The Influence of Genetic Variation on Susceptibility of Common Bottlenose Dolphins (*Tursiops truncatus*) to Harmful Algal Blooms

by

Kristina Marstrand Cammen

University Program in Ecology
Duke University

Date:_______________________
Approved:

____
Andrew J. Read, Supervisor

____
Richard T. Di Giulio

____
Eric P. Palkovacs

____
Patricia E. Rosel

____
Randall S. Wells

Dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the University Program in Ecology
in the Graduate School of Duke University

2014

ABSTRACT

The Influence of Genetic Variation on Susceptibility of Common Bottlenose Dolphins (*Tursiops truncatus*) to Harmful Algal Blooms

by

Kristina Marstrand Cammen

University Program in Ecology
Duke University

Date:_______________________
Approved:

____
Andrew J. Read, Supervisor

____
Richard T. Di Giulio

____
Eric P. Palkovacs

____
Patricia E. Rosel

____
Randall S. Wells

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy in the University Program in Ecology
in the Graduate School of Duke University

2014

ABSTRACT

The capacity of marine organisms to adapt to natural and anthropogenic stressors is an integral component of ocean health. Harmful algal blooms (HABs), which are one of many growing threats in coastal marine ecosystems, represent a historically present natural stressor that has recently intensified and expanded in geographic distribution partially due to anthropogenic activities. In the Gulf of Mexico, HABs of *Karenia brevis* occur almost annually and produce neurotoxic brevetoxins that have been associated with large-scale mortality events of many marine species, including the common bottlenose dolphin (*Tursiops truncatus*). The factors resulting in large-scale dolphin mortality associated with HABs are not well understood, particularly in regards to the seemingly different impacts of HABs in geographically disjunct dolphin populations. My dissertation investigates a genetic basis for resistance to HABs in bottlenose dolphins in central-west Florida and the Florida Panhandle. I used both genome-wide and candidate gene approaches to analyze genetic variation in dolphins that died putatively due to brevetoxicosis and live dolphins from the same geographic areas that survived HAB events. Using restriction site-associated DNA sequencing, I identified genetic variation that suggested both a common genetic basis for resistance to HABs in bottlenose dolphins across the Gulf coast of Florida and regionally specific resistance. Many candidate genes involved in the immune, nervous, and detoxification systems were found in close genomic proximity to survival-associated polymorphisms throughout the bottlenose dolphin genome. I further investigated two groups of

candidate genes, nine voltage-gated sodium channel genes selected because of their putative role in brevetoxin binding and four major histocompatibility complex (MHC) loci selected because of their genomic proximity to a polymorphism exhibiting a strong association with survival. I found little variation in the sodium channel genes and conclude that bottlenose dolphins have not evolved resistance to HABs via mutations in the toxin binding site. The immunologically relevant MHC loci were highly variable and exhibited patterns of genetic differentiation among geographic regions that differed from neutral loci; however, genetic variation at the MHC also could not fully explain variation in survival of bottlenose dolphins exposed to HABs. In my final chapter, I consider the advantages and drawbacks of the genome-wide approach in comparison to a candidate gene approach and, as laid out in my dissertation, I recommend using both complementary approaches in future investigations of adaptation in genome-enabled non-model organisms.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been possible without the contributions of many collaborators and the support of colleagues, friends, and family members. I am grateful for the support of my advisor Andy Read, whose constant faith in my abilities and continual encouragement to “think big” pushed me to take on challenges in both my research and professional development. In addition to Andy, I feel fortunate to have a diverse committee, including Patty Rosel, Randy Wells, Eric Palkovacs and Rich Di Giulio, who have contributed broadly to my thinking on conservation, genetics, evolution, and toxicology. Patty and Randy both welcomed me openly into their research groups, whom I acknowledge further below. In particular I would like to thank Patty, who has been a valued mentor and collaborator, for all of the insight she provided during my time in her lab and on many phone calls over the past years.

Most of my lab work was conducted at the Duke University Marine Conservation Molecular Facility, directed by Tom Schultz. I have greatly appreciated that Tom’s office door is always open to students, and his encouragement and initial tool development made it possible to include a genomic sequencing component in this research. I would also like to thank the many other students, post-docs, and technicians who have worked alongside me in the MCMF over the past five years – Emily Argo, Bernie Ball, Sara Blinebry, Brianne Soulen, and Andrew Thaler. Thank you in particular to Sophie Plouviez for the advice she has offered on many aspects of the molecular analyses described in this dissertation.

There are two research groups outside of Duke who have been integral in the completion of this dissertation: the NOAA SEFSC Marine Mammal Molecular Genetics Lab run by Patty Rosel and the Chicago Zoological Society's Sarasota Dolphin Research Program run by Randy Wells. Over the past five years, I have spent time working with both groups and developed great research collaborations and friendships. In particular, I would like to thank Lynsey Wilcox for all of the training she provided me during my time working in the MMMGL and for the microsatellite and MHC data she has contributed to this thesis. I would also like to thank Brian Balmer who led our sample processing team during dolphin health assessments in Sarasota Bay.

Samples for this research were provided by the NOAA SEFSC Marine Mammal Tissue and DNA Archive, the Sarasota Dolphin Research Program, the Mote Marine Laboratory Stranding Investigations Program, and the North Carolina Marine Mammal Stranding Network. Samples were collected under the following Scientific Research Permits issued by the National Marine Fisheries Service: 522-1527, 522-1569, 522-1785, and 15543 issued to RSW; 932-1489 (as amended) issued to the NOAA Marine Mammal Health and Stranding Response Program. Toxin data that informed early stages of study design were generously provided by Michael Twiner, Deborah Fauquier, Leanne Flewelling, and Spencer Fire. Funding for the research was provided by the PADI Foundation, the American Fisheries Society, the Duke University Marine Lab Kaupe fund, Dolphin Quest, and the Oak Foundation. Graduate student fellowships were provided by the Duke University Scholars Program, the Oak Foundation, and Katherine Goodman Stern.

Last, but definitely not least, I would like to acknowledge my family and friends who have supported me throughout my graduate career. Thank you to my parents, Leon and Inge-Lise Cammen, for their encouragement over the years, and to my partner, Joshua Stoll, for his continual love and support. Thank you to Rebecca Gruby, Meagan Dunphy-Daly, Amy Freitag, Noëlle Boucquey, Reny Tyson, Corrie Curtice, Matthew Bowers, and the many others who have made my time as a PhD student memorable. The Duke Marine Lab is an amazing community in which I have made lasting friendships and professional connections that I hope to maintain for many years to come.

INTRODUCTION

Marine organisms utilize a range of mechanisms to adapt to natural and anthropogenic stressors in their environment. This capacity to adapt is an integral component of ocean health in marine ecosystems that have been drastically altered by human activities. Growing threats in coastal marine ecosystems, such as emerging diseases, fisheries exploitation, persistent pollutants, and, of particular relevance to my dissertation, the increasing occurrence of harmful algal blooms (HABs), likely provide strong selection pressures that may induce adaptation dependent on temporal and spatial factors.

While evolution was originally viewed as a slow and gradual process, it is now generally accepted that given the right conditions microevolution within a species or population can occur rapidly, oftentimes in years to decades (Stockwell *et al.* 2003). A growing number of studies have confirmed that evolution can occur on ecological time scales in both artificial selection experiments and in natural systems (Hairston Jr *et al.* 2005; Reznick & Ghalambor 2001; Thompson 1998). The majority of examples from natural systems describe adaptation in populations that are exposed to strong selective pressures of anthropogenic origin, such as fisheries-induced evolution (Allendorf & Hard 2009) or adaptation to chemical pollutants (Meyer & Di Giulio 2003). In contrast, few studies have investigated contemporary evolution due to natural threats. Naturally produced biotoxins present an interesting case that could demonstrate both historical and contemporary adaptation. Fossilized remains (Emslie & Morgan 1994; Pyenson *et al.*

2014) and archival records from early explorers and fishers (Bandelier 1905, cited by Magaña *et al*., 2003; Ingersoll 1882, cited by Heil & Steidinger 2009) suggest that HABs have been present in the marine environment for centuries. However, HABs have also recently expanded in geographic distribution and are therefore affecting previously unexposed populations (Van Dolah 2000).

A spatially heterogeneous threat, like HABs, may favor local adaptation. Local adaptation occurs primarily in systems with low gene flow, strong spatially heterogeneous selection and little temporal variation in selection (Kawecki & Ebert 2004). In addition, local adaptation is generally favored when there is some constraint on phenotypic plasticity (i.e., non-genetic responses to an environmental influence, such as developmental or behavioral changes) (Ghalambor *et al.* 2007). Marine systems are generally expected to be characterized by high gene flow due to the relative lack of barriers and broad dispersal capabilities of many species. However, there is increasing evidence of population structure and genetic differentiation in many marine species capable of wide dispersal (e.g., Rosel *et al.* 2009), and local adaptation has been observed in several fishes in high gene flow environments (Hemmer-Hansen *et al.* 2007; Nielsen *et al.* 2009). In aquatic environments, local adaptation has been described across a range of species threatened by disease (Cammen *et al.* 2011), fisheries exploitation (Allendorf & Hard 2009), pollutants (Meyer & Di Giulio 2003), and HABs (Connell *et al.* 2006).

Local adaptation can be detected using reciprocal transplant or common garden experiments. However, in systems where transplant and exposure experiments are not logistically or ethically feasible, it is possible to study adaptation at the molecular level by comparing genetic variation among populations that differ in habitat or comparing cases and controls within populations. In the growing field of evolutionary and ecological functional genomics, researchers investigate adaptation by studying how genes that underlie ecologically important traits affect fitness in the natural environment (Feder & Mitchell-Olds 2003). Thus, in my dissertation I utilize complementary genome-wide and candidate gene approaches to investigate adaptation to HABs in common bottlenose dolphins (*Tursiops truncatus*) in the Gulf of Mexico.

HAB occurrence and toxicity

In the past decade, algal blooms of species that produce biotoxins have increased in frequency and geographic distribution worldwide (Van Dolah 2000). Some of this observed increase in HABs may be attributable to improved surveillance of coastal waters, but in addition, climate change, nutrient run-off, and ship ballast water transfer have likely contributed to a true increase in HABs. Regardless of the cause, the increase in HABs poses a significant threat to marine species and humans inhabiting coastal areas. The toxins released by harmful algae cause seafood-related human poisonings and large-scale mortalities of marine species (Van Dolah 2000). As HABs increase worldwide, it becomes increasingly important that we understand how marine species have adapted to survive exposure to algal toxins.

The Gulf of Mexico experiences almost annual blooms of the dinoflagellate *Karenia brevis* (previously known as *Gymnodinium breve* and *Ptychodiscus brevis*). Spanish explorers first described these HABs, also known as red tides, in the 1500s, and since the late 1800s, blooms have been recorded almost every year in the eastern Gulf of Mexico along the southwest coast of Florida (Steidinger 2009). Consistent with the worldwide trend described above, red tides in the Gulf of Mexico have increased in geographic distribution and duration in recent years. Overall, *K. brevis* has become more abundant in the eastern Gulf of Mexico, increasing 13-18 fold in concentration (cells/L) from the period 1954-1963 to 1994-2002 (Brand & Compton 2007). Additionally, blooms have been reported in the northern Gulf of Mexico along the Florida Panhandle (Flewelling *et al.* 2005) and in the western Gulf of Mexico along the coast of Texas (Magaña *et al.* 2003). On average, blooms last two to four months, but is now not uncommon for a bloom to last between eight and twelve months (Kusek *et al.* 1999). Although historically blooms were most common in the late fall and early winter, recently blooms have begun to start in August and extend through the spring. This increase in blooms of *K. brevis* in the eastern Gulf of Mexico may be partially attributed to anthropogenic influences, particularly the increased flow of nutrients into the coastal waters (Brand & Compton 2007); however, the historical occurrence of red tides implies that agricultural nutrient input is not a necessary factor in bloom development (Walsh *et al.* 2006).

The increase in red tides threatens coastal ecosystems in the Gulf of Mexico. *K. brevis* produce brevetoxins (PbTx), which are known to be potent neurotoxins (Poli *et al.* 1986) and may have several other modes of toxicity (e.g., respiratory irritant (Singer *et al.* 1998), genotoxic (Sayer *et al.* 2005), immunotoxic (Bossart *et al.* 1998), teratogenic (Kimm-

Brinson & Ramsdell 2001)). In humans, brevetoxins cause neurotoxic shellfish poisoning (NSP) if ingested and respiratory irritation if inhaled (Kirkpatrick *et al.* 2004). Brevetoxins can be lethal in marine species, which experience greater exposure to *K. brevis*. HABs of *K. brevis* have been implicated in large-scale mortality events of fish, sea birds, sea turtles, and marine mammals. Neurological symptoms observed in affected manatees include disorientation, listlessness, and inability to submerge or maintain a horizontal position (O'Shea *et al.* 1991). Bottlenose dolphins that strand putatively due to brevetoxicosis are usually dead upon recovery, so the symptoms of brevetoxicosis in bottlenose dolphins are poorly understood.

Marine organisms may be exposed to brevetoxins directly in the water column, through inhalation of aerosolized toxin, or through the ingestion of *K. brevis* cells or contaminated prey (Bossart *et al.* 1998; Flewelling *et al.* 2005; Van Dolah 2000). Brevetoxins can be potently ichthyotoxic, but certain fish are able to survive high concentrations of *K. brevis* and accumulate brevetoxins (Naar *et al.* 2007). As brevetoxins biomagnify they become a threat for upper trophic level predators in these coastal ecosystems, such as the common bottlenose dolphin.

Bottlenose dolphin mortality

The bottlenose dolphins in the Gulf of Mexico at risk of red tide exposure include several coastal and estuarine populations (Waring *et al.* 2013). Bottlenose dolphin estuarine populations may be particularly susceptible to the impact of localized threats, like HABs, due to their relatively small size and high site fidelity to their respective bays along the Gulf coast (Scott *et al.* 1990). However, the dolphin metapopulation structure may facilitate local adaptation, potentially leading to increased resistance in some populations (Kawecki & Ebert 2004). High site fidelity and limited gene flow has resulted in low but significant genetic differentiation among dolphin populations along the Gulf coast of Florida at a fine geographic scale (Sellas *et al.* 2005).

The first association between red tides and bottlenose dolphin mortality was reported in 1946 (Gunter *et al.* 1948). Since 1999, four bottlenose dolphin unusual mortality events (UMEs) have been attributed to blooms of *K. brevis* in the Gulf of Mexico (Waring *et al.* 2013). These events were considered UMEs, as defined by the Marine Mammal Protection Act, because they were unexpected, impacted a large portion of the dolphin population, and deviated from the historical average number of strandings in this region (Waring *et al.* 2013). This dissertation focuses primarily on UMEs that occurred in 1999-2000 and 2004 in the Florida Panhandle and in 2005-2006 in central-west Florida.

From August 1999 to February 2000, approximately 120 bottlenose dolphins stranded concurrent with a red tide (16×10^6 *K.* brevis cells/L) in the Florida Panhandle (Mase *et al.* 2000; Twiner *et al.* 2012). Of those tested, many stranded dolphins were positive for brevetoxins at high concentrations in their stomach contents (e.g., partially digested fish), and liver, kidney, lung, and spleen tissues (Mase *et al.* 2000; Twiner *et al.* 2012). Histological examination indicated upper respiratory tract lesions, similar to those observed in manatees that have inhaled aerosolized brevetoxins (Bossart *et al.* 1998).

Additionally, some dolphins suffered from lymphoid tissue depletion, indicating possible immunosuppression as a result of brevetoxin exposure (Mase *et al.* 2000). In contrast to the 1999 event, no significant bloom (≤0.001 *K.* brevis cells/L) was observed during the 2004 Florida Panhandle UME that affected over 100 bottlenose dolphins (National Marine Fisheries Service 2004). However, many of these dolphins were found with brevetoxin-contaminated fish in their stomachs (Twiner *et al.* 2012), suggesting that toxin vectors (e.g., fish feeding in other waters with an undetected bloom) may allow for delayed or remote dolphin exposure and mortality (Flewelling *et al.* 2005).

Red tides occur almost annually in central-west Florida (Florida Fish and Wildlife Research Institute 2008; Steidinger 2009), but relatively little dolphin mortality associated with HAB exposure has been reported in this region. Interestingly, brevetoxins have been detected in live bottlenose dolphins that show no evidence of brevetoxicosis (Fire *et al.* 2008b). Only a single HAB-associated dolphin UME has occurred in central-west Florida in the past two decades, and this mortality event was associated with an unusually severe and prolonged HAB. During red tides that persisted from February to December 2005 and August to December 2006, at least 165 bottlenose dolphins stranded (pers. comm. J. Litz), in addition to over 300 sea turtles (Fauquier *et al.* 2013) and 80 manatees (Florida Fish and Wildlife Conservation Committee 2006, 2007).

The factors resulting in large-scale dolphin mortality associated with red tides are not well understood, particularly in regards to the seemingly different impacts of red tides in geographically disjunct dolphin populations. Bloom densities and toxicities are similar in central-west Florida and the Florida Panhandle, making it difficult to explain the geographical variation in the impact of red tides based solely on these characteristics of the HABs (Table 1). The range of *K. brevis* cell counts and brevetoxin concentrations observed during Florida Panhandle UMEs are regularly observed in central-west Florida without associated large-scale dolphin mortality (Van Dolah *et al.* 2003).

Additionally, previous studies of relevance to the trophic transfer of brevetoxins have found little evidence suggesting a difference in bottlenose dolphin exposure to the HAB toxins across geographic regions. Similar concentrations of brevetoxin were found in fish sampled during HABs in both Sarasota Bay in central-west Florida (Fire *et al.* 2008a) and St. Joseph Bay in the Florida Panhandle (Flewelling *et al.* 2005; Naar *et al.* 2007), and the stomach contents of stranded dolphins from both areas contained fish with comparable levels of brevetoxins (Fire *et al.* 2007; Flewelling *et al.* 2005; Hinton & Ramsdell 2008). Bottlenose dolphins positive for brevetoxins have stranded during red tides in central-west Florida outside of the 2005-06 UME, however, these strandings were comparatively few in number per red tide (Fire *et al.* 2007). These results suggest that bottlenose dolphins in central-west Florida may be exposed to brevetoxins even during red tides that do not result in large-scale mortality.

Table 1: Summary of *Karenia brevis* and brevetoxin (PbTx) concentrations in water, dolphin tissue, and fish sampled in the Florida Panhandle and central-west Florida. <dl: below detection limit, which ranged between sample type and detection assay. Adapted from Twiner *et al.* (2012).

				Brevetoxin concentration (ng/g or ng/ml)					
		max *K. brevis* cells/L (x10[6])	% tested PbTx pos	Liver	Urine	Feces	Blood	Stomach Contents	Fish (viscera)
FL Panhandle	1999-2000 UME[g]	16	52%	<dl-138				28-500+	
	2004 UME[g]	≤0.001	100%	<dl-104	<dl-143	<dl-1,152	<dl-16	<dl-29,126	60-1,500[d]
	2005-06 UME[g]	4.8	93%	<dl-278	<dl-48	<dl-143	<dl-35	<dl-2,724	31-386[e]
Central-west FL	Non-UME HABs, 2001-03[a]	105	84%	<dl-71	<dl-100	288-749		<dl-2896	
	Dolphins surviving HABs, 2003-05[c]	5.39	63%		<dl-5	<dl-231			4-5,188[b]
	2005-06 UME[f]	162	72%	<dl-788	<dl-54	<dl-774	<dl-13	<dl-26,390	

[a]Fire *et al.* (2007); [b]Fire *et al.* (2008a); [c]Fire *et al.* (2008b); [d]Flewelling *et al.* (2005); [e]Naar *et al.* (2007); [f]NOAA & FWC (unpubl. data), cited in Twiner *et al.* (2012); [g]Twiner *et al.* (2012).

The variation in survival among bottlenose dolphins exposed to red tides may be due to differences in individual susceptibility. Resistance to red tides may be affected by external factors, such as exposure to other HAB toxins or anthropogenic pollutants, or by internal factors, such as genetic variation. The former hypothesis has been previously investigated in the case of domoic acid (Twiner *et al.* 2011; Twiner *et al.* 2012), DDTs, and PCBs (Kucklick & Yordy 2010); these toxins have all been measured at approximately similar concentrations in bottlenose dolphins from central-west Florida and the Florida Panhandle. Additionally, preliminary studies in mouse models indicate that domoic acid at the levels currently measured in bottlenose dolphins in the Gulf of Mexico do not increase the potency of brevetoxins (Naar & Flewelling, unpubl. data, cited in Gaydos, 2006).

My dissertation investigates a genetic basis for resistance to red tides. Historical records suggest that HABs have occurred on a regular basis for centuries in central-west Florida (Florida Fish and Wildlife Research Institute 2008; Steidinger 2009), but have been comparatively infrequent in the Florida Panhandle (Gaydos, 2006; L. Flewelling, pers. comm.). As a result, dolphin populations in central-west Florida have been exposed to a consistently strong selective pressure for HAB resistance, while dolphins in the Florida Panhandle may be relatively naïve and therefore more susceptible to HABs (Van Dolah *et al.* 2003). I hypothesize that this difference in selection could have resulted in local adaptive divergence between the populations, favoring genotypes that confer HAB resistance in the central-west Florida dolphin population.

Research objectives

The overall objective of my dissertation was to investigate the role of genetic variation in apparent differences in survival of common bottlenose dolphins exposed to HABs in central-west Florida and the Florida Panhandle. To achieve this objective, I used both genome-wide and candidate gene approaches to analyze genetic variation in stranded dolphins that died putatively due to brevetoxicosis and live dolphins that had previously survived red tide exposure (Fig. 1). I analyzed DNA samples collected from live dolphins representing the estuarine populations of Sarasota Bay in central-west Florida and St. Joseph Bay in the Florida Panhandle, as well as the adjacent coastal populations. I compared samples from live dolphins collected in the Florida Panhandle to samples collected from dolphins that stranded during HAB-associated UMEs in 1999 and 2004. Similarly, I compared samples from live dolphins collected in central-west Florida to those from dolphins that stranded during a HAB-associated UME in 2005-2006, as well as dolphins that stranded during other red tides in central-west Florida between 1992 and 2003. Most samples were provided by the NOAA Southeast Fisheries Science Center (SEFSC) Marine Mammal Tissue and DNA Archives. Additional samples were provided by the Mote Marine Laboratory stranding program and collected during health assessments run by the Chicago Zoological Society's Sarasota Dolphin Research Program.

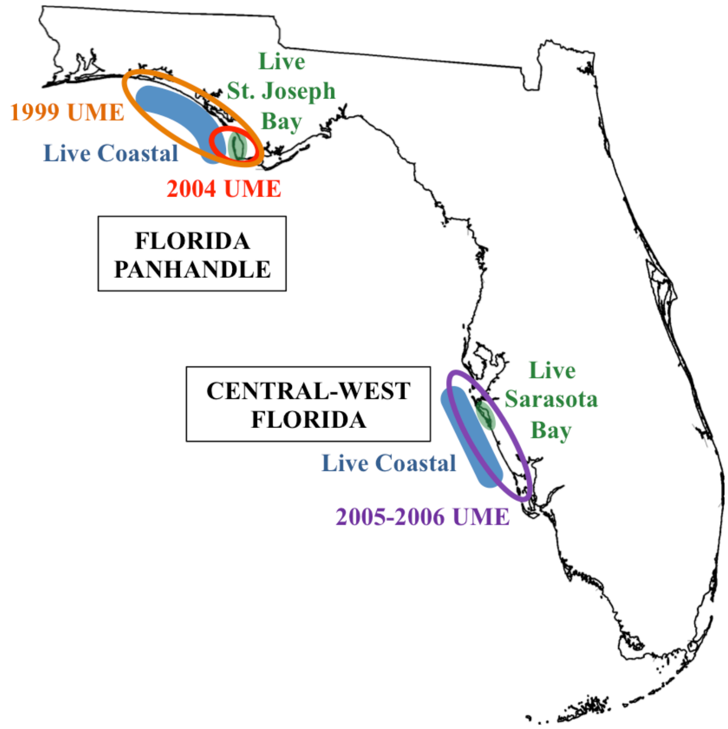

Figure 1: Map of approximate distribution of sampling locations, including live coastal and estuarine populations of common bottlenose dolphins from central-west Florida and the Florida Panhandle, and dolphins that died during unusual mortality events (UMEs) associated with red tides.

In Chapter 1, I describe a genome-wide approach to studying the association between genetic variation and survival in pools of live and dead bottlenose dolphin samples, as described above. I used restriction site-associated DNA (RAD) sequencing to investigate variation in 129,594 38-bp loci across the bottlenose dolphin genome. To improve our understanding of the functional mechanism of HAB resistance in bottlenose dolphins, I mapped all RAD loci to the dolphin genome and identified genes in close genomic proximity to survival-associated polymorphisms. This chapter concludes with a discussion of the genes that are involved in the immune, nervous, and detoxification systems and therefore may play a role in brevetoxin resistance. I further investigate several immune and nervous system candidate genes in the following chapters.

In Chapter 2, I characterize genetic variation at the putative brevetoxin binding site on the voltage-gated sodium channels of bottlenose dolphins. Across a wide range of terrestrial and marine species, neurotoxin resistance has independently evolved via mutations in toxin binding sites within the voltage-gated sodium channel gene complex (Anderson 1987; Bricelj *et al.* 2005; Geffeney *et al.* 2005; Jost *et al.* 2008a). I therefore tested the hypothesis that similar genetic variation in the putative binding site of brevetoxins in bottlenose dolphins explains differences in individual or population-scale resistance to harmful *K. brevis* blooms in the Gulf of Mexico.

In Chapter 3, I describe a similar candidate gene sequencing approach to investigate four major histocompatibility complex (MHC) loci in close genomic proximity to a significant RAD locus identified in Chapter 1. The MHC is a family of highly polymorphic genes that plays an important immunological role in antigen presentation (Germain 1994; Germain *et al.* 1996). Many studies have identified associations between genetic diversity at MHC genes and resistance to bacterial, viral, and parasitic pathogens (reviewed by Sommer 2005), but there has been relatively little prior investigation of the role of MHC diversity in resistance to environmental toxins (Cohen 2002; Ingale 2010).

Finally, in Chapter 4, I reflect upon the combination of genome-wide and candidate gene approaches used in this dissertation. My dissertation was initially conceived as a series of candidate gene investigations, but ongoing advances in genomic sequencing enabled me to add the genome-wide approach described in Chapter 1. In my concluding chapter, I offer insights into the promise of genomic sequencing for studies of adaptation in non-model organisms in their natural environment. I consider the advantages and drawbacks of the genome-wide approach in comparison to a candidate gene approach and, as laid out in my dissertation, I recommend using both complementary approaches to investigate adaptation in genome-enabled non-model organisms.

CHAPTER I:
GENOME-WIDE INVESTIGATION OF ADAPTATION TO HARMFUL ALGAL BLOOMS IN COMMON BOTTLENOSE DOLPHINS (*TURSIOPS TRUNCATUS*)

Introduction

The co-evolutionary arms race of pathogen evasion and host resistance has been well studied at the molecular level, particularly in plants (Anderson *et al.* 2010; Stahl & Bishop 2000). Similarly, studies investigating the interactions between toxic prey and their predators report co-evolution of toxicity and resistance (Becerra *et al.* 2009; Brodie Jr *et al.* 2002). A growing number of studies have investigated local adaptation of disease and toxin resistance in natural animal and plant populations at landscape scales (Brodie Jr *et al.* 2002; Cammen *et al.* 2011; Joost *et al.* 2007; Laine *et al.* 2011). As human activities alter the distributions of species, affecting host/pathogen and predator/prey interactions, it becomes increasingly important to understand how species have adapted to survive exposure to pathogens and toxins.

Harmful algal blooms (HABs) can be lethal to marine species and cause neurotoxic diseases in humans. These blooms appear to be increasing worldwide in frequency and geographic distribution, changes attributed partially to anthropogenic activities (Van Dolah 2000). In the Gulf of Mexico, HABs of the neurotoxin-producing dinoflagellate *Karenia brevis*, also known as red tides, occur almost annually along the coast of Florida. The brevetoxins produced by *K. brevis* bind with high affinity to voltage-gated sodium channels, resulting initially in neuron hyperexcitation and eventually anergy (Poli *et al.* 1986). As a result, brevetoxins cause neurotoxic shellfish poisoning in humans (Watkins *et al.* 2008) and have been associated with large-scale mortality of fish, sea birds, sea turtles, and marine mammals (Bossart *et al.* 1998; Flewelling *et al.* 2005; Kreuder *et al.* 2002; Walsh *et al.* 2010). Brevetoxins are ichthyotoxic but can bioaccumulate and biomagnify in some species (Naar *et al.* 2007), becoming a threat to upper trophic level predators, such as the common bottlenose dolphin (*Tursiops truncatus*).

Red tides have been reported for centuries along the coast of central-west Florida (Steidinger 2009), and this long history of blooms combined with the effects of exposure may have produced a selective pressure strong enough to induce evolved resistance, even in a marine mammal with long generation times. Populations of bottlenose dolphins may differ in their resistance to red tides, potentially due to the frequency of historical exposure (Twiner *et al.* 2012). HABs of *K. brevis* have been associated with several large-scale mortality events of bottlenose dolphins along the Gulf coast of Florida, particularly in the Florida Panhandle (Twiner *et al.* 2012), but red tides in central-west Florida often occur without associated dolphin mortality, and low levels of brevetoxins have been detected in live dolphins that do not express any symptoms of brevetoxicosis (Fire *et al.* 2008b).

Traditionally, molecular studies of disease resistance in natural populations have been limited to a few candidate genes and/or tens of neutral markers (e.g., microsatellites). Human genome-wide association studies compare hundreds of single nucleotide polymorphisms (SNPs) between cases and controls, but require a database of known SNPs (Hirschhorn & Daly 2005). Advances in sequencing methods now permit investigation of variation in thousands of SNPs across genomes of species without previously developed genomic resources. The application of such next-generation sequencing to non-model species and natural populations is growing rapidly with the development of genotyping-by-sequencing techniques. These techniques, which sequence only a portion of the genome, can be used to simultaneously identify and genotype thousands of SNPs (Davey *et al.* 2011). A growing number of studies have used these techniques to evaluate population structure, phylogenetic relationships, and local adaptation (Narum *et al.* 2013).

In this chapter, I extended the application of these genotyping-by-sequencing techniques to a study of disease resistance in natural populations. I searched for evidence of a genetic basis for brevetoxin resistance in bottlenose dolphins in the Gulf of Mexico using restriction site-associated DNA (RAD) sequencing. Dolphins that died during unusual mortality events (UMEs) due to HAB exposure were compared to dolphins from the same geographic areas that survived HAB events. I identified candidate genes that may be linked to brevetoxin resistance by searching the dolphin genome for immune, nervous, and detoxification genes adjacent to SNPs identified by RAD sequencing.

Methods

Samples

I conducted RAD sequencing on seven pools of bottlenose dolphin DNA samples defined by geographic sampling location, coastal or estuarine distribution, and survival during red tides (Fig. 2, Table 2). My study included a total of 156 samples from the following seven groups of bottlenose dolphins: dolphins that died during UMEs associated with red tides in the Florida Panhandle in 1999-2000 (N = 16) and 2004 (N = 35); live dolphins from the Florida Panhandle, including the estuarine population inhabiting St. Joseph Bay (N = 12) and the adjacent coastal population (N = 17); dolphins that died during red tides in central-west Florida between 1992 and 2003 (Fauquier *et al.* 2007; Fire *et al.* 2007) and during an unusually severe and prolonged red tide resulting in a UME in 2005-2006 (N = 25); and live dolphins from central-west Florida, including the estuarine population inhabiting Sarasota Bay (N = 25) and the adjacent coastal population (N = 26). All live dolphins were assumed to have survived at least one red tide based on their age and geographic distribution. Samples were categorized as coastal or estuarine based on photo-ID site fidelity (pers. comm. B. Balmer & R. Wells). It is not possible to determine whether dead dolphins originated from coastal or estuarine populations; I suspect a mix of the two populations in the dead dolphin sample pools.

Table 2: Sample pools representing coastal and estuarine dolphin populations and multiple UMEs associated with red tides from the Florida Panhandle and central-west Florida. Number of reads (following quality filtering) and loci (with >15 reads per locus) generated by RAD sequencing of these pooled bottlenose dolphin samples.

