Species Boundaries in a Broadcast Spawning Marine Invertebrate

by Marie Louise Nydam

SPECIES BOUNDARIES IN A BROADCAST SPAWNING MARINE INVERTEBRATE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Marie Louise Nydam

August 2010

SPECIES BOUNDARIES IN A BROADCAST SPAWNING MARINE INVERTEBRATE

Marie Louise Nydam, Ph. D.

Cornell University 2010

The first chapter of the thesis contains a mitochondrial cytochrome oxidase I (mtCOI) phylogeny of shallow-water species in the genus *Ciona.* The mtCOI sequences of Northeast Pacific/Mediterranean (Type A) and Northwest Atlantic (Type B) *Ciona intestinalis* differ by ~12% and *Ciona roulei* is nested within Type B. *Ciona savignyi* differs from all other haplotypes by 13-16%. A previously undescribed but morphologically distinct *Ciona* sp. found at the Banyuls-sur-Mer site was > 10% divergent from all other haplotypes.

The second chapter builds upon the mtCOI phylogeny and includes six nuclear genealogies for the genus *Ciona.* From these genealogies, I conclude that Type A and Type B are well-supported monophyletic groups. In spite of their morphological similarity, Type A vs. Type B divergences range from 0.035 to 0.124. In contrast, the morphologically distinct *C. roulei* is embedded within Type B in all genealogies, and *Ciona* sp. appears to be associated with Type B/*C. roulei* to the exclusion of Type A.

In the third chapter, I investigated the distribution of Type A and B in areas of potential sympatry to determine whether these two types occur together and if so, whether they show evidence of hybridization and introgression. Then I combine my data with other studies to investigate general patterns of reproductive isolation vs. divergence in marine broadcast spawners. Type A and B do occur sympatrically and their genomes show low levels of introgression. Type A and B may be near the upper

limit of the range of divergence values where introgression is still possible. However, introgression at divergence levels similar to those found in *Ciona* does occur, prompting questions about the strength of postmating prezygotic reproductive barriers in marine broadcast spawners.

In the fourth chapter, I identified three candidate sperm GRPs and used these to test whether reinforcement is occurring in this system by testing whether positive selection (as a proxy for prezygotic isolation) is stronger in sympatry than allopatry. While little evidence for reinforcement was found in these three candidate GRPs, tests such as those performed here may provide important insights into the process of speciation in marine broadcast spawners.

BIOGRAPHICAL SKETCH

Marie Louise Nydam was born in San Diego, CA, on November 19th, 1981. Marie was always interested in science, especially Marine Biology. This interest was sparked by trips to the intertidal zone with her father and sister when she was very young. She entered several projects in science fairs throughout middle school and high school, all of which involved the ecology of the intertidal zone. At La Jolla High School Marie participated in Science Olympiad, the Oceanic Club, and was captain of the National Ocean Science Bowl.

Marie entered the University of California, Davis, in 1999. In the spring of her sophomore year, she began working as a field and laboratory assistant to Randall Hughes, then a graduate student in the marine ecology laboratory of Dr. Jay Stachowicz. Marie worked with Randall Hughes until leaving for a spring semester at the Bodega Marine Laboratory her junior year. This spring semester was important in Marie's development as a scientist; it was then she realized that she was capable of doing independent research. During this semester she studied recruitment dynamics in the fouling communities of Bodega Harbor.

For the rest of her undergraduate career and afterwards as a postgraduate researcher employed by Dr. Stachowicz, Marie spent most of her time at Bodega. She continued to study the fouling communities, research that ultimately led to her first publication. She also assisted Stachowicz graduate students with their research in sea grass beds, subtidal rocky habitats, intertidal zones and mudflats. Marie was incredibly fortunate to be a part of the Stachowicz Lab; Jay, his graduate students, and postdocs treated her as a colleague and taught her how to do scientific research.

While research fouling community ecology, Marie realized that she was ultimately interested in how the species she studied had evolved to co-exist, not just

how they interacted in the present day. Her outstanding introductory evolution course with Dr. Rick Grosberg also influenced her to think about evolutionary timescales.

For graduate school, Marie decided to work with Dr. Richard Harrison on questions of speciation in a fouling community organism, the ascidian *Ciona intestinalis.* Although she has spent much of her Ph.D. at a lab bench, summer field work is still the best part about her job and she continues to be fascinated by the ecology and natural history of marine invertebrates. She will begin a postdoctoral position in August 2010, studying the population genetics of allorecognition loci in the ascidian *Botryllus schlosseri.*

To Jay Stachowicz and the members of the Stachowicz Laboratory 2001-2004:

Jarrett Byrnes

Randall Hughes

Kristin Hultgren

Amy Larson

Suzanne Olyarnik

Carol Thornber

For giving me an excellent start.

ACKNOWLEDGMENTS

I couldn't have asked for a better advisor than Dr. Rick Harrison. Rick cares a great deal for his students, both as people and as scientists. Living 3,000 miles away from friends and family was difficult especially during the long winters, and Rick's steady interest in my well-being was comforting. Graduate research can occasionally be stressful, but Rick never failed to alleviate that stress, always confidently assuring me that things would work out. I also appreciated Rick's generous financial support of my research, his willingness to give me the independence I crave, and his careful attention to my manuscripts.

I would also like to thank members of the Harrison Laboratory: Steve Bogdanowicz, Jose Andres, Ryan Thum, Sean Mullen, Erik Dopman, Mo Carter, Luana Maroja, Mo Carter, Jamie Walters, Ben Hamilton, Erica Larson, Findley Ransler and Mike Bronski. This lab group provided an incredibly supportive and intellectually stimulating environment; I am grateful to each and every member for his/her advice and friendship.

My committee members, Andy Clark and Chip Aquadro, offered both encouragement and constructive criticism throughout my Ph.D. I especially thank Andy for his continuing excitement about the *Ciona* system; every meeting with him was a pleasure.

This Ph.D. would have been substantially more difficult without logistical help from several people, including John Bishop, Nathan Clark, Aline Fiala-Medioni, Erin Newman-Smith, Laurent Leveque, Marta Pascual, the Smith Lab at UCSB, and Xavier Turon.

One of the primary reasons I came to Cornell for graduate school was the EEB graduate students I met at recruitment weekend. I've been so lucky to have such a

talented, funny, kind group of people as my friends. Special thanks to Dan Riskin, Amy Gonzales, Amelia Greiner, John Ratcliffe, Katie Wagner, Dena Vallano, Mari Kimura, Jen Moslemi, Marita Davison, Danica Lombardozzi, Dan and Krista Capps, and my first officemates in W343 Mudd (Krystal, Rulon, Jamie, Gretchen).

I would also like to thank my parents and sister, who encouraged me from the very beginning. My parents always said that they would support whichever pursuits/career I chose, as long as I was happy. My sister never fails to tell it like it is, and I am constantly appreciative of her insightful advice.

My boyfriend Josh has made my life infinitely better since he entered it. He reminds me to have balance in my life, and to embrace spontaneity. He is the reason I have maintained relatively calm while going through what is usually the stressful process of finishing and defending a thesis. I look forward to many more years of adventures with him.

My research was funded by an NSF Doctoral Dissertation Improvement Grant, the Cornell Graduate School, the Andrew W. Mellon Student Research Grant, Cornell University Sigma Xi, the American Museum of Natural History Lerner Gray Fund for Marine Research, the Cornell Ecology and Evolutionary Biology Department and the Orenstein Fund.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENEALOGICAL RELATIONSHIPS WITHIN AND AMONG SHALLOW-WATER *CIONA* SPECIES (ASCIDIACEA)

Abstract

In spite of historical and current interest in *Ciona intestinalis* and its congeners, little is known about evolutionary relationships among the members of the genus *Ciona.* Here 744-bp sequences of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene are used to examine phylogenetic relationships among three described species (*C. intestinalis, Ciona roulei, C. savignyi*) sampled from multiple coastal sites in the Northeast Pacific (California USA), Northwest Atlantic (from New Hampshire to Connecticut, USA), Northeast Atlantic (Sweden and The Netherlands), and Mediterranean (Banyuls-sur-Mer, France). The samples were collected in June-October 2005. The COI sequences of Northeast Pacific/Mediterranean (Type A) and Northwest Atlantic (Type B) *C. intestinalis* differ by ~12% and *C. roulei* is nested within Type B *C. intestinalis. C. savignyi* differs from all other haplotypes by 13-16%. A previously undescribed but morphologically distinct *Ciona* sp. found at the Banyuls-sur-Mer site was > 10% divergent from all other haplotypes. Although these data arise from a single gene study, they indicate that further elucidation of species relationships within the genus and of the species' distributions will be needed if continuing invasions and potential reproductive isolation are to be investigated.

Introduction

The ascidian *Ciona intestinalis* has served as a developmental model species

for over a century (Minganti 1948; Morgan 1904; Sordino 2000). Both *C. intestinalis* and its congener *C. savignyi* have invaded benthic communities throughout the temperate zone, necessitating expensive eradication programs (Lambert 2001; Lambert & Lambert 2003). Recently, the genome sequences of *C. intestinalis* and *C. savignyi* have become available (Dehal et al. 2002; Vinson et al. 2005), providing tools for comparative genomics and insights into vertebrate evolution (Boffelli et al. 2004; Johnson et al. 2004; Missal et al. 2005). Comparisons of the mitochondrial genomes of *C. intestinalis, C. savignyi,* and other tunicate species also have answered questions about chordate evolution (Gissi & Pesole 2003; Yokobori et al. 2003; Yokobori et al. 2005). Despite the historical and current interest in *C. intestinalis* and its congeners, we know very little about genealogical and phylogenetic relationships within and among members of this genus.

The genus *Ciona* comprises 12 known species, four found in shallow water and eight found exclusively in deep water (Harant & Vernieres 1933; Van Name 1945; Monniot & Monniot 1977; Monniot & Monniot 1983; Hoshino & Nishikawa 1985; Monniot & Monniot 1989; Monniot & Monniot 1990; Sanamyan 1998). Many of the deep water species have been described from only one specimen and most are difficult to collect. Here we focus on relationships among the common shallow-water species. *Ciona intestinalis* is thought to be endemic to Northeast Atlantic waters, where it was described (Linne 1767). It is widely distributed in the temperate zone presumably through transport in ballast water and on ship bottoms and now has a range that includes the Pacific Coasts of Japan and North America, New Zealand, Australia, and Chile, the Atlantic Coasts of North America, Europe, and South Africa, the Black Sea and the Mediterranean Sea (Kott 1952; Van Name 1945). *Ciona savignyi* (Herdman, 1882) is endemic to Japan, but has spread to the Pacific Coast of North America and the Atlantic Coast of Argentina (Hoshino & Nishikawa 1985). *Ciona roulei* (Lahille

1887) is found only in the Mediterranean Sea, where it has only been described from five sites along the coasts of France and Spain (Harant & Vernieres 1933). The fourth shallow water species, *C. edwardsi* (Roule, 1886), is rare and was not included in this study. Although the morphology of several *Ciona* species has been compared (Hoshino & Nishikawa 1985), no study thus far has examined phylogenetic relationships among these species.

Within *C. intestinalis,* two "types" have been recognized (Suzuki et al. 2005). Type A inhabits the Pacific Ocean and the Mediterranean Sea, while Type B inhabits the Atlantic Ocean. Type A and Type B are working names agreed upon by the community of *Ciona* biologists. The two types of *C. intestinalis* are partially reproductively isolated. Crosses using British Type B *C. intestinalis* eggs and Japanese Type A *C. intestinalis* sperm result in normal rates of fertilization, but the reciprocal cross yields many fewer fertilized eggs (Suzuki et al. 2005). While *C. roulei* and *C. intestinalis* are clearly distinct on morphological grounds (Harant & Vernieres 1933), crosses between these two species are also known to produce viable offspring. In particular, Mediterranean Type A *C. intestinalis* sperm x *C. roulei* eggs produced viable F1 tadpole larvae. However, fertilization fails in the reciprocal cross (Lambert et al. 1990). *C. intestinalis* and *C. savignyi* are completely reproductively isolated (Byrd & Lambert 2000).

Here we use mitochondrial DNA (mtDNA) sequence data to investigate the genealogical and phylogenetic relationships within and among *Ciona* species. In particular, we focus on the relationships and amounts of divergence between Type A and B *C. intestinalis*, and among *C. roulei, C. savignyi,* and *C. intestinalis*.

Materials and Methods

Sampling

Individuals from three described species (*C. intestinalis, C. roulei* and *C. savignyi)* and one unknown species (*Ciona* sp.) were collected by SCUBA or from docks (Table 1.1). *C. intestinalis* individuals were collected from five sites on the California Coast of North America, 12 sites on the Atlantic Coast of North America, and one site in the Northwestern Mediterranean (Banyuls-sur-mer, France). Additional individuals were obtained from two sites in the Northeast Atlantic (Fiskebäckskil, Sweden and Breskens Harbor, The Netherlands). *C. roulei* individuals were sampled from one site in the Northwestern Mediterranean (Banyuls-sur-mer, France). Individuals of a previously unidentified species (*Ciona* sp.) were also collected in the harbor at Banyuls-sur-mer. *C. savignyi* individuals were sampled from four sites on the California Coast of North America. Specimens of *C. roulei* and *C.* sp were deposited at the American Museum of Natural History. A sequence from *Halocynthia roretzi* (GenBank Accession No. NC_002177; (Yokobori et al. 1999) was used as the outgroup. Sequence data were obtained from 106 of the collected individuals, and the data set contained 107 sequences including the outgroup.

DNA extraction, amplification and sequencing

Ovaries were dissected from freshly-collected individuals, cut into several pieces, immediately preserved in DMSO (dimethyl sulfoxide), and ultimately (within 12 d) stored at -80°C until needed. We used ovaries because they are easily removed from the body and yield high-quality DNA. Samples from Sweden and The Netherlands were preserved and shipped in ethanol soon after collection. Upon arrival, the ovaries from these individuals were dissected and preserved in DMSO.

Table 1.1 Collection sites for *Ciona* species for which COI sequences were determined.

Species	Region/State or Country	Site	# of Individuals
Ciona intestinalis Type A	Northeast Pacific/CA	Newport Harbor	6
		Alamitos Bay	2
		Santa Barbara Harbor	4
		Half Moon Bay	7
		Sausalito	1
	Mediterranean/France	Banyuls-sur-Mer	7
Type B	Northwest Atlantic/NH to CT	Newcastle	2
		Gloucester	2
		Salem	2
		Beverly	3
		Winthrop	2
		Quincy	3
		Sandwich	1
		Buzzards Bay	3
		Woods Hole	4
		Westport Point	3
		Newport Shipyard	2
		Mystic	3
	Northeast Atlantic/Sweden	Fiskebäckskil	8
	Northeast Atlantic/The Netherlands	Breskens	15
Ciona roulei	Mediterranean/France	Banyuls-sur-Mer	6
Ciona savignyi	Northeast Pacific/CA	San Diego Bay	8
		Newport Harbor	1
		Santa Barbara Harbor	3
		Sausalito	6
Ciona sp.	Mediterranean/France	Banyuls-sur-Mer	2

Total DNA was extracted from the ovaries using the Qiagen DNeasy® Tissue Kit (Qiagen). Primers to amplify 856 bp of the mitochondrial cytochrome *c* oxidase subunit I (CO1) gene were developed from the consensus sequence of the published *C. savignyi* and *C. intestinalis* mitochondrial genomes (Consensus F: 5'GAGTAAGAACTGGRTGRACAGTTTAYCCTCC 3', Consensus R: 5'ATTAAAACTTAATCTAGTAAAAAGAGGRRATCAATGG 3'). PCR amplification was performed in a 20-µl total reaction volume with 2 mM $MgCl_2$, 0.2 mM dNTPs, 2 µl of 10x buffer, 0.2 µM of each primer, 0.8 U of *Taq* Polymerase (Gibco-BRL*)* and 2 µl of template DNA. The PCR protocol was as follows: 35 cycles (95ºC for 50 s, 48ºC for 1 min and 72 ºC for 1 min) and a final extension step at 72 ºC for 7 min, on an OmniGene (Hybaid) thermal cycler. PCR products (diluted up to 1:20 depending on the intensity of the PCR band on an agarose gel) were incubated with 1µl each of Exonuclease I and Antarctic Phosphatase at 37°C for 30 min, followed by 90°C for 10 min. The products were purified on Sephadex® columns (Sigma-Aldrich*)*. The purified product was sequenced with a Big Dye Terminator Cycle sequencing kit and an ABI-3700 automated sequencer (Applied Biosystems) using the primers listed above. All unique haplotypes (56 sequences) have been submitted to GenBank (Accession Numbers EF209056-EF209110).

Alignment

Sequences were edited and trimmed to 744 bp with SeqMan (DNASTAR) and aligned using the Clustal-W algorithm in MEGA 3.1 (Kumar et al. 2004). Alignments were confirmed visually, and no gaps were present in the final alignment. There is no evidence for nuclear pseudogenes in our dataset. Direct sequencing of PCR products yielded clean sequence and no stop codons were found in the inferred amino acid sequence.

Phylogenetic Analysis

Phylogenetic trees were constructed using Maximum Parsimony (MP), Maximum Likelihood (ML), and Neighbor Joining (NJ) methods. Analyses were performed using a data set comprised only of the unique haplotypes (56 sequences).

Parsimony analyses were performed using PAUP* 4.0 (Swofford 2003). Tree space was explored using a heuristic search, with 1000 replicates of random stepwise addition and TBR branch swapping. A strict consensus was obtained from this heuristic search. Support was assessed using bootstrap analysis (1000 replicates) with 10 replicates of random sequence addition for each of the bootstrap replicates.

The best-fit model of nucleotide substitution, determined by the hierarchical Likelihood Ratio Test in Modeltest 3.7 (Posada & Crandall 1998), was TIM+G. The TIM+G model was used to generate a Maximum Likelihood tree in PAUP* 4.0 (Swofford 2003). Bootstrap support was determined with 250 replicates of a heuristic search strategy, with an as-is addition sequence and TBR branch swapping.

PAUP* 4.0 (Swofford 2003) was also used to perform Neighbor Joining analyses. Distances based on the TIM+G model were used in the analysis, and confidence in the tree was assessed using 1000 bootstrap replicates.

Genetic Diversity

Average p-distances (total number of differences between two sequences divided by total length of sequence compared) within and between sites were calculated using PAUP* 4.0 (Swofford 2003). In addition, distances were estimated based on the TIM+G model. The number of haplotypes and haplotype diversity for each site were determined using DnaSP 4.10.4 (Rozas et al. 1999).

Population Structure

We characterized population structure within *Ciona* species and sites using an analysis of molecular variance (AMOVA) in Arlequin 3.0 (Schneider et al. 2000). Pairwise Φ_{ST} values between sites were calculated by AMOVA and p-values were obtained by a 110-replicate permutation test.

Results

With the exception of Mediterranean Type A, for which all individuals had the same haplotype, haplotype diversities within collection sites of *C. intestinalis* were high, ranging from 0.916 to 0.956 (Table 1.2). Haplotype diversity was similarly high in *C. roulei* (0.933), and only slightly lower in *C. savignyi* (0.752). Analyses based on AMOVA suggest that most variation was within sites, and that there was relatively little differentiation among sites within a region (Table 1.3; Northeast Pacific Type A *C. intestinalis* $\Phi_{ST} = -0.073$ ($p > 0.05$), Northwest Atlantic Type B *C. intestinalis* $\Phi_{ST} = 0.103$ ($p > 0.05$), *C. savignyi* $\Phi_{ST} = -0.099$ ($p > 0.05$)). But within both Type A and Type B *C. intestinalis*, differentiation existed between regions (Table 1.3; Type B $\Phi_{ST} = 0.226$ ($p < 0.0001$), Type A $\Phi_{ST} = 0.476$, ($p < 0.002$)).

In contrast to the limited variation among sites of the same species/type, there was substantial differentiation between Type A and B *C. intestinalis* and between *C. intestinalis* and *C. savignyi.* Within *C. intestinalis,* Type A and B differed in mtDNA sequence by ~12%. *C. savignyi* exhibited 13.5-16.2% mtDNA sequence divergence from either of the other *Ciona* species (Table 1.4).

Maximum Parsimony and Maximum Likelihood methods produced bootstrap majority-rule consensus trees with identical topologies. Both the 50% majority rule

Table 1.2 *Ciona* spp. haplotype diversity (calculated in DnaSP 4.10.4 [Rozas and Rozas 1999]) and number of haplotypes in each species or region within species

Species	Region	Haplotype Diversity (± SD)	Number of Haplotypes
C. intestinalis	Mediterranean	0	1
Type A	Northeast Pacific	0.916 ± 0.041	11
Type B	Northeast Atlantic	0.937 ± 0.033	14
	Northwest Atlantic	0.956 ± 0.026	21
C. roulei		0.933 ± 0.122	5
C. savignyi		0.752 ± 0.075	5
Ciona sp.		0	1

Table 1.3 Distribution of variation within and among sites and among regions resulting from AMOVA tests performed in Arlequin 3.0 (Schneider 2000). Regions and sites as in Table 1.1.

Type A *C. intestinalis*	% variation
Among regions	45.49
Among sites within regions	2.13
Within sites	52.38
Northeast Pacific Type A *C. intestinalis*	% variation
Among sites	-7.27
Within sites	107.27
Type B *C. intestinalis*	% variation
Among regions	5.37
Among sites within regions	17.24
Within sites	77.39
Northwest Atlantic Type B *C. intestinalis*	% variation
Among sites	10.27
Within sites	89.73
C. savignyi	% variation
Among sites	-9.91
Within sites	109.91
*Negative covariances can result from an absence of genetic structure.	

Table 1.4 Sequence divergence within and between Ciona species/regions within species. Above the diagonal: divergences corrected using the TIM+G model. Below the diagonal: uncorrected p-distances.

Species	Region	Mediterranean Type A	Northeast Pacific Type A	Northeast Atlantic Type B	Northwest Atlantic Type B	*C. roulei*	*C. savignyi*	*Ciona sp.*
C. intestinalis	Mediterranean	**0.000**	0.005	0.199	0.200	0.208	0.267	0.185
Type A	Northeast Pacific	0.005	**0.005**	0.201	0.201	0.210	0.267	0.190
Type B	Northeast Atlantic	0.122	0.123	**0.006**	0.007	0.010	0.324	0.216
	Northwest Atlantic	0.123	0.123	0.007	**0.007**	0.011	0.324	0.214
C. roulei		0.125	0.126	0.010	0.008	**0.004**	0.328	0.217
C. savignyi		0.136	0.135	0.161	0.161	0.162	**0.002**	0.276
Ciona sp.		0.108	0.109	0.127	0.125	0.126	0.140	**0.000**

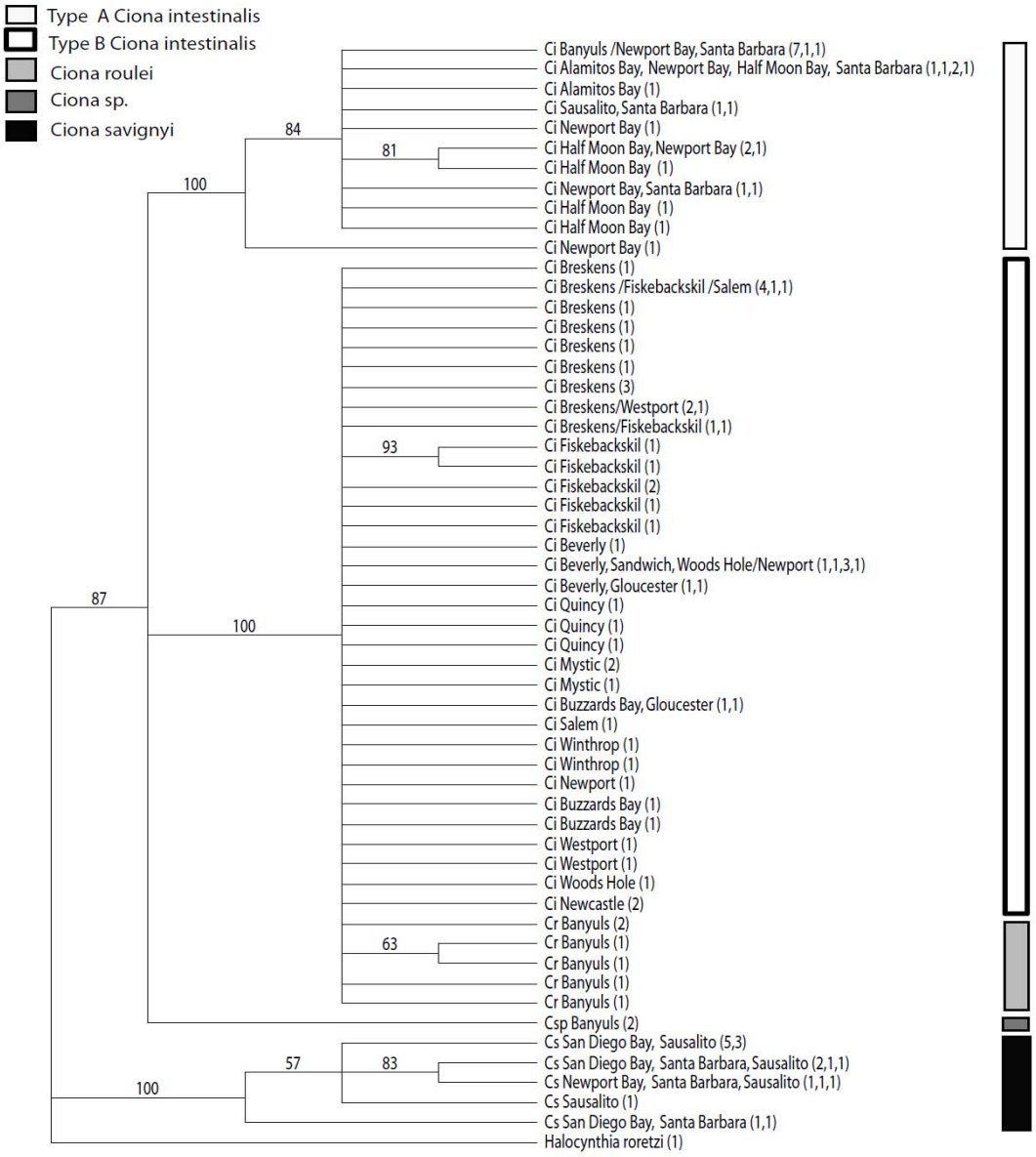

Figure 1.1 Bootstrap majority rule Maximum Parsimony tree generated from the 56 unique haplotypes found in 107 individuals (106 ingroup sequences and 1 outgroup sequence, listed in Materials and Methods). Numbers along branches indicate bootstrap percentages. Nodes with bootstrap percentages < 50% have been collapsed. Numbers in parentheses beside each site refer to the number of individuals sequenced from each site.

bootstrap MP tree (1000 replicates; Figure 1.1) and the 50% majority rule bootstrap ML tree (250 replicates; not shown) support the monophyly of *C. savignyi* and place it sister to a *C. intestinalis/C. roulei* clade. The *C. intestinalis/C. roulei clade* is divided into two well-supported sub-clades in both trees (Figure 1.1). One sub-clade, with 100% bootstrap support, defines Type A *C. intestinalis* and includes haplotypes from both the Northeast Pacific and the Mediterranean. A second sub-clade, also with 100% bootstrap support, includes Type B *C. intestinalis* from the Northeast and Northwest Atlantic, together with *C. roulei*. Thus Type B *C. intestinalis* is paraphyletic with respect to *C. roulei.* The NJ tree defines the same major clades revealed by the MP and ML analyses, and clearly shows that interspecific (or inter-type) differences dwarf intraspecific differences (Figure 1.2).

