

**Selection and Demographic History Shape the
Molecular Evolution of the Gamete Compatibility
Protein Bindin in *Pisaster* Sea Stars**

by

Iva Popovic

B.Sc. (Hons.), Simon Fraser University, 2010

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Biological Sciences
Faculty of Science

© Iva Popovic 2013

SIMON FRASER UNIVERSITY

Fall 2013

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Iva Popovic

Degree: Master of Science (Biological Sciences)

Title of Thesis: *Selection and demographic history shape the molecular evolution of the gamete compatibility protein bindin in Pisaster sea stars*

Examining Committee: Chair: Francis Law
Professor

Michael W. Hart
Senior Supervisor
Professor

Felix Breden
Supervisor
Professor

Isabelle M. Côté
Internal Examiner
Professor

Date Defended/Approved: November 25, 2013

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files ("Work") (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU's own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU's rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author's written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author's knowledge, infringe upon anyone's copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013

Abstract

Reproductive compatibility proteins have been shown to evolve rapidly under positive selection leading to reproductive isolation, despite the potential homogenizing effects of gene flow. I characterize the gene that encodes the gamete compatibility protein, *bindin*, for three broadcast spawning sea star species in the genus *Pisaster*, in a species-level comparison of *bindin* gene structure and molecular evolution. I then use phylogeographic patterns and variation in life history characteristics between *P. ochraceus* and *P. brevispinus* to test predictions about selection acting on *bindin* divergence among conspecific populations. I discover that divergence in the repetitive *bindin* domain structure may be partly influenced by concerted evolution within species. I find modest evidence of positive selection acting on *P. ochraceus* *bindin* alleles that is consistent with sexual conflict favoring selection for intraspecific *bindin* polymorphism, and can be explained in a demographic context of recent population expansions, coupled with the homogenizing effects of concerted evolution.

Keywords: *bindin*; gamete recognition; positive selection; sexual conflict; concerted evolution

To my loving grandmother, with all of her fingers crossed.

Acknowledgements

This work would not have been possible without the generous help of my collaborators and colleagues at Simon Fraser. This research was funded by the Natural Sciences and Engineering Research Council of Canada and by Simon Fraser University. I would like to give a big thanks to the FAB* Lab group for being a constant source of insight and inspiration, particularly Ben Sandkam, Will Stein and Kristen Westfall. This project would not have been possible without sea star samples provided by Siobhan Gray at the Bamfield Marine Sciences Centre and Bob Van Syoc at the California Academy of Sciences. Thank you to Felix Breden for his help throughout this process as a committee member and Isabelle Côté for her assistance with sample collection and for her interest and insight as the external examiner. John Wares and Peter Marko provided helpful feedback and advice on writing and data analysis as collaborators. I would also like to thank all of the Faculty and Staff of the Department of Biological Sciences for their support and assistance throughout my experience at Simon Fraser. Throughout my time at the Hart Lab, Carson Keever, Jenn Sunday and Susana Patiño have provided consistent support, advice, insight and friendship, without which this project would not have been possible. I would like to give a big thank you to Mike Hart not only for his constant mentorship, careful feedback, creativity and insistence on quality work, but also for taking me on as an undergraduate and introducing me to the wonderful world of marine invertebrates and evolution in the ocean. Finally, thanks are owed to my loving family for their continued love and support and to Will Atlas for his enthusiasm, endless support and patience through the many months of work on this thesis.

Table of Contents

Approval.....	ii
Partial Copyright Licence	iii
Abstract.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	viii
List of Figures	viii
Chapter 1. Overview	1
Chapter 2. Selection and Demographic History Shape the Molecular Evolution of the Gamete Compatibility Protein Bindin in <i>Pisaster</i> Sea Stars.....	6
2.1. Introduction	6
2.2. Methods	12
2.2.1. Study System	12
2.2.2. Population Sampling	13
2.2.3. Genetic Data	13
2.2.4. Sequence Alignment	16
2.2.5. Gene Genealogies	17
2.2.6. Repetitive Domain Analyses	18
2.2.7. Population Genetic Analyses	18
2.2.8. Tests of Selection.....	19
2.2.9. Coalescent Demographic Analysis.....	21
2.3. Results.....	23
2.3.1. Gene Structure.....	23
2.3.2. Repetitive Domain Analysis.....	25
2.3.3. Sequence Diversity	25
Bindin	25
Cytochrome c oxidase 1 (COI).....	27
2.3.4. Selection Analyses	27
2.3.5. Coalescent Demographic Analysis.....	28
2.4. Discussion.....	30
Chapter 3. Future Directions.....	51
References.....	55
Appendix A. Supporting Information	66

List of Tables

Table 1. Summary of collection sites and number of individuals and alleles sampled per location, per locus	39
Table 2. Primers sequences for each locus and location in gene	39
Table 3. Polymorphism statistics (including sites with gaps), number of variable nucleotide and amino acid sites, number of unique haplotypes, nucleotide diversity (π ; at individual sites), haplotype diversity (h) and number of alleles with indel variation.	40
Table 4. Summary of results from MEME and branch-sites REL tests of positive selection, implemented in HyPhy.	40
Table 5. Summary of branch-site models implemented in PAML. All codons inferred to be under selection correspond to sites within each separate species alignment.	41
Table 6. Summary of population parameter estimates inferred by IMA. ML effective population size estimates N_1 , N_2 and ancestral population size, N_A in 10^5 individuals, with confidence intervals (90% highest posterior densities, HPD) in parentheses. Divergence time in thousands of years (t). '?' indicates weakly defined upper or lower bounds on the posterior probabilities of estimated parameter values. The proportion of migrants into population1 (m_1) and population 2 (m_2) per generation.	42

List of Figures

Figure 1. Collection sites for five <i>P. ochraceus</i> and three <i>P. brevispinus</i> populations sampled in this study are shown on the map on the right. Dashed lines represent the northern and southern range boundaries for <i>P. brevispinus</i> . Solid lines represent the ranges boundaries for <i>P. ochraceus</i> . Refer to Table 1 for exact localities and sample sizes. Bindin haplotype networks for a) <i>P. ochraceus</i> and b) <i>P. brevispinus</i> are shown on the left. Each circle represents a unique haplotype; the area of the circle is proportional to the frequency of the haplotype; colors show the proportion of each haplotype from different localities shown on the map. Each line between haplotypes represents one inferred mutational step, slashes indicate an additional mutation, and numbers within lines specify the number of mutational steps greater than three. Isolated haplotypes represent alleles that differed from the rest of the sample by insertion-deletion differences and could not be connected to the rest of the haplotype network by nucleotide substitutions only.	43
--	----

- Figure 2.** a) *Pisaster ochraceus* bindin gene structure. b) Inferred amino acid sequence for the full-length *P. ochraceus* bindin; the N-terminal preprobindin sequence and the C-terminal conserved core region are in boxes and the furin-type cleavage site (RVRR) is in bold letters. c) Scaled diagram of repetitive domain structure variation in the second bindin exon for all three *Pisaster* species. Black regions represent non-repetitive terminal sequences; collagen-like regions are shown in blue; A repeats are shown in dark green; B repeats are shown in light green and are absent in *P. giganteus*. 44
- Figure 3.** a) Alignment of repeat domains in the second bindin exon for all three *Pisaster* species. b) Consensus Bayesian genealogy of repetitive domains with Bayesian posterior probabilities of partition at branch nodes. 'A' and 'B' denote separate clades grouping repeat types A and B. Circles at branch tips are colored corresponding to species (see legend in panel a) and numbered corresponding to repeat number counted in the 5' to 3' direction 45
- Figure 4.** Plots show variation in nucleotide diversity per base pair across the coding sequence for the second bindin exon for a) *P. ochraceus* and b) *P. brevispinus* populations, using the default settings (window length=100; step size=25). 47
- Figure 5.** Consensus Bayesian genealogies of bindin alleles for a) *P. ochraceus* and b) *P. brevispinus* rooted with a single *P. giganteus* bindin allele. The scale bar shows 0.4% sequence divergence. Numbers above branches represent Bayesian posterior probabilities of partition and numbers below the branches denote bootstrap support from 1000 replicates. Red branches were inferred to be experiencing episodic diversifying selection with $\omega \gg 1$, using the branch-site REL method implemented in HyPhy. Also refer to Table 4 and 5 for additional tests of positive selection 48
- Figure 6.** Posterior probability distributions of divergence times for bindin and COI in demographic units (years) between a) *P. ochraceus* population pairs; b) *P. brevispinus* population pairs. The mode of each posterior distribution indicates the maximum likelihood estimate for divergence time between each population pair. Note the wider prior distribution and older divergence times for *P. brevispinus*. 50

Chapter 1.

Overview

Speciation is the fundamental process responsible for the evolution of biodiversity. However, understanding the relationship between genetic changes and phenotypic diversity that can lead to differentially adapted populations and the formation of new species, is one of the most important challenges in explaining the speciation process (Springer et al. 2011). A recent surge of evolutionary research has focused on whole-genome scans that aim to identify the ‘genomic islands of divergence’ or genomic regions that underlie the origins of population genetic differentiation during speciation (Turner et al. 2005; Feder and Nosil 2010; Via 2012; Ellegren 2013). Likewise, experimental (e.g., Michel et al. 2010; Nosil and Feder 2012 for review) and theoretical (e.g., Tomaiuolo et al. 2007; Nosil and Flaxman 2010) studies have provided perspectives into the conditions favoring rapid evolution and adaptive divergence between populations in the absence of geographical barriers. Such work has emphasized speciation-with-gene-flow (Rice and Hostert 1993), where conspecific populations or hybridizing species diverge at specific parts of the genome in the face of ongoing dispersal and genetic exchange, which fundamentally oppose the diversifying effects of natural selection and genetic drift (Schwarz et al. 2009; Maroja et al. 2009; Smadja and Butlin 2011). Key examples have focused on divergence in reproductive traits involving differences or incompatibilities in reproductive morphology, chemosensory signaling, and mating behavior or host choice among populations or species (Sobel et al. 2010; Servedio et al. 2011; Bird et al. 2012). However, such examples most often involve model or terrestrial organisms with complex mating systems, highly sophisticated sensory signaling and behaviors that are governed by hundreds or thousands of genes and transcription factors (Hart 2012). Therefore, identifying genomic islands where differences among populations reflect the responses to selection on mating specificity that may underlie genetically based barriers to gene flow, represents a critical problem for our understanding of local adaptation and

speciation (Hart 2012). As a result, largely unresolved questions include describing the selective processes involved in generating the observed patterns of divergence (Rundle and Nosil 2005; Schluter 2009; Sobel et al. 2010), identifying which sets of genes are affected (Nosil and Schluter 2011; Lessios 2011), and resolving how specific changes in these genes mediate adaptive differences in male and female traits that can lead to reproductive barriers between conspecific populations or incipient species (Springer et al. 2011).

Despite being largely ignored by the Modern Synthesis (Love 2009; but see Mayr 1954), comparative studies focusing on organisms in the marine environment have offered some practical solutions to these evolutionary problems. Indeed, researchers have taken advantage of a rich diversity of sympatric sibling species (Marko 2005; Johannesson et al. 2010; Bird et al. 2011; Krug 2011) and mating systems (Hart et al. 2012; Purtiz et al. 2012) in the ocean to understand the evolutionary processes leading to population structure, the role of adaptive evolution in divergence and the biogeography of speciation. In particular, the diversity of broadcast spawning marine invertebrates provides at least two significant advantages to evolutionary ecologists interested in studying the complex relationship between selection and gene flow in speciation (Springer et al. 2011; Lessios 2011). First, many marine invertebrates with sessile or slow-moving adults, have a high capacity for widespread dispersal through long-distance movement of planktonic larval stages. Because strong oceanographic or geological obstacles to genetic exchange are often temporally unstable or permeable across marine seascapes and may be rare for marine animals with highly dispersive larvae, the effects of selection may have a primary role in promoting adaptive variation within species that can lead to reproductive barriers to gene flow between them (Lessios 2007). Therefore, patterns of sequence polymorphism in selectively neutral and protein-coding loci can be compared against the neutral population genetic expectations to make inferences about the demographic and selective processes that structure genetic variation and promote local adaptation (Clark et al. 2006).

Secondly, the relatively simple mating systems of broadcast spawning marine invertebrates, where individuals release great numbers of gametes into the plankton (estimated 200 billion sperm cells per single male sea urchin ejaculate; Swanson and Vacquier 2002), depend on mate recognition and fertilization specificity that is governed

primarily by the molecular interactions of recognition proteins associated exclusively with sperm and eggs (Hirohashi et al. 2008; Lessios 2011; Evans and Sherman 2013). Therefore, the responses to selective pressures involved in mating, such as sexual selection among competing sperm, sexual conflict between male and female gametes, and reinforcement against heterospecific fertilizations, are restricted to a few compatibility genes expressed on the surfaces of male and female gametes (Palumbi 2009; Lessios 2011; Vacquier and Swanson 2011). Such simple genetic systems provide an excellent framework in which it is possible to study the effects of local selection on phenotypic variation in reproductive proteins (at the level of amino acid changes) that may have direct implications for the molecular basis of prezygotic barriers to fertilization and for our understanding of the speciation process (Lessios 2011). Indeed, the earliest studies to find evidence of selection favoring the rapid divergence of reproductive proteins between closely related species have come from research on the sperm-expressed acrosomal proteins of broadcast spawning sea urchins and abalones (e.g., Vacquier et al. 1995; Metz and Palumbi 1996; Metz et al. 1998a; Biermann 1998; Swanson and Vacquier 2002; Turner and Hoekstra 2008; Vacquier and Swanson 2011). Such striking patterns of diversification have now been documented in reproductive genes of various taxonomic groups, including seminal fluid and female reproductive tract proteins in insects, seminal coagulants and sperm-egg binding proteins in copulating vertebrates, and self-incompatibility proteins and pollen coat genes in flowering plants (reviewed in Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008). The rapid evolution of reproductive genes across diverse sexually reproducing plants and animals, suggests that understanding how evolutionary processes act on simple broadcast spawning systems can have important implications for elucidating the role of selection in fertilization specificity at the level of molecules and genes, and may serve as a basis for predictions about genomic evolution during the earliest stages of reproductive isolation (Clark et al. 2006).

In this thesis, I focus on a group of ecologically important, broadcast spawning sea stars to investigate the potential roles of molecular, selective and demographic processes in shaping patterns of spatial variation and divergence in the gamete compatibility protein, *bindin*. *Bindin* is expressed in the sperm acrosomal vesicle (Vacquier and Moy 1977). Activation of the sperm is mediated through interactions with

the egg's extracellular matrix and causes exocytosis of the acrosome and exposure of its contents (including bindin) onto the acrosomal process, an extension of the sperm membrane (Vacquier et al. 1995). Bindin molecules coating the acrosomal process interact with glycoprotein receptors in the egg vitelline layer and function in the adhesion and fusion of the gamete membranes (Vacquier et al. 1995). Studies on the molecular evolution of bindin and its role in gametic specificity have largely focused on species-level comparisons among eleven genera of sea urchins, and to date, bindin has been documented in one Asteroid species, the sea star *Patiria miniata* (Patiño et al. 2009). However, there have been relatively few studies that have examined population-level adaptive bindin differences driven by positive selection within species (e.g., Geyer and Palumbi 2003; Geyer and Lessios 2009; Sunday and Hart 2013). As a result, understanding 'how' bindin differences evolve within species and the relative importance of the different selective processes implicated in the origins of bindin divergence, including the sensitivity of the rate of bindin evolution to variation among species or populations in the predicted strength of sexual conflict or reinforcement is imperfectly understood.

In this thesis, I first characterize the full-length gene encoding bindin, for all three species of the northeastern Pacific sea star genus *Pisaster*, which belongs to a relatively recent radiation within the taxonomic order Forcipulatacea (Mah and Foltz 2011). I describe bindin gene structure for *Pisaster ochraceus*, *Pisaster brevispinus* and *Pisaster giganteus* in a species-level comparison and investigate the molecular evolution of the highly repetitive bindin domain structure that is likely involved in species-specificity. I then present the results from a large-scale phylogeographic analysis of bindin spatial variation and patterns of positive selection in multiple populations of *P. ochraceus* and *P. brevispinus*. *Pisaster* species have overlapping spawning seasons and provide an excellent testing ground for distinguishing among alternative selective processes acting on bindin polymorphism because the current geographical distributions of *P. ochraceus* and *P. brevispinus* include allopatric and sympatric populations (where selection against hybridization may act on bindin) and these species are also expected to experience different forms and intensities of sexual selection and sexual conflict. Employing a combination of population genetic tools, codon model tests of selection and a series of coalescent-based population demographic analyses, I develop and test predictions

about the selective processes driving differences in the rate of binding evolution among populations and species, and consider how population demographic history has contributed to shaping the spatial distribution of binding variation among conspecific populations. This work contributes to our understanding of the interacting roles of selection, genetic drift and ongoing gene flow at loci underlying mating specificity and provides important perspectives on the molecular mechanisms and evolutionary processes that underlie local adaptation and population-level divergence in simple broadcast spawning mating systems.

Chapter 2.

Selection and Demographic History Shape the Molecular Evolution of the Gamete Compatibility Protein Bindin in *Pisaster* Sea Stars

2.1. Introduction

A long-standing goal of evolutionary biology is to understand the interacting roles of selective and demographic processes in shaping patterns of genetic diversity among populations, and how that can lead to the evolution of distinct species. However, a major source of debate has been the evolution of reproductive barriers in sympatry, where populations become genetically differentiated despite the potential homogenizing effects of gene flow (Pinho and Hey 2010; Smadja and Bultin 2011). Studies focusing on the earliest stages of lineage diversification have been critical in shedding light on the evolutionary conditions favoring divergence in sympatry (e.g., Martin and Hosken 2004; Messina and Jones 2011; Smadja and Butlin 2011 for review; Bird et al. 2012; Sunday and Hart 2013). Researchers have broadly focused on two key challenges: capturing the effects of selection on diverse loci as populations move along the “speciation continuum” and become genetically and reproductively differentiated (e.g., Nosil et al. 2009; Nosil 2012); and identifying barrier genes that maintain species boundaries among already diverged species during secondary contact (e.g., Won et al. 2005; Maroja et al. 2009; Pinho and Hey 2010). Both themes of research have made major advances in understanding the roles played by disruptive selection during ecological speciation (Schluter 2001; Rundle and Nosil 2005; Schluter 2009; Via 2009), the frequency and nature of sexual conflict (Rice and Holland 1997; Chapman et al. 2003; Tomaiuolo et al. 2007; Crespi and Nosil 2013) and sexual selection (Maan and Seehausen 2011) in genetic divergence, and the evolutionary conditions that increase the potential for reinforcement (Maroja et al. 2009; Pinho and Hey 2010). Among these developments,

genome-scale next generation sequencing and traditional genetic analyses have characterized genomic signatures of the origins of lineage sorting in parts of the genome specific to 'barrier' or 'speciation' loci, where relative reductions in gene flow may reflect differences among populations in the relative fitness of immigrant alleles specific to those loci (Turner et al. 2005; Nosil et al. 2009; Via 2012; Nosil and Feder 2012; Ellegren 2013). This collection of observations has challenged the traditional conceptualization that all loci experience the same demographic histories and that gene flow acts solely as a strong homogenizing force that prevents ongoing differentiation by genetic drift and local adaptation (Mayr 1942; Slatkin 1985). Therefore, a more complete understanding of the interaction between migration, genetic drift and selection at specific loci can provide important perspectives on the evolutionary processes that promote local adaptation and population-level divergence (Coyne and Orr 2004; Maroja et al. 2009; Pinho and Hey 2010; Feder and Nosil 2010).

Genes involved in reproductive incompatibilities have been a central focus of emerging research on patterns of genomic evolution in sympatric divergence and speciation (Schluter 2001; Coyne and Orr 2004; Via 2009; Lessios 2011; Nosil and Schluter 2011; Servedio et al. 2011). Selection on genes that underlie mating specificity and mate recognition may play an important role in the evolution of non-geographical barriers to gene flow between populations diverging in key phenotypic or life history traits that involve interactions between individuals such as mating behavior (Martin and Hosken 2004), host choice (Hawthorne and Via 2001; Messina and Jones 2011), mate signaling (color and vision, Kronfrost et al. 2006; biochemical, Smadja and Bultin 2008), or interactions between gametes such as cognate proteins involved in fertilization specificity (Clark et al. 2006; Vacquier and Swanson 2011; Sunday and Hart 2013). Genes encoding compatibility proteins have been shown to evolve rapidly in diverse groups of animals and plants (Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008), and the rapid divergence of interacting proteins involved in reproduction has also been implicated in the maintenance of species boundaries during hybridization in mammals (Paysuer and Nachman 2005) and insects (Maroja et al. 2009), and in reproductive incompatibilities between closely related species of marine invertebrates (e.g., Vacquier 1998; Hellberg and Vacquier 1999; Zigler et al. 2005; Clark et al. 2009).