Sample Pool	Collection Years	DNA Quality	N	Reads	Loci	Read Depth (avg ± sd)
Florida Panhandle						
Live coastal	2004-2005	Excellent	9	14,880,205	118,299	103.35 ± 64.54
Live coastal	2004-2005	Good	8	13,889,991	116,659	98.46 ± 61.85
Live estuarine[a]	2004-2006	Excellent	12	16,594,432	118,030	114.08 ± 73.78
1999 UME	1999	Good	16	9,236,187	110,902	59.37 ± 39.17
2004 UME	2004	Good	35	13,896,773	119,743	80.16 ± 54.99
Central-west Florida						
Live coastal	1999-2000	Good	26	3,073,044	62,391	34.79 ± 25.96
Live estuarine[b]	2009-2011	Excellent	25	17,269,321	119,823	112.4 ± 73.46
2005-2006 UME[c]	1992-2006	Excellent	13	15,101,244	119,454	100.60 ± 63.76
2005-2006 UME[c]	1992-2006	Good	12	15,386,415	120,991	97.05 ± 61.48
		Total	**156**	**119,327,612**	**129,594**	**88.69 ± 63.45**

[a]St Joseph Bay; [b]Sarasota Bay; [c]includes 9 dolphins that stranded during HABs between 1992-2003.

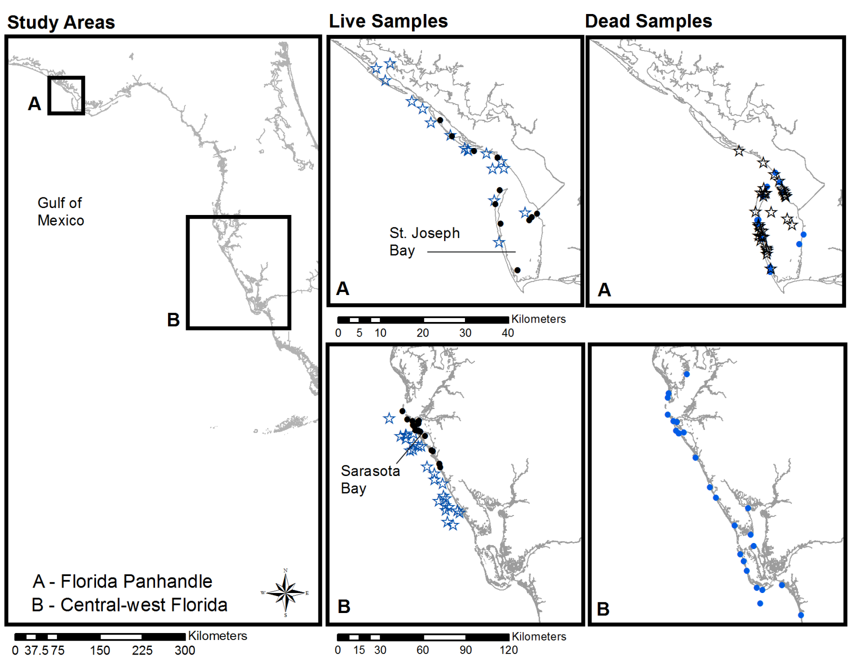

Figure 2: Map of study areas in the Florida Panhandle (A) and central-west Florida (B) indicating the location of sample collection for live coastal (blue stars) and estuarine (black circles) bottlenose dolphins and dolphin strandings during UMEs in the Panhandle in 1999 (blue circles) and 2004 (black stars) and during HABs between 1992 and 2006, including a UME in 2005-2006, in central-west Florida (blue circles).

Dolphin skin and genomic DNA samples were acquired from the Chicago Zoological Society's Sarasota Dolphin Research Program, the Mote Marine Laboratory Stranding Investigations Program, and the NOAA Southeast Fisheries Science Center (SEFSC) Marine Mammal Tissue and DNA Archive. I extracted DNA from 29 skin samples, representing primarily the Sarasota Bay population, using a Wizard® Genomic DNA Purification Kit (Promega) following manufacturer instructions. Genomic DNA provided by the NOAA SEFSC Marine Mammal Tissue and DNA Archive had been extracted following Rosel and Block (1996) using 250 µl volumes and Phase Lock gel tubes (5 Prime). The quality of genomic DNA was determined visually on agarose gels and quantified fluorometrically.

RAD library preparation

I prepared a RAD library as described in Etter *et al.* (2011) with the following minor modifications. Briefly, pools of 500 ng genomic DNA were digested with *SbfI-HF* (New England Biolabs). Pooling individuals prior to barcoding permits the use of less input DNA from each individual and increases the number of individuals that can be sequenced. Each pool was ligated to three P1 adaptors with unique 6 bp barcodes that differed by at least 2 bp from all other barcodes. Ligated DNA was sheared to an average size of 500 bp using a Bioruptor and size selected using AmpPure magnetic beads (Agencourt) (Lennon *et al.* 2010). The production PCR consisted of 18 cycles and product was purified using AmpPure beads. The RAD library was sequenced on a single lane of an Illumina HiSeq 50 bp single-end read run.

SNP identification and genotyping

I analyzed the resulting RAD sequences using *Stacks v0.99993* (Catchen *et al.* 2013). Reads were first demultiplexed and filtered for quality (minimum average Phred score of 10 across 6 bp sliding window; no errors in barcode or restriction site). Quality-filtered reads were then mapped to the bottlenose dolphin genome (Ttru_1.4/turTru2, GenBank Assembly ID: GCA_000151865.2) using Bowtie allowing up to three mismatches, but only a single best match to the genome. For each barcoded pool of samples, all reads that mapped to the same location were considered a "stack," also referred to as a RAD tag or locus. I only analyzed loci with greater than 15 reads per pooled sample.

SNPs were detected within each stack using a bounded maximum likelihood model (Catchen *et al.* 2013) adjusted to increase the likelihood of identifying a SNP at low minor allele frequency (lower bound = 0, upper bound = 0.001, alpha = 0.001). The model was capable of detecting a SNP when the minor allele was present in the pool at 10% or greater frequency. Previous analyses have confirmed that allele frequencies generated from pooled samples correlate highly with genotypes from direct sequencing (Van Tassell *et al.* 2008; Zavodna *et al.* 2013), and under certain conditions, sequencing pooled DNA allows for more accurate identification of SNPs and calculation of population allele frequencies (Futschik & Schlötterer 2010).

Live-dead comparisons

F_{ST} was estimated for pairwise comparisons of pools of live and dead dolphin samples assuming Hardy-Weinberg equilibrium in order to estimate heterozygosity from allele frequencies (Weir & Cockerham 1984). Loci were considered associated with survival if a Fisher's exact test for allele frequency difference was significant below an alpha threshold of 1×10^{-6}. I assessed similarities among live-dead comparisons by agglomerative hierarchical clustering using the complete linkage method (Shannon *et al.* 2003) and Euclidean distances (square root of sum of squared differences) between allele frequency differentials, implemented in R using the enhanced heat map function of the gplots package (v2.12.1).

I compared independent live-dead comparisons rather than combining pools (i.e., all live vs. all dead) due to known population differentiation and potential differences among UMEs. I conducted analyses both with and without loci with missing data. Missing data occurs in RAD sequencing datasets as a result of variable coverage due to restriction fragment length biases, restriction site heterozygosity, and PCR GC content bias (Davey *et al.* 2013).

Candidate gene identification

To identify potential candidate genes linked to the SNPs in my analysis, I aligned all RAD loci to the partially annotated Ensembl bottlenose dolphin genome (turTru1, Ensembl release 71, GenBank Assembly ID: GCA_000151865.1) and noted all genes within 100 kb of a RAD locus. The most recent bottlenose dolphin genome

(Ttru_1.4/turTru2) used to build “stacks” has not yet been annotated. Due to the fragmented nature of the current bottlenose dolphin genome assembly, the 200 kb search window around survival-associated RAD loci was often prematurely truncated and some candidate genes may be missing. Putative gene function was extrapolated from studies in humans and model systems, as reported in UniProt’s Protein Knowledgebase (Magrane & UniProt Consortium 2011). All genes were assigned gene ontology (GO) terms using g:profiler with Ensembl Genomes v16 (Reimand *et al.* 2011). I used existing knowledge of brevetoxins’ effects on the immune, nervous, and detoxification systems to narrow the list of candidate genes and identify those most likely to explain associations with survival observed in polymorphic RAD loci.

Results

RAD sequencing resulted in 119,327,612 quality sequences (NCBI Short Read Archive accession number SRP040141), of which the majority (87.12%) mapped to the dolphin genome at 129,594 locations, surrounding 70,354 *SbfI-HF* restriction enzyme recognition sites. 7,431 RAD loci contained a single bi-allelic polymorphic site with at least 10% minor allele frequency. I compared the allele frequencies of these polymorphic loci between pools of live and dead dolphins from the same region and observed significant differences in 5,107 (14.7%) comparisons (Table 3). The greatest number of differentiated loci was observed in the comparison of live coastal dolphins and dead dolphins from central-west Florida, but this was likely partially due to lower read coverage (average number of reads per locus) in this live sample (Table 2).

Table 3: Live-Dead Comparisons. Indicates the number of loci with significant differences in allele frequency between live and dead sample pools over the total number of loci sequenced at sufficient depth in both pooled samples. F_{ST} values represent the average (± SD) across all loci.

		FL Panhandle 1999 UME	FL Panhandle 2004 UME	Central-west FL 2005-06 UME*
FL Panhandle estuarine	Loci	636/5820	772/6286	
	F_{ST}	0.211 ± 0.259	0.188 ± 0.230	
FL Panhandle coastal	Loci	744/6053	1006/6551	
	F_{ST}	0.194 ± 0.236	0.174 ± 0.218	
Central-west FL estuarine	Loci			735/6668
	F_{ST}			0.131 ± 0.183
Central-west FL coastal	Loci			1215/3480
	F_{ST}			0.285 ± 0.272

*includes dolphins that stranded during HABs between 1992-2003.

I considered loci potentially linked to survival if they were consistently (e.g., one allele was always significantly more frequent in the live samples than the dead samples) differentiated in at least two of six comparisons between live and dead sample pools. This included 565 out of 2,962 polymorphic RAD loci that were sequenced to sufficient depth in all seven sample pools (i.e., loci with no missing data) (Fig. 3). Fifty-six loci were significant in the two comparisons within central-west Florida but no comparisons within the Florida Panhandle, suggesting regional differences in survival-associated genetic variation. For five of these loci (3693, 27718, 68597, 77572, 101552), the lack of difference between live and dead dolphins in the Florida Panhandle may be explained by little within-region variation and fixation or near fixation (90-100% frequency) of the allele associated with mortality in central-west Florida, which is consistent with the apparent overall susceptibility of dolphins in the Florida Panhandle. Six loci were significantly differentiated in only the four comparisons within the Florida Panhandle;

for one of these loci (2053), the allele associated with survival in the Florida Panhandle was nearly fixed in central-west Florida, consistent with the apparent overall resistance of dolphins in central-west Florida.

Additional loci were significant in only the two comparisons of the Florida Panhandle live samples to the 1999 UME (58 loci) or to the 2004 UME (61 loci). These numbers of loci that were significantly differentiated within only central-west Florida or only a single mortality event were greater than would be expected by chance (20.53 ± 4.32 SD loci), suggesting different mechanisms of toxicity among red tides and/or resistance among dolphin populations. No RAD locus was significantly differentiated in all six live-dead comparisons. However, of those loci with some missing data (i.e., sample pools that were not sequenced to sufficient depth), 36 RAD loci were significantly differentiated in all live-dead comparisons that could be evaluated.

To better understand the biological relevance of these survival-associated polymorphisms I mapped all RAD loci to the bottlenose dolphin genome and identified genes within 100 kb of each locus. Of the 7,039 genes found near a RAD locus (42.5% of all genes currently annotated in the dolphin genome), 78 genes were linked to a SNP with significant differences in allele frequency in all live-dead comparisons sequenced to sufficient depth. An additional 79, 4, 101, and 77 genes were linked to SNPs with significant differences in allele frequency in live-dead comparisons from only central-west Florida, the Florida Panhandle, the Panhandle 1999 UME, and the Panhandle 2004 UME, respectively (Appendix A).

Variation at either a single gene or a family of genes may result in resistance, so I analyzed these sets of candidate genes based both on individual function and gene ontology (GO). I found many candidate genes potentially related to brevetoxin exposure, including genes involved in T cell signaling, ion transport, and detoxification (Table 4). The group of 78 candidate genes surrounding RAD loci significantly differentiated in all comparisons overrepresented (hypergeometric probability < 0.01) several biologically relevant GO terms, including "cellular response to stress" (GO:0033554), "major histocompatibility class II protein complex" (GO:0042613), "DNA base-excision repair" (GO: 0006284), and "generation of neurons" (GO:0048699). Significant GO overrepresentation indicates that a larger number of genes with these ontologies were found in this group of 78 candidate genes than would be expected by chance given the presence of these GO terms in the 7,039 genes assessed by RAD sequencing.

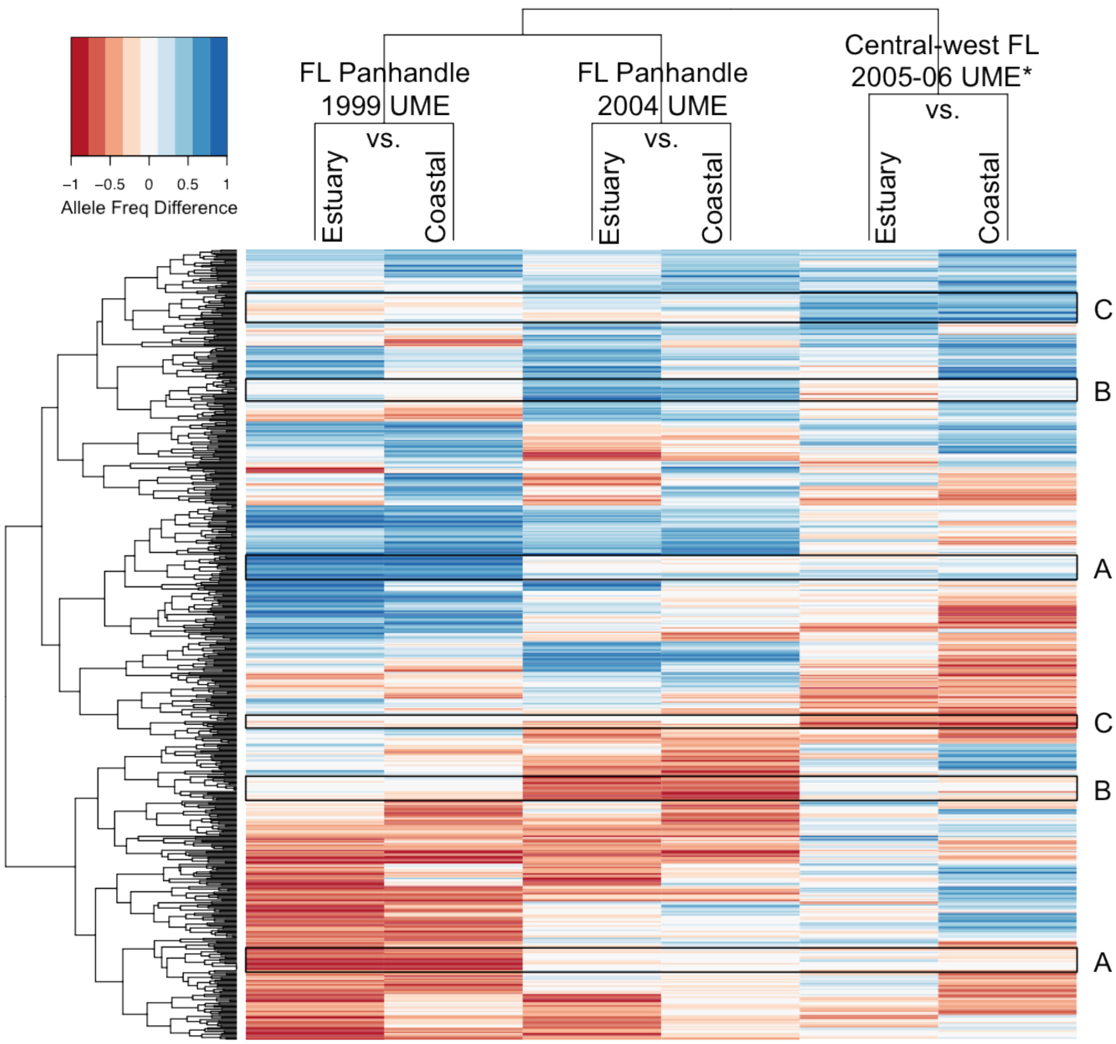

Figure 3: Heatmap indicating allele frequency differences between pools of live and dead dolphin samples at 565 polymorphic RAD loci sequenced in all sample pools and with significant differentiation in at least two live-dead comparisons. Phylograms show results of hierarchical clustering based on Euclidean distances between allele frequency differentials. Outlined regions highlight loci with significant differences in allele frequency in one region or UME: (A) FL Panhandle 1999 UME, (B) FL Panhandle 2004 UME, (C) Central-west FL.

Table 4: Candidate genes related to the immune, nervous, and detoxification systems. All listed genes were identified within 100 kb of a RAD locus that exhibited patterns of allele frequency difference between pools of live and dead dolphins consistent with a genetic association with red tide resistance in: all live-dead comparisons sequenced to sufficient read depth (*all*, included loci with missing data); all and only live-dead comparisons within a single geographic region (central-west Florida [*CWFL*] or Panhandle [*Pan*], excluded loci with missing data); all and only live-dead comparisons involving a single Panhandle UME (*1999 UME*, *2004 UME*, excluded loci with missing Panhandle data).

Gene Symbol and Name		Function	Comparisons
Immune System			
MHC-DRA	Major histocompatibility complex, class II, DR α	antigen presentation	all
MHC-DRB	Major histocompatibility complex, class II, DR β	antigen presentation	all
MHC-DQA	Major histocompatibility complex, class II, DQ α	antigen presentation	all
NFATC1	Nuclear factor of activated T-cells, cytoplasmic 1	cytokine induction in T-cells	all
PTPN7	Protein tyrosine phosphatase, non-receptor type 7	lymphocyte development	all
IGSF21	Immunoglobin superfamily, member 21	immune response receptor	CWFL
IL17REL	Interleukin 17 receptor E-like	inflammatory cytokine receptor	CWFL
CSF2RA	Colony-stimulating factor receptor subunit α	cytokine receptor	CWFL
SH2D4A	SH2 domain-containing protein 4A	T-cell development & function	CWFL
CKLF	Chemokine-like factor	inflammatory response	CWFL
PIM3	Serine/threonine-protein kinase pim-3	tumorigenesis	CWFL
PIM1	Serine/threonine-protein kinase pim-1	tumorigenesis	2004 UME
CD5	T-cell surface glycoprotein CD5	regulates T-cell proliferation	1999 UME
TAGAP	T-cell activation Rho GTPase-activating protein	T-cell activation	1999 UME
Immune & Nervous Systems			
NOTCH1	Neurogenic locus notch homolog protein 1	lymphocyte function; neuron differentiation	all
PTK2B	Protein-tyrosine kinase 2-beta	humoral immune response; regulation of Ca/K transport	1999 UME

Nervous System: Signal Transduction			
P2RX2	Purinergic receptor P2X type 2	ligand-gated ion channel; synaptic transmission	all
GALR3	Galanin receptor type 3	neuropeptide receptor; modulates action potentials	all, 2004 UME[1]
IL1RAPL1	Interleukin-1 receptor accessory protein-like 1	inhibits voltage-gated Ca channel	CWFL
KCNK13	Potassium channel subfamily K member 13	voltage-gated K channel	CWFL
SHC2	SHC transforming protein 2	signaling adapter	CWFL
SLC17A8	Vesicular glutamate transporter 3	neurotransmitter transporter	CWFL
GRIA2	Glutamate receptor 2	glutamate-gated ion channel; excitatory synaptic transmission	1999 UME
Nervous System: Development			
SCRIB	Protein scribble homolog	neuron morphogenesis	all
RACGAP1	Rac GTPase activating protein 1	neuron proliferation	all
FAIM2	Fas apoptotic inhibitory molecule 2	regulates neuronal apoptosis; cerebellar development	all
FARP2	FERM, RhoGEF & pleckstrin domain protein 2	nervous system development	all
SUN2	Sad1 and UNC84 domain containing 2	neuron migration	CWFL
MAPT	Microtubule-associated protein tau	neuronal polarity	CWFL
ATP2B2	Plasma membrane Ca-transporting ATPase 2	neuron differentiation	CWFL
PBX3	Pre-B-cell leukemia transcription factor 3	nervous system development	CWFL
SIM2	Single-minded homolog 2	nervous system development	CWFL
ODZ4	Teneurin-4	nervous system development	CWFL
Detoxification			
MGST1	Microsomal glutathione S-transferase 1	conjugation during detox.	all
GCLM	Glutamate-cysteine ligase regulatory subunit	gluatathione biosynthesis	Pan

[1]a single gene was identified in multiple groups if near multiple RAD loci.

Discussion

Survival-associated genetic variation

Using a genotyping-by-sequencing approach to analyze thousands of SNPs across the common bottlenose dolphin genome, I was able to identify genetic variation associated with survival in bottlenose dolphins exposed to HABs of neurotoxin-producing *K. brevis*. While a subset of SNPs showed significant associations with survival in dolphins from both central-west Florida and the Florida Panhandle, other SNPs were significant in only a single region. These results are consistent with both a common genetic basis for resistance to HABs in bottlenose dolphins across the Gulf coast of Florida and regionally specific resistance. I found many loci associated with survival in only central-west Florida, consistent with previous records of red tide occurrence and bottlenose dolphin mortality that suggest populations in this region are more resistant to red tides (Fire *et al.* 2007; Twiner *et al.* 2012), possibly due to more frequent historical exposure.

In addition to differences between geographic regions, I found differences in survival-associated genetic variation between mortality events within the Florida Panhandle region. These differences may reflect variation in the toxicity of HABs or variation in which dolphin populations were impacted. Although dolphins that stranded in 1999 and 2004 tested positive for similar levels of brevetoxin (Twiner *et al.* 2012), bloom dynamics differed between the two mortality events. Strandings occurred concurrently with a HAB in 1999, but elevated *K. brevis* concentrations were never recorded in 2004, suggesting exposure occurred via ingestion of prey that had been exposed to a HAB occurring elsewhere (Flewelling *et al.* 2005). For this reason among others, although dolphins were recovered with high brevetoxin concentrations in similar locations during both Florida Panhandle UMEs, the stranded dolphins may have originated from multiple genetically differentiated populations (Balmer *et al.* 2008; Sellas *et al.* 2005). These differences highlight the potential variability among HABs and dolphin mortality events that result in the complex picture of brevetoxicosis resistance described here. This complexity necessitated my broad genome-wide investigation of resistance and analysis of multiple live-dead comparisons rather than a single combined analysis of survivors and non-survivors.

Candidate genes

Most of our prior understanding of brevetoxin's effects comes from *in vitro* and model organism studies. From these studies, we know that brevetoxins are primarily neurotoxic, but can also have immunotoxic and genotoxic effects (Poli *et al.* 1986; Sayer *et al.* 2005). In the present study, I used a genome-wide association study to investigate adaptation to brevetoxin exposure in bottlenose dolphins. Inferring linkage with polymorphic RAD loci associated with survival, I identified candidate genes related to immune, nervous, and detoxification systems.

Immune system

The effects of brevetoxin exposure on the immune system have been described in naturally exposed marine animals (Bossart *et al.* 1998; Kreuder *et al.* 2002; Walsh *et al.*

2010; Walsh *et al.* 2005) and studied *in vivo* and *in vitro* in the laboratory. Brevetoxins appear to be both immunotoxic and an activator of inflammation (Murrell & Gibson 2009; Sayer *et al.* 2005). Immunotoxicity may result from the induction of DNA damage in lymphocytes; accordingly, I identified differences between live and dead dolphins at RAD loci in proximity with candidate genes involved in DNA base-excision repair (e.g., DNA directed polymerase epsilon and uracil-DNA glycosylase). Additionally, I identified candidate genes related to both inflammatory and T-cell mediated immune responses that are linked to survival-associated loci in bottlenose dolphins (Table 4). For example, three major histocompatibility complex (MHC) genes (DRA, DRB, and DQA) surround a single RAD locus (840) exhibiting nearly fixed differences between sequenced pools of live and dead dolphins (Panhandle live coastal vs. 2004 UME: $F_{ST} = 1$, $p < 1x10^{-27}$; Panhandle live coastal vs. 1999 UME: $F_{ST} = 1$; $p < 1x10^{-22}$; central-west Florida live estuarine vs. 2005-2006 UME: $F_{ST} = 0.92$, $p < 1x10^{-23}$). Variation at the MHC, which plays a critical role in antigen presentation to T cells, may moderate brevetoxin's immunotoxic effects (Chang *et al.* 2008). Variation at these MHC genes is further described in Chapter 3.

Nervous system

Brevetoxins are known neurotoxins that target the voltage-gated sodium channels in excitatory neurons (Poli *et al.* 1986). Four sodium channel genes were assessed (i.e., within 100 kb of a polymorphic RAD locus) in my RAD sequencing analysis, but none of these loci exhibited an association with survival. Instead, other genes involved in neuron signaling, including ion channels and neurotransmitter receptors (e.g., purinergic receptor P2RX2, galanin receptor GAL3, interleukin-1 receptor accessory protein IL1RAPL1, potassium channel KCNK123, glutamate transporter SLC17A8, and glutamate receptor GRIA2), were identified as potential candidates for evolved resistance to red tides in bottlenose dolphins (Table 4). These channels may affect sodium concentrations or neuronal depolarization indirectly and thereby moderate the effects of brevetoxin binding; however, further sequencing of these candidate genes in dolphins and additional studies of brevetoxin's cellular effects on the nervous system in model species are required to fully understand the role these candidate genes may play in brevetoxin resistance. In addition to neuron signaling, genes involved in neuronal development were overrepresented in the list of candidate genes derived from RAD sequencing (Table 4). Genes such as PBX3, SIM2, and ODZ4, which are involved in nervous system development, may play a previously undescribed role in brevetoxin resistance.

Detoxification system

Brevetoxin metabolism and detoxification occurs through the classic phase I and II stages, involving cytochrome P450 enzymes and glutathione S-transferases (Radwan & Ramsdell 2006; Radwan *et al.* 2005; Walsh *et al.* 2009; Washburn *et al.* 1996). My study identified two potential glutathione-related candidate genes (MGST1 and GCLM, Table 4). Previous studies have described elevated levels of microsomal GSTs (MGSTs) in aquatic species exposed to cyanobacterial toxins (Pflugmacher *et al.* 2005), but not brevetoxins (Washburn *et al.* 1994; Washburn *et al.* 1996). The brevetoxin detoxification pathway in dolphins may differ from those previously described in non-mammalian species. Alternatively, these glutathione genes may also play a role in the immune system, cytokine production, and oxidative stress (Fratelli *et al.* 2005). Previous studies have found evidence of brevetoxin-induced oxidative stress both *in vitro* (Walsh *et al.* 2009) and in loggerhead sea turtles exposed to red tides (Walsh *et al.* 2010), but this mechanism of brevetoxicity has not yet been described in bottlenose dolphins.

Conclusion

My study provides one of the first examples of using RAD sequencing to study toxin resistance in a natural population, and I see great potential for this approach in other studies of adaptation to disease and anthropogenic stressors. RAD sequencing simultaneously identifies and genotypes thousands of SNPs throughout a genome in multiple barcoded samples representing individuals or pools with a shared phenotype. Genetic variation associated with adaptation (e.g., red tide resistance) identified via RAD sequencing is rarely located within a coding region (McCarthy & Hirschhorn 2008). However, when an annotated reference genome is available, variation observed within a RAD locus can be linked to nearby genes regulating or encoding adaptation-related proteins. Here, I have described evidence for a genetic basis for resistance to HABs in bottlenose dolphins, and my characterization of candidate genes linked to survival-associated genetic variation provides both justification and direction for future studies of resistance to HAB toxins in marine mammals.

CHAPTER II: LACK OF VARIATION IN VOLTAGE-GATED SODIUM CHANNELS OF COMMON BOTTLENOSE DOLPHINS (*TURSIOPS TRUNCATUS*) EXPOSED TO NEUROTOXIC ALGAL BLOOMS

Introduction

Neurotoxin resistance has independently evolved in a variety of species via mutations in toxin binding sites within the voltage-gated sodium channel (SCN) gene complex. Voltage-gated sodium channels are membrane-spanning pores that play a key role in the generation of action potentials and neuronal depolarization (Yu & Catterall 2003). Neurotoxins that bind to sodium channels may block ion transfer, alter voltage-sensitivity, or prevent channel inactivation, all of which would alter neuron depolarization and the transmission of electric stimuli between neurons or across neuromuscular junctions (Anderson *et al.* 2005). However, a single amino acid change in an SCN gene can drastically change toxin binding affinity. Such mutations have been observed to induce neurotoxin resistance in a phylogenetically diverse group of organisms (Anderson 1987; Bricelj *et al.* 2005; Geffeney *et al.* 2005; Jost *et al.* 2008b). Neurotoxins, particularly tetrodotoxins, are a common mechanism of predator deterrence. In these systems, neurotoxin resistance is a necessity in the prey species that sequester the toxins and a likely outcome in predators that feed often on toxic prey. Additionally, in the context of human-induced evolution, many insects have evolved resistance to neurotoxic pesticides (Dong 2007).

There are numerous examples of neurotoxin-resistant species in marine environments in which harmful algae produce neurotoxins (Anderson *et al.* 2005). The most common of these toxins in US waters include domoic acid produced by *Pseudonitzschia* sp. primarily in the eastern Pacific Ocean, saxitoxin produced by *Alexandrium* sp. in the Gulf of Maine, and brevetoxin produced by *Karenia* sp. in the Gulf of Mexico (Van Dolah 2000). These toxins become a threat to marine organisms and humans that consume seafood, primarily shellfish, when the harmful algae occur in dense concentrations (>$1x10^3$ cells/L). Such harmful algal blooms (HABs) have been naturally occurring for centuries in some areas, but they are increasing in frequency and geographic range in US waters and worldwide (Van Dolah 2000). Some of this observed increase in HABs may be attributable to improved surveillance of coastal waters, but in addition, climate change, nutrient run-off, and ship ballast water transfer have likely contributed to a true increase in HABs, which may strengthen selective pressure for the evolution of neurotoxin resistance in marine species.

HABs, also known as red tides, of the dinoflagellate *Karenia brevis* occur almost annually along the west coast of Florida (Florida Fish and Wildlife Research Institute 2008; Steidinger 2009). *K. brevis* produces brevetoxins, a group of lipid-soluble compounds with a cyclic polyether backbone, referred to as PbTx from a previous taxonomic classification of the source dinoflagellate, *Ptychodiscus brevis* (Baden 1989). Brevetoxins are potent neurotoxins that bind with high affinity to voltage-gated sodium channels, causing dose-dependent depolarization and hyper-excitation (Poli *et al.* 1986).

The long ladder-like brevetoxin molecules span the sodium channel, affecting both the extracellular voltage sensor and the intracellular inactivation gate (Gawley *et al.* 1995). This results in inappropriate opening of the sodium channel and inhibition of channel inactivation. Neurons are first hyper-stimulated due to the persistent activation, but then exhibit a decrease in action potential and eventually become anergic (Huang *et al.* 1984).

In humans, inhalation of aerosolized brevetoxins can cause respiratory irritation (Backer *et al.* 2003), and ingestion of seafood containing brevetoxins can result in neurotoxic shellfish poisoning (NSP) (Watkins *et al.* 2008). The most common neurological symptom of NSP is paresthesias, or numbness and tingling (Watkins *et al.* 2008). In marine species, which experience greater exposure to *K. brevis*, brevetoxins can be lethal. HABs of *K. brevis* are often associated with large fish kills and mortalities of bottlenose dolphins, manatees, sea birds, and sea turtles (Bossart *et al.* 1998; Redlow *et al.* 2003; Van Dolah 2000). Neurological symptoms observed in affected manatees include disorientation, listlessness, and inability to submerge or maintain a horizontal position (O'Shea *et al.* 1991). Organisms may be exposed to brevetoxins directly in the water column, through inhalation of aerosolized toxin, or through the ingestion of *K. brevis* cells or contaminated prey (Bossart *et al.* 1998; Flewelling *et al.* 2005; Van Dolah 2000). Brevetoxins can be potently ichthyotoxic, but certain fish are able to survive high concentrations of *K. brevis* and accumulate brevetoxins (Naar *et al.* 2007). As brevetoxins biomagnify they become a threat for upper trophic level predators in these coastal ecosystems, such as the common bottlenose dolphin (*Tursiops truncatus*).

In the past two decades, four large-scale unusual mortality events (UMEs) of bottlenose dolphins have been attributed to blooms of *K. brevis*. Each event involved the mortality of over 100 dolphins that stranded dead, often during or shortly after a red tide was observed in the area. Of those tested, many stranded dolphins were positive for brevetoxins at high concentrations in their stomach contents and livers (Twiner *et al.* 2012). Three of these mortality events occurred in the Florida Panhandle and only one occurred in central-west Florida (NOAA & FWC, unpublished data, cited in Twiner *et al.*, 2012). Central-west Florida experiences HABs on an almost annual basis, but dolphin populations in this region, such as the well-studied population in Sarasota Bay, typically suffer little toxin-related mortality (Fire *et al.* 2007). The one mortality event that occurred in this region was associated with an unusually severe and prolonged HAB (Fauquier *et al.* 2013).

