribucion y densidad de percebes Pollicipes elegans (Crustacea: Cirripedia) en el mediolitoral peruano (Yasila, Piura; Chilca, Lima). *Boletin-Instituto del Mar del Peru*, **12**, 5-22.
Keever, C.C., Sunday, J., Puritz, J.B., Addison, J.A., Toonen, R.J., Grosberg, R.K. & Hart, M.W. (2009) Discordant Distribution of Populations and Genetic Variation in a Sea Star with High Dispersal Potential. *Evolution*, **63**, 3214-3227.
Knowlton, N. (1993) Sibling Species in the Sea. *Annual Review of Ecology and Systematics*, **24**, 189-216.
Knowlton, N. & Weigt, L.A. (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 2257-2263.

Laguna, J.E. (1990) Shore Barnacles (Cirripedia, Thoracica) and a Revision of Their Provincialism and Transition Zones in the Tropical Eastern Pacific. *Bulletin of Marine Science*, **46**, 406-424.
Lessios, H.A. (2008) The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. *Annual Review of Ecology Evolution and Systematics*, **39**, 63-91.
Lessios, H.A., Kessing, B.D. & Pearse, J.S. (2001) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin Diadema. *Evolution*, **55**, 955-975.
Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.
Lindberg, D.R. (1991) Marine Biotic Interchange between the Northern and Southern Hemispheres. *Paleobiology*, **17**, 308-324.
Liu, Z.H. & Herbert, T.D. (2004) High-latitude influence on the eastern equatorial Pacific climate in the early Pleistocene epoch. *Nature*, **427**, 720-723.
Marko, P.B. (1998) Historical allopatry and the biogeography of speciation in the prosobranch snail genus Nucella. *Evolution*, **52**, 757-774.
Marko, P.B. (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular ecology*, **13**, 597-611.
Marko, P.B., Hoffman, J.M., Emme, S.A., McGovern, T.M., Keever, C.C. & Cox, L.N. (2010) The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular ecology*, **19**, 146-169.
Marlow, J.R., Lange, C.B., Wefer, G. & Rosell-Mele, A. (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science*, **290**, 2288-+.
Mcmillan, W.O. & Palumbi, S.R. (1995) Concordant Evolutionary Patterns among Indo-West Pacific Butterflyfishes. *Proceedings of the Royal Society B-Biological Sciences*, **260**, 229-236.
Miner, B.G. (2002) Are the two physiological races of Pollicipes polymerus (Cirripedia) genetically divided along the California coast? *Invertebrate Biology*, **121**, 158-162.
Minin, V.N., Bloomquist, E.W. & Suchard, M.A. (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular biology and evolution*, **25**, 1459-1471.
Newman, W.A. (1992b) Biotic Cognates of Eastern Boundary Conditions in the Pacific and Atlantic: Relicts of Tethys and Climatic Change. *Proceedings of the San Diego Society of Natural History*, **16**, 1-7.
Newman, W.A. & Foster, B.A. (1987) Southern-Hemisphere Endemism among the Barnacles - Explained in Part by Extinction of Northern Members of Amphitropical Taxa. *Bulletin of Marine Science*, **41**, 361-377.
Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885.
Nylander, J. (2004) MrModeltest v2. *Evolutionary Biology Centre, Uppsala University*,
Palumbi, S.R. (1994) Genetic-Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics*, **25**, 547-572.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, **37**, 637-669.
Paulay, G. (1990) Effects of Late Cenozoic Sea-Level Fluctuations on the Bivalve Faunas of Tropical Oceanic Islands. *Paleobiology*, **16**, 415-434.
Pennington, J.T., Mahoney, K.L., Kuwahara, V.S., Kolber, D.D., Calienes, R. & Chavez, F.P. (2006) Primary production in the eastern tropical Pacific: A review. *Progress in Oceanography*, **69**, 285-317.
Phillips, B.F. & Mcwilliam, P.S. (1986) The Pelagic Phase of Spiny Lobster Development. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2153-2163.
Pulquerio, M.J.F. & Nichols, R.A. (2007) Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution*, **22**, 180-184.
Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, **19**, 2092-2100.
Richmond, R. (1990) The effects of the El Niño/Southern Oscillation on the dispersal of corals and other marine organisms. *Global ecological consequences of the 1982-83 el niño-southern oscillation* (ed. by P.W. Glynn), pp. 127-140. Elsevier, Amsterdam, The Netherlands.
Rincon-Martinez, D., Lamy, F., Contreras, S., Leduc, G., Bard, E., Saukel, C., Blanz, T., Mackensen, A. & Tiedemann, R. (2010) More humid interglacials in Ecuador during the past 500 kyr linked to latitudinal shifts of the equatorial front and the Intertropical Convergence Zone in the eastern tropical Pacific. *Paleoceanography*, **25**
Robertson, D.R. & Cramer, K.L. (2009) Shore fishes and biogeographic subdivisions of the Tropical Eastern Pacific. *Marine Ecology Progress Series*, **380**, 1-17.
Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.
Rosenblatt, R. (1967) The zoogeographic relationships of the marine shore fishes of tropical America. **5**, 579-592.
Rosenblatt, R.H. & Walker, B.W. (1962) The marine shore-fishes of the Galapagos Islands. *California Academy of Sciences, Occasional Papers*, **44**, 97-106.
Sanford, E., Roth, M.S., Johns, G.C., Wares, J.P. & Somero, G.N. (2003) Local selection and latitudinal variation in a marine predator-prey interaction. *Science*, **300**, 1135-1137.
Schubart, C.D., Diesel, R. & Hedges, S.B. (1998) Rapid evolution to terrestrial life in Jamaican crabs. *Nature*, **393**, 363-365.
Smouse, P.E., Long, J.C. & Sokal, R.R. (1986) Multiple-Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic zoology*, **35**, 627-632.
Springer, V.G. (1959) Systematics and zoogeography of the ciclidae fishes of the subtribe Labrisomini Hubbs. *Institute of Marine Science Publications*, **5**, 417-492.
Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, **57**, 758-771.
Stephens Jr, J.S. (1963) A revised classification of the blennioid fishes of the American family Chaenopsidae. **68**, 1-165.

Swanson, W.J. & Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, **3**, 137-144.
Swofford, D.L. (2003) PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, **10**, 512-526.
Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, **17**, 57-86.
Thompson, J.D., Gibson, T. & Higgins, D.G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2.3. 1-2.3. 22.
Valentine, J.W. & Jablonski, D. (1983) Larval Adaptations and Patterns of Brachiopod Diversity in Space and Time. *Evolution*, **37**, 1052-1061.
Van Syoc, R.I. (1994) Genetic divergence between subpopulations of the eastern Pacific goose barnacle Pollicipes elegans: mitochondrial cytochrome c subunit 1 nucleotide sequences. *Molecular marine biology and biotechnology*, **3**, 338-346.
Walker, B.W. (1960) The distribution and affinities of the marine fish fauna of the Gulf of California. *Systematic Zoology*, **9**, 123-133.
Walther, K., Crickenberger, S.E., Marchant, S., Marko, P.B. & Moran, A.L. (2013) Thermal tolerance of larvae of Pollicipes elegans, a marine species with an antitropical distribution. *Marine Biology*, **160**, 2723-2732.
Wares, J.P. (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular ecology*, **11**, 1131-1144.
Wiley, E.O. (1988) Vicariance Biogeography. *Annual Review of Ecology and Systematics*, **19**, 513-542.
Williams, S.T. & Reid, D.G. (2004) Speciation and diversity on tropical rocky shores: A global phylogeny of snails of the genus Echinolittorina. *Evolution*, **58**, 2227-2251.

Table 2.1: Molecular diversity and neutrality tests at the cytochrome c oxidase mitochondrial gene in populations of *Pollicipes elegans.*

Population	Location	*n*	Hap	π ± SD	H ± SD	*k* ± SD	S	R_2	Tajima's *D*	Fu's F_S
MEX1	Gaspareño, Mexico	25	21	0.006 ± 0.004	0.983 ± 0.017	3.713 ± 1.958	27	**0.0754**	**-1.784**	-**18.585**
MEX2	Migriño, Mexico	21	15	0.005 ± 0.003	0.938 ± 0.040	2.924 ± 1.611	25	**0.0753**	**-2.217**	**-10.107**
MEX3	Melaque, Mexico	24	21	0.008 ± 0.004	0.978 ± 0.024	4.486 ± 2.311	37	**0.0396**	**-2.091**	**-17.272**
SAL1	Mizata, El Salvador	25	19	0.010 ± 0.006	0.970 ± 0.022	5.880 ± 2.940	30	0.0827	-0.972	**-9.086**
SAL2	Taquillo, El Salvador	15	13	0.011 ± 0.006	0.981 ± 0.031	6.400 ± 3.249	24	0.1062	-0.552	**-5.275**
SAL3	Jucuarán, El Salvador	31	19	0.008 ± 0.005	0.942 ± 0.025	4.976 ± 2.513	29	0.0733	-1.127	**-7.804**
PER1	El Arco, Peru	20	14	0.007 ± 0.004	0.889 ± 0.068	4.237 ± 2.217	22	0.0811	-1.214	**-6.080**
PER2	Islilla, Peru	27	13	0.006 ± 0.004	0.889 ± 0.041	3.527 ± 1.870	16	0.1012	-0.522	**-3.840**

n, number of diploid individuals; Hap, number of haplotypes, π, nucleotide diversity; SD, standard deviation; H, haplotype diversity; k, number of haplotypes; S, number of segregation sites; R_2, Ramos-Onsins and Rozas index; Tajima's D; Fu's F_S statistic. Significant departures from the null model of neutrality or population growth are shown in boldface.

Table 2.2: Fixation indices and percentage of variation explained by each source for groups of populations identified by SAMOVA within each clade (all results are significant at $P < 0.05$ and shown in boldface).F_{SC}, variance among subpopulations within groups; F_{ST}, variance among subpopulations relative to the total variance; F_{CT}, variance among groups relative to the total variance.

		Fixation indices			Percentage variation		
K	Group compositions[§]	F_{SC}	F_{ST}	F_{CT}	Among groups	Among populations	Within populations
2	(MEX1, MEX2, MEX3) - (SAL1, SAL2, SAL3, PER1, PER2)	0.010	**0.526**	**0.522**	52.190	0.460	47.350
3	(MEX1, MEX2) - (MEX3) - (SAL1, SAL2, SAL3, PER1, PER2)	0.009	**0.495**	**0.491**	49.070	0.470	50.450
4	(MEX1) - (MEX2) - (MEX3) - (SAL1, SAL2, SAL3, PER1, PER2)	**0.013**	**0.481**	**0.475**	47.450	0.690	51.860
5	(MEX1) - (MEX2) - (MEX3) - (SAL2) - (SAL1, SAL3, PER1, PER2)	**0.002**	**0.447**	**0.446**	44.580	0.120	55.300
6	(MEX1) - (MEX2) - (MEX3) - (SAL2) - (PER1) - (SAL1, PER1, PER2)	-0.001	**0.414**	**0.414**	41.450	-0.040	58.590
7	(MEX1, MEX2) - (MEX3) - (SAL1) - (SAL2) - (SAL3) - (PER1) - (PER2)	-0.007	**0.384**	0.388	38.790	-0.400	61.610

[§] Population names as in **Table 2.1**

Table 2.3: Partial Mantel test results for the correlation between geographic and genetic distances, controlling for putative barrier to gene flow inferred in SAMOVA and for geographical distance after 10,000 randomizations. Z, standard score; r, correlation coefficient.

Model	Z	r	*P*
Correlation of genetic and geographic distance	59398.9071	0.8320	0.0001
Correlation of genetic distance and SAMOVA barrier	17.6221	0.8927	< 0.0001
Partial correlation of genetic and geographic distance controlling for SAMOVA barrier		0.6934	0.0001
Partial correlation of genetic and SAMOVA barrier controlling for geographical distance		0.8108	0.0107

Table 2.4: Cross validation and multiple phylogeographic model selection results under feed forward neural network method. The model with the highest posterior probability (PP) is highlighted in grey.

Model	δ	Samples	Cross validation (%)	Posterior Probability
NOSS	0.001	4,000	82	0.063
	0.002	8,000	81	0.1159
	0.005	20,000	79	0.1167
	0.01	40,000	80	0.115
SOSS	0.001	4,000	93	0.0019
	0.002	8,000	97	0.018
	0.005	20,000	90	0.0429
	0.01	40,000	96	0.0553
TOEMS	0.001	4,000	74	0.9344
	0.002	8,000	77	0.8589
	0.005	20,000	70	0.8272
	0.01	40,000	78	0.8165
TOEPS	0.001	4,000	84	0.0007
	0.002	8,000	84	0.0072
	0.005	20,000	86	0.0133
	0.01	40,000	86	0.0131

δ, tolerance level

Table 2.5: Parameter estimates in the approximate Bayesian computation (ABC) analysis of Tropical Origin, Early Mexican Split (TOEMS) model. Parameter estimates correspond to the weighted mode of the posterior samples and in parenthesis the 95% highest posterior density interval.

δ	Samples	t_0	Anc_0	t_1	Anc_1
0.001	1000	55,390 (38,255 – 80,924)	0.029 (0.0185 – 0.0789)	259,501 (213,471 – 396,795)	0.575 (0.128 – 4.5464)
0.002	2000	63,218 (41,889 – 87,683)	0.029 (0.0129 – 0.1758)	253,161 (202,816 – 404,574)	0.596 (0.1543 – 6.7856)
0.005	5000	61,521 (36,828 – 85,487)	0.044 (0.0217 – 0.4215)	282,141 (197,992 – 398,472)	0.511 (0.1185 – 7.3452)
0.01	10000	53,066 (30,772 – 89,702)	0.073 (0.0183 – 3.4424)	291,902 (181,899 – 372,680)	1.558 (0.2049 – 19.4613)

δ, tolerance level; t_0, Anc_0, t_1, Anc_1 as described on the methods and in **Fig. 2.2**

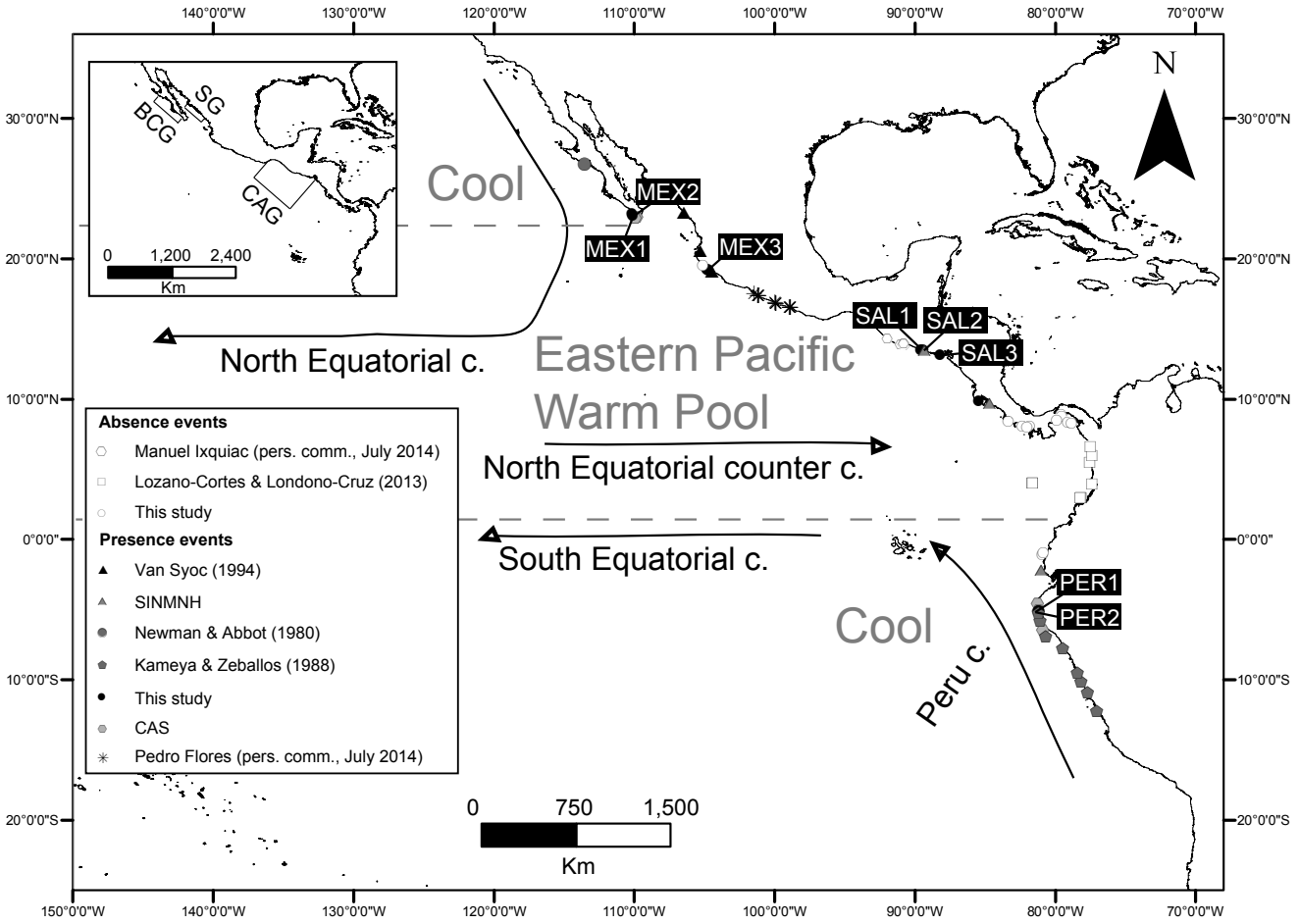

Figure 2.1: Map of the eastern tropical Pacific showing distribution of *Pollicipes elegans* and major surface current systems. Major current systems from Pennington *et al.* (2006). Presence and absence events as indicated in figure legend. Major regions without rocky intertidal habitat in the tropical eastern Pacific, the Sinaloan Gap (SG), the Central American Gap (CAG), and the mainland of Baja California Gap (BCG) (Hastings 2000) are shown on the inset.

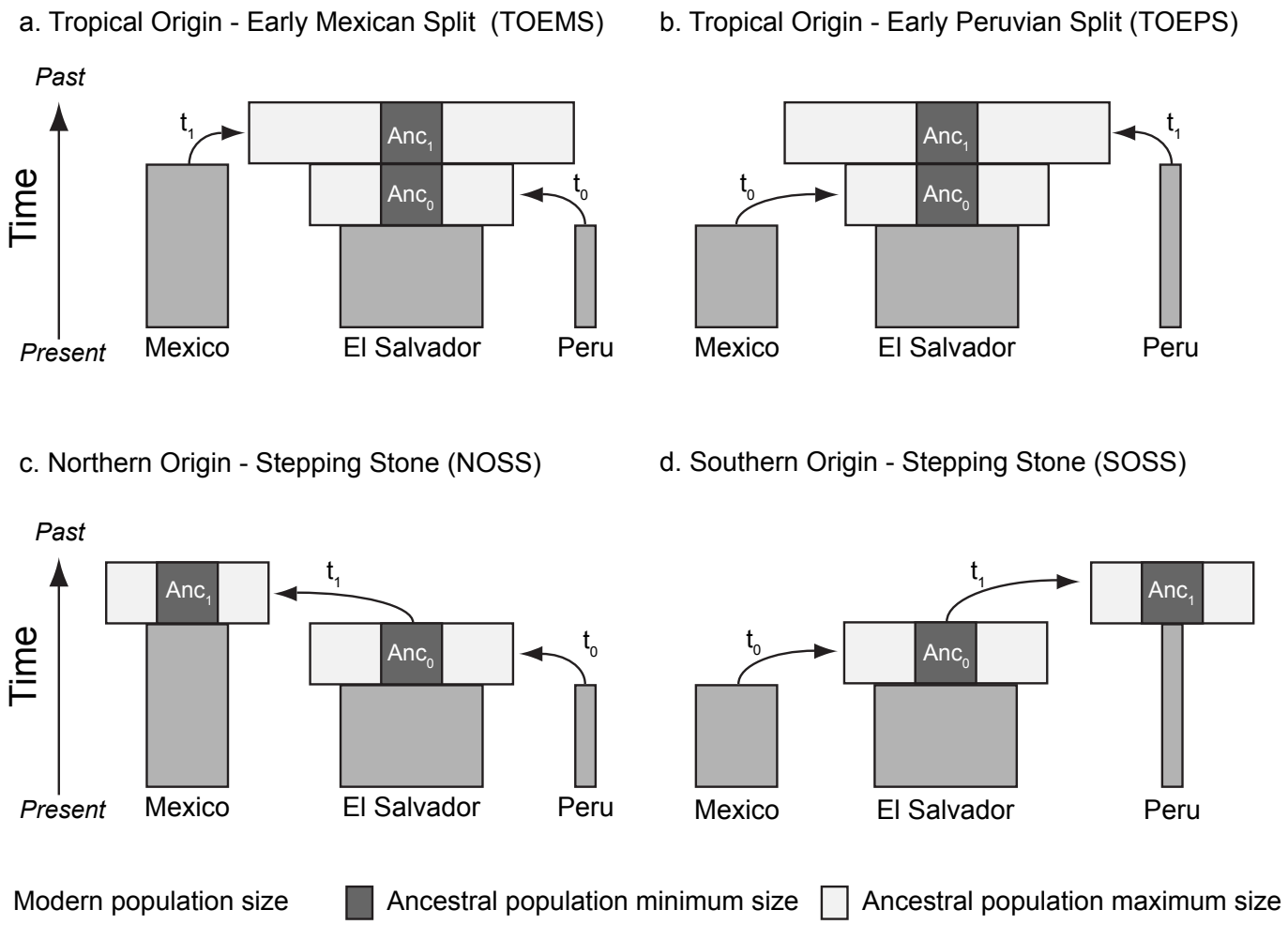

Figure 2.2: Population models used in BSSC simulations followed by Approximate Bayesian Computation to evaluate the plausibility of vicariance and dispersal as the origin of antitropical distribution of *Pollicipes elegans.* $T_{0\text{-}1}$ time of splitting event; $Anc_{0\text{-}1}$ Ancestral N_e at the moment of splitting.