Among externally fertilizing marine species, direct biological interactions between the sexes and mate recognition are restricted to a few proteins expressed on the surfaces of male and female gametes (Lessios 2011). Therefore, gamete recognition proteins are obvious potential targets of selection and may have a primary role in the evolution of prezygotic reproductive isolation in the marine environment where geological or other physical barriers to gene flow are often absent, transitory, or porous in nature (Palumbi 2004; Lessios 2007; Turner and Hoekstra 2008). In addition to maintaining barriers among existing species, male-expressed reproductive ligands, in particular the sperm acrosomal proteins bindin and lysin, show signatures of positive selection for high intraspecific sequence polymorphism in many, but not all genera (Metz and Palumbi 1996; Biermann 1998; Zigler et al. 2005; Springer and Crespi 2007; Moy et al. 2008; but see Metz et al. 1998b; Yang et al. 2000; Yang and Swanson 2002; Zigler and Lessios 2003, 2004; Hart et al. 2012), and have been shown to be targets of positive selection for population-level differences within species of sea urchins (Geyer and Palumbi 2003), mussels (Riginos et al. 2006), and sea stars (Sunday and Hart 2013) that broadcast spawn their gametes into the plankton. Evidence of diversification in response to population-specific selective pressures suggests that proteins involved in gamete recognition and fertilization may be important candidates in the origins of genetic differentiation across the genome, although the key selective mechanisms by which gamete recognition proteins evolve in different taxa remain unclear.

Research among echinoid genera has made notable progress in characterizing the role of selection on species-level variation in bindin (Vacquier and Swanson 2011; Lessios 2011; Vacquier 2012). Bindin lives a double life by mediating sperm-egg adhesion and serving as a sperm ligand for fusion to the egg membrane (Vacquier and Moy 1977; Metz et al. 1994; reviewed in Hirohashi et al. 2008). Comparative analyses of bindin suggest that positive selection for high relative rates of amino acid substitutions and variation in the number and length of repetitive elements among alleles has acted to strengthen fertilization specificity in a species-specific manner (Biermann 1998; Levitan and Ferrell 2006; Palumbi 1999; Zigler 2008; Levitan and Stapper 2009), and that bindin divergence may be an important mechanism for reproductive incompatibility among species (Zigler et al. 2005; Swanson and Vacquier 2002). However, a key unresolved question involves the origins of bindin differences and reproductive isolation between

closely related congeneric species: do they arise initially in an early part of the speciation process as differences among conspecific populations, or long after the initial stages of speciation and the evolution of genomic or developmental incompatibilities by reinforcement against hybridization in secondary contact among divergent species? Some sea urchin genera (*Tripneustes*, *Lytechinus* and *Arbacia*; Zigler and Lessios 2003, 2004; but see Lessios et al. 2012) consist of ecologically differentiated species in which speciation events seem to correspond to shifts in habitat use or other ecological traits (Lessios 2007; Lessios et al. 2012; Binks et al. 2012). Those examples implicate ecological speciation (Schluter 2009) as the original source of population genomic divergence and of postzygotic reproductive isolation, with selection against maladaptive hybridization or mating costs in secondary contact as a possible source of rapid evolution and divergence at *bindin* and other loci involved in prezygotic reproductive isolation (Dobzansky 1940; Servedio and Noor 2003; Geyer and Palumbi 2003; McCartney and Lessios 2004; Coyne and Orr 2004; Lessios 2007; Pinho and Hey 2010). In contrast, other sea urchin studies seem to rule out a significant role for sympatry, secondary contact, and reinforcement against hybridization in the evolution of *bindin* divergence and reproductive isolation (Geyer and Lessios 2009, Calderon et al. 2009, 2010; Pujolar and Pogson 2011; Lessios et al. 2012), and in these cases an alternative source of selection leading to prezygotic reproductive isolation is needed to account for *bindin* divergence and speciation (Palumbi 2009; Lessios 2011).

Among the selective mechanisms put forward to explain such cases of non-ecological speciation (Schluter 2009) associated with positive selection on gamete recognition loci are sexual selection and intersexual conflict over the rate of fertilization and polyspermy (Gould and Stephano 2003; Levitan and Ferrell 2006; Palumbi 2009). Directional selection among males under conditions of sperm competition may favor adaptations in male reproductive proteins for more efficient swimming, binding to the egg extracellular coat, penetration of the egg extracellular layers, and fusion with the egg membrane (see Levitan 2004; Palumbi 2009). If high rates of sperm contact often lead to fatal polyspermy for eggs (and low reproductive success for females), then females may experience antagonistic selection to counter adaptations among males and sperm by altering their egg surface sperm receptors (Frank 2000; Palumbi 2009). One possible result of this kind of sexual conflict is a coevolutionary 'arms race' among sperm and egg

compatibility loci, which can drive the evolution of high allelic variation and diversification within populations by frequency-dependent selection (Gavrilets and Waxman 2002; Gavrilets and Hayashi 2005; Haygood 2004; Levitan and Ferrell 2006; Palumbi 1999; Palumbi 2009; Levitan and Stapper 2009; Tomaiuolo and Levitan 2010). Strong support for this hypothesis comes from studies under natural spawning conditions and laboratory induced fertilization experiments in sea urchins, which reveal a density-dependent interaction among bindin genotype (Palumbi 1999; Levitan and Ferrell 2006) or specific amino acid changes (Levitan and Stapper 2009) and reproductive success.

Differences in specific mating system variables have been used to explain variation in the susceptibility to polyspermy and signatures of positive selection among lineages of closely related sea urchins (*Strongylocentrotus*; reviewed in Levitan 2006; Levitan 2008; Pujolar and Pogson 2011). Theory suggests that high sperm densities and strong sperm competition should lead to high rates of polyspermy and strong conflict between the sexes over the rate of sperm-egg contact, and select for traits such as higher sperm velocity in males and smaller egg size (smaller sperm targets) in females. In contrast, under conditions of low sperm density and limited sperm competition, sexual conflict should be weak and both sexes should experience selection favoring high compatibility between all sperm and all eggs (Levitan 1993; Levitan 2006). Therefore, the form and intensity of sexual conflict and the signatures of positive selection may vary among species that commonly experience high-density spawning, high local sperm concentrations and strong sperm competition, in comparison to species in which sperm competition and sexual conflict are predicted to be relatively weak (Levitan 2006; Levitan 2008; Pujolar and Pogson 2011).

To investigate the selective processes that may shape patterns of divergence at compatibility loci within and between congeneric species, I characterize the gene that encodes bindin in all three species of the sea star genus *Pisaster*: *P. ochraceus*, *P. brevispinus* and *P. giganteus* with overlapping geographic ranges in the northeastern Pacific Ocean. I describe bindin gene structure in a species-level comparison and analyze its molecular evolution. I then investigate patterns of positive selection among specific codons and lineages of bindin alleles in multiple populations of two of these species and compare them to patterns of divergence in the mitochondrial *cytochrome c oxidase 1 (COI)* gene to understand how patterns of bindin variation within species differ

between *P. ochraceus* and *P. brevispinus*, which are expected to experience different forms and intensities of sexual selection and sexual conflict. I predict that the intensity and extent of positive selection is stronger in *P. ochraceus*, whose life history characteristics are more consistent with a selective regime of high sperm competition and polyspermic conditions. If selection against hybridization is a significant source of selection for bindin divergence, an expected outcome of bindin evolution in response to reinforcement would be a pattern of bindin differentiation between *P. ochraceus* populations that are sympatric and allopatric with *P. brevispinus* populations, or higher rates of bindin evolution when species occur in sympatry compared to allopatry (Coyne and Orr 2004; Lessios 2007).

First, I compare the full-length bindin coding sequence among species and analyze the evolutionary relationships among the repetitive domains of the variable second bindin exon. The comparison of the highly repetitive bindin domain structure suggests that concerted evolution of repetitive domains has a strong effect on bindin divergence among species and bindin variation within species. Second, I examine population variation in the second bindin exon of two species: I show that positive selection acts on bindin allelic variation in the abundant and widespread keystone predator *P. ochraceus* but not in *P. brevispinus*. This difference is consistent with predictions based on life history differences between the two species that should lead to more intense sperm competition and higher polyspermy risk in *P. ochraceus*. Third, I show that there are no strong differences in patterns of positive selection or bindin divergence between *P. ochraceus* populations that are sympatric with *P. brevispinus* compared to populations that are found to the north or south of the *P. brevispinus* geographic range. This similarity suggests that the modest evidence for positive selection acting on *P. ochraceus* bindin evolution is likely not caused by reinforcement against hybridization with *P. brevispinus*. Fourth, I combine bindin and COI data in isolation-with-migration models to explore the historical demographic context of these positive selection results. Parameter values estimated from the model fit suggest that the relatively weak signal of positive selection on within-species bindin divergence among *P. ochraceus* alleles can be accounted for in part by relatively recent population divergence and northward population expansions that may be coupled with the potential homogenizing effects of concerted evolution on bindin divergence in this genus.

2.2. Methods

2.2.1. Study System

Pisaster species are broadly sympatric along the western coast of North America. *Pisaster ochraceus* (Brant 1835) has the widest range, and occurs in wave-exposed rocky intertidal habitats at middle and lower tidal heights from Prince William Sound, Alaska to Cedros Island, Baja California, Mexico (Lambert 2000). This species is a dominant invertebrate predator in the intertidal, where it plays an important ecological role in regulating species richness and community dynamics (Paine 1974). Adult individuals are abundant and can be found in high-density aggregations during feeding and spawning in the spring and summer months (Paine 1974). *Pisaster brevispinus* (Stimpson 1857) is the largest species in this genus but has a much smaller range from Sitka, Alaska to Santa Barbara, California (Lambert 2000). Individuals are commonly found at low spawning densities and are generally widely dispersed in low intertidal and subtidal habitats on sandy-mud substrates where they scavenge and prey on bivalves, snails and sand dollars (Farmanfarmaian et al. 1958, Smith 1961). Both species have separate sexes. *P. ochraceus* and *P. brevispinus* females generally contain mature oocytes (*P. brevispinus* 160-170 μm ; *P. ochraceus* 150-160 μm ; Strathmann 1987) and spawn during the months of May to August and March to August, respectively (Strathmann 1987). Interspecific fertilization can be induced under laboratory conditions (M. Hart unpublished observations), however hybridization has not been reported in nature. Fertilized eggs develop into bipinnaria larvae that swim and feed in the plankton for several weeks (Strathmann 1987). Although the larval duration for *P. brevispinus* has not been documented, the larvae of *P. ochraceus* can stay in the water column for 6 to 8 weeks (Strathmann 1987) and are expected to travel long distances (Harley et al. 2006). *Pisaster giganteus* inhabit low rocky intertidal habitats, (Farmanfarmaian et al. 1958) and are generally absent north of Monterey Bay, California (Ocean Biogeographic Information System; <http://iobis.org>), however the range and abundance of this species is relatively understudied. Because *P. giganteus* is relatively rare, its geographic distribution is poorly known, and the binding of this species is particularly difficult to sequence due to the highly homogeneous repetitive domain structure of the second binding exon, our population comparisons focus on *P. brevispinus* and *P. ochraceus*.

2.2.2. Population Sampling

Samples were collected from a total of 36 *P. ochraceus* and 22 *P. brevispinus* individuals from eight localities from Alaska, British Columbia and California (Table 1; Fig. 1). *Pisaster ochraceus* individuals were sampled from two populations from the northern (Cordova, AK) and southern (La Jolla, CA) extremes of its range where *P. brevispinus* has not been observed, and from three populations where it is sympatric with *P. brevispinus* (Bamfield, BC; Vancouver, BC; Hopkins Marine Station, CA). *Pisaster brevispinus* is only found in sympatry with *P. ochraceus*, therefore, all populations that were sampled for this study coexist with *P. ochraceus*. I used *P. ochraceus* samples from Hopkins and La Jolla (Harley et al. 2006) and from Cordova (Marko et al. 2010) that were collected for previously published mtDNA surveys; *P. brevispinus* samples from California (San Francisco Bay) and a single *P. giganteus* sample used to obtain a full-length bindin allele sequence were obtained from the California Academy of Sciences, CA (CAS accession numbers: 149761, 157354A, 157354B, 162445, 162449, 172146, 116600). All other samples were collected for this study from the field; tissues samples consisted of 10-20 tube feet preserved in 70-95% ethanol.

2.2.3. Genetic Data

For each individual sea star, genomic DNA was extracted from a single preserved tube foot using a proteinase K digestion with a 2x cetyltrimethyl ammonium bromide incubation (CTAB; Grosberg et al. 1996), followed by a chloroform extraction and ethanol precipitation. Extracted DNA was quantified and diluted in dH₂O to a desirable concentration of 10ng/μL. The initial full-length bindin alleles were amplified from *P. ochraceus* and *P. giganteus* using terminal primers CSCD and LAR (Table 2), which correspond to conserved amino acid motifs in the preprobindin and core domain (respectively) of sea urchins (Gao et al. 1986; Minor et al. 1991) and the sea star *Patiria miniata* (Patiño et al. 2009). An alternative preprobindin primer, ADAV (Table 2) was designed to amplify bindin in *P. brevispinus*. Thermal cycling conditions for both sets of primers were 94° (1:00) for one cycle; 94° (0:15), 58° (0:30), 68° (5:00) for 33 cycles; and final elongation 70° (10:00). The polymerase chain reaction (PCR) amplification cocktail consisted of 0.5μL Clontech 50x Advantage Taq polymerase mix and 2.5μL of

10x Advantage buffer; 19.5µL dH₂O; 0.5µL DNTP mix (10mM); 0.5µL forward and reverse primers (10µM); 1µL DNA. Amplicons were checked using 1% agarose gel electrophoresis and consisted of a single band of about 5 Kb (5027 bp in length for *P. ochraceus*, 4974 bp for *P. giganteus*, and 4859 bp for *P. brevispinus*). PCR amplicons were immediately cloned using a TOPO TA Cloning Kit (with PCR 4-TOPO vector; Invitrogen). Ten to 20 clones were screened for correct insert size, cleaned for sequencing using a DNA purification system (Wizard Plus SV Minipreps; Promega), and sequenced using universal plasmid primers by a commercial sequencing service (www.operon.com). I confirmed the identity of those sequences based on blastx sequence similarities to *Patiria miniata* bindin (GenBank accession ACJ70121.1; Patiño et al. 2009). Custom forward and reverse primers were used to sequence along the remaining length of the gene (Table 2). Due to the highly homogeneous repetitive structure of the second bindin exon of *P. giganteus*, this gene region was not accessible with custom sequencing primers. Instead, I re-amplified the second exon using the primers LRD and PGE, which correspond to the closest non-repetitive amino acid motifs flanking the highly repetitive domain (Table 2). I used the same PCR cocktails, thermal cycling conditions and cloning methods as described above and sequenced through the tandem repeats using universal plasmid primers. I characterized some general features of the bindin coding sequence structure by searching for peptide cleavage sites in ProP (<http://www.cbs.dtu.dk/services/ProP>). I used the statistical analysis of protein sequences (SAPS; http://www.isrec.isb.ch/software/SAPS_form.html) package to estimate the molecular weight and amino acid composition of the predicted mature bindin molecule.

The population-level analyses were focused on the highly repetitive second exon in bindin, where previous studies have documented highly significant population-level patterns of variation (Sunday and Hart 2013). In total, I cloned and sequenced 70 *P. ochraceus* and 40 *P. brevispinus* bindin alleles (Table 1). I designed two PCR primers, DTL and TRQ (Table 2), to amplify the second exon for both *P. ochraceus* and *P. brevispinus* populations. Thermal cycling conditions used were 94° (1:00) for one cycle; 94° (0:15), 71° (0:30), 68° (2:00) for 26 cycles; and final elongation 70° (10:00). The PCR amplification cocktail used is described above. Amplicons consisted of a single band of 2154-2388 bp in length for *P. ochraceus* and 2144-2163 bp for *P. brevispinus*

which correspond to the second bindin exon and about 50-200 bases pairs of 3' and 5' flanking introns. Slight modifications to the PCR thermal cycles, annealing temperature and DNA concentrations were made for a few samples to increase amplification strength or to minimize nonspecific banding. Six to 10 clones were screened for correct insert size and sequenced with forward and reverse universal plasmid primers. Alleles were inferred based on these sequences. Two or three copies of each inferred allele were sequenced with custom internal primers (Table 2) to minimize PCR and cloning errors, and unique nucleotide substitutions that differed from the consensus sequence of that allele were classified as sequence errors (see Sunday and Hart 2013). Additional alleles were sequenced if single nucleotide differences or sequence ambiguities could not be resolved and true polymorphisms could not be confirmed by majority rule. Only one allele was sequenced and included in the data set for two *P. ochraceus* individuals in which nucleotide differences among the consensus bindin sequences indicated a second allele, but the full length or true polymorphisms for that second allele could not be resolved. If sequences from all clones sequenced with plasmid primers had no polymorphisms, two to six clones were sequenced with internal primers to distinguish among potential alleles. If all sequenced clones represented the same allele, then this individual was considered a homozygote and two identical copies of this allele were included in the data set; PCR products from potential homozygotes were purified using Exo-SAP-IT PCR product cleanup kit (Affymetrix, Inc) and direct sequenced to confirm the absence of double nucleotide peaks. Chromatograms were proofread in 4Peaks v. 1.7.2 (A. Griekspoor and Tom Groothuis, mekentosj.com.). Two *P. brevispinus* individuals from San Francisco Bay could not be amplified, cloned and sequenced with high quality; therefore these samples were not included in the analysis.

I also amplified an 816 bp region of COI using primers from Harley et al. (2006). Although some *P. ochraceus* individuals from Hopkins and La Jolla were already sequenced previously for COI by Harley et al. (2006), several of the sequences available from GenBank contained ambiguous nucleotide sites and the data were reported for only a 543 bp region of the gene. Therefore, I sequenced COI haplotypes for 35 *P. ochraceus* and 20 *P. brevispinus* individuals that were also sequenced for bindin in this study (Table 1; one Hopkins sample could not be amplified). The thermal cycling conditions used are described in Keever et al. (2009). I used a PCR reaction cocktail of 0.2 μ L Tsg

DNA polymerase and 5.0 μ L of 10x reaction buffer; 34.8 μ L dH₂O; 1.0 μ L DNTP mix (10mM); 2.0 μ L MgCl₂ (50mM); 1.0 μ L forward and reverse primers (10 μ M); 5 μ L DNA. All mtDNA sequences were checked on agarose gels, purified, and direct sequenced as described above.

2.2.4. Sequence Alignment

Bindin sequences were trimmed to the coding sequence only and visualized in Se-AI v2.0 (Rambaut 2002). Although bindin had strong sequence similarity within each species, length variation among relatively few haplotypes complicated the alignment certainty among highly similar repetitive regions. I used a codon-based alignment algorithm in PRANK (Löytynoja and Goldman 2008; <http://code.google.com/p/prank-msa/>), with default alignment parameter settings, to infer homology among codons, because the detection of positive selection can be sensitive to alignment error (Jordan and Goldman 2012). PRANK uses a codon model to align protein-coding DNA, and simultaneously compares nucleotide and amino acid similarity among sequences. Therefore, for regions of DNA where amino acid similarity is low, the model is able to produce extra phylogenetic information by back-translating and aligning the nucleotides to maximize the similarity of the translated protein. This method has been shown to increase the accuracy of aligning truly homologous codons and to improve the power of detecting site-wise positive selection in downstream analyses, while maintaining low levels of false positives even in alignments of alleles with a high frequency of insertion-deletion differences (when compared to other alignment methods; Jordan and Goldman 2012). I used PRANK to generate intraspecific alignments for *P. ochraceus* and *P. brevispinus* bindin alleles; I then compared them to alignments that included a single *P. giganteus* bindin allele as an outgroup (sister species to *P. ochraceus*; Mah and Foltz 2011). All alignments generated by PRANK were manually inspected and only minor adjustments were necessary. Because the second exon is highly repetitive and varies in the number of repetitive domains among species (Fig. 2c), the orthology relationships among individual repeats is uncertain and I could not generate an interspecific haplotype alignment that included all three species with high confidence. Therefore, I limited these analyses to the intraspecific bindin alignments. Mitochondrial COI sequences for both

species had high sequence similarity and were all the same length. Therefore, it was not necessary to use an alignment algorithm to infer homology among codons.