The factors resulting in large-scale dolphin mortality associated with red tides are not well understood, particularly in regards to the seemingly different impacts of red tides in geographically disjunct dolphin populations. Bloom densities and toxicities are similar in central-west Florida and the Panhandle (Twiner *et al.* 2012), and, as summarized in Table 1 in the Introduction, previous studies have found similar brevetoxin levels in prey fish and dolphin stomach contents and tissue samples in the two regions. These findings suggest that dolphins in both regions may be similarly exposed to brevetoxins during red tides.

The differences observed in HAB-associated dolphin mortality between the Florida Panhandle and central-west Florida may be partially explained by a difference in historical HAB frequency. HABs have occurred on a regular basis for centuries in central-west Florida (Florida Fish and Wildlife Research Institute 2008; Steidinger 2009), but have been comparatively infrequent in the Florida Panhandle (pers. comm. L. Flewelling; Gaydos, 2006). As a result, dolphin populations in central-west Florida have been exposed to a consistently strong selective pressure for HAB resistance, while dolphins in the Florida Panhandle may be relatively naïve and therefore more susceptible to HABs (Van Dolah *et al.* 2003). This difference in selection could have resulted in local adaptive divergence between the populations, more strongly favoring genotypes that confer HAB resistance in the central-west Florida dolphin populations. Similar local adaptation to neurotoxins is seen in the cases of soft shell clam resistance to algal-produced saxitoxins (Connell *et al.* 2006) and garter snake resistance to toxic newts (Brodie III & Brodie Jr 1991). In both cases, resistance is the result of a mutation at the toxin binding site in the voltage-gated sodium channels. In this chapter, I test the hypothesis that genetic variation in the voltage-gated sodium channels of bottlenose dolphins explains differences among individuals or populations in resistance to brevetoxins produced by HABs of *K. brevis*.

Brevetoxins bind to site 5 on the alpha subunit of voltage-gated sodium channels (SCNA) (Cestele & Catterall 2000), which are composed of four homologous domains, each containing six transmembrane segments (Yu & Catterall 2003). The brevetoxin binding site is located on transmembrane segments in two separate domains (DIS6 and DIVS5) encoded by two separate SCNA exons (Widmark *et al.* 2011). I investigated genetic variation at these two exons in nine structurally different sodium channel isoforms, which exhibit unique patterns of tissue expression (Catterall 2000). Some studies have shown tissue selectivity in brevetoxin binding (Bottein Dechraoui & Ramsdell 2003; Bottein Dechraoui *et al.* 2006), but others have suggested that the evolution of resistance to neurotoxins takes place across all types of voltage-gated sodium channels (Jost *et al.* 2008b). My objective was to characterize genetic variation at all putative brevetoxin binding sites in bottlenose dolphins and determine if variation at these sites can explain apparent differences among individuals or populations in resistance to brevetoxins.

Methods

Samples

My study included DNA samples from the following groups of bottlenose dolphins: live dolphins from the Florida Panhandle, including the estuarine population found primarily in St. Joseph Bay (N=18) and the adjacent coastal population (N=19); live dolphins from central-west Florida, including the estuarine population found in Sarasota Bay (N=27) and the adjacent coastal population (N=28); dolphins that died during UMEs associated with red tides in the Florida Panhandle in 1999 (N=18) and 2004 (N=37); and dolphins that died during red tides in central-west Florida between 1992 and 2006 (N=27, including 18 dolphins that died during the 2005-2006 UME). All live dolphins were assumed to have survived at least one red tide based on their age and geographic distribution, and samples were categorized as coastal or estuarine based on photo-ID site fidelity (pers. comm. B. Balmer & R. Wells). Dolphins that died during non-UME red tides were included if toxin analyses suggested brevetoxicosis as the putative cause of death (Fauquier *et al.* 2007; Fire *et al.* 2007). I acquired dolphin skin and genomic DNA samples from the Chicago Zoological Society's Sarasota Dolphin Research Program, the Mote Marine Laboratory's Stranding Investigations Program, and the NOAA SEFSC Marine Mammal Tissue and DNA Archive.

I also included eight skin samples from coastal bottlenose dolphins that stranded in North Carolina between 2002 and 2006 to represent an outgroup that has rarely, if ever, been exposed to red tides. These samples were acquired from the North Carolina Marine Mammal Stranding Network.

DNA was extracted from skin samples using the Wizard® Genomic DNA Purification Kit (Promega) or phenol-chloroform extraction (Rosel & Block 1996) using 250 μl volumes and Phase Lock gel tubes (5 Prime). Genetic variation was initially analyzed in a sample set of 70 dolphins, including 8-10 samples from each of the eight groups. If non-synonymous mutations were detected in this initial sample set, genetic variation was further analyzed in the total sample set of 182 individuals.

Genetic analyses

Primers were designed to amplify the two brevetoxin binding sites on nine voltage-gated sodium channel isoforms identified in the annotated Ensembl bottlenose dolphin genome (turTru1, release 71, GenBank Assembly ID: GCA_000151865.1) (Table 5). The brevetoxin binding sites are located within domain I transmembrane spanning region 6 (DIS6) and domain IV transmembrane spanning region 5 (DIVS5) (Widmark *et al.* 2011). Primers were designed to amplify the complete exon (198-210 bp) containing DIS6 and a 171-186 bp fragment encoding DIVS4 and DIVS5, which are found on an exon that is too long (1217-3824 bp) to sequence in its entirety. In general, SCNA is encoded by 26 exons; however, the number of exons differs slightly between isoforms and the presence of nonconventional introns can confuse traditional algorithms for exon identification (Widmark *et al.* 2011). Here, I use the exon numbers as reported in NCBI AceView (Thierry-Mieg & Thierry-Mieg 2006) and the following literature (Chen *et al.* 2000; Dib-Hajj *et al.* 1999; Escayg *et al.* 2000; George Jr *et al.* 1993; Klugbauer *et al.* 1995; Plummer *et al.* 1998; Rabert *et al.* 1998; Wang *et al.* 1996). DIS6 is located on exons 9, 10, or 11; DIVS5 is located on exons 24, 26, 27, or 28.

Table 5: Primers used to amplify the putative brevetoxin binding sites (DIS6 and DIVS5) on bottlenose dolphin voltage-gated sodium channel alpha subunits (SCNA). The first set of primers amplify the DIS6-containing exon (9, 10, or 11); the second set of primers amplify the portion of the exon (24, 26, 27, or 28) that spans DIVS4 and DIVS5.

	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product Size (bp)
Exon containing DIS6			
SCN1A_exon9	AAAAATAATTCTGGTGTTCTTATTTTCA	TTTGGTGAATTTACCTGAGCTG	250
SCN2A_exon9	AAATATCTCTTCCATTTTTGCAGAC	GGCTCGAAGTTTGCTACTATACC	252
SCN3A_exon11	CCTAGACGTTACGTGCTGCT	AAGCAGCTTTGCTTATTATCACTTA	241
SCN4A_exon9	CCGCTCCCAGACCCTTC	CTCCTCTCCCTGAGTCCAGA	241
SCN5A_exon10	GCTGGGCAACTTGGGATTTA	AGGTGAAGGCTTGTTCTGGA	323
SCN8A_exon10	TCAGCCTGATTCGTTTGATG	AATCAGGTGGGGCCTTAGTC	327
SCN9A_exon10	TCAAGTTGAAATGATTTACCGAAA	CAGAATTGCTTCCTTTGACCA	448
SCN10A_exon9	TCTCTCTTCTTTCTTTTGAATCTGC	CACAAAGATCCACCCCTCAC	250
SCN11A_exon9	GTGCTGTGTCCCTTGCAGAC	CCCAAGCAGACCCCACT	255
Exon spanning DIVS4-S5			
SCN1A_exon26	TCTTTTATCCATCCATAGGTATGTTTC	TCAGGTGGTTTACTGTTGAGAA	385
SCN2A_exon26	CCATAGGTATGTTTCTGGCCG	AGTCTGGAGGTCCACTATTAAGAA	375
SCN3A_exon28	TTCTGTATTGTTCCATAGGTATGTTTC	CAATTCCAGCTTCCTTTTTAACAT	271
SCN4A_exon24	ATCCAGAAATACTTCGTGTCACC	GTCGTCGATGCCCGACT	237
SCN5A_exon28	CGGGAGCCATTGCTGTT	GCACTGTGCTCTCAGACATCAT	403
SCN8A_exon27	CTTAGGAATGTTCCTGGCTGA	GGTTTAGGATGGGCAGCA	359
SCN9A_exon27	GCAGATTTGATAGAAAGGTATTTTG	CAACTGAGCTTCCAGGATGA	394
SCN10A_exon27	GGTTCCCACAGTCCCCTCTA	CCTGGTGTTTTCTGCAATTCT	425
SCN11A_exon26	CCTTGAAAACCCAGGCACAC	GTGGTTATCTGGAAGAGGCAGA	304

All PCRs were conducted in a final volume of 20 μl and included 12-50 ng DNA, 1x PCR buffer (20 mM Tris pH 8.8, 50 mM KCl, 0.1% Triton X-100, 0.2 mg/mL BSA NEB purified), 2-4 mM $MgCl_2$, 0.2 mM dNTPs, 0.25 μM of each primer, and 1 unit Taq (Bioline). Only PCRs for SCN1A included 4 mM $MgCl_2$. All PCRs were run using a touchdown temperature profile consisting of an initial step of 94°C for 3 min, followed by 30 cycles of 94°C for 15 sec, 70°C for 15 sec decreasing by 0.5°C each cycle, and 72°C for 30 sec, followed by 25 cycles with a 55°C annealing temperature, and a final step of 72°C for 5 min. 10 μl of all successful PCR products were enzymatically purified using 1 unit Exonuclease I and 0.25 units Antarctic Phosphatase (New England Biolabs) incubated at 37°C for 1 hr. Purified PCR products were sequenced in both directions. BigDye sequencing reactions were conducted in a final volume of 10 μl including 1-5.2 μl purified PCR product, 1 μM primer, 1x BigDye Buffer (40 mM Tris pH 9, 5 mM $MgCl_2$, 1.5 M Betaine), 0.44 μl BDX64 (MCLAB), and 0.06 μl BigDye v1.1 (Duke IGSP). The sequencing profile consisted of an initial step of 94°C for 30 sec, followed by 25 cycles of 94°C for 15 sec, 50°C for 15 sec, and 60°C for 3 min. Sequencing products were purified using AmpPure magnetic beads (Agencourt) and sequenced on an Applied Biosystems ABI 3730xl automated sequencer.

I manually edited and aligned sequences in CodonCode Aligner v3 (CodonCode Corporation). Polymorphisms were identified as two chromatogram peaks in a single position, clearly observed in both forward and reverse sequences. All unique alleles that were not observed in multiple homozygous individuals were confirmed by sequencing

16-24 clones produced using pGEM®-T Easy Vectors with JM109 competent *E. coli* cells (Promega) following manufacturer instructions. Alleles were confirmed only if observed in multiple clones.

I conducted phylogenetic analyses in MEGA v5.2 (Tamura *et al.* 2011) using the maximum likelihood method with the evolutionary model that best described the nucleotide substitution pattern (maximum likelihood fit with lowest Bayesian information criterion). I used the Kimura 2-parameter evolutionary model (Kimura 1980), assuming that 40.9% of sites are evolutionarily invariable, to infer evolutionary history and conducted 1,000 bootstrap replicates (Felsenstein 1985). The tree was generated from 393 bp of combined nucleotide sequences from the DIS6-containing exon (9, 10, or 11) and the portion of the exon (24, 26, 27, or 28) spanning DIVS4-S5. For variable isoforms, I included the most common variant in the phylogenetic analysis.

I tested for significant effects of geographic region (central-west Florida vs. Panhandle) and survival on SCNA allele frequency using multifactorial logistic models implemented in JMP Pro 11.0.0 (SAS Institute Inc. 2014). I also assessed the power of this study to detect an association between genetic variation and survival using a case/control power analysis conducted in PGA (Menashe *et al.* 2008). Specifically, I calculated the minimum minor allele frequency, given the total sample size (94 controls, 83 cases), that would provide 80% power to detect a relative risk of 2 if the prevalance of disease is between 0.01 and 0.20. Relative risk refers to the ratio of the probability of a dolphin with a “susceptible” allele dying due to brevetoxin exposure as compared to the probability of a dolphin dying that does not have the “susceptible” allele. Disease prevalance describes the proportion of the population that experiences brevetoxicosis and was estimated as a range of values because of the current lack of understanding of brevetoxicosis disease etiology and prevalence in the bottlenose dolphin populations.

Results and Discussion

Sodium channel isoforms

I characterized genetic variation of the putative brevetoxin binding site in nine sodium channel isoforms in bottlenose dolphins and found that variation among isoforms conformed to expectations based on the general history of SCN gene duplication (Catterall *et al.* 2005). There was no evidence to suggest any of the sequences represented pseudogenes (i.e., no amino acid translations contained stop codons). Historical gene duplications in tetrapods resulted in four groups of SCNs: 1, 2, 3, and 9; 5, 10, and 11; 4; and 8 (Widmark *et al.* 2011). This grouping is generally reflected in the maximum likelihood tree built using only 393 bp of each bottlenose dolphin SCNA isoform (Fig. 4). Each cluster of SCNs is found on a separate chromosome across all mammalian genomes previously investigated (Widmark *et al.* 2011). I cannot confirm this chromosomal arrangement in bottlenose dolphins due to the fragmented nature of the currently available bottlenose dolphin genome, but as the bottlenose dolphin genome assembly is further improved, I expect to see the same chromosomal arrangement of SCN genes.

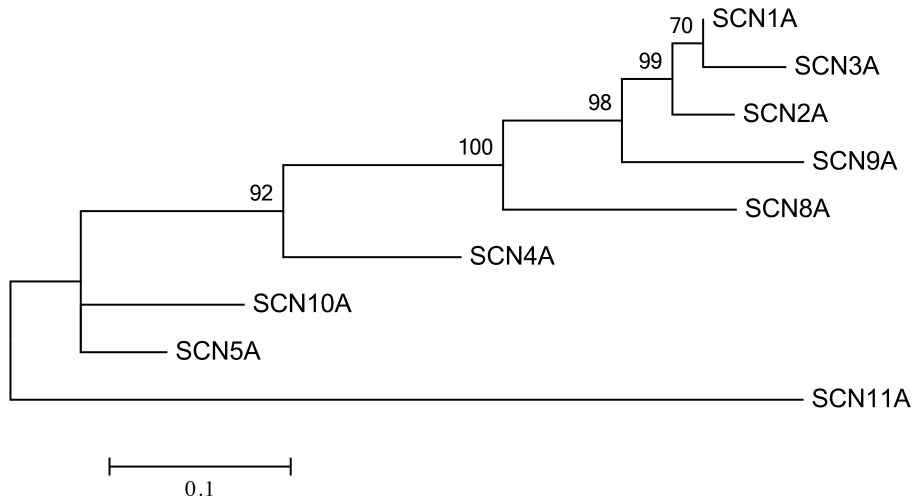

Figure 4: Maximum likelihood tree of bottlenose dolphin voltage-gated sodium channel genes generated from the combined nucleotide sequences of the DIS6-containing exon (9, 10, or 11) and the portion of the exon (24, 26, 27, or 28) spanning DIVS4-S5. For variant isoforms, the most common variant is displayed. Bootstrap values (percentage of 1,000 replicates) are displayed next to branches. Branch lengths represent number of substitutions per site from a total of 393 bp.

Sodium channel variation

Resistance-inducing mutations in the voltage-gated sodium channels are expected to be non-synonymous mutations that result in an amino acid residue substitution and corresponding change in protein structure. Of the 18 SCNA exons sequenced, only two (SCN5A_exon10 and SCN11A_exon26, Table 6) contained a non-synonymous mutation in the initially screened sample set of 70 dolphins, including 8-10 individuals representing live estuarine and coastal dolphins from the Florida Panhandle and central-west Florida, dolphins that died during UMEs associated with red tides in Florida, and a non-exposed outgroup of stranded dolphins from North Carolina. SCN5 and SCN11 belong to the same SCN cluster (Widmark *et al.* 2011) and often express some level of tetrodotoxin resistance (Catterall *et al.* 2005). Brevetoxins have been experimentally demonstrated to target the cardiac sodium channels (SCN5); however, type B brevetoxins (PbTx-2 and -3), which are the most common toxin type observed during red tides (Baden 1989), show lower relative affinity for heart sodium channels in comparison to skeletal (SCN4) sodium channels (Bottein Dechraoui & Ramsdell 2003; Bottein Dechraoui *et al.* 2006). Studies have not yet tested brevetoxin binding affinity to SCN11, which is found primarily in dorsal root ganglion neurons (Catterall *et al.* 2005).

Neither non-synonymous mutation occurred in a transmembrane spanning region where brevetoxin putatively binds (DIS6 or DIVS5). The non-synonymous mutation in SCN5A occurred in the intracellular loop between domains I and II, and the non-synonymous mutation in SCN11A occurred in DIVS4 (Fig. 5). The fourth transmembrane segment of each domain regulates sodium channel voltage-sensitivity, but this voltage-sensitivity is primarily attributed to positive amino acids, lysine and arginine (Widmark *et al.* 2011), neither of which were affected by the observed non-synonymous mutation in DIVS4 that encoded a change in amino acid from phenylalanine to leucine.

Five other SCNA exons contained synonymous mutations (Table 6, Fig. 5). The most frequently variable genes, SCN1A and SCN2A, belong to the same SCN cluster (Widmark *et al.* 2011) and are broadly expressed in neurons and highly tetrodotoxin-sensitive (Catterall *et al.* 2005). The synonymous mutations found in these SCNA exons, or others, may be linked to protein-changing mutations elsewhere in the sodium channel genes, beyond the putative brevetoxin binding site evaluated in my study.

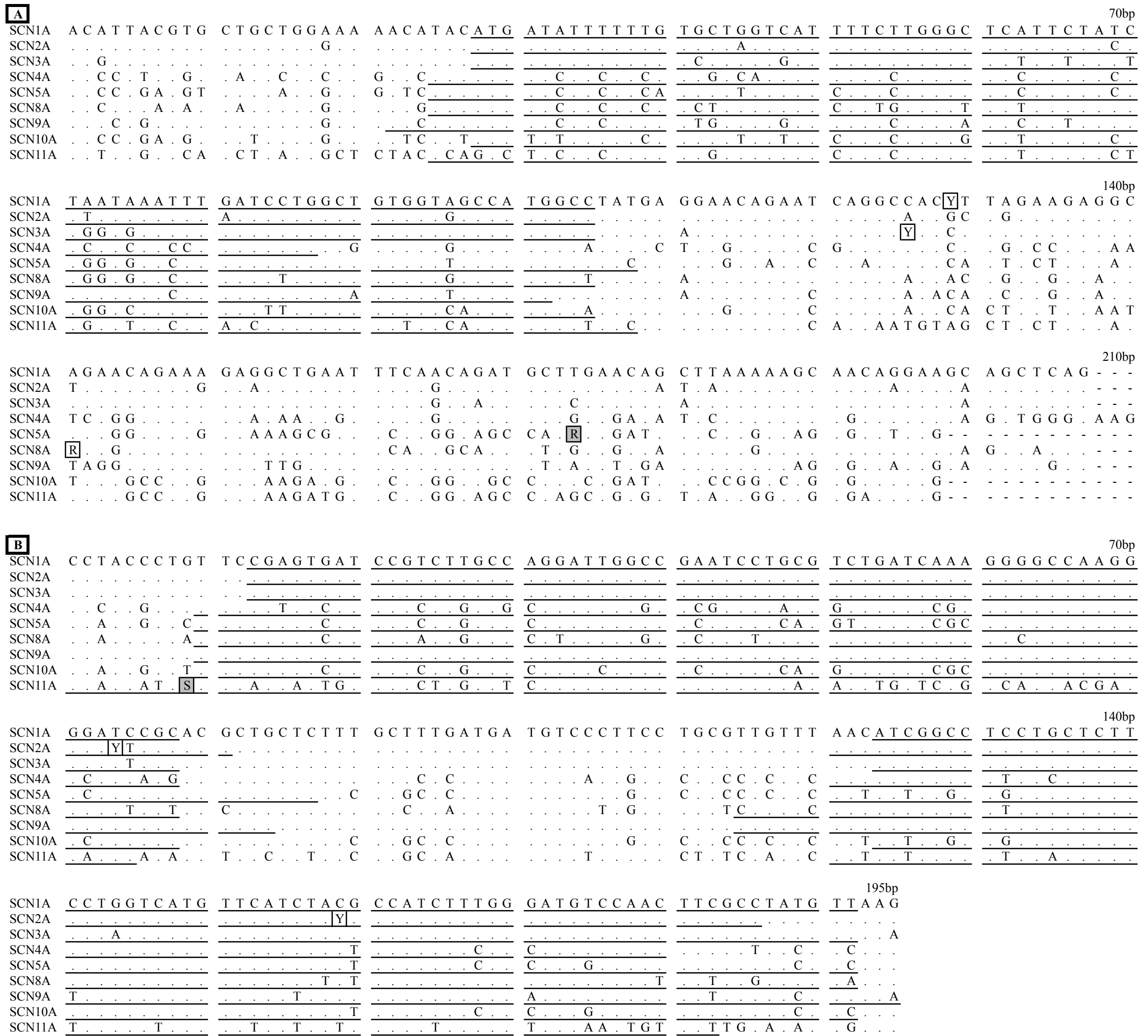

Figure 5: Alignment of bottlenose dolphin voltage-gated sodium channel nucleotide sequences from [A] the exon (9, 10, or 11) containing DIS6 (GenBank accession nos. KJ621083- KJ621095) and [B] the portion of exon (24, 26, 27, or 28) spanning DIVS4 and DIVS5 (GenBank accession nos. KJ621096- KJ621108). Dots indicate nucleotide identity with SCN1A; dashes represent missing sequence. Transmembrane segments are underlined. Polymorphic positions are represented by a degenerate base outlined by a square; non-synonymous polymorphisms are outlined and shaded in gray. Reading frame begins at position 1 for both sets of sequences. Note that one additional synonymous polymorphism (VGSC5s2, Y at position 278) is not depicted here because it is found in exon 28 beyond DIVS5.

Table 6: Summary of variation observed at the bottlenose dolphin voltage-gated sodium channel exons spanning the putative brevetoxin binding sites (DIS6 and DIVS5). All exons with polymorphic sites were tested for differences in allele frequency between regions (central-west Florida and the Florida Panhandle) and survival (dolphins that died during and survived following HABs); P values (bold if significant) from likelihood ratio chi-square tests are reported for the individual terms and an interaction term.

		# of Mutations		# of	Minor	Likelihood Ratio Chi-Square Tests		
	N	Syn.	Non-syn.	Alleles	Allele Freq.	Region	Survival	Reg.*Surv.
Exon containing DIS6								
SCN1A_exon9	70	1	0	2	0.18	0.069	0.958	0.554
SCN2A_exon9	70	0	0	1				
SCN3A_exon11	70	1	0	2	0.01	0.999	0.999	0.999
SCN4A_exon9	70	0	0	1				
SCN5A_exon10	182	0	1	2	0.01	**0.001**	1.000	1.000
SCN8A_exon10	70	1	0	2	0.01	0.998	0.998	0.998
SCN9A_exon10	70	0	0	1				
SCN10A_exon9	70	0	0	1				
SCN11A_exon9	70	0	0	1				
Exon spanning DIVS4-S5								
SCN1A_exon26	70	0	0	1				
SCN2A_exon26	70	2	0	3	0.10/0.19	**0.001**	0.234	0.535
SCN3A_exon28	70	0	0	1				
SCN4A_exon24	70	0	0	1				
SCN5A_exon28	70	1	0	2	0.05	**0.010**	1.000	1.000
SCN8A_exon27	70	0	0	1				
SCN9A_exon27	70	0	0	1				
SCN10A_exon27	70	0	0	1				
SCN11A_exon26	182	0	1	2	0.03	0.924	0.949	0.898

Effects of geographic region and survival

Genetic variation was rare among the sampled bottlenose dolphins, even in SCNA exons with polymorphisms (Table 6). Two minor alleles (SCN3A_exon11 and SCN8A_exon10) were only observed once in the initial sample set of Florida dolphins, and thus can have little effect on the observed differences in survival among individuals. Given the near parity of the number of live (34) and dead (28) Florida samples analyzed, I would expect to find alleles strongly associated with survival at closer to 50% frequency. A case/control power analysis suggests this study has 80% power to detect a relative risk of 2 with marker allele frequencies of 0.09-0.18. Thus, for at least SCN1A_ex9 and SCN2A_ex26, the most frequently variable SCNA exons, there should be an 80% probability of detecting an association with survival if there is one. However, I observed no significant effects of survival or the region:survival interaction term on SCNA allele frequency (Table 6).

Allele frequencies varied significantly between geographic regions (central-west Florida vs. Panhandle) in three cases (SCN5A_exon10, $\chi^2 = 12.53$, $p < 0.001$; SCN2A_exon26, $\chi^2 = 13.41$, $p < 0.01$; SCN5A_exon28, $\chi^2 = 6.69$, $p < 0.01$, Table 6). In both SCN5A exons the difference between regions was due to rare private alleles observed only in the Panhandle, but this difference could not be attributed to linked variation between exons as the private alleles were not observed in the same individuals. The difference between regions that I observed in rare SCNA alleles is more likely a reflection of limited gene flow between geographic regions than the cause of different levels of HAB resistance between regions. Other bottlenose dolphin populations elsewhere in the Gulf of Mexico exhibit significant fine-scale genetic differentiation (Sellas *et al.* 2005).

The lack of genetic variation in bottlenose dolphin voltage-gated sodium channels within the Gulf of Mexico could be explained by widespread evolved resistance to HABs. If this were the case, I would expect to see differences between the Florida samples and the non-exposed outgroup from North Carolina. No exonic variation was observed uniquely in North Carolina (although I observed a mutation in the intronic region upstream of SCN9A_exon10 only in North Carolina samples). SCNA minor alleles were almost never observed in the samples from North Carolina (except SCN8A_exon10 minor allele that was observed in a single North Carolina sample). Overall, my data do not support widespread SCN-based genetic resistance to red tides in bottlenose dolphins along the Gulf coast of Florida, although I acknowledge that my sample size from North Carolina was small. Instead, my results suggest similar selection pressures acting to conserve the existing voltage-gated sodium channel genetic structure across geographically disjunct bottlenose dolphin populations.

Discussion of lack of survival-associated SCN variation

Highly conserved SCNA sequences across bottlenose dolphins along the Gulf coast of Florida highlight the biological importance of the voltage-gated sodium channels, which are highly evolved and seemingly exhibit little variation in protein structure. In other cases of neurotoxin resistance, fitness tradeoffs are observed in conjunction with SCN amino acid substitutions (Brodie III & Brodie Jr 1999). These costs explain why resistance-inducing mutations are only observed in natural populations frequently exposed to neurotoxins. These costs may also explain why I do not observe SCNA mutations associated with brevetoxin resistance in bottlenose dolphins. Any change in the voltage-gated sodium channels that affects muscular control may be too costly for this species, or the fitness advantage may not outweigh the cost of mutation.

My findings may be explained by a difference in the strength or consistency of selection pressure between the dolphin-brevetoxin system and other studied systems that have evolved neurotoxin resistance. In many other systems, predators have evolved neurotoxin resistance to prey that sequester toxins as a defense mechanism (e.g., Brodie *et al.*, 2002). However, bottlenose dolphin prey do not sequester toxins as a defense mechanism, and dolphins are separated by at least one trophic level from the toxin-producing algae. In addition, HABs are episodic events, although low levels of toxin can be detected in the prey fish months after the termination of a bloom (Fire *et al.* 2008a).

Individual or population-level variation in bottlenose dolphin resistance to HABs may alternatively be explained by non-genetic hypotheses. For example, bottlenose dolphins may behaviorally adapt to the presence of HABs and/or utilize foraging strategies that minimize their exposure. Changes in dolphin foraging behavior have been observed during HABs, likely as a result of variation in prey abundance (Bowen 2011; McHugh *et al.* 2011; Powell & Wells 2011). Certain foraging behaviors may increase the likelihood of exposure to HABs; for example, eating discarded bycatch from commercial fishing vessels provides an avenue of exposure to toxins that originate in distant blooms (National Marine Fisheries Service 2004).

Chronic low levels of exposure may also lead to the acquisition of resistance. Low levels of brevetoxins have been measured in dolphins in Sarasota Bay even when HABs are not present (Fire *et al.* 2008b). These low levels of toxins may be explained by long-term retention of the toxin or by prolonged exposure to continual sources of toxin, such as prey species that store toxin (Fire *et al.* 2008a) or toxin adsorbed to plant material (Flewelling *et al.* 2005). Little research has been conducted on the effects of chronic exposure to background levels (<1,000 *K. brevis* cells/L) of this toxin in marine species.

My findings do not allow us to rule out a genetic basis for resistance to brevetoxins in bottlenose dolphins. Genetic variation associated with survival may be located in other parts of the dolphin genome, including other parts of the SCN genes outside the brevetoxin binding site. A genotyping-by-sequencing analysis of genome-wide variation in bottlenose dolphin HAB survivors and non-survivors has identified several other candidate genes related to the immune system, neuron development and detoxification (Chapter I). Such genomic approaches, which do not target specific genes, provide an excellent complement to the more traditional candidate gene approaches, such as those employed in this chapter.

Conclusion

My study found little genetic variation in the putative brevetoxin binding site in the voltage-gated sodium channels in bottlenose dolphins. Variation was rare and not associated with survival in dolphins exposed to HABs in the Gulf of Mexico. In contrast to many other previously studied neurotoxin-resistant species, I conclude that bottlenose dolphins have not evolved resistance to HABs via mutations in the brevetoxin binding site on the voltage-gated sodium channels.

CHAPTER III:
FROM GENOME-WIDE TO CANDIDATE GENE: AN INVESTIGATION OF VARIATION AT THE MAJOR HISTOCOMPATIBILITY COMPLEX IN COMMON BOTTLENOSE DOLPHINS EXPOSED TO HARMFUL ALGAL BLOOMS

Introduction

The major histocompatibility complex (MHC) is a family of highly polymorphic genes that plays an important immunological role in antigen presentation. Found primarily on the surface of antigen-presenting cells, such as dendritic cells and macrophages, MHC class II molecules bind peptides from extracellular pathogens and stimulate immune responses mediated by $CD4^{+}$ T lymphocytes (Germain 1994; Germain *et al.* 1996). The process of antigen binding by MHC molecules is specific and determined largely by residues in the peptide-binding region (PBR), or groove, of the MHC. Selection on MHC genes in this region is driven both by pathogen presence in a frequency-dependent manner (Takahata & Nei 1990) and by heterozygote advantage (Doherty & Zinkernagel 1975) resulting in balancing selection. Many studies have identified associations between genetic diversity at MHC genes and resistance to bacterial, viral, and parasitic pathogens (reviewed by Sommer 2005).

Relatively few studies have investigated the role of MHC diversity in resistance to environmental toxins (Cohen 2002; Ingale 2010). However, the genome-wide association study described in Chapter 1 suggested a potential role of genetic variation at the MHC in resistance of common bottlenose dolphins (*Tursiops truncatus*) to toxins produced by harmful algal blooms (HABs), or red tides. In the Gulf of Mexico, blooms of the brevetoxin-producing dinoflagellate *Karenia brevis* have been associated with large-scale mortality events of several marine species, including common bottlenose dolphins (Flewelling *et al.* 2005). The most commonly observed effects of brevetoxin exposure are neurotoxicity (O'Shea *et al.* 1991; Poli *et al.* 1986; Watkins *et al.* 2008) and respiratory distress (Bossart *et al.* 1998; Singer *et al.* 1998), but brevetoxins may also induce genotoxic (Sayer *et al.* 2005), immunotoxic (described below), and teratogenic (Kimm-Brinson & Ramsdell 2001) effects.

The effects of brevetoxin exposure on the immune system have been described in naturally exposed marine animals (Bossart *et al.* 1998; Kreuder *et al.* 2002; Walsh *et al.* 2010; Walsh *et al.* 2005) and studied *in vivo* and *in vitro* in the laboratory. Brevetoxins have been found in both macrophages and lymphocytes of manatees (*Trichechus manatus latirostris*) exposed to HABs (Bossart *et al.* 1998). *In vitro*, brevetoxin exposure in macrophages induces an inflammatory response (Sas & Baatz 2010), and in lymphocytes, low dose brevetoxin exposure stimulates cell proliferation, but high doses cause DNA damage and inhibit cell proliferation (Murrell & Gibson 2009). Voltage-gated sodium channels, the biological target of brevetoxins (Poli *et al.* 1986), are expressed on both macrophages and lymphocytes (Roselli *et al.* 2006), and although the role of sodium channels in immune function is still poorly understood, it is possible that they facilitate the immunotoxic effects of brevetoxins (Murrell & Gibson 2009).