Ciona sp.

Three individuals from Banyuls-sur-Mer Harbor, France, had haplotypes that were distinctive from those seen in all other individuals (Table 1.4, Figures 1.1 and 1.2). Sequence divergences between these haplotypes and haplotypes found in *C. intestinalis* and *C. savignyi* ranged from 10.8-14.0% (Table 1.4). These individuals (hereafter referred to as *Ciona* sp.) were also morphologically distinct from *C. intestinalis* and *C. roulei* found in Banyuls-sur-mer. Furthermore, they did not match morphological or ecological descriptions of *C. edwardsi* (the fourth shallow-water species) from the region (Copello 1981; Fiala-Medioni 1974). The tunics of these individuals were reddish-orange and opaque, and the body clear. In contrast, the *C. roulei* individuals had clear tunics and deep red bodies, while *C. intestinalis* individuals (from this region) had clear tunics and milky-white, opaque bodies. The tunics of *Ciona* sp. were very thin, well-attached, and difficult to remove, in contrast to the tunics of *C. intestinalis* and *C. roulei.* The three individuals assigned to *Ciona*

Figure 1.2 Neighbor Joining phylogram generated from the 56 unique haplotypes found in 107 individuals (106 ingroup sequences and 1 outgroup sequence, listed in Materials and Methods). Numbers along each branch indicate the number of changes along that branch and (in parentheses) the bootstrap percentages. Bootstrap percentages < 50% not shown. Numbers in parentheses beside each site refer to the number of individuals sequenced from each site.

sp. were three to four centimeters long fully extended with tunic intact, which was considerably smaller than *C. intestinalis* and *C. roulei* from this region. The buccal and atrial siphons were proportionally longer compared to those of *C. intestinalis* and *C. roulei*, and extended half the length of the body. The atrial siphon had six red pigment dots, and the buccal siphon had at least six. Photos of *Ciona* sp. have been placed on the Dutch Ascidians website: http://www.ascidians.com.

Discussion

Mitochondrial DNA sequence data confirm that *C. intestinalis* and *C. savignyi* exhibit substantial genetic divergence. The mtDNA phylogeny also shows *C. roulei* haplotypes to be embedded within and not clearly distinct from haplotypes found in Type B *C. intestinalis*. Furthermore, Type A and Type B *C. intestinalis* are exclusive groups (Figures 1.1 and 1.2), with sequence divergence that suggests a most recent common ancestor in the Pliocene, assuming a molecular clock rate of about $3\%^{/my}$ (estimated from mitochondrial DNA clocks in three echinoderm genera: (Lessios et al. 2001; Lessios et al. 1999; McCartney et al. 2000).

Suzuki et al. 2005 compared 120 kb of cosmid sequence from British Type B *C. intestinalis* with the whole genome sequence from a Northeast Pacific Type A *C. intestinalis*. This revealed a broad range of sequence divergence, with predicted coding regions showing the least divergence. Twenty four randomly chosen ESTs from British Type B *C. intestinalis* showed 95.2% average sequence identity to orthologues found in the Japanese Type A *C. intestinalis* cDNA database (Suzuki et al. 2005). Unfortunately, the selection of genomic regions in these analyses does not allow comparisons with patterns of divergence in other marine invertebrate taxa, because of the difficulty in identifying homologous gene regions. However,

mitochondrial COI data allow us to compare distances between Type A and B *C. intestinalis* to distances between other closely related marine taxa.

We searched the literature for studies that estimated amounts of interspecific and intraspecific mtDNA COI sequence divergence between congeneric marine invertebrate species. These comparisons used a variety of distance corrections, depending on what authors judged to be the most appropriate model of molecular evolution. Only a very few studies provided uncorrected sequence divergence. We calculated an average distance, regardless of correction method. Based on 464 comparisons of 101 nominal species of cnidarians, annelids, arthropods, bivalves, and echinoderms, the average interspecific divergence is 0.159 ± 0.002 (data from (Arndt et al. 1996; Biermann et al. 2003; Bucklin et al. 1999; Hart et al. 1997; Hellberg 1998; Hill et al. 2001; Hurtado et al. 2002; Landry et al. 2003; Lessios et al. 1999; McCartney et al. 2000; Metz 1998; O'Foighil et al. 1998; Zigler & Lessios 2004; Zigler et al. 2003). The average intraspecific difference is 0.109 ± 0.01, based on 80 comparisons of 45 nominal species of cnidarians, arthropods, annelids, bivalves, echinoderms and urochordates (Ganz & Burton 1995; Edmands et al. 1996; Glover et al. 2005; Hart et al. 1997; Hill et al. 2001; Hoeh et al. 1996; Hurtado et al. 2004; Jolly et al. 2005; King et al. 1999; Lee 2000; Lessios et al. 1999; Medina et al. 1999; O'Foighil et al. 1998; Palumbi et al. 1997; Tarjuelo et al. 2001; Tarjuelo et al. 2004; van Syoc 1994). The average TIM+G corrected distance between Type A and B *C. intestinalis* (0.2, Table 1.4) is twice as high as the average corrected intraspecific distance for marine invertebrates, and is even higher than the average corrected interspecific distance. Although the intraspecific comparisons from the literature may be biased towards species with substantial differentiation between sites, correcting such a bias would only increase the disparity between *C. intestinalis* and other marine invertebrates. These data highlight the substantial differentiation between Type A and

B *C. intestinalis* compared with intraspecific differentiation in other marine invertebrates.

Several other lines of evidence also point to substantial differentiation between Type A and B *C. intestinalis.* Hoshino and Nishikawa (1985) recognized differences between *C. intestinalis* from Naples and Japan (Type A) and the Northeast Atlantic (Type B) in the rate of change with growth of the number of inner longitudinal vessels. In addition, British Type B *C. intestinalis* and Japanese Type A *C. intestinalis* are partially reproductively isolated. Although crosses between British Type B *C. intestinalis* eggs and Japanese Type A *C. intestinalis* showed normal fertilization rates (as estimated by the frequency of "normal two-cell stage embryos"), the reciprocal cross yielded many fewer fertilized eggs (Suzuki et al. 2005).

Given this genetic, morphological and reproductive evidence, we tentatively advocate the separation of *C. intestinalis* into two species according to the phylogenetic species concept. This cryptic speciation in *C. intestinalis* is consistent with recent studies of other marine taxa in which molecular data uncover cryptic species previously hidden by anthropogenic transport or morphological similarity (Dawson & Jacobs 2001; Holland et al., 2004).

Ciona roulei is not a monophyletic group with respect to mtDNA haplotypes; some *C. roulei* haplotypes are more closely related to Type B *C. intestinalis* haplotypes than to other *C. roulei* haplotypes. Crosses between *C. roulei* and Mediterranean Type A *C. intestinalis* are successful in one direction (Lambert et al. 1990), leading the authors to suggest that *C. roulei* and *C. intestinalis* diverged more recently than *C. intestinalis* and *C. savignyi* (between which hybrids have not been produced). This distinctness of *C. savignyi* is strongly supported by the mitochondrial COI data. The position of *C. roulei* nested within the larger Type B *C. intestinalis* clade suggests that *C. roulei* was recently derived from Type B *C. intestinalis*.

Phylogenetic analyses lend support to a hypothesis that Northeast Atlantic Type B *C. intestinalis* individuals moved into the Mediterranean Sea, where they diverged from the ancestral Type B *C. intestinalis*, giving rise to *C. roulei.* Reproductive compatibility between *C. roulei* and Type B *C. intestinalis* has not been investigated; successful crosses between *C. roulei* and Type B *C. intestinalis* would reinforce the Northeast Atlantic origin of *C. roulei*'s ancestors.

The substantial intraspecific divergence in *C. intestinalis*, the placement of *C. roulei* within the Type B *C. intestinalis* clade, and the phylogenetically distinct *Ciona* sp. highlight the evolutionary and taxonomic ambiguities within the genus *Ciona*. Type A and B *C. intestinalis* are 12% divergent at the mitochondrial COI locus, and yet they exhibit partial reproductive compatibility. These two types are sympatric in southern England, and exhibit pre-zygotic reproductive isolation at this site (P. Sordino, personal communication). Similarly, mtDNA COI sequences differ by 12.5% between the sympatric *C. roulei* and Mediterranean Type A *C. intestinalis*, but viable F1 hybrids are produced from one of the two reciprocal crosses. Although they occur along the same coastline, this pair of species may be ecologically isolated. Nothing is known about reproductive isolation between *C. roulei* and Type B *C. intestinalis*. Differences in compatibility between individuals from allopatric and sympatric sites of these two species will help clarify the evolutionary processes generating reproductive isolation in *Ciona*.

While successful crosses between other pairs of ascidian species have been well documented in the laboratory (Jeffery & Swalla 1990; Lambert et al. 1981), hybrids between species have never been identified in nature. This lack of natural hybridization suggests that pre-zygotic isolating mechanisms may have a large role in the maintenance of reproductive isolation in ascidians. Even so, genetically distinct ascidian species such as *C. intestinalis* and *C. savignyi* have been confused historically

(Hoshino & Nishikawa 1985), raising the possibility that hybrids could be morphologically cryptic as well. These uncertainties necessitate further study of species level relationships within *Ciona* in the context of improved understanding of intrinsic barriers to gene exchange.

Acknowledgments

We thank A. Fiala-Medioni, P. Romans, and J-C. Roca at the Observatoire Océanologique de Banyuls-sur-mer for their generous intellectual and logistic support. We are also indebted to X. Turon and M. Pascual for helpful discussions and travel assistance, to A. Gittenberger and C. Dahlberg for Northeast Atlantic *C. intestinalis* samples, to J. Friel and C. Dardia for sample preservation advice, to B. Swalla, C. and G. Lambert, J. Carlton and P. Sordino for illuminating *Ciona* discussions. We thank members of the Harrison Lab, A. Clark, B. Swalla, P. Sordino, J. Grassle and two anonymous reviewers for providing comments on the manuscript. This research was funded by a Cornell University Sigma Xi Grant to M.L.N.

REFERENCES

Arndt, A., Marquez, C., Lambert, P., and M.J. Smith. 1996. Molecular phylogeny of eastern Pacific sea cucumbers (Echinodermata: Holothuroidea) based on mitochondrial DNA sequence. Molec. Phylogenet. Evol. 6:425-437.

Biermann, C. H., Kessing, B.D., and S.R. Palumbi. 2003. Phylogeny and development of marine model species: strongylocentrotid sea urchins. Evol. Dev. 5:360-371.

Boffelli, D., Weer, C.V., Weng, L., Lewis, K., Shoukry, M.I., Pachter, L., Keys, D.N., and E.M. Rubin. 2004. Intraspecies sequence comparisons for annotating genomes. Genome Res. 14:2406-2411.

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

Byrd, J., and C.C. Lambert. 2000. Mechanism of the block to hybridization and selfing between the sympatric ascidians Ciona intestinalis and Ciona savignyi. Mol. Repro. Dev. 55:109-116.

Chabry, L. 1887. Contribution a l'embrologie normale teratologique des ascidies simples. J. Anat. Physiol. Norm. Path. 23:167-319.

Copello, M. 1981. *Ciona edwardsi* (Roule, 1886) Espece Littorale de Mediterranee Distincte de *Ciona intestinalis* (Linne, 1767). Vie Milieu 31:243-253.

Dawson, M, and D. Jacobs. 2001. Molecular evidence for cryptic species of *Aurelia*

aurita (Cnidaria, Scyphozoa). Biol. Bull. 200:92-96.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., and A. Di Gregorio. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298:2157-2167.

Edmands, S., Moberg, P.E., and R.S. Burton. 1996. Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. Mar. Biol. 126:443-450.

Fiala-Medioni, A. 1974. Ascidians of the Rocky Benthos of Banyuls-sur-Mer. Faunistic Inventory and Ecology Notes. Vie Milieu 24:193-207.

Ganz, H. H., and R.S. Burton. 1995. Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod *Tigriopus californicus.* Mar. Biol. 123:821-827.

Gissi, C, and G. Pesole. 2003. Transcript mapping and genome annotation of ascidian mtDNA using EST data. Genome Res. 13:2203-2212.

Glover, A. G., Goetze, E., Dahlgren, T.G., and C.R. Smith. 2005. Morphology, reproductive biology and genetic structure of the whale-fall and hydrothermal vent specialist, *Bathykurila guaymasensis* Pettibone, 1989 (Annelida: Polynoidae). Mar. Ecol. 26:223-234.

Harant, H., and P. Vernieres. 1933. Tuniciers. Faune de France 27:1-93.

Hart, M. W., Byrne,M., and M.J. Smith. 1997. Molecular phylogenetic analysis of life history evolution in Asterinid starfish. Evolution 51:1848-1861.

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

Hill, M. W., Broman, K.W., Stupka, E., Smith, W.C., Jiang, D., and A. Sidow. 2008. The *C. savignyi* genetic map and its integration with the reference sequence facilitates insights into chordate genome evolution. Genome Res. 18:1369-1379.

Hoeh, W. R., Stewart, D.T., Sutherland, B.W. and E. Zouros. 1996. Cytochrome *c* Oxidase Sequence Comparisons Suggest an Unusually High Rate of Mitochondrial DNA Evolution in *Mytilus* (Mollusca: Bivalvia). Mol. Biol. Evol. 13:418-421.

Holland, B. S., M.N. Dawson, G.L. Crow and D.K. Hofmann. 2004. Global phylogeography of Cassiopea (Scyphozoa: Rhizostomeae): Molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. Mar. Biol. 145:1119-1128.

Hoshino, Z. I., and T. Nishikawa. 1985. Taxonomic studies of *Ciona intestinalis* (L.) and its allies. Publ. Seto Mar. Biol. Lab. 30:61-79.

Hurtado, L. A., Lutz, R.A., and R.C. Vrijenhoek. 2004. Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. Mol. Ecol. 13:2603-2615.

Hurtado, L. A., Mateos, M., Lutz, R.A., and R.C. Vrijenhoek. 2002. Molecular evidence for multiple species of *Oasisia* (Annelida: Siboglinidae) at eastern Pacific hydrothermal vents. Cah. Biol. Mar. 34:377-380.

Jeffery, W. R., and B.J. Swalla. 1990. Anural development in ascidians: evolutionary modification and elimination of the tadpole larva. Sem. Dev. Biol. 1:253-261.

Johnson, D. S., Davidson, B., Brown, C., Smith, W., and A. Sidow. 2004. Noncoding regulatory sequences of *Ciona* exhibit strong correspondence between evolutionary constraint and functional importance. Genome Res. 14:2448-2456.

Jolly, M. T., Jollivet, D., Gentil, F., Thiebaut, E., and F. Viard. 2005. Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the North coast of France. Heredity 94:23-32.

King, T. L., Eackles, M.S., Gjetvaj, B., and W.R. Hoeh. 1999. Intraspecific phylogeography of Lasmigona subviridis (Bivalvia: Unionidae): conservation implications of range discontinuity. Mol. Ecol. 8:S65.

Kott, P. 1952. The Ascidians of Australia. Aust. J. Mar. Fresh. Res. 3:206-233.

Kott, P. 1990. The Australian Ascidiacea Pt 2, Aplousobranchia (1). Memoirs of the Queensland Museum 29:1-266.

Kott, P. 1997. Tunicates. Pp. 1092-1255 *in* S. A. Shepherd, Thomas, I.M., ed. The Tunicates. Marine Invertebrates of Southern Australia. Part 1. Government Printer, Adelaide, South Australia.

Kumar, S., Tamura, K., and M. Nei. 2004. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. Brief. Bioinform. 5:150-163.

Lambert, G. 2001. A global overview of ascidian introductions and their possible

impact on the endemic fauna. Pp. 249-257 *in* H. Sawada, ed. The Biology of Ascidians. Springer-Verlag, Tokyo.

Lambert C.C, and G. Lambert. 2003. Persistence and differential distribution of nonindigenous ascidians in harbors of the Southern California Bight. Mar. Ecol. Prog. Ser. 259:145-161.

Lambert, G., Lambert, C.C., and D. Abbott. 1981. *Corella* species in the American Pacific Northwest: distinction of *C. inflata* Huntsman, 1912 from *C. willmeriana* Herdman 1898 (Ascidiacea, Phlebobranchiata). Can. J. Zool. 59:1493-1504.

Lambert, C.C, LaFargue, F., and G. Lambert. 1990. Preliminary note on the genetic isolation of *Ciona* species (Ascidiacea, Urochordata). Vie Milieu 40:293-295.

Landry, C., Geyer, L.B., Arakaki, Y., Uehara, Y., and S.R. Palumbi. 2003. Recent speciation in the Indo-West Pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. Proc. R. Soc. Lond. B 270:1839-1847.

Lee, C. E. 2000. Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate 'populations'. Evolution 54:2014-2027.

Lessios, H. A., Kessing, B.D., 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., Kessing, B.D., Robertson, D.R., and G. Paulay. 1999. Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land

barriers and ocean currents. Evolution 53:806-817.

Linne, C. 1767. Systema Naturae. Vindobonae: Typis Ioannis Thomae, 1767-1770.

McCartney, M. A., Keller, G., and H.A. Lessios. 2000. Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. Mol. Ecol. 9:1391–1400.

Medina, M., Weil, E., and A.M. Szmant. 1999. Examination of the Montastraea annularis species complex (Cnidaria: Scleractinia) using ITS and COI sequences. Marine Biotechnol. 1:89-97.

Metz, E. C. 1998. Mitochondrial DNA and Bindin Gene Sequence Evolution among Allopatric Species of the Sea Urchin Genus *Arbacia*. Mol. Biol. Evol. 15:185-195.

Minganti, A. 1948. Interspecific Fertilization in Ascidians. Nature 161:643-644.

Missal, K., Rose, D., and P. Stadler. 2005. Non-coding RNAs in *Ciona intestinalis*. Bioinformatics 21:i77-i78.

Monniot, C. 1990. Ascidies de Nouvelle-Caledonie. 8. Phlebobranches (suite). B. Mus. Natl. Hist. Nat. 12:491-515.

Monniot, C., and F. Monniot. 1977. Tuniciers benthiques profonds du Nord-Est Atlantique. Resultats des campagnes Biogas. B. Mus. Natl. Hist. Nat. 323:695-720.

Monniot, C., and F. Monniot. 1983. Ascidies antarctiques et subantarctiques: morphologie et biogeographie. Mem Mus Natn Hist Nat Ser A 125: 1-168.

Monniot, C., and F. Monniot. 1989. Ascidians collected around the Galapagos Islands using the Johnson-Sea-Link research submersible. P. Biol. Soc. Wash. 102:14-32.

Morgan, T.H. 1904. Self-fertilization induced by artificial means. J. Exp. Zool. 1:135-178.

O'Foighil, D., Gaffney, P.M., Wilbur, A.E., and T.J. Hilbish. 1998. Mitochondrial cytochrome oxidase I gene sequences support an Asian origin for the Portuguese oyster *Crassostrea angulata*. Mar. Biol. 131:497-503.

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L., and N. Tachino. 1997. Speciation and population genetic structure in tropical Pacific sea urchins. Evolution 51:1506-1517.

Posada, D., and K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Sanamyan, K. 1998. Ascidians from the north-western Pacific region. 5. Phlebobranchia. Ophelia 49:97-116.

Schneider, S., Roessli, D., and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Sordino, P., Heisenberg, C.P., Cirino, P., Toscano, A., Giuliano, P., Marino, R., Pinto,

M.R., and R. De Santis. 2000. A mutational approach to the study of development of the protochordate *Ciona intestinalis* (Tunicata, Chordata). Sarsia 85:173-176.

Suzuki, M., Nishikawa, T., and A. Bird. 2005. Genomic Approaches Reveal Unexpected Difference within Ciona intestinalis. J. Mol. Evol. 61:627-635.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b 10. Sinauer Associates, Sunderland.

Tarjuelo, I., Posada, D., Crandall, K. A., Pascual, M. and X. Turon. 2001. Cryptic species of *Clavelina* (Ascidiacea) in two different habitats: harbours and rocky littoral zones in the northwestern Mediterranean. Mar. Biol. 139:455-462.

Tarjuelo, I., Posada, D., Crandall, K. A., Pascual, M. and X. Turon. 2004. Phylogeography and speciation of colour morphs in the colonial ascidian *Pseudodistoma crucigaster*. Mol. Ecol. 13:3125-3136.

Turon, X., and S. López-Legentil. 2004. Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata. Mol. Phylogenet. Evol. 33:309-320.

Van Name, W. 1945. The north and south American ascidians. B. Am. Mus. Nat. Hist. 84:1-476.

van Syoc, R. J. 1994. Genetic divergence between subpopulations of the eastern Pacific goose barnacle *Pollicipes elegans*: mitochondrial cytochrome c subunit 1 nucleotide sequences. Mol. Mar. Biol. Biotech. 3:338-346.

Vinson, J., Jaffe, D., O'Neill, K., Karlsson, E., Stange-Thomann, N., Anderson, S.,

Mesirov, J., Satoh, N., Satou, Y., Nusbaum, C., Birren B., Galagan, J., and E. Lander. 2005. Assembly of polymorphic genomes: Algorithms and application to *Ciona savignyi*. Genome Res. 15:1127-1135.

Yokobori, S., Oshima, T., and H. Wada. 2005. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol. Phylogenet. Evol. 34:273-283.

Yokobori, S., Ueda, T., Feldmaier-Fuchs, G., Paabo, S., Ueshima, R., Kondow, A., Nishikawa, K., and K. Watanabe. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics 153:1851-1862.

Yokobori, S., Y. Watanabe, and T. Oshima. 2003. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): Comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. J. Mol. Evol. 57: 574-587

Yokobori S-i, Oshima T, and H. Wada. 2005. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol. Phylogenet. Evol. 34:273-283.

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., Raff, E.C., Popodi, E., Raff, R.A., and H.A. Lessios. 2003. Adaptive evolution of binding in the genus *Heliocidaris* is correlated with the shift to direct development. Evolution 57:2293-230

CHAPTER 2

POLYMORPHISM AND DIVERGENCE WITHIN THE ASCIDIAN GENUS *CIONA*

Abstract

The genus *Ciona,* a widely distributed group of solitary ascidians, has long been an important model in embryology and developmental biology. *Ciona* has also recently attracted the attention of evolutionary biologists because of the remarkably high levels of heterozygosity found within single individuals. Surprisingly, genealogical relationships in *Ciona* have received little attention. Here, we expand our knowledge of relationships among the members of the *Ciona* genus and estimate levels of polymorphism in natural populations.

Previous studies have documented the outgroup status of *Ciona savignyi* among the shallow water *Ciona* and revealed the existence of two distinct forms (Types A and B) of the widespread *C. intestinalis*. Here, using gene genealogies of six nuclear gene loci, we show Type A and B to be well-supported monophyletic groups. In spite of their morphological similarity, Type A vs. Type B divergences range from 0.035 to 0.124. In contrast, the morphologically distinct *C. roulei* is embedded within Type B in all genealogies, and a new species, *Ciona* sp., appears to be associated with Type B/*C. roulei* to the exclusion of Type A. Levels of polymorphism in natural populations are similar to levels reported in other organisms that are considered to be highly polymorphic.

Introduction

The genus *Ciona,* a widely distributed group of solitary ascidians, has been of

interest to biologists for over a century, largely because two of its members, *Ciona intestinalis* (Linne 1767) and *Ciona savignyi* (Herdman 1882) are model organisms in developmental biology and embryology. Recently published genome sequences of both species (Dehal et al. 2002; Vinson et al. 2005) have increased the utility of these species as models in developmental biology and have also attracted the attention of population geneticists because of the remarkably high levels of heterozygosity found within the single individuals for which whole genome sequences have been produced. Allelic polymorphism across the entire genome in *C. intestinalis* is 1.2% (including both Single Nucleotide Polymorphisms (SNPs) and insertion/deletions) (Dehal et al. 2002). This level of polymorphism is fifteen times that of *Homo sapiens* (Dehal et al. 2002) and three times that of the Japanese puffer fish, *Fugu rubripes,* which has been described as having a high level of polymorphism (Aparicio et al. 2002; Dehal et al. 2002). The average genome-wide SNP heterozygosity in *C. savignyi*, again based on a single individual, is 4.5% (Small et al. 2007).

Surprisingly, given the prominence and utility of these ascidians as models, genealogical and phylogenetic relationships within and among *Ciona* species have received little attention. As we uncover the evolutionary history of these species, our understanding of the complexity of this group continues to increase.

For example, it is now evident that *C. intestinalis* comprises two distinct, highly divergent entities (Suzuki et al. 2005; Caputi et al. 2007; Nydam & Harrison 2007). Type A *C. intestinalis* (the type for which the genome sequence is available), is thought to be native to the Northwest Pacific Ocean and is now found throughout the Pacific Ocean, the Mediterranean Sea, the Atlantic coast of South Africa, and the Black Sea (Van Name 1945; Kott 1952). Type B *C. intestinalis,* originally described by Linnaeus from the North Atlantic Ocean (Linne 1767), is found in the Western Atlantic Ocean (Nydam & Harrison 2007). The two types are ~12% divergent at the

mitochondrial COI gene (cytochrome oxidase I, uncorrected p-distance, (Nydam & Harrison, 2007)).