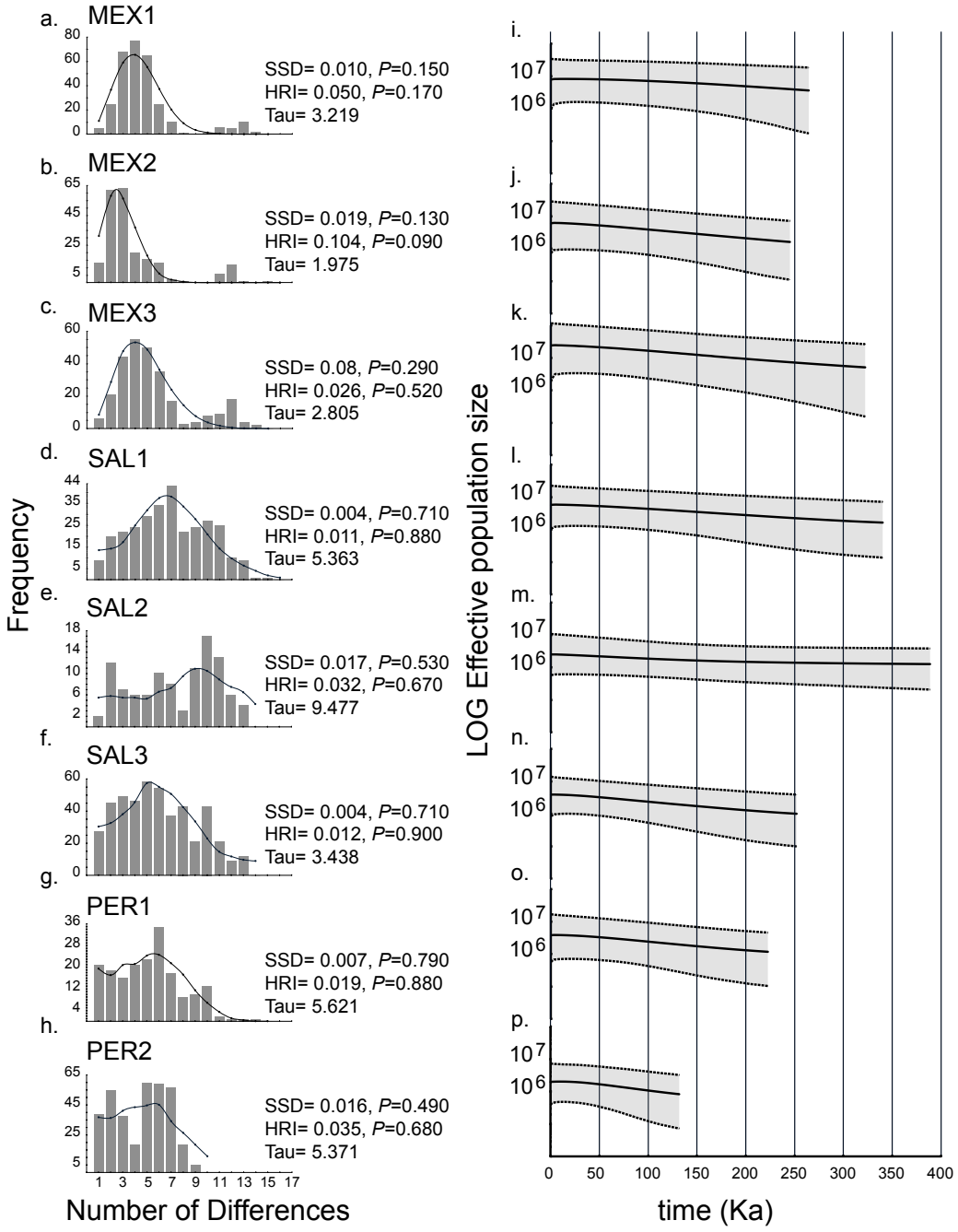

Figure 2.3: Mismatch distributions for each sampling site of *Pollicipes elegans* (a-h). Bars represent observed values and lines indicate the mismatch distribution expected from a sudden expansion model. SSD and HRI correspond to sum of squared deviations and Harpending's ruggedness index respectively. Bayesian skyride plots of the logarithm of N_e change through time (i-p). The black line indicates the mean posterior N_e through time. The gray area represents the 95% HPDI, taking into account coalescent model and phylogenetic uncertainty. Sampling site mismatch distribution (a-h) corresponds to skyride plots in (i-p)

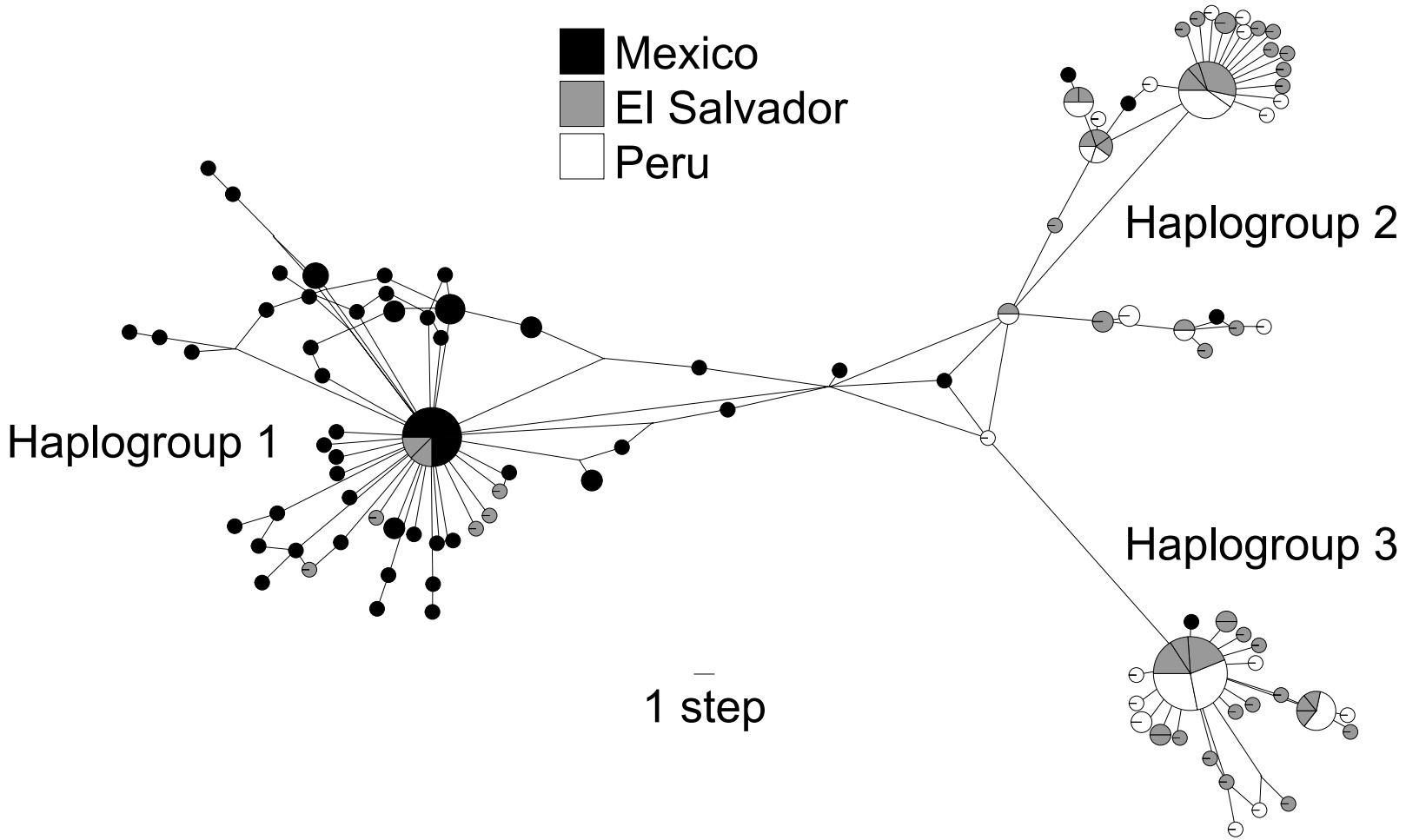

Figure 2.4: Unrooted median-joining network of haplotypes based on cytochrome c oxidase sequences in *Pollicipes elegans* populations. Area of circles is proportional to haplotype frequency; smallest circle represent one haplotype. Sample origin is color coded as indicated in the figure legend.

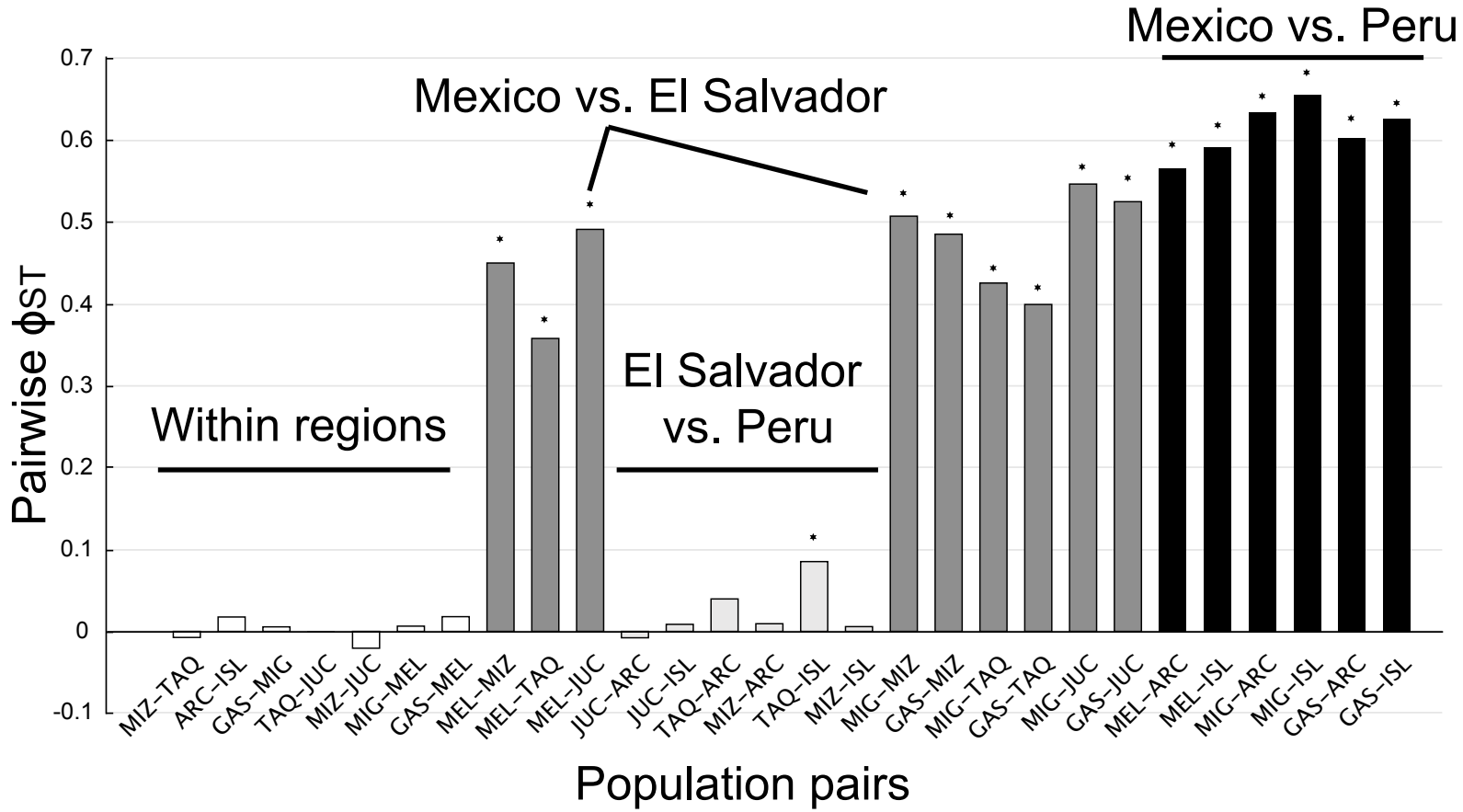

Figure 2.5: Pairwise Φ_{ST} among *Pollicipes elegans* populations. Population differentiation is presented from lower to higher geographical distance. Black bars represent population comparisons between Mexico and Peru populations. Grey bars represent comparisons between Mexico and El Salvador. Light grey bars correspond to El Salvador and Peru comparisons. White bars correspond to pairs of comparison between populations within the same region (*, $P < 0.05$). Population pairs are sorted in the x axis from smaller to larger geographical distance.

CHAPTER THREE

CRYPTIC DIVERSITY IN THE GOOSENECK BARNACLE *POLLICIPES ELEGANS* AS RESULT OF RECENT VICARIANCE IN THE TROPICAL EASTERN PACIFIC

ABSTRACT

Species are the most basic units of classification and diversity measurements in biology. However, the delimitation of species is often difficult because recently separated biological species may not yet have accumulated many distinguishing phenotypic differences. Species delimitation based on analyses of multilocus datasets have transitioned from the identification of species based on reciprocally monophyletic groups (no shared diversity among cryptic species) to the recognition of species before lineages completely sort at neutral loci (species delimitation in the presence of incomplete linage sorting), allowing the detection of speciation early in the speciation process. *Pollicipes elegans* is a species of gooseneck barnacle with a disjunct distribution that spans multiple biogeographic provinces and distinctly different sea surface temperature regimes in the tropical eastern Pacific. To test the existence of cryptic diversity in *P. elegans* I surveyed nucleotide variation and analyzed sequence data from six nuclear loci and one mitochondrial gene. Results from a combination of phylogenetic and phylogeographic species delimitation methods provide strong evidence that *P. elegans* contains two cryptic species. I therefore recommend separation of populations in the California/Mexican transition zone and the Mexican zoogeographic province, and populations south of the Central American Gap into two species.

INTRODUCTION

Species are the most basic unit of biological classification, and correct species delimitation is fundamental for understanding the ecological and evolutionary processes that generate and maintain diversity (Fouquet *et al.*, 2014). Species delimitation is also important in conservation for the identification of species and regions of conservation priority, such as rare endemic species and hot-spots of biodiversity (Sites Jr & Marshall, 2003; Sites & Marshall, 2004). Despite the clear importance of understanding species boundaries, delimiting species is often a difficult task because speciation is a continuum process where populations can represent lineages at different stages of differentiation and distinctiveness (De Queiroz, 2007). If isolation is strong or long enough, genetic drift or natural selection will eventually result in the fixation of observable traits – i.e. diagnostic characters – that can be used for the identification of evolutionary independent lineages (Lande, 1976). However, because phenotypic differences that can serve as diagnostic traits may lag far behind reproductive isolation, some species may remain morphologically cryptic despite functioning as ecologically distinct species (Giarla, 2013).

During speciation, ancestral lineages will eventually sort themselves into reciprocally monophyletic clades through a combination of stochastic (i.e., mutation and genetic drift) and deterministic (i.e., selection) processes (Avise, 2000). However, it may take a long time for reproductively isolated species to achieve reciprocal monophyly at enough genes to be easily detected (Pamilo & Nei, 1988; Maddison, 1997; Degnan & Rosenberg, 2009). This process is known as incomplete lineage sorting and occurs when

ancestral polymorphisms have not sorted to monophyly between species through differential extinction of alleles or haplotypes.

The use of molecular markers has proven useful in the identification and delimitation of species (Hebert *et al.*, 2003), particularly in the identification of taxa that are phenotypically plastic or morphologically cryptic. Because most of the molecular markers used in population genetics and phylogeography play no role in the speciation process, detection of species is only possible once these neutral genes have stochastically sorted into reciprocally monophyletic clades (Ryder, 1986; Moritz, 1994; Baum & Shaw, 1995). Incomplete lineage sorting (i.e. retention and stochastic sorting of ancestral polymorphisms) is expected to be most conspicuous in species with large effective population sizes and recent divergence times, as is often the case with many marine species (Rosenberg, 2003; Eytan & Hellberg, 2010). For example, snapping shrimps of the genus *Alpheus* found on opposite sides of the Isthmus of Panama have to yet reach complete monophyly despite complete isolation for more than 3 MY and evidence of strong pre- and post-mating isolation mechanisms (Knowlton *et al.*, 1993). However, discordant gene genealogies can provide information about species boundaries even when incomplete lineage sorting (ILS) is present (Hey & Nielsen, 2004; Maddison & Knowles, 2006; Knowles & Carstens, 2007).

Recent methodological advances in phylogenetic species delimitation (Liu *et al.*, 2009; Heled & Drummond, 2010; Yang & Rannala, 2010) utilize coalescent theory to infer species boundaries (or the species tree) when there is gene tree discordance. These methods model the divergence of populations across many loci, such that

populations, even those lacking reciprocally monophyletic markers as a consequence of incomplete linage sorting, can be delimited as evolutionarily independent lineages or species. Different coalescent-based methods for species delimitation use different algorithms. For example, *BEAST integrates the data of multilocus genes considering the data given a gene tree (Felsenstein likelihood), the likelihood of a gene tree given a species tree (multispecies coalescent prior), and a prior in the species tree (Birth Death model or Yule model). Nevertheless, all methods provide a probability for the clade being a separate evolutionary unit with a unique coalescent history across loci. Later, a predefined threshold is used to delimit evolutionary significant units (ESUs) of diversity, a concept consistent with the Evolutionary Species Concept (Simpson, 1951) and the General Lineage Concept of species (De Queiroz, 2005b). As with stricter phylogenetic species concepts that require complete reciprocal monophyly, species identified with coalescent Species Delimitation Methods (SDMs) may not be good biological species (i.e. lacking strong, intrinsic isolating mechanisms); however, an absence of gene flow is what causes species to progress toward reciprocal monophyly.

Here I explore the presence of cryptic diversity produced by vicariance in the gooseneck barnacle *Pollicipes elegans,* a marine taxon with a fragmented distribution in the tropical eastern Pacific (**Fig. 3.1**). The nominal *P. elegans* inhabits the low intertidal zone of wave exposed rocky shores in the tropical eastern Pacific. The species is more abundant in the peripheral portions of its range (**Fig. 3.1**) than in the center of its latitudinal range (**Fig. 3.1**), showing conspicuous gaps in the distribution between El Salvador and both Mexico and Peru (**Fig. 3.1**). A previous phylogeographic study based

on a single mitochondrial gene (Chapter 2) provided evidence that peripheral populations in Mexico, to the north of the Central American Gap (CAG), are genetically isolated from all populations to the south (**Fig. 2.1**, Chapter 2). Although these results were suggestive of cryptic diversity within *P. elegans* (Chapter 2), data from only a single locus that is nearly, but not completely reciprocally monophyletic, does not provide a fully comprehensive answer to the question of whether *P. elegans* consists of one or two evolutionarily separate lineages. Here, I expand on Chapter 2 to explicitly test for diversity in cryptic lineages in *P. elegans* and re-test the conclusions of historical biogeography using a multilocus dataset.

MATERIAL AND METHODS

Geographic sampling, DNA isolation and gene sequencing

Mitochondrial DNA (mtDNA) sequences were electronically retrieved from NCBI GenBank database under accession numbers KF958514-KF958701. Nuclear data were obtained from specimens collected from eight sites throughout the known geographic distribution of *P. elegans*, targeting populations at the north, center, and south portions of the distribution range (**Fig. 3.1**). At each sampling site, muscle samples from the peduncle of 50 live barnacles were preserved in 70% ethanol. Genomic DNA was extracted by overnight proteinase K digestion in 2X cetyl timethyl ammonium bromide (CTAB) followed by two chloroform extractions and ethanol precipitation. Nuclear genes were amplified and sequenced with universal exon-primed intron crossing (EPIC) markers Jarman *et al.* (2002). Intronic sequences included: (1) ATP Synthetase Subunit α (ATPSα), a nuclear-encoded protein complex that uses differences in proton

concentration to synthesize ATP from ADP; (2) Signal Recognition Particle 54-kDA subunit (SRP54,) a polypeptide complex that binds to the signal sequence of newly synthesized protein, assisting in transportation to the endoplasmic reticulum; and (3) Zinc Metalloproteinase (ZMP), a protein involved in cleaving intercellular peptide signaling molecules (Alberts *et al.*, 1994; Pan & Rubin, 1997). Multicopy nuclear ribosomal interspacers 1 and 2, and 5.8S rDNA sequences were amplified using primers POL18S1F and POL18S2R (Quinteiro *et al.*, 2011) and included in the analyses. Polymerase chain reaction (PCR) was carried out in 20 µL volumes containing 1x PCR buffer, 3 mM MgCl2, 0.4 mM each dNTPs, 0.5 U of GoTaq DNA polymerase (Promega), 0.3 µM each primer, and approximately 50 ng of genomic DNA.

Amplification conditions consisted of 94°C for 2 min followed by 5 cycles of 94°C for 30 s, 42°C for 1:30 min, and 72°C for 1 min. These first 5 cycles were followed by 35 cycles of 94° for 30 sec, 49°C for 1 min, 72°C for 1 min, with a final incubation step at 72°C for 5 min. PCR products were purified using QIAquick spin filter columns (QIAgen) and sequenced in both directions on an ABI-3730 sequencer (Applied Biosystems).

Alignment, phasing and recombination detection

Clear nuclear gene sequences containing multiple heterozygous sites and no insertions or deletions (or indels) were analyzed in PHASE v.2.1.1 to resolve the gametic phase of each sequence. Input files were prepared for PHASE (Stephens & Donnelly, 2003; Stephens & Scheet, 2005) using the online software package SeqPHASE (Flot, 2010). Allelic combinations phased with 0.90 pp (posterior probability) or higher were retained. Sequences containing indels or low pp allelic phase were cloned to resolve

constituent allelic sequences using the Invitrogen TOPO TA Cloning Kit (with pGEM®-T Easy Vector System I; Promega, Madison, WI). Six to 10 clones were screened for correct insert size, cleaned using a DNA purification system (Wizard Plus SV Minipreps; Promega, Madison, WI) and sequenced using the pUC/M13 sequencing primers. Putative recombinant regions were excluded using the software IMgc (Woerner *et al.*, 2007). Matrices for all gene fragments were edited using BioEdit (Hall, 1999) and aligned with ClustalW (Thompson *et al.*, 2002). Length variable data in rDNA was aligned with the program MAFFT (Katoh *et al.*, 2005) using the G_INS-I alignment algorithm. The program Gblocks (Castresana, 2000) was used to remove alignment ambiguous regions in the rDNA with the following settings: minimum number of sequences for a conserved position and flanking regions: 98, maximum number of contiguous non-conserved positions: 8, minimum length of a block: 5, and allowed gap positions: with half. All sequences were inspected for base calling errors with the forward and reverse strands and consensus sequences generated using Sequencher 4.7 (Gene Codes Corporation).

Individual Gene Trees

The Bayesian Information Criterion (BIC) in PartitionFinder v.1.1.1 (Lanfear *et al.*, 2012) was used to find the best sequence partition strategy and the best-fitting model of nucleotide substitution for each partition. GARLI v.0.96 (Zwickl, 2006) was used to estimate individual gene trees using maximum likelihood (ML). Bootstrap support for ML trees was generated with 1000 replicate searches, each consisting of two independent searches. Gene trees were converted to ultrametric trees using the R

v.3.1.1 (R Core Team 2014) package ape v.3.1.4 function chronos (Paradis *et al.*, 2004) and later used as input for some of the species delimitation methods described below.

Divergence time, effective population size and migration estimates

I estimated the posterior densities of the demographic parameters time of divergence (t), theta (θ), and migration (m) by sampling the coalescent in a Bayesian framework with the program IMA2 (Hey, 2010), which is based on an isolation-with-migration (IM) model (Nielsen & Wakeley, 2001; Hey & Nielsen, 2007; Hey, 2010). Because IMA2 requires know the phylogenetic relationships among populations when more than 2 populations are analyzed, I performed pairwise analyses among all populations to obtain the most recent estimate of effective θ for inclusion in the BP&P analysis (see below). The mean of trans-isthmian estimates of COI divergence rates (2%) from the most recently separated species among crustacean lineages (Knowlton & Weigt, 1998; Schubart *et al.*, 1998) and a generation time of one year estimated for the congeneric species *P. pollicipes* (Cruz & Araujo, 1999; Cruz *et al.*, 2010) were used to obtain the results in demographic units (i.e. number of individuals, time in years and migration as number of migrants per generation). Initial runs of 40 independent heated chains (Geyer, 1991) were performed to assess if the priors were suitable and if the heating conditions were appropriate. Then, two independent runs with 8 hours burn-in time and 64 hours in L-mode were used to produce the final results. Stationarity was assessed by visual inspection of the splitting time trend plots and by comparisons between parameter estimates generated from genealogies from the first and the second half of the run (Hey, 2010).