2.2.5. Gene Genealogies

I used the Model Selection tool implemented on the Datamonkey webserver (www.datamonkey.org; Delpont et al. 2010) to infer the best-fit nucleotide substitution model for the bindin and COI haplotype data. The best model of molecular evolution (based on AIC comparisons) for each bindin alignment was determined to be F81 (Felsenstein 1981). For the COI alignments, the best-fit model of evolution was HKY85 (Hasegawa et al. 1985). The GARD method for recombination detection in HyPhy (Kosakovsky Pond et al. 2011;) was used to screen the bindin alignment, but GARD found no evidence of recombination in the second exon of bindin for either species.

Intraspecific bindin genealogies were constructed using maximum likelihood (ML) and Bayesian approaches. All duplicate haplotypes were removed and the gene trees for both species were generated including and excluding a single bindin sequence from *P. giganteus* as a root. I used MEGA 5.0 (Tamura et al. 2011) to estimate a ML bindin gene tree for both species, under a Jukes-Cantor substitution model. Gaps were treated as missing data (complete deletion), rates among categories were gamma distributed with 4 rate categories, and branch support for the phylogeny was assessed by bootstrapping with 1000 replicates. Bayesian phylogenetic analyses were performed using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). I applied an F81 substitution model and ran the two parallel MCMC (Markov chain Monte Carlo) searches with four chains, for 3×10^6 generations, sampling every 1000 generations; after a burnin of 200000 a 50% majority rule consensus tree and posterior probability values of the nodes were estimated using the remaining trees from the posterior distribution. Run length and convergence of multiple chains was assessed using the average standard deviation of split frequencies between runs in Tracer v.1.5 (Rambaut and Drummond 2007). Maximum likelihood and Bayesian methods generated almost identical topologies, with only minor differences in relationships between two terminal branches of the phylogeny; therefore, only the Bayesian gene trees are shown (Fig. 5). In addition, the intraspecific genealogies generated by both Bayesian and ML methods did not change when they

were rooted with a *P. giganteus* sequence. Likewise, the COI topology for either species was not altered when a genealogy was inferred using a combined species alignment.

2.2.6. Repetitive Domain Analyses

Repetitive domains were defined in RADAR (Heger and Holm 2000; Goujon et al. 2010) using the longest bindin allele from each species alignment and the single allele sequenced for *P. giganteus*. To investigate the evolutionary relationships among repetitive domains within and between species, I treated each repeat domain as a separate sequence and created an alignment of repeats within a single bindin allele and combined this alignment across all three species (Fig. 3a; e.g., Meeds et al. 2001). I generated a phylogeny of repetitive domains using Bayesian sampling (Ronquist and Huelsenbeck 2003) and applied a HKY85 (Hasegawa et al. 1985) substitution model to the data (Fig. 3b). I looked for phylogenetic patterns among repeats consistent with an evolutionary mode of concerted evolution, in which the processes of gene conversion or misalignment and unequal cross-over produce repeat paralogs from existing repetitive units within alleles (Smith 1976; Dover 1982; Elder and Turner 1995; Swanson and Vacquier 1998; McAllister and Werren 1999; Carmon et al. 2010). I looked for evidence for two key characteristics indicative of this process: 1) repetitive domains are expected to have a higher degree of sequence similarity within species, relative to repeats in corresponding regions of the coding sequence between species (Swanson and Vacquier 1998; Meeds et al. 2001); 2) unequal cross-over events are expected to be suppressed closest to the unique sequences flanking the repetitive array (Stephans 1989); therefore, nucleotide substitutions that occur within terminal repeat domains are expected to be maintained and may lead to both higher sequence divergence among terminal repeats relative to other internal repeat domains, and greater sequence similarity between adjacent internal repeats within species (Durfy and Willard 1989; McAllister and Werren 1999; also see Carmon et al. 2007).

2.2.7. Population Genetic Analyses

I used DnaSP v. 4.10 (Rozas et al. 2003) to calculate nucleotide diversity from individual segregating sites and pairwise comparisons, and haplotype diversity indices including sites within alignment gaps. The sliding window method was used to calculate

and plot variation in nucleotide diversity per base pair across the bindin coding sequence, using the default settings (window length=100; step size=25; Fig. 4). Allele frequencies, Tajima's D (Tajima 1989) values and population differentiation in bindin and COI (Φ_{ST} ; Weir and Cockerham 1984) were calculated among all populations in Arlequin v. 3.11 (Excoffier et al. 2005). The most appropriate substitution models were applied to the data and pairwise Φ_{ST} values were tested for significant departures from zero using 10,000 random permutations of the data. Bonferroni corrections were used to adjust p-values for multiple pairwise comparisons. An analysis of molecular variance (AMOVA) as implemented in Arlequin (significance estimated with 10,000 permutations) was used to determine the partitioning of genetic diversity in bindin and COI as the percentage of the total genetic variation among populations, and among individuals within populations with no geographical groupings. TCS 1.21 (Clement et al 2000) was used to construct and visualize parsimony networks of bindin and COI haplotypes using a 95% confidence of connection limit (Fig. 1).

2.2.8. Tests of Selection

I characterized selection acting on coding sequences using codon models that estimate the rate of non-synonymous nucleotide substitutions (nonsilent) per nonsynonymous site, dN , compared to the rate of synonymous substitutions (silent) per synonymous site, dS , among specific bindin lineages and codon sites, using the maximum-likelihood method codeml in PAML 4.4 (Yang et al. 2007). Because the rate of synonymous substitutions (dS) provides an approximation of the rate of neutral evolution (ie. reflects mutation), the ratio $\omega = dN/dS$ can be interpreted in terms of positive selection ($\omega > 1$) for amino acid changing substitutions or purifying selection ($\omega < 1$) against amino acid changes. I used the unrooted intraspecific ML and Bayesian bindin phylogenies that consisted of only unique alleles. The branch-site method in codeml (Yang et al. 2005; Zhang et al. 2005) implements a maximum likelihood model of codon evolution to compare $\omega = dN/dS$ among sets of foreground and background branches in the genealogy and to identify individual codons, alleles and lineages evolving under positive selection for high relative rates of amino acid substitutions with $\omega > 1$. The selection model allows all of the sites along a foreground branch or a set of lineages in the gene tree to vary in ω and belong to one of three site classes: ω between 0 and 1

(class 0), $\omega=1$ (class 1) and ω estimated for the foreground branches (class 2a, 2b), relative to the background rate of evolution which can reflect both selection (2a) and neutral evolution (2b). Positive selection can be inferred if the rate class ω_2 is >1 and the likelihood value of a selection model is significantly greater than the likelihood value of the null model, without selection, where the estimated rate of dN/dS for the foreground lineages is set to neutral selection or $\omega=1$ (using a likelihood ratio test with one degree of freedom, the difference in the number of ω parameters between the null and selection model). The standard likelihood ratio test statistic ($-2\Delta l$), twice the log-likelihood difference, was compared to the X^2 distribution ($p=0.05$) to determine if the selection model was a significantly improved likelihood fit to the sequence data, relative to the null model. The branch-site test then uses a Bayesian inference to estimate the posterior probability that individual codons in the sequence alignment belong to the rate class with $\omega >1$ and are subject to positive selection (Yang et al. 2005); for these tests I used a critical posterior probability value of $p>0.90$. For all codeml analyses I used pairwise deletion of sequence gaps to estimate dN and dS for sites within regions of the alignment that included insertion-deletions differences among some alleles.

I used the ML and Bayesian bindin genealogies for each species and performed a series of PAML analyses based on intraspecific population-level hypotheses to test a hypothesis of differential rates of positive selection among branches or among sites for each single population in the foreground, with all of the other populations in the background. For each population I defined the foreground set of lineages as each of the terminal branches that lead to single alleles, and any internal branches that lead to clades of alleles sampled from that population. Additional analyses to identify individual sites and branches under positive selection in bindin and COI were carried out using two methods implemented in HyPhy that employ maximum likelihood-based analyses to fit different models of codon evolution and dN/dS rate variation to the sequence data. I used the MEME method (Murrel et al. 2012) to detect the presence of diversifying selection on individual codon sites among some unspecified branches in the genealogies. The branch-site REL method (Kosakovsky Pond et al. 2011) was used to identify sets of lineages in the bindin gene trees that differ in the proportion of unspecified codons experiencing positive selection and the average value of dN/dS at $p<0.05$. These analyses were also carried out using the ML bindin topology and the

most likely bindin gene tree generated from the Bayesian analysis, to assess if differences among the tips of the topology would affect the inference of positive selection.

2.2.9. Coalescent Demographic Analysis

The isolation-with-migration model in IMA (Nielsen and Wakeley 2001; Hey and Nielsen 2007; Hey and Nielsen 2004; <http://genfaculty.rutgers.edu/hey/software#IMa2>) was used to characterize the population demographic history underlying bindin divergence among *P. ochraceus* populations and among *P. brevispinus* populations. The software implements a non-equilibrium isolation-with-migration coalescent population model and jointly estimates six demographic parameters. Effective population sizes, ancestral population size, divergence time, and directional migration rates are inferred by Bayesian sampling (Markov chain Monte Carlo algorithm) that explores many highly probable genealogies (given the data) and parameter values (given the genealogies) for each locus from a pair of populations (or species) that share a common ancestor (IMa method; Hey and Nielsen 2007; Hey and Nielsen 2004; Hart and Marko 2010 for review). I performed a series of IMA analyses using different combinations of geographically adjacent population pairs from north to south across the sampled geographic range for each species. For each population pair I combined the bindin haplotype data with sequence data for COI into a two-locus analysis (Dataset S2). Preliminary single-locus analyses using only the COI haplotype data did not contain enough sequence diversity to independently calculate reliable posterior probability distributions and maximum likelihood estimates (MLE) for all demographic parameters. Therefore, parameter estimates based on population demography alone (ie. the balance between geneflow and genetic drift) could not be used to qualitatively estimate the effects of selection on the relative rates of geneflow and population divergence specific to the bindin locus. The maximum likelihood estimates for population parameters were scaled to the neutral mutation rate (μ). Model parameters were converted to population parameters in demographic units (population size estimates N_1 , N_2 , N_A as number of individuals, divergence time t in years) using a mtDNA mutation rate calibration of $\mu=2.812 \times 10^{-6}$ gene⁻¹ year⁻¹ based on a divergence rate of 4.56% per 3 million years estimated by Lessios et al. (2001) for the least-divergent sister species of *Diadema* and

Meoma sea urchins separated by the Isthmus of Panama. This mutation rate has been used for coalescent analyses for other sea star species (McGovern et al. 2010; Puritz et al. 2012) and likely represents the most accurate mutation rate compared to other geminate Panamanian sea urchin species pairs that may have diverged from each other before the divergence of *Diadema* species (Knowlton and Weigt 1998; Marko 2002; Hickerson et al. 2006). The HKY finite sites mutation model was applied to the data (Hasegawa et al. 1985) and a generation time of 5 years was used as the average age of breeding in adults in these species of long-lived sea stars (Menge 1975; Strathmann 1987; also see Marko et al. 2010 and McGovern et al. 2010).

Several short runs were performed to identify the range of appropriate prior distributions for each parameter, and then priors were adjusted such that each posterior parameter distribution was contained within the lower and upper bounds of the prior distribution. The MCMC search was optimized such that the chains sufficiently 'mix' or explore all of the possible highly probable genealogies and parameter values. I modified the number of chains in the parameter space ($n=100$) and a geometric heating scheme was applied to the search to improve the rate of mixing for multilocus data. Trees were sampled every 100 steps after a burnin of 2-3 million steps. To identify sufficient chain mixing and convergence of each search, I looked for high ESS (effective sample size) values greater than 100; ESS values are estimates of the number of independent observations that have been sampled for each parameter and are good indicators that run durations are sufficient and that Markov chains are mixing and independently exploring the parameter space. I also observed no autocorrelation trends in the model probability plots for the t parameter (splitting time) and the summary state statistic (L[P]), high swapping rates between adjacent chains, and consistent parameter estimates between early and late sets (set values) of saved genealogies. Importantly, I checked for consistency in parameter estimates within and among all independent runs carried out for each population pair and compared them to a final long run of 4-8 million steps. I used a sample of 50,000 gene trees saved from the last MCMCMC run in an additional analysis, 'Load Trees Modes' or L-mode, in IMA. The L-mode analysis carries out likelihood ratio tests to determine whether a series of nested models with fewer parameters is a significantly worse fit to the data than the fully parameterized model (Θ_1 , Θ_2 , Θ_A , m_1 , m_2). The main purpose for testing hypotheses about differences among

parameter values is to ask whether the migration value estimates in both directions are significantly different from each other, and if they are significantly different than the null hypothesis of zero gene flow (Hey and Nielson 2007).

2.3. Results

2.3.1. Gene Structure

The gene coding for bindin was highly similar in overall structure in all three *Pisaster* species and consisted of three exons and three protein domains: a cysteine-rich 5' preprobindin region, a central repetitive domain and the conserved 3' functional core (Patiño et al. 2009; Fig. 2a). The longest coding sequence from the conserved preprobindin motif CSCD to the core domain was 2196 bp (732 codons) in *P. ochraceus*, 2142 bp (714 codons) in *P. giganteus* and 2034 bp (678 codons) in *P. brevispinus* (see Fig. 2b). The predicted amino acid sequence had high similarity to the preprobindin and core regions of *P. miniata* bindin (expectation value $E=2 \times 10^{-11}$; Patiño et al. 2009). The 86 amino acid preprobindin sequence (Fig. 2b) is highly conserved among *Pisaster* species and contains just three nucleotide differences (between *P. brevispinus* and the other two species), two of which confer nonsynonymous substitutions (codons 55 and 74). In addition to the CSCD motif, the preprobindin region contains two other cysteine residues (codons 31 and 50) that are shared with *P. miniata* (Patiño et al. 2009). *Pisaster* preprobindin contains an intron ranging from 1677 bp to 1704 bp. Based on bindin cDNA from another forcipulate sea star, *Evasterias trochelii* (S. Patiño, unpublished data), this intron is predicted to occur within a conserved tyrosine (codon 53) in the preprobindin sequence.

ProP identified a probable furin-type cleavage motif RVRR (codons 83-86) following the preprobindin region, identical to the furin-type cleavage site RVRR in the sand dollar *Encope stokesii* (Zigler and Lessios 2003) and similar to the RARR motif in *P. miniata* (Patiño et al. 2009) where the N-terminal preprobindin sequence is predicted to be enzymatically cleaved from the mature molecule in the sperm acrosomal vesicle (Fig. 2b; Gao et al. 1986; Patiño et al. 2009; Zigler and Lessios 2003). The mature peptide downstream of the cleavage site is predicted to be 70 kDa, slightly smaller than

the estimated size of the presumed bindin protein isolated from the acrosomal vesicle of *P. ochraceus* sperm (95 kDa; Christen 1985). The most common amino acids in mature bindin are lysine (26.3%) alanine (13.2%) and glycine (10.7%). *Pisaster* mature bindin contains no cysteine or tryptophan residues, similar to other echinoderm taxa for which the mature bindin protein has been sequenced (Zigler and Lessios 2003; Patiño et al. 2009).

Mature bindin gene structure is dominated by a central repetitive region (described below) located in the second exon between the preprobindin sequence and the core domain that varies among species in the number of repeating domains defined in RADAR, with up to 16 repeats in *P. ochraceus*, 10 repeats in *P. giganteus* and nine repeats in *P. brevispinus* (Fig. 2c). Species also varied in the length of glycine- and lysine-rich collagen-like regions that separated the other parts of the repetitive domains (Fig. 2c). The collagen-like regions consisted of variable numbers of GKGRKK repeats that are similar to the collagen-like domains in the bindin of *P. miniata* and may facilitate self-affinity among bindin molecules following the acrosomal reaction during sperm and egg adhesion (Vacquier and Moy 1977; Patino et al 2009).

The predicted bindin core domain immediately downstream from a conserved leucine residue (also the predicted splice site for the second intron of 1027-1038 bp) had high sequence similarity to the invariant core region of *P. miniata*, with only a singleton amino acid difference (a serine-asparagine polymorphism) that is conserved in charge and polarity, and was similar to the central core region of several sea urchin species with the lowest expectation value of $E=4 \times 10^{-4}$ when compared to *Pseudoboletia* sp. (AFB82032.1). The 43 amino acid residues of the core domain were very highly conserved among all three *Pisaster* species, with only one synonymous replacement in *P. brevispinus*. The LGLLLRHLRHH motif, corresponding to part of the “B18” core domain of sea urchins, was identical among *Pisaster* species and identical to the same B18 regions of all sea urchin taxa (Zigler and Lessios 2003; Vacquier 2012) and other sea stars (Patiño et al. 2009) in which bindin has been sequenced (Vacquier and Swanson 2011). This domain has been shown to have a functional role in sperm-egg membrane fusion (Zigler 2008; Vacquier 2012) and has been predicted to be the binding substrate for the sea star egg bindin receptor OBi1 (Hart and Foster 2013; M. Hart, unpublished data).

2.3.2. Repetitive Domain Analysis

RADAR identified all non-collagen repetitive domains as a single repeat type (see Fig. 3a). Repetitive domains of *P. ochraceus* and *P. brevispinus* included both long repeats (30 codons; dark green, labeled 'A' in Fig. 2c) and short repeats (20-21 codons; light green, labeled 'B' in Fig. 2c), although the B repeats were not present in *P. giganteus*. Bayesian phylogenetic analyses showed that A and B repeat domains were grouped into two strongly supported clades (Fig. 3b). The overall similarity in gene structure among species, with multiple copies of distinct repetitive domain types, suggests that this repetitive structure is ancestral for the three *Pisaster* species, and that repeat copies in the same location within the gene structure across species might be orthologs and more closely related to each other than to paralogous repeat copies in other parts of the same allele. Instead, I found that the A repeats had overall high homogeneity among all repeat copies, with no strong pattern of higher or lower sequence similarity among repeats within species compared to between species. Low variation among A repeats (ranging from 91%-100% sequence similarity) limited the resolution of phylogenetic relationships among repeat copies within clade A (Fig. 3b). Therefore, the distinction between A repeats that are orthologous and originated in a common ancestor, versus repeats that originated independently within each species by unequal cross-over or another molecular process, could not be inferred. The genealogy of repetitive domains, however, showed that B repeats are more uniform within species than between them, with the exception of the first B repeat (1B) in *P. brevispinus*, which is more closely related to B repeats found in *P. ochraceus* (Fig. 3b). Among the five B repeats found in *P. ochraceus*, the three middle (internal) repeats (2B, 3B, 4B) grouped into a single clade (Fig. 3b), consistent with the predicted effects of concerted evolution within species (Elder and Turner 1995; Swanson and Vacquier 1998; McAllister and Werren 1999; Meeds et al. 2001).

2.3.3. Sequence Diversity

Bindin

Bindin alleles differed by insertion-deletion mutations that produced variation in the length of the second exon ranging from 1728-1962 bp in *P. ochraceus*, including five individual heterozygotes with insertion-deletion differences between alleles involving part

or all of one or more *A* or *B* repetitive domains or collagen-like sequence motifs (GKGRKK). I found fewer length differences in the smaller sample of *P. brevispinus* individuals and alleles (and I sequenced only a single allele of *P. giganteus*). All *P. ochraceus* individuals were heterozygous for the second exon of *bindin*; the aligned coding sequence was 1980 bp (660 amino acids) with 35 variable nucleotide sites and 26 amino acid polymorphisms (Table 3). I found extensive allele sharing among populations; out of 70 sequenced *bindin* alleles, there were only 25 haplotypes and only 13 gene copies were not shared among populations. The *P. brevispinus* *bindin* alleles ranged in length from 1719-1737 bp and varied only in the number of collagen-like sequences motifs (GKGRKK); the alignment was 1737 bp (579 amino acids) and was slightly less polymorphic with 17 variable sites and 14 amino acid polymorphisms (Table 3). Out of 20 *P. brevispinus* individuals sequenced, 4 were homozygous. Among 40 sequenced alleles, there were 16 haplotypes. One common haplotype was found at a relatively high frequency ($n=18$); this haplotype was inferred as ancestral by TCS and was present in all three sampled locations (Fig. 1b).