A third taxon, *Ciona roulei* (Lahille 1887), is endemic to the Northwestern Mediterranean Sea, having been found at five locations along the French coast, near the Spanish border (Harant & Vernieres 1933; Fiala-Medioni 1974). This species is morphologically distinct from all other *Ciona* species (Harant & Vernieres, 1933). In a mtCOI tree, *C. roulei* is embedded within Type B, i.e. Type B is paraphyletic with respect to *C. roulei* (Nydam & Harrison 2007).

In a previous study (Nydam & Harrison 2007), three morphologically distinct individuals (*Ciona* sp.) were collected from a depth of <5 meters in Banyuls harbor, southern France. These individuals were determined to be genetically distinct from all other *Ciona* species based on mtCOI sequence data (10.9 and 12.6% uncorrected p-distance from Type A and Type B, respectively) (Nydam & Harrison 2007). However, the phylogenetic placement of *Ciona* sp. within the *Ciona* genus remained unresolved using this single marker.

A fourth species among the shallow-water *Ciona* species is *C. savignyi,* which was first described from the Western Pacific Ocean (Herdman 1882) and subsequently invaded the Eastern Pacific Ocean (Hoshino & Nishikawa, 1985). Based on mtCOI sequencing, *C. savignyi* is a well supported monophyletic group that occupies a basal position in the phylogeny as a sister group to the *C. intestinalis/C. roulei/C.* sp clade (Nydam & Harrison 2007).

While the relationships in the mtCOI tree are well supported, inferences of phylogenetic relationships from a single marker should not be considered definitive. A genealogy from a single locus reflects the particular evolutionary history of that locus, which may not be the evolutionary history of the genome as a whole (Maddison 1997; Nichols 2001). Therefore, multiple markers need to be examined to construct a

species phylogeny with more confidence.

In the present study, we examine patterns of genetic variation within and among the shallow water members of the genus *Ciona.* Gene genealogies based on six nuclear loci allow us to determine whether patterns revealed by mtCOI are consistent across the nuclear genome; we focus on the distinctness of the two types of *C. intestinalis*, the paraphyly of Type B with respect to *C. roulei*, and the uncertain relationship between *Ciona* sp. and its congeners. We also estimate polymorphism in population samples of Type A and B *C. intestinalis* and in *C. roulei* for each of the six nuclear loci and compare these with estimates for *Ciona* from whole genome sequences and with levels of heterozygosity in natural populations of other taxa.

Materials and Methods

DNA extraction, amplification and sequencing

Ovaries were dissected from freshly-collected individuals, cut into several pieces, immediately preserved in DMSO (dimethyl sulfoxide), and ultimately (within 12 d) stored at -80°C until needed. We use ovaries as they are accessible by simple dissection and ovary tissue is easily lysed for DNA extraction. Samples from Sweden and The Netherlands were preserved and shipped in ethanol soon after collection. Upon arrival, the ovaries from these individuals were dissected and preserved in DMSO. Total DNA was extracted from the ovaries using the Qiagen DNeasy® Tissue Kit (Qiagen Corporation, Santa Clarita, CA).

To provide a mitochondrial COI tree that reflects a better sample of the geographic range of the *Ciona* species, we sequenced thirteen *C. intestinalis* (from Japan and the English Channel) and 7 *C. savignyi* individuals (from Japan) and added these sequences to the mitochondrial COI data matrix used previously (Nydam &

Harrison 2007). The nuclear gene genealogies each included a 22-individual subset of the individuals present in the mtCOI tree plus 10 additional *C. intestinalis* samples from the southern UK, northern France and Japan (Table 2.1). A *C. savignyi* individual from the Eastern Pacific Ocean served as the outgroup individual for each nuclear genealogy.

Approximately 1 kilobase from each of six nuclear genes was amplified using primers developed from the published Type A genome sequence. The genes are: Vesicular acetylcholine transporter (*vAChTP*), Cellulose synthase (*CiCesA*), Fibroblast Growth Factor orthologous to vertebrate Fibroblast Growth Factor 4/5/6 (*Ci-Fgf4/5/6*), Forkhead (*Ci-fkh*) (5' regulatory region), Jade (*jade*) and Patched (*Ci-Patched).* Loci were selected to represent a range of polymorphism, based on aligning whole genome shotgun reads from the Type A genome. This was done so that the six loci could be used for genealogical analyses, as well as for analyses of levels of polymorphism. Four of the loci were considered to have "low" levels of polymorphism: (*vAChTP*, *CiCesA, Ci-fkh,* and *Ci-Patched*), and two were considered to have "high" levels of polymorphism (*Ci-Fgf4/5/6* and *jade*). For *Ci-Fgf4/5/6* and *jade*, primers developed from Type A individuals did not consistently amplify Type B individuals; Type B individuals were amplified and sequenced using specific primers developed from preliminary sequences of several Type B individuals. *Ciona sp.* and *C. savignyi*-specific primers were also developed for several loci. Sequences from three of the loci (*vAChTP*, *CiCesA, jade*) include both coding and noncoding regions, whereas sequences from *Ci-Fgf4/5/6*, *Ci-fkh, Ci-Patched* are entirely noncoding (Table 2.2). Primer sequences and thermocycling conditions for the six nuclear loci are available from the authors. Primers and PCR amplification conditions for the

Table 2.1 Collection sites for *Ciona* species

Species	Region/State or Country	Site	# of Individuals
Ciona intestinalis	Northwest Pacific/Japan	Morotsu	1
Type A		Nishiura	1
		Onagawa	1
		Shikoku Island	1
		Yokohama	1
	Northeast Pacific/CA	Newport Harbor	1
		Alamitos Bay	1
		Santa Barbara Harbor	1
		Half Moon Bay	1
		Sausalito	1
	English Channel/United Kingdom	Falmouth	1
		Plymouth	1
	Mediterranean/France	Banyuls-sur-Mer	2
Type B	Northwest Atlantic/NH to CT	Newcastle	1
		Gloucester	1
		Winthrop	1
		Mystic	1
	Northeast Atlantic/Sweden	Fiskebäckskil	3
	Northeast Atlantic/The Netherlands	Breskens	3
	English Channel/United Kingdom	Falmouth	1
		Plymouth	1
	English Channel/France	Granville	1
Ciona roulei	Mediterranean/France	Banyuls-sur-Mer	3
Ciona sp.	Mediterranean/France	Banyuls-sur-Mer	2

Table 2.2: Percent of noncoding sequence for each gene

	vAChTP	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*
Type A *C. intestinalis*	19	31	100	100	77	100
Type B *C. intestinalis/Ciona roulei*	19	52	100	100	91	100

mtCOI sequences have been published previously (Nydam & Harrison 2007).

PCR amplification of nuclear genes was performed in a 10-µl total reaction volume with 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µl of 10x buffer (50mM KCl, 20mM Tris (pH 8.4)), 0.2 µM of each primer, 0.08 U of *Taq* Polymerase (Gibco-BRL*)* and 1 µl of template DNA. To obtain clean sequence data, cloning of PCR products was necessary for the *Ci-Fgf4/5/6* and *jade* loci and for certain individuals at the other nuclear loci. Cloning was performed using a pGEM®-T kit (Promega Corporation, Madison, WI). PCR products (obtained directly from DNA or from clones) were incubated with 0.25µl each of Exonuclease I and Shrimp Antarctic Phosphatase at 37°C for 30 min, followed by 90°C for 10 min. The products were purified on Sephadex® columns (Sigma-Aldrich). The purified product was sequenced with a Big Dye Terminator Cycle sequencing kit and an Automated 3730 DNA Analyzer (Applied Biosystems) using the primers listed above. All unique haplotypes have been submitted to GenBank (Accession Numbers XXX-XXX). Sequences were edited, trimmed and aligned with Aligner (CodonCode Corporation, Dedham, MA).

Analyses

We examined each locus for evidence of intragenic recombination, calculating the minimum number of recombination events (R_m) in DnaSP 5.0 (Rozas et al. 2003). For each locus, we used 1,000 replicate coalescent simulations in DnaSP 5.0 to obtain a confidence interval for the R_m value and a probability that the true R_m is less than or equal to the observed R_m.

For each locus which showed evidence of intragenic recombination we created recombination networks in the program SplitsTree4 (Huson & Bryant 2006). Phylogenetic trees were constructed using Maximum Likelihood (ML) and Bayesian inference, for each locus alone, and for all loci concatenated. Analyses for each locus

were performed using a data set composed only of the unique haplotypes. While we are aware that ML and Bayesian methods may not be entirely appropriate for loci showing evidence of intragenic recombination, it is only by using these methods that we can assess confidence in the topologies.

The best-scoring ML tree for each locus/concatenated loci and bootstrap support for each node on this tree were obtained using the program RAxML v. 7.0.0 on the CIPRES web portal (Stamatakis et al. 2008). The GTR+G likelihood model of nucleotide substitution was used in all analyses (General Time Reversible + Gamma Rate Distribution, RAxML only supports GTR-based models of nucleotide substitution). The concatenated data set was partitioned by locus, so α-shape parameters, GTR-rates, and base frequencies were estimated separately for each locus. All nodes with less than <50% support were then collapsed using TreeView 1.6.6 (Page 1996).

Bayesian analyses were performed with MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003). The GTR+G model of nucleotide substitution was applied to all data sets (Nset = 6) so that Bayesian trees could be compared to ML trees. In the concatenated data set, data were partitioned and the parameters for each locus (α-shape, GTR-rates, and base frequencies) were determined independently. Each analysis was run for 10 million generations, with sampling every 1000 generations. The first 2,000 trees were eliminated as burn-in and a 50% majority-rule consensus tree was created from the remaining 8,000 trees using PAUP*4.0 (Swofford 2003). The MrBayes runs were carried out using the resources of the Computational Biology Service Unit at Cornell University which is partially funded by the Microsoft Corporation.

P-distances (the proportion of nucleotide sites that differ between two sequences) were calculated for every Type A vs. Type B combination; average p-

distance between the two types was then calculated in MEGA 4.0 (Tamura et al. 2007). Both uncorrected p-distances and p-distances corrected for the locus-specific best likelihood model of nucleotide substitution (as determined by AIC (Akaike Information Criterion) in Modeltest 3.7 (Posada & Crandall 1998) were calculated. Additionally, the number of net nucleotide substitutions per site between Type A and B (D_a) was calculated for each locus in DnaSP 5.0 (Rozas et al. 2003) in order to correct for within-type variation (Nei 1987).

We tested the position of *C. roulei* within the Type B clade by obtaining two best-scoring ML trees in RAxML using the GTR+G model for each gene (Stamatakis et al. 2008). One tree was constrained by Type B monophyly, the other by *C. roulei* monophyly. The likelihood of each constraint tree was compared to the unconstrained ML tree using the Shimodaira-Hasegawa (S-H) test in PAUP*4.0 (Swofford 2003) with 10,000 resampling estimated log-likelihood (RELL) bootstrap replicates under a GTR+G model. These tests could not be performed for *vAChTP* because Type B and *C. roulei* individuals shared haplotypes, and for *Ci-Fgf4/5/6* in the case of *C. roulei* monophyly because only 1 *C. roulei* individual was sequenced for *Ci-Fgf4/5/6.*

Similar methodology was used to investigate the placement of *Ciona* sp. within each genealogy, with a best-scoring ML tree produced under the constraint that Type B/*C. roulei* and *Ciona* sp. formed a monophyletic group, to the exclusion of Type A.

We calculated two measures of nucleotide polymorphism within Type A, Type B and *C. roulei*: nucleotide diversity (π) and theta ($4N_e\mu$) estimated from segregating sites (θ_w) using DnaSP 5.0 (Rozas et al. 2003).

Results

Significant levels of intragenic recombination were detected at five of the six

Table 2.3: Intragenic recombination, measured as Rm (minimum # of recombination events)

	vAChTP	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*
Type A	0	0	4	2	7	5
Type B	0	6	19	18	21	18

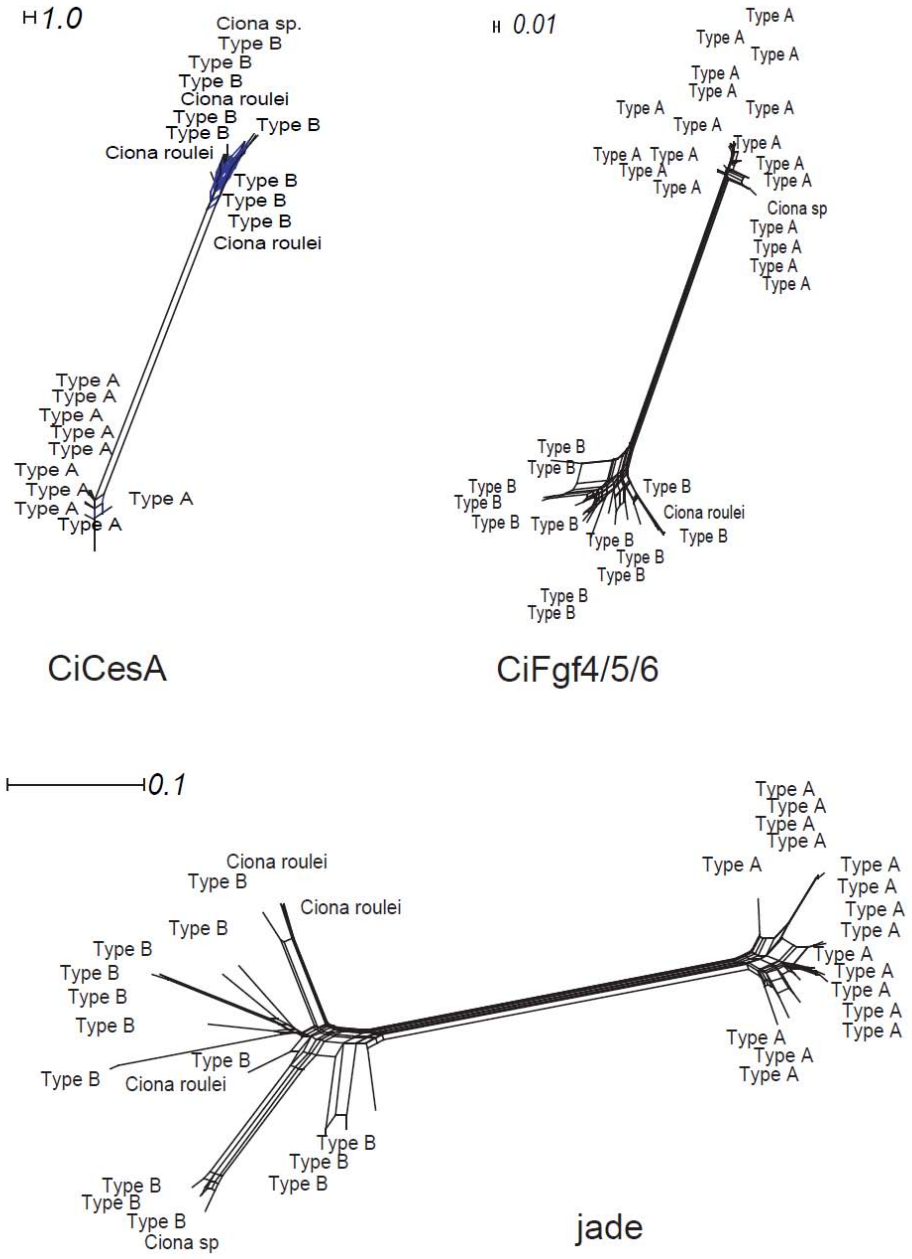

Figure 2.1 Recombination networks for *CiCesA, Ci-Fgf4/5/6* and *jade.* The scale bar in upper left hand corner of each network is proportional to the amount of molecular difference between the groups.

loci (Table 2.3), with Type B showing more recombination than Type A for all loci. The average minimum number of recombination events (R_m) across all six loci was 3 for Type A and 13.67 for Type B. Recombination networks created in SplitsTree4 (Huson & Bryant 2006) revealed Types A and B within *C. intestinalis* to be very distinct, with *C. roulei* embedded within Type B (Figure 2.1). Because SplitsTree4 (Huson & Bryant 2006) was unable to create recombination networks when all variable sites were included for *Ci-fkh* and *Ci-Patched* and when no recombination events were detected in *vAChTP*, only recombination networks for *CiCesA*, *Ci-Fgf4/5/6* and *jade* are shown in Figure 2.1.

The published mtCOI genealogy for the genus *Ciona* revealed Type A and Type B individuals as members of two distinct and well-supported clades (100% bootstrap support for each) (Nydam & Harrison 2007). This result, including the 100% bootstrap support for these clades, also holds in the updated mtCOI tree, which includes 13 new *C. intestinalis* individuals from Japan and the English Channel (tree not shown). The reciprocal monophyly of Type A and B is also evident when we use ML or Bayesian approaches to construct nuclear gene trees: either the Type A clade, the Type B clade, or both the Type A and B clades are well supported although we acknowledge Bayesian posterior probabilities are often high (Figure 2.2). There are no instances in which a Type A individual is found in a Type B clade, or vice versa (Figure 2.1 and 2.2).

Among the nuclear gene loci, divergence estimates between the Type A and Type B groups varied by locus, from 0.035 for *vAChTP* to 0.116 for *Ci-Fgf4/5/6* (Table 2.4, uncorrected p-distances). Corrected p-distances ranged from 0.039 for

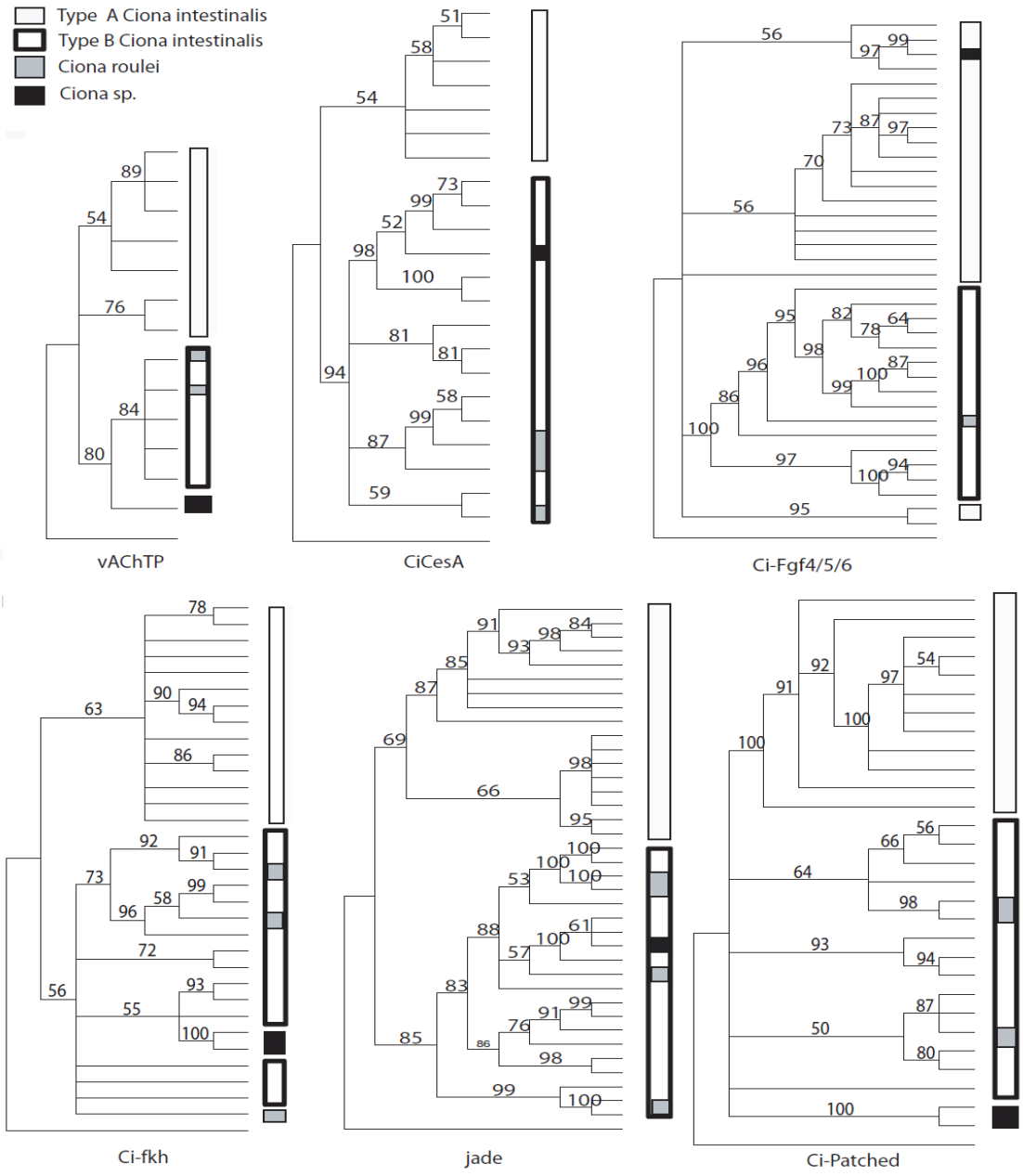

Figure 2.2 50% majority-rule consensus trees created in PAUP*4.0 from trees obtained by MrBayes 3.1.2. Values on the branches are Bayesian posterior probabilities. Nodes with a posterior probability value less than 50 were collapsed. The outgroup is *Ciona savignyi.*

Table 2.4 Uncorrected and corrected p-distances and Da between each C. intestinalis type and between each C. intestinalis type and Ciona sp. Uncorrected p-distance is the proportion of nucleotide sites that differ between two sequences, and corrected p-distance is adjusted by taking the locus-specific best likelihood model of nucleotide substitution into account. Da is the number of net nucleotide substitutions per site between species.

Species		*vAChTP*	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*	*mtCOI*
Type A vs. Type B	Uncorrected p-distance	0.035	0.077	0.116	0.051	0.113	0.077	0.124
	Corrected p-distance	0.039	0.092	0.151	0.055	0.139	0.081	0.156
	Da	0.035	0.072	0.09	0.029	0.08	0.051	0.102
Type A *C. intestinalis* vs. *Ciona* sp.	Uncorrected p-distance	0.191	0.079	0.019	0.184	0.125	0.241	0.109
	Corrected p-distance	0.413	0.104	0.024	0.307	0.161	0.388	0.187
	Da	0.193	0.085	0.015	0.193	0.129	0.229	0.099
Type B *C. intestinalis/C. roulei* vs. *Ciona* sp.	Uncorrected p-distance	0.177	0.014	0.115	0.176	0.069	0.229	0.125
	Corrected p-distance	0.35	0.015	0.094	0.286	0.081	0.358	0.232
	Da	0.184	0.007	0.097	0.176	0.03	0.207	0.121

vAChTP to 0.151 for *Ci-Fgf4/5/6*; D_a varied from 0.029 for *Ci-fkh* to 0.09 for *Ci-Fgf4/5/6* (Table 2.4). In all trees and networks, Type B is consistently paraphyletic with respect to *C. roulei*, a pattern which was evident in the original mtCOI tree (Nydam & Harrison 2007). Where multiple *C. roulei* individuals are included (all trees or networks except those for *Ci-Fgf4/5/6*), these individuals never form a monophyletic group within the Type B clade (Figure 2.1). The S-H tests found no statistical differences for any of the genes between the unconstrained tree and the tree constrained by Type B monophyly ($p > 0.05$), (*CiCesA*: $p = 0.1828$, *Ci-Fgf4/5/6*: $p = 0.4595$, *Ci-fkh*: $p = 0.4393$, *jade:* $p = 0.4989$, mtCOI: $p = 0.4636$, *Ci-Patched:* $p = 0.4675$). The S-H test also found no statistical differences between the unconstrained tree and the tree constrained by *C. roulei* monophyly for three genes ($p > 0.05$), (*Ci-fkh*: $p = 0.1333$, mtCOI: $p = 0.4944$, *Ci-Patched:* $p = 0.3529$). However, two of the genes showed a significantly better likelihood score for the unconstrained tree than the tree constrained by *C. roulei* monophyly ($p > 0.05$), (*CiCesA*: $p = 0.0309$, *jade:* $p = < 0.001$).

Using only the mtCOI marker, the relationship of *Ciona* sp. to its congeners is not clearly defined; MP and ML methods are unable to resolve the relationships among *Ciona* sp., Type A and Type B/*C. roulei,* whereas NJ methods placed *Ciona* sp. as sister to the *C. intestinalis* clade (Nydam & Harrison 2007). In the nuclear gene trees, where *Ciona* sp. individuals can be associated with one type or the other, they are always associated with Type B, except the *Ci-Fgf4/5/6* tree where *Ciona* sp. is in the Type A clade (Figure 2.1; ML trees are not shown). Divergence estimates are smaller for Type B vs. *Ciona* sp. than for Type A vs. *Ciona* sp. for all loci except *Ci-Fgf4/5/6* and mtCOI (Table 2.4).

In order to assess support for defined topologies with respect to placement of *Ciona* sp., we compared the unconstrained best-scoring ML tree for each locus to the

best-scoring tree constrained by Type B/*C. roulei/Ciona* sp. monophyly. For four of the loci, the unconstrained tree was not significantly different from the constrained tree ($p > 0.05$), (*vAChTP*: $p = 0.3167$, *CiCesA*: $p = 0.1323$, *Ci-Fgf4/5/6*: $p = 0.2684$, *Ci-fkh*: $p = 0.4369$). For the remaining three loci (*jade*, mtCOI, *Ci-Patched*), the unconstrained tree was significantly better supported than the constrained tree ($p < 0.05$), (*jade*: $p = 0.0304$, mtCOI: $p = 0.000$, *Ci-Patched*: $p = 0.0411$).

The concatenated data set gave a different result (with respect to the placement of *Ciona* sp.) using ML and Bayesian algorithms. In ML analysis *Ciona* sp. is a clade distinct from the well-supported Type A and Type B/*C. roulei* clades; but relationships among these clades cannot be resolved (Figure 2.3a). However, in the Bayesian analysis, Type B/*C. roulei/Ciona* sp. form a well-supported monophyletic group, as does Type A (Figure 2.3b). Each of the loci used in the concatenated data set have a substantial number of phylogenetically informative sites, so we view these concatenated trees as the product of information assembled from across the genome, rather than from one or two dominant loci.