Species Delimitation Methods

Coalescent-based species delimitation methods (SDMs) require partitioning samples into separate lineages (putative species) before testing species limits; therefore, coalescent SDMs determine the number of species by attempting to fuse lineages established *a priori*. To initially delineate cryptic lineage diversity in this system, I explored species partitioning using two approaches: patterns of genetic differentiation and a coalescent tree-based method. Genetic differentiation was first estimated among all populations by pairwise calculation of Φ_{ST} for each gene separately and for the concatenated dataset using ARLEQUIN v.3.5.1.3 (Excoffier & Lischer, 2010). Partitioning was then explored with the Bayesian MCMC sampler implemented in *BEAST v.2.1.3 (Bouckaert *et al.*, 2014). *BEAST generates posterior samples by simultaneously estimating gene and species trees under a hierarchical coalescent model while allowing for independent evolutionary processes in each genomic region. BEAUTi version 1.7.5 (Heled & Drummond, 2010) was used to create XML-formatted input files for *BEAST. Substitution models were chosen as previously described and were unlinked across genes, with parameters estimated separately for each gene. As a needed prior in *BEAST, sampling sites were assigned to species groups and evolutionary rates were estimated under a Yule process (Yule, 1925). A Yule model was chosen as the species tree prior, which assumes a constant lineage birth rate for each branch in the tree. This tree prior is most suitable for trees describing the relationships between individuals from different species and is often thought of as describing the net rate of speciation (Yang & Rannala, 2006). Species tree estimations were carried out with the assumption

of a strict molecular clock (Heled & Drummond, 2010). Data sets were run for 50 million generations in BEAST, sampling every 5,000 generations. Analyses were performed twice. Post-burn-in trees were combined with the program LogCombiner, and chains were assumed to converge when the average standard deviation of split frequencies was found to be < 0.01. The maximum clade credibility tree with posterior probabilities for each node was computed with the programTreeAnnotator. Log files were evaluated in TRACER v.1.5 (Heled & Drummond, 2010). The species tree was calculated using TreeAnnotator v.1.7.2 with a burn-in of 5000 trees. FigTree v.1.3.1 was used to visualize node ages, branch lengths and posterior probabilities of the consensus tree.

With groupings of putative species established, species limits were then tested using spedeSTEM v.1.0 (Ence & Carstens, 2011) and BP&P (Rannala & Yang, 2003; Yang & Rannala, 2010). SpedeSTEM finds the full species tree with the maximum likelihood in which each putative species is first assumed to be a separate lineage; spedeSTEM then finds alternative species trees by collapsing two or more sister-lineages into one. The best species tree is then chosen using the Akaike information criteria. SpedeSTEM assumes constant population sizes along the species tree (Burnham & Anderson, 2002) and that the gene trees (estimated with GARLI) are correct. Since spedeSTEM requires all lineages to be represented across loci (i.e., no missing data), all ZMP variants were removed (due to missing specimen data) resulting in a 4 locus dataset for this analysis. Each analysis included 500 replicates using average θ (genetic diversity or $4N_eu$) values estimated by pairwise demographic parameter estimates of two populations using IMa2 (Hey, 2010).

Species limits were also tested with BP&P, which uses the reversible-jump Markov chain Monte Carlo (rjMCMC) algorithm to test hypotheses of alternate species delimitations. BP&P differs from spedeSTEM primarily in that BP&P assumes a fixed species guide tree (provided by the user, in this case from *BEAST at nodes with pp=1.0) but takes into account uncertainty in gene tree reconstruction. Like spedeSTEM, BP&P also assumes no admixture following speciation and takes into account incomplete linage sorting between species. The rjMCMC algorithm either splits or joins putative species at each node on a guide tree, and posterior probabilities are generated for each collapsed subtree of the input guide tree. Posterior probabilities (pp) > 0.95 are interpreted as strong support in favor of a speciation event (Leaché & Fujita, 2010). Multiple analyses with varying priors of θ and τ were run to discern how varying the ancestral population sizes and root ages influenced the results. Species delimitation was run for 500,000 generations with sampling every 5 generations and a burn-in period of 10,000 generations using the algorithm 0 and the parameter fine-tune (ε) set to 10. Each analysis was run four times with different starting seeds to ensure stationarity of the rjMCMC chain, and nodal probabilities of the four independent runs were averaged for the reported guide trees. Due to limitations in the algorithm when handling large datasets (different results among replicate analyses), I pruned the multilocus dataset by removing duplicated haplotypes within sampling sites.

RESULTS

Molecular data

Complete aligned matrices included 210 rDNA (803 bp), 274 SRP54 (154 bp), 322 ATPSα (314 bp), 156 ZMP sequences (i.e. nuclear alleles) and 590 bp of haploid COI (mtDNA) sequences from 188 individuals. The ZMP sequences had three length variants (213, 442, 109 bp) that did not overlap when aligned. When mapped to the transcriptome of *P. elegans* (Chapter 4), sequences aligned with different genes and were considered as separate loci to be used in our analyses (**Table 3.2**). Results from IMgc indicated intralocus recombination in all but one nuclear gene (ZMP32). General information of datasets with and without putative recombining regions can be found in **Table 3.2**.

Gene Trees

The unrooted median joining haplotype networks of non-recombinant fragments showed a strong association of haplotypes with geography in two of the seven genes analyzed. COI and rDNA showed a congruent pattern of differentiation with respect to geography. Haplogroups 1 and 2 from the rDNA and haplogroup 1 from COI were composed primarily of individuals collected in Mexico (northern portion of distribution range) (**Fig. 3.2**); with few haplotypes shared with individuals collected in El Salvador (center of the distribution range) and no haplotypes shared with individuals collected in Peru (southern end of the distribution range). Haplogroup 3 of the rDNA and haplogroups 2 and 3 of COI were dominated by individuals collected in El Salvador and Peru, with substantially fewer individuals from Mexico (**Fig. 3.2**). In contrast, the five

remaining nuclear genes showed no clear associations between haplotypes and geography, with multiple haplotypes shared across all three regions where *P. elegans* occurs (**Fig. 3.2**).

The maximum likelihood mid-point rooted trees of the seven genomic regions showed heterogeneity in topology and support (**Fig. 3.4**). Nodes that separated individuals collected in Mexico from individuals collected in Peru and El Salvador were detected in COI and rDNA with strong bootstrap support (**Fig. 3.4**). Phylogenetic trees of the remaining nuclear genes showed lack of resolution with no clear geographical patterns and low node support.

Genetic Structure

As evident in the gene trees, the nuclear introns showed weaker patterns of genetic structure than the mitochondrial and ribosomal DNA (**Fig. 3.3**). However, strong and significant genetic differentiation was found in a few pairwise comparisons of nuclear introns. The patterns of differentiation of the concatenated dataset agreed with the mtDNA patterns, showing significant differentiation between individuals collected in Mexico when compared with populations sampled in the central and south portion of the distribution range, and low and non-significant differentiation between sites south of the CAG (**Fig. 2.1**, Chapter 2).

Coalescent-based demographic parameter estimates

Likelihood ratio tests from the IM analysis showed evidence for migration between *P. elegans* sites in eight of 28 pairwise comparisons. Significant migration between cryptic species was found at sites PER1 to MEX2 (Nm=0.38), PER2 to MEX3

(Nm=1.94), PER2 to MEX2 (Nm=0.33), SAL3 to MEX3 (Nm=1.44), and within species between sites PER1 to SAL2 (Nm=2.76), PER2 to SAL1 (Nm=5.12), SAL3 to SAL1 (Nm=2.45), MEX3 to MEX1 (Nm=4.49) forward in time. Migration direction at seven of the eight significant results was from south to north, a result that is not consistent with the patterns of oceanographic circulation of the region. In all other comparisons, the joint posterior density for m increased asymptotically as m approached zero, indicating either no gene flow across populations or a lack of migration signal in the data. Divergence time shows an overlap in the posterior density distribution for MEX3-PER2 (~150 to 450 ka) and SAL2-MEX3 (~100 to 300 ka). However, the divergence time between SAL3-PER2 had a more recent and narrower joint posterior density distribution (~30 to 200 ka) (**Fig. 3.6**). Average estimates of N_E based on pairwise comparisons for each collecting site are summarized in **Table 3.1**. Towards the north and center of the species' range, N_E is large (~1,200,00 to 3,900,000), especially when compared to sites in Peru (~850,000).

Species Trees and Species Delimitation Analyses

Node support by BP&P using the *BEAST inference tree indicated the presence of two species (**Fig. 3.5**). The first species corresponds to populations in the north, including the complete range of the California/Mexican Transition Zone (CMTZ) and the Mexican Zoogeographic Province (MP), ending at the northern end of the Central American Gap (CAG) (**Fig. 3.1**). The second species includes populations across a wider distribution range, including all populations south of the CAG, including the Panamic Zoogeographic Province (PP) and the Panamic/Peruvian Transition Zone (PPTZ) (**Fig.**

3.1). SpedeSTEM results were in complete agreement with those inferred with BP&P, showing the lowest AIC score associated with the species tree that collapsed all sites of the northern portion of the range together as a different species than all sites collected south of the CAG (PP and PPTZ zones). The best AIC species tree model from spedeSTEM had a probability weight of ωi = 0.469, a value that is substantially higher than the second best model ωi = 0.006 (**Table 3.4**). This best-fit model clearly indicates that *P. elegans* as currently described corresponds to two cryptic but genetically distinct species.

DISCUSSION

Using a combination of phylogeographic and coalescent-based methods and a larger multilocus data set, the analyses here confirm both the patterns of deep genetic differentiation (Van Syoc, 1994) and ancient divergence times (Chapter 2) reported elsewhere. They also show that the populations of the northern portion of the range of *P. elegans* represent a cryptic species differentiated from all populations to the south of the CAG, in the PP and the PPTZ.

Species delimitation

I used two species delimitation methods to evaluate the presence of cryptic diversity within *P. elegans*. Although the details of the species delimitation algorithm differ, these results indicate that individual gene trees provided the greatest support for a species tree model composed of two species, not the eight groups established *a priori* as sample localities in the case of spedeSTEM, or four groups found in the inference recovered by *BEAST and tested using BP&P. The two cryptic species found within *P.*

elegans were also supported by the pairwise Φ_{ST} patterns of genetic structure at the mtDNA and to some degree by the differentiation at the rDNA locus. The higher and more significant levels of differentiation for mtDNA were not unexpected, as mtDNA has a lower effective population size than nuclear genes and is expected to differentiate faster via genetic drift (Birky *et al.*, 1989).

Allopatric speciation in presence of gene flow

A common assumption in species delimitation methods based on coalescent theory is that species have diverged from a common ancestral species without gene flow after divergence (Camargo *et al.*, 2012). I tested this assumption with the full multilocus data set using IMa2, which showed low but significant levels of gene flow within and between cryptic species in *P. elegans* ranging from 0.33 to 1.94 Nm per generation between the two cryptic species and from 2.45 to 5.12 within the two cryptic species. In the four pairwise population comparisons in which gene flow was detected across the phylogenetic break that delimits the two cryptic species, migration rates were low. Simulation studies have shown that Nm > 10 is required to maintain drift connectivity (Lowe & Allendorf, 2010), meaning that Nm < 10 will to lead to differentiation via genetic drift. Therefore, the rates of gene flow observed are consistent with two cryptic species diverging via genetic drift. In that sense, SDMs should be robust to small amounts of gene flow because populations will move to reciprocal monophyly even with small amounts of gene flow (Nm<10).

Species Delimitation Methods limitations

Although the generation of new tools for data collection and analysis allows scientists to discover new diversity like never before, the discovery of cryptic diversity must become linked to species description and formal taxonomy. Species delimitation results needs to be examined carefully because of the serious consequences that incorrect identification and classification could have in conservation efforts (Bernardo, 2011). For example, the results found by different SDMs may provide contradictory results (different number of species) due to the simplifying assumptions of each algorithm. In those cases, inferences drawn from species delimitation studies should be conservative, focus only on delimitations that are congruent among multiple methods of delimitation (Carstens *et al.*, 2013), and be further examined in a taxonomic formal description.

In the *P. elegans* species delimitation, I performed a two-step analysis to discover cryptic diversity. Given that initial population divergence starts with differentiation in allele frequencies and secondly, with random lineage sorting and mutation that further differentiates lineages during speciation (Hey & Pinho, 2011), methods based on allele frequency clustering are expected to detect lineages earlier than SDMs approaches. First, I identified the smallest clusters of populations based on pairwise Φ_{ST} and the multispecies coalescent criterion of *BEAST. Then I used SDM to test if these clusters represent independent evolutionary lineages based on the pattern of allele coalescence in gene genealogies (BP&P) and independently evaluated the best species tree model using spedeSTEM. My results were congruent with two cryptic

species within *P. elegans,* and I plan to formally describe this species in a forthcoming article.

Local adaptation to the thermal gradient in the tropical eastern Pacific

Local adaptation could play a critical role in the process of speciation of *P. elegans* cryptic species. The lack of adult habitat availability in the Central American Gap - an area dominated by sandy beaches and estuaries interrupted by only short stretches of rocky coastline (Hastings, 2000; Robertson & Cramer, 2009) - may have prevented migration and favor local adaptation in each isolated cryptic species. However, *P. elegans* have an extended planktonic larval period (25-35 days) consisting of six naupliar stages and a cypris (Barnes, 1996; S. Crickenberger, personal communication, July 2013) suggesting that other factors such as temperature could contribute to the current lack of gene flow between regions.

If temperature is the primary factor limiting the colonization and persistence of species across the CAG, signatures of thermal adaptation in larval stages should be detected among populations of cryptic species of *P. elegans.* Consistent with this idea, larval thermal tolerance of Mexico cryptic species of *P. elegans* is lower than local thermal adapted larvae south of the CAG cryptic species providing evidence of a potential physiological barrier to gene flow into the tropics, but not outside the tropics (Walther *et al.*, 2013).

CONCLUSIONS

My study has led to the detection of cryptic lineage diversity within the gooseneck barnacle *P. elegans*, and both populations north of the intertropical

convergence zone and south of the CAG should be considered species. These results have been tested and confirmed through multiple approaches to species delimitation using multilocus data. A previous study (Chapter 2) showed that Mexican and Peruvian populations have a common tropical origin driven by variance during the Pleistocene, a history that fits the expectations of the tropical diversity engine, in which tropical regions supply diversity to extra-tropical peripheral regions (Jablonski *et al.*, 2006; Jablonski *et al.*, 2013). In agreement with those findings, species delimitation confirms the presence of at least two cryptic species for which divergence time and migration estimates support *P. elegans* as a novel model of extra-tropical divergence (cryptic species of Mexico) with a tropical origin.

Future studies should address if natural selection has been a factor driving speciation in *P. elegans*. Marine species often have large effective population sizes (Waples, 1998; Palsboll *et al.*, 2007). With large effective population sizes, even very low migration or dispersal may remove the evidence of population differentiation at neutral loci, but not at locally selected adaptive loci (Allendorf *et al.*, 2010). Selection acts more efficiently on large populations than small populations because the effects of genetic drift are weaker in large populations, and selection coefficients are more likely to be greater than migration (Pampoulie *et al.*, 2006; Hemmer-Hansen *et al.*, 2007).

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. (1994) Molecular biology of the cell. *Garland, New York*, 139-194.

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697-709.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press.

Barnes, M. (1996) Pedunculate Cirripedes of the genus Pollicipes. *Oceanography and Marine Biology, Vol 34*, **34**, 303-394.

Baum, D.A. & Shaw, K.L. (1995) Genealogical perspectives on the species problem. *Experimental and Molecular Approaches to Plant Biosystematics*, **53**, 289-303.

Bernardo, J. (2011) A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. *Systematics Association Special Volume Series*, **78**, 380-438.

Birky, C.W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation, and drift-equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, **121**, 613-627.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, **10**, e1003537.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

Camargo, A., Morando, M., Avila, L.J. & Sites, J.W. (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834-2849.

Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, **17**, 540-552.

Cruz, T. & Araujo, J. (1999) Reproductive patterns of Pollicipes pollicipes (Cirripedia : Scalpellomorpha) on the southwestern coast of Portugal. *Journal of Crustacean Biology*, **19**, 260-267.

Cruz, T., Castro, J.J. & Hawkins, S.J. (2010) Recruitment, growth and population size structure of Pollicipes pollicipes in SW Portugal. *Journal of experimental marine biology and ecology*, **392**, 200-209.

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

De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic biology*, **56**, 879-886.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332-340.
Ence, D.D. & Carstens, B.C. (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473-480.
Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.
Eytan, R.I. & Hellberg, M.E. (2010) Nuclear and Mitochondrial Sequence Data Reveal and Conceal Different Demographic Histories and Population Genetic Processes in Caribbean Reef Fishes. *Evolution; international journal of organic evolution*, **64**, 3380-97.
Flot, J.F. (2010) SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162-166.
Fouquet, A., Cassini, C.S., Baptista Haddad, C.F., Pech, N. & Rodrigues, M.T. (2014) Species delimitation, patterns of diversification and historical biogeography of the Neotropical frog genus Adenomera (Anura, Leptodactylidae). *Journal of Biogeography*, **41**, 855-870.
Geyer, C.J. (1991) Markov-Chain Monte-Carlo Maximum-Likelihood. *Computing Science and Statistics*, 156-163.
Giarla, T.C. (2013) *Systematics, Biogeography, and Phylogeography of Thylamys Mouse Opossums, a Recent Radiation of Neotropical Marsupials*. UNIVERSITY OF MINNESOTA,
Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* (ed by, pp. 95-98.
Hastings, P.A. (2000) Biogeography of the Tropical Eastern Pacific: distribution and phylogeny of chaenopsid fishes. *Zoological Journal of the Linnean Society*, **128**, 319-335.
Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 313-321.
Heled, J. & Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570-580.
Hemmer-Hansen, J., Nielsen, E.E., Frydenberg, J. & Loeschcke, V. (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (Platichthys flesus L.). *Heredity*, **99**, 592-600.
Hey, J. (2010) Isolation with Migration Models for More Than Two Populations. *Molecular biology and evolution*, **27**, 905-920.
Hey, J. & Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D-persimilis. *Genetics*, **167**, 747-760.
Hey, J. & Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785-2790.
Hey, J. & Pinho, C. (2011) Population genetics and objectivity in species diagnosis. *Evolution*,

Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102-106.
Jablonski, D., Belanger, C.L., Berke, S.K., Huang, S., Krug, A.Z., Roy, K., Tomasovych, A. & Valentine, J.W. (2013) Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10487-10494.
Jarman, S.N., Ward, R.D. & Elliott, N.G. (2002) Oligonucleotide primers for PCR amplification of coelomate introns. *Marine Biotechnology*, **4**, 347-355.
Katoh, K., Kuma, K.-i., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, **33**, 511-518.
Knowles, L.L. & Carstens, B.C. (2007) Delimiting species without monophyletic gene trees. *Systematic biology*, **56**, 887-895.
Knowlton, N. & Weigt, L.A. (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 2257-2263.
Knowlton, N., Weigt, L.A., Solorzano, L.A., Mills, D.K. & Bermingham, E. (1993) Divergence in proteins, mitochondrial-DNA, and reproductive compatibilty across the isthmus of Panama. *Science*, **260**, 1629-1632.
Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 314-334.
Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, **29**, 1695-1701.
Leaché, A.D. & Fujita, M.K. (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B: Biological Sciences*, rspb20100662.
Liu, L., Yu, L.L., Pearl, D.K. & Edwards, S.V. (2009) Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology*, **58**, 468-477.
Lowe, W.H. & Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038-3051.
Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.
Maddison, W.P. & Knowles, L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21-30.
Moritz, C. (1994) Defining evolutionarily-significant-units for conservation. *Trends in Ecology & Evolution*, **9**, 373-375.
Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885.
Palsboll, P.J., Berube, M. & Allendorf, F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution*, **22**, 11-16.
Pamilo, P. & Nei, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568-583.
Pampoulie, C., Ruzzante, D.E., Chosson, V., Jorundsdottir, T.D., Taylor, L., Thorsteinsson, V., Danielsdottir, A.K. & Marteinsdottir, G. (2006) The genetic structure of Atlantic cod (Gadus morhua) around Iceland: insight from

microsatellites, the Pan I locus, and tagging experiments. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2660-2674.
Pan, D. & Rubin, G.M. (1997) Kuzbanian controls proteolytic processing of Notch and mediates lateral inhibition during Drosophila and vertebrate neurogenesis. *Cell*, **90**, 271-280.
Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.
Quinteiro, J., Rodríguez-Castro, J., López, P., López-Jurado, L.F., González-Henríquez, N. & Rey-Méndez, M. (2011) Morphological and molecular characterization of a new species of Atlantic stalked barnacle (Scalpelliformes: Pollicipedidae) from the Cape Verde Islands. *Journal of the Marine Biological Association of the United Kingdom*, **91**, 1537-1545.
Rannala, B. & Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645-1656.
Robertson, D.R. & Cramer, K.L. (2009) Shore fishes and biogeographic subdivisions of the Tropical Eastern Pacific. *Marine Ecology Progress Series*, **380**, 1-17.
Rosenberg, N.A. (2003) The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465-1477.
Ryder, O.A. (1986) Species conservation and systematics - the dilemma of subspecies. *Trends in Ecology & Evolution*, **1**, 9-10.
Schubart, C.D., Diesel, R. & Hedges, S.B. (1998) Rapid evolution to terrestrial life in Jamaican crabs. *Nature*, **393**, 363-365.
Simpson, G.G. (1951) The species concept. *Evolution*, **5**, 285-298.
Sites Jr, J.W. & Marshall, J.C. (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, **18**, 462-470.
Sites, J.W. & Marshall, J.C. (2004) Operational criteria for delimiting species. *Annual Review of Ecology Evolution and Systematics*, **35**, 199-227.
Stephens, M. & Donnelly, P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162-1169.
Stephens, M. & Scheet, P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449-462.
Thompson, J.D., Gibson, T. & Higgins, D.G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2.3. 1-2.3. 22.
Van Syoc, R.I. (1994) Genetic divergence between subpopulations of the eastern Pacific goose barnacle Pollicipes elegans: mitochondrial cytochrome c subunit 1 nucleotide sequences. *Molecular marine biology and biotechnology*, **3**, 338-346.
Walther, K., Crickenberger, S.E., Marchant, S., Marko, P.B. & Moran, A.L. (2013) Thermal tolerance of larvae of Pollicipes elegans, a marine species with an antitropical distribution. *Marine Biology*, **160**, 2723-2732.
Waples, R.S. (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438-450.

Woerner, A.E., Cox, M.P. & Hammer, M.F. (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851-1853.
Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212-226.
Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264-9269.
Yule, G.U. (1925) A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 21-87.
Zwickl, D. (2006) *GARLI, vers. 0.951. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Ph. D. dissertation, University of Texas, Austin, Texas, USA,

Table 3.1: Sampling locality information. Province classification based on Pitombo and Bourton (2007) and Laguna (1990). HiPt N_E, maximal probability posterior peak for effective population size.