Brevetoxin's effects on the immune system may also result from their inhibitory action on cathepsins (Sudarsanam *et al.* 1992). Cathepsins play an important role in enabling MHC antigen presentation, and inhibition of these enzymes can reduce T lymphocyte response to infection (Katunuma *et al.* 2003). It is possible that polymorphisms in MHC genes may counteract or intensify the effects of brevetoxin-induced cathepsin inhibition on the T-cell response (Chang *et al.* 2008).

MHC gene expression has been previously evaluated in bottlenose dolphins as part of a study testing for unique expression profiles of dolphin stress and immune system genes among four dolphin populations, including two from the Gulf of Mexico, Sarasota Bay and St. Joseph Bay, Florida, where HABs have been associated with large-scale mortality events of bottlenose dolphins. Mancia *et al.* (2010) found significant differences in the expression of the MHC DRA gene among dolphin populations. Genetic variation in this system may be the result of differences in environmentally driven selection or genetic drift. Estuarine populations of bottlenose dolphins in this region exhibit high site fidelity and experience limited gene flow, resulting in significant genetic differentiation at a fine geographic scale (Sellas *et al.* 2005). This metapopulation structure makes individual populations particularly susceptible to localized threats, like HABs, but also enables local adaptation, such as for increased resistance to local threats (Kawecki & Ebert 2004).

Dolphin populations in Sarasota Bay and the adjacent coastal waters of central-west Florida experience more frequent HAB exposure than populations in St. Joseph Bay and the adjacent coastal waters of the Florida Panhandle. HABs of *K. brevis* initiate along the Western Florida Shelf (Steidinger 1975) and occur almost annually along the central-west Florida shoreline (Florida Fish and Wildlife Research Institute 2008; Steidinger 2009). Bottlenose dolphins in this region appear to be relatively resistant to HABs; only one large-scale unusual mortality event (UME), which occurred during an unusually severe and prolonged HAB in 2005-2006, has been attributed to brevetoxin exposure in central-west Florida (Twiner *et al.* 2012). In contrast, multiple UMEs associated with brevetoxin exposure have occurred in the Florida Panhandle, where HABs occur less frequently (pers. comm. L. Flewelling; Gaydos 2006). In the Florida Panhandle, bottlenose dolphin UMEs associated with brevetoxin exposure occurred in 1999-2000, 2004, and 2005-2006 (Mase *et al.* 2000; National Marine Fisheries Service 2004; Twiner *et al.* 2012).

I investigated a possible genetic basis to the apparent difference in resistance to brevetoxins between individuals and populations of bottlenose dolphins in the Gulf of Mexico using a genome-wide approach, described in Chapter 1. The approach, which utilized a genotyping-by-sequencing (GBS) technique, provided an unbiased search across the genome for variation associated with survival. However, most variants were found in non-coding portions of the genome, and it can be difficult to understand the biological importance of such genetic variation. Following a genome-wide study, candidate genes surrounding phenotype-associated single nucleotide polymorphisms (SNPs) can be sequenced to identify linked functional variation (McCarthy &

Hirschhorn 2008; Wilkening *et al.* 2009). In this chapter, I investigate genetic variation at four MHC loci that are located in genomic proximity to a SNP that exhibited nearly fixed differences between dolphins that died during HABs and those from nearby estuarine and coastal populations that survived (Fig. 6). In comparison to genetic variation at neutral microsatellite markers, used to estimate the influence of genetic drift, I investigate variation at the MHC DRA, DRB, DQA, and DQB loci in live and dead dolphins from central-west Florida and the Panhandle.

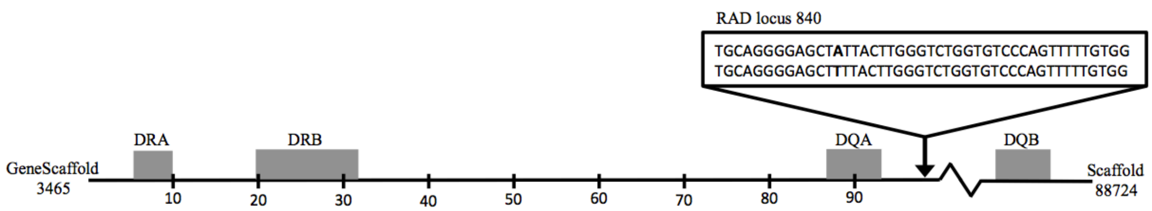

Figure 6: Genomic location of four MHC loci surrounding a survival-associated SNP. Sequence of restriction site-associated DNA (RAD) locus 840 highlights a SNP at position 13. Scaffold numbers refer to genome turTru1 (Ensembl, release 71, GenBank Assembly ID: GCA_000151865.1). Marks along scaffold represent distance in kb. The exact distance from RAD locus 840 to DQB is not available.

Methods

Samples

I compared genetic diversity in dolphins that died during three UMEs associated with red tides to dolphins from nearby coastal and estuarine populations that survived. My study included a total of 215 samples from the following seven groups of bottlenose dolphins: dolphins that died during UMEs associated with red tides in the Florida Panhandle in 1999-2000 (N=21) and 2004 (N=40); live dolphins from the Florida Panhandle, including the estuarine population inhabiting St. Joseph Bay (N=34) and the adjacent coastal population (N=22); dolphins that died during UMEs associated with red tides in central-west Florida in 2005-2006 (N=31); and live dolphins from central-west Florida, including the estuarine population inhabiting Sarasota Bay (N=25) and the adjacent coastal population (N=42). All live dolphins were assumed to have survived at least one red tide based on their age and geographic distribution. Samples were categorized as coastal or estuarine based on photo-ID site fidelity (pers. comm. B. Balmer & R. Wells). Dolphin skin and genomic DNA samples were acquired from the Chicago Zoological Society's Sarasota Dolphin Research Program, the Mote Marine Laboratory Stranding Investigations Program, and the NOAA Southeast Fisheries Science Center (SEFSC) Marine Mammal Tissue and DNA Archive.

Samples were analyzed jointly by myself at the Duke University Marine Conservation Molecular Facility (MCMF) and by P. Rosel, L. Wilcox, and I. T. Rodriguez at the NOAA SEFSC Marine Mammal Molecular Genetics Laboratory (MMMGL). I conducted all sequencing and genotyping for the DQA and DRA loci, and generated 27% of the DQB and DRB sequences and microsatellite genotypes. The remaining genotypes were provided by MMMGL.

DNA extraction

Genomic DNA provided by the NOAA SEFSC Marine Mammal Tissue and DNA Archive had been extracted following Rosel and Block (1996) using 250 µl volumes and Phase Lock gel tubes (5 Prime). I extracted DNA from an additional 29 skin samples, representing primarily the Sarasota Bay population, using a Wizard® Genomic DNA Purification Kit (Promega) following manufacturer instructions.

MHC genotyping

Individual PCRs were used to amplify the following portions of the alpha and beta chains of two MHC class II proteins in each sample: exon 2 of DQA (228 bp), DQB (172 bp), and DRB (238-246 bp); and exon 3 (272 bp), intron 3 (299 bp) and exon 4 (104 bp) of DRA (675 bp total). Exon 2 was targeted for most MHC loci due to its role in peptide binding (Brown *et al.* 1993); however, previous studies identified very little variation in exon 2 of DRA in bottlenose dolphins (Arbanasic *et al.* 2012) and other cetaceans (Xu *et al.* 2008; Xu *et al.* 2007; Xu *et al.* 2009), so I therefore targeted exons 3 and 4 of DRA where SNPs have been reported in other species (Chen *et al.* 2010; Sherry *et al.* 2001) and in other MHC loci in bottlenose dolphins (Yang *et al.* 2007). Exons 2 and 3 encode extracellular domains and exon 4 encodes the transmembrane domain of MHC molecules (van der Poel *et al.* 1990).

These four MHC loci were selected due to their close genomic proximity to restriction site-associated DNA (RAD) locus 840 (Fig. 6), which contains a SNP exhibiting nearly fixed differences between pools of dolphins that died putatively due to HAB exposure and pools of dolphins that survived previous red tide exposure (Chapter 1). DRA, DRB, and DQA are located within 100 kb of RAD locus 840 on GeneScaffold 3465 (Ensembl, release 71, turTru1, GenBank Assembly ID: GCA_000151865.1), while DQB is located on Scaffold 88724. In more complete genomes, DQB is found within the gene cluster including DQA, DRA, and DRB (Horton *et al.* 2004), so I assume GeneScaffold 3465 and Scaffold 88724 are in close proximity in the dolphin genome even though low genomic sequence coverage prevents their alignment. Assuming the MHC class II gene cluster is intact in bottlenose dolphins, variation at the DQB locus may also be linked to SNP variation in RAD locus 840.

PCR primers, reagents, and temperature profiles are reported in Table 7. PCR products were genotyped using either single-strand conformation polymorphism (SSCP) or direct sequencing.

Single-strand conformation polymorphism (SSCP)

Rosel and colleagues at the MMMGL initially attempted to use SSCP to genotype DQB and DRB in 89 samples. Amplified products were prepared for SSCP by combining 1 μl of PCR product (diluted by up to 1:50), 0.5 μl of freshly made 0.3 M sodium hydroxide, 1 μl of GeneScan 500 TAMRA Size Standard (Applied Biosystems), and 10.5 μl of Hi-Di formamide. This mixture was denatured at 95°C for 5 min and then cooled on ice for at least 5 min before electrophoresis on an ABI 310 Genetic Analyzer using 3% GeneScan Polymer (Applied Biosystems) and 10% glycerol in 1x TBE buffer. Runs were maintained at 13kV and 30°C or 40°C for DRB or DQB, respectively. SSCP profiles were analyzed using the 310 GeneScan Analysis Software v3.1.2 (Applied Biosystems).

All unique SSCP patterns were cloned for sequencing using the StrataClone PCR Cloning Kit (Agilent). Following cloning, Rosel and colleagues found that multiple alleles had similar SSCP peak patterns and determined it would be necessary to sequence most samples for accurate genotyping. Only 23 DQB genotypes in the final dataset were determined by SSCP; the remaining genotypes were determined by direct sequencing.

Direct sequencing

Following amplification, PCR products were prepared for sequencing using purification and sequencing protocols that differed slightly between the MCMF and MMMGL. In the MCMF, I purified PCR and sequencing products using enzymes and AmpPure magnetic beads (Agencourt), respectively. To enzymatically purify 10 µl of PCR product, I added 1 unit Exonuclease I and 0.25 units Antarctic Phosphatase (New England Biolabs) and incubated the mixture at 37°C for 1 hr. Purified products were sequenced using BigDye v1.1 (Duke IGSP) (Table 7) on an ABI3730xl DNA Analyzer. In the MMMGL, Rosel and colleagues purified PCR and sequencing products using LMP gel extraction and ethanol precipitation, respectively. Purified products were sequenced using a BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems) on an Applied Biosystems ABI3130 Genetic Analyzer. To ensure consistency between genotypes produced in the two labs, we compared genotypes from positive controls, and all sequences for a given locus were edited by the same person. I edited all DQA and DRA sequences using CodonCode Aligner v3 (CodonCode Corporation), and L. Wilcox edited all DQB and DRB sequences using Sequencher v4 (Gene Codes).

All unique alleles that were not observed in multiple homozygous individuals were confirmed as outlined by Marsh *et al.* (2005) by sequencing an average of 10 clones produced using the StrataClone PCR Cloning Kit (Agilent) in the MMMGL or 16-24 clones produced using pGEM®-T Easy Vectors with JM109 competent *E. coli* cells (Promega) in the MCMF. Alleles were confirmed only if observed in multiple clones. In the end, all but one allele that was identified was cloned and sequenced. The only allele not confirmed by sequencing, Tutr-DQA*07, was observed in six homozygotes from this study and has been previously described in bottlenose dolphins (GenBank: KF751589).

I compared all allele sequences to existing cetacean MHC alleles in NCBI's GenBank sequence database (http://www.ncbi.nlm.nih.gov/genbank/) to identify transpecific polymorphisms. At the time of this comparison, February 2014, there were 21 DQA exon 2 sequences representing 12 cetacean species, 245 DQB exon 2 sequences representing 38 cetacean species, and 107 DRB sequences representing 21 cetacean species. These numbers are likely an overestimate of the number of true MHC alleles in the database due to redundancy and cloning error. DRA exons 3-4 with the intermediate intron have not been previously studied in cetaceans.

If an exact match was found to a bottlenose dolphin allele already in GenBank, the allele is referred to throughout this chapter using the existing name. The naming of DQB alleles was somewhat complicated by redundancy in GenBank and variation among GenBank sequences in intron sequence that I did not sequence. I therefore refer to some DQB alleles using multiple numbers (e.g., Tutr-DQB*01/12) when matching sequences have been contributed to the database and named by different authors or when GenBank sequences match across exon 2 and differ only in flanking intronic sequence. All new alleles were named following the convention of Klein *et al.* (1990) and numbered in sequence following those already listed in GenBank to avoid further redundancy (Tutr-DQA*11-17, Tutr-DQB*25-27, Tutr-DRB*08-19). DRA alleles (Tutr-DRA*01-05) were defined and numbered based on exon variation, but when additional intronic variation was present, a lowercase letter was added following the number (e.g., Tutr-DRA*01a).

Microsatellite genotyping

Most samples were genotyped at 19 microsatellite loci: Ttr04, Ttr11, Ttr19, Ttr34, Ttr48, Ttr58, Ttr63, TtrFF6 (Rosel *et al.* 2005), MK5, MK6, MK8, MK9 (Krützen *et al.* 2001), EV14, EV37, EV94 (Valsecchi & Amos 1996), TexVet5, TexVet7 (Rooney *et al.* 1999), KWM12a (Hoelzel *et al.* 1998), and Ppho130 (Rosel *et al.* 1999). Microsatellite genotyping followed methods described in Rosel *et al.* (2009). Eleven samples from the central-west Florida 2005-2006 UME did not have available microsatellite genotypes. I therefore ran all statistical analyses comparing MHC and microsatellite diversity with and without these samples. I observed no difference between these analyses so I present only results including all 215 samples, where eleven samples from the central-west Florida 2005-2006 UME are coded as having missing alleles for all microsatellite loci.

Analyses

To assess the evolutionary relationship among MHC alleles, I conducted phylogenetic analyses in MEGA v5.2 (Tamura *et al.* 2011). For each locus, I selected the evolutionary model that best described the nucleotide substitution pattern (maximum likelihood fit with lowest Bayesian information criterion) and then accordingly built a maximum likelihood tree with 1,000 bootstrap replicates (Felsenstein 1985). Evolutionary models included the Jukes-Cantor model (Jakes & Cantor 1969), Kimura 2-parameter model (Kimura 1980), and Tamura 3-parameter model (Tamura 1992). All trees were rooted using a bovine MHC sequence as an outgroup.

I evaluated selection on each locus by comparing the number of non-synonymous (d_N) and synonymous (d_S) substitutions per site across the full exon sequence and including and excluding the putative peptide-binding region (PBR). PBR sites were defined as in Brown *et al.* (1993). These analyses were conducted using the Nei-Gojobori method with Jukes-Cantor correction (Nei & Gojobori 1986) implemented in MEGA v5.2 (Tamura *et al.* 2011).

All microsatellite and MHC loci were tested for Hardy-Weinberg Equilibrium (HWE) and linkage disequilibrium using the Markov chain method implemented in Genepop v4.2 (Raymond & Rousset 1995) with 1,000 dememorization steps followed by 100 batches with 1,000 iterations each. Microsatellite loci were tested for neutrality using the Lositan Selection Workbench (Antao *et al.* 2008; Beaumont & Nichols 1996) with 50,000 simulations, assuming a stepwise mutation model and a forced 'neutral' mean F_{ST}.

I evaluated MHC diversity among groups of dolphin samples using estimates of individual heterozygosity, genetic differentiation, and allele frequency. Heterozygosity was evaluated as the average number of alleles per individual for MHC loci and as internal relatedness (IR) (Amos *et al.* 2001) and standardized multi-locus heterozygosity (MLH) (Coltman *et al.* 1999) for microsatellite loci. IR and MLH were calculated using the Rhh package (Alho & Valimaki 2012) in R (R Core Team 2013). I used diveRsity (Keenan *et al.* 2013) in R to calculate pairwise estimators of G'_{ST} (Hedrick 2005), which use $\hat{H}_S$ and $\hat{H}_T$ to incorporate sample size into the estimate (Nei & Chesser 1983). G'_{ST} provides a measure of differentiation standardized for heterozygosity, which enables a relatively unbiased comparison of statistics among loci that differ in heterozygosity, including microsatellites and nuclear genes. I further tested the effects of region (central-west Florida vs. Panhandle) and survival (live vs. dead) on allele frequency of each MHC and microsatellite locus using multifactorial logistic models implemented in JMP Pro 11.0.0 (SAS Institute Inc. 2014). Only alleles present at >5% allele frequency in the complete sample set were included in these models.

For one MHC locus in which allele frequencies appeared to be significantly affected by survival, I also tested for differences in the frequency of each SNP variant found within the PBR using Fisher's Exact Tests. This SNP analysis more closely paralleled the analysis of SNPs identified in RAD loci, as described in Chapter 1. I was therefore able to compare the variation previously observed at a SNP within RAD locus 840 with the variation described in this chapter at MHC SNPs in close genomic proximity (Fig. 6).

Table 7: PCR primers, reagents, and temperature profiles for amplification and sequencing of MHC loci. Following amplification, genotypes were determined by single strand conformation polymorphism (SSCP) or direct sequencing (Seq).

Locus	Primers	Primer Sequence (5'-3') or Source	Reagents		Temperature Profile
DQB	DQB1[1]	Tsuji *et al.* 1992 as reported in	25 ng DNA, 1x High Fidelity PCR Buffer		95°C 2min
SSCP	DQB2[1]	Murray *et al.* 1995	0.2 µM of each primer		95°C 30sec
			0.15 mM dNTPs, 2.5 mM $MgSO_4$	10x	64°C* 20sec
			0.4U Platinum *Taq* High Fidelity		72°C 1min
					*dec by 1°C every cycle
					95°C 40sec
				25x	54°C 30sec
					72°C 1min
					72°C 10min
DQB	DQB1	Tsuji *et al.* 1992 as reported in	25 ng DNA, 1x High Fidelity PCR Buffer		95°C 2min
Seq	DQB2	Murray *et al.* 1995	0.2 µM of each primer		95°C 30sec
			0.15 mM dNTPs, 2.5 mM $MgSO_4$	35x	62°C 30sec
			1U Platinum *Taq* High Fidelity		72°C 1min
			BSA if needed		72°C 10min
DRB	DRBAMP-A[1]	Munguia-Vega *et al.* 2007	25 ng DNA, 1x High Fidelity PCR Buffer		94°C 2min
SSCP	DRB3b[1]	Baker *et al.* 2006	0.3 µM of each primer		94°C 30sec
			0.15 mM dNTPs, 1.5 mM $MgSO_4$	35x	52°C 30sec
			0.4U Platinum *Taq* High Fidelity		72°C 1min
					72°C 10min
DRB[2]	DRBAMP-A	Munguia-Vega *et al.* 2007	25 ng DNA, 1x High Fidelity PCR Buffer		94°C 2min
Seq	DRB3b	Baker *et al.* 2006	0.3 µM of each primer		94°C 30sec
			0.15 mM dNTPs, 3.0 mM $MgSO_4$	35x	52°C 30sec
			1U Platinum *Taq* High Fidelity		72°C 1min
			BSA if needed		72°C 10min

DRB[2]	DRBAMP-A	Munguia-Vega *et al.* 2007	25 ng DNA, 1x High Fidelity PCR Buffer		94°C 2min
Seq	DRBe2	CTCACCTCGCCGCTGCAC	0.3 μM of each primer		94°C 15sec
			0.15 mM dNTPs, 1.5mM $MgSO_4$	30-35x	48°C 15sec
			0.5U Platinum *Taq* High Fidelity		72°C 15sec
			0.08 mg/ml BSA		72°C 10min
DRA	DRAe34f	TGTCCCCCAGTACCTCCAGA	25 ng DNA, 1x MCMF PCR Buffer[3]		94°C 3min
Seq	DRAe34r	TTTGCGCACACCCTTGAT	0.25 μM of each primer		94°C 15sec
			0.2 mM dNTPs, 2.0 mM $MgCl_2$	55x	70-55°C* 15sec
			1U *Taq*		72°C 30sec
			BSA if needed		72°C 5min
		*beginning at 70°C dec by 0.5°C every cycle for first 30 cycles, then 55°C for 25 cycles			
DQA	DQAex2F	Heimeier *et al.* 2009	25 ng DNA, 1x MCMF PCR Buffer[3]		94°C 3min
Seq	DQAex2R-mod[4]	GAGGCAGAATGGTGGACAC	0.2 μM of each primer		94°C 20sec
			0.2 mM dNTPs, 2.5 mM $MgCl_2$	30x	58°C 15sec
			1U *Taq*		72°C 20sec
			BSA if needed		72°C 10min
MCMF Sequencing Reaction			1x MCMF BigDye Buffer[3]		94°C 30sec
(10 μl)			1 μM primer, 0.44 μl BDX64 (MCLAB)		94°C 15sec
			0.06 μl BigDye v1.1 (Duke IGSP)	25x	50°C 15sec
			1-5.2 μl purified PCR product		60°C 3min
MMMGL Sequencing Reaction			1x ABI Sequencing Buffer		96°C 1min
(10 μl)			0.32 μM primer		96°C 10sec
			1 μl BigDye v1.1 (ABI)	25x	50°C 5sec
			2 μl purified PCR product		60°C 4min

[1]For SSCP, forward and reverse primers were fluorescently labeled with 6-FAM and TET, respectively.

[2]Samples were initially sequenced with DRB3b, until sequence analyses identified a subset of samples exhibiting poor sequence using DRB3b. These poor samples and all remaining samples (N = 175) were amplified and sequenced with DRBe2.

[3]PCR Buffer: 20 mM Tris pH 8.8, 50 mM KCl, 0.1% Triton X-100, 0.2 mg/mL BSA; BigDye Buffer: 40 mM Tris, 5 mM $MgCl_2$, 1.5 M Betaine.

[4]DQAex2R was slightly modified (removal of four 5′ bases and addition of three 3′ bases) to improve reverse sequence quality.

Results

MHC sequence variation

I identified 10 DQA alleles (Fig. 7 and 11), 11 DQB alleles (Fig. 8 and 12), 5 DRA alleles (Fig. 9 and 13), and 12 DRB alleles (Fig. 10 and 14) in bottlenose dolphins from the Gulf of Mexico. Three DQA alleles and eight DQB alleles were identical to alleles that had been previously described in bottlenose dolphins from the Adriatic Sea, coastal waters of Taiwan, and Japanese aquariums (Kita *et al.* 2010; Yang *et al.* 2008). Additionally, DQA allele Tutr-DQA*11 was identical in sequence to exon 2 of complete DQA transcript variants 1 and 2 (GenBank accession nos. XM_004317915 and XM_004317916), and Tutr-DQA*12 was identical in sequence to Tutr17-DQA (GenBank: EU024798 (Heimeier *et al.* 2009)) in 167 bp, but this deposited sequence did not cover the complete exon 2 of DQA (missing first 78 bp). Two DQB alleles were identical to DQB alleles from other delphinids (Tutr-DQB*04/14 and Tuad-DQB*05 [GenBank: EU698977] (Xu *et al.* 2009); Tutr-DQB*13/24 and Stco-DQB*06 [GenBank: EU698969] (Xu *et al.* 2009)).

No individual contained more than two alleles at any one locus, suggesting that only a single locus was amplified in all cases. Additionally, maximum likelihood phylogenetic analyses of MHC alleles do not support multiple lineages within any locus (Fig. 15-18).

The pattern and type of nucleotide substitutions in the DRB and DQB alleles suggested that these two loci encoding MHC beta chains have experienced balancing selection, particularly in the PBR (Table 8). d_N was significantly greater than d_S, and this difference was greater in PBR sites than non-PBR sites. Less difference was observed between d_N and d_S across DQA alleles, and DRA showed no significant difference between d_N and d_S, though the sequence did not include the PBR. MHC beta chains were also more variable than alpha chains in the number of alleles and average nucleotide distance among alleles (Table 8).

Population genetic analyses

Both MHC and microsatellite loci were tested for deviations from HWE within each sample and for linkage disequilibrium across the entire sample set. In all but one case (DQB in Sarasota Bay, $P < 0.001$), loci showed no significant deviations from HWE, following a sequential Bonferroni correction for multiple tests (Rice 1989). No pairs of microsatellite loci appeared linked, but, as expected due to their proximity in the genome, tests for linkage disequilibrium among almost all pairs of MHC loci were highly significant. When testing for neutrality, only one microsatellite locus, Ttr48, was found to be a significant outlier ($P < 0.01$). Analyses of population differentiation described below were therefore run both with and without Ttr48; I found no difference in conclusions drawn from the results of both tests and therefore only present results from analyses including all microsatellite loci.

Geographic region

I compared genetic diversity at MHC and microsatellite loci between live dolphins from central-west Florida and the Panhandle. In multifactorial logistic models, region (central-west Florida vs. Panhandle) had a significant effect on allele frequency for all MHC loci and almost all microsatellite loci (Table 10). DQB and DRB loci showed different patterns of genetic differentiation (G'_{ST}) between central-west Florida and the Panhandle than those observed at putatively neutral microsatellite markers (Fig. 19A). Genetic differentiation at DQB was greater than microsatellite differentiation across all four comparisons of coastal and estuarine populations in central-west Florida and the Panhandle. Genetic differentiation at DRB was also greater than microsatellite differentiation in comparisons of the central-west Florida estuarine population to the Panhandle populations, but lower in comparisons of the central-west Florida coastal population to the Panhandle populations.

Survival

I compared genetic diversity at MHC and microsatellite loci between live and dead dolphins and found a significant effect of survival on DRB allele frequency overall (χ^2 = 27.02, df = 7, $P < 0.001$, Table 10). The differences in allele frequency neared significance in three of six individual comparisons of dolphins that died during three UMEs associated with red tides, each compared to live dolphins from nearby coastal and estuarine populations (Fig. 20). Genetic differentiation (G'_{ST}) at DRB was greater than microsatellite differentiation in five out of six live-dead comparisons (Fig. 19B). However, the observed differences could not be attributed to differences in frequency of any single DRB allele or non-synonymous SNP within the PBR.

There was no evidence of a significant effect of survival on any other MHC locus (Table 10, Fig. 19B). A significant effect of survival was observed for one microsatellite locus, EV37 ($\chi^2 = 37.14$, df = 5, $P < 0.0001$). All analyses were re-run without this locus, but there was little difference in average microsatellite G'_{ST} when this locus was removed. No locus showed a significant interaction effect between region and survival (Table 10).

Heterozygosity

I found no significant difference in the average number of MHC alleles per individual between live and dead bottlenose dolphins in the six comparisons described above (Table 9). Similarly, I found no significant difference in multi-locus microsatellite heterozygosity between live and dead dolphins (Table 9). These results suggested that neither heterozygote advantage at these MHC loci nor inbreeding depression play a significant role in variable HAB resistance in bottlenose dolphins. However, overall I did find slightly higher heterozygosity in central-west Florida than the Panhandle when all live and dead samples from each region were pooled for analysis (central-west FL: avg MLH = 1.03, sd = 0.16; FL Panhandle: avg MLH = 0.98, sd = 0.13; t = -2.29, $P < 0.05$).

Table 8: Sequence diversity among MHC alleles. The average proportion (p-distance) of nucleotide sites that differ among alleles, and the number of non-synonymous (d_N) and synonymous (d_S) substitutions per site for each MHC locus across the full exon sequence and including and excluding the putative peptide-binding region (PBR).

				Overall				PBR				non-PBR			
	Exon	Alleles	p-distance (SE)	Length (aa)	d_N (SE)	d_S (SE)	$d_N > d_S$ P-value	Length (aa)	d_N (SE)	d_S (SE)	$d_N > d_S$ P-value	Length (aa)	d_N (SE)	d_S (SE)	$d_N > d_S$ P-value
DQA	2	10	0.028 (0.007)	75	0.036 (0.011)	0.009 (0.009)	0.008	17	0.071 (0.037)	0.008 (0.008)	0.023	58	0.025 (0.010)	0.010 (0.010)	0.051
DQB	2	11	0.075 (0.012)	57	0.105 (0.023)	0.012 (0.007)	0	14	0.299 (0.073)	0.034 (0.025)	0	43	0.052 (0.020)	0.006 (0.006)	0.024
DRB	2	12	0.085 (0.010)	79	0.111 (0.020)	0.028 (0.010)	0	22	0.290 (0.065)	0.083 (0.034)	0.001	57	0.055 (0.016)	0.007 (0.005)	0.003
DRA	3-4	5	0.004 (0.002)	125	0.003 (0.002)	0.015 (0.007)	0.140								

Table 9: Heterozygosity at MHC and microsatellite loci. Evaluated as the average number of alleles per individual for four MHC loci and as internal relatedness (IR) and standardized multi-locus heterozygosity (MLH) across 19 microsatellite loci.

	DQB		DRB		DQA		DRA		IR		MLH	
	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
Central-west FL coastal	1.81	0.40	1.93	0.26	1.83	0.38	1.29	0.46	0.02	0.13	1.05	0.15
Central-west FL estuarine	1.60	0.50	1.72	0.46	1.68	0.48	1.12	0.33	0.02	0.15	0.99	0.16
Central-west FL 2005-06 UME	1.77	0.43	1.84	0.37	1.71	0.46	1.19	0.40	0.01	0.13	1.03	0.16
FL Panhandle coastal	1.86	0.35	1.77	0.43	1.77	0.43	1.05	0.21	0.05	0.15	0.97	0.16
FL Panhandle estuarine	1.79	0.41	1.88	0.33	1.71	0.46	1.09	0.29	0.04	0.15	0.98	0.17
FL Panhandle 2004 UME	1.83	0.38	1.85	0.36	1.80	0.41	1.13	0.33	0.03	0.16	1.00	0.16
FL Panhandle 1999 UME	1.76	0.44	1.81	0.40	1.71	0.46	1.05	0.22	0.09	0.12	0.95	0.12

Table 10: The effects of region (central-west Florida vs. Panhandle) and survival (live vs. dead) on allele frequency of four MHC loci and 19 microsatellite loci. Results of multifactorial logistic models. Bold *P* values remained significant following sequential Bonferroni correction within each factor.