The π values averaged across all six nuclear loci are 0.0094 (Type A), 0.0361 (Type B), and 0.0324 (*C. roulei*). The θ_w values are 0.0104 (Type A), 0.0388 (Type B), and 0.03 (*C. roulei*). Levels of polymorphism were highly variable across the six nuclear loci (Table 2.5). The two loci determined to have high level of heterozygosity in the Type A genome (*Ci-Fgf/4/5/6* and *jade*) had the highest π and θ_w values for nearly all loci and species.

Discussion

Relationships within Ciona: molecules and morphology

Our data suggest that levels of morphological and molecular differentiation are

Figure 2.3 Trees obtained from concatenating the seven loci into one data set. The outgroup is *Ciona savignyi*. a) Maximum Likelihood: The best-scoring ML tree was obtained using the program RAxML v. 7.0.0. Values on the branches are bootstrap values. All nodes with less than <50% support collapsed. b) Bayesian: 50% majority-rule consensus trees created in PAUP*4.0 from trees obtained by MrBayes 3.1.2. Values on the branches are Bayesian posterior probabilities. Those nodes with a posterior probability value less than 50 were collapsed.

Table 2.5: Polymorphism within each species. Polymorphism is measured by π, nucleotide diversity, calculated as the weighted average of the proportion of nucleotide differences among all sequences in the population, and Θw, (4Neμ) calculated from the number of segregating sites.

Species	Measure	*vAChTP*	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*
Type A	π	0.0027	0.0024	0.0097	0.0070	0.0225	0.0130
	Θ_w	0.0023	0.0029	0.0117	0.0094	0.0220	0.0164
Type B	π	0.0006	0.0340	0.0460	0.0382	0.0616	0.0363
	Θ_w	0.0009	0.0488	0.0446	0.0473	0.0529	0.0382
C. roulei	π	0.0009	0.0190	NA*	0.0343	0.0811	0.0266
	Θ_w	0.0009	0.0185	NA*	0.0343	0.0699	0.0266

* Only one *C. roulei* individual was sequenced for *Ci-Fgf4/5/6*

not correlated within the shallow-water *Ciona* species. Morphologically similar entities (Type A and Type B) are highly differentiated at all loci examined, whereas morphologically distinct "species" (*C. roulei* and Type B) cannot be distinguished based on the seven loci we surveyed. Indeed, *C. roulei* is always embedded within Type B and never forms a monophyletic group within the Type B clade for any of the loci used.

Previous workers have always considered *C. roulei* morphologically distinct from *C. intestinalis* (Lahille 1890; Harant & Vernieres 1933; Fiala-Medioni 1974; Lambert et al. 1990). Harant & Vernieres (1933) and Lahille (1890) compared *C. roulei* to sympatric *C. intestinalis* (Type A) using multiple individuals from several populations of each species; these studies remain the only detailed morphological descriptions of *C. roulei* (Lahille 1890; Harant & Vernieres 1933). Although several of these distinguishing traits are variable within *C. intestinalis,* these variations do not overlap with *C. roulei*'s phenotype. For instance, *C. pulchella*, (subsequently synonymized with *C. intestinalis* (Hoshino & Nishikawa 1985)), from Devon, Cornwall and Guernsey (English Channel coast) was described as a new species distinct from *C. intestinalis* owing in part to its coloration (reddish, pale yellow or hyaline white) (Alder & Hancock 1907). The coloration of this morph is distinct from *C. roulei,* however (personal observation; Lahille 1890; Harant & Vernieres 1933). A orange morph of Type B was also described by Millar (Millar 1953), but this morph has little or no pigment in the transverse bars of the branchial sac (Millar 1953), whereas the reddish coloration of *C. roulei* is due to pigment in the peribranchial and transverse vessels of the branchial sac (Lahille 1890). Likewise, the tunic of *C. intestinalis* has been described as both "smooth and soft" (Alder & Hancock 1907) and "wrinkled" (Harant & Vernieres 1933); the tunic morphology of *C. intestinalis* varies with location, habitat and depth (personal observation). But *C. roulei*'s tunic is

consistently smooth (personal observation; Lahille 1890; Harant & Vernieres 1933).

Although these morphological differences were based on comparisons with Type A, the two types of *C. intestinalis* are extremely similar morphologically. In fact, they were not distinguished at all before molecular data became available, despite a thorough anatomical study of two populations of each type (Hoshino & Nishikawa 1985). Furthermore, the few traits by which they may differ are not the same traits that differentiate *C. roulei* and Type A (Caputi et al. 2007). Therefore, the morphological traits that separate Type A and *C. roulei* must also separate Type B and *C. roulei*.

Because *Ciona roulei* and Type B are genetically inseparable yet morphologically distinct, genetic and morphological data conflict with regard to whether *C. roulei* should be considered a species separate from Type B. According to the Biological Species Concept, the inability of two taxa to interbreed confers species status on both taxa, regardless of genetic similarity (Mayr 1995). Therefore, the presence of reproductive incompatibilities would enable us to label *C. roulei* as a species distinct from Type B. While we know that sympatric *C. roulei* and Type A are incompatible in one direction (Lambert et al. 1990), compatibility between the allopatric *C. roulei* and Type B has not been determined as *C. roulei* is now difficult to obtain. Without data on reproductive isolation between these two types, we have no clear evidence supporting species status for *C. roulei.*

The only evidence for the distinctness of *C. roulei* and Type B lies in their morphological differences, as several genetic observations fail to discriminate these two taxa. First, for all the loci for which multiple *C. roulei* individuals were sequenced, haplotypes found in *C. roulei* are interspersed in the Type B clade (*i.e.* some haplotypes of *C. roulei* are more similar to Type B haplotypes than to other *C. roulei* haplotypes). At one locus, *vaChTP,* Type B and *C. roulei* even share

haplotypes. Second, at two of five loci tested, an unconstrained tree had a statistically better likelihood than a tree constrained by *C. roulei* monophyly. Therefore, it is likely that *C. roulei* is recently derived from Type B individuals that invaded the Mediterranean Sea. It is possible that the observed morphological differences might reflect the new environmental conditions encountered in the Mediterranean Sea. Without additional data, we cannot distinguish between scenarios in which these differences represent genetic changes in response to a new selection regime and scenarios that involve direct responses to the environment (phenotypic plasticity).

Ciona sp., which is found in the Mediterranean Sea, appears to be more closely related to Type B than Type A. However, while six loci result in trees in which *Ciona* sp. is associated with Type B rather than Type A, the S-H tests show that only four loci provide significant support for an association of *Ciona* sp. with Type B/*C. roulei* rather than as sister to a *C. intestinalis/C. roulei* clade. Trees based on the concatenated data give different results depending on which tree-building algorithm is used. Therefore, our results lead us to support an association of *Ciona* sp. with Type B/*C. roulei* to the exclusion of Type A. However, analyses that include additional *Ciona* sp. individuals and additional loci are necessary to resolve these relationships.

The morphology of *Ciona* sp. is distinct from that of *Ciona* located in this region (for a description see (Nydam & Harrison 2007), including *Ciona edwardsi* (Roule, 1886), a species which is extremely rare (Copello 1981), and for which specimens could not be obtained. In the case of *Ciona* sp., genetic and morphological data agree that this species is distinct from other *Ciona,* although no data are available on reproductive compatibilities between this species and other *Ciona* species.

Levels of polymorphism

Based on whole-genome annotation and analyses from single individuals, both

Ciona intestinalis and *C. savignyi* have previously been shown to have very high levels of single-nucleotide polymorphisms, but levels of heterozygosity vary widely across the genome (Dehal et al. 2002; Small et al. 2007). Indeed, it was argued (Small et al. 2007) that *C. savignyi* harbors the highest levels of SNP variation in a multi-cellular organism. In *Ciona,* high levels of variation have been attributed to large effective population size. This explanation has been proposed not only for sea squirts but also for the Pacific oyster, *Crassostrea gigas*, (Sauvage et al. 2007) and the nematode *Caenorhabditis remanei* (Cutter et al. 2006).

For the loci *Ci-Fgf4/5/6, Ci-fkh and Ci-Patched,* the sequences we obtained are entirely noncoding, so we can compare π for these three loci to π for noncoding sites in the oyster and nematode loci. Average π values across loci are 0.038 in oysters (Sauvage et al. 2007) and 0.051 in nematodes (Cutter et al. 2006), whereas averages for Type A, Type B, and *C. roulei* are 0.01, 0.04 and 0.03. θ_w values are also similar between nematodes and *Ciona* for noncoding sites (not calculated for oysters).

The variability in amounts of population-level polymorphism for all three Ciona "species" among the six nuclear loci examined in this paper confirms the pattern seen in the single Type A individual from which genome sequence was obtained. Not only is this variability also present in populations of Type B and *C. roulei*, but levels of diversity in these taxa are substantially higher than in Type A at all six loci.

We cannot currently explain the difference in polymorphism between Type A and B *C. intestinalis.* Available evidence does not support the notion that differences in current effective population size can explain the remarkable amount of variation found in *Ciona intestinalis*. Type A, as a cosmopolitan species, likely has a larger effective population size than the geographically restricted Type B. And although a larger number of Type A than Type B populations are invasive and may have experienced a reduction of genetic diversity, the analysis of Boffelli et al. shows both

native and invasive Type B populations with much higher heterozygosity than either invasive or native Type A populations (2004).

Levels of polymorphism in *C. roulei* are similar to those in Type B, even though the *C. roulei* estimates were obtained from only 2-3 individuals, whereas Type B estimates were obtained from 11-13 individuals. Given the paraphyly of Type B with respect to *C. roulei,* we previously proposed a scenario in which Type B individuals invaded the Mediterranean Sea from the North Atlantic Ocean (Nydam & Harrison 2007). This invasion must have occurred relatively recently, given that *C. roulei* haplotypes are embedded within Type B haplotypes. Although the signature of a bottleneck may be obscured by population expansion, the current diversity of *C. roulei* does not support the idea that the ancestors of this species suffered a substantial reduction in genetic diversity due to a bottleneck effect when entering the Mediterranean Sea.

The genus *Ciona* provides an excellent system for investigating the evolution of molecules and morphology. *Ciona* species pairs represent the entire range of possible combinations of morphological and genetic divergence. Type A and B are cryptic species: divergent at the DNA sequence level but morphologically nearly identical. Discovery of cryptic species in the marine environment is accelerating with the widespread use of molecular tools; examining the biology of existing cryptic species will allow us to understand how and why substantial molecular divergence has evolved without concomitant morphological change. In contrast, Type B and *C. roulei* are genetically indistinguishable but morphologically divergent. This species pair provides an opportunity to address questions in evolutionary ecology related to local adaptation and phenotypic plasticity. Finally, *Ciona* sp. is both morphologically and molecularly distinct from all other *Ciona* spp. The discovery of a previously unknown and morphologically distinct *Ciona* species in a harbor immediately adjacent

to a marine laboratory highlights how little is known about a genus that includes important model organisms. Surely a continued effort to understand diversity within the genus *Ciona* will increase the utility of *C. intestinalis* and *C. savignyi* as models for diverse biological processes. Levels of polymorphism in natural populations are similar to levels reported in other organisms that are considered to be highly polymorphic (nematodes and oysters). This study provides the first estimate of population-level heterozygosity in *Ciona* and provides a critical first step in elucidating the origin and maintenance of these extraordinary high levels of polymorphism.

Acknowledgements

We would like to thank members of the Harrison Lab, C. Aquadro, and A. Clark for giving advice and commenting on an earlier version of the manuscript, D. Chong for help with data collection, J. Bishop for providing logistical assistance and G. Miller-Messner for collecting the Japanese samples. This work was funded by an AMNH Lerner-Gray Grant for Marine Research, a Cornell University Research Travel Grant to M.L.N., and by NSF grant DEB-0639904 to R.G.H.

REFERENCES

Alder, J., Hancock, A. 1907. The British Tunicata. Volume 2. Ray society, London.

Aparicio, S., Chapman, J., Stupka, E. et al. 2002. Whole Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes*. Science 297:1301-1310.

Boffelli, D., Weer, C.V., Weng, L., Lewis, K., Shoukry, M.I., Pachter, L., Keys, D.N., Rubin, E.M. 2004. Intraspecies sequence comparisons for annotating genomes. Genome Res. 14:2406-2411.

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M. and Sordino, P. 2007. Cryptic speciation in a model invertebrate chordate. Proc. Natl. Acad. Sci. USA 104:9364-9369.

Copello, M. 1981. Ciona Edwardsi (Roule, 1886) Espece Littorale de Mediterranee Distincte de Ciona intestinalis (Linne, 1767). Vie Milieu 31:243-253.

Cutter, A., Baird, S.E., Charlesworth, D.A. 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei.* Genetics 174:901-913.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A. 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298:2157-2167.

Fiala-Medioni, A. 1974. Ascidians of the Rocky Benthos of Banyuls-sur-Mer. Faunistic Inventory and Ecology Notes. Vie Milieu. 24:193-207.

Harant, H., and P. Vernieres. 1933. Tuniciers. Faune de France 27:1-93.

Herdman, W. A. 1882. Report on the Tunicata collected during the voyage of the HMS Challenger during the years 1873-1876. Part I: Ascidiae simplices. Report on the Scientific Results of the Voyage of HMS Challenger 6:236-237.

Hoshino, Z. I., Nishikawa, T. 1985. Taxonomic studies of Ciona intestinalis (L.) and its allies. Publ. Seto Mar. Biol. Lab. 30:61-79.

Huson, D. H., Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Johnson, D. S., Davidson, B., Brown, C., Smith, W., Sidow, A. 2004. Noncoding regulatory sequences of *Ciona* exhibit strong correspondence between evolutionary constraint and functional importance. Genome Res. 14:2448-2456.

Kott, P. 1952. The Ascidians of Australia. Australian Journal of Marine and Freshwater Research. 3: 206-233.

Lahille, F. 1890. I. Recherches sur les Tuniciers II. Propositions donnees par la Faculte. Faculte des Sciences Paris. Toulouse.

Lambert, C. C., LaFargue, F., Lambert, G. 1990. Preliminary note on the genetic isolation of Ciona species (Ascidiacea, Urochordata). Vie Milieu 40:293-295.

Linne, C. 1767. Systema Naturae. Vindobonae: Typis Ioannis Thomae, 1767-1770.

Maddison, W. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523-536.

Mayr, E. 1995. Species, classification, and evolution. Pp. 3-12 *in* R. Arai, Kato, M.,

and Doi, Y., ed. Biodiversity and Evolution. National Science Museum Foundation, Tokyo.

Millar, R. H. 1953. Ciona. The University Press of Liverpool, Liverpool.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nichols, R. 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16:37-45.

Nydam, M.L, Harrison, R.G. 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). Mar. Biol. 151:1839-1847.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12:357-358.

Posada, D., Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Ronquist, F., Huelsenbeck, J.P. 2003. MrBayes3: Bayesian inference under mixed models. Bioinformatics 19:1572-1574.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Sauvage, C., Bierne, N., Lapegue, S., Boudry, P. 2007. Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. Gene 406:13-22.

Small, K. S., Brudno, M., Hill, M.M., Sidow, A. 2007. Extreme genomic variation in a natural population. P. Nat. Acad. Sci. USA 104:5698-5703.

Stamatakis, A. 2008. A Rapid Bootstrap Algorithm for the RAxML Web-Servers. Syst. Biol. 57:758-771.

Suzuki, M., Nishikawa, T., and Bird, A. 2005. Genomic Approaches Reveal Unexpected Difference within Ciona intestinalis. J. Mol. Evol. 61:627-635.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b 10. Sinauer Associates, Sunderland.

Tamura, K., Dudley, J., Nei, M., Kumar, S. 2007. Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596-1599.

Van Name, W. 1945. The north and south American ascidians. Bull. Am. Mus. Nat. Hist. 84:1-476.

Vinson, J., Jaffe, D., O'Neill, K., Karlsson, E., Stange-Thomann, N., Anderson, S., Mesirov, J., Satoh, N., Satou, Y., Nusbaum, C., Birren B., Galagan, J., Lander, E. 2005. Assembly of polymorphic genomes: Algorithms and application to *Ciona savignyi*. Genome Res. 15:1127-1135.

CHAPTER 3

INTROGRESSION DESPITE SUBSTANTIAL DIVERGENCE IN A BROADCAST SPAWNING MARINE INVERTEBRATE

Abstract

Understanding the relationship between reproductive isolation and time since divergence is critical to our understanding of speciation. One group for which we know little about the relationship between hybridization/introgression and time since divergence is the marine broadcast spawners. Here, we investigate the distribution of closely related cryptic species of marine broadcast spawners (Type A and B *Ciona intestinalis*) in areas of potential sympatry to determine whether these two types occur together and if so, whether they show evidence of hybridization and introgression. Then we combine our data with other studies to investigate general patterns of reproductive isolation vs. divergence in marine broadcast spawners.

We found that Type A and B *C. intestinalis* occurred sympatrically in 2007, and that 21 individuals show evidence of introgression in sympatry (out of ~500 sympatric individuals). Type A and B *C. intestinalis* are 12.4% divergent at mitochondrial COI (mtCOI) and between 3.5-11.6% divergent at six nuclear loci, and in comparison with other marine broadcast spawning species at mtCOI, these two types may be near the upper limit of the range of divergence values where introgression is still possible. However, introgression at divergence levels similar to those found in *Ciona* does occur, prompting questions about the strength of postmating prezygotic reproductive barriers in marine broadcast spawners.

Introduction

Understanding the relationship between reproductive isolation and time since divergence is critical to our understanding of speciation. Therefore, examining hybridization and introgression between species pairs with different times since divergence has been an active area of research since the 1970s. The first studies, by Wilson and colleagues (Wilson et al. 1974; Prager & Wilson 1975) used immunological distances between albumin proteins as a proxy for time since divergence in pairs of mammal, bird and frog species pairs. Subsequent studies, in groups as diverse as amphibians, angiosperms, fish, insects and sea urchins use distances based on allozyme or DNA sequence data to estimate time since divergence (Coyne & Orr 1989; Coyne & Orr 1997; Mendelson 2003; Moyle et al. 2004; Price & Bouvier 2002; Sasa et al. 1998; Zigler et al. 2005). These studies have shown that rates of hybridization and introgression generally decline with increasing time since divergence (Edmands 2002). However, this decline varies across taxa, with certain groups showing evidence of hybridization and introgression despite substantial time since divergence (Lamb & Avise 1986; Freyhof et al. 2005; Koblmuller et al. 2007).

One group for which we know very little about the relationship between hybridization/introgression and time since divergence is the marine broadcast spawners. Broadcast spawners release sperm and egg into the water column, where fertilization occurs. Therefore, behavioral and mechanical barriers are absent in broadcast spawning organisms. Temporal isolation can be an important premating barrier to fertilization between marine species (Levitan et al. 2004), but if two broadcast spawning species are not temporally isolated, postmating prezygotic barriers (involving gamete recognition) are often the only barriers preventing interspecific fertilization. In some taxa, hybridization and introgression have been observed despite

the existence of these prezygotic barriers (Levitan 2002; Harper & Hart 2005). Because broadcast spawning species possess fewer and possibly less efficient prezygotic barriers to hybridization and introgression, gene flow might be expected to persist even after long intervals of separation. Thus, broadcast spawning species are an interesting group in which to examine the relationship between hybridization/introgression and time since divergence.

The widespread and invasive broadcast spawning ascidian *Ciona intestinalis* has been shown to comprise two distinct, highly divergent entities not distinguished until recently (Suzuki et al. 2005; Caputi et al. 2007; Nydam & Harrison 2007). Although morphologically cryptic (Caputi et al. 2007), these two types, hereafter referred to as Type A and B, are ~12% divergent at the mtCOI locus (Nydam & Harrison 2007). Type A (the type for which a genome has been sequenced) is thought to be native to the Northwestern Pacific Ocean and has invaded the Eastern Pacific Ocean, the Mediterranean Sea, the Atlantic coast of South Africa, and the Black Sea (Van Name 1945; Kott 1952). Type B, thought to be native to the Northern Atlantic Ocean (Linne 1767; Monniot & Monniot 1994), has invaded the Western Atlantic Ocean (Nydam & Harrison 2007). If Type A has invaded the native range of Type B, these two types potentially overlap along the Atlantic coast of France and the English Channel coasts of France and England.

A recent study combined fertilization data with microsatellite and sequence data from several populations to conclude that these two types should be considered different species (Caputi et al. 2007). In the laboratory, both prezygotic (Suzuki et al. 2005) and postzygotic barriers (Caputi et al. 2007) occur between Type A and B individuals from allopatric populations. Genetic data from Caputi et al. (2007) confirmed previous observations (Suzuki et al. 2005; Nydam & Harrison 2007) that the two types were both distinct and genetically divergent from one another.

However, only two locations in the area of potential overlap were sampled. Therefore, the extent of hybridization or introgression (if any) between these two highly divergent mitochondrial types is unknown.

Here, we investigate the distribution of Type A and B in areas of potential sympatry to determine whether these two types occur together and if so, whether they show evidence of hybridization and introgression. Then we combine our data with other studies to investigate general patterns of reproductive isolation vs. divergence in marine broadcast spawners. We focus on introgression throughout our study and in comparisons with other studies. Most studies examine the relationship of hybridization to time since divergence because hybridization is often easier to quantify than introgression. We feel, however, that introgression is more informative for our understanding of the speciation process, as only introgression provides definitive evidence of gene flow.

Materials and Methods

Sampling

In the summer of 2007, approximately 20 individuals were collected from each of 19 sampling locations, from La Rochelle to Granville (France) and Falmouth to Ramsgate (United Kingdom). No *C. intestinalis* were found in Arcachon, France (the southernmost site in the study). In the summer of 2009, approximately 50 individuals were collected from each of the 2007 locations that were found to contain both types and/or individuals of mixed ancestry. One marina was sampled at each location, with the exception of Plymouth, United Kingdom, where two marinas (Queen Anne's Battery and Sutton Harbour) and one laboratory (seawater system of the Marine Biological Laboratory) were sampled. Color of the spermiduct and the spermiduct

papillae were noted for each individual, as these features are potential characters for type discrimination (Caputi et al. 2007).

Type Determination

Because morphological features do not allow reliable discrimination of Types A and B, we identified type-specific SNPs for each of seven loci, based on data for individuals from allopatric populations. The loci are: Vesicular acetylcholine transporter (*vAChTP*), Cellulose synthase (*CiCesA*), Fibroblast Growth Factor orthologous to vertebrate Fibroblast Growth Factor 4/5/6 (*Ci-Fgf4/5/6*), Forkhead (*Ci-fkh*) (5' regulatory region), Jade (*jade*), mtCOI, and Patched (*Ci-Patched)*. For each locus, a restriction enzyme was chosen that would cut a Type A sequence but not a Type B sequence (or vice versa) based on the sequence difference between the types at the SNP site. All seven gene regions were amplified in all individuals collected in the zone of sympatry (primers and thermocycling conditions available from the authors). Each PCR product was then cut with the appropriate restriction enzyme and run on an agarose gel. The type of each individual was determined from the bands displayed, with individuals of known type serving as controls. Where the banding pattern could not be unambiguously assigned to a type, the PCR product for that locus was sequenced to determine the type. The *Ci-fkh* locus gave erratic banding patterns with several different SNP/enzyme combinations and thus digests of this locus were not involved in type determination.

Sequencing of Individuals possessing both Type A and Type B alleles

For those individuals for which type appeared inconsistent across loci, the PCR product for each locus was cloned using the pGEM®-T kit and up to eight clones were sequenced. Additionally, new primers were developed for each locus (in slightly

different locations from the primers used to create the genealogies and type the sympatric zone individuals). For each locus in each individual of putative mixed ancestry, PCR products from these new primers were cloned and sequenced to circumvent any possible PCR bias of the original primer for one of the two alleles. PCR products were incubated with 0.25µl each of Exonuclease I and Shrimp Antarctic Phosphatase at 37°C for 30 min, followed by 90°C for 10 min. The products were purified using CleanSeq beads (Agencourt). The purified product was sequenced with a Big Dye Terminator Cycle sequencing kit and an Automated 3730 DNA Analyzer (Applied Biosystems). All unique haplotypes have been submitted to GenBank (Accession Numbers XXX-XXX). Sequences were edited, trimmed and aligned with Aligner (CodonCode Corporation, Dedham, MA).

Calculation of Hardy-Weinberg Equilibrium

For each locus, we calculated Hardy-Weinberg Equilibrium for each population where both Type A and B individuals were found. Calculations were done using a web tool (Rodriguez et al. 2009).

Principal Components Analysis of Individuals with mixed ancestry

The genotype of each individual (A/A, A/B, or B/B) at each of the 7 loci was determined from sequence data. No distinction between different Type A or B alleles at a particular locus was made for this analysis. The statistical package ade4 (Dray & Dufour 2007) in R 2.10.0 was then used to perform a principal components analysis on the genotypes of the 21 individuals of mixed ancestry.

Results

In 2007, Type A and B individuals were found to coexist in six locations: two on the Atlantic coast of France (Concarneau and Camaret-sur-mer), one on the English Channel coast of France (Perros-Guirec) and three on the English Channel coast of the United Kingdom (Falmouth, Plymouth (both marinas), and Torquay) (Figure 3.1a). In two of the sympatric locations (Concarneau, France and Plymouth, UK), Type A individuals represented the majority, in Perros-Guirec, France the two types were present in equal numbers and in the remaining three sites (Camaret-sur-mer, France, Falmouth/Torquay, UK) Type B individuals outnumbered Type A individuals (Figure 3.1a). Although we found Type A individuals in many locations, no locations contained only Type A individuals. In 2009, when only these six sympatric locations were re-sampled, we found that pure Type A had completely disappeared from all locations, except for one individual in Camaret-sur-mer (Figure 3.1b). Many locations contained only Type B individuals; every individual from these locations was homozygous for the B allele (*i.e.* a single band was seen on the restriction digest gel).

799 individuals were surveyed across the 2 years. 55 individuals were pure Type A, 723 individuals were pure Type B, and 21 individuals carried alleles characteristic of both Type A and B (Table 3.1). The individuals with both Type A and B alleles came from six locations (Camaret-sur-mer, Perros-Guirec and Granville, France, and Falmouth, Plymouth, and Torquay, UK). Two *C. intestinalis* individuals of mixed ancestry were collected from Granville; it is therefore likely that Granville is or was a sympatric location even though no pure Type A individuals were found there.