Site acronym	Province	Locality	Average HiPt N_E
MEX1	Californian-Mexican T.Z.	Gaspareño, Mexico	2,040,854
MEX2	Californian-Mexican T.Z.	Migriño, Mexico	996,106
MEX3	Mexican-Panamic T.Z.	Melaque, Mexico	3,913,112
SAL1	Panamic	Mizata, El Salvador	1,794,830
SAL2	Panamic	Taquillo, El Salvador	1,293,363
SAL3	Panamic	Jucuarán, El Salvador	3,359,630
PER1	Panamic-Peruvian T.Z.	El Arco, Peru	820,039
PER2	Panamic-Peruvian T.Z.	Islilla, Peru	890,413

Table 3.2: Species delimitation and non-coalescent analyses dataset information. Locus information includes name of locus, length (in base pairs), number of sequences, number of unique alleles, number of segregating sites, number of parsimony informative sites, and model of DNA sequence evolution. Numbers in parentheses correspond to the dataset before removal of possible recombinant regions. Collapsed by haplotype dataset was performed retaining at least one haplotype type per sampling site

Locus	bp	Sequences	Collapsed	h	S	PI	Model
COI mtDNA	590	188	135	105	87	40	GTR + gamma (pos 1)
						1	F81 (pos 2)
						3	TrNef + I (pos 3)
rDNA	425 (803)	194 (210)	65	43 (80)	52 (97)	22	K80
ATPSα	224 (314)	278 (322)	76	22 (82)	25 (44)	20	K80
SRP	141 (154)	265 (274)	53	19 (33)	20 (24)	14	K80
ZMP32	213	40	16	11	11	10	K80
ZMP41	155 (442)	34 (44)	17	14 (28)	20 (74)	18	K80
ZMP80	109	70 (72)	27	16 (20)	16 (17)	9	HKY

Table 3.3: spedeSTEM validation results. Results from spedeSTEM validation approach show model, number of free parameters (k), AIC score, AIC differences (Δi), likelihood of the model, and probability of the model (*ω*i)

Model	k	AIC	Δi	Model-likelihood	*ω*i
(MEX1, MEX2, MEX3) (SAL1, SAL2, SAL3, PER1, PER2)	1	11925.50291	0	1	0.469420272
(MEX1, MEX2, MEX3)(SAL2, SAL3, PER1, PER2)(SAL1)	2	11934.13155	8.62864	0.013375642	0.006278797
(SAL3, PER1, PER2)(MEX1, MEX2, MEX3)(SAL1)(SAL2)	3	11933.47747	7.97456	0.018550102	0.008707794
(MEX1, MEX2, MEX3)(SAL3)(SAL2)(PER1, PER2)(SAL1)	4	11946.04899	20.54608	3.45522E-05	1.62195E-05
(MEX1, MEX2, MEX3)(PER2)(SAL3)(PER1)(SAL1)(SAL2)	5	11945.65539	20.15248	4.20673E-05	1.97472E-05
(MEX1)(PER2)(SAL3)(PER1)(SAL2)(MEX2, MEX3)(SAL1)	6	11970.47875	44.97584	1.71246E-10	8.03863E-11
(MEX2)(PER2)(SAL3)(PER1)(SAL2)(MEX3)(MEX1)(SAL1)	7	11976.62675	51.12384	7.9177E-12	3.71673E-12

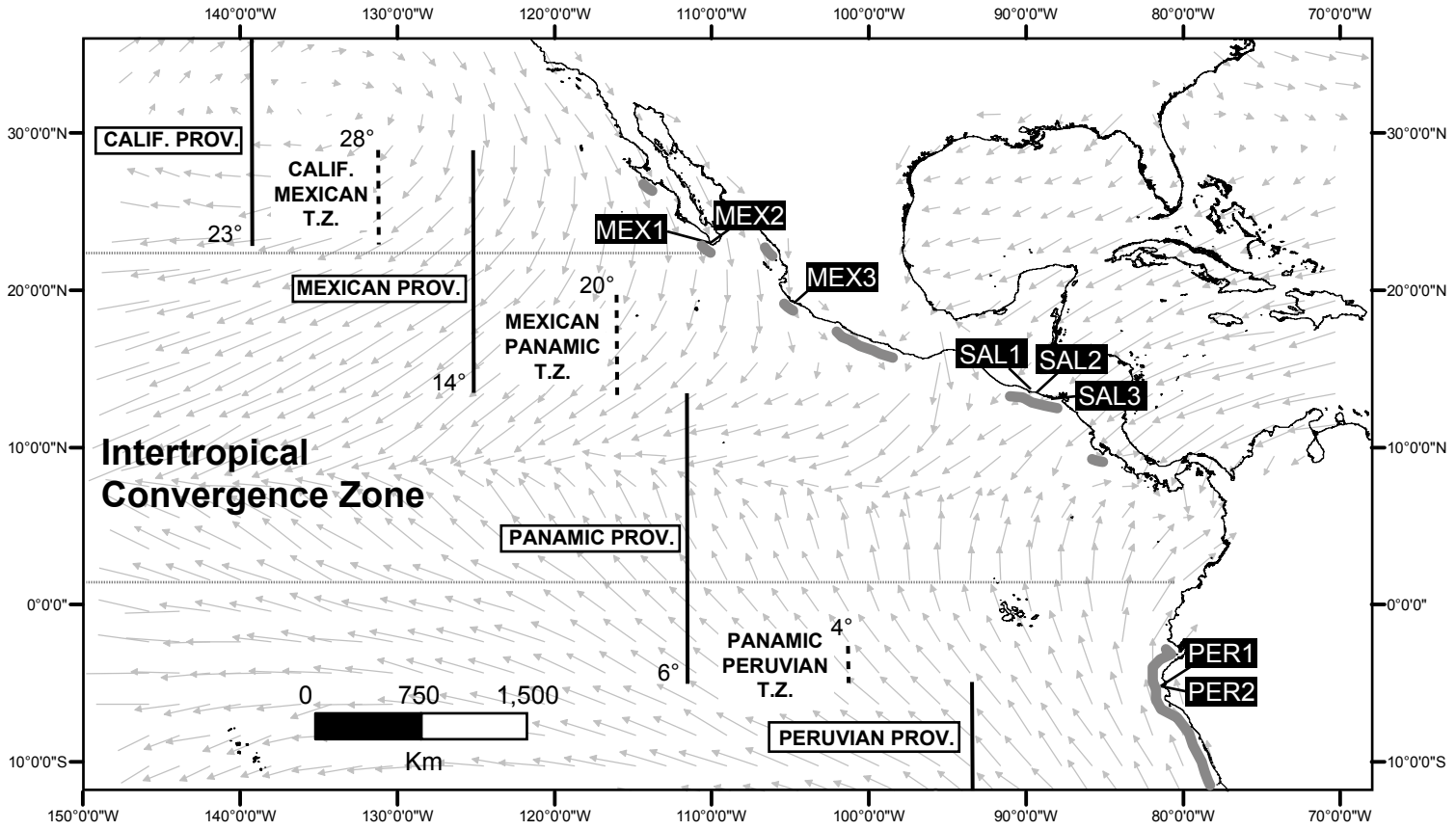

Figure 3.1: Map of the eastern tropical Pacific region showing distribution of *Pollicipes elegans*, wind stress and provinces. Grey blobs represent the known distribution of *P. elegans* after (Chapter 2). Grey arrows represent wind stress climatology data after (Pennington, *et al*. 2006). Provinces are defined after Laguna (1990).

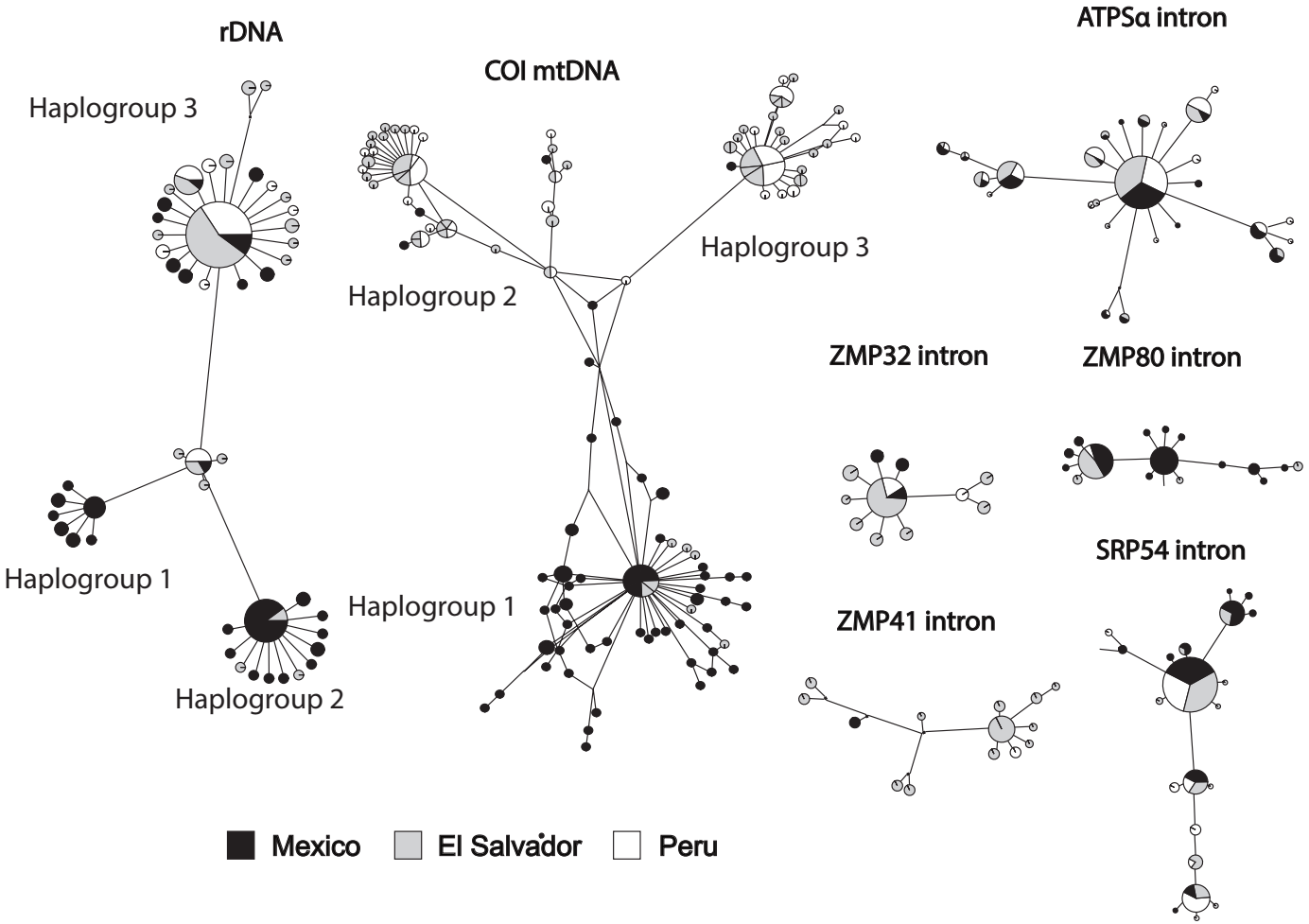

Figure 3.2: Unrooted median-joining network of haplotypes based on cytochrome c oxidase sequences, ATPSα, SRP54, ZMP32, ZMP41, ZMP80 and rDNA in *Pollicipes elegans* populations. Area of circles is proportional to haplotype frequency; smallest circle represent one haplotype. Sample origin is color coded as indicated in the figure legend.

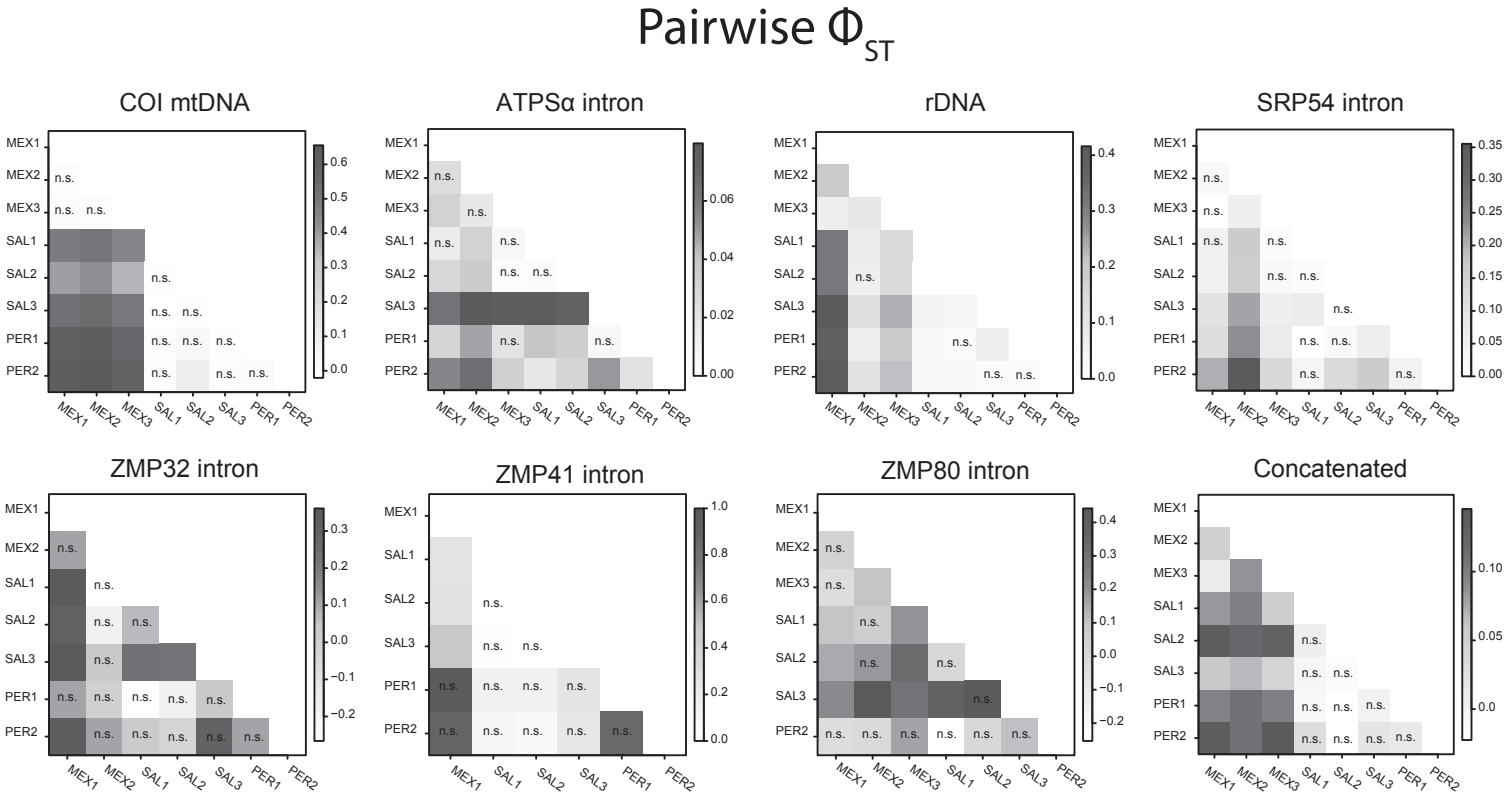

Figure 3.3: Heat map of pairwise Φ_{ST} among *Pollicipes elegans* populations for nuclear, mitochondrial and concatenated datasets. Darker colors represent higher genetic differentiation. Differentiation scale color code varies per individual gene. ns; non-significant genetic differentiation at α= 0.05.

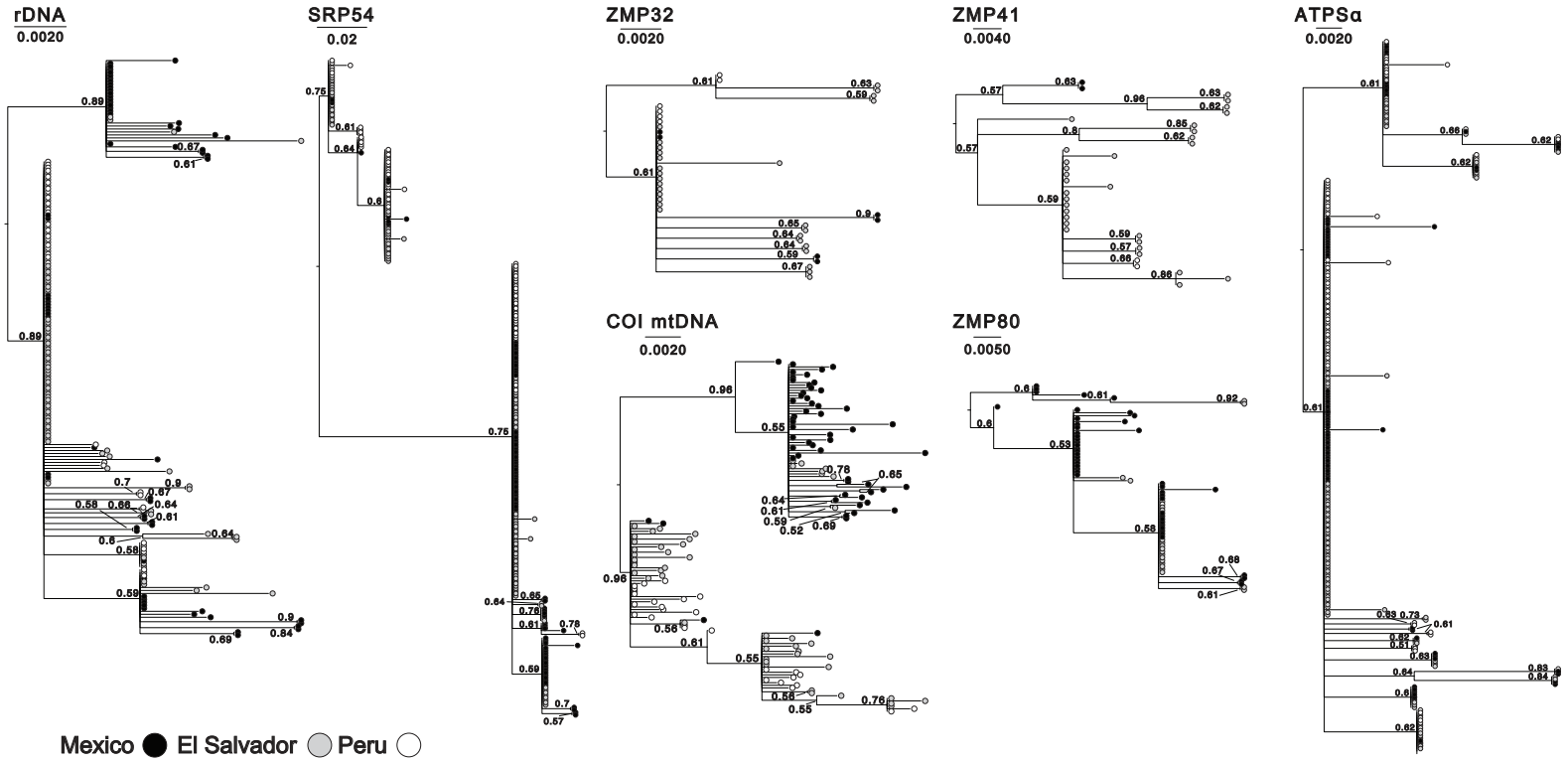

Figure 3.4: Maximum likelihood tree inference was generated from partial sequences of the mitochondrial cytochrome oxidase 1 (COI) gene and the nuclear genes ATPSα, SRP54, ZMP32, ZMP41, ZMP80 and rDNA of *Pollicipes elegans* populations. Confidence support above the main nodes corresponds to bootstrap value. Sample origin is color coded as indicated in the figure legend. Figure is cut-off and numbers are unreadable.

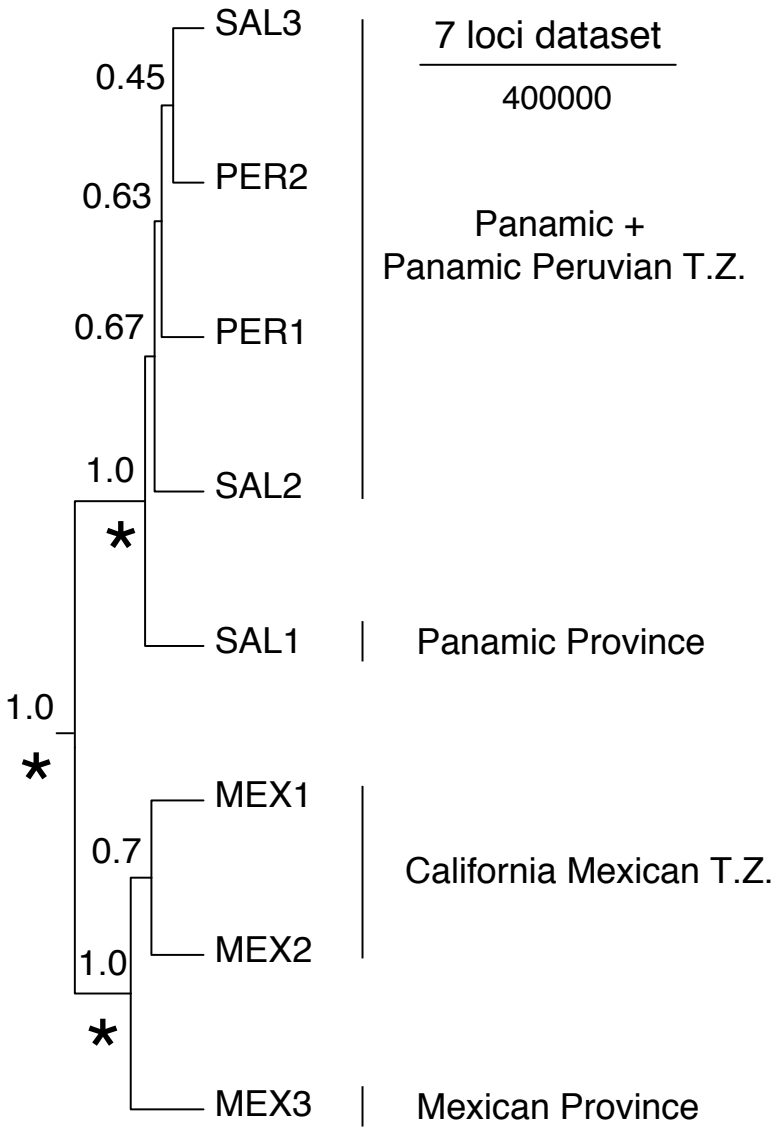

Figure 3.5: Species delimitation results. BP&P results represented on species tree estimated with *BEAST using the complete nuclear and mtDNA dataset. Numerical values above nodes represent posterior probabilities (*pp)* values obtained with *BEAST (*pp* > 0.95 supports putative species); asterisks below nodes denote species delimitation test in BP&P analyses with strong support for groups that correspond to different species. Scale bar represents time in years and vertical lines represent biogeographical provinces and zones after Laguna (1990).