Haplotype diversity (h) was high within all populations across the geographical range studied, but slightly higher in *P. ochraceus* ($h= 0.9231-0.9615$) than in *P. brevispinus* ($h=0.7857-0.8167$). Nucleotide diversity was low and ranged from 0.00131-0.00249 among all populations (Table 3). Across all of the populations sampled, I observed the highest nucleotide diversity (~ 0.006) contained within the C- and N-terminal repeat domains in *P. ochraceus* and the C-terminal repeat domain in *P. brevispinus* (Fig. 3c). In both species, Tajima's D values were generally negative (Table A1), which may reflect an excess of rare alleles indicating population expansions or a signature of positive or purifying selection, however these statistics were not significantly different from neutrality ($p>0.1$) in all but two southern localities (SF Bay and La Jolla; Table A1). There were no fixed nucleotide or indel differences among populations. An analysis of molecular variance (AMOVA) emphasized high within-population variability, and suggested that the percentage of genetic variation that could be attributed to variation among populations was not significantly different from zero (Table A3a). Pairwise population differentiation (Φ_{ST}) was generally low (-0.05023 to 0.04830 in *P. ochraceus*; -0.006 to 0.07935 in *P. brevispinus*) and not significantly different from zero

in any pairwise comparisons (Table A2a), including *P. ochraceus* populations that are sympatric or allopatric with a congener.

Cytochrome c oxidase 1 (COI)

Although the mtDNA sample sizes in this study were relatively small, I found sequence diversity among COI haplotypes in *P. ochraceus* similar to those in Harley et al. (2006). The *P. ochraceus* alignment included 13 variable nucleotides sites, a singleton amino acid polymorphism, and seven unique haplotypes among 35 sequences (Table 3). Variation among 20 *P. brevispinus* sequences included 14 variable nucleotide sites, one variable amino acid site and seven unique haplotypes (Table 3). Generally, I observed variable COI haplotype diversities ($h=0.2857-0.8571$) for all populations and low nucleotide diversity ranging from 0.00035-0.00710 (Table 3). Tajima's D values were generally negative and all were non-significant. An AMOVA analysis found high within-population variability and attributed 22.45% ($p<0.01$) of genetic variance to variation among populations in *P. ochraceus* and 5.09% (NS) among *P. brevispinus* populations (Table A3b). Both species were characterized by high COI homogeneity (Φ_{ST} not significantly different from zero; Table A2b). I found the highest Φ_{ST} values (0.263-0.312; NS) between *P. ochraceus* sequences sampled in Cordova relative to all other conspecific populations. These findings are in line with significant genetic differentiation previously observed in *P. ochraceus* sampled from Cordova (Marko et al. 2010).

2.3.4. Selection Analyses

Codon models revealed a weak but statistically significant signal of positive selection ($\omega>1$) favoring a high relative rate of amino acid substitution differences among *P. ochraceus* bindin alleles. MEME identified two sites (codons 557 and 624) in both ML and Bayesian bindin gene trees to be under strong positive selection in the *P. ochraceus* alignment (Table 4). The branch-site REL analysis inferred two terminal branches in *P. ochraceus* to be under strong positive selection ($\omega>>1$, that is, a very large but imprecisely estimated value); the first was a private allele found at Hopkins, and the second lineage was a highly derived allele (based on a haplotype network inferred in TCS with a 95% confidence limit) sampled only once in both La Jolla and Cordova ($\omega=1256$) (Table 4; Fig. 5); this particular allele contained an eight amino acid deletion

following a lysine-serine substitution at the positively selected site at codon 624. Although the estimated ω along these branches was exceptionally high, the branch site REL model identified few (0.01) or no sites with a high probability of $\omega > 1$ along either lineage, suggesting that the signature of selection among *P. ochraceus* alleles may be relatively weak. Branch-site models in PAML (Table 5) corroborated the tests of selection in MEME and estimated a high mean dN/dS ratio across a small proportion (0.5%) of codons with $\omega > 1$, when alleles sampled from Cordova were in the foreground and all other populations were constrained as background lineages in the ML ($\omega = 197$) and Bayesian ($\omega = 455$) topologies. Both analyses identified one site (codon 624) with a posterior probability $> 95\%$ for belonging in a separate rate class with $\omega \gg 1$. Tests for positive selection among *P. ochraceus* alleles sampled in Hopkins and La Jolla also yielded a significantly better fit to the data compared to the null model in both the ML and Bayesian topologies, and estimated a high foreground rate of positive selection (ω ranging from 44.8-107) among a few codons. However, these analyses could not identify any individual sites with a significant posterior probability of evolving under positive selection (Table 5).

The MEME and branch-site REL methods did not identify any specific sites or branches with $\omega > 1$ among *P. brevispinus* lineages within *P. brevispinus* (Table 4). The series of branch-site codon models that compared signals of positive selection among *P. brevispinus* populations could not reject the null model in all likelihood ratio tests when each population was specified in the foreground (Table 5). Additional tests of positive selection in the COI genealogy implemented in HyPhy did not find any evidence for $dN > dS$ among branches or codon sites at this locus.

2.3.5. Coalescent Demographic Analysis

Multilocus analyses among five geographically adjacent *P. ochraceus* population pairs (Fig. 6a; Table 6; Dataset S1) sampled in this study revealed a clear difference in estimated splitting times among southern (California) populations and populations from northern regions (British Columbia and Alaska). The posterior probability distributions for divergence time (t) between Hopkins and La Jolla populations had MLEs of $t \sim 21,000$ years (Fig. 6a; Table 6), suggesting that southern populations last shared a common

ancestor near the end of the last glacial maximum (LGM) about 20,000 year ago. All other adjacent pairs of populations sampled from Hopkins to Cordova consistently had the highest likelihood modes for relatively recent divergence times < 5000 years (Fig. 6a), although the upper confidence limit on the MLEs for t were relatively old and overlapped with the LGM (Table 6). Such recent divergence times were consistent among all runs, despite the relatively slow mutation rate applied to the COI data, compared to other mutation rates reported in the literature (Knowlton and Weigt 1998; Marko 2002; Hickerson et al. 2006); a faster mutation rate calibration would lead us to infer even more recent divergence times among northern *P. ochraceus* populations.

In contrast, the isolation-with-migration model revealed variable population histories among three *P. brevispinus* population pairs (Fig. 6b; Table 6; Dataset S1). Estimated population splitting times among populations predated the LGM and were ~ 40,000 years for Bamfield and Port Moody and ~270,000 years before present for Bamfield and San Francisco Bay. Among populations from Port Moody and San Francisco Bay, divergence time also predated the LGM ~ 34,000 years since they last shared a panmictic ancestor, indicating that *P. brevispinus* populations likely survived in deeper waters in northern refugia during the LGM. However, it would be desirable to test this pattern of an old north-south divergence using data from additional southern populations.

The small number of loci (2) in this study limited my ability to estimate all parameters in the same isolation-with-migration model (Edwards and Beerli 2000; Marko and Hart 2011). As a result, the posterior probability distributions for effective population size (N_e) were poorly resolved among all population pair analyses in both species; all posterior distributions had a low likelihood mode and a flat posterior distribution of low non-zero probabilities for very large population sizes. Extending the q_1 and q_2 parameter priors did not resolve the wide range of non-zero parameter value probabilities. It is likely that because of the large effective population sizes that are common for broadcast spawning species with dispersive larvae (Marko et al. 2010) it is difficult to capture the relatively high, within-population diversity with relatively small sample sizes ($n \leq 16$) for just two loci. Nevertheless, the maximum likelihood N_e estimates (the modes in the posterior probability distributions) for all *P. ochraceus* populations were consistently ~20,000-50,000 individuals, which was about half the magnitude estimated by Harley et

al. (2006) from mutation-drift equilibrium statistics (Table 6). The isolation-with-migration model also estimated relatively large ancestral population sizes of ~200,000 individuals between the northern populations sampled in this study (Lighthouse Park, Bamfield, Cordova), compared to smaller estimates (~ 55,000 individuals) among southern populations (Table 6). Effective population size estimates for *P. brevispinus* were much larger (~ 300,000 individuals) and ancestral N_e was ~ 150,000 for all population pairs (Table 6).

The directional migration parameters (m_1 , m_2) for all population pairs had the highest likelihood values for the lowest bin in the posterior distribution, with the proportion of gene copies that are new migrants every generation ranging from 0.00097 to 0.02090 among all *P. ochraceus* populations and 0.0042 to 0.0127 among *P. brevispinus* populations (Table 6). However, because all probabilities for gene flow parameter estimates were relatively low, and low nonzero probabilities were estimated for a wide range of very high (and very low) gene flow parameter values, it was necessary to use relatively wide prior distributions to characterize the posterior probabilities distributions. Therefore, the posterior probability distributions for most migration parameter estimates had poorly defined upper and lower bounds (Figure A1). Likelihood ratio tests in L-mode yielded significantly poorer fits to all seven nested models that were fixed for the null gene flow parameter ($m=0$) for all population pair analyses in both species, with the exception of two populations pairs in which m estimates were not significantly different from $m=0$ (immigration into Lighthouse park, *P. ochraceus*; and into Bamfield, *P. brevispinus*). The L-mode results, coupled with weakly defined lower bounds for migration parameters (including zero) and the absence of clearly defined posterior distributions with non-zero highest likelihood modes, suggest that gene flow is generally very low among populations on the spatial scale sampled in this study for both species.

2.4. Discussion

Genes encoding reproductive compatibility proteins have been shown to evolve rapidly under positive selection and have been implicated in the primary divergence among conspecific populations and in the maintenance of reproductive boundaries

between already diverged species in sympatry. I characterized the gene coding for bindin for all three sea star species in the genus *Pisaster* and analyzed its molecular evolution within and between species. *Pisaster* bindin is similar in structure to that of other sea star genera studied to date, and is highly variable in a central repetitive domain that separates the preprobindin sequence and the conserved functional core. I discover that evolution of the central repetitive domain among species may be partly influenced by the processes of concerted evolution within species, leading to significant differences in the copy number of repetitive domains among all three *Pisaster* species and homogenization of variation among paralogous repeat copies within species. Despite the potential effects of concerted evolution on within-species polymorphism, I find a weak signal of positive selection on bindin divergence among *P. ochraceus* alleles compared to *P. brevispinus*, with no significant genetic differentiation among alleles and current biogeographic patterns of selection and bindin variation that cannot be explained by ongoing reinforcement. Several evolutionary hypotheses can account for the absence of a strong signal of positive selection on bindin at the level of conspecific populations: the absence of phenotypic selection, a relatively weak response to selection in the context of small population sizes, the homogenizing effects of gene flow on the signal of substitutions and selection acting on bindin, protein diversification by an evolutionary mechanism other than an excess of nonsynonymous differences among alleles that can be detected by codon models, or insufficient time for the effects of selection to manifest as bindin diversification. I find evidence supporting the latter two potential explanations for the relatively modest signal of positive selection I detected among bindin codons in *P. ochraceus* and I consider how these findings may lend a more complete understanding of the source of bindin divergence in *Pisaster* species.

Comparisons of codon model results among *P. ochraceus* populations consistently identified one site (codon 624) under positive selection and indicated relatively weak diversifying selection among populations and across two terminal bindin lineages leading to alleles sampled in populations allopatric and sympatric with *P. brevispinus*. The inference of positive selection on alleles sampled from allopatric populations, along with a lack of bindin differentiation among sympatric and allopatric populations of *P. ochraceus*, suggests that reinforcement during secondary contact may not be an important selective mechanism for bindin divergence among these congeners.

Patterns of positive selection that are inconsistent with reinforcement have also been shown in gamete recognition proteins of other species of marine invertebrates. Clark et al. (2007) found a signature of positive selection in two paralogs of the sperm acrosomal protein lysin in an abalone subspecies (*Haliotis tuberculata coccinea*) that does not coexist with any other congeners. Geographic variation of the sperm protein M7 lysin among three hybridizing *Mytilus* species also showed no obvious patterns of differentiation consistent with reinforcement as a major source of selection on this gene (Riginos et al. 2006). Among sea urchins, positive selection in bindin was detected among allopatric and sympatric populations of *Echinometra lucunter* and its sister species *Echinometra viridis*, across their geographic distribution (Geyer and Lessios 2009; also see Geyer and Palumbi 2005), and the allopatric congeners *Paracentrotus lividus* and *Paracentrotus gaimardi* show highly variable bindin sequences that are influenced by positive selection (Calderon et al. 2009; Calderon et al. 2010). Similarly, phylogeographic analyses among both allopatric (Metz et al. 1998b) and sympatric hybridizing species (Lessios 2012) in the genus *Arbacia* identified no evidence of selection on bindin, suggesting a pattern that is also inconsistent with the hypothesis of ongoing reinforcement against hybridization.

Conservative population-level comparisons of positive selection could not single out any sites or branches with statistical support for $\omega > 1$ among bindin lineages within *P. brevispinus*. These results suggest a lack of ongoing positive selection among bindin alleles within that species. In comparison to statistically significant divergence under selection among some *P. ochraceus* alleles, the general absence of diversifying selection among *P. brevispinus* alleles is consistent with the predicted effect of lower spatial density of spawning adults and sperm-limiting conditions for fertilization, which are expected to decrease the potential for a sexual conflict over polyspermy (Levitan 1993; Levitan 2006).

Although I found statistically significant support for positive selection on bindin divergence among *P. ochraceus* populations, the small number of alleles and codons under selection is unexpected because *P. ochraceus* are characteristically found along shallow rocky shores in high-density aggregations during the summer months when spawning takes place, and this species seems more likely to experience high sperm concentrations, strong sperm competition, and strong selection by sexual conflict over

fertilization rates (and polyspermy) similar to other intertidal species (e.g., abalone, Stephano 1992; Babcock and Keesing 1999). One potential explanation for the relatively limited evidence for positive selection among bindin codons and lineages in *P. ochraceus* is methodological: the capacity of codon models to identify population differences in positively selected lineages or sites may be limited by the degree of resolution of intraspecific genealogies (e.g., Sunday and Hart 2013). If within-population polymorphism is relatively low (i.e., shallow branches), the true relationships among haplotypes may be harder to resolve, and confounding polytomies may limit the potential to identify signals of selection along terminal lineages (Anisimova et al. 2001). Furthermore, because methods that estimate dN/dS by maximum likelihood were originally designed for analyses of divergent interspecific sequences, ω describes the relative rate of nonsynonymous versus neutral or synonymous fixation events (Zhang et al. 2005; Kryazhimskiy and Plotkin 2008). Therefore, the power to detect $dN/dS > 1$ is relatively low among conspecific samples where differences between sequences are segregating polymorphisms within populations, with shallow differences among alleles rather than fixed differences among species, and the results of such analyses may depend on the location (shallow versus deep) of a particular lineage in the genealogy (Kosakovsky Pond et al. 2011).

Another potential explanation for modest positive selection on *Pisaster* bindin is the effect of concerted evolution on bindin divergence among conspecific alleles. Lineage-specific concerted evolution has been shown for other surface recognition molecules that interact with other proteins, including the egg receptor for abalone sperm lysin, pollen coat genes, fungal glycoproteins involved in immunorecognition, and egg extracellular coat proteins in *Drosophila* (e.g., Galindo et al. 2002; Fiebig et al. 2004; Johanneson et al. 2005; Carmon et al. 2007). The process of concerted evolution may lead to interspecific diversity in the copy number of repetitive units among species, and ultimately leads to the elimination of substitution differences and homogenization of repeat structure within species through the molecular mechanisms of gene conversion or unequal crossing over at meiosis (Smith 1976; Dover 1982; Elder and Turner 1995; Swanson and Vacquier 1998; McAllister and Werren 1999; Carmon et al. 2010). I found significant variation in the number of repeating domains between *Pisaster* species, with 13-16 repeats in *P. ochraceus*, 10 repeats in *P. giganteus* and nine repeats in *P.*

brevispinus, although the overall length of the second bindin exon and the repetitive nature of the gene structure were relatively conserved among species (Fig. 2c). My analysis of the central repetitive structure among all three *Pisaster* species identified two repeat domain clades, *A* and *B*, which were strongly supported in Bayesian inference (Fig. 3b). I found evidence suggesting repetitive domain evolution by concerted evolution within species, as indicated by inferred homology and higher sequence similarity among *B* repeat domains within each species than between repeat domains corresponding to the same region of the coding sequence among species.

A second prediction of this hypothesis is that mispairing of repeat units and unequal cross-over events are expected to be suppressed closest to unique non-repetitive sequences at the ends of repeating arrays (Stephan 1989). Therefore, nucleotide substitutions that occur in the terminal repeat domains should be maintained across populations of alleles and may lead to higher sequence divergence at the terminal repeats relative to other internal repeat domains (McAllister and Werren 1999; Meeds et al 2001; Carmon et al. 2007). The diversity in repeat structure in *P. ochraceus* was consistent with this prediction, with three internal *B* repeats (*2B*, *3B*, *4B*) grouped into a single clade (Fig. 3b) and the highest nucleotide diversities observed across the C- and N-terminal repeat domains (Fig. 4). The lack of homogenization in repeat *1B* in *P. brevispinus* is also consistent with concerted evolution within species, suggesting that repeat *1B* has not been homogenized since *P. brevispinus* last shared a common ancestor with *P. ochraceus*. However, the highest bindin nucleotide diversity among all *P. brevispinus* populations sampled in this study was found only in the C-terminal region (Fig. 4).

While unequal crossover can be a major source of allelic diversity through re-shuffling of existing variation or insertion and deletion of repeat copies within alleles (McAllister and Werren 1999; Meeds et al. 2001), gene conversion and the elimination of sequence polymorphism (within and between alleles) by concerted evolution effectively reduces the potential to detect the signal of diversifying selection within a population and across the entire length of a gene (Gay et al. 2007). Importantly, the observation that the site (codon 624) inferred to be under selection in *P. ochraceus* is positioned in the last repetitive domain, and that this site is followed by a polymorphic indel of eight amino acids flanking the non-repeating sequence at the C-terminal end of the second exon,

also supports the prediction that repeat copies at the end of a repetitive array should be less likely to experience gene conversion or homogenization by concerted evolution. Therefore, a signal of positive selection ($\omega > 1$) averaged across all sites in a specific lineage is expected to be weak if an excess of substitution differences among alleles is likely to be limited to sites at the end of a repeating array. Interestingly, codon 624 in *P. ochraceus* occurs in the same part of the bindin gene structure as the positively selected codon 842 in two geographically isolated populations of the sea star *Patiria miniata* (Sunday and Hart 2013). Positive selection has also acted on polymorphism in the same corresponding gene region in sea urchin bindin, ~40 amino acids upstream of the central core domain in six species from the genus *Echinometra* (Metz and Palumbi 1996; McCartney and Lessios 2002, 2004; Zigler et al. 2005; Vacquier and Swanson 2011) and among *Stongylocentrotus* sea urchins (Biermann 1998).

Positive selection for bindin divergence on the C-terminal end of the repetitive domain in *P. ochraceus* may underlie functional variation and sequence divergence in a potentially species-specific manner that is similar to the mechanism by which the vitelline envelope receptor for lysin (VERL) evolves in abalone species (Swanson and Vacquier 1998). VERL is a major egg coat glycoprotein and consists of 22 tandem repeats that are ~153 residues in length (Swanson and Vacquier 1998; Galindo et al. 2002). While repeats 3-22 evolve neutrally and have been homogenized by concerted evolution, such that repeats within species are more similar to each other than to repeats between species (Galindo et al. 2002), sequence divergence and positive selection has been observed in the N-terminal repeats 1 and 2, suggesting that they evolve independently, and likely function in species specificity of sperm-egg interactions (Galindo et al. 2003; Clark et al. 2009).

In addition to the effect of the molecular mechanisms by which bindin evolves, my results suggest that population demographic effects can also influence the spatial distribution of adaptive variation in bindin and other coding sequences that might otherwise be expected to show a strong signal of divergence under selection. Tests of population divergence showed no significant differentiation in bindin among populations, even over large spatial scales, with no differences in the pattern of differentiation among *P. ochraceus* populations that are or are not sympatric with a congener. Similar patterns have been previously documented for COI spatial variation in *P. ochraceus*, despite

striking variation in color morphology across the range of this species (Marko et al. 2010; Harley et al. 2006; Stickle et al. 1992; also see Pankey and Wares 2009), and populations are characterized by a general lack of genetic structure for nuclear genes, mitochondrial DNA, and allozymes from 22 populations across the entire geographic range from California to Alaska, which includes two known biogeographic breaks found in other marine species from the northeastern Pacific (Dawson et al. 2001; Sotka et al. 2004; Harley et al. 2006; Kelly and Palumbi 2010). Such low levels of divergence and widespread allele sharing have been largely interpreted as high levels of gene flow among populations by ongoing long distance dispersal of larvae that can stay in the water column for 6 to 8 weeks (Strathmann 1987). However, genetic similarity between populations can also reflect haplotype or allele sharing due to incomplete lineage sorting of ancestral polymorphisms that were present in both populations before they recently diverged, rather than by virtue of ongoing gene flow (Marko and Hart 2011).