All loci that we examined have both Type A and B alleles (Table 3.1). A majority Type A individual has Type A alleles at 4 or more of the 7 loci we sequenced. A majority Type B individual has Type B alleles at 4 or more of the 7 loci

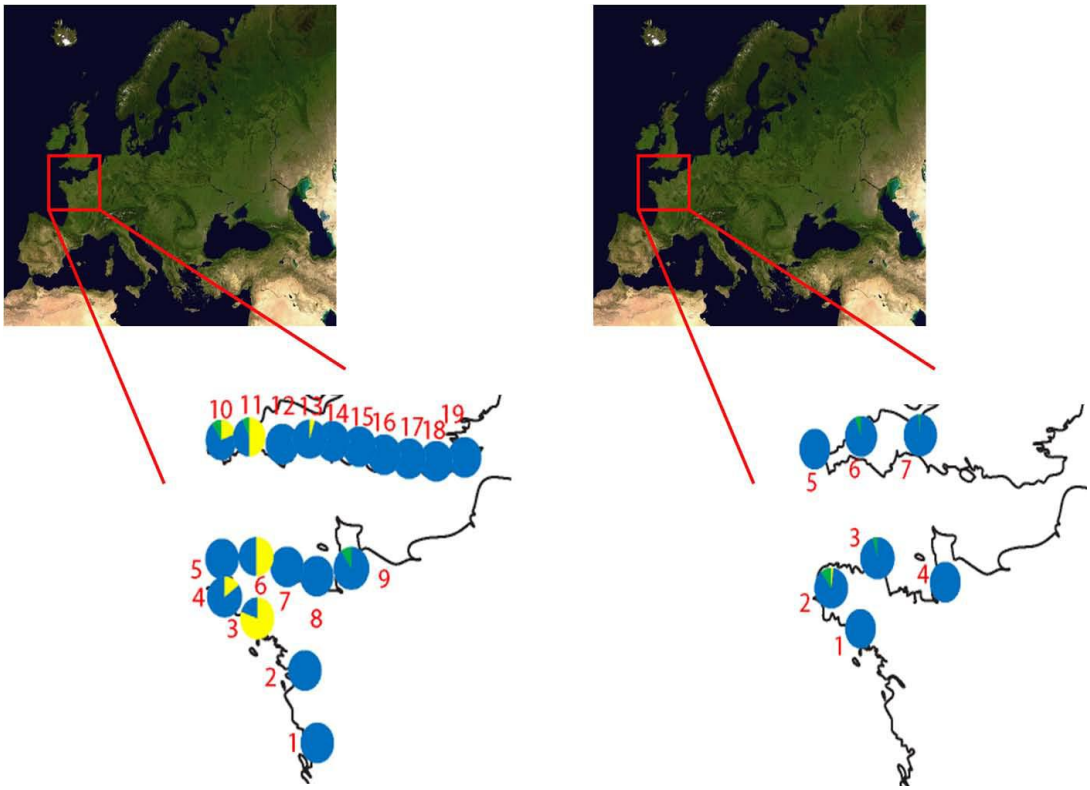

Figure 3.1 (a): Distribution of Type A and B in the sympatric zone (2007) Type A*:* Yellow, Introgressed: Green, Type B: Blue **1:** La Rochelle **2:** Pornic **3:** Concarneau **4:** Camaret-sur-mer **5:** Brest **6:** Perros-Guirec **7:** Lezardrieux **8:** Saint-Servan **9:** Granville **10:** Falmouth **11:** Plymouth Marinas **12:** Plymouth MBA seawater system **13:** Torquay **14:** Poole **15:** Hamble **16:** Gosport **17:** Portsmouth **18:** Brighton **19:** Dover (b): Distribution of Type A and B in the sympatric zone (2009) Type A*:* Yellow, Introgressed: Green, Type B: Blue **1:** Concarneau **2:** Camaret-sur-mer **3:** Perros-Guirec **4:** Granville **5:** Falmouth **6:** Plymouth **7:** Torquay

Table 3.1 Introgressed individuals and genotype at each locus organized by introgressed loci (last column). Type A homozygotes are shown in blue, Type B homozygotes in red, and Type A/B heterozygotes in purple. The number after each Type A or B designation refers to a haplotype (e.g. *vaChTP* has four Type A haplotypes and two Type B haplotypes).

Individual	Location	*vAChTP*	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*	mtCOI	Majority Type (# of loci)	Minority Type (# of loci)
CiCSM37	Camaret, FR	A2/A2	A4/A4	A4/A4	A2/A2	A4/B7	A3/A3	A1	A (6)	A/B (1)
CiCSM51	Camaret, FR	A1/A1	A3/A3	A5/A5	A4/A4	A5/B8	A4/A4	A1	A (6)	A/B (1)
CiCSM57	Camaret, FR	A4/B1	A2/B7	A13/B10	A9/B10	B19/B19	A6/B10	A2	A/B (5)	A (1), B/B (1)
CiFM1	Falmouth, UK	B1/B1	B5/B5	B5/B5	B5/B5	B14/B14	B5/B5	A1	B (6)	A (1)
CiFM2	Falmouth, UK	A1/B2	A1/A1	A1/A1	A1/A1	A1/A1	A1/A1	A1	A (6)	A/B (1)
CiGr8	Granville, FR	A3/B1	A1/A1	A12/B9	A2/A2	A13/A13	A6/B9	A1	A (4)	A/B (3)
CiPG33	Perros-Guirec, FR	A1/A1	A3/A3	A2/A2	A2/A2	A2/B5	A2/A2	A1	A (6)	A/B (1)
CiPG36	Perros-Guirec, FR	A1/A1	A1/A1	A3/A3	A3/A3	A3/B6	A2/A2	A1	A (6)	A/B (1)
CiQAB21	Plymouth, UK	A1/A1	A3/A3	A6/A6	A5/A5	A6/B9	A5/A5	A1	A (6)	A/B (1)
CiQAB26	Plymouth, UK	A1/A1	A3/A3	A7/A7	A6/A6	A7/B10	A2/A2	A1	A (6)	A/B (1)
CiQAB28	Plymouth, UK	A1/A1	A3/A3	A8/A8	A7/A7	A8/B11	A6/A6	A1	A (6)	A/B (1)
CiQAB43	Plymouth, UK	A1/A1	A1/A1	A9/A9	A8/A8	A9/B12	A7/A7	A1	A (6)	A/B (1)
CiQAB45	Plymouth, UK	A1/A1	A3/A3	A10/A10	A7/A7	A10/B13	A6/A6	A1	A (6)	A/B (1)
CiTQ38	Torquay, UK	A1/A1	A4/A4	A11/A11	B7/B7	A11/B16	A8/A8	A1	A (5)	B/B (1), A/B (1)
CiCSM58	Camaret, FR	B1/B1	A2/B2	B2/B2	B2/B2	B2/B2	B2/B2	B2	B (6)	A/B (1)
CiCSM60	Camaret, FR	B1/B1	A1/B3	B3/B3	B3/B3	B3/B3	B3/B3	B3	B (6)	A/B (1)
CiGr10	Granville, FR	A3/B1	B6/B6	B8/B8	B9/B9	B18/B18	A6/B8	B5	B (5)	A/B (2)
CiQAB3	Plymouth, UK	A1/B1	A5/A5	B6/B6	B6/B6	B15/B15	B6/B6	B2	B (5)	A/A (1), A/B (1)
CiQAB12	Plymouth, UK	B1/B1	A3/B4	B4/B4	B4/B4	B4/B4	B4/B4	B4	B (6)	A/B (1)
CiQAB46	Plymouth, UK	B1/B1	A3/A3	B7/B7	B8/B8	A12/B17	B7/B7	B3	B (5)	A/A (1), A/B (1)
CiSHM11	Plymouth, UK	A1/B1	B1/B1	B1/B1	B1/B1	B1/B1	B1/B1	B1	B (6)	A/B (1)

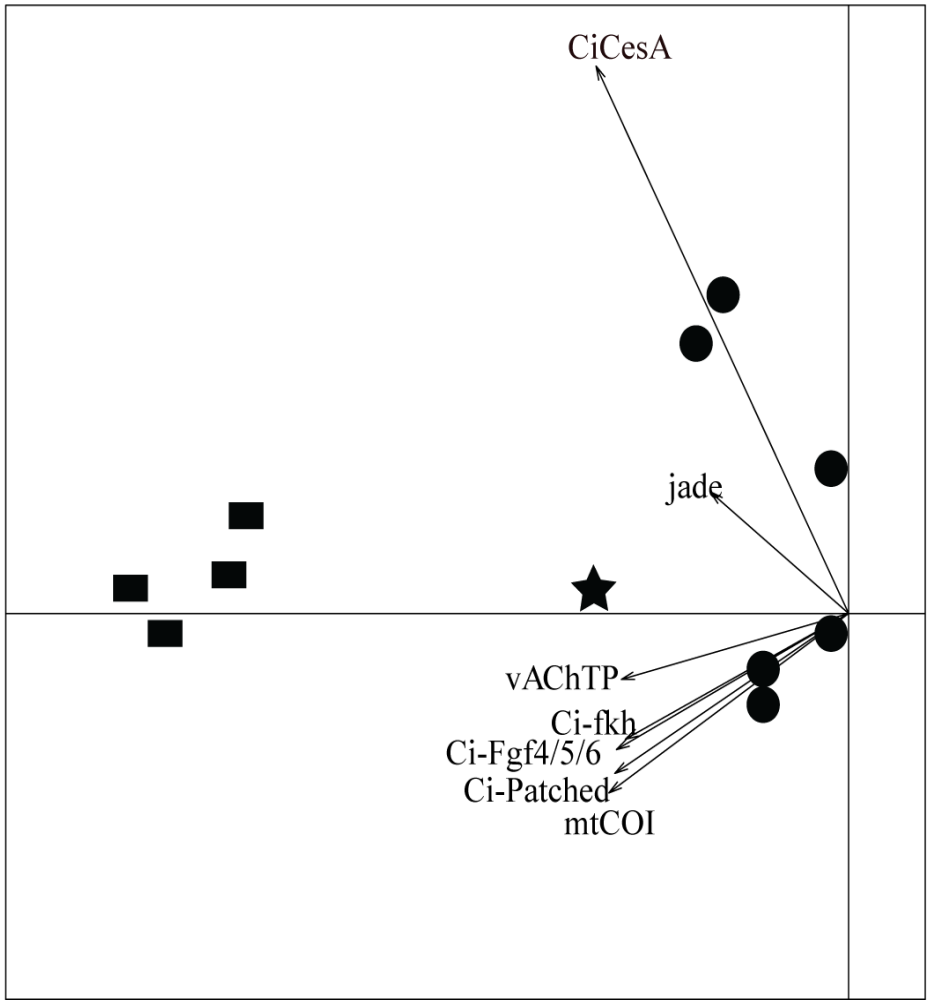

Figure 3.2 Scatterplot of the two largest principal components of the data. Rectangles represent individuals that are Type A at ≥ 4 loci, circles represent individuals that are Type B at ≥ 4 loci. The star represents the single individual that was heterozygous at ≥ 4 loci. The arrows correspond to the seven markers (six nuclear and mtCOI) used to perform the principal components analysis (*e.g.* the individuals represented by the two circles closest to the *CiCesA* label differ from other majority Type B individuals because they are Type A at *CiCesA*).

we sequenced. Of the individuals with both Type A and Type B alleles, 12 are majority Type A, 8 are majority Type B, and one is heterozygous at a majority of the loci. A principal components analysis (Figure 3.2) illustrates this result: the majority Type A individuals group together, as do the majority Type B individuals, while the majority heterozygous individual (CiCSM57) is located between these two groups.

Although Type A and B alleles are present for each locus, the seven loci appear to exhibit different rates of introgression between Type A and B genomes. Among the individuals of mixed ancestry, the *jade* locus was Type A in a Type B background (or vice versa) in 11/21 individuals, *vAChTP* and *CiCesA* loci in 5/21 individuals, and the remaining four loci in only 1 or 2 individuals (Table 3.1). The direction of introgression is, in some instances, asymmetric; for the *jade* locus in particular, alleles from Type B are often present in individuals with Type A genetic background, but Type A *jade* alleles are rarely found in a Type B background. This locus is therefore located in a region of the Type B genome that is likely to introgress into a Type A background. The opposite pattern is seen in the *CiCesA* locus – the Type A allele is always found in a Type B background. When *vAChTP* introgresses, alleles move in both directions: alleles of Type A are found in the Type B background and vice versa.

For each of the loci that introgressed in 5 or more individuals (*vAChTP*, *CiCesA,* and *jade*), multiple alleles were involved in the introgression events (Table 3.1). Regarding the two *vAChTP* individuals that are majority Type A, two different Type B alleles introgressed. For the three *vAChTP* individuals that are majority Type B, two different Type A alleles introgressed. In the five *CiCesA* individuals that are majority Type B, four different Type A alleles introgressed. And finally for *jade,* each individual has a unique *jade* Type A and/or Type B allele, so again multiple alleles were involved in the introgression events. These data provide evidence that

individuals of mixed ancestry derive from multiple hybridization events.

As only one pure Type A individual was found in the 2009 sampling, Hardy-Weinberg Equilibrium was calculated for the 2007 data only. All populations violated Hardy-Weinberg Equilibrium for all loci, with the exception of Granville, France, which was in Hardy-Weinberg Equilibrium for three loci (Table 3.2). In all cases, Hardy-Weinberg Equilibrium is violated owing to a deficit of heterozygotes.

Four morphological classes of *C. intestinalis* were found in the North Atlantic and English Channel, three of which were discussed previously (Caputi et al. 2007): WW (uncolored spermiducts and white-pigmented or absent spermiduct papillae), WO (uncolored spermiducts and orange-pigmented spermiduct papillae), OW (orange-pigmented spermiducts and white-pigmented or absent spermiduct papillae), and OO (orange-pigmented spermiducts and orange-pigmented spermiduct papillae). Of the individuals collected for this study, 98% of WW individuals (297/303), 98% of OW individuals (65/66) and 95% of OO individuals (20/21) were pure or majority Type B. 23% of individuals with the WO phenotype were pure or majority Type A individuals (49/209). The majority of the individuals of mixed ancestry for which phenotype could be scored were WO (84%, 16/19), 11% were WW (2/19) and 5% were OW (1/19).

Discussion

Formation of the current zone of sympatry

The 2007 distribution of Type A and B as determined from this study provides some insight into the establishment of an area of sympatry in the Northeast Atlantic Ocean. Because Type B was first described as *Ciona intestinalis* in the Northeast Atlantic Ocean (Linne 1767), Type B is considered native to the locations sampled in

Table 3.2 Hardy-Weinberg Equilibria. Numbers of individuals in each genotypic class for each sympatric population at each locus, and the p-value associated with the chi-squared value with one degree of freedom. Only 2007 data are represented, as pure Type A individuals were not present in 2009 samples.

vAChTP	AA	AB	BB	Result	p-value
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	0	2	18	HWE observed	>0.05
Sutton Harbour Plymouth	16	1	3	HWE violated	<0.001
Queen Anne's Plymouth	6	1	13	HWE violated	<0.001
Falmouth	4	1	15	HWE violated	<0.001
Torquay	1	0	19	HWE violated	<0.001
CiCesA	AA	AB	BB		
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	1	0	19	HWE violated	<0.001
Sutton Harbour Plymouth	16	0	4	HWE violated	<0.001
Queen Anne's Plymouth	7	1	12	HWE violated	<0.001
Falmouth	5	0	15	HWE violated	<0.001
Torquay	1	0	19	HWE violated	<0.001
Ci-Fgf4/5/6	AA	AB	BB		
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	0	1	19	HWE observed	>0.05
Sutton Harbour Plymouth	16	0	4	HWE violated	<0.001
Queen Anne's Plymouth	6	0	14	HWE violated	<0.001
Falmouth	5	0	15	HWE violated	<0.001
Torquay	1	0	19	HWE violated	<0.001
Ci-fkh	AA	AB	BB		
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	1	0	19	HWE violated	<0.001
Sutton Harbour Plymouth	16	0	4	HWE violated	<0.001
Queen Anne's Plymouth	6	0	14	HWE violated	<0.001
Falmouth	5	0	15	HWE violated	<0.001
Torquay	1	0	19	HWE violated	<0.001
jade	AA	AB	BB		
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	1	0	19	HWE violated	<0.001
Sutton Harbour Plymouth	16	0	4	HWE violated	<0.001
Queen Anne's Plymouth	6	0	14	HWE violated	<0.001
Falmouth	5	0	15	HWE violated	<0.001
Torquay	1	0	19	HWE violated	<0.001
Ci-Patched	AA	AB	BB		
Concarneau	16	0	4	HWE violated	<0.001
Camaret	3	0	17	HWE violated	<0.001
Perros-Guirec	10	0	10	HWE violated	<0.001
Granville	0	2	18	HWE observed	>0.05
Sutton Harbour Plymouth	16	0	4	HWE violated	<0.001
Queen Anne's Plymouth	6	0	14	HWE violated	<0.001
Falmouth	5	0	15	HWE violated	<0.001

this study. Type A, however, is cosmopolitan, having invaded a large portion of the temperate oceans through anthropogenic transport (Lambert & Lambert 1998), which is likely how it reached the English Channel.

Although Type A is abundant in the Mediterranean Sea (Kott 1990; Caputi et al. 2007), its absence in the locations along the Atlantic coast of France sampled in 2007 argues against the Mediterranean as a source of dispersal into the Northern Atlantic. In a similar hybrid zone, *Mytilus galloprovincialis* occurs from the Mediterranean up the Atlantic coast of the Iberian Peninsula into the Atlantic coast of France and the English Channel, where it hybridizes with *M. edulis* (Daguin et al. 2001; Bierne et al. 2003). However, the *M. galloprovincialis* on the Iberian Peninsula are genetically distinct from Mediterranean *M. galloprovincialis* and it is the former that are thought to be the source for the hybridizing individuals along the Atlantic coast of France (Daguin et al. 2001). Because Type A was restricted to the westernmost sites in the southern UK (Falmouth, Plymouth and Torquay), and to the four westernmost sites in France (Concarneau, Camaret-sur-mer, Perros-Guirec (this study, 2007), and Brest (previously reported by Caputi et al. 2007)), we suggest that Type A was introduced directly to the western English Channel.

The two most plausible mechanisms for Type A introduction into this area are by ship (as larvae that metamorphose into juveniles in the ballast tank or ship's sea chest) or by attachment to shellfish transported to the area for aquaculture. The oyster *Crassostrea gigas* was introduced to the Atlantic and Bretagne coasts of France from Japan beginning in the late 1960s and later from the Pacific Coast of the United States and Canada (Gruet et al. 1976; Grizel & Héral 1991). Both of these source areas contain Type A populations, which could easily have attached to the imported oyster shells.

Type A would have had many opportunities to spread in the English Channel

after arriving in the Northern Atlantic. Ports with international shipping traffic (such as Plymouth and Brest) can serve as excellent starting places for transport within a region (Wasson et al. 2001). Ships making short trips in and around the Northern Atlantic (including extensive ferry traffic in the English Channel) could transport eggs and larvae in their ballast water and adults settled on their hulls.

The presumed ease of intra-Channel transport is not consistent with the 2007 distribution of Type A, which is restricted to the coast of Bretagne and the Western English Channel. Genetic data from this study support a scenario whereby Type A and B have been hybridizing for several generations (see below), so it is unlikely that the absence of Type A *C. intestinalis* from locations east of Torquay is due to a very recent arrival in the English Channel with little time to spread eastward. A plausible explanation for the restricted distribution of Type A in 2007 relates to temperature within the English Channel and Northern Atlantic around the coast of Bretagne. Winter mean sea surface temperatures are 4°C colder in the eastern Channel than in the western Channel (Hayward & Ryland 1995). Type A individuals are less tolerant of the colder temperatures encountered in the eastern English Channel than are the native Type B. The lower limit of temperature tolerance for Type A in the Gulf of Naples is 8°C, while Type B individuals on the West coast of Norway and Sweden can survive at 1°C (Dybern 1965).

It is possible that individuals of Type A are able to survive the colder winter temperatures in the eastern English Channel but that their growth and/or fitness are compromised, providing a competitive advantage to Type B individuals. The positive correlation between water temperature and aspects of fitness in Type A has been well documented (Sentz-Braconnot 1966; Yamaguchi 1975; Cirino et al. 2002).

When the sites that contained both types and/or individuals of mixed ancestry were re-sampled in 2009, pure Type A individuals were not found except for one

individual in Camaret-sur-mer, France. Sampling was limited to harbors, so Type A may still be present in deeper refugia, but the shift in the observed distribution was nevertheless dramatic. This shift is surprising, given that 35% of the *C. intestinalis* at these sites in 2007 was Type A (with a maximum of 80% Type A at Concarneau). Whether the Type A individuals declined due to environmental factors or were outcompeted by the Type B individuals is not known. While Type A is broadly invasive throughout the temperate oceans, this species has been known to decline after invasion, although no studies have addressed the causes of these declines. Type A records exist from all Australian ports (Kott 1990), but as of 1997 Type A could only be found in Port Phillip Bay, Victoria (Kott 1997) and subsequently in a single location in southwestern Australia (McDonald 2004). Type A was found in the Eastern Pacific as early as 1915 (Ritter & Forsyth 1917), but *C. savignyi* now occupies more sites in Southern California than Type A, despite having been first recorded in the area in 1985 (Lambert & Lambert 1998, 2003).

The fitness advantage of Type B over Type A is particularly noticeable at the *jade* gene, as 10 of the 12 majority Type A individuals have a B allele at *jade*. Although a large portion of the sequenced used was non-coding (77% in Type A, 91% in Type B, see Chapter 2, Table 2.2), the sequence we used is presumably closely linked to the adjacent coding regions. Homologues of the *Ciona jade* protein have been identified in mouse, human, zebrafish and a puffer fish (*Fugu rubripes*); the high degree of conservation throughout these lineages implies a critical function for this protein (Tzouanacou et al. 2003, Figure 6). This protein contains PHD zinc finger domains, which are found in proteins involved in chromatin-mediated transcriptional regulation (Aasland et al. 1995). *Jade* has been implicated in the development of the anterior-posterior axis in the mouse embryo (Tzouanacou et al. 2003), as well as the suppression of renal cancer in humans (Zhou et al. 2005).

Although Type A and B existed in sympatry in several English Channel locations in 2007, we acknowledge that this zone of sympatry may not be stable given the near-absence of Type A in 2009. However, our sampling was limited to harbors; we did not sample individuals from deeper benthic populations. Dramatic seasonal or annual fluctuations in abundances of harbor populations of *C. intestinalis* and other ascidians have been reported; harbor populations are often re-colonized from deeper water benthic populations (Lambert & Lambert 1998; Svane & Havenhand 1993). Future sampling of this area, including deeper water populations, is necessary to determine whether Type A continues to persist in this area.

Hybridization and Introgression vs. Ancestral polymorphism

The observation that 21 *C. intestinalis* individuals have Type A alleles in a Type B genetic background, or vice versa, almost certainly reflects recent introgression rather than incomplete lineage sorting of ancestral polymorphism. First, all but one of the individuals with both Type A and B alleles were found at sites where the two types were sympatric as recently as 2007. If incomplete lineage sorting of ancestral polymorphism explains patterns of variation, individuals with both Type A and B alleles should be found across the distribution of Type A and B, not restricted to areas of sympatry (Hare & Avise 1998; Masta et al. 2002; McGuire et al. 2008). Second, if sharing of Type A and B alleles results from persistence of ancestral polymorphisms, it is unlikely that single individuals would carry both Type A and B alleles at more than one locus. One would not expect 2-3 unlinked loci to show the same pattern of lineage sorting between two highly divergent lineages, given that lineage sorting is a random process. Third, the high divergence values between Type A and B (12.4% at mtCOI) make incomplete lineage sorting of ancestral polymorphism unlikely; assuming neutral evolution, the loss of polymorphism and the

fixation of type-specific alleles are a functional of the time since divergence (Maddison 1997; Wendel & Doyle 1998). Thus while introgression seems the likeliest explanation for the patterns we see, it is also possible that our allopatric individuals sampled thus far do not represent the total genetic variation within each type.

The pattern of introgression, whereby mixed-type individuals seem to be the product of multiple generations of backcrossing to either pure Type A or B individuals, suggests that these two types have been hybridizing for several years (Plymouth populations have 2 generations/yr, 3 if summer water temperatures are warmer than average (Dybern 1965)). The presence of only advanced generation backcross hybrids could result from two scenarios: 1) A relatively old hybrid zone with no ongoing hybridization between Type A and B or 2) A hybrid zone of at least several generations in duration with recent hybridization between Type A and B. Although we cannot rule out the possibility that previous gene flow has ceased to occur, introgression largely confined to locations of sympatry provides evidence for current gene flow (Coyne & Orr 2004).

Many hybrid populations in other taxa consist of few or no F1 individuals and numerous backcrossed individuals (Harrison & Bogdanowicz 1997; Bierne et al. 2003; Kronforst et al. 2006). The widely acknowledged explanation for this pattern is that pre or postzygotic isolating barriers between the parental species limit the production of F1 offspring. But once formed, F1 individuals will likely backcross to parental types; backcrossing may occur more readily than initial hybridization, because the interacting gametes share more of their genome than do gametes derived from the two parental types (Arnold 1997). Mallet (2005) acknowledges that there are few empirical data to support or refute this explanation for the paucity of F1s relative to backcrossed individuals, but he cites studies in butterflies that show F1 offspring

are more difficult to produce than backcrossed individuals (Mallet et al. 1998; Naisbit et al. 2003).

Because *C. intestinalis* is a model laboratory organism in developmental biology and a genome sequence is available for Type A, discrimination of the two genetically divergent types in the field has been an important goal since these two types were recognized. This discrimination becomes especially critical in sympatric populations from the Bretagne region of France, from which many European laboratories collect experimental animals. We have found that white pigmentation of the spermiduct, when coupled with white pigmentation of the spermiduct papillae/absence of spermiduct papillae, indicates a Type B individual 98% of the time, but coupled with orange pigmentation of the spermiduct papillae could indicate an individual of either type (23% Type A, 77% Type B). Orange pigmentation of the spermiduct, regardless of the presence or pigmentation of the spermiduct papillae, indicates a Type B individual 98% of the time. White pigmentation of spermiduct papillae/absence of the papillae, regardless of the color of the spermiduct, indicates a Type B individual 89% of the time. In contrast, orange pigmentation of spermiduct papillae, when coupled with white pigmentation of the spermiduct, could indicate an individual of either type (23% Type A, 77% Type B), but when coupled with orange pigmentation of the spermiduct, indicates a Type B individual 95% of the time (however, these OO individuals are extremely rare: only 21 total). To summarize: the vast majority of Type A individuals sampled have white pigmentation of the spermiduct coupled with orange pigmented papillae (WO), whereas Type B individuals fall into all morphological classes. From a morphological perspective, the vast majority of WW, OW and OO individuals is Type B, whereas 23% of WO individuals is Type A and 77% are Type B.