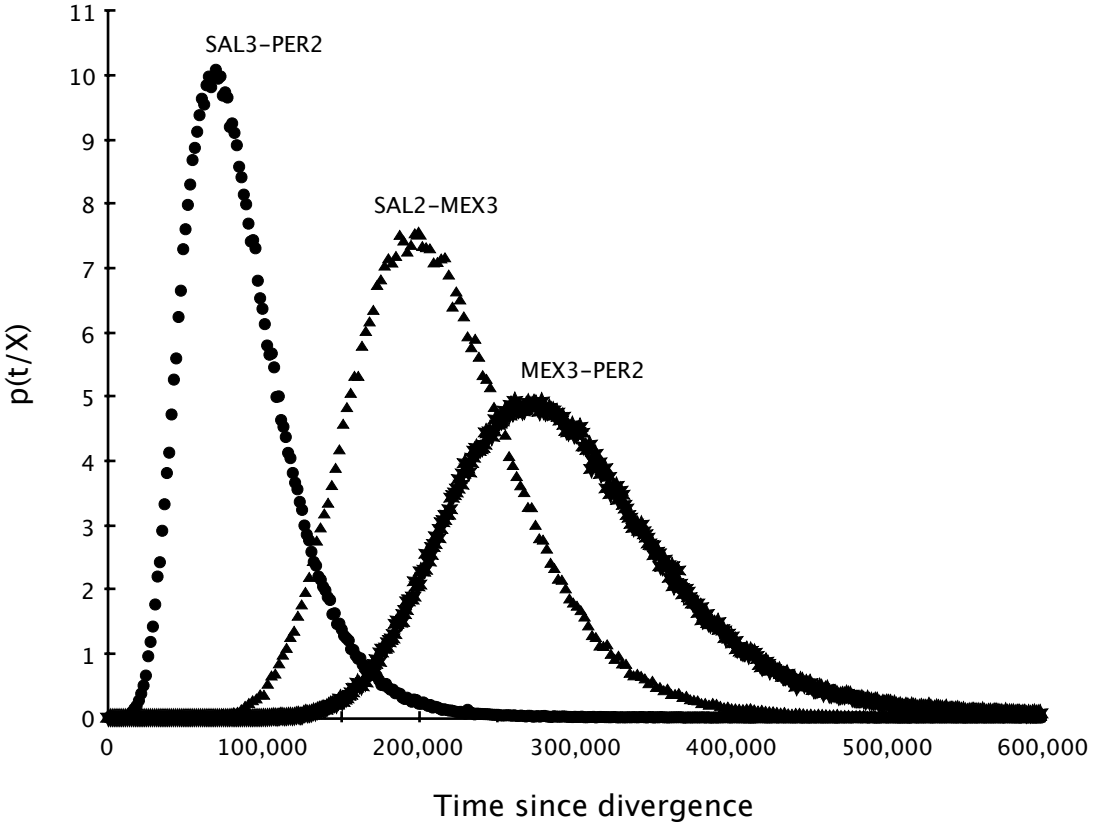

Figure 3.6: Joint posterior density plots obtained from IMA2 analyses based on nuclear and mtDNA data of *Pollicipes elegans*. The most recent divergence time estimate between pairs of populations is presented among regions.

CHAPTER FOUR

THE TRANSCRIPTOME OF THE PACIFIC GOOSENECK BARNACLE *POLLICIPES ELEGANS* OBTAINED FROM RECENT DIVERGENT POPULATIONS: INSIGHTS INTO THE MOLECULAR BASIS FOR LOCAL ADAPTATION FOLLOWED BY ALLOPATRIC SPECIATION

ABSTRACT

Identifying functional genetic variation of ecological and evolutionary importance is fundamental to understanding the role of local adaptation and its relation with speciation. The barnacle *Pollicipes elegans* seems an ideal system to scan the genome for signatures of selection because recently diverged species are found in contrasting regimes of seasonal temperature and have remained isolated for enough time to be recognized as cryptic species. Using RNA sequencing data from individuals collected from Mexico and El Salvador, I built the first transcriptome draft for the group and mine the dataset for microsatellite and single nucleotide polymorphisms (SNPs) markers. Additionally, I mapped the RNA-seq data from each individual back to the transcriptome and compared the frequencies of SNPs to identify genomic regions potentially under natural selection. A total of 11 outlier genes under spatial diversifying selection could be detected, and their function is discussed based on the putative biological function of the genomic regions.

INTRODUCTION

Advances in DNA-sequencing technologies have opened the possibility of understanding adaptation at the molecular level. Although a wide range of statistical methods have been developed to infer selection (Watterson, 1977; Hudson *et al.*, 1987; Tajima, 1989; Fu, 1996), identifying functional genetic variation of ecological and evolutionary importance has been proven difficult given the difficulty of separating demographic history of populations from

signals of selection (Nielsen, 2001). However, with the development of next-generation sequencing technologies, now it is possible to survey hundreds or thousands of loci that can be used to identify genomic regions or "outlier loci" that exhibit patterns of variation that deviate from the rest of the genome. This method is capable of distinguishing natural selection from other genomic effects such as genetic drift and gene flow using the assumption that patterns of genetic variation common to thousands of genes are most likely driven by demographic processes (Vasemagi & Primmer, 2005). Since the development of next-generation sequencing, discovery of genes and genomic regions associated with local adaptation beyond model taxa has become possible (reviewed in Stapley *et al.*, 2010; Udpa *et al.*, 2014), allowing identification of the genetic mechanisms that drive phenotypic divergence in taxa such as three-spined stickleback (Deagle *et al.*, 2012), adaptation to temperature in the isopod *Tigriopus californicus* (Schoville *et al.*, 2012) and the Atlantic cod *Gadus morhua* (Bradbury *et al.*, 2010), and flowering time in sunflowers (Renaut *et al.*, 2014).

The Pacific gooseneck barnacle *Pollicipes elegans* (Lesson, 1831) is a marine species that inhabits wave-exposed rocky shores in the tropical eastern Pacific. With a distribution range that spans over 4,000 km, *P. elegans* is well suited for studying the importance of local selection associated with allopatric speciation for several reasons. First, *P. elegans* has a disjunct distribution across a steep latitudinal thermal gradient such that populations are geographically isolated in regions with different sea surface temperatures (SSTs). SST is consistently higher at the center of the distribution range (i.e. El Salvador) than in the peripheral range (i.e. Mexico and Peru), with strong differences in seasonal variation of the sea surface temperature and nutrient availability notably higher in the peripheral populations (**Fig. 1.1, 4.1**). Second, an

estimate of the divergence time between central and peripheral populations suggests that populations of *P. elegans* in Mexico separated from those in El Salvador approximately 350 ka, with negligible gene flow between the two regions since the split (Chapters 2,3). Finally, multilocus species delimitation methods found strong support for the existence of two cryptic species within the nominal *P. elegans* (Chapter 3). Together, an absence of gene flow between recently separated populations living in contrasting environmental conditions provide a unique opportunity to characterize the earliest steps in functional genetic divergence between incipient species of *P. elegans*.

My study had three main objectives. First, I assembled the first draft transcriptome of *P. elegans* using RNA sequencing data from individuals collected in Mexico and El Salvador. Second, I mined the transcriptome dataset for new genomic tools for the study of this non-model species: microsatellite and single nucleotide polymorphisms (SNPs). Finally, I compared the frequencies of SNPs to identify genomic regions potentially under selection. The results are discussed based on the putative biological function of the genomic regions differentially expressed and under selection.

MATERIAL AND METHODS

Sampling of Pollicipes elegans from Mexico and El Salvador.

Live adult barnacles were obtained from wave exposed low intertidal sites in La Libertad, El Salvador on March 11 of 2013 and Baja California, Mexico, on September 20 of 2013 during low tide (**Fig. 4.1**). Animals were transported overnight in a cooler to Clemson University and kept in a closed system aquaria at 25°C, pH 8.0-8.2, and salinity of 31 ppm for at least one-week prior dissection of muscle tissue from the peduncle. Total RNA was extracted

from live tissue from four individuals per population using Trizol® (Invitrogen, USA) and a purification column system (Qiagen, USA) following the manufacturer's protocol. Total RNA sample aliquots were frozen in liquid nitrogen and shipped on dry ice to the Genome Sequencing Analysis Facility (GSAF) at University of Texas, Austin. Confirmation of sample quality and concentration was conducted using the Agilent 2100 Bioanalyzer instrument (Agilent Technologies, USA).

Library preparation and sequencing

Sample preparation for Illumina sequencing was performed by GSAF using the TruSeq RNA-Seq Sample Prep kit (Illumina Inc., USA) and a custom protocol detailed in Podnar *et al.* (2014) . Briefly, poly-A RNA was isolated from total RNA and fragmented using the NEBNext magnesium RNA fragmentation module (New England Biolabs, USA). cDNA synthesis was performed using SuperScript® III reverse transcriptase (Invitrogen, USA) followed by end repair and addition of adenosienes to the 3' ends. Adapters were ligated to the cDNA and fragments were purified and size selected (~200 ± 25 bp) using Agencourt AMPure XP beads (Beckman Coulter, USA) prior to PCR enrichment. The eight libraries were quantified using a Bioanalyzer 2100 (Aligent Technologies, USA), pooled in equimolar amounts, and sequenced on two lanes on the Illumina HiSeq2000 (Illumina Inc.) as paired-end reads of length 101 bp. After trimming adaptors and removing sequences shorter than 36 bp with Trimmomatic v.0.32 (Bolger *et al.*, 2014), less than one percent of the raw data (0.68%) was excluded (**Table 4.1**). All raw reads were submitted to the NCBI Sequence Read Archive. Assembly was performed with the Trinity pipeline using filtered paired reads from all eight libraries pooled into one dataset.

De novo transcriptome assembly, gene prediction, and annotation

Transcriptome assembly was performed with the Trinity package (Grabherr *et al.*, 2011). The Trinity *de novo* RNAseq assembly pipeline consists of three steps (Inchworm, Chrysalis, and Butterfly) and was executed using default parameters and implementing the –PasaFly flag, since it provides a balance between the number of isoforms reconstructed (i.e. alternative splicing variants, herein transcripts), and sensitivity for full length transcript reconstruction (i.e. longest alternative splicing isoform, herein unicontig) (Brian Hass August 2014, personal communication). Assembly was completed on a computer node with 16 CPU and 62 gb of RAM on the Palmetto cluster high performance computing (HPC) at Clemson University. Trinity pipeline output consists of a single FASTA file with alternative spliced isoforms (transcripts) for each unicontig model. Coding sequence was predicted from each transcript using Transdecoder. This method builds a Markov model that compares a randomization of the 500 longest open reading frames (ORF) against a randomization of all ORFs. Only the putative ORFs that outscore the other reading frames are retained in the gene prediction step. Transcripts were annotated using the UniProt SwissProt database, Pfam-A, eggNOG, and gene ontology utilizing the Trinotate annotation pipeline. Transcripts were first aligned against the UniRef-Swissprot database using blast+ v.2.2.29 with an e-value cutoff of 1.0E-5. Protein domains were identified by searching the Pfam_A database with HMMER 3.0, a domain noise cutoff, and signal peptides and transmembrane regions annotated with SignalP 4.1 and TMHMM 2.0, respectively. The resulting outputs were loaded into a Trinotate database where eggnog and Gene Ontology (GO) terms were added and the resulting annotation set was exported as a delimited file for further analysis. The resulting annotated genes and transcript sequences were submitted to NCBI as Transcriptome Shotgun Assembly.

Read mapping and N50

Transcript abundance was estimated at both unicontig and transcript level for each sample using RSEM. This method is appropriate because it can provide estimates for when multiple isoforms map back to the same sequences. The align_and_estimate_abundance.pl script distributed with Trinity uses bowtie v.0.12.7 to map the reads to the transcriptome, and after sorting the bam file, estimates the abundance of each fragment using RSEM. Assessment of the quality of the transcriptome assembly was estimated later using the count_features_given_MIN_FPKM_threshold.pl and count_N50_given_MIN_FPKM_threshold.pl. These scripts aim to identify the number of unicontig and transcripts given a threshold (set to FPKM> 0.5) to obtain a summary statistic of counts and N50 excluding fragments with low expression and likely low biological significance.

SSRs, SNPs identification, and selection inference using outlier FST method

Simple Sequence Repeats (SSRs) can be found within protein-coding genes, their untranslated regions (UTRs), and introns. The expansion or contractions of SSRs within those regions can lead to phenotypic changes, by gain or loss of gene function via frame shift mutation if SSRs are within coding regions, regulation of gene expression or disruption of splicing, for SSRs within UTRs, or affecting gene transcription, mRNA splicing, or export to cytoplasm when SSRs are found in intronic regions (Li *et al.*, 2004). Simple Sequence Repeats (SSRs) were identified in the final assembly using MSDB v.2.4.1 (Du *et al.*, 2013). Settings were adjusted to require a minimum of 12, 7, 5, 4, 4, and 4 perfect repeats for mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs. PCR primers in the flanking regions of the detected SSRs were designed using Primer3 v.1.1.4 (Untergasser *et al.*, 2012) leaving settings to default and

changing primer product size between 150 and 350 bp, minimum GC content of 30 %, minimum primer length of 20 bp, melting temperature between 52 and 62 °C, and a maximum melting temperature difference between primers of 4 °C. To identify single nucleotide polymorphisms (SNPs), the GATK Best Practices workflow for SNP and indel calling on RNAseq data was used (Auwera *et al.*, 2013). This pipeline handles splice junctions by mapping each sample to the transcriptome using STAR aligner in two steps (Engstrom *et al.*, 2013). In the first pass, splice junctions are identified, and this information is used to guide the final alignment. The SAM files produced were processed through Picard by adding read group information, sorting the reads, and marking duplicates before indexing. After this step, the SplitNCigarReads tool was used to split reads into exon segments (removing Ns but maintaining group information) and hard-clipping any sequence overhanging into the intronic region that may cause false variant calls. Finally, call variant sites were performed using the HaplotypeCaller tool and filtered based on Fisher Strand values (FS > 30) and quality by depth values (QD < 2.0) (McKenna *et al.*, 2010; DePristo *et al.*, 2011; Auwera *et al.*, 2013).

Finally, I performed F_{ST} outlier analysis using Bayescan 2.1 (Foll & Gaggiotti, 2008) to identify SNPs that show greater divergence than expected under neutrality between Mexican and Salvadorian populations of *P. elegans*. The analysis was performed three times with 50,000 iterations using default settings. Outliers were identified using a false discovery rate (FDR) threshold of 0.05 based on SNPs shared by all individuals per population and a quality score of 30. Outlier SNPs were annotated against the SwissProt Uniprot database using significantly more stringent criteria (e-value < 1E-20) with Trinotate.

RESULTS

Sequencing, quality filtering and de novo transcriptome assembly

The number of reads obtained per library and summary statistics for the transcriptome assembly was similar to other *de novo* crustacean transcriptome projects (Jung *et al.*, 2011; Zeng *et al.*, 2011; Ning *et al.*, 2013; Lenz *et al.*, 2014). A total of approximately 123 million paired 101 bp reads were obtained from Illumina Hiseq 2000 sequencing, totaling over 70.14 GB of data. On average, 16,540,624 pair reads were obtained from each library with no over-representation of a particular sample in the raw sequence dataset (**Table 4.1**). The assembly yielded 237,038 contig transcripts and a GC content of 52.98%, with an N50 length of 785 bases and with 28,669 contigs greater than 1000 bp (**Fig. 4.2**).

Assessment of the assembly quality was performed by gathering summary statistics at 0.64 reads per kilobase of transcripts per million mapped reads (FPKM). This method removes the effect of low expressed genes that are usually not considered in down-stream analyses. Overall, the number of genes and transcripts was similar across samples but not per population. On average 31,720 genes and 39,168 transcripts could be detected in samples collected in El Salvador (**Table 4.2**), and 65,156 genes and 72,495 transcripts for samples collected in Mexico (**Table 4.2**)**.** Similarly, N50 length of transcript and genes was higher in samples collected in Mexico (1,451 bp and 962 bp, respectively) than samples collected in El Salvador (1,750 bp and 1,197 bp, respectively) (**Table 4.2**).

Gene annotation and differential gene expression analyses

Successful annotation strongly depends on the quality of the assembly. Annotation success of each transcript and unicontig varied depending on the database used, ranging from

3,970 successful annotations using the Signal Peptides database (1.6% of total transcripts) to 78,034 annotations based on the Uniprot database (32% of total transcripts). Moreover, I found 1,734 proteins that each match a trinity transcript by > 80% and 90% of their protein lengths; 5,718 represent nearly full-transcripts having > 80% alignment coverage and 3,984 are covered by more than 90% of their protein lengths suggesting assembly quality was appropriate. A summary of annotation hits of each isoform is present in **Table 4.3**.

SNPs, SSR and Outlier F_{ST} identification

MSDB identified 29,216 putative simple sequence repeats (SSR) distributed across the trinity contigs. Mono- and trinucleotide SSRs were the most frequent motifs representing approximately 66 and 27% of the total SSRs, respectively (**Table 4.4**). The hexanucleotide motif was the least frequent representing 1.7 percent of the total. Of the total SSRs, approximately 76% (22,270) primer pairs were designed using PRIMER3 (**Appendix B**). The GATK pipeline identified an average of 160,227 SNPs and 7,895 indels per sample. The proportion of homozygotes and heterozygotes, including the number of private SNPs was similar across all samples (**Table 4.5**). Filtering of SNPs based on a quality score of at least 30 and requiring SNPs to be shared by at least 4 individuals per population reduced the initial SNP dataset to a subset of 44,578 loci for F_{ST} outlier analysis.

Outlier test for selected vs. neutral variation using BAYESCAN revealed 11 putative outliers that showed outstanding genetic differentiation when compared with the SNPs distribution expected for loci under neutrality. Divergence was driven by spatially varying selection with an alpha value ranging from 1.97 to 2 (positive values of alpha indicate diversifying selection, while negative values correspond to balancing or purifying selection) with

an average 0.33, and an average F_{ST} of 0.06 for all other loci (**Fig. 4.3**). Completely fixed loci were not detected across samples. Annotation of the loci under selection against the Swisprot Unigene database (E-value <1e-20) identified seven of the loci as follows: (1) RNA polymerase II associated protein; (2) Alpha/beta hydrolase domain containing protein 14B; (3) Serine/threonine-protein kinase 40; (4) cytochrome c oxidase subunit 3; (5) Aminopeptidase N, protein trapped in endoderm-1; (6) protein trapped in endoderm; and (6) Dynein heavy chain cytoplasmic. Annotation based on the Pfam database was positive in four genes as follows: (1) TPR 11 (PF13414.1), which form scaffolds to mediate protein–protein interactions and often the assembly of multiprotein complexes (Lamb *et al.*, 1995) (2) Abhydrolase 5 (PF12695.2), the alpha/beta hydrolase fold includes proteases, lipases, peroxidases, esterases, epoxide hydrolases and dehalogenases (Nardini & Dijkstra, 1999), (3) Peptidase M1 (PF01433.15) a family of aminopeptidases which members differ widely in specificity, hydrolysing acidic, basic or neutral N-terminal residues (Rawlings & Barrett, 1995); and (4) 7TM GPCR Srsx; (PF10320.4), that correspond to members of the seven-transmembrane G-protein-coupled receptor class that mediates chemoreception in *Caenorhabditis elegans* (Troemel *et al.*, 1995).

DISCUSSION

The implementation of RNA-seq technology has enabled the generation of genomic resources quickly and efficiently for multiple non-model species (Ekblom & Galindo, 2011). For two incipient species of barnacle in the eastern Pacific, I sequenced the transcriptome from transcribed mRNA from four individuals per population. The *de novo* assembly resulted in the identification of a total of 237,038 transcripts with N50 length size of 785 bp and approximately

one third of the unicontigs annotated according to their homology matches against public databases. Additional transcriptome mining for molecular markers yielded a total of 29,216 SSRs that adds to recently developed microsatellite markers (Plough & Marko, 2014). The large number of annotated genes and molecular markers developed for *P. elegans* establishes the base for future investigations on functional genomics in this group.

Although it is difficult to predict *a priori* the number of sequences needed to detect all transcripts [because transcript expression varies depending on treatment or tissue (Wang *et al.*, 2009)], I can estimate coverage of the *P. elegans* transcriptome based on an average of 16.5 million reads per individual (Table 4.1). The amount of DNA contained within a haploid nucleus (C-value) is not known for *P. elegans*, but it is known for the congeneric species *P. polymerus* (Sowerby, 1833), which has a C-value of 0.9 pg (Bachmann & Rheinsmi.El, 1973), which translates into an estimated genome size of more than 880 Mb (conversion factor 1 pg=978 Mb) (Dolezel *et al.*, 2003). Assuming the genome size of *P. polymerus* is similar to *P. elegans,* the Illumina reads represent a sequencing coverage of approximately 20-fold per sample and 160-fold when all samples are considered together.

Evidence of spatially varying selection in P. elegans

The 45,578 high quality SNPs used for outlier F_{ST} identified at least 11 genes under diversifying selection. Six of the 11 genes were found to have annotation information. The RNA polymerase II associated protein forms an interface between the RNA polymerase II enzyme and chaperone/scaffolding protein and it may be required to connect regulators of protein complex formation (Jeronimo *et al.*, 2007). Similarly, the tetratricopeptide repeats correspond to structural motifs that form scaffolds to mediate protein-protein interactions and often the

assembly of multiprotein complexes (Blatch & Lassle, 1999). These two proteins may have important implications in the synthesis of proteins at different thermal optima. The Abhydrolase domain containing 5 functions as an acyltransferase for the synthesis of phosphatidic acid, the major intermediate in membrane and storage lipid biosynthesis and is a coactivator of adipocyte triglyceride lipase (Ghosh *et al.*, 2008), and the peptidase M1 plays a role in the removal of single amino acids from the amino terminus of small peptides, with the main function of proteolysis. These metabolic proteins can be associated to specific environments as the ones observed in Mexico and El Salvador. Finally, the 7TM GPCR Srsx is a main protein responsible for chemoreception including paracrine and endocrine processes. Although adult barnacles are sessile organisms, chemoreception may be important in early developmental stages, having possibly important consequences in the population fitness finding a mate (Plough & Marko, 2014) or responding to predators (Lively, 1986).

CONCLUSIONS

RNA-seq of recently separated populations of *P. elegans* allowed the characterization of the transcriptome of a gooseneck barnacle, a non-model marine invertebrate of great interest as an ecological engineer, as an artisanally harvested species, and potentially as a model for thermal adaptation in evolutionary biology. The new data generated here points to the direction of candidate genes associated to local adaptation that are associated with the early origin of cryptic diversity within the nominal *P. elegans.* The first examination of genes under selection in this species identifies 11 genes under diversifying selection with functions that include protein binding, hydrolase activity, cytochrome-c oxidase activity, metallopeptidase activity and G protein coupled receptor activity. The large set of SSR and SNP markers will help

to better understand genome-wide patterns of adaptive variation in this eastern tropical marine species.

REFERENCES

Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D. & Thibault, J. (2013) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, 11.10. 1-11.10. 33.

Bachmann, K. & Rheinsmi.El (1973) Nuclear DNA amounts in Pacific crustacea. *Chromosoma*, **43**, 225-236.

Blatch, G.L. & Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932-939.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I.G., Snelgrove, P.V.R., Morris, C.J., Gregory, R.S., Hardie, D.C., Hutchings, J.A., Ruzzante, D.E., Taggart, C.T. & Bentzen, P. (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3725-3734.

Deagle, B.E., Jones, F.C., Chan, Y.F., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 1277-1286.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. & Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491-+.

Dolezel, J., Bartos, J., Voglmayr, H. & Greilhuber, J. (2003) Nuclear DNA content and genome size of trout and human. *Cytometry Part A*, **51A**, 127-128.

Du, L., Li, Y., Zhang, X. & Yue, B. (2013) MSDB: A User-Friendly Program for Reporting Distribution and Building Databases of Microsatellites from Genome Sequences. *Journal of Heredity*, **104**, 154-157.

Ekblom, R. & Galindo, J. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1-15.

Engstrom, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Raetsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigo, R., Bertone, P. & Consortium, R. (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, **10**, 1185-+.

Foll, M. & Gaggiotti, O. (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, **180**, 977-993.

Fu, Y.X. (1996) New statistical tests of neutrality for DNA samples from a population. *Genetics*, **143**, 557-570.