		Region		Survival		Reg*Surv	
	df	X^2	*P*	X^2	*P*	X^2	*P*
DQA	4	14.196	**0.0067**	4.323	0.3591	0.357	0.9858
DQB	4	39.037	**<0.0001**	4.303	0.3665	1.981	0.7390
DRB	7	31.514	**<0.0001**	27.020	**0.0003**	11.267	0.1274
DRA	1	9.094	**0.0026**	0.517	0.4712	0.645	0.4219
Ttr04	5	19.648	**0.0015**	6.902	0.2280	8.551	0.1284
Ttr11	5	25.563	**0.0001**	3.867	0.5687	6.152	0.2917
Ttr19	1	3.898	0.0484	0.187	0.6656	5.217	0.0224
Ttr34	2	11.146	**0.0038**	3.693	0.1580	3.437	0.1793
Ttr48	1	34.324	**<0.0001**	0.061	0.8045	0.015	0.9015
Ttr58	2	10.685	**0.0048**	1.760	0.4147	3.399	0.1828
Ttr63	4	27.249	**<0.0001**	10.274	0.0361	15.861	0.0032
TtrFF6	2	1.838	0.3998	1.197	0.5495	0.233	0.8901
MK5	4	23.669	**<0.0001**	2.045	0.7274	5.485	0.2411
MK6	5	6.070	0.2993	3.889	0.5655	8.305	0.1402
MK8	3	37.329	**<0.0001**	0.265	0.9665	1.849	0.6043
MK9	3	24.178	**<0.0001**	8.421	0.0381	8.340	0.0384
EV14	4	6.299	0.1779	7.092	0.1311	4.379	0.3572
EV37	5	29.446	**<0.0001**	37.141	**<0.0001**	3.732	0.5886
EV94	2	6.080	0.0478	7.486	0.0237	2.493	0.2876
TexVet5	3	8.254	0.0410	81.74	0.0426	2.759	0.4302
TexVet7	3	13.097	**0.0044**	1.090	0.7794	2.540	0.4681
KWM12a	4	42.103	**<0.0001**	3.154	0.5324	2.736	0.6029
Ppho130	4	36.238	**<0.0001**	15.364	0.0040	14.68	0.0054

```
Tutr-DQA*02  GCACAACTAT CTACCAGTCT TTTGGGCCCT CTGGCCAGTA CACCCAAGAA [50]
Tutr-DQA*03  .......... .......... .......... .......... ......C...
Tutr-DQA*07  .......... .......... .......... .......... ..........
Tutr-DQA*11  .......... .......... .......... .......... ..........
Tutr-DQA*12  .......... .......... .......... .......... ..........
Tutr-DQA*13  .......... .......... .......... .........T ....AC....
Tutr-DQA*14  .......... .......... .......... .......... ......C...
Tutr-DQA*15  .......... .......... .......... .......... ..........
Tutr-DQA*16  .......... .......... .......... .........T ....AC....
Tutr-DQA*17  .......... .......... .......... .......... ......C...

Tutr-DQA*02  TTTGATGGAG ATGAGCTGTT TTATGTGGAC CTGGAGAAGA AGGAGACTGT [100]
Tutr-DQA*03  .......... .....AA... .......... .......... ..........
Tutr-DQA*07  .......... .......... .......... .......... ..........
Tutr-DQA*11  .......... .......... .......... .......... ..........
Tutr-DQA*12  .......... .......... .......... .......... ..........
Tutr-DQA*13  .......... .......... .......... .......... ..........
Tutr-DQA*14  .......... .......... .......... .......... ..........
Tutr-DQA*15  .......... .......... .......... .......... ..........
Tutr-DQA*16  .......... .......... .......... .......... ..........
Tutr-DQA*17  .......... ......A... .......... .......... ..........

Tutr-DQA*02  CTGGCGGCTG CCTGTGTTTA GCGAATTCGC AGGTTTTGAC CCACAGGGTG [150]
Tutr-DQA*03  .......... .......... .......... .A........ ......C...
Tutr-DQA*07  .......... .......... ........A. .A........ ..........
Tutr-DQA*11  .......... .......... .......... .......... ..........
Tutr-DQA*12  .......... .......... .......... .......... ..........
Tutr-DQA*13  .......... .......... .......... .A........ ..........
Tutr-DQA*14  .......... .......... .......... .A........ ..........
Tutr-DQA*15  .......... .......... .......... .......... ..........
Tutr-DQA*16  .......... .......... ........A. .A........ ..........
Tutr-DQA*17  .......... .......... .......... .A........ ..........

Tutr-DQA*02  CCCTGAGAAA CACAGCTGTG GCGAAACAAA ACTTGGATAT CCTGATTAAA [200]
Tutr-DQA*03  .......... ..T....... .T......T. .......... ..........
Tutr-DQA*07  .......... ..T....... .T......T. .......... ..........
Tutr-DQA*11  .......... ..T....... .T......T. .......... ..........
Tutr-DQA*12  .......... ..T....... .T......T. .......... ..........
Tutr-DQA*13  .......... ..T....... ......T.C. C......... .A........
Tutr-DQA*14  .......... ..T....... .......GC. C......... .A........
Tutr-DQA*15  .......... ..T....... .T......T. .......... .A........
Tutr-DQA*16  .......... ..T....... .......GC. C......... .A........
Tutr-DQA*17  .......... ..T....... ........T. .......... .A........

Tutr-DQA*02  CGCTCCAACT TTACCCCTGT TACCAATG [228]
Tutr-DQA*03  .......... .......... ........
Tutr-DQA*07  .......... .......... ........
Tutr-DQA*11  .......... .......... ........
Tutr-DQA*12  T......... .......... ........
Tutr-DQA*13  .......... .......... ........
Tutr-DQA*14  T......... .......... ........
Tutr-DQA*15  .......... .......... ........
Tutr-DQA*16  .......... .......... ........
Tutr-DQA*17  .......... .......... ........
```

Figure 7: Nucleotide sequence alignment of 10 DQA exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico (GenBank accession nos. KF751586, KF751589, KF751590, KJ722648-KJ722654). Dots indicate nucleotide identity with Tutr-DQA*02. Underlined sites are part of the putative peptide-binding region. Reading frame begins at position 3.

```
Tutr-DQB*25     CACGGAGCGG GTGCGGTTCG TGGACAGATA CATCTATAAT CGGGAGGAGT [50]
Tutr-DQB*26     .......... ......GC.A ..AG...... .........C ..........
Tutr-DQB*27     .......... ......G... .........C .........C ..........
Tutr-DQB*01/12  .......... ......GG.A ..AG...... .........C ..........
Tutr-DQB*04/14  .......... ......C... ....A..... .........C ..........
Tutr-DQB*05     .......... ......CA.. ..AG...... .........C ..........
Tutr-DQB*06     .......... ......G... .........C .........C ..........
Tutr-DQB*07     .......... .......... .......... .........C ..........
Tutr-DQB*08     .......... .......... .......... .........C ..........
Tutr-DQB*09/22  .......... ......C... ..AG....A. .........C ..........
Tutr-DQB*13/24  .......... ......CA.. ..AG...... .........C ..........

Tutr-DQB*25     TCGTGCGCTT CGACAGCGAC GTGGGCGAGC ACCGGGCGGT GACCGAGCTG [100]
Tutr-DQB*26     A......... .......... .........T T......... ..........
Tutr-DQB*27     .A........ .......... .........T .......... ..........
Tutr-DQB*01/12  A......... .......... .........T T......... ..........
Tutr-DQB*04/14  .......... .......... .........T T......... ..........
Tutr-DQB*05     A......... .......... .........T T......... ..........
Tutr-DQB*06     .A........ .......... .........T T......... ..........
Tutr-DQB*07     .......... .......... .........T T......... ..........
Tutr-DQB*08     A......... .......... .......... .......... ..........
Tutr-DQB*09/22  .G........ .......... .........T T......... ..........
Tutr-DQB*13/24  A......... .......... .........T T......... ..........

Tutr-DQB*25     GGCCGGCCGG ACGCCGAGTA CTGGAACAGC CAGAAGGACA TCCTGGAGCG [150]
Tutr-DQB*26     .......GTA C......... .......... ...G...... ........GA
Tutr-DQB*27     .......... .......... .......... .......... ..........
Tutr-DQB*01/12  .......... .......... ..TC...... .......... .........A
Tutr-DQB*04/14  .......... .....AC... .......G.. ...G...... ........GA
Tutr-DQB*05     .......... .....A.... .......... ...G...... ..........
Tutr-DQB*06     .......... .......... .......... .......... ..........
Tutr-DQB*07     .......... ........C. .......... .......... ..........
Tutr-DQB*08     .......... .......... .......... .........T ..........
Tutr-DQB*09/22  .......G.. C......... .......... .G.......T A........T
Tutr-DQB*13/24  .......... .......... ..TC...... .......... ........GA

Tutr-DQB*25     GAAACGGGCC GAGCTGGACA CG [172]
Tutr-DQB*26     .G........ .C.G...... .. [172]
Tutr-DQB*27     .......... .......... .. [172]
Tutr-DQB*01/12  .G........ .C.G...... .. [172]
Tutr-DQB*04/14  .G........ .C........ .. [172]
Tutr-DQB*05     .......... ..CG...... G. [172]
Tutr-DQB*06     .......... .......... .. [172]
Tutr-DQB*07     .G..T..... ..CGG..... .. [172]
Tutr-DQB*08     .......... .......... .. [172]
Tutr-DQB*09/22  ..G....... ...G...... .. [172]
Tutr-DQB*13/24  .G........ ..CG...... G. [172]
```

Figure 8: Nucleotide sequence alignment of 11 DQB exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico (GenBank accession nos. KJ722710-KJ722712, EF690293/AB302055, AB302047/AB302057, AB302048, AB302049, AB302050, AB302051, AB302052/AB302065, AB302056/AB302067). Some sequences assigned multiple allele numbers due to redundancy in GenBank or variation among GenBank sequences in intron sequence. Dots indicate nucleotide identity with Tutr-DQB*25. Underlined sites are part of the putative peptide-binding region. Reading frame begins at position 2.

```
Tutr-DRA*01a AGTGACTGTG CTCTCAAACA CCCCTGTGGA ACTGGGAGAG CCCAACGTCC [50]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ......A...
Tutr-DRA*03  .......... .......... .......... .......... ......A...
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... ........T. .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a TCATCTGTTT CATCAACAAG TTTTCTCCAC CAGTGATCAA TGTCACGTGG [100]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ......A...
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a CTTCAAAATG GCAAACCTGT CACCACAGGA GTGTCAGAGA CAGTCTTCCT [150]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a GCCCAGGAAT GACCACCTTT TCCGCAAGTT TCACTACCTC CCCTTCGTGC [200]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a CCTCAGCCGA GGATGTCTAT GACTGCAAGG TGGAGCACTG GGGTTTGGAG [250]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

                                      ¶
Tutr-DRA*01a GAGCCTCTTC TCAAGCACTG GGGTAAGAAC CATCCTTCAA TCTCCTGTAT [300]
Tutr-DRA*01b .......... .......... .......... T......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... T......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a TTTATGGTTT TCTCCTGTGA TGCATGTGTC TGGCCCTTTG GGGCCACAAG [350]
Tutr-DRA*01b ....C..... .......... C......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  ....C..... .......... C......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a TCCCCTAACT GTTTCATAAC AGCTGCTCCA ATTCCTGTTC CTGGTCTCCC [400]
Tutr-DRA*01b .......... .......... .......... .......... ..........
Tutr-DRA*01d .......... .......... .......... .......... ..........
Tutr-DRA*02  .......... .......... .......... .......... ..........
Tutr-DRA*03  .......... .......... .......... .......... ..........
Tutr-DRA*04  .......... .......... .......... .......... ..........
Tutr-DRA*05  .......... .......... .......... .......... ..........

Tutr-DRA*01a ATTGCTCAGC TTTTCCCACT TGTAATCTGT AATTCCCTTA ATACGTCATC [450]
Tutr-DRA*01b .......... ...C...... .......... .......... ..........
Tutr-DRA*01d .......... ...C...... .......... .......... ..........
Tutr-DRA*02  .......... ...C...... .......... .......... ..........
Tutr-DRA*03  .......... ...C...... .......... .......... ..........
Tutr-DRA*04  .......... ...C...... .......... .......... ..........
Tutr-DRA*05  .......... ...C...... .......... .......... ..........
```

Figure 9: Nucleotide sequence alignment of 5 DRA alleles identified in bottlenose dolphins from the Gulf of Mexico (GenBank accession nos. KJ722655, KJ722656, KJ722658, KJ722661-KJ722664). Sequences span exons 3-4 and the intermediate intron. DRA alleles defined by exon variation; Tutr-DRA*01a, *01b, and *01d differ only in intron sequence. Lines above sequence delineate exons 3 and 4. ¶ indicates start and end of the intermediate intron. Dots indicate nucleotide identity with Tutr-DRA*01a. Reading frame begins at position 2.

```
Tutr-DRA*01a  CTGTCTTCCT CTTATTTAGT TGATGAAAAA ATATTCTTTT GACTAAGTAC [500]
Tutr-DRA*01b  .......... .......... .......... .......... ..........
Tutr-DRA*01d  .......... .......... .......... .......... ..........
Tutr-DRA*02   .......... .......... .......... .......... ..........
Tutr-DRA*03   .......... .......... .......... .......... ..........
Tutr-DRA*04   .......... .......... .......... .......... ..........
Tutr-DRA*05   .......... .......... .......... .......... ..........

Tutr-DRA*01a  CCTATATTTG GTCCTAAGTG CTGCTCCCTA CCACATGCAT TACTTCTGTA [550]
Tutr-DRA*01b  .......... .......... .......... .......... ..........
Tutr-DRA*01d  .......... .......... .......... .......... ..........
Tutr-DRA*02   .......... .......... .......... .......... ..........
Tutr-DRA*03   .........T .......... .......... .......... ..........
Tutr-DRA*04   .......... .......... .......... .......... ..........
Tutr-DRA*05   .......... .......... .......... .......... ..........

                                    ¶
Tutr-DRA*01a  CTCTGATTTA TTTTTCCCCA GAGTATGAAG CTCCAACCCC CCTCACAGAG [600]
Tutr-DRA*01b  .......... .......... .......... .......... ..........
Tutr-DRA*01d  .......... .......... .......... .......... ..........
Tutr-DRA*02   .......... .......... .......... .......... ..........
Tutr-DRA*03   .......... .......... .......... .......... ..........
Tutr-DRA*04   .......... .......... .......... .......... ..........
Tutr-DRA*05   .......... .......... .......... .......... ..........

Tutr-DRA*01a  ACCACAGAGA ACGCAGTGTG TGCCCTGGGC CTGGTTATGG CTCTGGTGGG [650]
Tutr-DRA*01b  .......... .......... .......... .......... ..........
Tutr-DRA*01d  .......... .......... .......... .......... ..........
Tutr-DRA*02   .......... .......... .......... .......... ..........
Tutr-DRA*03   .......... ....C..... .......... ......G... ..........
Tutr-DRA*04   .......... .......... .......... .......... ..........
Tutr-DRA*05   .......... .T........ .......... .......... ..........

Tutr-DRA*01a  CATCATTGTT GGGACGATCT TCATC [675]
Tutr-DRA*01b  .......... .......... ..... [675]
Tutr-DRA*01d  .......... .......... ..... [675]
Tutr-DRA*02   .......... .......... ..... [675]
Tutr-DRA*03   .......... .......... ..... [675]
Tutr-DRA*04   .......... .......... ..... [675]
Tutr-DRA*05   .......... .......... ..... [675]
```

Figure 9 continued.

```
Tutr-DRB*08  TATCAGTTCA AGGCCGAGTG TCGTTTCTCT AATGGGACAG AGCGGGTGCG[50]
Tutr-DRB*09  ........T. ..TT...... ..A....... .......... ..........
Tutr-DRB*10  ......G.A. ...G...... ..A....... .......... ..........
Tutr-DRB*11  A.......T. .......... ..A....... .......... ..........
Tutr-DRB*12  ........T. ...G...... ..A....... .......... ..........
Tutr-DRB*13  ........T. ..T....... ..A....... .........C ..........
Tutr-DRB*14  ........T. .......... ..A....... .......... ..........
Tutr-DRB*15  A.......T. ..T..T.C.. ..AC...... .......... ..........
Tutr-DRB*16  ........A. ...G...... ..A....... .......... ..........
Tutr-DRB*17  ........T. .......... ..A....... .......... ..........
Tutr-DRB*18  ........A. .......... .......... .......... ..........
Tutr-DRB*19  ........T. ...G...... ..A....... .......... ..........

Tutr-DRB*08  GTTCGTGGTG AGACGCATGT ATAACCGGGA GGAGTACGTG CGCTTCGACA[100]
Tutr-DRB*09  .C..T....T ...GA...C. ......A... ....C..... ....A.....
Tutr-DRB*10  .C..T....T ...GA...C. ......A... ....C..... ....A.....
Tutr-DRB*11  ..A.T..A.T ...GA...C. ......A... ....C.G... ....A.....
Tutr-DRB*12  .C..T....T ...GA...C. .......... .......... ....A.....
Tutr-DRB*13  ..A.T....T ...GA.T.C. ......A... ....C.G... ..........
Tutr-DRB*14  .C..T....T ...GA...C. .......... ....AC.C.. ....A.....
Tutr-DRB*15  ........A. .A..A..C.. .......... .......... ..........
Tutr-DRB*16  .C..T....T ...AA...C. .......... .....T.... ....A.....
Tutr-DRB*17  ..A.T....T ...GA...C. ......A... ....C..... ....A.....
Tutr-DRB*18  .......... .......... .......... .....T.C.. ..........
Tutr-DRB*19  .GC.T....T ...GA...C. ......A... ....C..... ....A.....

Tutr-DRB*08  GCGACGTGGG CGAGTACCGG GCGGTGACCG AGCTGGGCCG GCGGACCGCC[150]
Tutr-DRB*09  .......... .......... .......... .......... .C.GA.....
Tutr-DRB*10  .......... .......... .......... .......... .C.GA.....
Tutr-DRB*11  .......... .......... .......... .......... .C.GA.....
Tutr-DRB*12  .......... .......... .......... .......... ..........
Tutr-DRB*13  .......... .......... .......... .......... .C.GA.....
Tutr-DRB*14  .......... .......... .......... ......C... .A.GAG....
Tutr-DRB*15  .......... .......... ........A. .......... .C.GA.....
Tutr-DRB*16  .......... ...C...... .......... .......... .C..T.....
Tutr-DRB*17  .......... .......... .......... .......... .C.GA.....
Tutr-DRB*18  .......... .......... .......... .......... ..A.......
Tutr-DRB*19  .......... .......... .......... .......... .C.GA.....

Tutr-DRB*08  GAGTACTGGA ACAGCCAGAA GGACCTCCTG GAGCGGCACC GGGCCGAGGT[200]
Tutr-DRB*09  .......... .......... .......... ....A.AGA. ..........
Tutr-DRB*10  .......... .......... .......... ......AGA. ..........
Tutr-DRB*11  .......... .......... .......... ....A.A... ..........
Tutr-DRB*12  ...A.GC... .......... .......... ......AGA. ..........
Tutr-DRB*13  .......... .......... .......... ....A.AGA. .....A....
Tutr-DRB*14  ......C... .......... .......... ....A.A... ..........
Tutr-DRB*15  .......... .......... .......... ....A.A.A. A....A....
Tutr-DRB*16  ...A.GCT.. .......... .......... ......AGA. ..........
Tutr-DRB*17  .......... .......... .......... .C..A.AGA. ..........
Tutr-DRB*18  ...A...... .......... ...T...... ....A.G.A. A.....G...
Tutr-DRB*19  .......... .......... .......... ......AGA. ..........

Tutr-DRB*08  GGACACGGTG TGCAGACACA ACTACAGGGT TGTGGAGA [238]
Tutr-DRB*09  .......... .......... .....G.... ..G..... [238]
Tutr-DRB*10  .......... .......... .......... ..G..... [238]
Tutr-DRB*11  .......TAC ......T... .......... ........ [238]
Tutr-DRB*12  .......... .........G .......... ..G..... [238]
Tutr-DRB*13  .......TAC .......... .....G.... ........ [238]
Tutr-DRB*14  .......TAC .......... .....G...G ........ [238]
Tutr-DRB*15  .......TAC .......... .....G.... ........ [238]
Tutr-DRB*16  .......... .......... .....G.... ..G..... [238]
Tutr-DRB*17  .......... .......... .....G...G ........ [238]
Tutr-DRB*18  .......TAC .......... .......... ..AT...C [238]
Tutr-DRB*19  .......... .......... .......... ..AT...C [238]
```

Figure 10: Nucleotide sequence alignment of 12 DRB exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico (GenBank accession nos. KJ722698-KJ722709). Dots indicate nucleotide identity with Tutr-DRB*08. Underlined sites are part of the putative peptide-binding region. Reading frame begins at position 1.

```
                  ↓                     ↓   ↓   ↓       ↓ ↓                   ↓                ↓   ↓↓↓   ↓
Tutr-DQA*02  TTIYQSFGPS GQYTQEFDGD ELFYVDLEKK ETVWRLPVFS EFAGFDPQGA [50]
Tutr-DQA*03  .......... ....H..... .K........ .......... ...S....R.
Tutr-DQA*07  .......... .......... .......... .......... ..TS......
Tutr-DQA*11  .......... .......... .......... .......... ..........
Tutr-DQA*12  .......... .......... .......... .......... ..........
Tutr-DQA*13  .......... ..F.T..... .......... .......... ...S......
Tutr-DQA*14  .......... ....H..... .......... .......... ...S......
Tutr-DQA*15  .......... .......... .......... .......... ..........
Tutr-DQA*16  .......... ..F.T..... .......... .......... ..TS......
Tutr-DQA*17  .......... ....H..... .Q........ .......... ...S......

Tutr-DQA*02  LRNTAVAKQN LDILIKRSNF TPVTN [75]
Tutr-DQA*03  ...I..V.H. .......... .....
Tutr-DQA*07  ...I..V.H. .......... .....
Tutr-DQA*11  ...I..V.H. .......... .....
Tutr-DQA*12  ...I..V.H. ......C... .....
Tutr-DQA*13  ...I....YT ...M...... .....
Tutr-DQA*14  ...I....RT ...M..C... .....
Tutr-DQA*15  ...I..V.H. ...M...... .....
Tutr-DQA*16  ...I....RT ...M...... .....
Tutr-DQA*17  ...I....H. ...M...... .....
```

Figure 11: Amino acid sequence alignment of 10 DQA exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico. Dots indicate amino acid identity with Tutr-DQA*02. Arrows above sequence indicate sites that are part of the peptide-binding region.

```
Tutr-DQB*25     TERVRFVDRY IYNREEFVRF DSDVGEHRAV TELGRPDAEY WNSQKDILER KRAELDT [57]
Tutr-DQB*26     .....AMS.. ......Y... ......F... .....RT... ....E....E E..AV..
Tutr-DQB*27     .....V...S ......L... ......Y... .......... .......... .......
Tutr-DQB*01/12  .....GMS.. ......Y... ......F... .......... F........Q E..AV..
Tutr-DQB*04/14  .....L.E.. .......... ......F... .........T ..G.E....E E..A...
Tutr-DQB*05     .....H.S.. ......Y... ......F... .........K ....E..... ...DV.R
Tutr-DQB*06     .....V...S ......L... ......F... .......... .......... .......
Tutr-DQB*07     .......... .......... ......F... .........H .......... EW.DG..
Tutr-DQB*08     .......... ......Y... .......... .......... ......F... .......
Tutr-DQB*09/22  .....L.S.N ......L... ......F... .....RA... ...R..Y..L R...V..
Tutr-DQB*13/24  .....H.S.. ......Y... ......F... .......... F........E E..DV.R
```

Figure 12: Amino acid sequence alignment of 11 DQB exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico. Some sequences assigned multiple allele numbers due to redundancy in GenBank or variation among GenBank sequences in intron sequence. Dots indicate amino acid identity with Tutr-DQB*25. Arrows above sequence indicate sites that are part of the peptide-binding region.

```
Tutr-DRA*01  V T V L S N T P V E  L G E P N V L I C F  I N K F S P P V I N  V T W L Q N G K P V  T T G V S E T V F L [50]
Tutr-DRA*02  . . . . . . . . . .  . . . . . . I . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*03  . . . . . . . . . .  . . . . . . I . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*04  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*05  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .

                                                                                               ¶
Tutr-DRA*01  P R N D H L F R K F  H Y L P F V P S A E  D V Y D C K V E H W  G L E E P L L K H W  E Y E A P T P L T E [100]
Tutr-DRA*02  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*03  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*04  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRA*05  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .

Tutr-DRA*01  T T E N A V C A L G  L V M A L V G I I V  G T I F I 25]
Tutr-DRA*02  . . . . . . . . . .  . . . . . . . . . .  . . . . .
Tutr-DRA*03  . . . . . . . . . .  . . V . . . . . . .  . . . . .
Tutr-DRA*04  . . . . . . . . . .  . . . . . . . . . .  . . . . .
Tutr-DRA*05  . . . . . . . . . .  . . . . . . . . . .  . . . . .
```

Figure 13: Amino acid sequence alignment of 5 DRA alleles spanning exons 3-4 identified in bottlenose dolphins from the Gulf of Mexico. Dots indicate amino acid identity with Tutr-DRA*01. ¶ indicates boundary between exons 3 and 4.

```
             ↓   ↓   ↓                                    ↓    ↓   ↓       ↓ ↓                   ↓                   ↓
Tutr-DRB*08  Y Q F K A E C R F S  N G T E R V R F V V  R R M Y N R E E Y V  R F D S D V G E Y R  A V T E L G R R T A [50]
Tutr-DRB*09  . . . . F . . H . .  . . . . . . . L L .  . D I . . Q . . H .  . Y . . . . . . . .  . . . . . . . P D .
Tutr-DRB*10  . . V . G . . H . .  . . . . . . . L L .  . D I . . Q . . H .  . Y . . . . . . . .  . . . . . . . P D .
Tutr-DRB*11  N . . . . . . H . .  . . . . . . . Y L I  . D I . . Q . . Q .  . Y . . . . . . . .  . . . . . . . P D .
Tutr-DRB*12  . . . . G . . H . .  . . . . . . . L L .  . D I . . . . . . .  . Y . . . . . . . .  . . . . . . . . . .
Tutr-DRB*13  . . . . S . . H . .  . . . Q . . . Y L .  . D F . . Q . . Q .  . . . . . . . . . .  . . . . . . . P D .
Tutr-DRB*14  . . . . . . . H . .  . . . . . . . L L .  . D I . . . . . T L  . Y . . . . . . . .  . . . . . A . Q E .
Tutr-DRB*15  N . . . S Y . H . .  . . . . . . . . . E  K H T . . . . . . .  . . . . . . . . . .  . . . K . . . P D .
Tutr-DRB*16  . . L . G . . H . .  . . . . . . . L L .  . N I . . . . . F .  . Y . . . . . H . .  . . . . . . . P I .
Tutr-DRB*17  . . . . . . . H . .  . . . . . . . Y L .  . D I . . Q . . H .  . Y . . . . . . . .  . . . . . . . P D .
Tutr-DRB*18  . . L . . . . . . .  . . . . . . . . . .  . . . . . . . . F L  . . . . . . . . . .  . . . . . . . . . .
Tutr-DRB*19  . . . . G . . H . .  . . . . . . . A L .  . D I . . Q . . H .  . Y . . . . . . . .  . . . . . . . P D .

               ↓ ↓       ↓    ↓     ↓ ↓    ↓     ↓     ↓ ↓    ↓ ↓
Tutr-DRB*08  E Y W N S Q K D L L  E R H R A E V D T V  C R H N Y R V V E [79]
Tutr-DRB*09  . . . . . . . . . .  . Q R . . . . . . .  . . . . . G . G .
Tutr-DRB*10  . . . . . . . . . .  . . R . . . . . . .  . . . . . . . G .
Tutr-DRB*11  . . . . . . . . . .  . Q N . . . . . Y .  . Y . . . . . . .
Tutr-DRB*12  . K R . . . . . . .  . . R . . . . . . .  . . . D . . . G .
Tutr-DRB*13  . . . . . . . . . .  . Q R . . K . . . Y  . . . . . G . . .
Tutr-DRB*14  . . R . . . . . . .  . Q N . . . . . . Y  . . . . . G G . .
Tutr-DRB*15  . . . . . . . . . .  . Q K Q . K . . . Y  . . . . . G . . .
Tutr-DRB*16  . K L . . . . . . .  . . R . . . . . . .  . . . . . G . G .
Tutr-DRB*17  . . . . . . . . . .  A Q R . . . . . . .  . . . . . G G . .
Tutr-DRB*18  . N . . . . . . F .  . Q E Q . G . . . Y  . . . . . . . D .
Tutr-DRB*19  . . . . . . . . . .  . . R . . . . . . .  . . . . . . . D .
```

Figure 14: Amino acid sequence alignment of 12 DRB exon 2 alleles identified in bottlenose dolphins from the Gulf of Mexico. Dots indicate amino acid identity with Tutr-DRB*08. Arrows above sequence indicate sites that are part of the peptide-binding region.

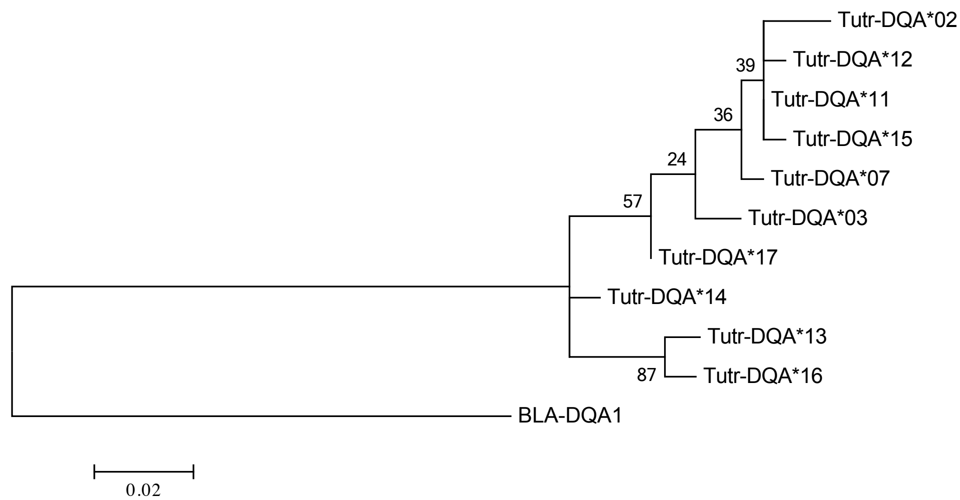

Figure 15. Maximum likelihood tree of bottlenose dolphin DQA exon 2 alleles. Evolutionary history was inferred using the Kimura 2-parameter model with a discrete Gamma distribution with five categories to model evolutionary differences among sites. Tree was rooted using bovine BLA-DQA1 (GenBank: NM_001013601) as an outgroup. Bootstrap values (percentage of 1,000 replicates) are displayed next to branches. Branch lengths represent number of substitutions per site.

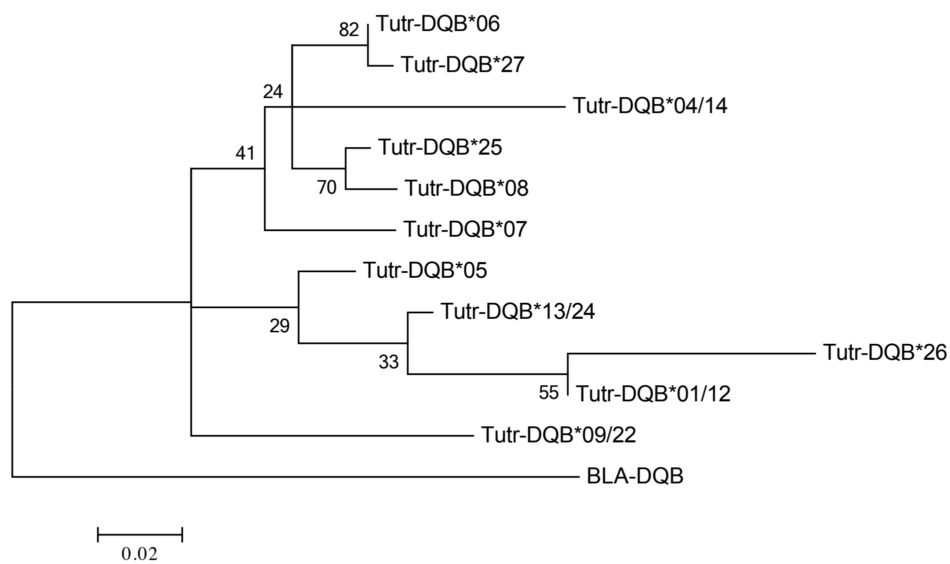

Figure 16. Maximum likelihood tree of bottlenose dolphin DQB exon 2 alleles. Evolutionary history was inferred using the Tamura 3-parameter model with a discrete Gamma distribution with five categories to model evolutionary differences among sites. Tree rooted using bovine BLA-DQB (GenBank: NM_001034668) as an outgroup. Bootstrap values (percentage of 1,000 replicates) are displayed next to branches. Branch lengths represent number of substitutions per site.

Figure 17. Maximum likelihood tree of bottlenose dolphin DRA alleles spanning exons 3-4 and the intermediate intron. Evolutionary history was inferred using the Kimura 2-parameter model with uniform rates among sites. Tree rooted using bovine BLA-DRA (GenBank: M30120) as an outgroup. Bootstrap values (percentage of 1,000 replicates) are displayed next to branches. Branch lengths represent number of substitutions per site.

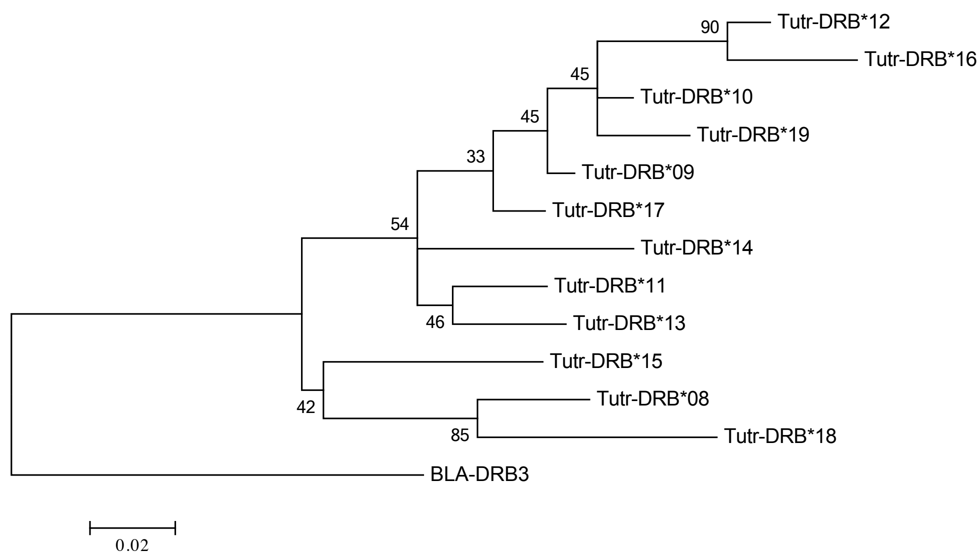

Figure 18. Maximum likelihood tree of bottlenose dolphin DRB exon 2 alleles. Evolutionary history was inferred using the Jukes-Cantor model with a discrete Gamma distribution with five categories to model evolutionary differences among sites. Tree rooted using bovine BLA-DRB3 (GenBank: AB558425) as an outgroup. Bootstrap values (percentage of 1,000 replicates) are displayed next to branches. Branch lengths represent number of substitutions per site.