Previous morphological analyses found that Type A individuals always had

orange-pigmented spermiduct papillae except those sampled in the southern UK (Plymouth) (Caputi et al. 2007). The results presented here support this observation: the vast majority of Type A individuals had orange-pigmented papillae; those that did not were from populations on both sides of the English Channel. The same published analysis found that Type A individuals always had an uncolored spermiduct; we found this to be true (with only two exceptions). Our results also agree with the observation (Caputi et al. 2007) that the spermiducts of Type B individuals can be orange or uncolored, but don't agree with the finding that the spermiduct papillae of this type are never orange-colored.

Reproductive isolation vs. divergence in marine broadcast spawners

Our data provide evidence for introgression between two deeply divergent forms within *C. intestinalis,* a broadcast spawning marine invertebrate. As in many other externally fertilizing marine invertebrates (Swanson & Vacquier 1997; Hellberg & Vacquier 1999), reproductive isolation between *Ciona* species can be attributed to species specific gamete interactions (Byrd & Lambert 2000). But hybridization and introgression can occur in spite of these barriers (Gardner 1997; Levitan 2002; Harper & Hart 2005), providing support for the idea that broadcast spawning may lead to frequent hybridization and introgression (Gardner 1997). Is *C. intestinalis* a special case, or do other marine broadcast spawning species pairs show evidence for introgression despite deep divergences?

To investigate the relationship between divergence and introgression in marine broadcast spawners, we compiled mtCOI divergence data for 39 pairs of taxa from the literature for which the presence/absence of introgression has been assessed or was considered likely/unlikely. All sequence data were taken from Genbank and all

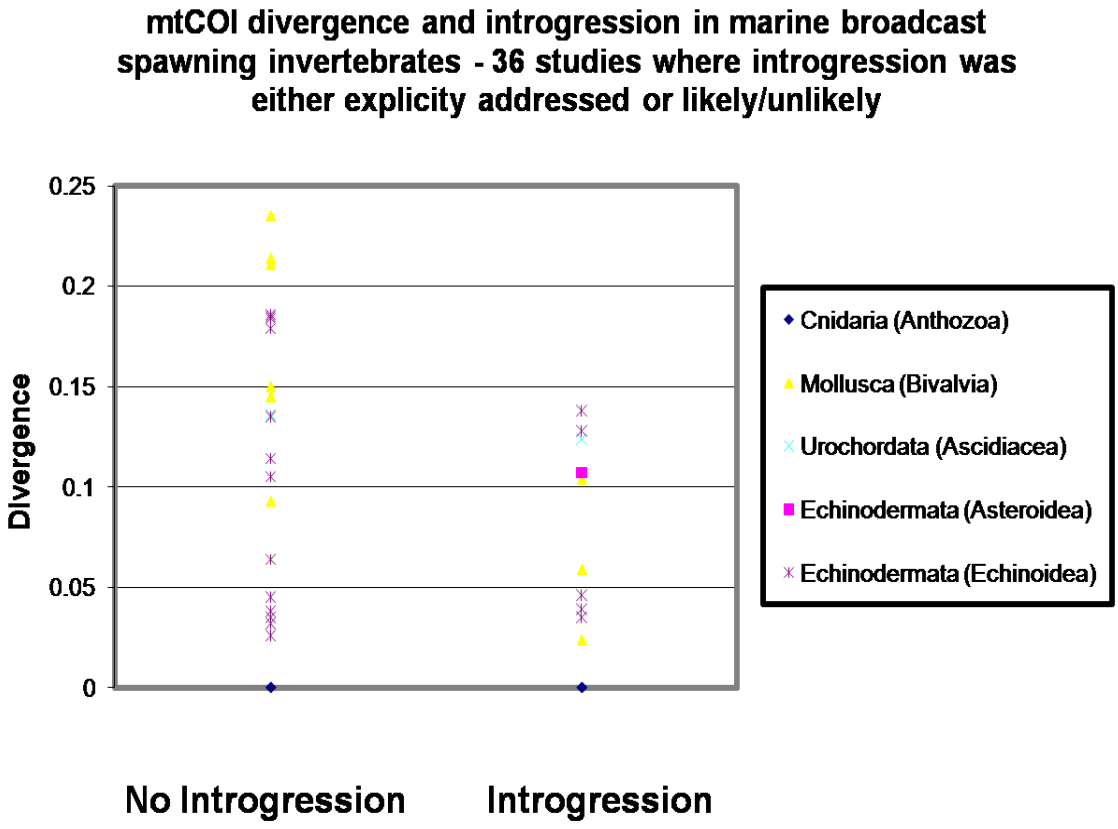

Figure 3.3 Uncorrected mitochondrial COI p-distance vs. introgression in 36 pairs of marine broadcast spawning invertebrates.

Table 3.3 All p-distances and introgression information used in Figure 3.2

Phylum (Class)	Species 1	Species 2	Uncorrected p-distance	Introgression
Cnidaria (Anthozoa)	*Acropora cervicornis*	*Acropora palmata*	0	Yes
	Montastrea annularis	*Montastrea faveolata*	0	Unlikely
	Montastrea annularis	*Montastrea franksi*	0	Unlikely
	Montastrea faveolata	*Montastrea franksi*	0	No evidence
	Platygyra daedalea	*Platygyra lamellina*	0	Likely
	Platygyra daedalea	*Platygyra pini*	0	Likely
	Platygyra lamellina	*Platygyra pini*	0	Likely
Echinodermata (Asteroidea)	*Asterias forbesi*	*Asterias rubens*	0.107	Yes
Echinodermata (Echinodea)	*Diadema savignyi*	*Diadema setosum*	0.138	Yes
	Diadema savignyi	*Diadema paucispinum*	0.035	Yes
	Diadema setosum	*Diadema paucispinum*	0.128	Yes
	Echinometra sp. A	*E.* sp. B (*mathaei)*	0.032	Unlikely
	Echinometra sp. A	*Echinometra* sp. D (*oblonga*)	0.035	No evidence
	Echinometra. sp. B (*mathaei)*	*Echinometra* sp. D (*oblonga*)	0.026	No evidence
	Echinometra. sp. B (*mathaei)*	*Echinometra* sp. C	0.038	No evidence
	Echinometra lucunter	*Echinometra viridis*	0.045	No evidence
	Heliocidaris erythrogramma	*Heliocidaris tuberculata*	0.135	No evidence
	Meridiastra calcar	*Parvulastra exigua*	0.186	Unlikely
	Meridiastra gunnii	*Parvulastra exigua*	0.179	Unlikely
	Pseudechinus albocinctus	*Pseudechinus huttoni*	0.046	Likely
	Pseudechinus albocinctus	*Pseudechinus novaezealandiae*	0.184	Unlikely
	Pseudechinus huttoni	*Pseudechinus novaezealandiae*	0.185	Unlikely
	Stronglyocentrotus droebachiensis	*Stronglyocentrotus pallidus*	0.039	Yes
	Stronglyocentrotus droebachiensis	*Strongylocentrotus franciscanus*	0.105	Unlikely
	Stronglyocentrotus droebachiensis	*Stronglyocentrotus droebachien*	0.064	Unlikely
	Strongylocentrotus franciscanus	*Stronglyocentrotus droebachien*	0.114	Unlikely
Mollusca (Bivalvia)	*Crassostrea angulata*	*Crassostrea virginica*	0.211	Unlikely
	Crassostrea ariakensis	*Crassostrea virginica*	0.235	Unlikely
	Crassostrea gigas	*Crassostrea angulata*	0.024	Yes
	Crassostrea gigas	*Crassostrea ariakensis*	0.145	No evidence
	Crassostrea gigas	*Crassostrea sikamea*	0.093	No evidence
	Crassostrea gigas	*Crassostrea virginica*	0.214	Unlikely
	Crassostrea sikamea	*Crassostrea ariakensis*	0.15	No evidence
	Haliotis laevigata	*Haliotis rubra*	0.104	Yes
	Macoma balthica (Baltic Sea)	*Macoma balthica* (North Sea)	0.059	Yes
Chordata (Ascidiacea)	Type A *Ciona intestinalis*	Type B *Ciona intestinalis*	0.124	Yes
	Type A *Ciona intestinalis*	*Ciona savignyi*	0.136	No evidence

distances were uncorrected p-distance as calculated in MEGA 4.0 (Tamura et al. 2007). Among the introgressing species pairs, Type A/Type B *C. intestinalis* has one of the highest divergence levels (Figure 3.3), but there are sea urchin species pairs that introgress despite having higher divergence levels than Type A/Type B *C. intestinalis* (Table 3.3). There are no introgressing species pairs beyond the 13.8% divergence level, even though divergences continue up to 23.5%. This trend suggests that Type A/Type B *C. intestinalis* may be near the upper limit of the range of divergence values where introgression is still possible. Several studies have suggested that tunicates such as *C. intestinalis* have a faster rate of molecular evolution than other organisms (Winchell et al. 2002; Yokobori et al. 2005; Delsuc et al. 2006). Therefore, the time since divergence between Type A and Type B may be shorter for a given mtCOI divergence than for other marine broadcast spawners. However, the lack of an adequate fossil record for ascidians prevents us from addressing this possibility. But the fact that Type A and B and other broadcast spawning marine invertebrates show evidence of introgression despite substantial divergence is consistent with the idea that these organisms may have fewer or weaker interspecific prezygotic barriers.

Although mtCOI is a useful proxy for time since divergence between species pairs, this gene does not have any direct relationship to the reproductive barriers that contribute to speciation. In fact, a lack of correlation between mtCOI divergence and prezygotic reproductive isolation has been noted in sea urchins (Zigler et al. 2005) and between divergence at other non-reproductive markers and prezygotic reproductive isolation in sea urchins and oysters (Lessios 1984; Lessios & Cunningham 1990; Gaffney & Allen 1993). In contrast, there is a negative correlation between divergence in the nonsynonymous sites of the *bindin* gene, which encodes a protein involved in the binding of the sperm to the egg, and gamete compatibility (an important measure of postmating prezygotic isolation) in sea urchins (Zigler et al.

2005). Species in the sea urchin genus *Echinometra* (a large component of the Zigler et al. 2005 study), show gamete incompatibility, thought to be a result of positive selection on bindin, despite low levels of genetic divergence (Geyer & Palumbi 2003). The positive selection acting on bindin, leading to postmating prezygotic isolation despite very little genetic divergence, may be a byproduct of evolution in sympatry (Geyer & Palumbi 2003) as this pattern is not seen in allopatric sea urchin taxa (Metz 1998; Zigler & Lessios 2003; Zigler & Lessios 2004). Further characterization of interspecific divergence in bindin and other reproductive proteins will lead to a more complete understanding of the evolution of reproductive isolation in these organisms.

In conclusion, examination of Type A and B populations in the zone of potential sympatry found these types co-existing in six locations and evidence of gene flow in three locations. Not only are Type A and B cryptic species, they show evidence of recent introgressive hybridization despite substantial divergence. Examining the relationship of mtCOI divergence to introgressive hybridization in other marine broadcast spawning species pairs reveals that introgression at divergence levels similar to those found in Type A and B *C. intestinalis* does occur, prompting questions about the strength of postmating prezygotic reproductive barriers in this group.

Acknowledgements

We would like to thank members of the Harrison Lab for giving molecular advice and commenting on an earlier version of the manuscript, John Bishop for providing logistical assistance, Derrick Chong for help with data collection, and Gabrielle Miller-Messner for collecting the samples from Japan. This work was funded by an AMNH Lerner-Gray Grant for Marine Research, a Cornell University Research Travel Grant and an NSF Doctoral Dissertation Improvement Grant to M.L.N. and R.G.H.

REFERENCES

Aasland, R., Gibson, T.J., and A.F. Stewart. 1995. The PHD finger: Implications for chromatin-mediated transcriptional regulation. Trends in Biochemical Sciences 20: 56-59.

Arnold, M. L. 1997. Natural hybridization and evolution. Oxford Univ. Press, New York.

Bierne, N., P. Borsa, D. Daguin, D. Jollivet, F. Viard, P. Bonhomme, and P. David. 2003. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. Mol. Ecol. 12:447-461.

Byrd, J., and C. C. Lambert. 2000. Mechanism of the block to hybridization and selfing between the sympatric ascidians *Ciona intestinalis* and *Ciona savignyi*. Mol. Repro. Dev. 55:109-116.

Caputi, L., N. Andreakis, F. Mastrototaro, P. Cirino, M. Vassillo, and P. Sordino. 2007. Cryptic speciation in a model invertebrate chordate. Proc. Natl. Acad. Sci. USA 104:9364-9369.

Cirino, P., A. Toscano, D. Caramiello, A. Macina, V. Miraglia, and A. Monte. 2002. Laboratory culture of the ascidian *Ciona intestinalis* (L.): A model system for molecular developmental biology research. Mar. Mod. Elec. Rec. Marine Biological Laboratory, Woods Hole, MA.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila.* Evolution 43:362-381.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51:295-303.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Daguin, C., F. Bonhomme, and P. Borsa. 2001. The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis,* as revealed by intron length polymorphism at locus *mac-1.* Heredity 86:342-354.

Delsuc, F., H. Brinkmann, D. Chourrout, and H. Phillippe. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965-968.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22:1-20.

Dybern, B. I. 1965. The life cycle of *Ciona intestinalis* (L.) f. tipica in relation to the environmental temperature. Oikos 16:109-131.

Edmands, S. 2002. Does parental divergence predict reproductive compatibility? Trends Ecol. Evol. 17:520-527.

Freyhof, J., D. Lieckfeldt, C. Pitra, and A. Ludwig. 2005. Molecules and morphology: Evidence for introgression of mitochondrial DNA in Dalmatian cyprinids. Mol. Phylogenet. Evol. 37:347-354.

Gaffney, P. M., and S. K. Allen, Jr. 1993. Hybridization among *Crassostrea* species: a review. Aquaculture 116:1-13.

Gardner, J. P. A. 1997. Hybridization in the sea. Adv. Mar. Biol. 31:1-78.

Geyer, L., and S. Palumbi. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution 57:1049-1060.

Grizel, H., and M. Héral. 1991. Introduction into France of the Japanese oyster (*Crassostrea gigas*). Journal du Conseil International pour l'Exploration de la Mer 47:399–403.

Gruet, Y., M. Héral, and J. M. Robert. 1976. Premières observations sur l'introduction de la faune associée sur la côte atlantique française. Cah. Biol. Mar. 17:173-

184.

Hare, M. P., and J. C. Avise. 1998. Population structure in the American oyster as inferred by nuclear gene genealogies. Mol. Biol. Evol. 15:119-128.

Harper, F. M., and M. W. Hart. 2005. Gamete compatibility and sperm competition affect paternity and hybridization between sympatric *Asterias* Sea Stars. Biol. Bull. 209:113-126.

Harrison, R. G., and S. B. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493-505.

Hayward, P. J., and J.S. Ryland. 1995. Handbook of the Marine Fauna of North-West Europe. Oxford University Press, Oxford.

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.

Koblmuller, S., N. Duftner, K. M. Sefc, M. Aibara, M. Stipacek, M. Blanc, B. Egger, and C. Sturmbauer. 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika - the result of repeated introgressive hybridization. BMC Evol. Biol. 7:7-19.

Kott, P. 1952. The ascidians of Australia. Aus. J. Mar. Fresh. Res. 3:206-233.

Kott, P. 1990. The Australian Ascidiacea Pt 2, Aplousobranchia (1). Memoirs of the Queensland Museum 29:1-266.

Kott, P. 1997. The Tunicates. In: Shepherd, S.A., Thomas, I.M., (eds). Marine Invertebrates of Southern Australia, Part 1. Government Printer, Adelaide, South Australia, p. 1092-1255.

Kronforst, M. R., L. G. Young, L. M. Blume, and L. Gilbert. 2006. Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. Evolution 60:1254-1268.

Lamb, T., and J. C. Avise. 1986. Directional introgression of mitochondrial DNA in a hybrid population of tree frogs: The influence of mating behavior. Proc. Natl. Acad. Sci. USA 83:2526-2530.

Lambert, C. C., and G. Lambert. 1998. Nonindigenous ascidians in southern California harbors and marinas. Mar. Biol. 130:675-688.

Lambert, C. C., and G. Lambert. 2003. Persistence and differential distribution of nonindigenous ascidians in the harbors of the Southern California Bight. 259: 145-161.

Lessios, H. A. 1984. Possible prezygotic reproductive isolation in sea urchins separated by the Isthmus of Panama. Evolution 38:1144-1148.

Lessios, H. A., and C. W. Cunningham. 1990. Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the Isthmus of Panama. Evolution 44:933-941.

Levitan, D. R. 2002. The relationship between conspecific fertilization success and reproductive isolation among three congeneric sea urchins. Evolution 56:1599-1609.

Levitan, D. R., H. Fukami, J. Jara, D. Kline, T. M. McGovern, K. E. McGhee, C. A. Swanson, and N. Knowlton. 2004. Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the *Montastraea annularis* species complex. Evolution 58:308-323.

Maddison, W. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523-536.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol 20:229-237.

Mallet, J., W. O. McMillan, and C.D. Jiggins. 1998. Mimicry and warning color at the boundary between races and species. Pp. 390-403 *in* D. J. Howard, and Berlocher, S.H., ed. Endless Forms: Species and Speciation. Oxford University

Press, Oxford.

Masta, S. E., B. Sullivan, T. Lamb, and E. J. Routman. 2002. Phylogeography, species boundaries, and hybridization among toads of the *Bufo americanus* group. Mol. Phylogenet. Evol. 24:302-314.

McDonald, J. 2004. The invasive pest species *Ciona intestinalis* (Linnaeus, 1767) reported in a harbour in southern Western Australia. Mar. Pollut. Bull. 49: 854-874.

McGuire, J. A., C. W. Linkem, M. S. Koo, D. W. Hutchison, A. K. Lappin, D. I. Orange, J. Lemos-Espinal, B. R. Riddle, and J. R. Jaeger. 2008. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evolution 61:2879-2897.

Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma*). Evolution 57:317-327.

Metz, E. C. 1998. Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. Mol. Biol. Evol. 15:185-195.

Monniot, C., and F. Monniot. 1994. Additions to the inventory of eastern tropical Atlantic ascidians; arrival of cosmopolitan species. Bull. Mar. Sci. 54:71-93.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. Evolution 58:1195-1208.

Naisbit, R. E., C. D. Jiggins, J. Mallet. 2003. Mimicry: developmental genes that contribute to speciation. Evol. Dev. 5:269-280.

Nydam, M. L., and R. G. Harrison. 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). Mar. Biol. 151:1839-1847.

Nydam, M.L., and R.G. Harrison. 2010. Polymorphism and divergence within the

ascidian genus *Ciona.* Mol. Phyl. Evol. doi:10.1016/j.ympev.2010.03.042

Prager, E. M., and A. C. Wilson. 1975. Slow evolutionary loss of the potential for interspecific hybridization in birds: A manifestation of slow regulatory evolution. Proc. Natl. Acad. Sci. USA 72:200-204.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083-2089.

Ritter, W. E., and R. A. Forsyth. 1917. Ascidians of the littoral zone of southern California. Univ. Calif. Publs. Zool. 16:439-512.

Rodriguez, S., T. R. Gaunt, and I. N. M. Day. 2009. Hardy-Weinberg Equilibrium testing of biological ascertainment for Mendelian randomization studies. Am. J. Epidem. 169:505-514.

Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811-1820.

Sentz-Braconnot, E. 1966. Donnees ecologiques sur la fixation d'Invertebres sur des plaques immergees dans la rade de Villefranche-sur-mer. Int. Revue ges Hydrobiol. 51:461-484.

Suzuki, M., T. Nishikawa, and A. Bird. 2005. Genomic approaches reveal unexpected difference within *Ciona intestinalis.* J. Mol. Evol. 61:627-635.

Svane, I., and J. N. Havenhand. 1993. Spawning and dispersal in *Ciona intestinalis* (L.). Pubbl Staz zool Napoli (I: Mar Ecol) 14:53-66.

Swanson, W. J., and V. D. Vacquier. 1997. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc. Natl. Acad. Sci. USA 94:6724-6729.

Tzouanacou, E., Tweedie, S., and V. Wilson. 2003. Identification of *Jade1*, a Gene Encoding a PHD Zinc Finger Protein, in a Gene Trap Mutagenesis Screen for Genes Involved in Anteroposterior Axis Development. Molecular and Cellular

Biology 23:8553-8562.

Van Name, W. 1945. The north and south American ascidians. B. Am. Mus. Nat. Hist. 84:1-476.

Wasson, K., C. J. Zabinc, L. Bedinger, M. C. Diaz, and J. S. Pearse. 2001. Biological invasions of estuaries without international shipping: the importance of intraregional transport. Biol. Conserv. 102:143-153.

Wendel, J. F., and J. J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Pp. 265–96 *in* D. E. Soltis, Soltis, P.S., Doyle, J.J., ed. Molecular systematics of plants II: DNA sequencing. Kluwer, Dordrecht, the Netherlands.

Wilson, A. C., L. R. Maxson, and V. M. Sarich. 1974. Two types of molecular evolution. Evidence from studies of interspecific hybridization. Proc. Natl. Acad. Sci. USA 71:2843-2847.

Winchell, C. J., J. Sullivan, C. B. Cameron, B. J. Swalla, and J. Mallatt. 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol. 19:762-776.

Yamaguchi, M. 1975. Growth and reproductive cycles of the marine fouling ascidians *Ciona intestinalis*, *Styela plicata*, *Botrylloides violaceus*, and *Leptoclinum mitsukurii* at Aburztsubo-Moroiso Inlet (Central Japan). Mar. Biol. 29:253-259.

Yokobori, S., Oshima, T., Wada, H. 2005. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol. Phylogenet. Evol. 34:273-283.

Zhou, M.I., Foy, R.L., Chitalia, V.C., Zhao, J., Panchenko, M.V., Wang, H., and H.T. Cohen. 2005. Jade-1, a candidate renal tumor suppressor that promotes apoptosis. Proc. Natl. Acad. Sci. USA 102: 11035-11040.

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. McCartney, D. R. Levitan, and H. A. Lessios. 2005. Sea urchin bindin divergence predicts gamete compatibility. Evolution 59:2399-2404.

CHAPTER 4

REPRODUCTIVE PROTEIN EVOLUTION IN TYPE A AND B *C. INTESTINALIS*

Abstract

Rapid evolution, and in many cases positive selection, has become a well-known feature of gamete recognition proteins (GRPs) in marine broadcast spawners. But in most cases the evolutionary process or processes responsible for this pattern remain elusive. One of the mechanisms often cited as a potential driver of rapid evolution of GRPs in marine broadcast spawners is reinforcement, a strengthening of prezygotic isolation caused by postzygotic isolation when two species come into secondary contact after diverging in allopatry. The clearest way to test whether reinforcement is occurring is to test whether prezygotic isolation is stronger in sympatry than allopatry.

Candidate GRPs from two lineages of *C. intestinalis* (Type A and B) are evolving more rapidly than control proteins, consistent with patterns seen in insects and mammals. d_n/d_s ratios are higher in sympatric than allopatric populations in two of the three candidate GRPs tested, although this result may be driven by d_s rather than d_n. None of the polymorphism statistics showed significant differences between sympatric and allopatric populations. While little evidence for reinforcement was found in these three candidate GRPs, tests such as those performed here may provide important insights into the process of speciation in marine broadcast spawners.

Introduction

While a substantial number of studies have found evidence for positive selection on gamete-recognition proteins (GRPs) in marine broadcast spawners, the underlying evolutionary forces responsible for these patterns remain elusive. Potential processes include assortative mating, microbial resistance, reinforcement, sexual selection and sexual conflict (reviewed in Swanson & Vacquier 2002), and distinguishing between these possibilities presents a considerable challenge. In marine invertebrates, gamete-recognition proteins are often major players in reproductive isolation between species or incipient species (*e.g.* Geyer & Palumbi 2003; Springer & Crespi 2007). Therefore, an investigation into the processes that lead to rapid evolution of these proteins will provide critical insights into speciation in the marine environment.

One of the mechanisms often cited as a potential driver of rapid evolution of GRPs in marine broadcast spawners is reinforcement; and some of the most prominent cases of reinforcement occur in marine broadcast spawners (Coyne & Orr 2004). Reinforcement, as originally envisioned by Dobzhansky, is defined as a strengthening of prezygotic isolation caused by postzygotic isolation when two species come into secondary contact after diverging in allopatry (Dobzhansky 1937). Offspring of heterospecific matings are somehow unfit, so natural selection favors those individuals that mate with conspecifics. Butlin made an important distinction between two scenarios: one in which the hybrids between the two species are completely sterile or inviable (and therefore no gene flow occurs between the two species) and the other in which the hybrids are not completely sterile or inviable (and gene flow can occur) (Butlin 1987). In the first scenario, the two species are already completely reproductively isolated; he suggested this scenario should therefore be termed

"reproductive character displacement", not reinforcement (Butlin 1987).

The clearest way to test whether reinforcement is occurring is to test whether prezygotic isolation is stronger in sympatry than allopatry (Coyne & Orr 2003). Reinforcement does not occur in allopatry, as hybrids cannot be produced. While this test has been applied to many different taxa (*e.g.* Coyne & Orr 1989,1997; Matute 2010; Saetre et al. 1997; Wullschleger et al. 2002), reinforcement has never been tested in this way in marine broadcast spawners. Evidence for reinforcement in these taxa is indirect; explicit comparisons of prezygotic isolation between two species in sympatry vs. allopatry have not been performed.