Ghosh, A.K., Ramakrishnan, G., Chandramohan, C. & Rajasekharan, R. (2008) CGI-58, the causative gene for Chanarin-Dorfman syndrome, mediates acylation of lysophosphatidic acid. *Journal of Biological Chemistry*, **283**, 24525-24533.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-U130.

Hudson, R.R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R., Blanchette, M., Price, D.H. & Coulombe, B. (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Molecular Cell*, **27**, 262-274.

Jung, H., Lyons, R.E., Dinh, H., Hurwood, D.A., McWilliam, S. & Mather, P.B. (2011) Transcriptomics of a Giant Freshwater Prawn (Macrobrachium rosenbergii): De Novo Assembly, Annotation and Marker Discovery. *Plos One*, **6**

Lamb, J.R., Tugendreich, S. & Hieter, P. (1995) Tetratrico Peptide Repeat interactions - to TPR or not to TPR. *Trends in Biochemical Sciences*, **20**, 257-259.

Lenz, P.H., Roncalli, V., Hassett, R.P., Wu, L.-S., Cieslak, M.C., Hartline, D.K. & Christie, A.E. (2014) De Novo Assembly of a Transcriptome for Calanus finmarchicus (Crustacea, Copepoda) - The Dominant Zooplankter of the North Atlantic Ocean. *Plos One*, **9**

Lesson, R. (1831) Histoire naturelle des Mollusques, Annélides et Vers recueillis dans le voyage autour du monde de la corvette de sa majesté, la "Coquille," exécuté pendant les années 1822, 23, 24 et 1825, sous le commandement du capitaine Duperrey. *Paris, Zoologie*, **2**, 25-471.

Li, Y.C., Korol, A.B., Fahima, T. & Nevo, E. (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*, **21**, 991-1007.

Lively, C.M. (1986) Predator-induced shell dimorphism in the acorn barnacle Chthamalus anisopoma *Evolution*, **40**, 232-242.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.

Nardini, M. & Dijkstra, B.W. (1999) alpha/beta hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology*, **9**, 732-737.

Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641-647.

Ning, J., Wang, M., Li, C. & Sun, S. (2013) Transcriptome Sequencing and De Novo Analysis of the Copepod Calanus sinicus Using 454 GS FLX. *Plos One*, **8**

Plough, L.V. & Marko, P.B. (2014) Characterization of Microsatellite Loci and Repeat Density in the Gooseneck Barnacle, Pollicipes elegans, Using Next Generation Sequencing. *Journal of Heredity*, **105**, 136-142.

Podnar, J., Deiderick, H., Huerta, G. & Hunicke-Smith, S. (2014) Next-Generation Sequencing RNA-Seq Library Construction. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, **106**, 4.21.1-4.21.19.

Rawlings, N.D. & Barrett, A.J. (1995) Evolutionary families of metallopeptidases. *Proteolytic Enzymes: Aspartic and Metallo Peptidases*, **248**, 183-228.

Renaut, S., Owens, G.L. & Rieseberg, L.H. (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, **23**, 311-324.

Schoville, S.D., Barreto, F.S., Moy, G.W., Wolff, A. & Burton, R.S. (2012) Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod Tigriopus californicus. *Bmc Evolutionary Biology*, **12**

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P. & Slate, J. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705-712.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A. & Bargmann, C.I. (1995) Divergent 7 transmembrane receptors are candidate chemosensory receptors in c-elegans. *Cell*, **83**, 207-218.

Udpa, N., Ronen, R., Zhou, D., Liang, J., Stobdan, T., Appenzeller, O., Yin, Y., Du, Y., Guo, L., Cao, R., Wang, Y., Jin, X., Huang, C., Jia, W., Cao, D., Guo, G., Claydon, V.E., Hainsworth, R., Gamboa, J.L., Zibenigus, M., Zenebe, G., Xue, J., Liu, S., Frazer, K.A., Li, Y., Bafna, V. & Haddad, G.G. (2014) Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. *Genome Biology*, **15**

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G. (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research*, **40**

Vasemagi, A. & Primmer, C.R. (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.

Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57-63.

Watterson, G.A. (1977) Heterosis or neutrality. *Genetics*, **85**, 789-814.

Zeng, V., Villanueva, K.E., Ewen-Campen, B.S., Alwes, F., Browne, W.E. & Extavour, C.G. (2011) De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. *Bmc Genomics*, **12**

Table 4.1: Number reads obtained from each sample using the Illumina HiSeq 2000 sequencing

Sample	Lane	Raw (pairs only)	Filtered by Trimmomatic	Mapped back to assembly (including single reads)
SAL1	L2	6,803,109	62,287	10,251,606
SAL1	L3	7,223,422	61,881	11,001,014
SAL2	L2	7,910,351	60,251	11,982,464
SAL2	L3	8,380,552	60,135	12,817,833
SAL3	L2	7,917,840	111,545	12,261,648
SAL3	L3	8,359,602	109,435	13,087,004
SAL4	L2	6,069,706	49,773	9,244,766
SAL4	L3	6,492,316	49,616	9,978,107
MEX1	L2	8,082,711	52,485	10,646,425
MEX1	L3	8,483,837	52,362	11,292,090
MEX2	L2	10,690,972	72,138	14,021,284
MEX2	L3	11,113,636	71,772	14,731,143
MEX3	L2	8,620,630	71,382	11,294,628
MEX3	L3	9,024,471	69,798	11,970,951
MEX4	L2	8,410,217	71,874	9,969,600
MEX4	L3	8,741,622	71,059	10,500,721
Total		132,324,994	1,097,793	185,051,284

Table 4.2: Counts of longest isoform (unicontig) and transcripts (alternative splicing isoforms) with associated N50 length at FPKM = 0.64.

	Unicontig		Transcripts	
Sample	N50	Entries	N50	Entries
SAL1	1,325	39,542	1,650	47,466
SAL2	1,484	28,863	1,787	36,078
SAL3	1,456	31,296	1,751	39,321
SAL4	1,539	27,179	1,810	33,808
MEX1	969	64,260	1,200	71,320
MEX2	1,096	48,814	1,255	52,833
MEX3	899	73,075	1,179	82,415
MEX4	882	74,474	1,152	83,410

Table 4.3: Annotation summary based on transcripts assembled with Trinity.

Database	Number of annotations	Percent of total
Proteins (Uniprot)	78,034	31.52
Protein families (Pfam)	37,533	15.16
Signal Peptides (SignalP4.1)	3,970	1.6
Transmembrane domains (TMHMM)	8,121	3.28
eggNOG	53,303	21.53
GO	77,533	31.32

Table 4.4: Summary information of simple sequence repeats (SSRs) identified using MSDB

Nucleotide	Total Counts	Total Length(bp)	Average Length(bp)	Frequency(%)	Density(bp/Mb)
mononucleotide	17,009	293,028	17.23	58.22	2,253
dinucleotide	1,949	33,826	17.36	6.67	260
trinucleotide	8,070	135,180	16.75	27.62	1,040
tetranucleotide	854	15,720	18.41	2.92	121
pentanucleotide	835	19,145	22.93	2.86	147
hexanucleotide	499	14,712	29.48	1.71	113

Table 4.5: Summary information of single nucleotide polymorphism (SNP) identified using the GATK pipeline for RNA-seq. Hom AA, homozygous for a single alternate allele (e.g. both alleles have the same mutation); Hom RR, homozygous reference; Het AA, both alleles are non-reference but they are not the same allele; Het RA, one reference allele and one alternate allele.

Sample	SNP count	indel count	private	missing	Hom AA	Hom RR	Het AA	Het RA
SAL1	149,899	6,976	58,705	342,056	62,741	287,076	417	93,658
SAL2	160,493	7,557	67,176	312,922	66,643	305,038	448	100,897
SAL3	175,537	8,236	81,218	299,182	64,114	303,049	470	119,133
SAL4	137,506	6,401	50,725	357,167	60,321	284,923	397	83,140
MEX1	156,175	8,266	59,942	340,164	65,355	281,402	413	98,614
MEX2	142,731	7,572	54,129	373,331	58,100	262,374	410	91,733
MEX3	186,865	9,386	77,400	284,930	68,633	304,833	512	127,040
MEX4	172,612	8,764	73,968	317,240	67,370	287,401	511	113,426

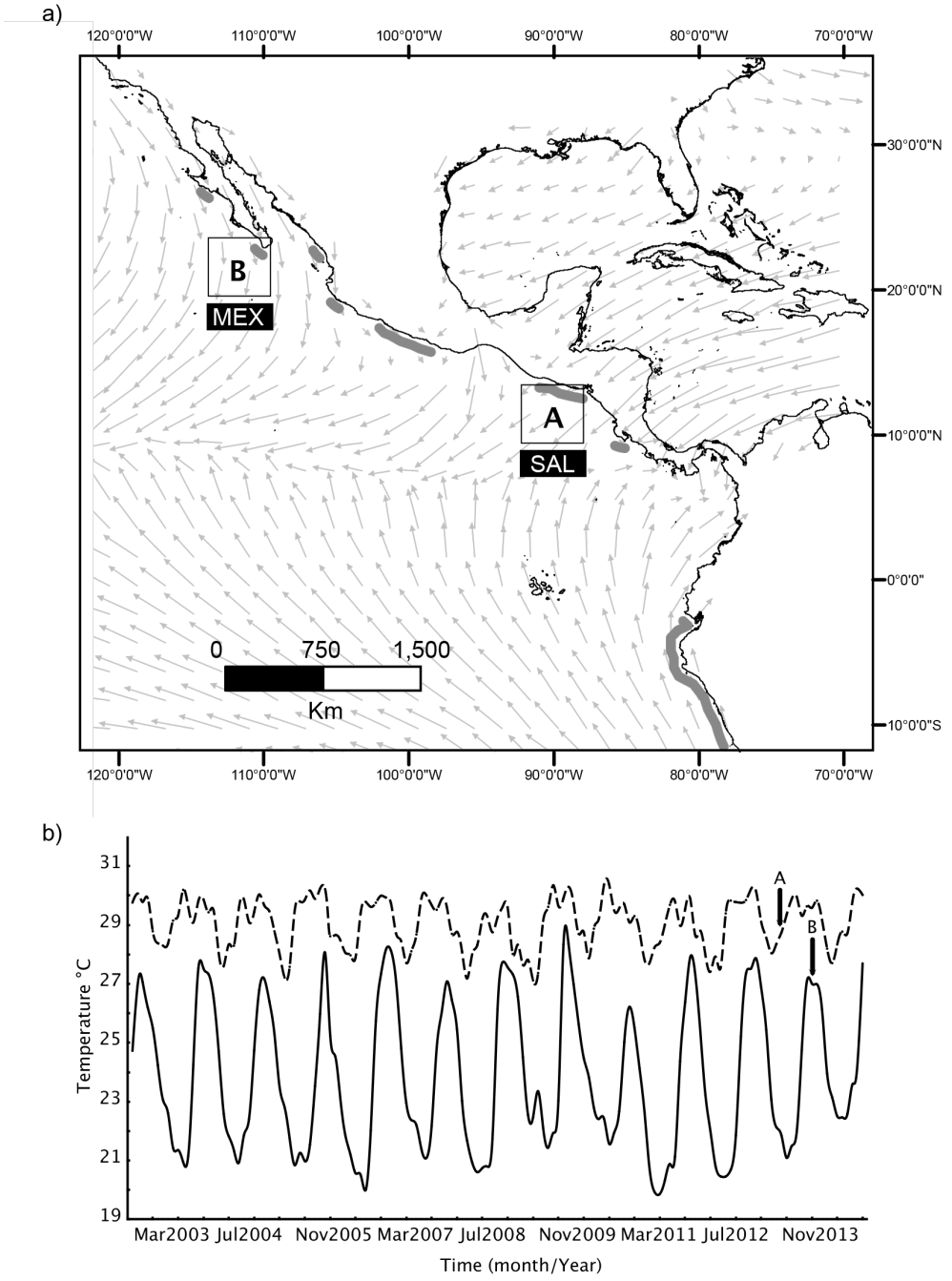

Figure 4.1: Map of the eastern tropical Pacific region showing distribution of *Pollicipes elegans*, wind stress, collections sites and sea surface temperature profiles. a) Grey blobs represent the known distribution of *P. elegans* after (Chapter 2). Grey arrows represent wind stress climatology data after (Pennington, *et al*. 2006); b) Profiles of average sea surface temperature from July 2002 to July 2014 at the collection sites. Arrows represent time of the collection.

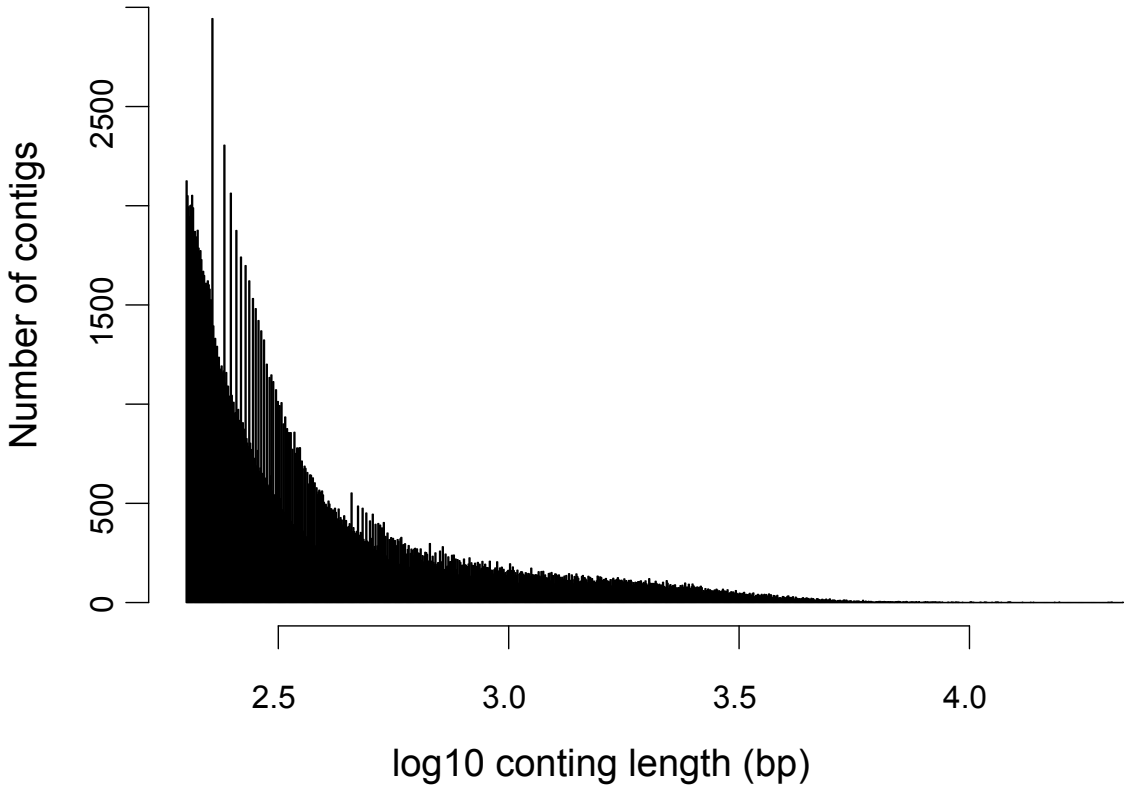

Figure 4.2: Distribution of transcript size for *Pollicipes elegans* transcriptome data

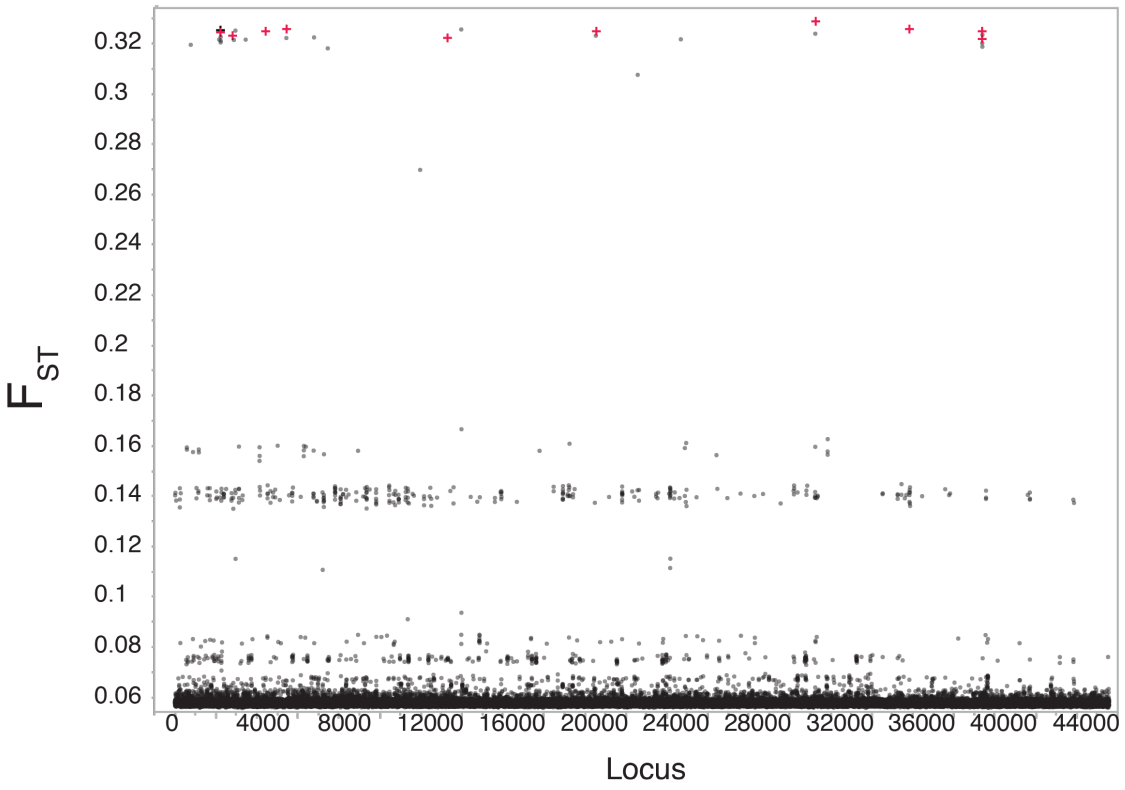

Figure 4.3: Outlier F_{ST} analysis. The 11 outlier SNPs identified in the samples are shown in red.

CHAPTER FIVE

CONCLUSIONS AND FUTURE DIRECTIONS

The goal of this dissertation was to examine the potential roles of spatial isolation, sea surface temperature change, and natural selection on the process of speciation in the tropical eastern Pacific, aiming to evaluate how these factors may contribute to the tropical diversity engine. To address this complex question, I used the marine barnacle *Pollicipes elegans* as a model to characterize the process of diversification at the interface between tropical and extra-tropical regions. First, I reconstructed the demographic history of the species to distinguish between trans-tropical dispersal and the formation of peripheral isolates by vicariance, to understand the biogeographic significance of *P. elegans* distribution. Second, with evidence of a very strong differentiation – but not complete lineage sorting of mtDNA haplotypes – I used multilocus data to test for the presence of cryptic species within *P. elegans*. Both methods of species delimitation that I used provided strong support for speciation having taken place between tropical and extra-tropical populations within the nominal *P. elegans*. Finally, I used transcriptome sequencing to scan the genome for genes that show evidence of natural selection, and that could potentially be associated with spatial and temporal temperature variation, aiming to evaluate their role in local adaptation associated with the process of speciation.

The historical biogeography of *P. elegans* (Chapter 2) provides strong evidence that the disjunct distribution of this species represents three main populations, an oldest in the center of the distribution range (El Salvador populations) and two others composed by peripheral populations at the edge of extra-tropical regions to the north (Mexico) and south (Peru) of the warm eastern tropical Pacific convergence. A large contribution of my thesis was distinguishing between dispersal and vicariance as explanations for the disjunct distribution of *P. elegans* in the eastern Pacific. By using Approximate Bayesian Computation I found support for a tropical origin of the peripheral (extra-tropical) populations of *P. elegans* in Mexico and Peru, products of two vicariant events during the Pleistocene, one that separated populations from Mexico and El Salvador approximately 350 ka ago and a second more recent event approximately 100 ka ago, that separated populations of Peru from El Salvador. Since Mexican and Peruvian populations have a common tropical origin, the historical biogeography of *P. elegans* fits the expectations of the tropical diversity engine. Moreover, migration estimates based on the mitochondrial data suggest no evidence of gene flow across the thermal gradients to the north and south of the tropical eastern Pacific, what provides ideal conditions for allopatric speciation.

In chapter 3 I tested this hypothesis by using multilocus data and species delimitation methods and found strong support for two cryptic species within *P. elegans*. One composed of populations in Mexico and, a second species composed of all other populations to the south, from El Salvador to Peru. Divergence time and migration

estimates including the additional nuclear genes showed agreement with Chapter 2 in terms of divergence time between regions, but was different for the migration estimates, showing low but significant signatures of gene flow between and within cryptic species. Even thought the mtDNA showed no evidence of gene flow, these results are not surprising given that mtDNA with a haploid genome may have lost signatures of gene flow by genetic drift. The low and limited gene flow observed between central and peripheral populations was similar in both peripheral populations, however, the length of time Mexican populations have been isolated with limited gene flow is much longer than the Peruvian peripheral populations, and likely promoted that cryptic diversity could be detect only in the north peripheral range of the species distribution.

Finally, in Chapter 4 I performed a genomic survey and identified eleven genes under spatially divergent selection between the two species. The genes under selection code for proteins that have functions that include protein binding, hydrolase activity, cytochrome-c oxidase activity, metallopeptidase activity and G protein coupled receptor activity. These candidate genes may represent loci that are involved in local adaption between tropical and extra-tropical regions in *P. elegans*. Considering the scenario of restricted gene flow and prolonged isolation of these cryptic species, environmental drivers that promote genomic patterns of speciation could be related with temperature and need further attention. The information generated in this research hopes to

contribute reliable models that can predict the effects of temperature in species diversity, as expected today with the pace of human-induced climate change.

In answer to the overall objective posed at the beginning of my dissertation, the current distribution of *P. elegans* is likely the result of a combination of historical and contemporaneous effects, however divergence timing across regions strongly supports that temperature change during Pleistocene is the main cause of differentiation causing speciation at the transition zone between tropical and extra-tropical populations in the northern hemisphere, and that local adaptation product of spatially diversifying selection have contribute to the isolation of the cryptic species found within *P. elegans*.

FUTURE DIRECTIONS

Understanding the interaction between natural selection and gene flow is an exciting subject to potentially address in the future with *P. elegans*. In my dissertation, I used transcriptome sequencing to detect signatures of natural selection, and provided candidate genes that could be ecologically relevant with respect to spatial variation in sea surface temperature. To better understand if the mechanisms of local adaptation are indeed associated with temperature, future studies could potentially perform laboratory enzyme assays under controlled conditions as a proxy to evaluate thermal adaptation in response to thermal stress. It also could be interesting to see if similar patterns of local adaptation are also observed at Peruvian populations, given that environmental conditions in Peru are similar to the ones observed in Mexico and the divergence time is more recent (convergent evolution). A second area of research that

could be investigated is where in the life history of *P. elegans* – larvae, newly settled juveniles, or adults - the candidate genes are affected by selection. In the future, common garden experiments designed to target the effects of temperature in the genomic response and fitness across the developmental stages of *P. elegans* is a good alternative. By characterizing signatures of divergent natural selection through the life history of *P. elegans*, it may be would be possible to understand how natural selection affects the fitness of *P. elegans* and its relevance for the process of speciation.