I used the coalescent-based method in IMA to distinguish between shared ancestral polymorphism from allele sharing caused by recurrent gene flow among geographically adjacent conspecific populations (Hart and Marko 2010 and Kuhner 2009 for review). This method is particularly informative for species from biogeographic areas like the northeast Pacific that have been strongly influenced by temporally varying processes such as Pleistocene glacial cycles that can drive haplotype frequencies and population genetic structure far from the evolutionary equilibrium between mutation, gene flow, and genetic drift (McGovern et al. 2010; Marko and Hart 2011). Demographic analyses among three *P. brevispinus* population pairs had the highest likelihood estimates for older population splitting times ~ 34,000-270,000 years before present, that predated or overlapped with the end of the LGM ~ 20,000 years ago (Fig. 6b). Glacial persistence in northern refugia and deep coalescent histories in this biogeographic region have been shown to be especially common for species inhabiting lower intertidal habitats (Marko et al. 2004; 2010). A series of IMA analyses among five adjacent *P. ochraceus* populations that were not genetically differentiated revealed recent divergence times <5000 years among populations sampled from the northern half of the geographic range and older estimated splitting times (~21,000 years) between two southern populations sampled in California (Fig. 6a). This pattern of relatively shallow demographic histories among northern populations is consistent with a model of recent

post-glacial recolonization from southern populations into higher latitudes, and confirms the recent findings of Marko et al. (2010) that this species has experienced significant population expansions over the last 20,000 years in the northeastern Pacific. The isolation-with-migration model also estimated low gene flow among all populations, suggesting that few migrants successfully travel among populations each generation, with no clear pattern of asymmetric gene flow towards southern or northern directions for both species. Such estimates are consistent with low levels of current-mediated larval dispersal and population connectivity previously predicted in *Patiria miniata* from the same biogeographic region (J. Sunday, unpublished data). Together, these findings indicate that a combination of recent divergence, relatively large population sizes and very low gene flow can explain the genetic similarity observed among *P. ochraceus* populations in a demographic context of relatively recent northern expansions along the northeastern Pacific coast.

A much stronger pattern of population-specific positive selection for bindin divergence has been identified in *Patiria miniata*, among two populations that occur north and south of a relatively old (~280,000 years) phylogeographic separation in the northeastern Pacific (Keever et al. 2009; McGovern et al. 2010; Sunday and Hart 2013). Sunday and Hart (2013) found high allelic diversity and divergent selection for alternative bindin lineages and codons among northern and southern populations in the face of low ongoing gene flow estimated from non-coding phylogeographic markers (McGovern et al. 2010). Because *P. miniata* does not coexist with any closely related species, positive selection for divergent bindin variation does not reflect reinforcement against hybridization. If divergence between *P. miniata* populations instead reflects selection based on sexual conflict within populations, then this result may suggest that extended periods of time (with low gene flow) may be necessary for independent coevolutionary arms races among male and female gametes driven by sexual selection to produce patterns of high allelic diversity and bindin divergence between population that are detectable using codon models. This interpretation is directly relevant to my comparison of *Pisaster* species: high allelic diversity within all *P. ochraceus* populations sampled in this study (compared to generally low diversity at the mtDNA locus) is consistent with an arms race outcome of sexual conflict favoring selection for bindin polymorphism within populations (Gavrilets 2000; Gavrilets and Waxman 2002; Haygood 2004); however, a

few thousand years since the time of population splitting estimated for *P. ochraceus* in this study may not be enough time for the effects of positive selection through sexual conflict to lead to population differences in bindin, especially for large-bodied sea stars with long generation times and long life spans of several decades (Menge 1975). Therefore, my analyses of bindin variation within and between species of *Pisaster* sea stars may indicate lower and upper bounds on the temporal scale on which the effects of selection on bindin may be most likely to be detected: >5000 years (among *P. ochraceus* populations that show positive selection but no population divergence), and possibly not >280,000 years (between *Patiria miniata* populations that show strong positive selection and highly significant population bindin differentiation).

While this study aimed to distinguish among the effects of selection on bindin molecular evolution, functional studies identifying the correlation between intraspecific repeat variation (e.g., Minor et al 1991) or specific amino acid differences (e.g., Levitan and Stapper 2009), and sperm-egg compatibility during fertilization are required for a more complete understanding of the role of concerted evolution and the selective mechanisms that mediate species-specific fertilization and bindin divergence within and between *Pisaster* species. Repetitive structure variation caused by concerted evolution and tandem duplication or deletion of complete repeat domains (and other smaller indel differences) in the mature bindin of *Pisaster* species suggests that these sea stars may experience selection on bindin divergence in a manner similar to bindin in other broadcast spawning marine invertebrates (e.g., Biermann 1998; Zigler and Lessios 2003; analogous, Moy et al. 2008; Sunday and Hart 2013). Similar variation in the number and length of repetitive domains flanking the core region, and as few as 8-10 amino acid replacements, have been shown to be associated with gamete incompatibility among congeneric species of sea urchins (Zigler et al. 2005). Experimental and molecular studies of pairs of coevolving sperm and egg recognition proteins are needed to effectively characterize the molecular interactions between male and female proteins (e.g., Clark et al 2009; Tomaiuolo and Levitan 2010) and to sharpen our general understanding of how genetic changes are linked to ongoing coevolution and population-level divergence in sperm and egg compatibility.

Table 1. Summary of collection sites and number of individuals and alleles sampled per location, per locus

Species	Location	Code	Coordinates	Individuals	Alleles sampled	
					Bindin	COI
<i>P. ochraceus</i>	Cordova, AK	COR	60.54°N 145.70°W	7	14	7
	Bamfield, BC	BAM	48.83°N 125.14°W	7	14	7
	Lighthouse Park, BC	LHP	49.33°N 123.26°W	7	14	7
	Hopkins Marine Stn, CA	HOP	36.62°N 121.90°W	8	15	7
	La Jolla, CA	LA	32.84°N 117.28°W	7	13	7
<i>P. brevispinus</i>	Bamfield, BC	BAM	48.83°N 125.14°W	8	16	8
	Port Moody, BC	PM	49.28°N 122.83°W	8	16	8
	San Francisco Bay, CA	CAS	37.72°N 122.28°W	6	8	4

Table 2. Primers sequences for each locus and location in gene

Locus	Species	Name; Direction	Primer Oligonucleotides	Location in gene
Bindin	<i>P. ochraceus</i>	CSCD; forward	5'-CCTGTTCTGTGATCTGCTG-3'	exon 1
		DET; forward	5'-GGATGAGACTTGTGGATGATTGCTT-3'	intron 1
		AAV; forward	5'-AGGCTGCAGTCTGGTGACTT-3'	intron 1
		DTL; forward	5'-GACACCCTCCCACTGTTTTCACTGT-3'	intron 1
		LRD; forward	5'-GCAGTTTGAGAGATGCTGCTCATT-3'	exon 2
		EVS; reverse	5'-TGCTTAACTGAAACCTC-3'	exon 2
		PGE; reverse	5'-TGTATTAAGGGTAGCCGTTTCACCAGG-3'	exon 2
		TRQ; reverse	5'-AAGTGTGGGGAGCCTGTCTTGTG-3'	intron 2
		LEL; reverse	5'-GTATAGTTAGGGCTTGTGATAAGTAT-3'	intron 2
		LAR; reverse	5'-TCACCAATGCGAGCTAGAAGACTGG-3'	exon 3
	<i>P. brevispinus</i>	ADAV; forward	5'-GCTGATGCAGTATCCCATCATGGTGAT-3'	exon 1
		QLK; forward	5'-CCAGCTTAAGAACAAGAACAGTT-3'	exon 2
mtDNA	<i>P. ochraceus</i>	Poc-f-mt; forward*	5'-CTAATGATTGGCGCACCAGATA-3'	
		Poc-r-mt; reverse*	5'-GTAGTGAAAGTGGGCAACTACG-3'	

* (Harley et al. 2006)

Table 3. Polymorphism statistics (including sites with gaps), number of variable nucleotide and amino acid sites, number of unique haplotypes, nucleotide diversity (π ; at individual sites), haplotype diversity (h) and number of alleles with indel variation.

Locus	Species	Location	Variable Sites		Unique haplotypes	π	h	Alleles with indels	
			Nucleotide	Amino Acid					
Bindin	<i>P. ochraceus</i>	COR	16	13	10	0.00178	0.9451	2	
		BAM	16	12	9	0.00183	0.9231	0	
		LHP	20	13	9	0.00235	0.9231	1	
		HOP	20	16	11	0.00249	0.9333	0	
		LA	22	15	11	0.00199	0.9615	2	
		All Populations	35	26	25	0.00209	0.9346	5	
	<i>P. brevispinus</i>	BAM	11	10	9	0.00210	0.8167	3	
		PM	11	8	8	0.00148	0.8000	3	
		CAS	9	7	5	0.00131	0.7857	1	
		All Populations	17	14	16	0.00124	0.7897	7	
	mtDNA	<i>P. ochraceus</i>	COR	10	0	4	0.00710	0.8571	0
			BAM	2	0	3	0.00093	0.5238	0
			LHP	1	0	2	0.00035	0.2857	0
HOP			3	1	3	0.00175	0.7143	0	
LA			3	0	4	0.00146	0.8095	0	
All Populations			13	1	7	0.00285	0.6689	0	
<i>P. brevispinus</i>		BAM	7	1	4	0.00279	0.7500	0	
		PM	10	0	5	0.00502	0.8571	0	
		CAS	3	0	3	0.00306	0.8333	0	
		All Populations	14	1	7	0.00389	0.8316	0	

Table 4. Summary of results from MEME and branch-sites REL tests of positive selection, implemented in HyPhy.

Species; Analysis	Genealogy	Positively Selected Codons		Branches with $\omega > 1$
		MEME ($p > 0.1$)	Branch-site REL ($p > 0.5$)	Branch-site REL ($p > 0.5$)
<i>P. ochraceus</i> ;	ML	557V, 624K		2 lineages
	Bayesian	557V, 624K		2 lineages
<i>P. brevispinus</i> ;	ML	0		0
	Bayesian	0		0

Table 5. Summary of branch-site models implemented in PAML. All codons inferred to be under selection correspond to sites within each separate species alignment.

Species; Analysis	Genealogy	Branch-site Hypothesis	-2ΔlnL	Parameter Estimate (ω)	Positively Selected Sites ($\omega > 1$; $p \geq 0.90$)
<i>P. ochraceus</i> ; Population-level	ML	Cordova	8.34*	$\omega_2=197$	624K (p=0.948)
		Bamfield	0	$\omega_2=1$	-
		Lighthouse Park	0	$\omega_2=1$	-
		Hopkins	5.11*	$\omega_2=44.8$	-
		La Jolla	4.23*	$\omega_2=77.4$	-
	Bayesian	Cordova	9.99*	$\omega_2=455$	624K (p=0.971)
		Bamfield	0	$\omega_2=1$	-
		Lighthouse Park	0	$\omega_2=1$	-
		Hopkins	5.03*	$\omega_2=47.0$	-
		La Jolla	5.12*	$\omega_2=107$	-
<hr/>					
<i>P. brevispinus</i> ; Population-level	ML	Bamfield	2.18	$\omega_2=188$	-
		Port Moody	0	$\omega_2=1$	-
		SF Bay	1.24	$\omega_2=1$	-
	Bayesian	Bamfield	0	$\omega_2 \gg 1$	-
		Port Moody	0	$\omega_2=1$	-
		SF Bay	0	$\omega_2=1$	-

* significant ($df=1$)

Table 6. Summary of population parameter estimates inferred by IMA. ML effective population size estimates N_1 , N_2 and ancestral population size, N_A in 10^5 individuals, with confidence intervals (90% highest posterior densities, HPD) in parentheses. Divergence time in thousands of years (t). '?' indicates weakly defined upper or lower bounds on the posterior probabilities of estimated parameter values. The proportion of migrants into population 1 (m_1) and population 2 (m_2) per generation.

Species	Populations	N_1	N_2	N_A
<i>P. och</i>	COR—BAM	0.594 (0.155-14.5)	0.578 (0.089-14.5)	2.11 (1.27-3.46)
	COR—LHP	0.529 (0.089-12.1)	0.419 (0.007-12.0)	1.80 (0.945-3.31)
	BAM—LHP	0.226 (0.029-2.42)	0.189 (0.024-2.40)	0.491 (0.002-2.21)
	LHP—HOP	0.293 (0.076-6.36)	0.271 (0.083-6.39)	0.623 (0.048-1.26)
	HOP—LA	0.489 (0.173-8.72)	0.479 (0.183-8.79)	0.581 (0.002-1.26)
<i>P. bre</i>	BAM—PM	2.12 (0.455-14.8)	1.88 (0.604-14.9)	1.76 (0.733-3.48)
	PM—CAS	2.83 (0.446-262)	2.23 (0.447-85.3)	1.45 (0.531-3.79)
	BAM—CAS	2.99 (1.06-169)	4.27 (1.74-84.6)	1.25 (0.446-0.342)

Species	Populations	t (kyr)	m_1	m_2
<i>P. och</i>	COR—BAM	4.02 (0.211?-67.6?)	0.012 (0.012-20.3)	0.018 (0.018-29.4)
	COR—LHP	2.69 (0.897-270)	0.014 (0.014-23.3)	0.014 (0.014-23.8)
	BAM—LHP	3.59 (0.717?-878?)	0.021 (0.021-33.5)	0.021 (0.021-33.8)
	LHP—HOP	4.29 (0.598?-309?)	0.021 (0.021-34.7)	0.021 (0.021-35.1)
	HOP—LA	21.3 (0.644-337)	0.009 (0.009-16.1)	0.009 (0.009-16.0)
<i>P. bre</i>	BAM—PM	42.8 (1.29-399)	0.005 (0.005-7.82)	0.005 (0.005-8.10)
	PM—CAS	33.9 (1.48-441)	0.013 (0.004-6.79)	0.004 (0.004-7.23)
	BAM—CAS	272 (1.39-529)	0.005 (0.005-7.57)	0.005 (0.005-7.51)

Figure 1. Collection sites for five *P. ochraceus* and three *P. brevispinus* populations sampled in this study are shown on the map on the right. Dashed lines represent the northern and southern range boundaries for *P. brevispinus*. Solid lines represent the ranges boundaries for *P. ochraceus*. Refer to Table 1 for exact localities and sample sizes. Bindin haplotype networks for a) *P. ochraceus* and b) *P. brevispinus* are shown on the left. Each circle represents a unique haplotype; the area of the circle is proportional to the frequency of the haplotype; colors show the proportion of each haplotype from different localities shown on the map. Each line between haplotypes represents one inferred mutational step, slashes indicate an additional mutation, and numbers within lines specify the number of mutational steps greater than three. Isolated haplotypes represent alleles that differed from the rest of the sample by insertion-deletion differences and could not be connected to the rest of the haplotype network by nucleotide substitutions only.

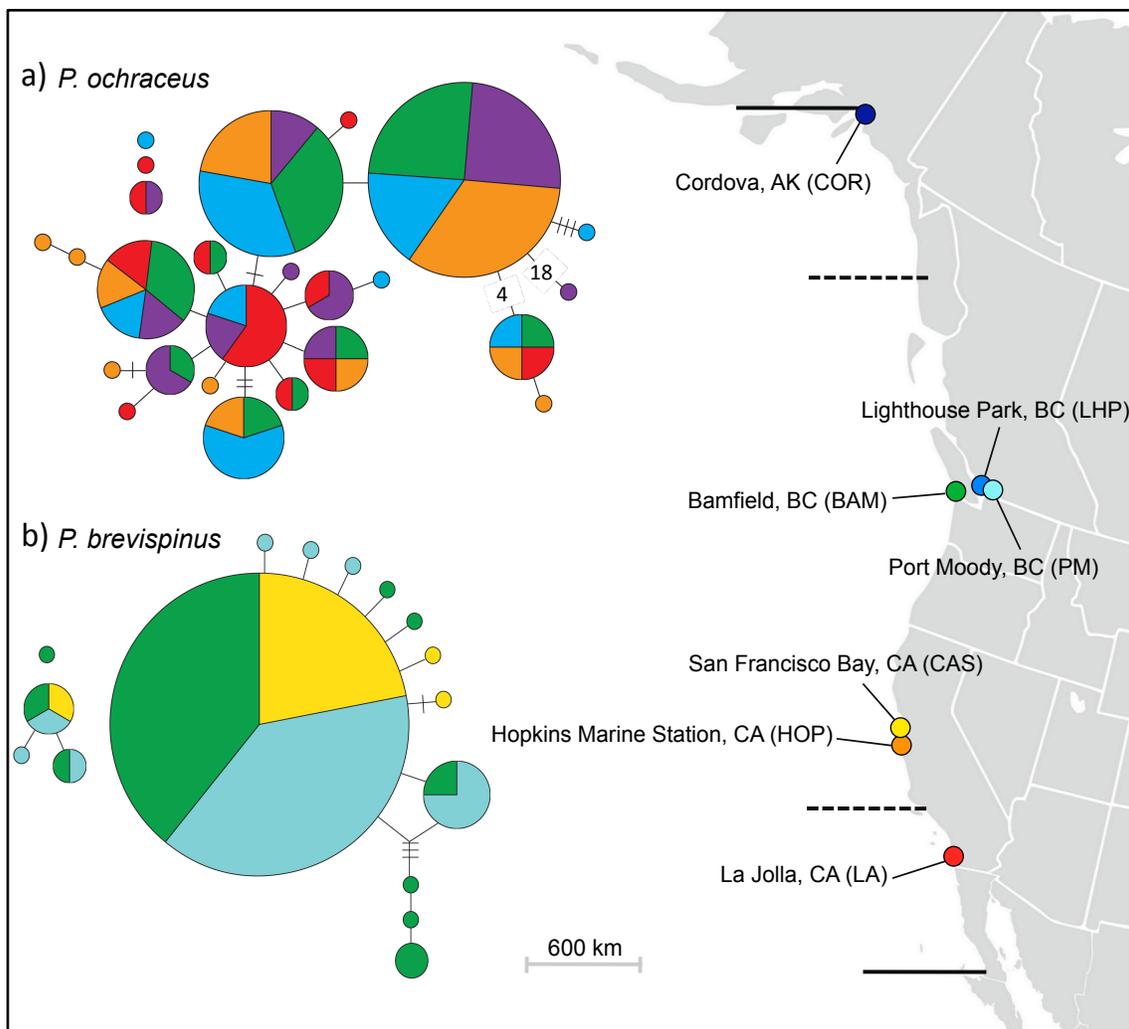
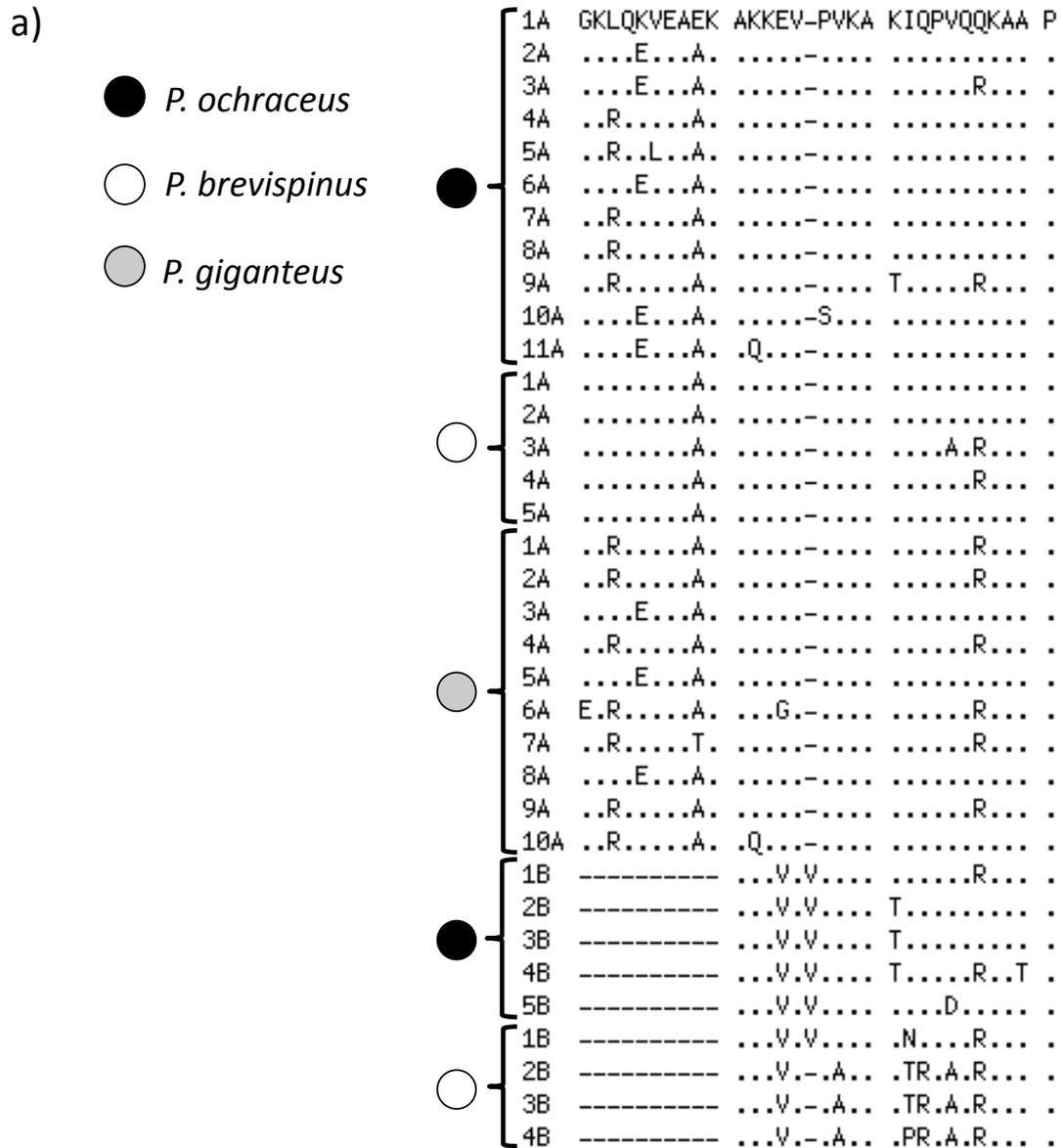