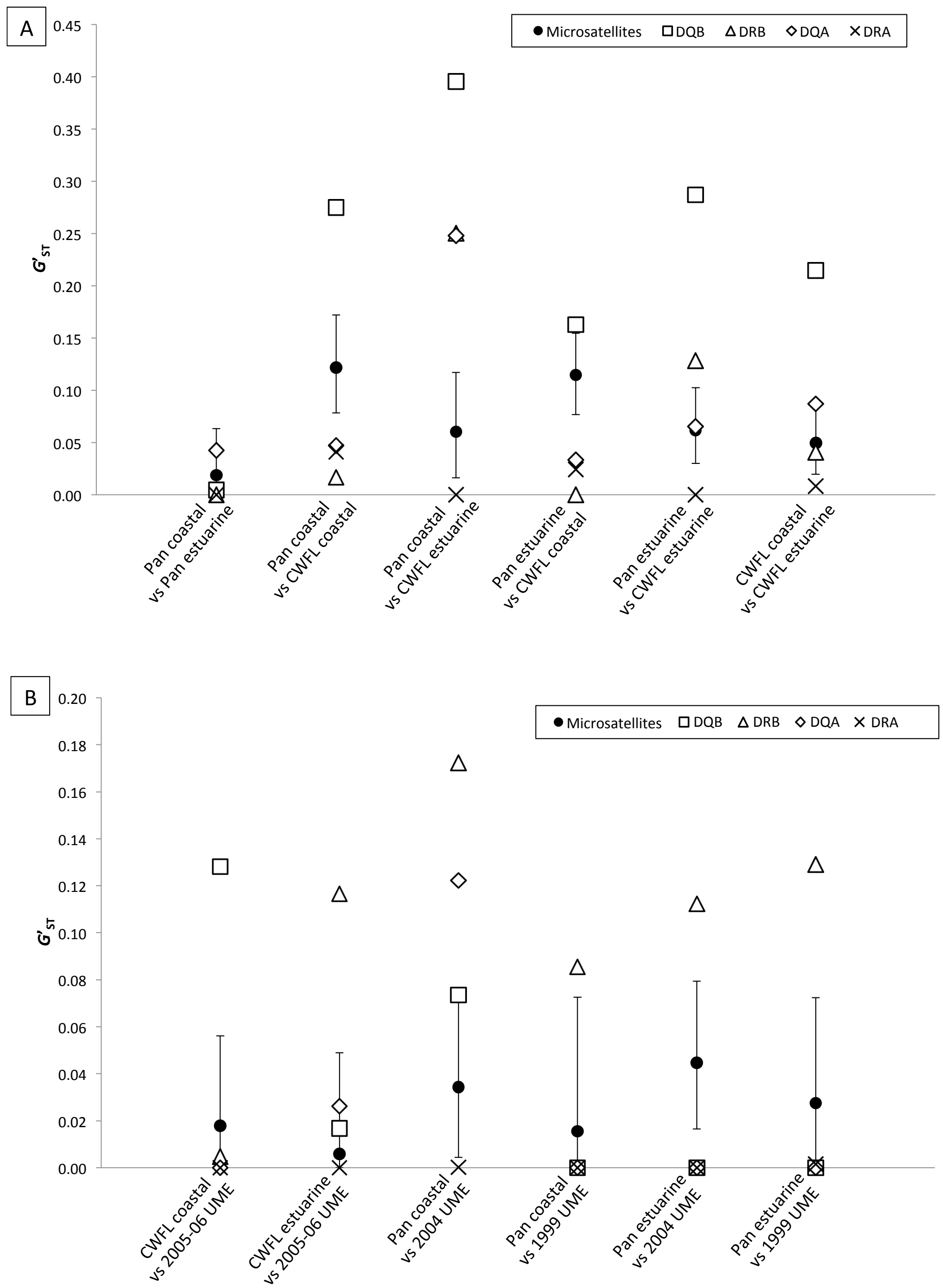

Figure 19: Genetic differentiation, G'_{ST}, between A) live estuarine and coastal dolphins from central-west Florida (CWFL) and the Florida Panhandle (Pan) and B) live and dead dolphins from UMEs associated with red tides. Pairwise G'_{ST} was averaged across 19 microsatellites and calculated independently for four MHC loci. Error bars around microsatellite G'_{ST} represent the 95% confidence interval (1,000 bootstraps by locus).

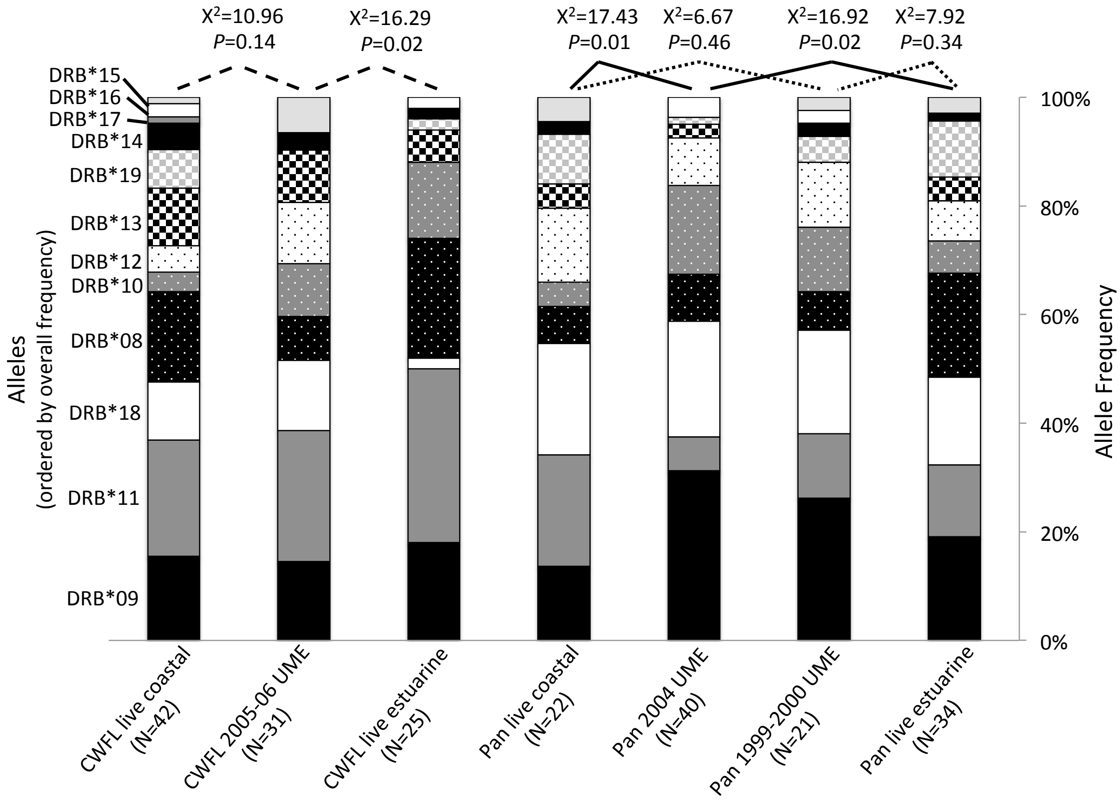

Figure 20: Frequency of DRB alleles in bottlenose dolphin samples. Connecting lines above columns indicate pairs of dolphin groups tested for differences in allele frequency (χ^2 and *P* values). Live dolphins from coastal and estuarine populations were compared to dolphins that died during UMEs associated with red tides in central-west Florida (CWFL) in 2005-2006 (dashed line) and in the Florida Panhandle (Pan) in 2004 (solid line) and 1999-2000 (dotted line).

Discussion

Previously described associations between genetic variation and survival in bottlenose dolphins exposed to HABs suggested that variation in the immunologically relevant MHC genes may play a role in variation in brevetoxin resistance in bottlenose dolphins (Chapter 1). In my investigation of this hypothesis in dolphins from central-west Florida and the Florida Panhandle, I found that some MHC loci were under balancing selection and exhibited patterns of genetic differentiation among geographic regions that differed from neutral loci. However, genetic variation at the four MHC loci studied here could not fully explain variation in survival in bottlenose dolphins exposed to HABs.

MHC diversity in bottlenose dolphins

The diversity identified in MHC loci in bottlenose dolphins in the Gulf of Mexico is comparable to variation described in other cetaceans and some terrestrial mammals. Early studies suggested that marine mammals had generally low MHC diversity (Murray *et al.* 1995; Slade 1992; Trowsdale *et al.* 1989), but a number of studies in the last 15 years have identified MHC diversity in cetaceans comparable to that described in terrestrial mammals (Baker *et al.* 2006; Murray & White 1998; Xu *et al.* 2010; Xu *et al.* 2009). These recent studies suggest that the hypotheses initially proposed to limit MHC diversity, including lower pathogen presence in marine environments (Slade 1992) and demographic factors, (e.g., founder effects and bottlenecks), cannot be applied across all marine mammals. Additionally, although Slade (1992) initially questioned the functionality of the MHC in marine mammals due to his observations of low diversity, recent studies have confirmed expression of DQA, DQB and DRB in dolphins (Heimeier *et al.* 2009; Yang *et al.* 2007), and almost all studies find evidence of balancing selection (e.g., high d_N/d_S ratios (Hughes & Nei 1989) and transpecific polymorphisms (Klein 1987)) at MHC loci in cetaceans.

A recent review of MHC diversity at the DQB exon 2 across 33 cetacean species found that MHC in cetaceans is under weaker balancing selection than in artiodactyls, but there is also greater variation in selection among cetaceans, with some species (e.g., mysticetes) showing evidence of stronger selection at DQB than others (e.g., ziphiids, monodontids, sperm whale, finless porpoise) (Villanueva-Noriega *et al.* 2013). Delphinids were found in both high (short-finned pilot whale, North Pacific white-sided dolphin) and low (Hector's dolphin, bottlenose dolphin) selection groups. Overall, it appears that the MHC plays a similarly important immunological role in marine and terrestrial mammals.

DRB was the most variable locus in bottlenose dolphins in the Gulf of Mexico, followed by DQB, DQA, and finally DRA, which showed little variation. These results are in accordance with previous studies in humans that found greatest selection intensity on DRB, followed by DQB and then DQA (Satta *et al.* 1994). DRA exon 2 is monomorphic in humans (Hughes & Nei 1989). Similar patterns of diversity among MHC loci have also been observed in other marine and terrestrial mammals (e.g., Berggren *et al.* 2005; Heimeier *et al.* 2009; Munguia-Vega *et al.* 2007; Niranjan *et al.* 2010; Xu *et al.* 2008; Xu *et al.* 2007; Xu *et al.* 2009).

The least variable locus, DRA, was previously found to differ in expression among dolphin populations in the coastal waters of the northwest Atlantic Ocean and Gulf of Mexico (Mancia *et al.* 2010). In contrast, the more variable DQA did not differ in expression among dolphin populations; expression of DQB and DRB was not evaluated in this study. These findings suggest, as has been described elsewhere, that variation in gene expression is not necessarily linked to genetic variation in the coding region itself (Vasemagi & Primmer 2005). Variation in promotor regions or other genes upstream in the MHC pathway may be responsible for the observed variation in DRA expression, which could also impact immunological function and disease resistance (Ting & Trowsdale 2002).

No locus exhibited evidence of duplication in bottlenose dolphins from the Gulf of Mexico, although Yang *et al.* (2007) described multiple DRB loci in bottlenose dolphins (*T. truncatus* and *T. aduncus*) from the Western Pacific. This inconsistency may be due to methodological differences between the two studies, including different primers and starting material (e.g., DNA vs. RNA), or reflect real differences in the number of duplicated loci among populations of the same species, as has been observed for DQB in the finless porpoise (Du *et al.* 2010; Xu *et al.* 2007).

The pattern of nonsynonymous and synonymous mutations among MHC alleles confirms that diversity at almost all MHC loci in bottlenose dolphins is likely the result of balancing selection. Transpecific polymorphisms (i.e., identical alleles in other species) were identified for only the DQB locus, probably because it is currently the best described MHC class II locus in cetaceans. These transpecific polymorphisms likely resulted from the conservation of allelic variation over evolutionary time (Klein 1987), although introgressive hybridization cannot be ruled out (Caballero & Baker 2010; Sylvestre & Tasaka 1985).

Effects of geographic region and survival

Most studies of cetacean MHC to date have focused solely on describing genetic diversity at these immunologically important genes. Several studies suggest that differences in MHC diversity among populations may reflect differences in pathogen pressure (Du *et al.* 2010; Murray *et al.* 1995; Vassilakos *et al.* 2009; Xu *et al.* 2010), but no study has directly tested for an association between MHC diversity and disease susceptibility in cetaceans. In one case, researchers found a marginally significant difference in DQB allele frequency between stranded and bycaught bottlenose dolphins in Taiwan, but could not evaluate specific causes of death (Yang *et al.* 2008).

I found that MHC diversity in bottlenose dolphins in the Gulf of Mexico was significantly affected by both geographic region and survival following red tides. Regional differentiation at neutral markers is expected given genetic differentiation previously described among dolphin populations in the Gulf of Mexico (Sellas *et al.* 2005). However, genetic differentiation between groups of live samples from central-west Florida and Panhandle populations at both DRB and DQB loci was significantly different from that observed across microsatellite markers (Fig. 19). Interestingly, these two MHC loci exhibited different patterns of differentiation, suggesting that selection may be acting differently on the two loci. DQB variation differed more than expected based on neutral markers in all comparisons between the two regions. In contrast, DRB variation differed in comparisons of each of the Florida Panhandle populations to only the estuarine but not coastal central-west Florida population. This may be consistent with limited mortality of the estuarine Sarasota Bay dolphin population during the 2005-2006 UME (pers. comm. R. Wells).

Geographic variation in MHC diversity may be attributed to spatial differences in pathogen pressure, such as from the brevetoxin-producing harmful algae, *K. brevis*. I therefore evaluated the role of MHC diversity in survival of bottlenose dolphins exposed to HABs across six comparisons of live and dead dolphins. Dead dolphins represented three UMEs associated with HABs, one from central-west Florida and two from the Florida Panhandle. Dolphins from each UME were compared to live dolphins from coastal and estuarine dolphin populations from the same geographic region. Genetic differentiation between live and dead dolphins was significantly greater at the DRB locus than microsatellites for almost all live-dead comparisons. This finding suggests a potential role of the DRB in variation in survival of bottlenose dolphins exposed to HABs; however, neither heterozygosity nor a single allele appears to induce resistance to brevetoxins.

Link to genome-wide study

The MHC loci described above were targeted for sequencing in these dolphin populations to investigate functional variation at candidate genes surrounding a previously identified survival-associated RAD locus, which exhibited nearly fixed SNP differences between live and dead dolphins in a similar set of samples. However, no MHC locus approached this nearly fixed difference in allele frequency among dolphin groups. It is possible that variation within the RAD locus is linked to variation elsewhere in the MHC or to other nearby genes. The former seems relatively unlikely considering that I investigated the region of MHC genes generally under greatest balancing selection, the PBR (Hughes & Nei 1989), in three highly polymorphic MHC loci. There are no other annotated genes in the bottlenose dolphin genome on the scaffold containing the survival-associated RAD locus. However, the dolphin genome is not yet fully annotated and the scaffolds are relatively short, suggesting that there could be other genes in the region that I did not sequence.

Alternatively, the nearly fixed differences at the RAD locus could be a technical artifact due to biased sequencing of only a few individuals in a pooled sample. During RAD library preparation, samples were pooled prior to barcoding and allele frequencies were determined by the count of SNP variants in the pooled sample, rather than from individually determined genotypes. This method has been shown previously to produce robust estimates of allele frequency (Futschik & Schlötterer 2010; Van Tassell *et al.* 2008; Zavodna *et al.* 2013), but this was not confirmed in my study. Confirmation would require re-amplification and sequencing of the RAD locus in individual dolphins. If the association between SNP variation and survival was confirmed, further studies could investigate variation in additional MHC exons or other candidate genes identified by alignment of the RAD locus to a more complete and better annotated genome (e.g., cow).

Although next-generation sequencing has greatly advanced our abilities to conduct association studies, this study highlights some of the difficulties we still face in applying these techniques to non-model organisms. GBS facilitates the identification of SNPs associated with phenotype (e.g., disease resistance) across the genome, but as seen in Chapter 1, this variation is most often located in non-coding regions. Following genome-wide analyses, candidate gene studies may still be required to link phenotype-associated SNPs with nearby functional variation that offers biologically relevant predictive power. However, candidate gene studies face numerous logistical limitations that limit the number of genes that can be feasibly investigated (Tabor *et al.* 2002). Of particular relevance to studies of non-model organisms, post-GBS candidate gene studies require clearly annotated genomes. As the price of genome sequencing continues to decline, whole genome re-sequencing (Bentley 2006) may offer a more effective and attractive approach to association studies, one that combines the goals of genome-wide GBS and candidate gene sequencing.

CHAPTER IV:
COMPLEMENTARY GENOME-WIDE AND CANDIDATE GENE APPROACHES TO THE STUDY OF ADAPTATION IN NON-MODEL ORGANISMS

There is a growing interest in better understanding how genes that underlie ecologically important traits affect fitness of individuals in their natural environment (Feder & Mitchell-Olds 2003). Research into adaptation relies on the investigation of genotype-phenotype associations, which have been traditionally studied with quantitative trait loci (QTL) (Complex Trait Consortium 2003) and candidate gene sequencing (Tabor *et al.* 2002). However, both approaches are somewhat limited in their application to studies of non-model organisms in the natural environment. QTL studies require controlled crosses or known pedigrees, and candidate gene sequencing has primarily focused on a small number of genes (Piertney & Webster 2010).

Recently, technological advances in genomic sequencing have enabled new avenues of molecular research in ecological adaptation. The development of single nucleotide polymorphism (SNP) arrays, particularly for humans, has largely replaced QTL and candidate gene approaches with genome-wide association studies (GWAS) (Hirschhorn & Daly 2005). GWAS provide a relatively unbiased scan of the genome for SNP variation associated with phenotype and are particularly powerful in systems with highly developed genomic resources, including annotation. In contrast to QTL and candidate gene studies, GWAS require no pedigree information or *a priori* hypothesis of the biological pathways involved in disease susceptibility. GWAS can therefore lead to the discovery of significant novel or unexpected genes.

Genome-wide analyses are becoming increasingly feasible in non-model organisms in the natural environment (Tautz *et al.* 2010). Next-generation, or massively parallel, DNA sequencing has drastically decreased the cost per bp of sequencing (Sboner *et al.* 2011), facilitating new genome sequencing projects and the application of these techniques not only to individuals, but also to whole populations. Genomic techniques enable whole genome sequencing, large-scale marker-based genotyping (i.e., SNP arrays) and genome-wide genotyping-by-sequencing (GBS) (Allendorf *et al.* 2010). Rather than relying on previously characterized SNPs, GBS techniques allow for simultaneous identification and genotyping of SNPs in organisms with minimal genomic resources (Davey *et al.* 2011).

These primarily SNP-based studies deviate from the prominent use of microsatellites over the past decade theoretically, by differences in mutation model (Ellegren 2004; Morin *et al.* 2009), and logistically, by a shift in laboratory techniques and computational analyses. In some cases, the shift from microsatellites to genome-wide studies simply increases the number of evaluated markers, improving our power to describe neutral population parameters (Allendorf *et al.* 2010; Hess *et al.* 2011; Leffler *et al.* 2012). In other cases, genome-wide studies facilitate new research, particularly within studies of adaptation where genome-wide markers enable the simultaneous investigation of both neutral and functional variation across the genome (Allendorf *et al.* 2010).

In this final chapter of my dissertation, I reflect on the promise of genomic sequencing for studies of non-model organisms, specifically in investigations of adaptation and genotype-phenotype associations. Drawing primarily on my case study of a genetic basis for resistance to red tides in bottlenose dolphins, I consider the advantages and drawbacks of the genome-wide approach in comparison to a candidate gene approach. Although some researchers have begun to question the need for candidate gene sequencing given recent advances in genomic sequencing (Wilkening *et al.* 2009), others recommend using genome-wide approaches in various combinations with other association mapping techniques (Kim *et al.* 2010; Nadeau & Jiggins 2010; Stinchcombe & Hoekstra 2008). Similarly, I suggest using the complementary approaches of GWAS and candidate gene sequencing in combination for a more powerful investigation of adaptation in genome-enabled non-model organisms. In addition to the genome-wide approach, I discuss the need for both post-GWAS and GWAS-independent candidate gene studies (Fig. 21).

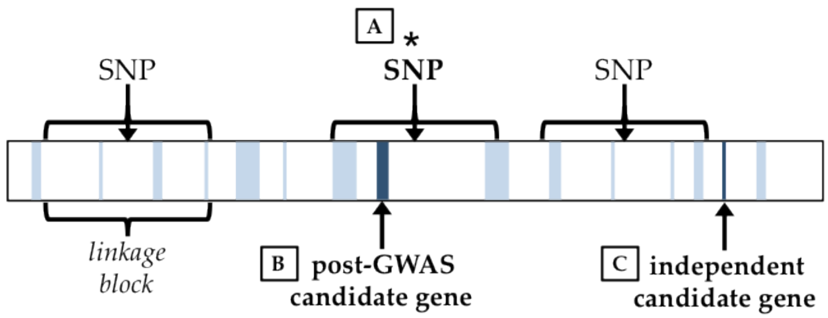

Figure 21: Three approaches to the study of genetic variation that underlies ecologically important traits. A) The genome-wide approach identifies and genotypes SNPs located throughout the genome, primarily in non-coding regions. A SNP exhibiting a significant association with phenotype is demarcated with an asterisk. B) Following a genome-wide association study (GWAS), candidate genes located in putative linkage blocks surrounding a phenotype-associated SNP can be further investigated. Shaded regions represent genes. C) Additional candidate genes can be selected for sequencing independent of the genome-wide approach and based on an *a priori* hypothesis of genotype-phenotype association.

Defining non-model organisms

Prior to discussing the techniques, it is important to understand the group of organisms to which these approaches may be applied. The definition of a "non-model" organism has begun to change as next-generation sequencing technologies enable genomic research across species. Traditionally, model organisms were defined by their ability to be studied in a lab (e.g., short generation time and large numbers of offspring), their similarity to humans, or their economic importance. More recently, genomic model organisms have been selected primarily for their small genome size (e.g., pufferfish (Brenner *et al.* 1993)), and non-model organisms are considered those with no or very limited genomic resources (Ekblom & Galindo 2011; Seeb *et al.* 2011). I consider the criteria of no reference genome too conservative in this era of rapidly growing numbers of sequenced genomes. Instead, I merge the definitions of ecological and genomic model organisms and consider non-model organisms to be those that are difficult or impossible to experimentally manipulate and that have relatively limited genomic resources (e.g., incomplete genome sequencing or annotation). Organisms with reference genomes, or organisms for which there are closely related species with reference genomes, are considered genome-enabled (Kohn *et al.* 2006). This designation may apply to non-model organisms as I have defined them above if the reference genome is incomplete in sequence or annotation.

By these definitions, the common bottlenose dolphin (*Tursiops truncatus*) is considered a genome-enabled non-model organism. It is logistically, legally, and ethically impossible to manipulate the natural environment of bottlenose dolphins and, although the bottlenose dolphin genome was the first marine mammal genome to be sequenced (Lindblad-Toh *et al.* 2011), genomic resources for this species remain incomplete. The first release of the bottlenose dolphin genome (turTru1.0, Ensembl release 71, GenBank Assembly ID: GCA_000151865.1) was sequenced at low coverage (2.59x) using Sanger sequencing, and annotated using the Ensembl gene-build pipeline involving whole genome alignment to the Ensembl-annotated human genome. There are currently 16,550 annotated coding genes in the bottlenose dolphin genome, but due to the low coverage of this genome, this number is likely an underestimate. The most recent bottlenose dolphin genome (Ttru_1.4/turTru2, GenBank Assembly ID: GCA_000151865.2) has been supplemented with 3.5x 454 sequencing and 30x Illumina sequencing. This genome has been assembled into 240,901 scaffolds with an N50 of

116,261 bp, but has not yet been annotated. Scaffolds, which may contain gaps in sequence, are reconstructed portions of the genome that are built from alignments of short sequence reads based on sequence similarity and known distances between sequences (e.g., end-sequenced clones of known length). An N50 score describes the scaffold length for which 50% of all bases are found in scaffolds of that length or greater.

The publicly available bottlenose dolphin genome greatly enables studies such as mine of the molecular ecology of bottlenose dolphins, but the incomplete annotation and large number of scaffolds limit some genomic analyses. To complete the bottlenose dolphin genome, further sequencing and transcriptomics would be required to fill in sequence gaps, align scaffolds into larger chromosomes, and identify additional genes across the genome. Low coverage genomes are becoming increasingly common for non-model organisms, such as the bottlenose dolphin, but few non-model genomes are completely sequenced and annotated at this point.

Genome-wide approach

GBS techniques, which permit simultaneous identification and genotyping of thousands of SNPs across the genome, are particularly popular in studies of non-model organisms because they do not require a reference genome (although adaptation studies are limited in non-genome-enabled organisms; see *Anonymity*). Several authors have predicted that GBS techniques will soon dominate SNP genotyping studies for population genomics of non-model organisms, but they recognize the value of SNP arrays for long-term studies (Allendorf *et al.* 2010; Seeb *et al.* 2011). Here, GBS refers to multiple protocols that utilize reduced representation libraries, including restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008; Elshire *et al.* 2011; Van Tassell *et al.* 2008). GBS investigates only a portion of the genome, primarily regions that flank restriction enzyme recognition sites. These small fractions of the genome contain thousands of SNPs, which are sufficient for many analyses (Perry 2014). GBS can be used in studies of population structure, phylogeography, phylogenetic reconstruction, and association mapping in a variety of organisms (reviewed in Ekblom & Galindo 2011; Narum *et al.* 2013).

The primary advantage to genome reduction techniques, such as GBS, is the high coverage that can be obtained from a standardized subset of the genome across multiple individuals. Sequencing less of the genome per individual enables multiplexing of individuals into a single sequencing run (see *Samples*) while still producing the read depth necessary for SNP discovery (see *Coverage*). Coverage, or read depth, refers to the number of sequencing reads that map to a single genomic location in each sample. This genome reduction strategy addresses a common tradeoff that must be considered in next-generation sequencing studies, to sequence a greater number of individuals at low to moderate coverage (5-20x) or fewer individuals at high coverage (>20x) (Nielsen *et al.* 2011). Sample selection and coverage, two important components of designing a successful genome-wide study of adaptation in non-model organisms, are discussed in further detail in the following sections.

Samples – quantity & quality

GBS techniques typically require between 100 ng and 1 μg of high quality genomic DNA per sample if building your own library (Elshire *et al.* 2011; Etter *et al.* 2011), or up to 3 μg of genomic DNA per sample if using a commercial sequencing facility (e.g., Floragenex, Cornell University Institute for Genomic Diversity). These sample requirements, which are significantly greater than those required for Sanger sequencing (<25 ng, poor to high quality DNA), may be restrictive for some non-model organisms from which it can be difficult to collect sufficient quantity and quality DNA samples. However, this limitation can be overcome by pooling DNA from multiple individuals (see *Multiplexing*) or by using DNA capture techniques prior to sequencing (Mamanova *et al.* 2010). DNA capture can enrich endogenous DNA in a sample through hybridization to target sequences (e.g., from a known reference genome) that allow for removal of non-target DNA (Gnirke *et al.* 2009).

Several studies have now successfully sequenced genomic DNA samples of limited quantity or quality, including samples from stranded marine mammals and samples collected using noninvasive techniques (e.g., fecal samples (Perry *et al.* 2010)). I sequenced pools of 500 ng of genomic DNA from both excellent (high molecular weight DNA with minimal shearing) and good (sheared DNA with some high molecular weight) quality DNA collected from live and dead bottlenose dolphins and did not observe any difference in sequence coverage; however, I did not sequence any poor (sheared DNA with no high molecular weight DNA) quality DNA. DNA quality was determined prior to sequencing on an agarose gel, where shearing (i.e., fragmentation) is visible as a smear below a high molecular weight band of DNA.

Samples – multiplexing

To increase the number of samples in a single sequencing run, genomic DNA from multiple individuals (up to 96, with potential to increase (Elshire *et al.* 2011)) can be ligated to uniquely barcoded adapters that enable individual assignment of sequences following pooled sequencing. Genomic DNA from individuals can also be pooled prior to barcoding based on population or phenotype. For example, I pooled DNA from dolphins based on geographic sampling location, coastal or estuarine distribution, survival (live or stranded), and DNA quality. Pools of DNA included 8 to 35 individuals each. This approach of pooling DNA prior to barcoding may be appropriate for population level comparisons or analyses of genetic variation associated with a specific phenotype (Cheng *et al.* 2012; Huang *et al.* 2012; Kim *et al.* 2010; Spurgin *et al.* 2011; Zavodna *et al.* 2013), particularly when faced with sample or funding limitations.

Kim *et al.* (2010) recommend pooling samples in association studies using next-generation sequencing data because of the increased accuracy of allele frequency estimates in studies with larger numbers of individuals, even when sequenced at relatively low coverage. The accuracy of allele frequency estimation in cases and controls largely determines the power to detect an association. To ensure accurate estimation of allele frequency, great care must be taken in measuring DNA concentrations and adding an equal concentration of DNA from each sample to the pool.

If successful, previous analyses have confirmed that allele frequencies generated from pooled samples correlate highly with genotypes from direct sequencing (Gautier *et al.* 2013; Rellstab *et al.* 2013; Zavodna *et al.* 2013), and under certain conditions, sequencing pooled DNA actually allows for more accurate identification of SNPs and calculation of allele frequencies and population genetic parameters (Futschik & Schlötterer 2010).

Pooling samples prior to barcoding is advantageous for the reasons described above, but there are also limitations that should be considered prior to using this approach. Several software packages commonly used to analyze RAD data (e.g., Stacks (Catchen *et al.* 2013)) are not yet designed for use with pooled samples and so customized analysis pipelines must be created. Additionally, because resulting sequences cannot be demultiplexed to the individual level, the data cannot be used downstream for any analyses that require individual genotypes, such as genetic population assignment or mixed stock analysis.

Coverage & SNP discovery

Sufficient sequence coverage is essential in analyzing GBS data. SNP identification and genotyping requires that one can distinguish between sequencing error (e.g., ~1% for Illumina HiSeq 2000) and true polymorphisms to characterize individuals as homozygotes or heterozygotes. Multiple algorithms have been developed to call SNPs and genotypes from next-generation sequencing data (Nielsen *et al.* 2011; Pabinger *et al.* 2013). SNP and genotype call accuracy can be improved by specifying base quality and read coverage filters. Additionally, likelihood-based methods may be further improved by incorporating priors from well characterized SNP databases, linkage disequilibrium, or genotype probabilities estimated assuming Hardy-Weinberg equilibrium (Nielsen *et al.* 2011). Fortunately, SNP and genotype calls in association studies should be relatively robust to errors as long as the error structure is similar for cases and controls (Nielsen *et al.* 2011).

Sequence coverage across GBS loci can be highly variable. Hypothetically, sequences should be equally distributed across all loci in all individuals, but in reality next-generation sequencing is highly stochastic. Variability in sequence coverage is further amplified by biases due to fragment length variation, restriction site heterozygosity, and PCR GC content (Davey *et al.* 2013). For example, in my study of bottlenose dolphins, I obtained sequences from 129,594 RAD loci that surrounded 70,354 *SbfI-HF* recognition sites in the dolphin genome. Each enzyme recognition site should result in two RAD loci, but 11,114 enzyme recognition sites were not sequenced in both directions. Among loci that were sequenced, there was large variation in read depth; after implementing a filter requiring greater than 15 reads per locus per pooled sample (see below), I obtained an average overall read depth of 88.69 ± 63.45 SD.

Prior to analysis of the bottlenose dolphin RAD sequencing data, I filtered by quality using a minimum average Phred score of 10 (i.e., 90% base call accuracy) across 6 bp sliding windows, and I filtered by minimum read depth to include only RAD loci with greater than 15 reads in at least two sample pools. SNPs were identified using a bounded multinomial-based maximum likelihood model implemented in Stacks

(Catchen *et al.* 2013); I reduced the upper bound on sequencing error rate at each nucleotide position to increase the likelihood of calling a SNP at low minor allele frequency. Manual confirmation of the resulting SNP calls suggested that the model was capable of detecting a SNP for minor alleles present at 10% or greater frequency.