APPENDICES

Appendix A

Supporting information Chapter 2

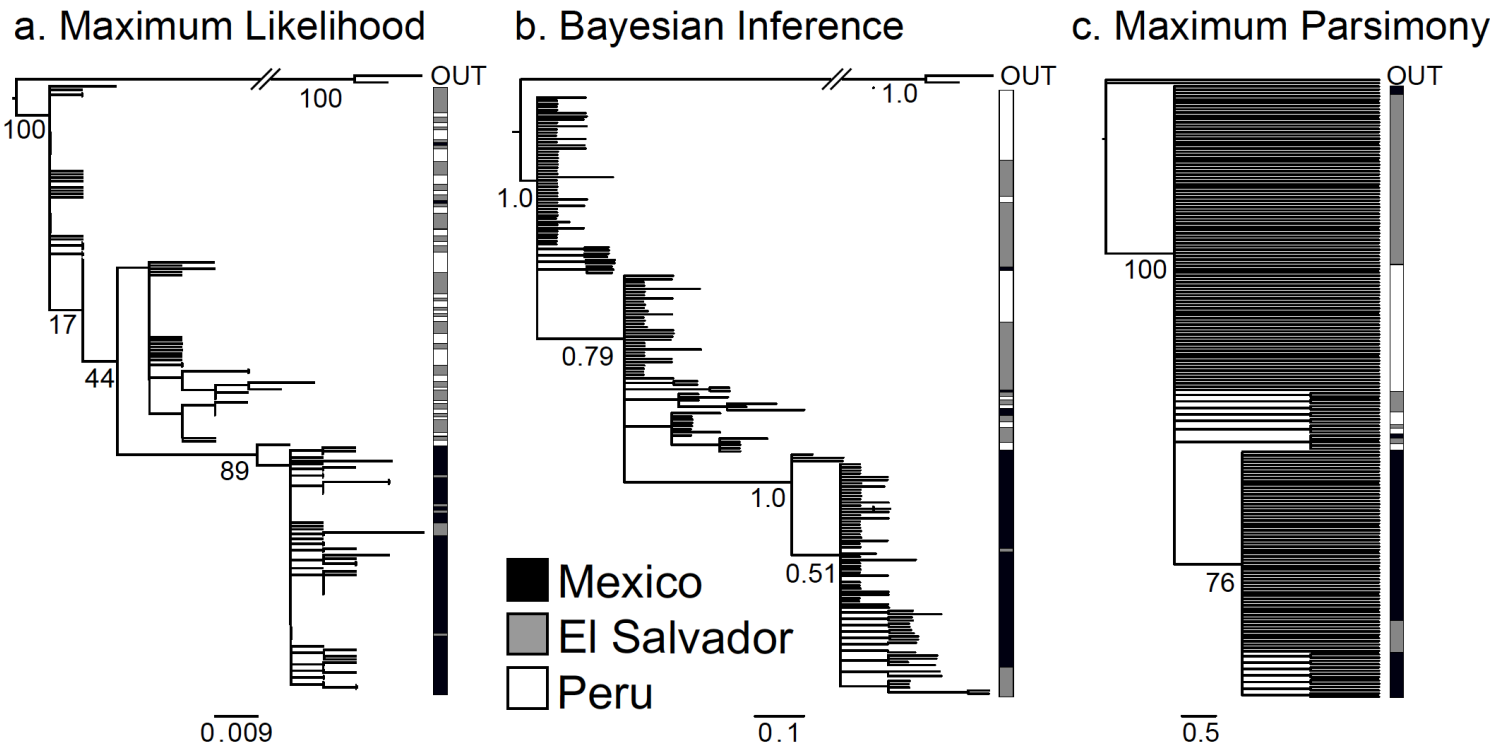

Figure A-1: a. Maximum likelihood tree; b. Bayesian inference tree; c. Maximum Parsimony tree. Tree inference was generated from partial sequences of the mitochondrial cytochrome oxidase 1 (COI) nucleotide sequence. Confidence support above the main nodes correspond to posterior probability for the Bayesian inference and bootstrap support for the likelihood and Parsimony inference. OUT represents sequences of *P. pollicipes* and *P. caboverdensis* used as outgroups in the analyses. Sample origin is color coded as indicated in the figure legend.

Table A-1: Likelihood ratio test (LLRtest) for migration rate parameters *m*, and population migration *2NM* between pairs of populations of *P. elegans*.

Population pairs[§]	*m*	LLRtest	*2NM*	LLRtest
PER1->PER2	0.02	0.00	0.00	0.00
PER1<-PER2	0.00	0.00	0.00	0.00
MEX1->PER1	0.00	0.00	0.00	0.00
MEX1<-PER1	0.00	0.00	0.00	0.00
MEX1->PER2	0.02	0.20	0.55	0.39
MEX1<-PER2	0.00	0.00	0.00	0.00
MEX1->SAL3	0.00	0.00	0.00	0.00
MEX1<-SAL3	0.00	0.00	0.00	0.00
MEX1->MEX3	0.00	0.00	0.00	0.00
MEX1<-MEX3	0.52	1.53[#]	15.31	1.29[#]
MEX1->MEX2	0.00	0.00	0.00	0.00
MEX1<-MEX2	0.00	0.00	0.00	0.00
MEX1->SAL1	0.00	0.00	0.00	0.00
MEX1<-SAL1	0.00	0.00	0.00	0.00
MEX1->SAL2	0.00	0.00	0.00	0.00
MEX1<-SAL2	0.00	0.00	0.00	0.00
SAL3->PER1	0.00	0.00	0.00	0.00
SAL3<-PER1	0.00	0.00	0.00	0.00
SAL3->PER2	0.28	0.18	0.00	0.00
SAL3<-PER2	0.00	0.00	0.00	0.00
MEX3->PER1	0.04	0.08	2.53	0.24
MEX3<-PER1	0.00	0.00	0.00	0.00
MEX3->PER2	0.07	7.10**	5.23	7.05**
MEX3<-PER2	0.00	0.00	0.00	0.00
MEX3->SAL3	0.09	2.97*	6.72	2.89*
MEX3<-SAL3	0.00	0.00	0.00	0.00
MEX3->SAL1	0.06	0.13	1.19	0.11
MEX3<-SAL1	0.00	0.00	0.00	0.00
MEX3->SAL2	0.00	0.00	0.00	0.00
MEX3<-SAL2	0.00	0.00	0.00	0.00
MEX2->PER1	0.04	0.58	0.53	0.95
MEX2<-PER1	0.00	0.00	0.00	0.00
MEX2->PER2	0.06	2.32[#]	0.75	3.47*
MEX2<-PER2	0.00	0.00	0.00	0.00
MEX2->SAL3	0.02	0.04	0.44	0.13

MEX2<-SAL3	0.00	0.00	0.00	0.00
MEX2->MEX3	0.00	0.00	0.00	0.00
MEX2<-MEX3	0.29	0.27	0.00	0.00
MEX2->SAL1	0.00	0.00	0.00	0.00
MEX2<-SAL1	0.00	0.00	0.00	0.00
MEX2->SAL2	0.00	0.00	0.00	0.00
MEX2<-SAL2	0.00	0.00	0.00	0.00
SAL1->PER1	0.00	0.00	0.00	0.00
SAL1<-PER1	0.00	0.00	0.00	0.00
SAL1->PER2	0.57	0.99	20.00	1.97[#]
SAL1<-PER2	0.00	0.00	0.08	0.02
SAL1->SAL3	0.00	0.00	1.47	0.07
SAL1<-SAL3	0.00	0.00	0.17	0.01
SAL1->SAL2	0.00	0.00	0.00	0.00
SAL1<-SAL2	0.00	0.00	0.00	0.00
SAL2->PER1	0.00	0.00	0.00	0.00
SAL2<-PER1	0.00	0.00	0.00	0.00
SAL2->PER2	0.00	0.00	0.00	0.00
SAL2<-PER2	0.00	0.00	0.00	0.00
SAL2->SAL3	0.00	0.00	0.00	0.00
SAL2<-SAL3	0.00	0.00	0.00	0.00

Significant values are shown in boldface.
ç, è, denote directionality of gene flow backwards in time.
*P < 0.05, **P < 0.01; [#] inconclusive results. LLRtest= Log-likelihood ratio test. None of the results were significant after Bonferroni multiple test correction (56 comparisons P=0.00089). [§] Sampling site names as in Table 2.1.

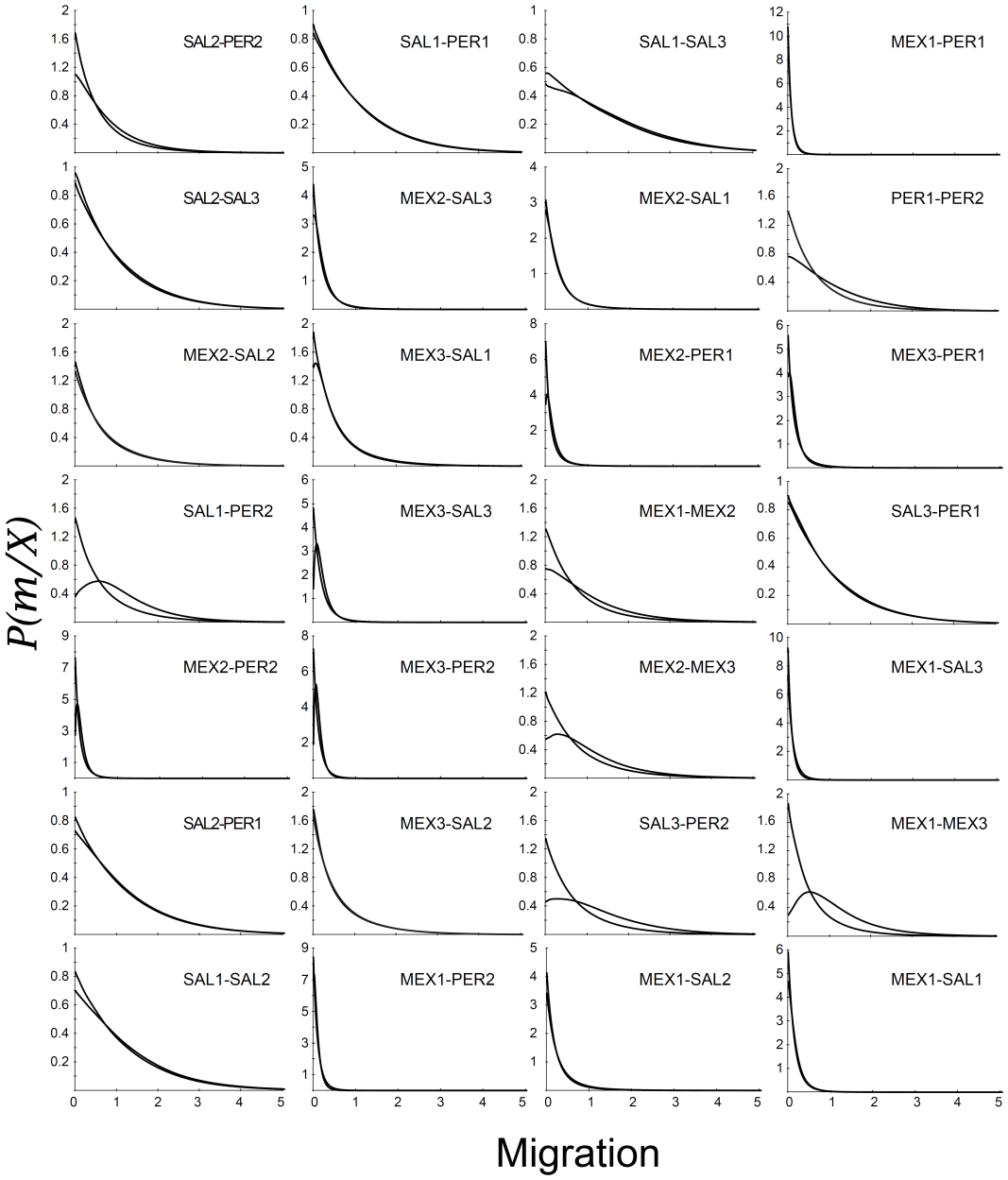

Figure A-2: Joint posterior density plots of parameter m between populations obtained from IMa2 analyses based on data from COI locus of *Pollicipes elegans*.

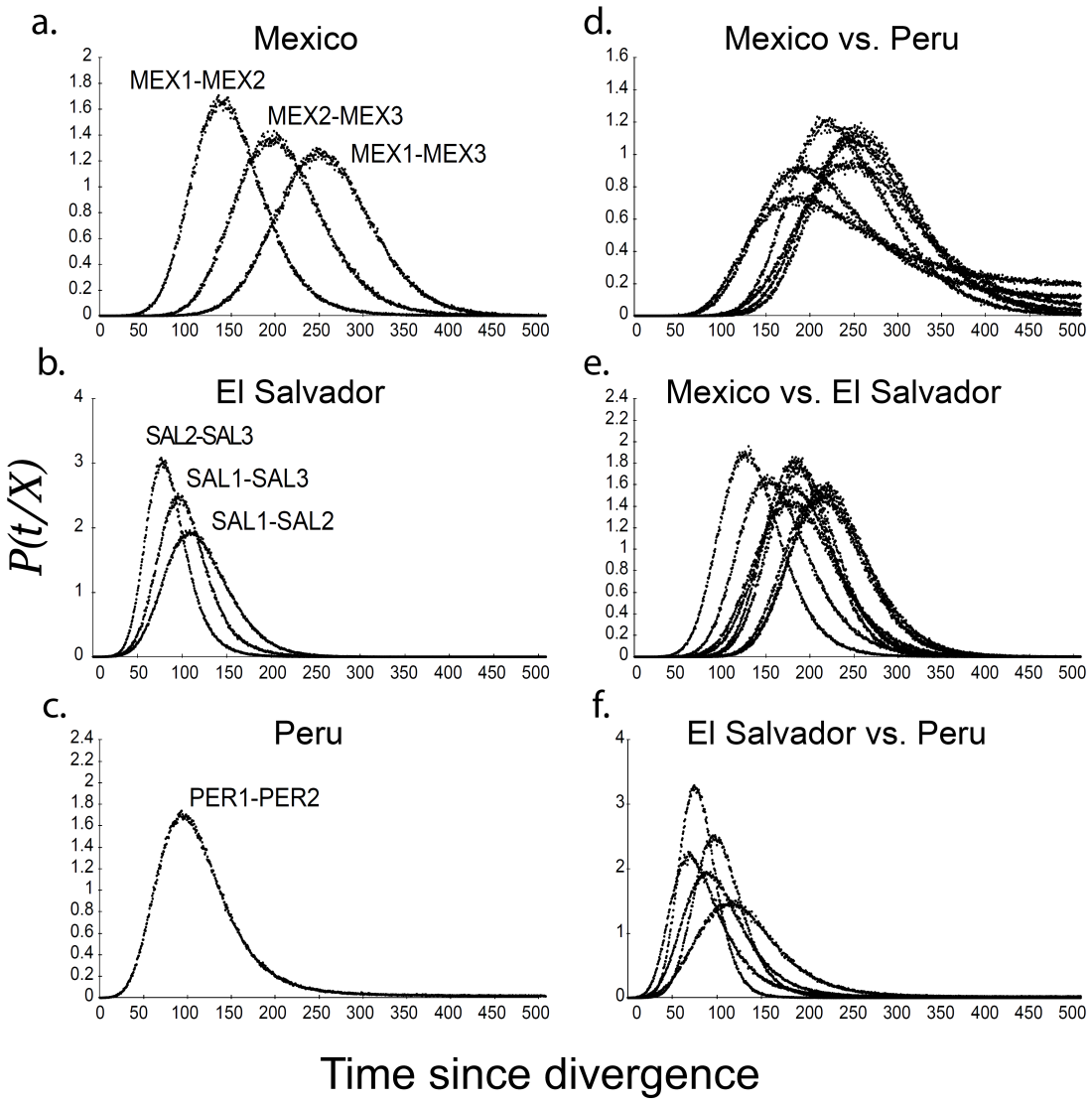

Figure A-3 :Joint posterior density plots obtained from IMa2 analyses based on data from COI locus of *Pollicipes elegans*. Divergence time between pairs of populations is presented among and within regions.

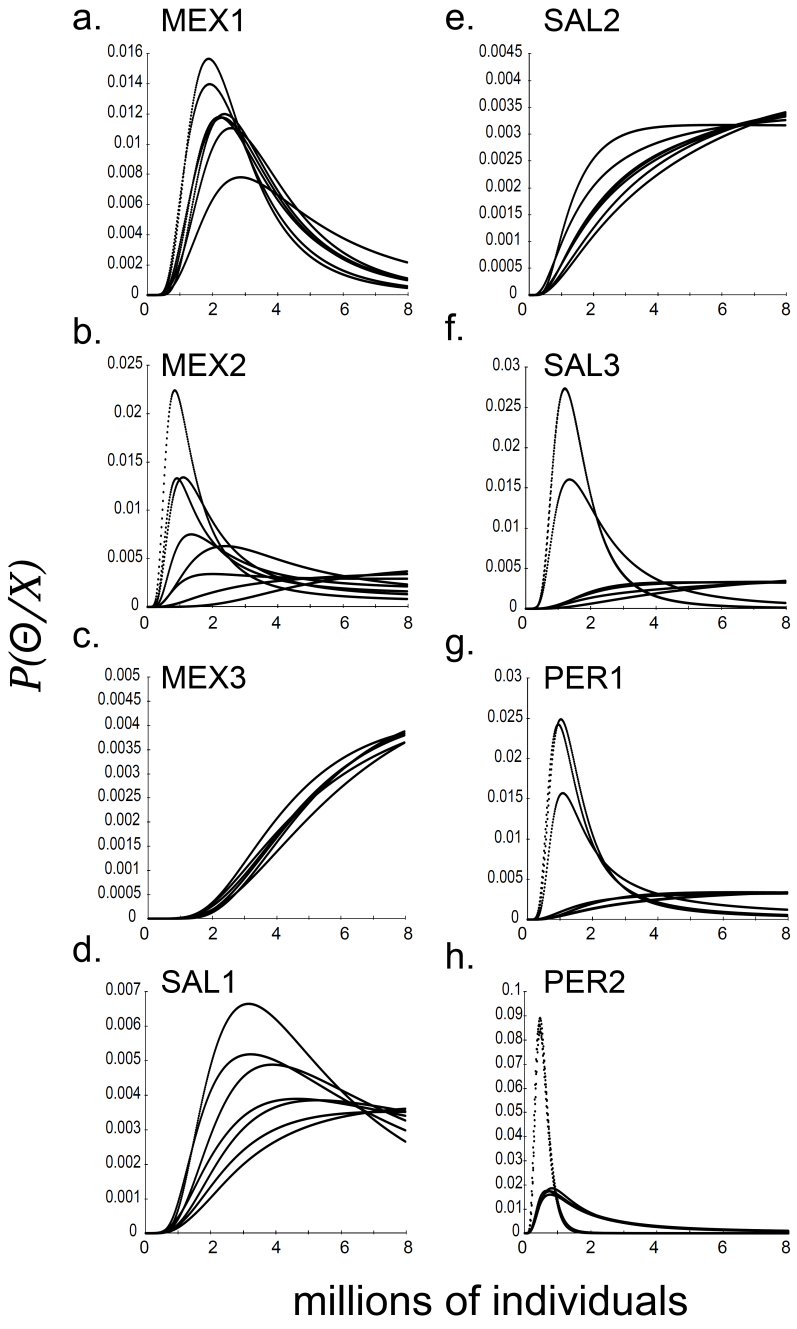

Figure A-4: Joint posterior density plots of millions of individuals obtained from IMa2 analyses based on data from COI locus of *Pollicipes elegans*. Each line in the plot corresponds to a pairwise comparison between populations.

Appendix B

References of combined Chapters

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. (1994) Molecular biology of the cell. *Garland, New York*, 139-194.

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697-709.

Anderson, C.N.K., Ramakrishnan, U., Chan, Y.L. & Hadly, E.A. (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733-1734.

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

Arntz, W.E., Gallardo, V.A., GutiÃ©rrez, D., Isla, E., Levin, L.A., Mendo, J., Neira, C., Rowe, G.T., Tarazona, J. & Wolff, M. (2006) El Niño and similar perturbation effects on the benthos of the Humboldt, California, and Benguela Current upwelling ecosystems. *Advances in geosciences*, **6**, 243-265.

Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D. & Thibault, J. (2013) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, 11.10. 1-11.10. 33.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press.

Bachmann, K. & Rheinsmi.El (1973) Nuclear DNA amounts in Pacific crustacea. *Chromosoma*, **43**, 225-236.

Barber, P.H. & Bellwood, D.R. (2005) Biodiversity hotspots: evolutionary origins of biodiversity in wrasses (Halichoeres : Labridae) in the Indo-Pacific and new world tropics. *Molecular Phylogenetics and Evolution*, **35**, 235-253.

Barnes, M. (1996) Pedunculate Cirripedes of the genus Pollicipes. *Oceanography and Marine Biology, Vol 34*, **34**, 303-394.

Baum, D.A. & Shaw, K.L. (1995) Genealogical perspectives on the species problem. *Experimental and Molecular Approaches to Plant Biosystematics*, **53**, 289-303.

Beaumont, M.A., Zhang, W.Y. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025-2035.

Bernardo, J. (2011) A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. *Systematics Association Special Volume Series*, **78**, 380-438.

Birky, C.W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation, and drift-equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, **121**, 613-627.

Blatch, G.L. & Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932-939.
Blum, M.G.B. & Francois, O. (2010) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, **20**, 63-73.
Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, **10**, e1003537.
Bowen, B.W. & Grant, W.S. (1997) Phylogeography of the sardines (Sardinops spp): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution*, **51**, 1601-1610.
Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I.G., Snelgrove, P.V.R., Morris, C.J., Gregory, R.S., Hardie, D.C., Hutchings, J.A., Ruzzante, D.E., Taggart, C.T. & Bentzen, P. (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3725-3734.
Briggs, J.C. (1974) *Marine zoogeography*. McGraw-Hill, New York.
Briggs, J.C. (1987) Antitropicality and vicariance. *Systematic Biology*, **36**, 206-207.
Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
Burridge, C.P. (2002) Antitropicality of Pacific fishes: molecular insights. *Environmental Biology of Fishes*, **65**, 151-164.
Camargo, A., Morando, M., Avila, L.J. & Sites, J.W. (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834-2849.
Carbajal, W., Castañeda, J., Castro, J., De la Cruz, J., Ramírez, P., Quevedo, M. & Zeballos, C. (1998) Informe del Seguimiento de las Pesquerías Artesanales. In: *Informe Progressivo*, p. 28. Instituto del Mar de Perú (IMARPE), Chiclayo, Perú.
Carlquist, S. (1981) Chance Dispersal. *American Scientist*, **69**, 509-516.
Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.
Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, **17**, 540-552.
Chown, S.L. & Gaston, K.J. (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology & Evolution*, **15**, 311-315.
Cowen, R.K. & Sponaugle, S. (2009) Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science*, **1**, 443-466.
Cowman, P.F. & Bellwood, D.R. (2013) Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of

hard versus soft barriers. *Proceedings of the Royal Society B-Biological Sciences*, **280**

Crisp, M.D., Isagi, Y., Kato, Y., Cook, L.G. & Bowman, D.M.J.S. (2010) Livistona palms in Australia: Ancient relics or opportunistic immigrants? *Molecular Phylogenetics and Evolution*, **54**, 512-523.