New information has allowed us to use the ascidian *Ciona intestinalis,* a longtime developmental model, to ask questions about processes that lead to the rapid evolution of GRPs. First, *C. intestinalis* actually comprises two distinct and divergent lineages, now termed Type A and Type B (Nydam & Harrison 2007, 2010b; Caputi et al. 2007). Type A is thought to be native to the Northwestern Pacific Ocean and has invaded the Eastern Pacific Ocean, the Mediterranean Sea, the Atlantic coast of South Africa, and the Black Sea (Van Name 1945; Kott 1952). Type B, thought to be native to the Northern Atlantic Ocean (Linne 1767; Monniot & Monniot 1994), has invaded the Western Atlantic Ocean (Nydam & Harrison 2007). Second, the ranges of Type A and B overlap in the English Channel and the Atlantic coast of France; a small amount of introgression occurs between Type A and B individuals from these sympatric populations (Nydam & Harrison 2010a). Although hybrid individuals from allopatric populations are sterile or inviable in the laboratory (Caputi et al. 2007), the evidence for introgression in nature shows these two lineages are not completely reproductively isolated. Nothing is known about pre or postzygotic barriers in sympatry.

Here, we identified a set of candidate GRPs from *C. intestinalis* sperm using proteomic and bioinformatic techniques. We then tested whether these proteins

evolve more rapidly than control proteins. We then choose a subset of these proteins that show evidence of positive selection, and directly test a reinforcement scenario by comparing divergence and polymorphism statistics in sympatric and allopatric populations of Type A and B *C. intestinalis* to ask whether signatures of positive selection are stronger in sympatric than allopatric populations. We caution that this system does not provide an ideal test of reinforcement, as we do not have evidence that these candidate GRPs are involved in prezygotic isolation, and Type A and B may have come into secondary contact very recently. However, we believe that this study could catalyze similar tests in other systems, to increase our knowledge of the evolutionary forces behind the widespread pattern of positive selection in the GRPs of marine broadcast spawners.

Materials and Methods

Identification of candidate GRPs from sperm: proteomics experiment

Sperm was collected from Type A individuals living in Santa Barbara, CA, filtered through a 70 μm nylon cell strainer (BD Biosciences) by centrifuging for 3 minutes at 3,000 rpm, and stored dry at -80C. Sperm samples from several different individuals were later pooled and diluted 5-fold in phosphate buffer*;* the concentration of this dilution was determined to be 915 μg/ml. 500 μl of this diluted sperm was shipped to the University of Victoria Genome BC Proteomics Centre for the experiments described below.

9.5 M urea, 50 mM NH_4HCO_3 and 0.2% SDS were added to the sample, which was then sonicated. The proteins then underwent disulphide reduction and sulphydryl alkylation (200 mM DTT and 200 mM iodoacetamide) and were digested overnight at 37C with 20mg trypsin (Promega). Samples were subsequently cleaned with a cation

exchange Cartridge Kit for cICAT (Applied Biosystems).

Strong cation exchange chromatography: 10 mM KH_2PO_4 (pH 2.7), 25% ACN buffer was added to the sample, which was then injected onto a Polysulphoethyl A strong cation exchange chromatography (SCX) column (PolyLC, Columbia, MD). The flow rate was set to 0.5 ml min^{-1}. After equilibration, a 0–35% gradient of 10 mM KH_2PO_4, 25% ACN, 0.5 M KCl was added for 30 min. Each SCX fraction was reduced and transferred to autosampler vials (Dionex/LC Packings, Amsterdam).

One-dimensional reversed-phase chromatography with online mass spectrometry: Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) analyses of the SCX fractions were performed in a Hybrid Quadruple-TOF LC–MS/MS mass spectrometer (QStar Pulsar I, MDS Sciex), with a nanoelectrospray ionization source (Proxeon, Odense, Denmark) fitted with a 10 mm fused-silica emitter tip (New Objective, Woburn, MA). A C18AQ Nano LC and a Zorbax C18 guard column (Agilent Technologies) performed the chromatographic separation. The ANALYST QS v. 1.1 software service pack (ABI MDS SCIEX, Concord, Canada) gathered the data.

Mass spectrometry data analyses: the information dependent acquisition file was viewed using ANALYST v. 1.1 software, and the peak lists were assembled with a built-in MASCOT script (1.6b16 ABI—Matrix Science Limited). Spectra with less than 10 peaks were discarded. MASCOT v. 2.0 (Matrix Science Limited) was used to analyzed the data. Spectrometry data were searched against a database of amino acid sequences from the CIPRO (*Ciona intestinalis* Protein) database (http://cipro.ibio.jp/new/).

Comparison of d_n/d_s values between candidate GRPs and control proteins

Of proteins identified by proteomic analysis, we selected 39 proteins (30

candidate GRPs and 9 control proteins) for further analysis using proteomic and bioinformatic approaches (see Results section). Genomic sequences for the genes encoding the 39 proteins were located by performing tblastn searches to the *Ciona intestinalis* Ensembl genome server (http://uswest.ensembl.org/Ciona_intestinalis/Info/Index). Primers were developed in coding regions, and the coding regions of all 39 proteins were sequenced from cDNA of two Type A and two Type B individuals, all from allopatric populations. Testis tissue from these four individuals was collected in 2008 and immediately placed in RNAlater (Ambion). The tissue/RNAlater was frozen at -80C within seven days of collection.

Total RNA was extracted from testis tissue with the RNAdvance Kit (Agencourt) and was used to synthesize single-stranded cDNA using SuperScript III reverse transcriptase (Invitrogen) and an oligo (dT) primer. A 5-fold dilution of the single-stranded cDNA was then PCR-amplified with TRsa and TS-PCR primers. The resulting PCR product was diluted 50-fold and used as the template for amplification of the coding regions for the 30 candidate GRPs proteins and 9 control proteins. The amplified coding regions were incubated with 0.25µl each of Exonuclease I and Shrimp Antarctic Phosphatase at 37°C for 30 min, followed by 90°C for 10 min. The products were purified using CleanSeq beads (Agencourt), and the purified products were sequenced with a Big Dye Terminator Cycle sequencing kit and an Automated 3730 DNA Analyzer (Applied Biosystems). Sequences were edited, trimmed and aligned with Aligner (CodonCode Corporation, Dedham, MA). Primers and cycling conditions are available from the authors.

Once the sequences were obtained, the codeml program in PAML 4.4 (Yang 2007) was used to obtain pairwise d_n/d_s values for each Type A vs. Type B combination (Type A Individual #1 vs. Type B Individual #1, Type A Individual #1

vs. Type B Individual #2, Type A Individual #2 vs. Type B Individual #1, Type A Individual #2 vs. Type B Individual #2). The average d_n/d_s value for all four combinations was calculated for each gene, and the d_n/d_s values of the putative GRPs and control proteins were found to be distributed non-normally using the Shapiro-Wilk test in R (version 2.10.1). The Shapiro-Wilk test is the most robust test of non-normality for small to medium sample sizes (Conover 1999, Shapiro & Wilk 1965). The d_n/d_s values of the candidate GRPs and control proteins were therefore compared using a one-tailed Mann-Whitney U test in R (version 2.10.1).

Sympatric vs. allopatric divergence analyses

Samples were collected in 2005-2009 from allopatric and sympatric populations of Type A and Type B. The allopatric population of Type A was located in Half Moon Bay, CA while sympatric populations of Type A were located in Perros-Guirec, France and Concarneau, France. Allopatric populations of Type B were located in Woods Hole, MA and Gosport, England.

Ovaries were dissected from freshly-collected individuals, cut into several pieces, immediately preserved in DMSO (dimethyl sulfoxide), and ultimately (within 10 d) stored at -80°C until needed. Total DNA was extracted from the ovaries using the Qiagen DNeasy® Tissue Kit (Qiagen Corporation, Santa Clarita, CA).

At least 10 individuals from each of 4 populations (allopatric Type A and B, sympatric Type A and B) were sequenced for the genes encoding three candidate GRPs: CIPRO37.40.1, CIPRO60.5.1, CIPRO100.7.1. The criteria for selecting these three candidate GRPs for the sympatric/allopatric comparison, from the 30 candidate GRPs used in the comparison of d_n/d_s values between GRPs and control proteins, is available in the Results section.

The same individuals were also sequenced for two control proteins:

CIPRO53.35.1 and mitochondrial cytochrome oxidase I (mtCOI). A signature of enhanced selection in sympatry vs. allopatry could be due to selective processes or demographic processes (e.g. recent population growth). But demographic processes would affect all genes, not just candidate GRPs. So proteins not involved in the fertilization process (control proteins) were also subjected to divergence analyses. CIPRO53.35.1 was selected as a control protein because it contains a domain similar to a ribosomal L32 protein domain and is expressed in many different tissues; whereas mtCOI is an enzyme in the electron transport chain of the mitochondria. For all nuclear genes the PCR product for each locus was cloned using the pGEM®-T kit and up to eight clones were PCR-amplified and sequenced as described above. For all loci, both alleles were identified for each individual.

Each sympatric Type A allele was randomly assigned to a sympatric Type B allele and each allopatric Type A allele to an allopatric Type B allele. The codeml program in PAML 4.4 (Yang 2007) was then used to obtain pairwise d_n/d_s values for these sympatric Type A vs. sympatric Type B and allopatric Type A vs. allopatric Type B pairs. The same process was completed using a different set of randomly assigned pairs, to ensure that results were independent of the allele pairings. The two randomizations will be referred to as Randomization 1 and 2. For each gene and each randomization, a Shapiro-Wilk test was used to test the normality of the pairwise d_n/d_s values. Where the Shapiro-Wilk test found evidence for non-normality, non-parametric one-tailed Mann-Whitney U tests were performed in R (version 2.10.1). Where the Shapiro-Wilk test did not find evidence for non-normality for pairwise d_n/d_s values, one-tailed t-tests were performed in R. But because normality can be difficult to assess in small sample sizes (Dytham 2003), we also conducted non-parametric one-tailed Mann-Whitney U tests in R for these randomizations where the Shapiro-Wilk test did not find evidence for non-normality.

Sympatric vs. allopatric polymorphism analyses

Both alleles of at least 10 individuals from each of four populations (allopatric Type A and B, sympatric Type A and B) were sequenced for three candidate gamete recognition genes and two control genes as described in the "Sympatric vs. allopatric divergence analyses" section.

For each population and each gene, the summary statistics θ and π were calculated in DnaSP 5.10.1 (Rozas et al. 2003). We also employed the following tests: McDonald-Kreitman (McDonald & Kreitman 1991), Tajima's D (Tajima 1989), Fu and Li's D* and F*, and Fay and Wu's H (Fay & Wu 2000) in DnaSP. Statistical significance of D, D*, F* and H were determined using 1,000 coalescent simulations in DnaSP. Estimates of per gene recombination for each population were made in DnaSP and were then imported into the simulations. 95% confidence intervals for D, D*, F* and H statistics were also recorded; sympatric and allopatric populations were determined to be significantly different for each statistics if the confidence intervals were non-overlapping.

Results

Identification of candidate GRPs from sperm: proteomics

161 proteins were found in the sperm; each of these proteins was subsequently identified using one or more of the following sections of CIPRO: the descriptive summary available for many proteins, the Pfam Domain search, and the BlastP search, or the GO (Gene Ontology) program. 144 of these proteins were determined unlikely to be GRPs; they are likely involved in the movement or metabolism of the sperm

(e.g. actin, dynein, myosin, tektin, α-tubulin, ATP-synthase, creatine kinase, enolase, malate dehydrogenase). The identities of these 144 proteins are available from the authors. Of the remaining 17 proteins, seven were likely GRPs, and 10 could not be identified as similar to any known proteins. Of these 17 proteins, four could not be analyzed because the corresponding gene could not be amplified or successfully sequenced from Type B individuals, and one could not be analyzed because the corresponding gene did not have significant tblastn hits to the genome (and therefore no primers could be designed). We selected the remaining 12 proteins for further analysis.

Identification of candidate GRPs proteins from sperm: bioinformatics

We also used a bioinformatic approach to identify potential GRPs. First, we accessed the functional classifications of Type A testis ESTs sequenced by Inaba et al. (2002) and selected 25 ESTs that could code for GRPs. We then located the genes corresponding to these ESTs with the KOG (EuKaryotic Orthologous Groups) tool provided on JGI's *C. intestinalis* genome browser (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html). These genes were then searched against the CIPRO database using blastx to identify resulting protein matches. 19 of these proteins were determined to be GRP candidates, but 10 failed to amplify and/or sequence in Type B individuals; the remaining nine were selected for further study.

Second, we located every protein in the CIPRO database that was identified as being expressed only in the testis tissue. We chose a subset of 10 of these proteins that were likely GRPs, and further investigated the nine proteins that amplified Type B cDNA. In total, we selected 12 candidate GRPs identified proteomically and 18 identified bioinformatically. We also selected 9 control proteins (not involved in the fertilization process) from the proteomics experiments to compare with the putative

GRPs.

Comparison of d_n/d_s values between candidate GRPs and control proteins

The d_n/d_s values for candidate GRPs are significantly higher than the d_n/d_s values for control proteins (Figure 4.1, $p = 0.004891$ using a one-tailed Mann-Whitney U Test).

However, PAML analyses assume that d_s values are constant across a sequence. If some sites across the sequence have unusually low d_s values, a d_n/d_s value greater than 0.5 could be inferred in the absence of positive selection (Pond and Muse 2005). Similarly, significantly higher d_n/d_s values for candidate gametic recognition genes than control genes could be the result of either higher d_n or lower d_s values in the candidate recognition genes, and only higher d_n values provide evidence of positive selection.

To address this issue, we performed two-tailed Mann-Whitney U test in R (the Shapiro-Wilk test found evidence for non-normality), comparing d_n values in candidate gamete recognition vs. control genes, and d_s values in candidate gamete recognition vs. control genes. d_n values were significantly different between candidate gamete recognition and control genes ($p = 0.002$), whereas d_s values were not ($p = 0.269$). These tests are consistent with the assumption that d_n, rather than d_s, is driving this pattern.

Figure 4.1 also shows that two proteins have a d_n/d_s ratio greater than 0.5, which is the value above which we consider positive selection to likely be occurring when conservative pairwise d_n/d_s comparisons are used. A study by Swanson et al. across many different taxa showed that if a pairwise comparison yielded a dn/ds of greater than 0.5, the value was often greater than 1 when more sensitive site-specific tests were used (Swanson et al. 2004). CIPRO37.40.1 was identified from the CIPRO

database as having testis-only expression, and contains domains similar to ricin-type beta-trefoil lectin domains ($d_n/d_s = 0.618$, $d_n = 0.054$, $d_s = 0.087$). CIPRO100.7.1 was

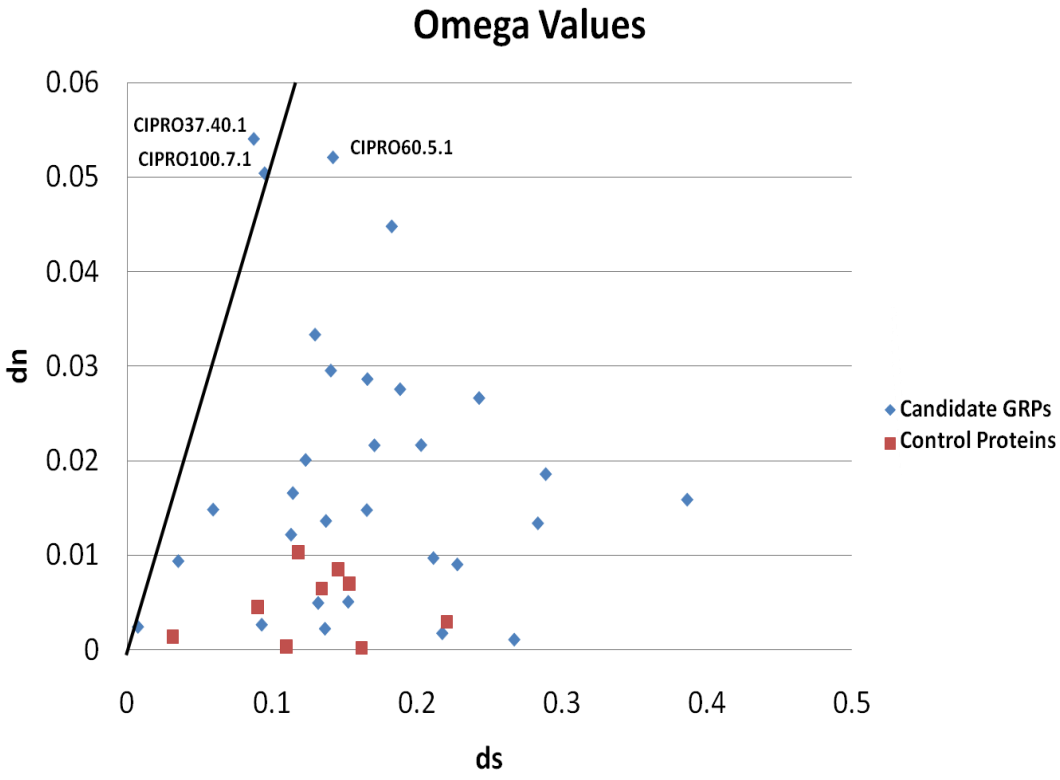

Figure 4.1 Pairwise d_n/d_s values (Type A vs. Type B *C. intestinalis*) for 30 candidate GRPs and 9 control proteins. The line is $d_n/d_s = 0.5$.

identified proteomically, and its function is unknown. CIPRO100.7.1 is a large protein (1,225 amino acids), and was therefore divided into three sections for sequencing. Sections 2, 3, and all 3 sections analyzed together had d_n/d_s values less than 0.5, so only Section 1 was analyzed in subsequent sympatric vs. allopatric comparisons. Section 1 of CIPRO100.7.1 had a d_n/d_s equal to 0.531, a d_n equal to 0.050, and a d_s equal to 0.095. One protein, CIPRO60.5.1, has a d_n/d_s ratio lower than 0.5, but was identified in the CIPRO database as a metalloproteinase and has a GO biological function of "sperm binding to zona pellucida" ($d_n/d_s = 0.366$, $d_n = 0.052$, $d_s = 0.142$). We chose these three proteins for the subsequent sympatric vs. allopatric comparisons because they showed evidence of positive selection (in the case of CIPRO37.40.1 and CIPRO100.7.1), or because their putative function was so clearly related to gamete recognition (in the case of CIPRO60.5.1).

Sympatric vs. allopatric divergence analyses

The results of the divergence analyses are shown in Table 4.1. For the candidate GRPs CIPRO37.40.1 and CIPRO60.5.1, d_n/d_s values were significantly higher in sympatric than allopatric populations using both parametric and non-parametric tests for both randomizations. For CIPRO37.40.1, Randomization 1: $p = 0.004$ (one-tailed t test) and $p = 0.008$ (one-tailed Mann Whitney U test), Randomization 2: $p = 0.002$ (one-tailed t test) and $p = 0.008$ (one-tailed Mann Whitney U test). For CIPRO60.5.1, Randomization 1: $p = 0.020$ (one-tailed t test) and $p = 0.018$ (one-tailed Mann Whitney U test), Randomization 2: $p = 0.018$ (one-tailed Mann Whitney U test). Sympatric and allopatric d_n/d_s values were not significantly different for the candidate GRP CIPRO100.7.1 for either randomization (Randomization 1: $p = 0.92$, one-tailed Mann Whitney U Test, Randomization 2: $p = 0.806$, one=tailed Mann Whitney U Test). For the control proteins, d_n/d_s values

Table 4.1 Sympatric vs. allopatric comparisons of dn/ds values for Randomization 1 and 2. d_n/d_s values for each test are averages of all the pairwise comparisons made for each test.

Candidate GRPs (Randomization 1)	d_n/d_s Sympatric	d_n/d_s Allopatric	Statistical Result	P-value (t-test, Mann-Whitney U test)
CIPRO37.40.1	0.498	0.405	Sympatric > Allopatric	(0.004, 0.008)
CIPRO60.5.1	0.468	0.441	Sympatric > Allopatric	(0.020,0.018)
CIPRO100.7.1	0.510	0.548	No difference	(NA, 0.920)
Candidate GRPs (Randomization 2)	d_n/d_s Sympatric	d_n/d_s Allopatric	Statistical Result	P-value (t-test, Mann-Whitney U test)
CIPRO37.40.1	0.498	0.413	Sympatric > Allopatric	(0.002, 0.008)
CIPRO60.5.1	0.471	0.438	Sympatric > Allopatric	(NA,0.018)
CIPRO100.7.1	0.520	0.543	No difference	(NA, 0.806)
Control Proteins (Randomization 1)	d_n/d_s Sympatric	d_n/d_s Allopatric	Statistical Result	P-value (t-test, Mann-Whitney U test)
CIPRO53.35.1	0.053	0.096	No difference	(NA, 0.998)
mtCOI	0.003	0.003	No difference	(NA, 0.436)
Control Proteins (Randomization 2)	d_n/d_s Sympatric	d_n/d_s Allopatric	Statistical Result	P-value (t-test, Mann-Whitney U test)
CIPRO53.35.1	0.042	0.098	No difference	(NA, 1)
mtCOI	0.003	0.003	No difference	(NA, 0.676)

were not significantly different between sympatry and allopatry for either randomization as determined by Mann Whitney U tests: (CIPRO53.35.1- Randomization 1: p = 0.998, Randomization 2: p = 1, mtCOI – Randomization 1: p = 0.436, Randomization 2: p = 0.676).

As discussed above, significantly higher d_n/d_s values in sympatry than allopatry could be the result of either higher d_n or lower d_s values in sympatry, and only higher d_n values provide evidence of positive selection.

To determine whether nonsynonymous or synonymous substitutions are responsible for this result, we performed two-tailed t-tests in R (the Shapiro-Wilk test did not find evidence for non-normality), comparing d_n values in sympatry vs. allopatry, and d_s values in sympatry vs. allopatry for two randomizations. For CIPRO37.40.1, d_n is not significantly different between sympatry and allopatry for either randomization (p values = 0.907, 0.369), but d_s is significantly different between sympatry and allopatry for both randomizations (p values = 0.021, 0.002). For CIPRO60.5.1, d_n is not significantly different between sympatry and allopatry for either randomization (p values = 0.636, 0.6), but d_s is significantly different between sympatry and allopatry for one randomization (p values = 0.045, 0.074). It is possible, therefore, that d_s, rather than d_n, is driving the pattern we see at these loci.

Sympatric vs. allopatric polymorphism analyses

The summary statistics are shown in Table 4.2. No consistent differences between sympatric and allopatric Type A or between sympatric and allopatric Type B can be discerned for any of these statistics for any of these candidate genes. Table 4.3 presents the results of the McDonald-Kreitman tests. Fixed nonsynonymous substitutions are not more common in sympatric than allopatric comparisons; these tests provide no evidence for positive selection on the genes encoding these three

Table 4.2 Summary Statistics. n is the number of alleles sequenced from each population.

Candidate GRP genes	Population	n	θ	π (total)	π (synonymous sites)	π (nonsynonymous sites)
CIPRO37.40.1	Sympatric Type A	28	0.026	0.022	0.042	0.016
	Allopatric Type A	24	0.013	0.012	0.013	0.010
	Sympatric Type B	16	0.019	0.016	0.041	0.008
	Allopatric Type B	8	0.043	0.050	0.109	0.032
CIPRO60.5.1	Sympatric Type A	20	0.002	0.002	0.000	0.003
	Allopatric Type A	22	0.024	0.018	0.030	0.015
	Sympatric Type B	24	0.035	0.034	0.057	0.025
	Allopatric Type B	20	0.021	0.026	0.042	0.019
CIPRO100.7.1	Sympatric Type A	24	0.011	0.009	0.017	0.005
	Allopatric Type A	26	0.007	0.007	0.012	0.004
	Sympatric Type B	22	0.018	0.020	0.031	0.015
	Allopatric Type B	8	0.017	0.022	0.030	0.017
Control genes						
CIPRO53.35.1	Sympatric Type A	22	0.008	0.006	0.027	0.001
	Allopatric Type A	32	0.006	0.005	0.020	0.000
	Sympatric Type B	32	0.013	0.011	0.043	0.002
	Allopatric Type B	28	0.016	0.015	0.042	0.007
mtCOI	Sympatric Type A	10	0.000	0.000	0.000	0.000
	Allopatric Type A	12	0.006	0.004	0.021	0.000
	Sympatric Type B	9	0.005	0.005	0.023	0.000
	Allopatric Type B	11	0.005	0.005	0.021	0.001

Table 4.3 Results of the McDonald Kreitman Tests for all genes. FS = Fixed Synonymous. PS = Polymorphic Synonymous. FN = Fixed Nonsynonymous. PN = Polymorphic Nonsynonymous.

Candidate GRP genes	Test	FS	PS	FN	PN	P value (Fisher's Exact Two Tailed Test)
CIPRO37.40.1	Sympatric Type A vs. Sympatric Type B	0	35	0	36	NA
	Allopatric Type A vs. Allopatric Type B	9	38	15	48	0.644214
CIPRO60.5.1	Sympatric Type A vs. Sympatric Type B	6	27	7	22	0.755749
	Allopatric Type A vs. Allopatric Type B	3	16	6	19	0.709518
CIPRO100.7.1	Sympatric Type A vs. Sympatric Type B	9	25	12	30	1
	Allopatric Type A vs. Allopatric Type B	5	12	8	17	1

Control genes	Test	FS	PS	FN	PN	P value (Fisher's Exact Two Tailed Test)
CIPRO53.35.1	Sympatric Type A vs. Sympatric Type B	0	18	0	7	NA
	Allopatric Type A vs. Allopatric Type B	4	17	1	11	0.630145
mtCOI	Sympatric Type A vs. Sympatric Type B	76	9	3	1	0.384631
	Allopatric Type A vs. Allopatric Type B	70	21	3	2	0.590441

candidate GRPs in sympatry. None of the statistics for which significance was determined by coalescent simulation (D, D*, F*, H) showed significant differences between sympatry and allopatry for any of the genes examined, as all of the confidence intervals overlapped (Table 4.4 for Tajima's D, Table 4.5 for D* and F*, Table 4.6 for H).