Figure 3. a) Alignment of repeat domains in the second bindin exon for all three *Pisaster* species. b) Consensus Bayesian genealogy of repetitive domains with Bayesian posterior probabilities of partition at branch nodes. 'A' and 'B' denote separate clades grouping repeat types A and B. Circles at branch tips are colored corresponding to species (see legend in panel a) and numbered corresponding to repeat number counted in the 5' to 3' direction



b)

- *P. ochraceus*
- *P. brevispinus*
- *P. giganteus*

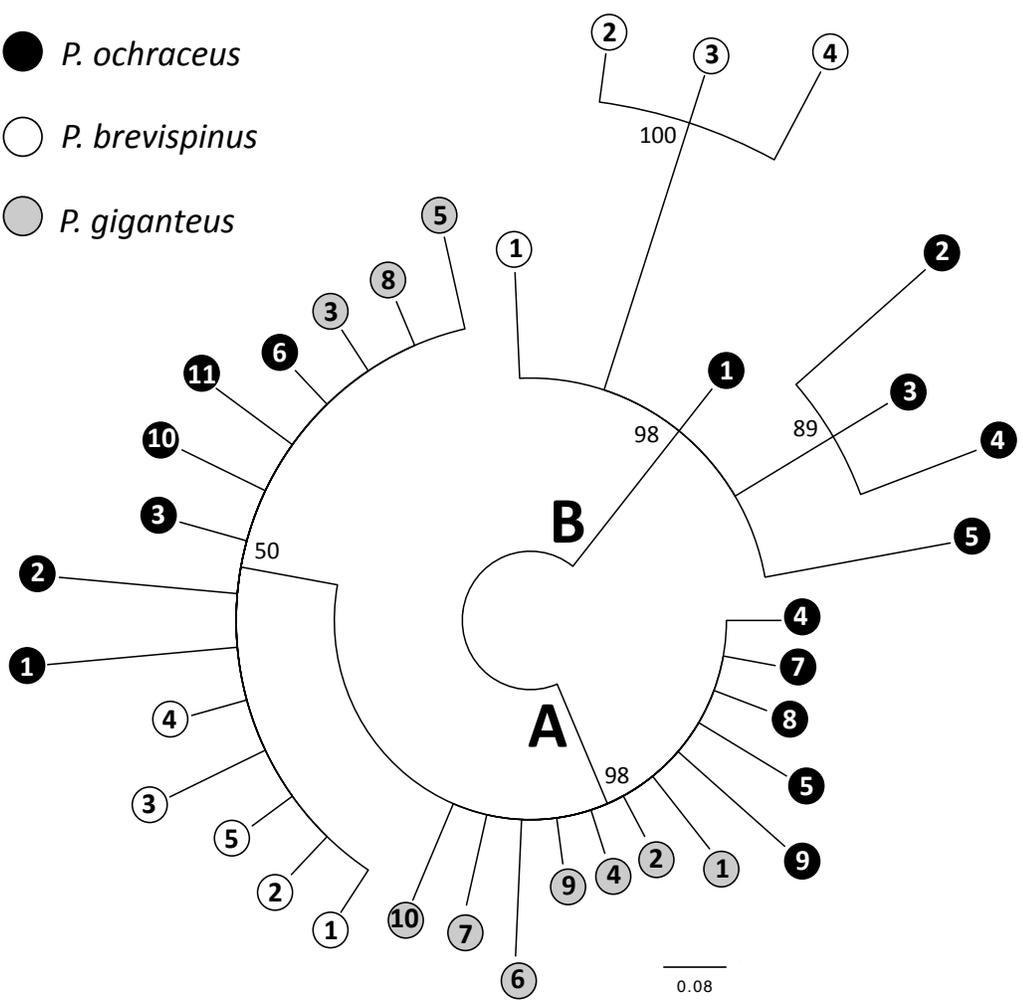


Figure 4. Plots show variation in nucleotide diversity per base pair across the coding sequence for the second bindin exon for a) *P. ochraceus* and b) *P. brevispinus* populations, using the default settings (window length=100; step size=25).

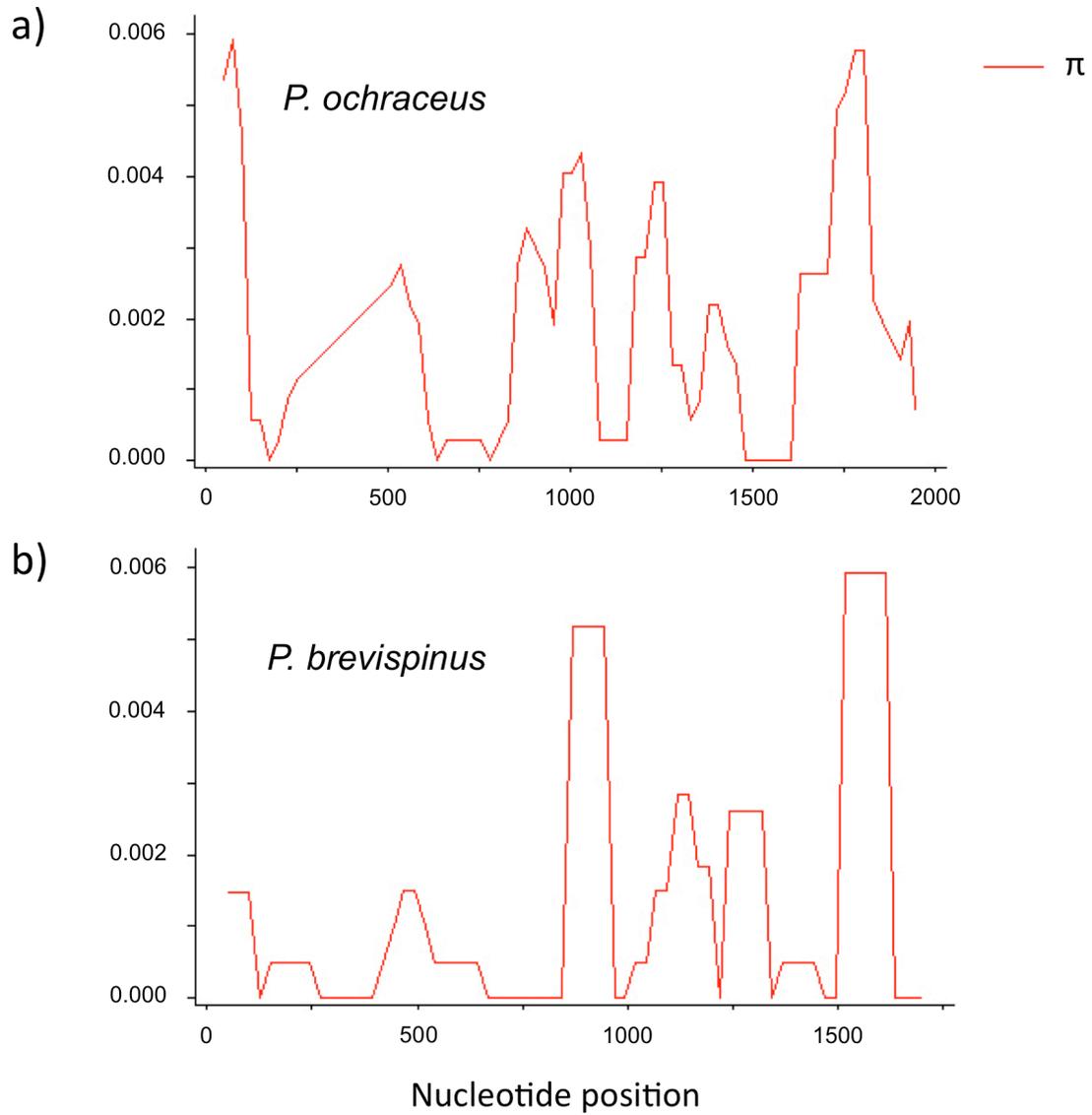
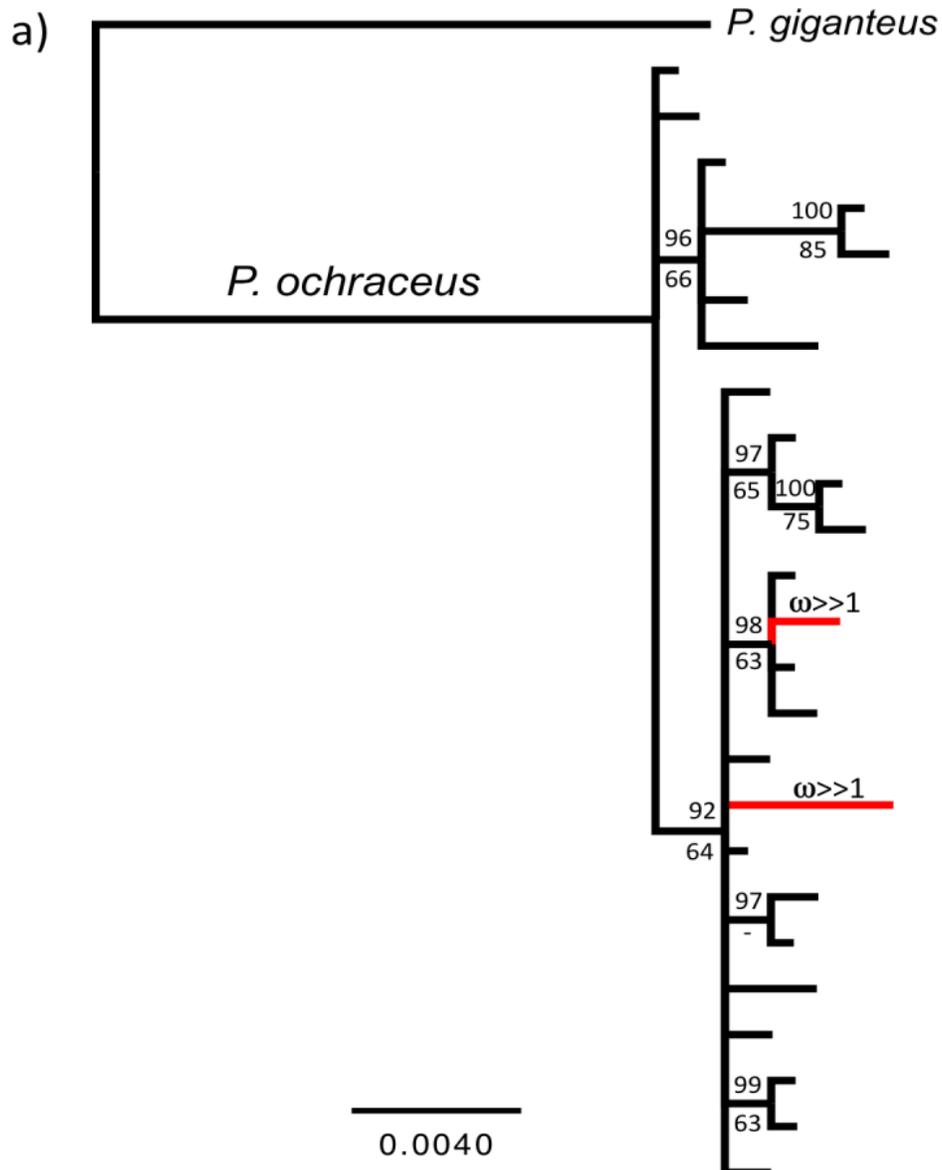


Figure 5. Consensus Bayesian genealogies of bindin alleles for a) *P. ochraceus* and b) *P. brevispinus* rooted with a single *P. giganteus* bindin allele. The scale bar shows 0.4% sequence divergence. Numbers above branches represent Bayesian posterior probabilities of partition and numbers below the branches denote bootstrap support from 1000 replicates. Red branches were inferred to be experiencing episodic diversifying selection with $\omega \gg 1$, using the branch-site REL method implemented in HyPhy. Also refer to Table 4 and 5 for additional tests of positive selection



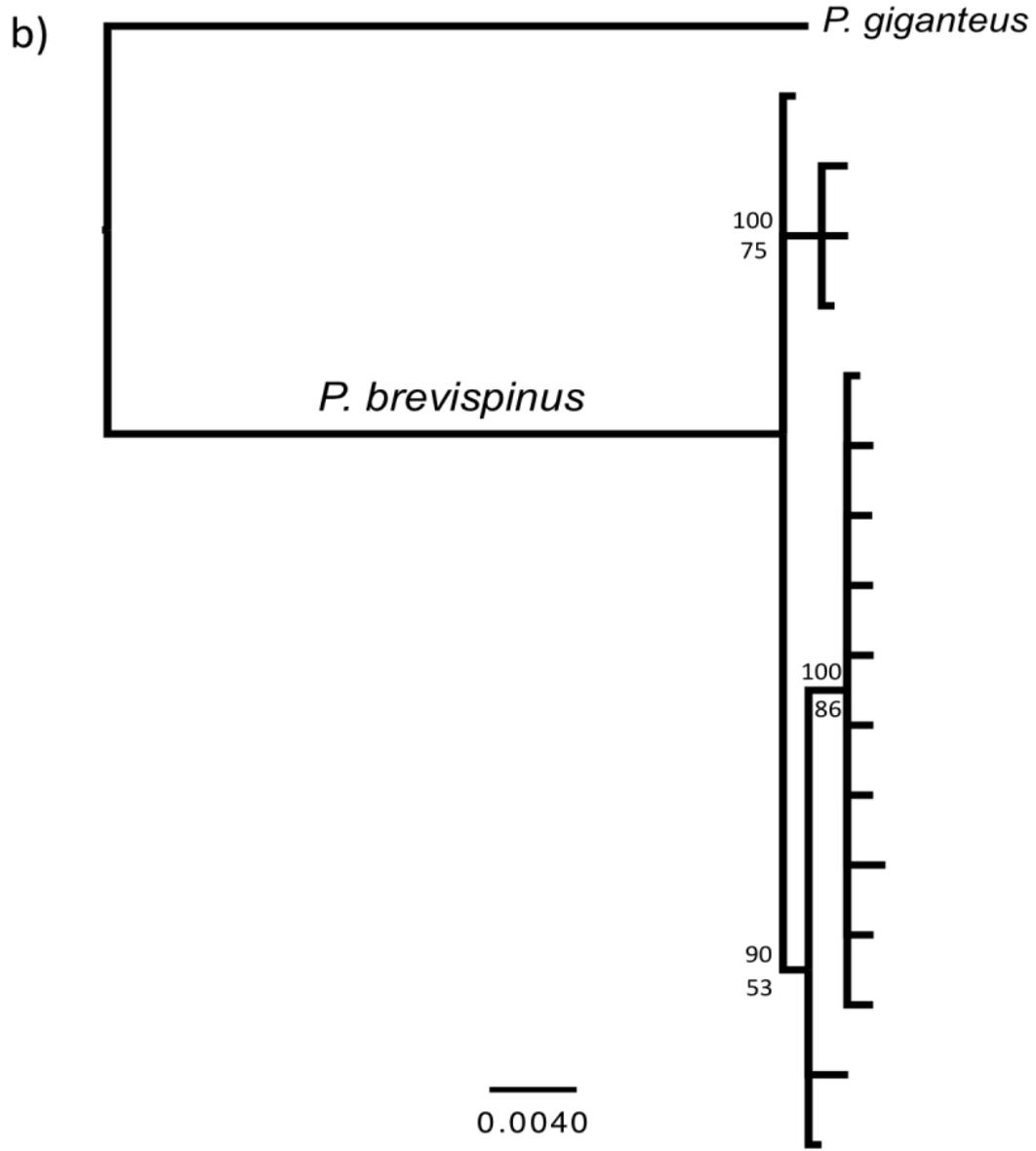
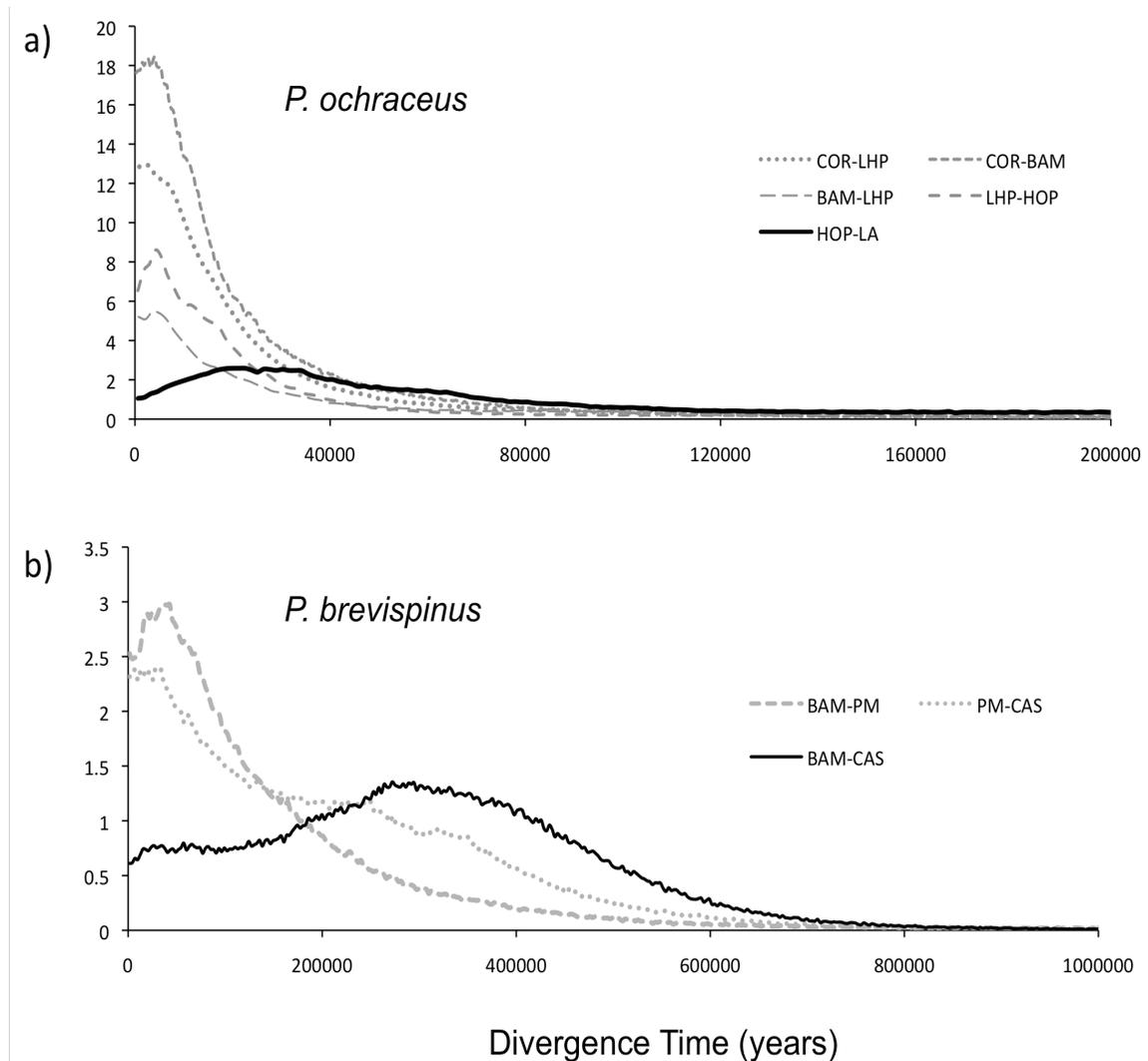


Figure 6. Posterior probability distributions of divergence times for bindin and COI in demographic units (years) between a) *P. ochraceus* population pairs; b) *P. brevispinus* population pairs. The mode of each posterior distribution indicates the maximum likelihood estimate for divergence time between each population pair. Note the wider prior distribution and older divergence times for *P. brevispinus*.