I used computational simulations to evaluate the effects of a minimum read depth filter. Bootstrap resampling (1,000 replicates) across simulated populations with minor allele frequencies of 10%, 30%, and 50% indicated that with a read depth of 15 the calculated mean allele frequency was within 0.003 of the true allele frequency with standard deviations from 0.08 to 0.13. Increasing the read depth minimum decreased the amount of variance in the estimate around the true allele frequency (Fig. 22). However, increasing the read depth minimum also increased the amount of missing data and thus decreased the number of loci available for analysis (Fig. 23). Because the goal of my analysis was to identify loci for which allele frequencies differed greatly between sample pools, I chose to accept some error around the allele frequency estimate in order to increase the proportion of the genome scanned for outliers. I recognize that the inclusion of loci with missing data made the analyses more susceptible to false positives. However, removing all loci with missing data would have greatly limited the genomic scope of this project and its capacity to identify loci potentially linked to adaptation in this system. In studies with more complete genomes and multiple individually barcoded samples, it may be possible to use linkage to impute missing genotypes (Marchini & Howie 2010).

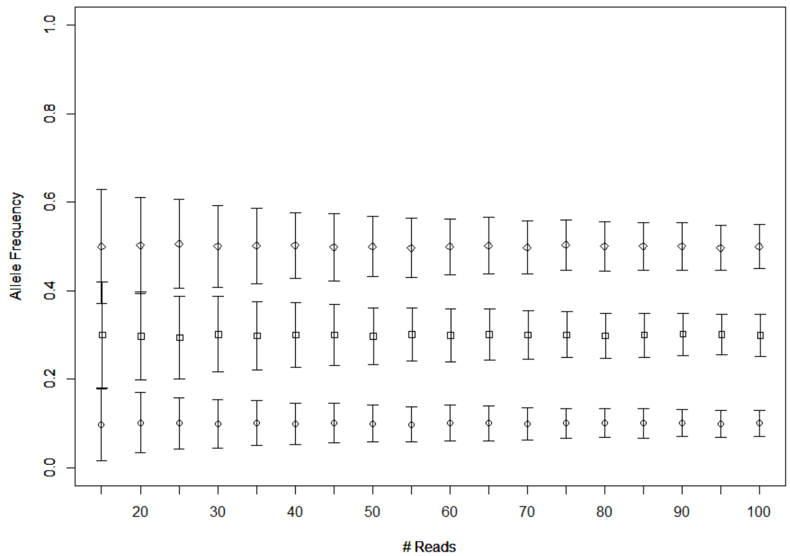

Figure 22: Mean allele frequency ± SD calculated by bootstrap resampling (1,000 replicates) across simulated populations with minor allele frequencies of 10% (circle), 30% (square), and 50% (diamond), indicating the decreasing amount of variance around allele frequency estimates as sample size (i.e., number of reads) increases.

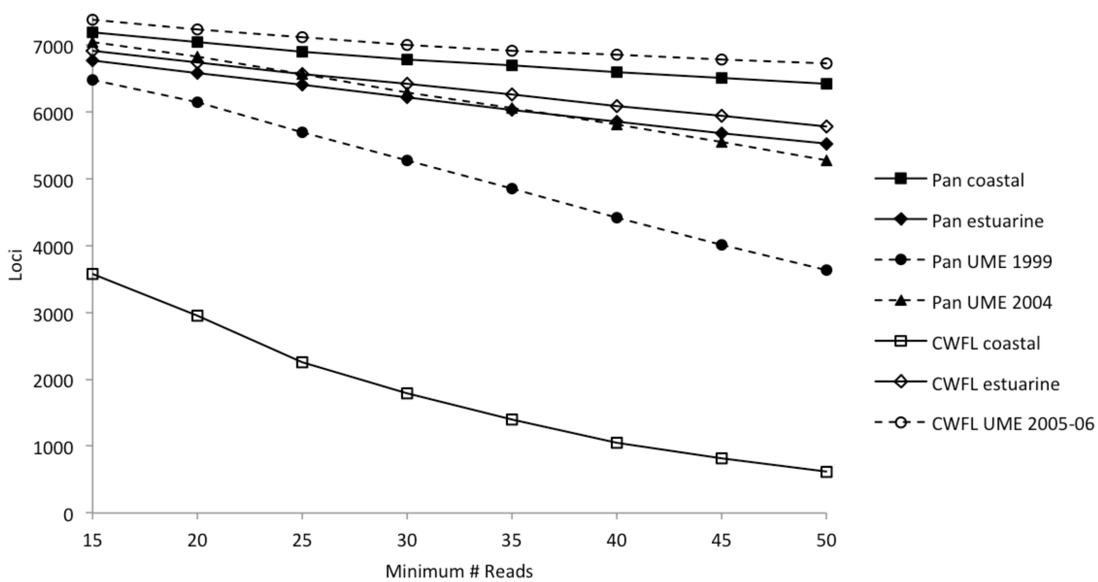

Figure 23: Number of loci per sample pool with 1 SNP and minor allele frequency greater than 10% sequenced in at least a given minimum number of reads, indicating the decreasing number of loci available for analysis as the minimum read depth cutoff increases. Samples include live coastal (squares) and estuarine (diamonds) bottlenose dolphins from central-west Florida (CWFL; open symbols) and the Florida Panhandle (Pan; solid symbols), as well as dolphins that stranded during UMEs in the Panhandle in 1999 (circles) and 2004 (triangles) and during red tides between 1992 and 2006, including a UME in 2005-2006, in central-west Florida (circles).

Adaptive loci

Loci linked to adaptation can be identified by comparing allele frequencies among populations that are adapted to different environments, or within a population between cases and controls. Loci with significant differences in allele frequency between groups or exhibiting significantly greater genetic differentiation (e.g., F_{ST}) than observed at neutral markers across the genome (i.e., “outliers” (Storz 2005)) may be under selection or, more likely, linked to genes under selection. It is however important to note that the number of phenotype-associated loci is likely to be inflated; GWAS are susceptible to false positives that result from extreme differences in allele frequency not due to selection (Allendorf *et al.* 2010; Nadeau & Jiggins 2010; Stinchcombe & Hoekstra 2008).

The number of false positives can be reduced by requiring that loci are significant in multiple tests (Hohenlohe *et al.* 2010); however, this assumes that the same loci will be responsible for adaptation across the entire sample set, and this assumption is not always valid (Hoekstra & Nachman 2003). In my study of bottlenose dolphin resistance to red tides, I conducted multiple comparisons between pools of live and dead dolphins from two geographic regions and three separate mortality events. Using multiple comparisons, I was able to identify loci that showed significant differences in allele frequency between live and dead dolphins across all comparisons, as well as loci that were significant in only one geographic region or during only one mortality event.

Anonymity

The primary disadvantage of GBS association studies is that most markers under investigation are anonymous (i.e., located in non-coding and/or non-annotated portions of the genome), and it can be difficult to understand the biological importance of GBS loci linked to phenotype. Given the small percentage of a genome that encodes proteins, it is highly unlikely that GBS loci associated with phenotype are located within a coding region. More commonly, the significant variants are linked to a gene regulating or encoding an adaptation-related protein.

GBS loci can be mapped to a reference genome of the study species or closely related species, if available, to identify potentially linked candidate genes in the surrounding genomic region (Fig. 21A). If known, average measures of linkage disequilibrium across the genome can be used to determine the size of the flanking genomic region or the density of markers required across the genome (Allendorf *et al.* 2010). Because of the generally small size of genome scaffolds for the bottlenose dolphin, I limited my search window to 100 kb from a RAD locus. Other studies have investigated out to 300 kb around a RAD locus (Hohenlohe *et al.* 2010).

This mapping approach is limited to genome-enabled organisms, and in particular genomes with annotation. For non-model species with no genome annotation resources, the identification of anonymous phenotype-associated markers enables quantification and comparison of loci involved in selection across samples and genotype to phenotype predictive power (Allendorf *et al.* 2010; McCarthy & Hirschhorn 2008).

However, one cannot make inferences about functional mechanisms underlying phenotype without annotation information. For genome-enabled organisms, this mapping approach will generate a long list of candidate genes, many of which are likely to be false positives. As described above, GWAS are susceptible to false positives (Allendorf *et al.* 2010; Nadeau & Jiggins 2010; Stinchcombe & Hoekstra 2008), and furthermore, within a putative linkage block surrounding a correctly identified phenotype-associated SNP, there are likely to be multiple genes not all of which encode the studied phenotype. For these reasons, additional analyses may be necessary to confirm the GWAS results and better understand the functional mechanism of genetic association (see *Combined approach*).

Candidate gene approach

In contrast to the genome-wide approach, which scans anonymous markers across the genome with no *a priori* hypothesis of which genes may be associated with phenotype, a candidate gene approach investigates regions of the genome that have a known function. The conservation of gene function across distinct lineages enables the investigation in non-model organisms of candidate genes whose functions have been previously characterized in other well-studied organisms (Fitzpatrick *et al.* 2005). Some researchers have questioned the need for candidate gene approaches in this genomic era (Wilkening *et al.* 2009), but others have argued that most studies that have successfully identified genes responsible for phenotype have utilized candidate gene approaches (Feder & Mitchell-Olds 2003; Stinchcombe & Hoekstra 2008). In some cases, for example studies of a genetic basis for susceptibility to infectious disease, candidate gene approaches may outperform GWAS (Amos *et al.* 2011). Only in species, like humans, with almost complete genome sequence and annotation, extensive knowledge of linkage across the genome, and well-developed SNP genotyping tools, is it possible to know that genome-wide approaches will screen all genes and biological pathways of potential interest (Wilkening *et al.* 2009).

The primary advantage of the candidate gene approach is the ability to test an informed hypothesis of an association between functional genetic variation and phenotype with relatively high statistical power (Amos *et al.* 2011). Inherent in hypothesis testing, however, is a logistical limitation to the number of hypotheses (i.e., candidate genes) that can be tested. Candidate gene studies in non-model systems have tended to focus on a relatively small number of candidate genes sequenced over a range of well-studied species (Piertney & Webster 2010). Accordingly, the primary critique of the candidate gene approach is its limited scope and inability to discover novel genes associated with phenotype.

Combined approach

By combining genome-wide and candidate gene approaches, we can address both critiques of the anonymity of the genome-wide approach and the limited scope of the traditional candidate gene approach. In contrast to a "bottom-up" approach, where candidate genes are selected based solely on the knowledge of function in other species, I recommend a "top-down" approach (Piertney & Webster 2010), where candidate genes are selected from genomic regions associated with a phenotype, which have been identified, for example, by GWAS (post-GWAS candidate genes; Fig. 21B). Only in the past several years have advances in next-generation sequencing facilitated such a top-down approach to candidate gene selection in non-model organisms. Now, high-throughput discovery of candidate genes can be the first step in a study of functional or adaptation genomics (Feder & Mitchell-Olds, 2003).

Even when using a top-down approach to identify candidate genes, the number of annotated genes in the region surrounding a phenotype-associated locus may be overwhelming. It is therefore reasonable to prioritize candidate genes based on functional relevance to the phenotype of interest and target the most functionally important components (e.g., exons) of a gene (McCarthy & Hirschhorn 2008). For example, in sequencing four major histocompatibility complex (MHC) class II loci in close proximity to a SNP associated with survival in bottlenose dolphins following red tide exposure, I prioritized the functionally important peptide-binding region (Chapter 3). In this case, I was able to characterize existing variation at the MHC in bottlenose dolphins from the Gulf of Mexico, which may have important health implications for these populations, but I was not able to link MHC variation to brevetoxin resistance to the degree expected following RAD sequencing.

The post-GWAS candidate gene approach does not negate the need to also consider candidate genes selected independently from genome-wide studies (Fig. 21C). GBS studies will not assay all genes, and in some cases, a strong *a priori* hypothesis of gene involvement in phenotype will justify sequencing genes that were not covered by GBS loci or were not identified as significant outliers in GBS studies. For example, in the case of bottlenose dolphin survival following exposure to red tides, there was strong justification for investigating genetic variation in the voltage-gated sodium channels. The neurotoxins produced by red tides bind to these sodium channels (Poli *et al.* 1986), and in many other systems, organisms have evolved resistance to neurotoxins via mutations in the toxin-binding site (Anderson 1987; Bricelj *et al.* 2005; Geffeney *et al.* 2005; Jost *et al.* 2008a). Of nine voltage-gated sodium channel isoforms, only four were in genomic proximity to a polymorphic RAD locus sequenced in Chapter 1, and none of these loci exhibited significant differences in allele frequency between live and dead dolphins. In Chapter 2, I therefore used a candidate gene approach to sequence the putative red tide toxin binding site on all nine sodium channel isoforms in bottlenose dolphins. I identified very little genetic variation in these regions and found no association between genetic variation and survival. In contrast to many other previously studied neurotoxin-resistant species, I concluded that bottlenose dolphins have not evolved resistance to red tides via mutations in the brevetoxin binding site on the voltage-gated sodium channels.

General limitations

Even with a combined genome-wide and candidate gene approach, it is important to recognize the limitations inherent in studying the genetic basis of complex traits, for example susceptibility to complex diseases (Hirschhorn & Daly 2005), particularly in non-model systems. These limitations represent multiple facets of adaptation research in non-model species in the natural environment that cannot be overcome by technological advances in genomic science. Perhaps the most important limitation is the inability to experimentally manipulate environments and design controls. Natural variation in complex phenotypes can be difficult to characterize, particularly in contrast to laboratory strains bred for certain characteristics. For example, in studies of disease, characterizing resistant and susceptible phenotypes can be unknowingly confounded by differences in exposure rate, a variable that cannot be controlled in the natural environment (Amos *et al.* 2011). In my case study of bottlenose dolphins, I chose to study a binomial phenotype, survival. I could not quantify the exact exposure of each dolphin to red tide toxin, but assumed relatively equal exposure among individuals across space and time given the available information on algal cell and toxin concentrations measured in water, prey fish, and dolphin tissue samples (Fire *et al.* 2007; Fire *et al.* 2008a; Flewelling *et al.* 2005; Hinton & Ramsdell 2008; Naar *et al.* 2007; Van Dolah *et al.* 2003).

Other limitations are inherent in the design of association studies. In studies of complex traits, which are likely polygenic, statistical power to detect associations may be hindered by small effect sizes (i.e., how much each gene contributes to the final phenotype) (Hirschhorn & Daly 2005). Additionally, once an association between genotype and phenotype has been identified, a causative relationship can only be assessed using experimental knockouts that artificially modify gene function (e.g., RNA interference (Hannon 2002)). Conclusions are therefore correlative, unless follow-up studies are performed in representative transgenic organisms.

Given advances in genomic sequencing, it has become increasingly easy to find loci associated with phenotype; however, it remains difficult to describe the underlying functional mechanisms of adaptation. Advances in genomic sequencing have greatly enabled genomics research in non-model species, but the majority of non-model genomes are still poorly annotated. Without annotation, it is only possible to identify anonymous loci associated with phenotype. Studies aiming to increase our understanding of genes that underlie ecologically important traits are effectively limited to genome-enabled species, and furthermore to species whose genomes, or genomes of closely related species, are at least partially annotated.

An alternate approach to GBS and candidate gene studies may be whole genome re-sequencing (Bentley 2006). With the continually decreasing price of genome sequencing (Sboner *et al.* 2011), it is becoming feasible to sequence at low coverage complete genomes from multiple individuals (e.g., Jones *et al.* 2012; Martin *et al.* 2013). The greater density of markers produced by these approaches enable identification of "genomic islands," regions that surround genes under divergent selection (Feder *et al.* 2012; Turner *et al.* 2005). The location of these islands may be more informative than individual phenotype-associated SNPs in identifying the underlying functional variation responsible for phenotype. Although whole genome re-sequencing will face many limitations of its own (e.g., low coverage SNP genotyping, computational requirements)

(Allendorf *et al.* 2010), genome re-sequencing may in the future become an effective alternative approach to association studies, which combines the goals of genome-wide GBS and candidate gene sequencing.

Conclusion

In studying the genetic basis for resistance to red tides in bottlenose dolphins, I demonstrated the potential power of a combined approach using genome-wide RAD sequencing and both post-GWAS and GWAS-independent candidate gene sequencing. I was able to identify RAD loci that suggested both a common genetic basis for resistance to HABs in bottlenose dolphins across the Gulf coast of Florida and regionally specific resistance. I conducted follow-up candidate gene studies of two groups of candidate genes, four MHC loci selected because of their genomic proximity to a SNP exhibiting a strong association with survival and nine voltage-gated sodium channels selected because of an *a priori* hypothesis of gene involvement. Variation in bottlenose dolphin mortality associated with HAB exposure could not be solely attributed to variation in either candidate gene group. However, the genome-wide analysis provides evidence that suggests a genetic basis for resistance to HABs in bottlenose dolphins and provides justification and direction for future studies of resistance to HAB toxins in marine mammals.

APPENDIX A: RAD SEQUENCING CANDIDATE GENES

Annotated genes within 100 kb of select RAD loci in the Ensembl bottlenose dolphin genome (turTru1, Ensembl release 71, GenBank Assembly ID: GCA_000151865.1). Candidate RAD loci were selected based on consistent and significant differences in allele frequency in: all live-dead comparisons sequenced to sufficient read depth (included loci with missing data); all and only live-dead comparisons within a single geographic region (central-west Florida or Panhandle, excluded loci with missing data); all and only live-dead comparisons involving a single Panhandle UME (1999 or 2004 UME, excluded loci with missing Panhandle data). Note: a single gene could be identified in multiple groups if within 100 kb of multiple candidate RAD loci.

Locus ID	Gene ID	Gene Symbol	Description
All live-dead comparisons sequenced to sufficient read depth, 78 genes			
840	ENSTTRG00000004130	HLA-DRA	major histocompatibility complex, class II, DR alpha
840	ENSTTRG00000004146	HLA-DRB	major histocompatibility complex, class II, DR beta
840	ENSTTRG00000004152	HLA-DQA	major histocompatibility complex, class II, DQ alpha
31997	ENSTTRG00000006960	CWF19L2	CWF19-like 2, cell cycle control (S. pombe)
44973	ENSTTRG00000004557	GORASP2	golgi reassembly stacking protein 2, 55kDa
44973	ENSTTRG00000004560	TLK1	tousled-like kinase 1
88058	ENSTTRG00000015215	SNED1	sushi, nidogen and EGF-like domains 1
88058	ENSTTRG00000015232	MTERFD2	MTERF domain containing 2
88058	ENSTTRG00000015236	PASK	PAS domain containing serine/threonine kinase
88058	ENSTTRG00000015244	PPP1R7	protein phosphatase 1, regulatory subunit 7
88058	ENSTTRG00000015247	ANO7	anoctamin 7
88058	ENSTTRG00000015251	HDLBP	high density lipoprotein binding protein
88058	ENSTTRG00000015252	SEPT2	septin 2
88962	ENSTTRG00000006920	TBC1D15	TBC1 domain family, member 15
88962	ENSTTRG00000006925	TPH2	tryptophan hydroxylase 2
95939	ENSTTRG00000007170	FAM196A	family with sequence similarity 196, member A
99316	ENSTTRG00000007583	NOTCH1	notch 1
99316	ENSTTRG00000007585	EGFL7	EGF-like-domain, multiple 7
99316	ENSTTRG00000007586	AGPAT2	1-acylglycerol-3-phosphate O-acyltransferase 2
99316	ENSTTRG00000007587	LCN10	lipocalin 10

106479	ENSTTRG00000010623	P2RX2	purinergic receptor P2X, ligand-gated ion channel, 2
106479	ENSTTRG00000010629	POLE	polymerase (DNA directed), epsilon, catalytic subunit
106479	ENSTTRG00000010651	PXMP2	peroxisomal membrane protein 2, 22kDa
106479	ENSTTRG00000010653	PGAM5	phosphoglycerate mutase family member 5
106479	ENSTTRG00000010654	ANKLE2	ankyrin repeat and LEM domain containing 2
113531	ENSTTRG00000004803	TRIOBP	TRIO and F-actin binding protein
113531	ENSTTRG00000004807	H1F0	H1 histone family, member 0
113531	ENSTTRG00000004808	GCAT	glycine C-acetyltransferase
113531	ENSTTRG00000004809	GALR3	galanin receptor 3
113531	ENSTTRG00000004811	ANKRD54	ankyrin repeat domain 54
113531	ENSTTRG00000004812	EIF3L	eukaryotic translation initiation factor 3, subunit L
113531	ENSTTRG00000004814	MICALL1	MICAL-like 1
116938	ENSTTRG00000016530	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1
121079	ENSTTRG00000017087	SCRIB	scribbled homolog (Drosophila)
121079	ENSTTRG00000017088	PUF60	poly-U binding splicing factor 60KDa
121079	ENSTTRG00000017089	NRBP2	nuclear receptor binding protein 2
124956	ENSTTRG00000011160	SLC15A5	solute carrier family 15, member 5
124956	ENSTTRG00000011168	MGST1	microsomal glutathione S-transferase 1
126328	ENSTTRG00000002377	FAM125B	family with sequence similarity 125, member B
126521	ENSTTRG00000015252	SEPT2	septin 2
126521	ENSTTRG00000015253	FARP2	FERM, RhoGEF and pleckstrin domain protein 2
126521	ENSTTRG00000015254	STK25	serine/threonine kinase 25
128900	ENSTTRG00000007047	GPR123	G protein-coupled receptor 123
128900	ENSTTRG00000007105	KNDC1	kinase non-catalytic C-lobe domain (KIND) containing 1
129239	ENSTTRG00000010460	SLC41A3	solute carrier family 41, member 3
129239	ENSTTRG00000010472	ALDH1L1	aldehyde dehydrogenase 1 family, member L1
133391	ENSTTRG00000011356	ARG1	arginase, liver
133391	ENSTTRG00000011360	MED23	mediator complex subunit 23
133391	ENSTTRG00000011371	ENPP3	ectonucleotide pyrophosphatase/phosphodiesterase 3
133870	ENSTTRG00000001000	NCKAP5L	NCK-associated protein 5-like

133870	ENSTTRG00000001001	BCDIN3D	BCDIN3 domain containing
133870	ENSTTRG00000001002	FAIM2	Fas apoptotic inhibitory molecule 2
133870	ENSTTRG00000001003	AQP2	aquaporin 2 (collecting duct)
133870	ENSTTRG00000001004	AQP5	aquaporin 5
133870	ENSTTRG00000001005	AQP6	aquaporin 6, kidney specific
133870	ENSTTRG00000001006	RACGAP1	Rac GTPase activating protein 1
133870	ENSTTRG00000021406	N/A	novel transcript
134248	ENSTTRG00000016530	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1
134756	ENSTTRG00000005008	SPIRE2	spire homolog 2 (Drosophila)
134756	ENSTTRG00000005011	TCF25	transcription factor 25 (basic helix-loop-helix)
134756	ENSTTRG00000005014	C5I7U7_TURTR	Melanocortin-1 receptor
134756	ENSTTRG00000005016	DEF8	differentially expressed in FDCP 8 homolog (mouse)
134756	ENSTTRG00000005017	N/A	novel transcript
134756	ENSTTRG00000005018	N/A	novel transcript
134756	ENSTTRG00000005019	N/A	novel transcript
136996	ENSTTRG00000010678	MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast)
136996	ENSTTRG00000010687	FTSJ2	FtsJ RNA methyltransferase homolog 2 (E. coli)
136996	ENSTTRG00000010692	NUDT1	nudix (nucleoside diphosphate linked moiety X)-type motif 1
136996	ENSTTRG00000010695	SNX8	sorting nexin 8
138930	ENSTTRG00000011217	GPR37L1	G protein-coupled receptor 37 like 1
138930	ENSTTRG00000011223	ARL8A	ADP-ribosylation factor-like 8A
138930	ENSTTRG00000011226	PTPN7	protein tyrosine phosphatase, non-receptor type 7
138930	ENSTTRG00000011233	LGR6	leucine-rich repeat containing G protein-coupled receptor 6
138930	ENSTTRG00000011238	UBE2T	ubiquitin-conjugating enzyme E2T (putative)
138930	ENSTTRG00000011239	PPP1R12B	protein phosphatase 1, regulatory subunit 12B
209576	ENSTTRG00000008058	UNG	uracil-DNA glycosylase
209576	ENSTTRG00000008072	ALKBH2	alkB, alkylation repair homolog 2 (E. coli)
209576	ENSTTRG00000008079	USP30	ubiquitin specific peptidase 30

Central-west Florida only, 79 genes			
8711	ENSTTRG00000004400	ATP2B2	ATPase, Ca++ transporting, plasma membrane 2
27718	ENSTTRG00000005626	CBY1	chibby homolog 1 (Drosophila)
27718	ENSTTRG00000005627	TOMM22	translocase of outer mitochondrial membrane 22 homolog (yeast)
27718	ENSTTRG00000005628	JOSD1	Josephin domain containing 1
27718	ENSTTRG00000005629	GTPBP1	GTP binding protein 1
27718	ENSTTRG00000005630	SUN2	Sad1 and UNC84 domain containing 2
27718	ENSTTRG00000005631	DNAL4	dynein, axonemal, light chain 4
27857	ENSTTRG00000015186	PDE1A	phosphodiesterase 1A, calmodulin-dependent
28267	ENSTTRG00000012584	SLC16A6	solute carrier family 16, member 6 (monocarboxylic acid transporter 7)
28267	ENSTTRG00000012586	ARSG	arylsulfatase G
28267	ENSTTRG00000012592	WIPI1	WD repeat domain, phosphoinositide interacting 1
29312	ENSTTRG00000011816	KIAA1024	KIAA1024
32820	ENSTTRG00000001384	SH2D4A	SH2 domain containing 4A
32820	ENSTTRG00000001385	CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1
33584	ENSTTRG00000007220	N/A	novel transcript
35079	ENSTTRG00000007857	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic
44584	ENSTTRG00000000617	FAM110D	family with sequence similarity 110, member D
44584	ENSTTRG00000000618	ZNF593	zinc finger protein 593
44584	ENSTTRG00000000619	CNKSR1	connector enhancer of kinase suppressor of Ras 1
44584	ENSTTRG00000000620	CATSPER4	cation channel, sperm associated 4
44584	ENSTTRG00000000622	CEP85	centrosomal protein 85kDa
44584	ENSTTRG00000000624	SH3BGRL3	SH3 domain binding glutamic acid-rich protein like 3
44584	ENSTTRG00000000625	UBXN11	UBX domain protein 11
44794	ENSTTRG00000013475	ABCA13	ATP-binding cassette, sub-family A, member 13
46174	ENSTTRG00000017402	PBX3	pre-B-cell leukemia homeobox 3
49835	ENSTTRG00000005013	N/A	novel transcript
52384	ENSTTRG00000001391	SCYL2	SCY1-like 2 (S. cerevisiae)

52384	ENSTTRG00000001393	SLC17A8	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 8
67401	ENSTTRG00000005754	PRR16	proline rich 16
68388	ENSTTRG00000011564	BUD13	BUD13 homolog (S. cerevisiae)
68388	ENSTTRG00000011577	ZNF259	zinc finger protein 259
68388	ENSTTRG00000011587	APOA5	apolipoprotein A-V
68597	ENSTTRG00000017162	PIM3	pim-3 oncogene
68597	ENSTTRG00000017163	IL17REL	interleukin 17 receptor E-like
69047	ENSTTRG00000007961	GSC	goosecoid homeobox
69401	ENSTTRG00000013506	PHAX	phosphorylated adaptor for RNA export
69401	ENSTTRG00000013509	C5ORF48	chromosome 5 open reading frame 48
69401	ENSTTRG00000013511	LMNB1	lamin B1
77925	ENSTTRG00000009102	N/A	novel transcript
77925	ENSTTRG00000009119	MAPT	microtubule-associated protein tau
77925	ENSTTRG00000009160	KANSL1	KAT8 regulatory NSL complex subunit 1
80111	ENSTTRG00000013579	EFCAB11	EF-hand calcium binding domain 11
80111	ENSTTRG00000013580	TDP1	tyrosyl-DNA phosphodiesterase 1
80111	ENSTTRG00000013581	KCNK13	potassium channel, subfamily K, member 13
82094	ENSTTRG00000011013	CDH5	cadherin 5, type 2 (vascular endothelium)
82094	ENSTTRG00000011030	BEAN1	brain expressed, associated with NEDD4, 1
82094	ENSTTRG00000011035	TK2	thymidine kinase 2, mitochondrial
82094	ENSTTRG00000011048	CKLF	chemokine-like factor
82094	ENSTTRG00000011054	N/A	novel transcript
82094	ENSTTRG00000011065	CMTM2	CKLF-like MARVEL transmembrane domain containing 2
82094	ENSTTRG00000011072	CMTM3	CKLF-like MARVEL transmembrane domain containing 3
82476	ENSTTRG00000017084	SIM2	single-minded homolog 2 (Drosophila)
82476	ENSTTRG00000017086	HLCS	holocarboxylase synthetase
84824	ENSTTRG00000005989	IL1RAPL1	interleukin 1 receptor accessory protein-like 1
85425	ENSTTRG00000001492	PRKAB1	protein kinase, AMP-activated, beta 1 non-catalytic subunit

85425	ENSTTRG00000001493	CIT	citron (rho-interacting, serine/threonine kinase 21)
90934	ENSTTRG00000005160	ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
90934	ENSTTRG00000005222	N/A	novel transcript
92092	ENSTTRG00000007752	HNRNPR	heterogeneous nuclear ribonucleoprotein R
92092	ENSTTRG00000007756	ZNF436	zinc finger protein 436
92092	ENSTTRG00000007762	N/A	novel transcript
92092	ENSTTRG00000007775	TCEA3	transcription elongation factor A (SII), 3
92092	ENSTTRG00000007788	ASAP3	ArfGAP with SH3 domain, ankyrin repeat and PH domain 3
100679	ENSTTRG00000007777	ITPRIPL1	inositol 1,4,5-trisphosphate receptor interacting protein-like 1
100679	ENSTTRG00000007784	NCAPH	non-SMC condensin I complex, subunit H
100679	ENSTTRG00000007812	ARID5A	AT rich interactive domain 5A
102567	ENSTTRG00000016126	C2CD4C	C2 calcium-dependent domain containing 4C
102567	ENSTTRG00000016131	SHC2	SHC (Src homology 2 domain containing) transforming protein 2
102567	ENSTTRG00000016137	ODF3L2	outer dense fiber of sperm tails 3-like 2
104138	ENSTTRG00000011943	CSF2RA	colony stimulating factor 2 receptor, alpha, low-affinity
106559	ENSTTRG00000002957	TRRAP	transformation/transcription domain-associated protein
106559	ENSTTRG00000002991	N/A	novel transcript
106559	ENSTTRG00000002994	SMURF1	SMAD specific E3 ubiquitin protein ligase 1
109462	ENSTTRG00000005874	LMF1	lipase maturation factor 1
109462	ENSTTRG00000005897	SOX8	SRY (sex determining region Y)-box 8
122606	ENSTTRG00000016204	NUDT7	nudix (nucleoside diphosphate linked moiety X)-type motif 7
122606	ENSTTRG00000016205	VAT1L	vesicle amine transport protein 1 homolog (T. californica)-like
128872	ENSTTRG00000000621	C15ORF26	chromosome 15 open reading frame 26
130555	ENSTTRG00000011643	IGSF21	immunoglobin superfamily, member 21
Panhandle only, 4 genes			
53245	ENSTTRG00000004204	DNTTIP2	deoxynucleotidyltransferase, terminal, interacting protein 2
53245	ENSTTRG00000004208	GCLM	glutamate-cysteine ligase, modifier subunit
30199	ENSTTRG00000011957	N/A	novel transcript
19599	ENSTTRG00000012144	PDE4DIP	phosphodiesterase 4D interacting protein

Panhandle 2004 UME only, 101 genes			
5108	ENSTTRG00000001008	ASIC1	acid-sensing (proton-gated) ion channel 1
5108	ENSTTRG00000001010	SMARCD1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1
5108	ENSTTRG00000001012	GPD1	glycerol-3-phosphate dehydrogenase 1 (soluble)
5108	ENSTTRG00000001013	N/A	novel transcript
5108	ENSTTRG00000001014	CERS5	ceramide synthase 5
5108	ENSTTRG00000001015	LIMA1	LIM domain and actin binding 1
11909	ENSTTRG00000010876	KRTAP8-1	keratin associated protein 8-1
11909	ENSTTRG00000010890	N/A	novel transcript
13904	ENSTTRG00000006318	SAMD12	sterile alpha motif domain containing 12
13936	ENSTTRG00000014069	ST8SIA5	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5
16242	ENSTTRG00000001409	CASR	calcium-sensing receptor
20297	ENSTTRG00000001863	N/A	novel transcript
25531	ENSTTRG00000008250	KIAA0146	KIAA0146
25531	ENSTTRG00000008265	PRKDC	protein kinase, DNA-activated, catalytic polypeptide
26350	ENSTTRG00000000384	N/A	novel transcript
37401	ENSTTRG00000016162	DPM3	dolichyl-phosphate mannosyltransferase polypeptide 3
37401	ENSTTRG00000016163	SLC50A1	solute carrier family 50 (sugar transporter), member 1
37401	ENSTTRG00000016164	EFNA1	ephrin-A1
37401	ENSTTRG00000016165	N/A	novel transcript
37401	ENSTTRG00000016166	EFNA4	ephrin-A4
37401	ENSTTRG00000016167	ADAM15	ADAM metallopeptidase domain 15
37401	ENSTTRG00000016168	DCST1	DC-STAMP domain containing 1
37401	ENSTTRG00000016169	DCST2	DC-STAMP domain containing 2
37401	ENSTTRG00000016171	ZBTB7B	zinc finger and BTB domain containing 7B
37401	ENSTTRG00000016172	LENEP	lens epithelial protein
37401	ENSTTRG00000016174	FLAD1	FAD1 flavin adenine dinucleotide synthetase homolog (S. cerevisiae)
37401	ENSTTRG00000016175	CKS1B	CDC28 protein kinase regulatory subunit 1B