Table 4.4 Tajima's D statistics for all genes

Candidate GRP genes	Population	D statistic	P-value	95% confidence interval
CIPRO37.40.1	Sympatric Type A	-0.078	0.281	"-1.68841 to 1.69923"
	Allopatric Type A	-0.004	0.157	"-1.20834 to 1.17"
	Sympatric Type B	-0.132	0.290	"-1.77 to 1.611"
	Allopatric Type B	0.006	0.092	"-0.99686 to 0.94395"
CIPRO60.5.1	Sympatric Type A	-0.018	0.541	"-1.51 to 1.66"
	Allopatric Type A	-0.087	0.561	"-1.69367 to 1.76737"
	Sympatric Type B	-0.058	0.528	"-1.45443 to 1.32636"
	Allopatric Type B	0.006	0.519	"-1.54796 to 1.72153"
CIPRO100.7.1	Sympatric Type A	-0.018	0.503	"-1.46162 to 1.35414"
	Allopatric Type A	0.009	0.510	"-1.3292 to 1.48969"
	Sympatric Type B	-0.014	0.515	"-0.92013 to 0.96708"
	Allopatric Type B	-0.042	0.526	"-1.35751 to 1.17643"
Control genes				
CIPRO53.35.1	Sympatric Type A	-0.057	0.544	"-1.68808 to 1.83296"
	Allopatric Type A	-0.002	0.511	"-1.37399 to 1.4227"
	Sympatric Type B	-0.027	0.524	"-1.42163 to 1.46487"
	Allopatric Type B	-0.061	0.545	"-1.31043 to 1.28035"
mtCOI	Sympatric Type A	-0.002	0.127	"-1.11173 to 1.43863"
	Allopatric Type A	-0.065	0.185	"-1.83094 to 1.77946"
	Sympatric Type B	-0.064	0.203	"-1.67754 to 1.75974"
	Allopatric Type B	-0.063	0.186	"-1.75914 to 1.8452"

Table 4.5 Fu and Li's D* and F* statistics for all genes

Candidate GRP genes	Population	D*	P-value	95% confidence interval	F*	P-value	95% confidence interval
CIPRO37.40.1	Sympatric Type A	-0.157	0.517	"-2.59 to 1.25985"	-0.111	0.496	"-2.498 to 1.52763"
	Allopatric Type A	-0.028	0.670	"-1.52677 to 1.239"	-0.063	0.589	"-1.58919 to 1.21221"
	Sympatric Type B	-0.034	0.578	"-2.24738 to 1.28647"	-0.090	0.540	"-2.66286 to 1.49659"
	Allopatric Type B	-0.024	0.734	"-1.00 to 0.87955"	0.007	0.612	"-1.03183 to 1.04566"
CIPRO60.5.1	Sympatric Type A	0.032	0.458	"-1.96617 to 1.25359"	-0.028	0.492	"-2.16055 to 1.477"
	Allopatric Type A	-0.040	0.432	"-2.464 to 1.25307"	-0.086	0.469	"-2.56615 to 1.49075"
	Sympatric Type B	-0.026	0.452	"-1.83617 to 1.17832"	-0.076	0.491	"-2.08 to 1.29"
	Allopatric Type B	-0.087	0.466	"-2.08275 to 1.17893"	-0.071	0.471	"-2.18321 to 1.38673"
CIPRO100.7.1	Sympatric Type A	-0.018	0.467	"-1.83829 to 1.22556"	-0.024	0.472	"-1.86692 to 1.35458"
	Allopatric Type A	-0.056	0.497	"-1.87585 to 1.23"	-0.070	0.506	"-2.09829 to 1.46478"
	Sympatric Type B	0.020	0.471	"-1.1613 to 0.99970"	-0.037	0.512	"-1.23262 to 1.02296"
	Allopatric Type B	-0.014	0.481	"-1.33924 to 1.16179"	-0.076	0.542	"-1.57039 to 1.29433"
Control genes							
CIPRO53.35.1	Sympatric Type A	-0.024	0.472	"-1.86692 to 1.35458"	0.013	0.456	"-2.20751 to 1.64394"
	Allopatric Type A	-0.070	0.506	"-2.09829 to 1.46478"	-0.012	0.501	"-1.68033 to 1.45582"
	Sympatric Type B	-0.037	0.512	"-1.23262 to 1.02296"	-0.057	0.504	"-1.91043 to 1.46351"
	Allopatric Type B	-0.076	0.542	"-1.57039 to 1.29433"	-0.004	0.474	"-1.88842 to 1.46422"
mtCOI	Sympatric Type A	0.008	0.762	"-1.24341 to 1.02623"	-0.039	0.786	"-1.34668 to 1.06879"
	Allopatric Type A	-0.006	0.440	"-2.229 to 1.40344"	-0.102	0.487	"-2.39875 to 1.54664"
	Sympatric Type B	-0.048	0.463	"-1.92508 to 1.43324"	-0.092	0.486	"-2.10684 to 1.58013"
	Allopatric Type B	-0.093	0.489	"-2.07471 to 1.42077"	-0.065	0.469	"-2.30147 to 1.50465"

Table 4.6 Fay and Wu's H test for all genes

Candidate GRP genes	Population	H statistic	P-value	95% confidence interval
CIPRO37.40.1	Sympatric Type A	-0.166	0.348	"-21.476 to 7.34392"
	Allopatric Type A	-0.040	0.434	"-6.39 to 4.01449"
	Sympatric Type B	0.114	0.333	"-16.6833 to 6.3"
	Allopatric Type B	0.233	0.430	"-16.64 to 11.64"
CIPRO60.5.1	Sympatric Type A	0.006	0.351	"-2.37427 to 1.222"
	Allopatric Type A	-0.216	0.338	"-20.5 to 7.41126"
	Sympatric Type B	-0.144	0.394	"-18.956 to 9.159"
	Allopatric Type B	-0.161	0.392	"-13.53684 to 5.69474"
CIPRO100.7.1	Sympatric Type A	0.080	0.387	"-6.28261 to 3.78261"
	Allopatric Type A	0.014	0.401	"-4.61538 to 2.75692"
	Sympatric Type B	0.174	0.436	"-8.8658 to 6.02597"
	Allopatric Type B	-0.055	0.425	"-11.00 to 6.07143"
Control genes				
CIPRO53.35.1	Sympatric Type A	-0.061	0.347	"-5.36797 to 2.07792"
	Allopatric Type A	0.003	0.413	"-2.61694 to 1.54839"
	Sympatric Type B	0.017	0.420	"-4.03226 to 2.54839"
	Allopatric Type B	-0.027	0.406	"-6.68783 to 3.35979"
mtCOI	Sympatric Type A	-0.001	0.718	"-1.06667 to 0.4444"
	Allopatric Type A	0.024	0.312	"-6.72727 to 2.4242"
	Sympatric Type B	-0.044	0.344	"-6.47222 to 2.58333"
	Allopatric Type B	-0.119	0.315	"-8.69091 to 2.7272"

Discussion

Comparison of d_n/d_s values between candidate GRPs and control proteins

Candidate GRPs in *C. intestinalis* are evolving more rapidly than control proteins, and this pattern is likely driven by substitutions at nonsynonymous sites. Rapid evolution has been documented at specific GRPs (*i.e.* lysin, VERL and bindin) in marine broadcast spawners, and d_n/d_s values are lower for mtCOI than the GRPs lysin and VERL for green and pink abalone (Clark et al. 2009). The pattern we see in *C. intestinalis* has also been documented in insects and mammals (e.g. butterflies: Walters & Harrison 2010; field crickets: Andres et al. 2008; mouse and human: Torgerson et al. 2002; primates: Wyckoff et al. 2000). This study suggests that a pattern of faster evolution in reproductive proteins than control proteins may apply to a wider group of organisms than previously imagined (*i.e.* external as well as internal fertilizers).

Evolution of candidate GRPs in Ciona intestinalis – no evidence for reinforcement

While d_n/d_s values were higher in sympatry than allopatry for CIPRO37.40.1 and CIPRO60.5.1, it appears that this result may be driven by differences in d_s. We therefore have no evidence that positive selection is enhanced in sympatry, and if these candidate GRPs are involved in prezygotic isolation, we have no evidence for enhanced prezygotic isolation. The polymorphism statistics likewise give no indication that reinforcement is driving the rapid evolution of these three proteins.

We cannot conclude from lack of evidence for reinforcement on CIPRO37.40.1, CIPRO60.5.1 and CIPRO100.7.1 that reinforcement is not occurring in this system. If reinforcement is indeed driving prezygotic isolation between Type A and B, there are several reasons why we might not have detected it in this study. First, primers for candidate GRPs were developed from the Type A genomic sequence and

were used to amplify and sequence both Type A and B individuals (the Type B genome has not been sequenced). But Type A and B are substantially divergent (p-distances: 0.124 at mtCOI, 0.035 to 0.116 for six nuclear loci; Nydam and Harrison 2010a), which could explain why 15 genes encoding GRP candidates could not be successfully amplified and/or sequenced in Type B individuals. It is possible that the genes that could not be amplified and/or sequenced (and were therefore excluded from the analyses) encode proteins that are evolving more rapidly between Type A and B than those that were included in the analyses. If this is the case, we may have missed proteins whose d_n/d_s values were greater than 0.5, proteins that would have been included in the sympatric vs. allopatric tests of reinforcement.

A second reason why reinforcement is possible in *C. intestinalis* despite lack of evidence from three rapidly evolving candidate GRPs is that these are not the proteins that are involved in prezygotic isolation. While components of the fertilization process in solitary ascidians such as *C. intestinalis* are well-characterized (Satoh 1994; Sawada et al. 2000), the genes and corresponding proteins responsible for species-specificity have not been identified as in other marine broadcast spawners.

Two sperm proteins that interact directly with the egg have been identified in *C. intestinalis,* but it is not known whether these proteins are involved in gamete recognition. The first protein is α L-fucosidase, which binds to the vitelline coat of the egg (Hoshi et al. 1985). Five of our candidate GRPs (CIPRO187.4.1, CIPRO19.75.1, CIPRO33.15.1, CIPRO552.7.1, and CIPRO58.12.1) had domains also found in α L-fucosidases, but none of these proteins were expressed in testis tissue, based on expression data in CIPRO. CIPRO187.4.1 could not be amplified from Type B, and the other four genes did not have d_n/d_s ratios > 0.5. A second protein is a chymotrypsin-like enzyme that may dissolve the vitelline coat of the egg (Marino et al. 1992). However, the amino acid sequence for this protein is not available and we

identified dozens of chymotrypsin-like proteins in the genome.

Alternatively, reinforcement may be driving enhanced prezygotic isolation, but before mating occurs, rather than at the fertilization step. We do not know whether Type A and B release gametes at the same time of day or in the same season in the English Channel. Since some gene flow has occurred (Nydam & Harrison 2010b), spawning must at least partially overlap. But if there is standing variation in the time and/or season of spawning, selection may be favoring those Type A individuals that spawn when the majority of Type B individuals do not.

It is also possible that reinforcement is not driving the evolution of prezygotic isolation in *C. intestinalis*. The process of reinforcement does not begin immediately after secondary contact, and a recent initiation of secondary contact in the English Channel could mean that there has not been enough time for the effects of reinforcement to be evident. Type B is native to Northern Europe and presumably a long-time resident of the English Channel. We do not know when Type A invaded the English Channel. The first published record of Type A in this area was in 2007 (Nydam and Harrison 2010b), but as Type A and B were only recognized in 2005 (Suzuki et al. 2005), Type A living in this area prior to 2005 would not have been distinguished from the native Type B. However, the introduction of Type A was likely human-mediated (Nydam & Harrison 2010b), a recent invasion on an evolutionary timescale. So the secondary contact between Type A and B might be too recent for reinforcement. However, evidence for reinforcement has been found in several instances where secondary contact is on the same time scale as that of Type A and B. For instance, in British Columbian sticklebacks, populations are so recent that phylogenetic relationships between them cannot be determined (Rundle & Schluter 1998). Similarly, the secondary contact between *Mus musculus musculus* and *Mus musculus domesticus* has only been in existence since sometime after the Neolithic

(9500 BCE) (Boursot et al. 1993).

Evolution of candidate GRPs in Ciona intestinalis – alternatives to reinforcement

Other explanations for this pattern of faster evolution of candidate sperm GRPs in Type A and B should be addressed, whether or not reinforcement is involved. The first explanation is sperm competition, which occurs in *Ciona* as it does in internal fertilizers. So selection could be acting on any proteins that determine how quickly sperm fertilize eggs: proteins involved in metabolism, motility, binding, penetration, etc. However, as Figure 4.1 shows that candidate GRPs specifically are evolving more rapidly than sperm proteins that are not candidate GRPs (control proteins), proteins directly involved in sperm-egg interactions are more likely to be experiencing directional selection than those involved in helping the sperm reach the egg.

Another process that could be leading to rapid evolution of sperm GRPs is sexual conflict, which occurs when the optimal outcomes of fertilization are different for sperm and eggs. For sperm, the optimal outcome is fertilization of an egg as quickly as possible. But fertilization of eggs by multiple sperm (polyspermy) results in developmental defects in many taxa. Therefore, the optimal outcome of fertilization for an egg may be slower fertilization, to avoid polyspermy. Ascidians like *C. intestinalis* often live in close proximity to many conspecific individuals (Lambert et al. 1997). Also, an individual usually sends sperm into the water column before eggs (ascidians are hermaphrodites). So eggs are released into a vast amount of sperm just spawned from many neighbors, making the risk of polyspermy very high. Perhaps in response to this risk, ascidians have evolved two separate blocks to polyspermy, whereas many other marine broadcast spawners have a single block (Lambert et al. 1997). Given these effective polyspermy blocks, sexual conflict resulting from polyspermy is not likely to be major driver of GRP evolution in

ascidians such as *C. intestinalis.*

Lastly, egg surface proteins could be changing rapidly to prevent pathogens from entering the egg. If the same proteins involved in preventing microbial attack are involved in sperm/egg recognition, this could lead to the rapid evolution of sperm proteins to keep up with the ever-changing egg proteins.

Positive selection on GRPs in other marine broadcast spawners: reinforcement

Some of the most rapidly evolving proteins yet discovered are GRPs in marine broadcast spawners (*e.g.* bindin in sea urchins, lysin in abalone and mussels). In sea urchins, the bindin protein facilitates sperm attachment to the egg and fusion of sperm and egg (Vacquier & Moy 1977). In three genera of sea urchin that contain sympatric species (*Echinometra, Heliocidaris,* and *Strongylocentrotus*), regions of bindin show evidence of positive selection (Lessios 2007 and references therein). In *Arbacia, Lytechinus* and *Tripneustes,* genera that do not contain sympatric species, bindin shows no evidence of positive selection (Lessios 2007 and references therein). This pattern is consistent with a reinforcement hypothesis (Lessios 2007; Palumbi 2009). However, if reinforcement is driving the evolution of GRPs between sympatric species, we would expect more nonsynonymous substitutions between species than within species (Zigler & Lessios 2003). This pattern is evident in *Heliocidaris* (Zigler et al. 2003), but not in *Echinometra* (Metz & Palumbi 1996) or *Stronglyocentrotus* (Debenham et al. 2000).

Stronger evidence for reinforcement comes from a study of *Echinometra oblonga,* which has populations that are sympatric and allopatric with *Echinometra* species C (Geyer & Palumbi 2003). Substantial divergence in bindin alleles between *E. oblonga* and *E.* sp. C. occurs where the two species are sympatric, but not where they are allopatric (Geyer & Palumbi 2003).

In abalone and mussels, sperm proteins known as lysins are involved in dissolution of the egg vitelline envelope, enabling the sperm to enter the egg. The best-characterized lysins are in the abalone genus *Haliotis.* An early study of 20 *Haliotis* species (19 sympatric and 1 allopatric species) found many pairwise comparisons with d_n/d_s values > 1 (Lee et al. 1995). A later study of 25 species corroborate the pairwise results of Lee et al. 1995 and also used maximum likelihood models of codon substitution to identify lineage and site-specific evidence of positive selection (Yang et al. 2000). Lineages containing sympatric or closely related species usually had d_n/d_s values > 1, whereas lineages with distantly related allopatric species always had d_n/d_s values < 1, a pattern consistent with reinforcement (Yang et al. 2000). The authors also note a d_n/d_s value > 1 for the two branches separating a group of Japanese species from two groups of Californian species; this speciation event was likely allopatric. So while a comprehensive study of *Haliotis* lysin presents a pattern that supports reinforcement, one should not conclude the action of reinforcement without explicitly testing whether d_n/d_s values are higher in sympatry than allopatry.

In the mussel *Mytilus galloprovincialis,* two divergent clades of Lysin-M7 have been found: G and G_D (Springer and Crespi 2007). Evidence of positive selection is seen between G and G_D, and within G_D (Springer & Crespi 2007). The divergence between the two clades is the result of rapid evolution in the G_D clade, and G_D alleles are found at higher levels in sympatric populations of *M. galloprovincialis* (where it hybridizes with *Mytilus edulis*) than in allopatric populations (Springer & Crespi 2007). Thus, secondary contact and subsequent gene flow between *M. galloprovincialis* and *M. edulis* may have resulted in the divergence of G_D (Springer & Crespi 2007). This pattern is consistent with a reinforcement scenario, as *M. galloprovincialis/M. edulis* hybrids are less fit than parentals (Bierne et al. 2006). However, the authors urge caution as the connection between Lysin-M7 G_D

divergence and prezygotic isolation has not been shown; many other evolutionary forces could lead to divergence in sympatry (Springer & Crespi 2007).

Differentiating between processes that lead to rapid evolution in marine GRPs

Rapid evolution, and in many cases positive selection, has become a well-known feature of GRPs in marine broadcast spawners. But in most cases the evolutionary process or processes responsible for this pattern remain elusive.

The process of reinforcement can easily be tested in an explicit manner (by comparing prezygotic isolation in sympatry vs. allopatry). One example of such a test exists in marine broadcast spawners: gametic compatibility (as measured by the concentration of sperm needed to fertilize 20% of the eggs) was compared between sympatric and allopatric populations of *Mytilus edulis* and *Mytilus trossulus* (Slaughter et al. 2008). But given the importance of GRPs in speciation in marine broadcast spawners (Palumbi 1992), this test should also be applied to GRPs directly.

The work discussed here provides the first explicit test of the reinforcement hypothesis (by comparing prezygotic isolation in sympatry vs. allopatry) where candidate GRPs are viewed as the source of the prezygotic isolation. Data from lysin in abalone and mussels, and from bindin in sea urchins, are compatible with a reinforcement hypothesis; this hypothesis can be tested in species pairs or groups that have both sympatric and allopatric populations. Such tests are particular in these taxa, as lysin and bindin have been shown to be directly involved in prezygotic isolation (not the case for the three candidate GRPs used in this study). Specific tests of the evolutionary processes behind the patterns of positive selection may provide important insights into the process of speciation in marine broadcast spawners.

Acknowledgements

We would like to thank Erin Newman-Smith at the Ascidian Stock Center (supported by grant # R24GM075049 from the National Institutes of Health) for supplying animals, the Smith Laboratory at UCSB and John Bishop at Plymouth MBA for providing logistical assistance, Dr. Kazuo Inaba for sharing information about the CIPRO database, Derek Smith and the staff of the University of Victoria Genome BC Proteomics Centre for generating the proteomics data, and members of the Harrison Laboratory for advice. This work was funded by NSF Doctoral Dissertation Grant DEB-0907862 to M.L.N. and R.G.H. and NSF Grant DEB-0639904 to R.G.H.

REFERENCES

Andres, J. A., Maroja, L.S., and R.G. Harrison. 2007. Searching for candidate speciation genes using a proteomics approach: seminal proteins in field crickets. P. Roy. Soc. B. 275:1975-1983.

Bierne, N., F. Bonhomme, P. Boudry, M. Szulkin, and P. David. 2006. Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. P. Roy. Soc. B. 273:1253-1260.

Boursot, P., Auffray, J.-C., Britton-Davidian, J. and F. Bonhomme. 1993. The evolution of house mice. Ann. Rev. Ecol. Sys. 24:119–152.

Butlin, R.K. 1987. Speciation by reinforcement. Trends Ecol. Evol. 2: 8-13.

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M. and P. Sordino. 2007. Cryptic speciation in a model invertebrate chordate. Proc. Natl. Acad. Sci. USA 104:9364-9369.

Clark, N.L., Gasper, J., Sekino, M., Springer, S.A., Aquadro, C.F., and W. J. Swanson. 2009. Coevolution of interacting fertilization proteins. PLOS Genetics e:1000570.

Conover, W.J. 1999. Practical Nonparametric Statistics. 3rd ed. Wiley Publishing.

Coyne, J. A., and H.A. Orr. 1989. Patterns of speciation in *Drosophila.* Evolution 43:362-381.

Coyne, J. A., and H.A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51:295-303.

Coyne, J. A., and H.A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland,

MA.

Debenham, P., M. A. Brzezinski, and K. R. Foltz. 2000. Evaluation of sequence variation and selection in the bindin locus of the red sea urchin, *Strongylocentrotus franciscanus*. J. Mol. Evol. 51:481–490.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

Dytham, C. 2003. Choosing and using statistics – a biologist's guide. 2nd edition. Blackwell Publishing, London, UK.

Fay, J., and C-I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Fu, Y.X., and W.H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Geyer, L., and S. Palumbi. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution 57:1049-1060.

Hoshi, M., de Santis R, Pinto MR, Cotelli F, and F. Rosati. 1985. Sperm glycosidases as mediators of sperm-egg binding in the ascidians. Zool. Sci. 2:65-69.

Inaba, K., Padma, P., Satouh, Y., Shin-I, T., Kohara, Y., Satoh, N., and Y. Satou. 2002. EST analysis of gene expression in testis of the ascidian Ciona intestinalis. Mol. Repro. Dev. 62:431-445.

Kott, P. 1952. The Ascidians of Australia. Aus. J. Mar. Fresh. Res. 3:206-233.

Lambert, C. C., H. Goudeau, C. Franchet, G. Lambert and M. Goudeau. 1997.

Ascidian eggs block polyspermy by two independent mechanisms, one at the egg plasma membrane, the other involving the follicle cells. Mol. Repro. Dev. 48:137-143.

Lee, Y.-H., T. Ota, and V. D. Vacquier. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol. Biol. Evol. 12:231-238.

Lessios, H. A. 2007. Reproductive isolation between species of sea urchins. Bull. Mar. Sci. 81:191–208.

Linne, C. 1767. Systema Naturae. Vindobonae: Typis Ioannis Thomae, 1767-1770.

Marino, R., De Santis, R., Hirohashi N, Hoshi M, Pinto MR, and N. Usui. 1992. Purification and characterization of a vitelline coat lysine from *Ciona intestinalis* spermatozoa. Mol. Repro. Dev. 32:383-388.

Matute, D.R. 2010. Reinforcement of gametic isolation in *Drosophila.* PLoS Biology 8: e1000341.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652-654.

Metz, E.C., Robles, S.R., and V.D. Vacquier. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc. Natl. Acad. Sci. USA 95:10676–10681.

Metz, E. C., and S.R. Palumbi. 1996. Positive Selection and Sequence Rearrangements Generate Extensive Polymorphism in the Gamete Recognition Protein Bindin. Mol. Biol. Evol. 13:397-406.

Monniot, C., and F. Monniot. 1994. Additions to the inventory of eastern tropical Atlantic ascidians: arrival of cosmopolitan species. Bull. Mar. Sci. 54:71-93.

Nydam, M.L, and R.G. Harrison 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). Mar. Biol. 151:1839-1847.

Nydam, M.L, and R.G. Harrison 2010a. Polymorphism and divergence within the ascidian genus *Ciona.* Mol. Phyl. Evol. 56:718-726.

Nydam, M.L, and R.G. Harrison 2010b. Introgression despite substantial divergence in a broadcast spawning marine invertebrate. Evolution. Submitted.

Palumbi, S.R. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. Heredity 102:66-76.

Palumbi, S. 1992. Marine speciation on a small planet. Trends Ecol. Evol. 7:114-118.

Pond, S.K., and S.V. Muse. 2005. Site-to-Site variation of synonymous substitution rates. Mol. Biol. Evol. 22:2375-2385.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Rundle, H. D., and D. Schluter. 1998. Reinforcement of stickleback mating preferences: Sympatry breeds contempt. Evolution 52:200-208.

Saetre, G.-P., Moum, T., Bures, S., Kral, M., Adamjan, M., and J. Moreno. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. Nature 387:589-592.

Satoh, N. 1994. Developmental Biology of Ascidians. 1st ed. Cambridge University

Press, Cambridge.

Sawada, H., Yokosawa, H., and C.C. Lambert, eds. 2001. The Biology of Ascidians. 1st ed. Springer, New York.

Shapiro S.S., and M.B. Wilk. 1965. An analysis of variance test for normality. Biometrika 52: 591-599.

Slaughter, C., McCartney, M.A., and P.O. Yund. 2008. Comparison of gamete compatibility between two blue mussel species in sympatry and in allopatry. Biol. Bull. 214:57-66.

Springer S.A., and B.J. Crespi. 2007. Adaptive gamete-recognition divergence in a hybridizing *Mytilus* population. Evolution 61:772–783.

Swanson, W. J., and V.D. Vacquier. 2002. Reproductive protein evolution. Annu. Rev. Ecol. Syst. 33:161-179.

Swanson, W. J., Wong, A., Wolfner, M.F., and C.F. Aquadro. 2004. Evolutionary Expressed Sequence Tag Analysis of *Drosophila* Female Reproductive Tracts Identifies Genes Subjected to Positive Selection. Genetics 168:1457-1465.

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

Torgerson, D.G., Kulathinal, R.J., and R.S. Singh. 2002. Mammalian Sperm Proteins Are Rapidly Evolving: Evidence of Positive Selection in Functionally Diverse Genes. Mol. Biol. Evol. 19:1973-1980.

Vacquier, V.D., and G.W. Moy. 1977. Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. Proc. Natl. Acad. Sci. USA 74:2456-2460.

Van Name, W. 1945. The north and south American ascidians. B. Am. Mus. Nat. Hist. 84:1-476.

Walters, J.R., and R.G. Harrison. 2010. Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. Mol. Biol. Evol. doi:10.1093/molbev/msq092

Wullschleger E.B., Wiehn J., and J. Jokela. 2002. Reproductive character displacement between the closely related freshwater snails *Lymnaea peregra* and *L. ovata*. Evol. Ecol. Res. 4:247–257.

Wyckoff, G.J., Wang, W., and C.-I. Wu. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24:1586-1591.

Yang, Z., Swanson, W.J., 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.

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., Raff, E.C., Popodi, E., Raff, R.A., and H.A. Lessios. 2003. Adaptive evolution of binding in the genus *Heliocidaris* is correlated with the shift to direct development. Evolution 57:2293-2302.