Chapter 3.

Future Directions

Comparative genetic studies in simple broadcast spawning mating systems have helped to improve our understanding of the molecular evolution of genes involved in sperm-egg compatibility and the evolutionary processes implicated in their divergence. Codon models have been successful in detecting site-specific signatures of positive selection, and identifying potentially functional regions of proteins involved in fertilization specificity without any prior molecular knowledge of the importance of these candidate regions (e.g., Sunday and Hart 2013). However, it is likely that codon model tests that are based on more conventional models of molecular evolution (ie. nucleotide substitutions), may overlook important cases of adaptation involving the processes of concerted evolution and tandem duplication of repetitive gene regions that appear to be a common phenomenon among proteins involved in reproduction and may have important functional effects on sperm-egg compatibility (e.g., Biermann 1998; Galindo et al. 2002; Fiebig et al. 2004). A better understanding of the evolutionary consequences of these molecular processes on within-species polymorphism will require estimating rates of gene conversion and crossover events from existing genetic variation (e.g., Johanneson et al. 2005; Gay et al. 2007) and developing phylogeny-based tests of positive selection that account for evolutionary change through the processes of concerted evolution. Furthermore, incorporating the mechanisms of interallelic gene conversion and chromosomal crossing-over into models of evolutionary change that describe how alleles change over time, and into coalescent population genetic models (e.g., Wiuf and Hein 2000), may inform more accurate inferences of genealogical history and locus-specific rates of gene flow (e.g., Sousa et al. 2013).

Despite tremendous advances in statistical tools and our general understanding of the evolution of fertilization, most studies, along with the work of this thesis, have focused on examining variation in single loci or only handfuls of genes that mediate sperm and egg molecular interactions. For example, other peptides that function at different stages of sperm and egg recognition, such as sperm receptors that bind to egg-

expressed pheromones during gamete chemoattraction, egg extracellular matrix glycoproteins and sperm receptors that induce the acrosomal reaction in the sperm head, other acrosomal proteins involved in egg vitelline layer lysis, and proteins that function in the egg vitelline layer to form the fertilization envelope, have been well studied by cell and developmental biologists (reviewed in Hirohashi et al. 2008; Vacquier 2012), but are still relatively unexplored by evolutionary ecologists (reviewed in Vacquier and Swanson 2011; also see Galindo et al. 2003; Pujolar and Pogson 2011; Hart 2012; Hart 2013). Evidence of positive selection on these genes may indicate evolutionary responses mediating sperm competition, fertilization rate or other outcomes of mating interactions (Evans and Sherman 2013). However, the relative importance of these other potential targets of selection for fertilization specificity is largely unknown, and calls for population genetic analyses that include more complete sets of interacting compatibility genes (Pujolar and Pogson 2011; Hart and Foster 2013).

A significant obstacle for identifying candidate genes involved in different stages of the fertilization process, and for studying the effects of selection on specific receptor-ligand pairs, has been the limited capacity of traditional PCR and targeted gene methods (Pujolar and Pogson 2011; Vacquier 2012; Hart 2012). For example, molecular variation in the large glycoprotein egg receptor for bindin (EBR1) in sea urchins and sea stars has been particularly difficult to study, because its long and complex gene structure that is made up of highly repetitive sequence domains and more than 50 exons, is not easily accessible by standard PCR methods, which favor relatively short amplicons of only a few thousand base pairs (Kamei and Glabe 2003; Pujolar and Pogson 2011; Hart 2013). Furthermore, other gamete recognition genes may have several paralogous copies (e.g., Hart 2012) or multiple isoforms (e.g., Moy et al. 2008; Hart and Foster 2013) that complicate population genetic analyses and explaining the effects of genetic variation on gamete compatibility between conspecific mates. Therefore, only a small number of studies have considered the evolution of genes functioning at more than one part of fertilization (e.g., Hellberg et al. 2012). Moreover, research on proteins coevolving with well studied compatibility loci, such as sea urchin bindin and mollusk lysin, have involved either comparisons of complete coding receptor sequences from one or two species (e.g., Galindo et al. 2002; Kamei and Glabe 2003) or population- and species-level analyses of sequence variation that have been limited to small portions or single exons

of the gene structure (e.g. Swanson and Vacquier 1998; Galindo et al. 2003; Clark et al. 2009; Pujolar and Pogson 2011).

Genomic and proteomic approaches that utilize high-throughput sequencing methods are becoming an increasingly accessible way to characterize and analyze variation in full-length coding sequences of multiple relevant compatibility genes for larger samples of populations and species (Dorus et al. 2006; Findlay and Swanson 2010; Hart and Foster 2013; Hart 2013). Comparative analyses of population variation in pairs of coevolving male and female loci are an important avenue of future research that could have profound implications for understanding among-locus variation in population genetic responses to selection and the effects on rates of gene flow specific to these parts of the genome involved in reproduction. Analyses of sequence data from next-generation sequencing methods will also be an effective way to identify larger sets of phylogeographic marker loci for more powerful estimates of demographic parameters, and differences in these estimates when compared to loci expected to be under selection. Similarly, exploring whether signatures of adaptive evolution are restricted to molecules involved in gamete recognition, or if such population genetic responses to selection are characteristic of all gonad-expressed genes, such as egg yolk genes not directly involved in fertilization (e.g., Prowse and Bryne 2012), will be useful in evaluating the relative contribution of specific compatibility genes to genomic divergence and reproductive barriers in sympatry (Lessios 2011).

While such genomic advances may improve our ability to identify interesting genes and rapidly evolving genomic regions, resolving outstanding questions about when functional divergence in reproductive genes takes place and their role in incipient speciation will require knowledge of the adaptive relevance of reproductive protein evolution on gamete compatibility. In particular, it will be important to determine if concerted evolution of repeating domains is involved in adaptive changes in sperm-egg compatibility and if there is selection for length variation in highly repetitive gamete recognition genes such as *bindin*. An integrative approach that combines population genetic studies of naturally occurring genetic variation, and functional experiments that analyze sperm-egg compatibility between known male and female genotypes under laboratory conditions of sperm competition and polyspermy risk will be particularly effective for measuring how selection acts on reproductive compatibility among natural

populations (Turner and Hoekstra 2008). Therefore, future endeavors in reproductive protein evolution and sympatric divergence will require a shift in research from comparisons among lineages that have already speciated, to the processes occurring among diverging populations within species (Palumbi 2009; Bird et al. 2012). An improved understanding of the functional effects of genetic variation on sperm-egg compatibility will not only be important for identifying commonalities in reproductive evolution among diverse mating systems and taxonomic groups, but it will be fundamental for understanding the potential for ongoing coevolution between male and female compatibility proteins to lead to reproductive divergence and the evolution of new species in sympatry.

References

- Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18:1585-1592.
- Babcock, R., and J. Keesing. 1999. Fertilization biology of the abalone *Haliotis laevis*: laboratory and field studies. *Can. J. of Fish. Aquat. Sci.* 56:1668-1678.
- Biermann, C. H. 1998. The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: *Strongylocentrotidae*). *Mol. Biol. Evol.* 15:1761-1771.
- Binks, R. M., J. Prince, J. P. Evans, and W. J. Kennington. 2012. More than bindin divergence: reproductive isolation between sympatric subspecies of a sea urchin by asynchronous spawning. *Evolution*. 66:3545-3557.
- Bird, C. E., B. S. Holland, B. W. Bowen, and R. J. Toonen. 2011. Diversification of sympatric broadcast-spawning limpets (*Cellana spp.*) within the Hawaiian archipelago. *Mol. Ecol.* 20:2128-2141.
- Bird, C. E., I. Fernandez-Silva, D. J. Skillings, and R. J. Toonen. 2012. Sympatric speciation in the post “modern synthesis” era of evolutionary biology. *Evol. Biol.* 39:158-180.
- Calderón, I., X. Turon, and H. A. Lessios. 2009. Characterization of the sperm molecule bindin in the sea urchin genus *Paracentrotus*. *J. Mol. Evol.* 68:366-376.
- Calderón, I., C. R. R. Ventura, X. Turon, and H. A. Lessios. 2010. Genetic divergence and assortative mating between colour morphs of the sea urchin *Paracentrotus gaimardi*. *Mol. Ecol.* 19:484-493.
- Carmon, A., M. Wilkin, J. Hassan, M. Baron, and R. MacIntyre. 2007. Concerted evolution within the *Drosophila dumpy* gene. *Genetics* 176:309-325.
- Carmon, A., M. Larson, M. Wayne, and R. MacIntyre. 2010. The rate of unequal crossing over in the *dumpy* gene from *Drosophila melanogaster*. *J. Mol. Evol.* 70:260-265.
- Chapman, T., G. Arnqvist, J. Bangham, and L. Rowe. 2003. Sexual conflict. *Trends in Ecol. Evol.* 18:41-47.
- Christen, R. 1985. Isolation of acrosomal vesicles and their surrounding membranes from starfish sperm. *Dev. Growth Differ.* 27:529-538.
- Clark, N. L., J. E. Aagaard, and W. J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. *Repro.* 131:11-22.
- Clark, N. L., G. D. Findlay, X. Yi, M. J. MacCoss, and W. J. Swanson. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. *Mol. Biol. Evol.* 24:2081-2090.

- Clark, N. L., J. Gasper, M. Sekino, S. A. Springer, C. F. Aquadro, W. J Swanson. 2009. Coevolution of interacting fertilization proteins. *PLoS Genet.* 5:e1000570.
- Clement, M., D. Posada, K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9:1657–1659.
- Crespi, B., and P. Nosil. 2013. Conflictual speciation: species formation via genomic conflict. *Trends Ecol. Evol.* 28:48-57.
- Coyne, J. A., H. A. Orr. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
- Dawson, M. N. 2001. Phylogeography in coastal marine animals: a solution from California? *J. Biogeogr.* 28:723-736.
- Delport, W., A. F. Poon, S. D. Frost, and S. L. K. Pond. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics.* 26:2455-2457.
- Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. *Amer. Nat.* 312-321.
- Dorus, S., S. A. Busby, U. Gericke, J. Shabanowitz, D. F. Hunt, and T. L. Karr. 2006. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat. Gen.* 38:1440-1445.
- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299:111-117.
- Durfy, S. J., and H. F. Willard. 1989. Patterns of intra-and interarray sequence variation in alpha satellite from the human X chromosome: evidence for short-range homogenization of tandemly repeated DNA sequences. *Genomics* 5:810-821.
- Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Elder, Jr. J. F., and B. J. Turner. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 297-320.
- Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. Wolf. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756-760.
- Evans, J. P., and C. D. Sherman. 2013. Sexual Selection and the Evolution of Egg-Sperm Interactions in Broadcast-Spawning Invertebrates. *Biol. Bull.* 224:166-183.
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* 1:47.
- Farmanfarmaian, A., A. C. Giese, R. A. Boolootian, and J. Bennett. 1958. Annual reproductive cycles in four species of west coast starfishes. *J. Exp. Zool.* 138:355-367.
- Feder, J. L., and P. Nosil. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64:1729-1747.

- Fiebig, A., R. Kimport, and D. Preuss. 2004. Comparisons of pollen coat genes across Brassicaceae species reveal rapid evolution by repeat expansion and diversification. *Proc. Nat. Acad. Sci. USA* 101:3286-3291.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368-376.
- Frank, S. A. 2000. Sperm competition and female avoidance of polyspermy mediated by sperm-egg biochemistry. *Evol. Ecol. Res.* 2:613-625.
- Galindo, B. E., G. W. Moy, W. J. Swanson, and V. D. Vacquier. 2002. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene* 288:111-117.
- Galindo, B. E., V. D. Vacquier, and W. J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. *Proc. Nat. Acad. Sci. USA.* 100:4639-4643.
- Gao, B., L. E. Klein, R. J. Britten, and E. H. Davidson. 1986. Sequence of messenger RNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization. *Proc. Nat. Acad. Sci. USA.* 83:8634-8638.
- Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* 403:886-889.
- Gavrilets, S., and D. Waxman 2002. Sympatric speciation by sexual conflict. *Proc. Nat. Acad. Sci. USA.* 99:10533-10538.
- Gavrilets, S., and T. I. Hayashi. 2005. Speciation and sexual conflict. *Evol. Eco.* 19:167-198.
- Gay, J., S. Myers, and G. McVean. 2007. Estimating meiotic gene conversion rates from population genetic data. *Genetics* 177:881-894.
- Geyer, L. B., and S. R. Palumbi. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* 57:1049-1060.
- Geyer, L. B., and S. R. Palumbi. 2005. Conspecific sperm precedence in two species of tropical sea urchins. *Evolution* 59:97-105.
- Geyer, L. B., and H. A. Lessios. 2009. Lack of character displacement in the male recognition molecule, bindin, in Atlantic sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* 26:2135-2146.
- Goujon M, H. McWilliam , W. Li, F. Valentin, S. Squizzato , J. Paern, and R. Lopez. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nuc. Acid. Res.* 2010 Jul, 38 Suppl: W695-9 doi: 10.1093/nar/gkq313
- Gould, M. C., and J. L. Stephano. 2003. Polyspermy prevention in marine invertebrates. *Micros. Res. Tech.* 61:379-388.
- Grosberg, R. K., D. R. Levitan, and B. B. Cameron. 1996. Characterization of genetic structure and genealogies using RAPD-PCR markers: a random primer for the novice and nervous. Pp. 67-100 in J. D. Ferraris and S. R. Palumbi, eds. *Molecular zoology: Advances, strategies, and protocols.* Wiley-Liss, New York.
- Harley, C. D. G., M. S. Pankey, J. P. Wares, R. K. Grosberg, and M. J. Wonham, 2006. Color polymorphism and genetic structure in the sea star *Pisaster ochraceus*. *Biol. Bull.* 211:248-262.

- Hart, M.W., and P. B. Marko. 2010. It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. *Integr. Comp. Biol.* 50:643–661
- Hart, M. W. 2012. Next-generation studies of mating systems evolution. *Evolution* 66:1675-1680.
- Hart, M. W., I. Popovic, and R. B. Emler. 2012. Low rates of bindin codon evolution in lecithotrophic *Heliocidaris* sea urchins. *Evolution* 66:1709-1721.
- Hart, M. W. 2013. Structure and evolution of the sea star egg receptor for sperm bindin. *Mol. Ecol.* 22:2143–2156
- Hart, M.W., and A. Foster. 2013. Highly expressed genes in gonads of the bat star *Patiria miniata*: gene ontology, expression differences, and gamete recognition loci. *Invertebr. Biol.* 132:241-250.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412: 904-907.
- Haygood, R. 2004. Sexual conflict and protein polymorphism. *Evolution* 58:1414-1423.
- Heger, A. and L. Holm. 2000. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* 41:224-237. doi: 10.1002/1097-0134(20001101)41:2<224::AID-PROT70>3.0.CO;2-Z
- Hellberg, M. E., and V. D. Vacquier. 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Mol. Biol. Evol.* 16:839-848.
- Hellberg, M. E., A. B. Dennis, P. Arbour-Reily, J. E. Aagaard, and W. J. Swanson. 2012. The Tegula tango: a coevolutionary dance of interacting, positively selected sperm and egg proteins. *Evolution*, 66:1681-1694.
- Hey, J, and R. Nielsen. 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, and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Nat. Acad. Sci. USA.* 104:2785–2790
- Hickerson, M. J., E. A. Stahl, and H. A. Lessios. 2006. Test for simultaneous divergence using approximate Bayesian computation. *Evolution.* 60:2435-2453.
- Hirohashi, N., N. Kamei, H. Kubo, H. Sawada, M. Matsumoto, and M. Hoshi. 2008. Egg and sperm recognition systems during fertilization. *Dev. Growth. Differ.* 50:S221-S238.
- Johannesson, H., J. P. Townsend, C. Y. Hung, G. T. Cole, and J. W. Taylor. 2005. Concerted evolution in the repeats of an immunomodulating cell surface protein, *SOWgp*, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. *Genetics* 171:109-117.

- Johannesson, K., S. H. Saltin, I. Duranovic, J. N. Havenhand, and P. R. Jonsson. 2010. Indiscriminate males: mating behaviour of a marine snail compromised by a sexual conflict? *PLoS ONE* 5:e12005.
- Jordan, G., and N. Goldman. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol.* 29:1125-1139.
- Kamei, N., and C. G. Glabe. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes Dev.* 17:2502-2507.
- Kelly, R. P., and S. R. Palumbi. 2010. Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One.* 5:e8594.
- Knowlton, N., and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 265:2257–2263.
- Kronforst, M. R., L. G. Young, L. M. Blume, L. E. and Gilbert. 2006. Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution* 60:1254-1268.
- Krug, P. J. 2011. Patterns of speciation in marine gastropods: A review of the phylogenetic evidence for localized radiations in the sea. *Amer. Malacol. Bull.* 29:169–186.
- Kosakovsky Pond, S. L., B. Murrell, M. Fourment, S. D. W. Frost, W. Delport, and K. Scheffler. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 24:1-13.
- Kryazhimskiy, S., and J. B. Plotkin. 2008. The population genetics of dN/dS . *PLoS Genet.* 4:e1000304.
- Kuhner, M. K. 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol. Evol.* 24:86-93
- Lambert, P. 2000. *Sea Stars of British Columbia, Southeast Alaska, and Puget Sound.* UBC Press.
- Lessios, H. A., B. D. Kessing, and J. S. Pearse. 2001. Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* 55:955-975.
- Lessios, H. A. 2007. Reproductive isolation between species of sea urchins. *Bull. Mar. Sci.* 81:191-208.
- Lessios, H. A. 2011. Speciation genes in free-spawning marine invertebrates *Integr. Comp. Biol.* 51:456-465.
- Lessios, H. A., S. Lockhart, R. Collin, G. Sotil, P. Sanchez-Jerez, K. S. Zigler, A. F. Perez, M. J. Garrido, L. B. Geyer, G. Bernardi, V. D. Vacquier, R. Haroun, and B. D Kessing. 2012. Phylogeography and bindin evolution in *Arbacia*, a sea urchin genus with an unusual distribution. *Mol. Ecol.* 21:130-144.
- Levitan, D. R. 1993. The importance of sperm limitation to the evolution of egg size in marine invertebrates. *Amer. Nat.* 517-536.