Cronin, T.W. & Forward, R.B. (1986) Vertical Migration Cycles of Crab Larvae and Their Role in Larval Dispersal. *Bulletin of Marine Science*, **39**, 192-201.

Cruz, T. & Araujo, J. (1999) Reproductive patterns of Pollicipes pollicipes (Cirripedia : Scalpellomorpha) on the southwestern coast of Portugal. *Journal of Crustacean Biology*, **19**, 260-267.

Cruz, T., Castro, J.J. & Hawkins, S.J. (2010) Recruitment, growth and population size structure of Pollicipes pollicipes in SW Portugal. *Journal of experimental marine biology and ecology*, **392**, 200-209.

Csillery, K., Francois, O. & Blum, M.G.B. (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475-479.

Csillery, K., Blum, M.G.B., Gaggiotti, O.E. & Francois, O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410-418.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life: Murray*. London.

de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F. & Beaufort, L. (2005) Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature*, **433**, 294-298.

de Queiroz, A. (2005a) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution*, **20**, 68-73.

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

De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic biology*, **56**, 879-886.

Deagle, B.E., Jones, F.C., Chan, Y.F., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 1277-1286.

Defeo, O., Castrejón, M., Ortega, L., Kuhn, A.M., Gutiérrez, N.L. & Castilla, J.C. (2013) Impacts of Climate Variability on Latin American Small-scale Fisheries. *Ecology and Society*, **18**, 30.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332-340.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. & Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491-+.

Diaz-Perez, A.J., Sequeira, M., Santos-Guerra, A. & Catalan, P. (2012) Divergence and biogeography of the recently evolved Macaronesian red Festuca (Gramineae) species inferred from coalescence-based analyses. *Molecular ecology*, **21**, 1702-1726.
Dolezel, J., Bartos, J., Voglmayr, H. & Greilhuber, J. (2003) Nuclear DNA content and genome size of trout and human. *Cytometry Part A*, **51A**, 127-128.
Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology*, **7**, 214.
Du, L., Li, Y., Zhang, X. & Yue, B. (2013) MSDB: A User-Friendly Program for Reporting Distribution and Building Databases of Microsatellites from Genome Sequences. *Journal of Heredity*, **104**, 154-157.
Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular ecology*, **11**, 2571-2581.
Ekblom, R. & Galindo, J. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1-15.
Ekman, S. (1953) *Zoogeography of the sea*. Sidgwick and Jackson, London,.
Ence, D.D. & Carstens, B.C. (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473-480.
Engstrom, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Raetsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigo, R., Bertone, P. & Consortium, R. (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, **10**, 1185-+.
Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.
Eytan, R.I. & Hellberg, M.E. (2010) Nuclear and Mitochondrial Sequence Data Reveal and Conceal Different Demographic Histories and Population Genetic Processes in Caribbean Reef Fishes. *Evolution; international journal of organic evolution*, **64**, 3380-97.
Fischer, A.G. (1960) Latitudinal variations in organic diversity. *Evolution*, **14**, 64-81.
Floeter, S., Rocha, L., Robertson, D., Joyeux, J., Smith-Vaniz, W., Wirtz, P., Edwards, A., Barreiros, J.P., Ferreira, C. & Gasparini, J.L. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22-47.
Flot, J.F. (2010) SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162-166.
Foll, M. & Gaggiotti, O. (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, **180**, 977-993.
Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, **3**, 294-299.

Fouquet, A., Cassini, C.S., Baptista Haddad, C.F., Pech, N. & Rodrigues, M.T. (2014) Species delimitation, patterns of diversification and historical biogeography of the Neotropical frog genus Adenomera (Anura, Leptodactylidae). *Journal of Biogeography*, **41**, 855-870.

Fu, Y.X. (1996) New statistical tests of neutrality for DNA samples from a population. *Genetics*, **143**, 557-570.

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

Geyer, C.J. (1991) Markov-Chain Monte-Carlo Maximum-Likelihood. *Computing Science and Statistics*, 156-163.

Ghosh, A.K., Ramakrishnan, G., Chandramohan, C. & Rajasekharan, R. (2008) CGI-58, the causative gene for Chanarin-Dorfman syndrome, mediates acylation of lysophosphatidic acid. *Journal of Biological Chemistry*, **283**, 24525-24533.

Giarla, T.C. (2013) *Systematics, Biogeography, and Phylogeography of Thylamys Mouse Opossums, a Recent Radiation of Neotropical Marsupials*. UNIVERSITY OF MINNESOTA,

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-U130.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* (ed by, pp. 95-98.

Hastings, P.A. (2000) Biogeography of the Tropical Eastern Pacific: distribution and phylogeny of chaenopsid fishes. *Zoological Journal of the Linnean Society*, **128**, 319-335.

Hawkins, B.A. & Porter, E.E. (2001) Area and the latitudinal diversity gradient for terrestrial birds. *Ecology Letters*, **4**, 595-601.

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

Heled, J. & Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570-580.

Hellberg, M.E. (1998) Sympatric sea shells along the sea's shore: The geography of speciation in the marine gastropod Tegula. *Evolution*, **52**, 1311-1324.

Hemmer-Hansen, J., Nielsen, E.E., Frydenberg, J. & Loeschcke, V. (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (Platichthys flesus L.). *Heredity*, **99**, 592-600.

Hey, J. (2010) Isolation with Migration Models for More Than Two Populations. *Molecular biology and evolution*, **27**, 905-920.

Hey, J. & Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D-persimilis. *Genetics*, **167**, 747-760.
Hey, J. & Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785-2790.
Hey, J. & Pinho, C. (2011) Population genetics and objectivity in species diagnosis. *Evolution*,
Hickerson, M.J. & Cunningham, C.W. (2005) Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (Xiphister) revealed by multilocus DNA analysis. *Evolution*, **59**, 344-360.
Hillebrand, H. (2004a) On the generality of the latitudinal diversity gradient. *American Naturalist*, **163**, 192-211.
Hillebrand, H. (2004b) Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series*, **273**, 251-267.
Hudson, R.R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.
Hunn, C.A. & Upchurch, P. (2001) The importance of time/space in diagnosing the causality of phylogenetic events: Towards a "chronobiogeographical" paradigm? *Systematic Biology*, **50**, 391-407.
Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, **23**, 254-267.
Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, **93**, 145-159.
Hutchison, D.W. & Templeton, A.R. (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898-1914.
Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102-106.
Jablonski, D., Belanger, C.L., Berke, S.K., Huang, S., Krug, A.Z., Roy, K., Tomasovych, A. & Valentine, J.W. (2013) Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10487-10494.
Jarman, S.N., Ward, R.D. & Elliott, N.G. (2002) Oligonucleotide primers for PCR amplification of coelomate introns. *Marine Biotechnology*, **4**, 347-355.
Jensen, J.L., Bohonak, A.J. & Kelley, S.T. (2005) Isolation by distance, web service. *Bmc Genetics*, **6**, 13.
Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R., Blanchette, M., Price, D.H. & Coulombe, B. (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Molecular Cell*, **27**, 262-274.

Jung, H., Lyons, R.E., Dinh, H., Hurwood, D.A., McWilliam, S. & Mather, P.B. (2011) Transcriptomics of a Giant Freshwater Prawn (Macrobrachium rosenbergii): De Novo Assembly, Annotation and Marker Discovery. *Plos One*, **6**

Kameya, A. & Zeballos, J. (1988) Distribucion y densidad de percebes Pollicipes elegans (Crustacea: Cirripedia) en el mediolitoral peruano (Yasila, Piura; Chilca, Lima). *Boletin-Instituto del Mar del Peru*, **12**, 5-22.

Katoh, K., Kuma, K.-i., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, **33**, 511-518.

Keever, C.C., Sunday, J., Puritz, J.B., Addison, J.A., Toonen, R.J., Grosberg, R.K. & Hart, M.W. (2009) Discordant Distribution of Populations and Genetic Variation in a Sea Star with High Dispersal Potential. *Evolution*, **63**, 3214-3227.

Knowles, L.L. & Carstens, B.C. (2007) Delimiting species without monophyletic gene trees. *Systematic biology*, **56**, 887-895.

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

Knowlton, N. & Weigt, L.A. (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 2257-2263.

Knowlton, N., Weigt, L.A., Solorzano, L.A., Mills, D.K. & Bermingham, E. (1993) Divergence in proteins, mitochondrial-DNA, and reproductive compatibilty across the isthmus of Panama. *Science*, **260**, 1629-1632.

Laguna, J.E. (1990) Shore Barnacles (Cirripedia, Thoracica) and a Revision of Their Provincialism and Transition Zones in the Tropical Eastern Pacific. *Bulletin of Marine Science*, **46**, 406-424.

Lamb, J.R., Tugendreich, S. & Hieter, P. (1995) Tetratrico Peptide Repeat interactions - to TPR or not to TPR. *Trends in Biochemical Sciences*, **20**, 257-259.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 314-334.

Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, **29**, 1695-1701.

Leaché, A.D. & Fujita, M.K. (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B: Biological Sciences*, rspb20100662.

Lenz, P.H., Roncalli, V., Hassett, R.P., Wu, L.-S., Cieslak, M.C., Hartline, D.K. & Christie, A.E. (2014) De Novo Assembly of a Transcriptome for Calanus finmarchicus (Crustacea, Copepoda) - The Dominant Zooplankter of the North Atlantic Ocean. *Plos One*, **9**

Lessios, H.A. (2008) The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. *Annual Review of Ecology Evolution and Systematics*, **39**, 63-91.

Lessios, H.A., Kessing, B.D. & Pearse, J.S. (2001) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin Diadema. *Evolution*, **55**, 955-975.
Lesson, R. (1831) Histoire naturelle des Mollusques, Annélides et Vers recueillis dans le voyage autour du monde de la corvette de sa majesté, la "Coquille," exécuté pendant les années 1822, 23, 24 et 1825, sous le commandement du capitaine Duperrey. *Paris, Zoologie*, **2**, 25-471.
Li, Y.C., Korol, A.B., Fahima, T. & Nevo, E. (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*, **21**, 991-1007.
Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.
Lindberg, D.R. (1991) Marine Biotic Interchange between the Northern and Southern Hemispheres. *Paleobiology*, **17**, 308-324.
Liu, L., Yu, L.L., Pearl, D.K. & Edwards, S.V. (2009) Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology*, **58**, 468-477.
Liu, Z.H. & Herbert, T.D. (2004) High-latitude influence on the eastern equatorial Pacific climate in the early Pleistocene epoch. *Nature*, **427**, 720-723.
Lively, C.M. (1986) Predator-induced shell dimorphism in the acorn barnacle Chthamalus anisopoma *Evolution*, **40**, 232-242.
Lowe, W.H. & Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038-3051.
Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.
Maddison, W.P. & Knowles, L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21-30.
Marko, P.B. (1998) Historical allopatry and the biogeography of speciation in the prosobranch snail genus Nucella. *Evolution*, **52**, 757-774.
Marko, P.B. (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular ecology*, **13**, 597-611.
Marko, P.B., Hoffman, J.M., Emme, S.A., McGovern, T.M., Keever, C.C. & Cox, L.N. (2010) The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular ecology*, **19**, 146-169.
Marlow, J.R., Lange, C.B., Wefer, G. & Rosell-Mele, A. (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science*, **290**, 2288-+.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.
Mcmillan, W.O. & Palumbi, S.R. (1995) Concordant Evolutionary Patterns among Indo-West Pacific Butterflyfishes. *Proceedings of the Royal Society B-Biological Sciences*, **260**, 229-236.

Miner, B.G. (2002) Are the two physiological races of Pollicipes polymerus (Cirripedia) genetically divided along the California coast? *Invertebrate Biology*, **121**, 158-162.
Minin, V.N., Bloomquist, E.W. & Suchard, M.A. (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular biology and evolution*, **25**, 1459-1471.
Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M. & Turelli, M. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315-331.
Moritz, C. (1994) Defining evolutionarily-significant-units for conservation. *Trends in Ecology & Evolution*, **9**, 373-375.
Nardini, M. & Dijkstra, B.W. (1999) alpha/beta hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology*, **9**, 732-737.
Newman, W. (1992a) Biotic cognates of eastern boundary conditions in the Pacific and Atlantic: relicts of Tethys and climate change. *Proceedings of the San Diego Society of Natural History* (ed by, pp. 1-7.
Newman, W.A. (1992b) Biotic Cognates of Eastern Boundary Conditions in the Pacific and Atlantic: Relicts of Tethys and Climatic Change. *Proceedings of the San Diego Society of Natural History*, **16**, 1-7.
Newman, W.A. & Killingley, J.S. (1985) The Northeast Pacific Intertidal Barnacle Pollicipes-Polymerus in India - a Biogeographical Enigma Elucidated by O-18 Fractionation in Barnacle Calcite. *Journal of Natural History*, **19**, 1191-1196.
Newman, W.A. & Foster, B.A. (1987) Southern-Hemisphere Endemism among the Barnacles - Explained in Part by Extinction of Northern Members of Amphitropical Taxa. *Bulletin of Marine Science*, **41**, 361-377.
Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641-647.
Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885.
Ning, J., Wang, M., Li, C. & Sun, S. (2013) Transcriptome Sequencing and De Novo Analysis of the Copepod Calanus sinicus Using 454 GS FLX. *Plos One*, **8**
Nylander, J. (2004) MrModeltest v2. *Evolutionary Biology Centre, Uppsala University*,
Palsboll, P.J., Berube, M. & Allendorf, F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution*, **22**, 11-16.
Palumbi, S.R. (1994) Genetic-Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics*, **25**, 547-572.
Pamilo, P. & Nei, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568-583.
Pampoulie, C., Ruzzante, D.E., Chosson, V., Jorundsdottir, T.D., Taylor, L., Thorsteinsson, V., Danielsdottir, A.K. & Marteinsdottir, G. (2006) The genetic

structure of Atlantic cod (Gadus morhua) around Iceland: insight from microsatellites, the Pan I locus, and tagging experiments. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2660-2674.
Pan, D. & Rubin, G.M. (1997) Kuzbanian controls proteolytic processing of Notch and mediates lateral inhibition during Drosophila and vertebrate neurogenesis. *Cell*, **90**, 271-280.
Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.
Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, **37**, 637-669.
Paulay, G. (1990) Effects of Late Cenozoic Sea-Level Fluctuations on the Bivalve Faunas of Tropical Oceanic Islands. *Paleobiology*, **16**, 415-434.
Pennington, J.T., Mahoney, K.L., Kuwahara, V.S., Kolber, D.D., Calienes, R. & Chavez, F.P. (2006) Primary production in the eastern tropical Pacific: A review. *Progress in Oceanography*, **69**, 285-317.
Phillips, B.F. & Mcwilliam, P.S. (1986) The Pelagic Phase of Spiny Lobster Development. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2153-2163.
Pianka, E.R. (1966) Latitudinal gradients in species diversity: a review of concepts. *American Naturalist*, **100**, 33-46.
Plough, L.V. & Marko, P.B. (2014) Characterization of Microsatellite Loci and Repeat Density in the Gooseneck Barnacle, Pollicipes elegans, Using Next Generation Sequencing. *Journal of Heredity*, **105**, 136-142.
Podnar, J., Deiderick, H., Huerta, G. & Hunicke-Smith, S. (2014) Next-Generation Sequencing RNA-Seq Library Construction. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, **106**, 4.21.1-4.21.19.
Pulquerio, M.J.F. & Nichols, R.A. (2007) Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution*, **22**, 180-184.
Quinteiro, J., Rodríguez-Castro, J., López, P., López-Jurado, L.F., González-Henríquez, N. & Rey-Méndez, M. (2011) Morphological and molecular characterization of a new species of Atlantic stalked barnacle (Scalpelliformes: Pollicipedidae) from the Cape Verde Islands. *Journal of the Marine Biological Association of the United Kingdom*, **91**, 1537-1545.
Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, **19**, 2092-2100.
Rannala, B. & Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645-1656.
Rawlings, N.D. & Barrett, A.J. (1995) Evolutionary families of metallopeptidases. *Proteolytic Enzymes: Aspartic and Metallo Peptidases*, **248**, 183-228.
Renaut, S., Owens, G.L. & Rieseberg, L.H. (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, **23**, 311-324.

Richmond, R. (1990) The effects of the El Niño/Southern Oscillation on the dispersal of corals and other marine organisms. *Global ecological consequences of the 1982-83 el niño-southern oscillation* (ed. by P.W. Glynn), pp. 127-140. Elsevier, Amsterdam, The Netherlands.
Rincon-Martinez, D., Lamy, F., Contreras, S., Leduc, G., Bard, E., Saukel, C., Blanz, T., Mackensen, A. & Tiedemann, R. (2010) More humid interglacials in Ecuador during the past 500 kyr linked to latitudinal shifts of the equatorial front and the Intertropical Convergence Zone in the eastern tropical Pacific. *Paleoceanography*, **25**
Robertson, D.R. & Cramer, K.L. (2009) Shore fishes and biogeographic subdivisions of the Tropical Eastern Pacific. *Marine Ecology Progress Series*, **380**, 1-17.
Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.
Rosenberg, N.A. (2003) The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465-1477.
Rosenblatt, R. (1967) The zoogeographic relationships of the marine shore fishes of tropical America. **5**, 579-592.
Rosenblatt, R.H. & Walker, B.W. (1962) The marine shore-fishes of the Galapagos Islands. *California Academy of Sciences, Occasional Papers*, **44**, 97-106.
Rosenzweig, M.L. (1992) Species diversity gradients: we know more and less than we thought. *Journal of mammalogy*, **73**, 715-730.
Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
Roy, K., Jablonski, D. & Valentine, J.W. (1994) Eastern Pacific molluscan provinces and latitudinal diversity gradient: no evidence for" Rapoport's rule". *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 8871.
Ryder, O.A. (1986) Species conservation and systematics - the dilemma of subspecies. *Trends in Ecology & Evolution*, **1**, 9-10.
Sanford, E., Roth, M.S., Johns, G.C., Wares, J.P. & Somero, G.N. (2003) Local selection and latitudinal variation in a marine predator-prey interaction. *Science*, **300**, 1135-1137.
Schoville, S.D., Barreto, F.S., Moy, G.W., Wolff, A. & Burton, R.S. (2012) Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod Tigriopus californicus. *Bmc Evolutionary Biology*, **12**
Schubart, C.D., Diesel, R. & Hedges, S.B. (1998) Rapid evolution to terrestrial life in Jamaican crabs. *Nature*, **393**, 363-365.
Simpson, G.G. (1951) The species concept. *Evolution*, **5**, 285-298.
Simpson, G.G. (1964) Species density of North American recent mammals. *Systematic zoology*, **13**, 57-73.

Sites Jr, J.W. & Marshall, J.C. (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, **18**, 462-470.
Sites, J.W. & Marshall, J.C. (2004) Operational criteria for delimiting species. *Annual Review of Ecology Evolution and Systematics*, **35**, 199-227.
Smouse, P.E., Long, J.C. & Sokal, R.R. (1986) Multiple-Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic zoology*, **35**, 627-632.
Springer, V.G. (1959) Systematics and zoogeography of the ciclidae fishes of the subtribe Labrisomini Hubbs. *Institute of Marine Science Publications*, **5**, 417-492.
Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, **57**, 758-771.
Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P. & Slate, J. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705-712.
Stebbins, G.L. (1974) Flowering plants: evolution above the species level. *London: Arnold xviii, 399p..Illustrations.General (KR, 197500089)*,
Stephens Jr, J.S. (1963) A revised classification of the blennioid fishes of the American family Chaenopsidae. **68**, 1-165.
Stephens, M. & Donnelly, P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162-1169.
Stephens, M. & Scheet, P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449-462.
Stevens, G.C. (1989) The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, **133**, 240-256.
Stevens, G.C. (1992) The elevational gradient in altitudinal range: an extension of Rapoport's latitudinal rule to altitude. *The American Naturalist*, **140**, 893-911.
Swanson, W.J. & Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, **3**, 137-144.
Swofford, D.L. (2003) PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, **10**, 512-526.
Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, **17**, 57-86.
Terborgh, J. (1973) On the notion of favorableness in plant ecology. *The American Naturalist*, **107**, 481-501.

Thompson, J.D., Gibson, T. & Higgins, D.G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2.3. 1-2.3. 22.
Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A. & Bargmann, C.I. (1995) Divergent 7 transmembrane receptors are candidate chemosensory receptors in c-elegans. *Cell*, **83**, 207-218.
Udpa, N., Ronen, R., Zhou, D., Liang, J., Stobdan, T., Appenzeller, O., Yin, Y., Du, Y., Guo, L., Cao, R., Wang, Y., Jin, X., Huang, C., Jia, W., Cao, D., Guo, G., Claydon, V.E., Hainsworth, R., Gamboa, J.L., Zibenigus, M., Zenebe, G., Xue, J., Liu, S., Frazer, K.A., Li, Y., Bafna, V. & Haddad, G.G. (2014) Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. *Genome Biology*, **15**
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G. (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research*, **40**
Valentine, J.W. & Jablonski, D. (1983) Larval Adaptations and Patterns of Brachiopod Diversity in Space and Time. *Evolution*, **37**, 1052-1061.
Van Syoc, R.I. (1994) Genetic divergence between subpopulations of the eastern Pacific goose barnacle Pollicipes elegans: mitochondrial cytochrome c subunit 1 nucleotide sequences. *Molecular marine biology and biotechnology*, **3**, 338-346.
Van Syoc, R.J., Fernandes, J.N., Carrison, D.A. & Grosberg, R.K. (2010) Molecular phylogenetics and biogeography of Pollicipes (Crustacea: Cirripedia), a Tethyan relict. *Journal of experimental marine biology and ecology*, **392**, 193-199.
Vasemagi, A. & Primmer, C.R. (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.
von Humbolt, A. (1808) *Ansichten der Natur mit wissenschaftlichen Erlauterungen Tubingen*. J. G. Cotta, Tubingen, Germany.
Walker, B.W. (1960) The distribution and affinities of the marine fish fauna of the Gulf of California. *Systematic Zoology*, **9**, 123-133.
Wallace, A.R. (1878) *Tropical nature, and other essays*. Macmillan and co., London.
Walther, K., Crickenberger, S.E., Marchant, S., Marko, P.B. & Moran, A.L. (2013) Thermal tolerance of larvae of Pollicipes elegans, a marine species with an antitropical distribution. *Marine Biology*, **160**, 2723-2732.
Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57-63.
Waples, R.S. (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438-450.
Wares, J.P. (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular ecology*, **11**, 1131-1144.
Watterson, G.A. (1977) Heterosis or neutrality. *Genetics*, **85**, 789-814.

Weir, J.T. & Schluter, D. (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574.
Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639-644.
Wiens, J.J., Graham, C.H., Moen, D.S., Smith, S.A. & Reeder, T.W. (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *American Naturalist*, **168**, 579-596.
Wiley, E.O. (1988) Vicariance Biogeography. *Annual Review of Ecology and Systematics*, **19**, 513-542.
Williams, S.T. & Reid, D.G. (2004) Speciation and diversity on tropical rocky shores: A global phylogeny of snails of the genus Echinolittorina. *Evolution*, **58**, 2227-2251.
Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273-309.
Woerner, A.E., Cox, M.P. & Hammer, M.F. (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851-1853.
Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212-226.
Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264-9269.
Yule, G.U. (1925) A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 21-87.
Zeng, V., Villanueva, K.E., Ewen-Campen, B.S., Alwes, F., Browne, W.E. & Extavour, C.G. (2011) De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. *Bmc Genomics*, **12**
Zwickl, D. (2006) *GARLI, vers. 0.951. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Ph. D. dissertation, University of Texas, Austin, Texas, USA,