- Levitan, D. R. 2004. Density-dependent sexual selection in external fertilizers: variances in male and female fertilization success along the continuum from sperm limitation to sexual conflict in the sea urchin *Strongylocentrotus franciscanus*. *Amer. Nat.* 164:298-309.
- Levitan, D. R. 2006. The relationship between egg size and fertilization success in broadcast-spawning marine invertebrates. *Integr. Comp. Biol.* 46:298-311.
- Levitan, D. R., and D. L. Ferrell. 2006. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. *Science*. 312:267-269.
- Levitan, D. R. 2008. Gamete traits influence the variance in reproductive success, the intensity of sexual selection, and the outcome of sexual conflict among congeneric sea urchins. *Evolution* 62:1305-1316.
- Levitan, D. R., and A. P. Stapper. 2009. Simultaneous positive and negative frequency-dependent selection on sperm binding, a gamete recognition protein in the sea urchin *Strongylocentrotus purpuratus*. *Evolution* 64:785-797.
- Love, A. C. 2009. Marine invertebrates, model organisms, and the modern synthesis: Epistemic values, evo-devo, and exclusion. *Theor. Biosci.* 128, 19–42.
- Löytynoja, A., and N. Goldman. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632-1635.
- Maan, M. E., and O. Seehausen. 2011. Ecology, sexual selection and speciation. *Ecol. Lett.* 14:591-602.
- Mah, C., and D. Foltz. 2011. Molecular phylogeny of the Forcipulatacea (Asterozoa: Echinodermata): systematics and biogeography. *Zool. J. Linn. Soc.* 162:646-660.
- Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Mol. Biol. Evol.* 19:2005-2021.
- Marko, P. B. 2004. 'What's larvae got to do with it?' Contrasting patterns of post-glacial population structure in two benthic marine gastropods with identical larval dispersal capabilities. *Mol. Ecol.* 13:597–611
- Marko, P. B. 2005. An intraspecific comparative analysis of character divergence between sympatric species. *Evolution* 59:554-564.
- Marko, P. B., J. M. Hoffman, S. A. Emme, T. M. McGovern, C. C. Keever, and N. L. Cox. 2010. The 'Expansion–Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? *Mol. Ecol.* 19:146-169.
- Marko, P. B. and M. W. Hart. 2011. The complex analytical landscape of gene flow inference. *Trends Ecol. Evol.* 26:448-456
- Martin, O. Y., and D. J. Hosken. 2004. Reproductive consequences of population divergence through sexual conflict. *Curr. Biol.* 14:906-910.
- Maroja, L. S., J. A. Andres, and R. G. Harrison. 2009. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*. 63:2999–3015.
- Mayr, E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

- Mayr, E. 1954. Geographic speciation in tropical echinoids. *Evolution* 8:1–18.
- McAllister, B. F., and J. H. Werren. 1999. Evolution of tandemly repeated sequences: What happens at the end of an array? *J. Mol. Evol.* 48:469-481.
- McCartney, M. A., H. A. Lessios. 2002. Quantitative analysis of gamete incompatibility between closely related species of neotropical sea urchins. *Biol. Bull.* 202:166-181.
- McCartney, M. A., and H. A. Lessios. 2004. Adaptive evolution of sperm binding tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* 21:732-745.
- McGovern, T. M, C. C. Keever, M. W. Hart, C. Sasaki, and P. B. Marko. 2010. Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Mol. Ecol.* 19:5043–5060
- Meeds, T., E. Lockard, and B. T. Livingston. 2001. Special evolutionary properties of genes encoding a protein with a simple amino acid repeat. *J. Mol. Evol.* 53:180-190.
- Menge, B. A. 1975. Brood or broadcast? The adaptive significance of different reproductive strategies in the two intertidal sea stars *Leptasterias hexactis* and *Pisaster ochraceus*. *Mar. Biol.* 31:87-100.
- Messina, F. J., and J. C. Jones. 2011. Inheritance of traits mediating a major host shift by a seed beetle, *Callosobruchus maculatus* (Coleoptera: Chrysomelidae: Bruchinae). *Ann. Entomol. Soc. Am.* 104:808–815.
- Metz, E. C, R. E. Kane, H. Yanagimachi, and S. R. Palumbi. 1994. Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. *Biol. Bull.* 187:23-34.
- Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. *Mol. Biol. Evol.* 13:397-406.
- Metz EC, Robles-Sikisaka R, Vacquier VD .1998a. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proc. Nat. Acad. Sci. USA* 95:10676–10681.
- Metz, E. C., G. Gómez-Gutiérrez, and V. D. Vacquier. 1998b. Mitochondrial DNA and binding gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Mol. Biol. Evol.* 15:185-195.
- Michel, A. P., S. Sim, T. H. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. *Proc. Nat. Acad. Sci. USA* 107:9724-9729.
- Minor, J. E., D. R. Fromson, R. J. Britten, and E. H. Davidson. 1991. Comparison of the binding proteins of *Strongylocentrotus franciscanus*, *Strongylocentrotus purpuratus*, and *Lytechinus variegatus*: sequences involved in the species specificity of fertilization. *Mol. Biol. Evol.* 8:781– 795.

- Moy, G. W., S. A. Springer, S. L. Adams, W. J. Swanson, and V. D. Vacquier. 2008. Extraordinary intraspecific diversity in oyster sperm bindin. *Proc. Nat. Acad. Sci. USA*. 105:1993-1998.
- Murrell, B., J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. Kosakovsky Pond. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8:e1002764.
- Nielsen, R. and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896
- Nosil, P., and S. M. Flaxman. 2011. Conditions for mutation-order speciation. *Proc. R. Soc. B. Biol. Sci.* 278(1704), 399-407.
- Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* 24:145-156.
- Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. *Trends Ecol. Evol.* 26:160-167.
- Nosil, P. 2012. *Ecological speciation*. Oxford University Press.
- Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. B. Biol. Sci.* 367:332-342.
- Paine, R. T. 1974. Intertidal community structure. *Oecologia*, 15: 93-120.
- Palumbi, S. R. 1994. Genetic divergence, reproductive isolation and marine speciation. *Annu. Rev. Ecol. Evol. Syst.* 25:547-572
- Palumbi, S. R. 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Nat. Acad. Sci. USA*. 96:12632-12637.
- Palumbi, S. R. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity*. 102:66-76.
- Pankey, M. S., and J. P. Wares. 2009. Overdominant maintenance of diversity in the sea star *Pisaster ochraceus*. *J. Evol Biol.* 22:80-87.
- Patiño, S., J. E. Aagaard, M. J. MacCoss, W. J. Swanson, and M. W. Hart. 2009. Bindin from a sea star. *Evol. Dev.* 11:376-381.
- Paysuer, B. A., and M. W. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biol. J. Linn. Soc.* 84:523-534.
- Pinho, C., and J. Hey. 2010. Divergence with gene flow: models and data. *Annu. Rev. Ecol. Evol. Syst.* 41:215-230.
- Prowse, T. A., and M. Byrne. 2012. Evolution of yolk protein genes in the Echinodermata. *Evol. Dev.* 14:139-151.
- Pujolar, J. M., and G. H. Pogson. 2011. Positive Darwinian selection in gamete recognition proteins of *Strongylocentrotus* sea urchins. *Mol. Ecol.* 20:4968–4982.
- Puritz, J. B., C. C. Keever, J. A. Addison, M. Byrne, M. W. Hart., R. K. Grosberg, and R. J. Toonen. 2012. Extraordinarily rapid life-history divergence between *Cryptasterina* sea star species. *Proc. R. Soc. B. Biol. Sci.* 279: 3914-3922.

- Rambaut, A. 2002. Se–Al Sequence Alignment Editor v2.0 Software. Department of Zoology, University of Oxford, Oxford, UK.
- Rambaut, A., and A. J. Drummond. 2007. Tracer v1. 4.
- Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* 1637-1653.
- Rice, W. R., and B. Holland. 1997. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol.* 41:1-10.
- Riginos, C., D. Wang, and A. J. Abrams. 2006. Geographic variation and positive selection on M7 lysin, an acrosomal sperm protein in mussels (*Mytilus spp.*). *Mol. Biol. Evol.* 23:1952-1965.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.
- Rundle, H. D., and P. Nosil. 2005. Ecological speciation. *Ecol. Lett.* 8:336-352.
- Schluter, D. 2001. Ecology and the origin of species. *Trends Ecol. Evol.* 16:372-380.
- Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737-741.
- Schwarz, D., H. Robertson, J. Feder, K. Varala, M. Hudson, G. Ragland, G.J, Hahn D.A and S. Berlocher. 2009. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics* 10:633.
- Servedio, M. R., and M. A. Noor. 2003. The role of reinforcement in speciation: theory and data. *Annu. Rev. Ecol. Evol. Syst.* 339-364.
- Servedio, M. R., G. Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: ‘magic’ but not rare? *Trends Ecol. Evol.* 26:389-397.
- Slatkin, M. 1985. Gene flow in natural populations. *Annu. Rev. Ecol. Evol. Syst.* 16:393-430.
- Smadja, C., and R. K. Butlin. 2008. On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity* 102:77-97.
- Smadja, C. M., and R. K. Butlin. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* 20:5123-5140.
- Smith, L. S. 1961. Clam-digging behavior in the starfish, *Pisaster brevispinus* (Stimpson 1857). *Behaviour* 148-153.
- Smith, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* 191:528-535.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295-315.

- Sotka, E. E., J. P. Wares, J. A. Barth, R. K. Grosberg, R. K., and S. R. Palumbi. 2004. Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Mol. Ecol.* 13:2143-2156.
- Sousa, V. C., M. Carneiro, N. Ferrand, and J. Hey. 2013. Identifying loci under selection against gene flow in isolation-with-migration models. *Genetics*, 194:211-233.
- Springer, S. A., and B. J. Crespi. 2007. Adaptive gamete recognition divergence in a hybridizing *Mytilus* population. *Evolution* 61:772-783.
- Springer, S. A., B. J. Crespi, and W. J. Swanson. 2011. Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture. *Mol. Ecol.* 20:2240-2257.
- Strathmann, M. F. 1987. Reproduction and development of marine invertebrates of the northern Pacific coast: data and methods for the study of eggs, embryos, and larvae. University of Washington Press.
- Stephan, W. 1986. Recombination and the evolution of satellite DNA. *Genet. Res.* 47:167-174.
- Stephano, J. L. 1992. A study of polyspermy in abalone. Abalone Fisheries, Biology and Culture. Fishing News Books, Cambridge, UK
- Stickle, W. B., D. W. Foltz, M. Katoh, and H. L. Nguyen. 1992. Genetic structure and mode of reproduction in five species of sea stars (Echinodermata: Asteroidea) from the Alaskan coast. *Can. J. Zool.* 70:1723-1728.
- Sunday, J. M., and M. W. Hart. 2013. Sea star populations diverge by positive selection at a sperm-egg compatibility locus. *Ecol. Evol.* doi:10.1002/ece3.487
- Swanson, W. J., and V. D. Vacquier. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* 281:710-712.
- Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3:137-144.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genet.* 123:585-595.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731-2739.
- Tomaiuolo, M., T. F. Hansen, and D. R. Levitan. 2007. A theoretical investigation of sympatric evolution of temporal reproductive isolation as illustrated by marine broadcast spawners. *Evolution* 61:2584-2595.
- Tomaiuolo, M., and D. R. Levitan. 2010. Modeling how reproductive ecology can drive protein diversification and result in linkage disequilibrium between sperm and egg proteins. *Amer. Nat.* 176:14-25.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3:e285.
- Turner, L. M., and H. E. Hoekstra. 2008. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev.* 52:769.

- Vacquier, V. D., and G. W. Moy. 1977. Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. *Proc. Nat. Acad. Sci. USA.* 74:2456-2460.
- Vacquier, V. D. 1998. Evolution of gamete recognition proteins. *Science* 281:1995-1998.
- Vacquier, V.D and W. J. Swanson. 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb. Perspect. Biol.* doi:10.1101/cshperspect.a002931
- Vacquier, V. D. 2012. The quest for the sea urchin egg receptor for sperm. *Biochem. Biophys. Res. Commun.* 425:583-587.
- Via, S. 2009. Natural selection in action during speciation. *Proc. Nat. Acad. Sci. USA.* 106:9939-9946.
- Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B. Biol. Sci.* 367:451-460.
- Weir, B.S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Wiuf, C., and J. Hein. 2000. The coalescent with gene conversion. *Genetics*, 155:451-462.
- Won, Y. J., A. Sivasundar, Y. Wang, and J. Hey. 2005. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. *Proc. Nat. Acad. Sci. USA.* 102:6581-6586.
- Yang, Z., W. J. Swanson, and V. D. Vacquier. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* 17:1446-1455.
- Yang, Z., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* 19:49-57.
- Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107-1118.
- Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586-1591
- Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22:2472-2479
- Zigler, K. S., and H. A. Lessios. 2003. Evolution of bindin in the pantropical sea urchin *Tripneustes*: comparisons to bindin of other genera. *Mol. Biol. Evol.* 20:220-231.
- Zigler, K. S., and H. A. Lessios. 2004. Speciation on the coasts of the new world: phylogeography and the evolution of bindin in the sea urchin genus *Lytechinus*. *Evolution* 58:1225-1241.
- Zigler, K. S., M. A. McCartney, D. R. Levitan, and H. A. Lessios. 2005. Sea urchin bindin divergence predicts gamete compatibility. *Evolution* 59:2399-2404.
- Zigler, K. S. 2008. The evolution of sea urchin sperm bindin. *Inter. J. Dev. Biol.* 52:791.

Appendix A.

Supporting Information

Figure A1. Posterior probability distributions of bidirectional migration parameter estimates between a) *P. ochraceus* population pairs; b) *P. brevispinus* population pairs

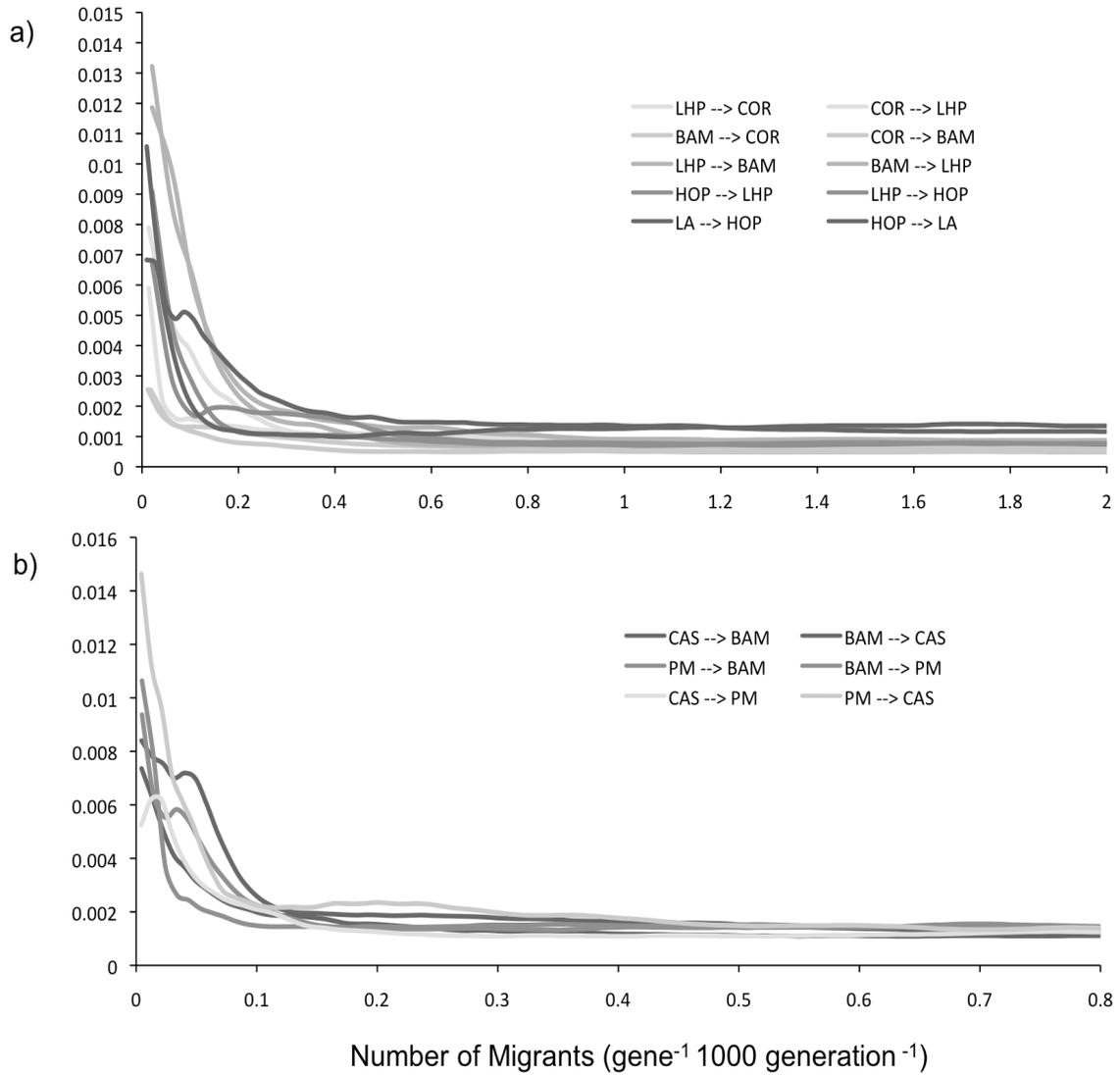


Table A1. Tajima's D statistics

<u>Species</u>	<u>Population name</u>	<u>Tajima's D</u>
<i>P. ochraceus</i>	Cordova, AK	-1.15169 NS
	Bamfield, BC	-1.06619 NS
	Lighthouse Park, BC	-1.05440 NS
	Hopkins, CA	-0.78153 NS
	La Jolla, CA	-0.90676 p=0.012

<u>Species</u>	<u>Population name</u>	<u>Tajima's D</u>
<i>P. brevispinus</i>	Bamfield, BC	0.54442 NS
	Port Moody, BC	-1.13785 NS
	SF Bay, CA	-1.63982 p=0.03

Table A2. Summary statistics for population pairwise differences (Φ_{ST}) and p-values

a) bindin

Species	Label	Population name
<i>P. ochraceus</i>	1:	Cordova, AK
	2:	Bamfield, BC
	3:	Lighthouse Park, BC
	4:	Hopkins, CA
	5:	La Jolla, CA

Population pairwise Φ_{ST}

Distance method: Jukes & Cantor

	1	2	3	4	5
1	0.00000				
2	-0.02435	0.00000			
3	0.00084	-0.04472	0.00000		
4	0.00459	-0.05023	-0.02730	0.00000	
5	-0.02088	0.02088	0.04730	0.04830	0.00000

Φ_{ST} p-values

	1	2	3	4	5
1	*				
2	0.71102+-0.0045	*			
3	0.38590+-0.0055	0.87467+-0.0033	*		
4	0.32472+-0.0046	0.97178+-0.0019	0.68439+-0.0049	*	
5	0.68973+-0.0049	0.21632+-0.0048	0.09148+-0.0029	0.10306+-0.0024	*

Species	Label	Population name
<i>P. brevispinus</i>	1:	Bamfield, BC
	2:	Port Moody, BC
	3:	SF Bay, CA

Population pairwise Φ_{ST}

Distance method: Jukes & Cantor

	1	2	3
1	0.00000		
2	0.05766	0.00000	
3	0.07935	-0.00600	0.00000

Φ_{ST} p-values

	1	2	3
1	*		
2	0.10761+-0.0028	*	
3	0.09900+-0.0029	0.43966+-0.0049	*

b) COI

Species	Label	Population name
<i>P. ochraceus</i>	1:	Cordova, AK
	2:	Bamfield, BC
	3:	Lighthouse Park, BC
	4:	Hopkins, CA
	5:	La Jolla, CA

Population pairwise Φ_{ST}

Distance method: Tamura & Nei

	1	2	3	4	5
1	0.00000				
2	0.28179	0.00000			
3	0.31160	-0.06937	0.00000		
4	0.27785	0.02405	0.23612	0.00000	
5	0.26369	-0.11117	0.02773	-0.03292	0.00000

Φ_{ST} p-values

	1	2	3	4	5
1	*				
2	0.07475+-0.0025	*			
3	0.05128+-0.0024	0.72775+-0.0043	*		
4	0.04445+-0.0022	0.39689+-0.0045	0.11553+-0.0031	*	
5	0.14375+-0.0033	0.99990+-0.0000	0.44283+-0.0046	0.59895+-0.0044	*

Species	Label	Population name
<i>P. brevispinus</i>	1:	Bamfield, BC
	2:	Port Moody, BC
	3:	SF Bay, CA

Population pairwise Φ_{ST}

Distance method: Tamura & Nei

	1	2	3
1	0.00000		
2	0.01158	0.00000	
3	0.09558	0.08527	0.00000

Φ_{ST} p-values

	1	2	3
1	*		
2	0.41372+-0.0041	*	
3	0.10692+-0.0032	0.23621+-0.0042	*

Table A3. Analysis of molecular variance (AMOVA; no population grouping) summaries for *P. ochraceus* and *P. brevispinus* bindin variation:

a) bindin

P. ochraceus:

Source of variation	Sum of squares	d.f.	Variance components	% of variation
Among populations	4	8.225	-0.01208 Va	-0.55
Within populations	65	144.648	2.22535 Vb	100.55

Va P-value = 0.55545+-0.00506

P. brevispinus:

Source of variation	Sum of squares	d.f.	Variance components	% of variation
Among populations	2	3.571	0.05822 Va	5.30
Within populations	37	38.496	1.04043 Vb	94.70

Va P-value = 0.08307+-0.00258

b) COI

P. ochraceus:

Source of variation	Sum of squares	d.f.	Variance components	% of variation
Among populations	4	11.512	0.27531 Va	22.45
Within populations	30	28.524	0.95078 Vb	77.55

Va P-value = 0.00752+-0.00080

P. brevispinus:

Source of variation	Sum of squares	d.f.	Variance components	% of variation
Among populations	2	4.155	0.08296 Va	5.09
Within populations	17	26.290	1.54650 Vb	94.91

Va P-value = 0.23089+-0.00410