37401	ENSTTRG00000016176	SHC1	SHC (Src homology 2 domain containing) transforming protein 1
37401	ENSTTRG00000016177	PYGO2	pygopus homolog 2 (Drosophila)
37401	ENSTTRG00000016178	PBXIP1	pre-B-cell leukemia homeobox interacting protein 1
45286	ENSTTRG00000007249	TMED2	transmembrane emp24 domain trafficking protein 2
45286	ENSTTRG00000007251	DDX55	DEAD (Asp-Glu-Ala-Asp) box polypeptide 55
45286	ENSTTRG00000007254	EIF2B1	eukaryotic translation initiation factor 2B, subunit 1 alpha, 26kDa
45286	ENSTTRG00000007259	GTF2H3	general transcription factor IIH, polypeptide 3, 34kDa
45286	ENSTTRG00000007263	TCTN2	tectonic family member 2
47361	ENSTTRG00000016538	N/A	novel transcript
50512	ENSTTRG00000003090	ACOXL	acyl-CoA oxidase-like
65556	ENSTTRG00000008454	GINS3	GINS complex subunit 3
65556	ENSTTRG00000008456	NDRG4	NDRG family member 4
65556	ENSTTRG00000008459	SETD6	SET domain containing 6
65556	ENSTTRG00000008460	CNOT1	CCR4-NOT transcription complex, subunit 1
69635	ENSTTRG00000007394	NUDT5	nudix (nucleoside diphosphate linked moiety X)-type motif 5
69635	ENSTTRG00000007395	SEC61A2	Sec61 alpha 2 subunit (S. cerevisiae)
69635	ENSTTRG00000007399	DHTKD1	dehydrogenase E1 and transketolase domain containing 1
69755	ENSTTRG00000009130	PAMR1	peptidase domain containing associated with muscle regeneration 1
74281	ENSTTRG00000017072	PLA2G2D	phospholipase A2, group IID
74281	ENSTTRG00000017073	PLA2G2C	phospholipase A2, group IIC
74281	ENSTTRG00000017074	UBXN10	UBX domain protein 10
79874	ENSTTRG00000009435	CRYBB1	crystallin, beta B1
79874	ENSTTRG00000009438	CRYBA4	crystallin, beta A4
79874	ENSTTRG00000009444	N/A	novel transcript
82467	ENSTTRG00000009986	ENKUR	enkurin, TRPC channel interacting protein
82467	ENSTTRG00000009989	THNSL1	threonine synthase-like 1 (S. cerevisiae)
82467	ENSTTRG00000009996	GPR158	G protein-coupled receptor 158
96607	ENSTTRG00000014578	TAB1	TGF-beta activated kinase 1/MAP3K7 binding protein 1
96607	ENSTTRG00000014579	N/A	novel transcript

96607	ENSTTRG00000014580	SMCR7L	Smith-Magenis syndrome chromosome region, candidate 7-like
96607	ENSTTRG00000014581	ATF4	activating transcription factor 4
96607	ENSTTRG00000014582	RPS19BP1	ribosomal protein S19 binding protein 1
101485	ENSTTRG00000004156	PTCHD4	patched domain containing 4
103332	ENSTTRG00000010107	ATRNL1	attractin-like 1
107603	ENSTTRG00000011043	LSM14B	LSM14B, SCD6 homolog B (S. cerevisiae)
107603	ENSTTRG00000011050	PSMA7	proteasome (prosome, macropain) subunit, alpha type, 7
107603	ENSTTRG00000011062	SS18L1	synovial sarcoma translocation gene on chromosome 18-like 1
107603	ENSTTRG00000011070	GTPBP5	GTP binding protein 5 (putative)
107603	ENSTTRG00000011074	HRH3	histamine receptor H3
107603	ENSTTRG00000011077	N/A	novel transcript
107603	ENSTTRG00000011082	OSBPL2	oxysterol binding protein-like 2
107603	ENSTTRG00000011089	ADRM1	adhesion regulating molecule 1
107603	ENSTTRG00000011095	LAMA5	laminin, alpha 5
109164	ENSTTRG00000003437	THBD	thrombomodulin
109164	ENSTTRG00000003438	N/A	novel transcript
109812	ENSTTRG00000012732	RAN	RAN, member RAS oncogene family
109812	ENSTTRG00000012734	GPR133	G protein-coupled receptor 133
113535	ENSTTRG00000004803	TRIOBP	TRIO and F-actin binding protein
113535	ENSTTRG00000004807	H1F0	H1 histone family, member 0
113535	ENSTTRG00000004808	GCAT	glycine C-acetyltransferase
113535	ENSTTRG00000004809	GALR3	galanin receptor 3
113535	ENSTTRG00000004811	ANKRD54	ankyrin repeat domain 54
113535	ENSTTRG00000004812	EIF3L	eukaryotic translation initiation factor 3, subunit L
113535	ENSTTRG00000004814	MICALL1	MICAL-like 1
113535	ENSTTRG00000004815	C22ORF23	chromosome 22 open reading frame 23
114040	ENSTTRG00000011013	CDH5	cadherin 5, type 2 (vascular endothelium)
116166	ENSTTRG00000001067	ANO1	anoctamin 1, calcium activated chloride channel
116166	ENSTTRG00000001068	N/A	novel transcript

116362	ENSTTRG00000012411	OSBPL7	oxysterol binding protein-like 7
116362	ENSTTRG00000012452	MRPL10	mitochondrial ribosomal protein L10
116362	ENSTTRG00000012463	LRRC46	leucine rich repeat containing 46
116362	ENSTTRG00000012472	SCRN2	secernin 2
116362	ENSTTRG00000012481	SP6	Sp6 transcription factor
116362	ENSTTRG00000012483	SP2	Sp2 transcription factor
116362	ENSTTRG00000012493	PNPO	pyridoxamine 5'-phosphate oxidase
119845	ENSTTRG00000000902	BRSK2	BR serine/threonine kinase 2
119845	ENSTTRG00000000905	MOB2	MOB kinase activator 2
121667	ENSTTRG00000009276	CECR1	cat eye syndrome chromosome region, candidate 1
122195	ENSTTRG00000003341	CSTF2T	cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant
122490	ENSTTRG00000011351	RSPH3	radial spoke 3 homolog (Chlamydomonas)
122490	ENSTTRG00000011352	TAGAP	T-cell activation RhoGTPase activating protein
122490	ENSTTRG00000011353	FNDC1	fibronectin type III domain containing 1
123285	ENSTTRG00000001604	FAM19A5	family with sequence similarity 19, member A5
123285	ENSTTRG00000001605	N/A	novel transcript
Panhandle 1999 UME only, 77 genes			
13711	ENSTTRG00000004002	SLIT3	slit homolog 3 (Drosophila)
22022	ENSTTRG00000002221	CSMD2	CUB and Sushi multiple domains 2
28252	ENSTTRG00000000041	VPS37C	vacuolar protein sorting 37 homolog C (S. cerevisiae)
28252	ENSTTRG00000000042	CD5	CD5 molecule
32737	ENSTTRG00000004658	GNB5	guanine nucleotide binding protein (G protein), beta 5
32737	ENSTTRG00000004710	MYO5C	myosin VC
36260	ENSTTRG00000002911	ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like
39584	ENSTTRG00000000442	N/A	novel transcript
39584	ENSTTRG00000000443	MOCS2	molybdenum cofactor synthesis 2
39584	ENSTTRG00000000445	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)
41209	ENSTTRG00000017279	FAM222A	family with sequence similarity 222, member A
45732	ENSTTRG00000016657	MTPAP	mitochondrial poly(A) polymerase

46197	ENSTTRG00000001976	FAM174B	family with sequence similarity 174, member B
49689	ENSTTRG00000010127	ADAMTS20	ADAM metallopeptidase with thrombospondin type 1 motif, 20
57394	ENSTTRG00000005687	SOX9	SRY (sex determining region Y)-box 9
57794	ENSTTRG00000000032	THBS1	thrombospondin 1
57794	ENSTTRG00000000033	FSIP1	fibrous sheath interacting protein 1
58173	ENSTTRG00000001481	EFCAB10	EF-hand calcium binding domain 10
58173	ENSTTRG00000001482	ATXN7L1	ataxin 7-like 1
64052	ENSTTRG00000003121	SHOX	short stature homeobox
64475	ENSTTRG00000002373	PTK2B	PTK2B protein tyrosine kinase 2 beta
64475	ENSTTRG00000002378	CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)
64475	ENSTTRG00000002379	EPHX2	epoxide hydrolase 2, cytoplasmic
65099	ENSTTRG00000014356	ACOX3	acyl-CoA oxidase 3, pristanoyl
65099	ENSTTRG00000014379	TRMT44	tRNA methyltransferase 44 homolog (S. cerevisiae)
65099	ENSTTRG00000014396	N/A	novel transcript
72202	ENSTTRG00000014484	ZFPM1	zinc finger protein, multitype 1
72397	ENSTTRG00000015104	HAUS5	HAUS augmin-like complex, subunit 5
72397	ENSTTRG00000015108	ATP4A	ATPase, H+/K+ exchanging, alpha polypeptide
72397	ENSTTRG00000015114	TMEM147	transmembrane protein 147
72397	ENSTTRG00000015124	GAPDHS	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic
72397	ENSTTRG00000015130	SBSN	suprabasin
72397	ENSTTRG00000015144	DMKN	dermokine
72397	ENSTTRG00000015144	DMKN	dermokine
72397	ENSTTRG00000015156	KRTDAP	keratinocyte differentiation-associated protein
72397	ENSTTRG00000015162	FFAR2	free fatty acid receptor 2
79657	ENSTTRG00000007327	LGR5	leucine-rich repeat containing G protein-coupled receptor 5
79657	ENSTTRG00000007328	ZFC3H1	zinc finger, C3H1-type containing
79657	ENSTTRG00000007330	THAP2	THAP domain containing, apoptosis associated protein 2
79752	ENSTTRG00000002431	CTNNA2	catenin (cadherin-associated protein), alpha 2
83604	ENSTTRG00000012364	SVOPL	SVOP-like

83604	ENSTTRG00000012381	ATP6V0A4	ATPase, H+ transporting, lysosomal V0 subunit a4
86439	ENSTTRG00000001759	RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1
86439	ENSTTRG00000001759	RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1
86439	ENSTTRG00000001759	RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1
90884	ENSTTRG00000011130	GLRB	glycine receptor, beta
90884	ENSTTRG00000011144	GRIA2	glutamate receptor, ionotropic, AMPA 2
92003	ENSTTRG00000001966	MTCH1	mitochondrial carrier 1
92003	ENSTTRG00000001967	FGD2	FYVE, RhoGEF and PH domain containing 2
92003	ENSTTRG00000001968	PIM1	pim-1 oncogene
93326	ENSTTRG00000015274	SASH1	SAM and SH3 domain containing 1
98199	ENSTTRG00000016272	WBSCR17	Williams-Beuren syndrome chromosome region 17
98199	ENSTTRG00000016273	CALN1	calneuron 1
101704	ENSTTRG00000005944	N/A	novel transcript
101704	ENSTTRG00000005950	CYP21A2	cytochrome P450, family 21, subfamily A, polypeptide 2
101704	ENSTTRG00000005953	TNXB	tenascin XB
101704	ENSTTRG00000005957	N/A	novel transcript
101704	ENSTTRG00000005958	ATF6B	activating transcription factor 6 beta
101704	ENSTTRG00000005959	N/A	novel transcript
101704	ENSTTRG00000005960	N/A	novel transcript
101704	ENSTTRG00000005961	PPT2	palmitoyl-protein thioesterase 2
101704	ENSTTRG00000005962	AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
101704	ENSTTRG00000021403	N/A	novel transcript
102085	ENSTTRG00000009051	SUGT1	SGT1, suppressor of G2 allele of SKP1 (S. cerevisiae)
102085	ENSTTRG00000009116	LECT1	leukocyte cell derived chemotaxin 1
102734	ENSTTRG00000009556	GSG1L	GSG1-like
102734	ENSTTRG00000009558	XPO6	exportin 6
106947	ENSTTRG00000012234	VAX2	ventral anterior homeobox 2
106947	ENSTTRG00000012236	ATP6V1B1	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B1
106947	ENSTTRG00000012239	ANKRD53	ankyrin repeat domain 53

106947	ENSTTRG00000012250	TEX261	testis expressed 261
112230	ENSTTRG00000002742	TLE3	transducin-like enhancer of split 3
114612	ENSTTRG00000002019	ADORA2B	adenosine A2b receptor
114612	ENSTTRG00000002020	ZSWIM7	zinc finger, SWIM-type containing 7
114612	ENSTTRG00000002021	TTC19	tetratricopeptide repeat domain 19
114612	ENSTTRG00000002022	NCOR1	nuclear receptor corepressor 1
121513	ENSTTRG00000003184	TRABD2B	TraB domain containing 2B

REFERENCES

Alho J, Valimaki K (2012) Rhh: Calculating multilocus heterozygosity and heterozygosity-heterozygosity correlation. *R package version 1.0.2*.

Allendorf FW, Hard JJ (2009) Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proceedings of the National Academy of Sciences* **106**, 9987.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews: Genetics* **11**, 697-709.

Amos W, Driscoll E, Hoffman JI (2011) Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings: Biological Sciences* **278**, 1183-1188.

Amos W, Wilmer JW, Fullard K, *et al.* (2001) The influence of parental relatedness on reproductive success. *Proceedings: Biological Sciences* **268**, 2021-2027.

Anderson J, Gleason C, Foley R, *et al.* (2010) Plants versus pathogens: an evolutionary arms race. *Functional Plant Biology* **37**, 499-512.

Anderson P (1987) Properties and pharmacology of a TTX-insensitive Na+ current in neurones of the jellyfish *Cyanea capillata*. *Journal of Experimental Biology* **133**, 231-248.

Anderson PAV, Roberts-Misterly J, Greenberg RM (2005) The evolution of voltage-gated sodium channels: were algal toxins involved? *Harmful Algae* **4**, 95-107.

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**, 323.

Arbanasic H, Curkovic S, Gomercic T, Gomercic M, Galov A (2012) Diveristy of MHC class II *DRA* gene in the Adriatic bottlenose dolphins (*Tursiops truncatus*). In: *Proceedings of the 11th Croatian Biological Congress* (eds Jelaska S, Klobucar G, Jelaska L, Levanic D, & Luksa Z), Croatian Biological Society, Zagreb, Croatia, pp. 151-152.

Backer L, Fleming L, Rowan A, *et al.* (2003) Recreational exposure to aerosolized brevetoxins during Florida red tide events. *Harmful Algae* **2**, 19-28.

Baden DG (1989) Brevetoxins: unique, polyether dinoflagellate toxins. *FASEB Journal* **3**, 1807-1817.

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One* **3**, e3376.

Baker CS, Vant M, Dalebout M, *et al.* (2006) Diversity and duplication of *DQB* and *DRB*-like genes of the MHC in baleen whales (suborder: *Mysticeti*). *Immunogenetics* **58**, 283-296.

Balmer BC, Wells RS, Nowacek SM, *et al.* (2008) Seasonal abundance and distribution patterns of bottlenose dolphins (*Tursiops truncatus*) near St. Joseph Bay, Florida, USA. *Journal of Cetacean Research and Management* **10**, 157-167.

Bandelier F (1905) *The Journey of Alvar Nunez Cabez De Vaca and his Companions from Florida to the Pacific, 1528-1536.* Allerton Book Co., New York.

Beaumont M, Nichols R (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 1619-1626.

Becerra JX, Noge K, Venable DL (2009) Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18062-18066.

Bentley DR (2006) Whole-genome re-sequencing. *Current Opinion in Genetics and Development* **16**, 545-552.

Berggren KT, Ellegren H, Hewitt GM, Seddon JM (2005) Understanding the phylogeographic patterns of European hedgehogs, *Erinaceus concolor* and *E. europaeus* using the MHC. *Heredity* **95**, 84-90.

Bossart G, Baden D, Ewing R, Roberts B, Wright S (1998) Brevetoxicosis in manatees (*Trichechus manatus latirostris*) from the 1996 epizootic: gross, histologic, and immunohistochemical features. *Toxicologic Pathology* **26**, 276-282.

Bottein Dechraoui M, Ramsdell J (2003) Type B brevetoxins show tissue selectivity for voltage-gated sodium channels: comparison of brain, skeletal muscle and cardiac sodium channels. *Toxicon* **41**, 919-927.

Bottein Dechraoui M, Wacksman J, Ramsdell J (2006) Species selective resistance of cardiac muscle voltage gated sodium channels: characterization of brevetoxin and ciguatoxin binding sites in rats and fish. *Toxicon* **48**, 702-712.

Bowen S (2011) *Diet of bottlenose dolphins Tursiops truncatus in the northwest Florida Panhandle and foraging behavior near Savannah, Georgia*. Master of Science, Savannah State University.

Brand LE, Compton A (2007) Long-term increase in *Karenia brevis* abundance along the Southwest Florida Coast. *Harmful Algae* **6**, 232-252.

Brenner S, Elgar G, Sanford R, *et al.* (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**, 265-268.

Bricelj V, Connell L, Konoki K, *et al.* (2005) Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. *Nature* **434**, 763-767.

Brodie III E, Brodie Jr E (1991) Evolutionary response of predators to dangerous prey: reduction of toxicity of newts and resistance of garter snakes in island populations. *Evolution* **45**, 221-224.

Brodie III E, Brodie Jr E (1999) Costs of exploiting poisonous prey: evolutionary trade-offs in a predator-prey arms race. *Evolution* **53**, 626-631.

Brodie Jr E, Ridenhour B, Brodie III E (2002) The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution* **56**, 2067-2082.

Brown J, Jardetzky T, Gorga J, *et al.* (1993) Three-dimensional struture of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39.

Caballero S, Baker CS (2010) Captive-born intergeneric hybrid of a Guiana and bottlenose dolphin: *Sotalia guianensis* x *Tursiops truncatus*. *Zoo Biology* **29**, 647-657.

Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W (2011) Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Molecular Ecology* **20**, 740-752.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catterall W (2000) From ionic currents to molecular review mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **26**, 13-25.

Catterall WA, Goldin AL, Waxman SG (2005) International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacological Reviews* **57**, 397-409.

Cestele S, Catterall W (2000) Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. *Biochimie* **82**, 883-892.

Chang ST, Linderman JJ, Kirschner DE (2008) Effect of multiple genetic polymorphisms on antigen presentation and susceptibility to *Mycobacterium tuberculosis* infection. *Infection and Immunity* **76**, 3221-3232.

Chen Y, Cunningham F, Rios D, *et al.* (2010) Ensembl variation resources. *BMC Bioinformatics* **11**, 293.

Chen Y, Dale T, Romanos M, *et al.* (2000) Cloning, distribution and functional analysis of the type III sodium channel from human brain. *European Journal of Neuroscience* **12**, 4281-4289.

Cheng C, White BJ, Kamdem C, *et al.* (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* **190**, 1417-1432.

Cohen S (2002) Strong positive selection and habitat-specific amino acid substituion patterns in *Mhc* from an estuarine fish under intense pollution stress. *Molecular Biology and Evolution* **19**, 1870-1880.

Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259-1267.

Complex Trait Consortium (2003) The nature and identification of quantitative trait loci. *Nature Reviews: Genetics* **4**, 911-916.

Connell LB, MacQuarrie SP, Twarog BM, Iszard M, Bricelj VM (2006) Population differences in nerve resistance to paralytic shellfish toxins in softshell clam, *Mya arenaria*, associated with sodium channel mutations. *Marine Biology* **150**, 1227-1236.

Davey JW, Cezard T, Fuentes-Utrilla P, *et al.* (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* **22**, 3151-3164.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews: Genetics* **12**, 499-510.

Dib-Hajj S, Tyrrell L, Escayg A, *et al.* (1999) Coding sequence, genomic organization, and conserved chromosomal location of the mouse gene Scn11a encoding the sodium channel NaN. *Genomics* **59**, 309-318.

Doherty P, Zinkernagel R (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**, 50-52.

Dong K (2007) Insect sodium channels and insecticide resistance. *Invertebrate Neuroscience* **7**, 17-30.

Du H, Zheng J, Wu M, Zhao Q, Wang D (2010) High MHC DQB variation and asymmetric allelic distribution in the endangered Yangtze finless porpoise, *Neophocaena phocaenoides asiaeorientalis*. *Biochemical Genetics* **48**, 433-449.

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

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews: Genetics* **5**, 435-445.

Elshire RJ, Glaubitz JC, Sun Q, *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6**, e19379.

Emslie S, Morgan G (1994) A catastrophic death assemblage and paleclimatic implications of Pliocene seabirds of Florida. *Science* **264**, 684-685.

Escayg A, MacDonald B, Meisler M, *et al.* (2000) Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. *Nature Genetics* **24**, 343-345.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology* **772**, 157-178.

Fauquier D, Flewelling L, Landsberg J, *et al.* (2007) Investigating brevetoxin-induced mortality in bottlenose dolphins stranded in central west Florida. Mote Marine Laboratory Technical Report No. 1222, 2007, 19 pp. Available at: https://dspace.mote.org/dspace/bitstream/2075/2313/1/MTR%201222.pdf

Fauquier DA, Flewelling LJ, Maucher J, *et al.* (2013) Brevetoxin in blood, biological fluids, and tissues of sea turtles naturally exposed to *Karenia brevis* blooms in central west Florida. *Journal of Zoo and Wildlife Medicine* **44**, 364-375.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.

Feder M, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Reviews: Genetics* **4**, 649-655.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the boostrap. *Evolution* **39**, 783-791.

Fire SE, Fauquier D, Flewelling LJ, *et al.* (2007) Brevetoxin exposure in bottlenose dolphins (*Tursiops truncatus*) associated with *Karenia brevis* blooms in Sarasota Bay, Florida. *Marine Biology* **152**, 827-834.

Fire SE, Flewelling LJ, Naar J, *et al.* (2008a) Prevalence of brevetoxins in prey fish of bottlenose dolphins in Sarasota Bay, Florida. *Marine Ecology Progress Series* **368**, 283-294.

Fire SE, Flewelling LJ, Wang ZH, *et al.* (2008b) Florida red tide and brevetoxins: Association and exposure in live resident bottlenose dolphins (*Tursiops truncatus*) in the eastern Gulf of Mexico, USA. *Marine Mammal Science* **24**, 831-844.

Fitzpatrick MJ, Ben-Shahar Y, Smid HM, *et al.* (2005) Candidate genes for behavioural ecology. *Trends in Ecology & Evolution* **20**, 96-104.

Flewelling L, Naar J, Abbott J, *et al.* (2005) Red tides and marine mammal mortalities. *Nature* **435**, 755-756.

Florida Fish and Wildlife Conservation Committee (2006) Save the Manatee Trust Fund: Annual Report 2005-2006 (ed. Reinert T). Available at: http://myfwc.com/research/manatee/trust-fund/annual-reports/2005-2006/

Florida Fish and Wildlife Conservation Committee (2007) Save the Manatee Trust Fund: Annual Report 2006-2007 (ed. Reinert T). Available at: http://myfwc.com/research/manatee/trust-fund/annual-reports/2006-2007/

Florida Fish and Wildlife Research Institute (2008) *164 years of suspected and documented red tides off Florida's west coast.* Available at: http://myfwc.com/research/redtide/monitoring/database/

Fratelli M, Goodwin LO, Orom UA, *et al.* (2005) Gene expression profiling reveals a signaling role of glutathione in redox regulation. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13998-14003.

Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* **186**, 207-218.

Gautier M, Foucaud J, Gharbi K, *et al.* (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology* **22**, 3766-3779.

Gawley R, Rein K, Jeglitsch G, *et al.* (1995) The relationship of brevetoxin 'length' and A-ring functionality to binding and activity in neuronal sodium channels. *Chemistry and Biology* **2**, 533-541.

Gaydos J (2006) Bottlenose Dolphins and Brevetoxins: A Coordinated Research and Response Plan. NOAA Technical Memorandum NMFS-OPR-32, 46 pp. Available at: http://www.nmfs.noaa.gov/pr/pdfs/health/bottlenose_brevetoxins.pdf

Geffeney S, Fujimoto E, Brodie E, Ruben P (2005) Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction. *Nature* **434**, 759-763.

George Jr A, Iyer G, Kleinfield R, Kallen R, Barchi R (1993) Genomic organization of the human skeletal muscle sodium channel gene. *Genomics* **15**, 598-606.

Germain R (1994) MHC-depedent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. *Cell* **76**, 287-299.

Germain R, Castellino F, Han R, *et al.* (1996) Processing and presentation of endocytically acquired protein antigens by MHC class II and class I molecules. *Immunological Reviews* **151**, 5-30.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**, 394-407.

Gnirke A, Melnikov A, Maguire J, *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**, 182-189.

Gunter G, Williams R, Davis C, Smith F (1948) Catastrophic mass mortalities of marine animals and coincidental phytoplankton bloom on the west coast of Florida, November 1946-August 1947. *Ecological Monographs* **18**, 309-324.

Hairston Jr NG, Ellner SP, Geber MA, Yoshida T, Fox JA (2005) Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters* **8**, 1114-1127.

Hannon G (2002) RNA interference. *Nature* **418**, 244-251.

Hedrick P (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.

Heil CA, Steidinger KA (2009) Monitoring, management, and mitigation of *Karenia* blooms in the eastern Gulf of Mexico. *Harmful Algae* **8**, 611-617.

Heimeier D, Baker CS, Russell K, *et al.* (2009) Confirmed expression of MHC class I and class II genes in the New Zealand endemic Hector's dolphin (*Cephalorhynchus hectori*). *Marine Mammal Science* **25**, 68-90.

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

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources* **11 Suppl 1**, 137-149.

Hinton M, Ramsdell JS (2008) Brevetoxin in two planktivorous fishes after exposure to *Karenia brevis*: implications for food-web transfer to bottlenose dolphins. *Marine Ecology Progress Series* **356**, 251-258.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews: Genetics* **6**, 95-108.

Hoekstra HE, Nachman M (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology* **12**, 1185-1194.

Hoelzel AR, Potter CW, Best PB (1998) Genetic differentiation between parapatric 'nearshore'and 'offshore'populations of the bottlenose dolphin. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 1177.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* **6**, e1000862.

Horton R, Wilming L, Rand V, *et al.* (2004) Gene map of the extended human MHC. *Nature Reviews: Genetics* **5**, 889-899.

Huang J, Wu C, Baden D (1984) Depolarizing action of a red-tide dinoflagellate brevetoxin on axonal membranes. *Journal of Pharmacology and Experimental Therapeutics* **229**, 615.

Huang Y, Zhu C, Li X, *et al.* (2012) Differentiated population structure of a genetically depauperate mangrove species *Ceriops tagal* revealed by both Sanger and deep sequencing. *Aquatic Botany* **101**, 46-54.

Hughes A, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 958-962.

Ingale A (2010) Antigenic epitopes prediction and MHC binder of a paralytic insecticidal toxin (ITX-1) of *Tegenaria agrestis* (hobo spider). *Open Access Bioinformatics* **2**, 97-103.

Ingersoll I (1882) On the fish mortality in the Gulf of Mexico. In: *Proceedings of the U.S. Natural Museum 4*, pp. 74-80.

Jones FC, Grabherr MG, Chan YF, *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Joost S, Bonin A, Bruford MW, *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.

Jost M, Hillis D, Lu Y, *et al.* (2008a) Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular Biology and Evolution* **25**, 1016.

Jost MC, Hillis DM, Lu Y, *et al.* (2008b) Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular Biology and Evolution* **25**, 1016-1024.

Jukes TH, Cantor, CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp. 21-132. Academic Press, New York.

Katunuma N, Matsunaga Y, Himeno K, Hayashi Y (2003) Insights into the roles of cathepsins in antigen processing and presentation revealed by specific inhibitors. *Biological Chemistry* **384**, 883-890.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.

Keenan K, McGinnity P, Cross T, Crozier W, Prodohl P (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782-788.

Kim SY, Li Y, Guo Y, *et al.* (2010) Design of association studies with pooled or un-pooled next-generation sequencing data. *Genetic Epidemiology* **34**, 479-491.

Kimm-Brinson K, Ramsdell J (2001) The red tide toxin, brevetoxin, induces embryo toxicity and developmental abnormalities. *Environmental Health Perspectives* **109**, 377.

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

Kirkpatrick B, Fleming LE, Squicciarini D*, et al.* (2004) Literature review of Florida red tide: implications for human health effects. *Harmful Algae* **3**, 99-115.

Kita Y, Inoko H, Shiina T, Tanaka S (2010) Development of newly paternity test method for dolphins using microsatellite and *MHC-DQB* markers. *Bulletin of Institute of Oceanic Research and Development, Tokai University* **31**, 1-8.

Klein J (1987) Origin of major histocompatibility copmlex polymorphism - the transspecies hypothesis. *Human Immunology* **19**, 155-162.

Klein J, Bontrop R, Dawkins R*, et al.* (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31**, 217-219.

Klugbauer N, Lacinova L, Flockerzi V, Hofmann F (1995) Structure and functional expression of a new member of the tetrodotoxin-sensitive voltage-activated sodium channel family from human neuroendocrine cells. *The EMBO Journal* **14**, 1084-1090.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Kreuder C, Mazet JAK, Bossart GD*, et al.* (2002) Clinicopathologic features of suspected brevetoxicosis in double-crested cormorants (*Phalacrocorax auritus*) along the Florida Gulf coast. *Journal of Zoo and Wildlife Medicine* **33**, 8-15.

Krützen M, Valsecchi E, Connor RC, Sherwin W (2001) Characterization of microsatellite loci in *Tursiops aduncus*. *Molecular Ecology Notes* **1**, 170-172.

Kucklick J, Yordy JE (2010) *A geographical perspective on bottlenose dolphin contamination.* Sarasota Dolphin Research Program, Sarasota, FL. Available at: http://sarasotadolphin.org/2010/01/06/a-geographical-perspective-on-bottlenose-dolphin-contamination/

Kusek K, Vargo G, Steidinger K (1999) *Gymnodinium breve* in the field, in the lab, and in the newspaper- a scientific and journalistic analysis of Florida red tides. *Contributions in Marine Science* **34**, 1-228.

Laine AL, Burdon JJ, Dodds PN, Thrall PH (2011) Spatial variation in disease resistance: from molecules to metapopulations. *Journal of Ecology* **99**, 96-112.

Leffler EM, Bullaughey K, Matute DR, *et al.* (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology* **10**, e1001388.

Lennon NJ, Lintner RE, Anderson S, *et al.* (2010) A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biology* **11**, R15.

Lindblad-Toh K, Garber M, Zuk O, *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-482.

Magaña HA, Contreras C, Villareal TA (2003) A historical assessment of *Karenia brevis* in the western Gulf of Mexico. *Harmful Algae* **2**, 163-171.

Magrane M, UniProt Consortium (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database: The Journal of Biological Databases and Curation* **2011**, bar009.

Mamanova L, Coffey AJ, Scott CE, *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods* **7**, 111-118.

Mancia A, Warr G, Almeida J, *et al.* (2010) Transcriptome profiles: Diagnostic signature of dolphin populations. *Estuaries and Coasts* **33**, 919-929.

Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. *Nature Reviews: Genetics* **11**, 499-511.

Marsh SG, Albert ED, Bodmer WF, *et al.* (2005) Nomenclature for factors of the HLA system, 2004. *Human Immunology* **66**, 571-636.

Martin A, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4087-4091.

Martin SH, Dasmahapatra KK, Nadeau NJ, *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**, 1817-1828.

Mase B, Jones W, Ewing R, *et al.* (2000) Epizootic in bottlenose dolphins in the Florida Panhandle: 1999-2000. In: *2000 Proceedings of the AAZV and IAAM Joint Conference*, pp. 522-525.

McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. *Human Molecular Genetics* **17**, R156-165.

McHugh KA, Allen JB, Barleycorn AA, Wells RS (2011) Severe *Karenia brevis* red tides influence juvenile bottlenose dolphin (*Tursiops truncatus*) behavior in Sarasota Bay, Florida. *Marine Mammal Science* **27**, 622-643.

Menashe I, Rosenberg P, Chen B (2008) PGA: power calculator for case-control genetic association analyses. *BMC Genetics* **9**, 36.

Meyer J, Di Giulio R (2003) Heritable adaptation and fitness costs in killifish (*Fundulus heteroclitus*) inhabiting a polluted estuary. *Ecological Applications* **13**, 490-503.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**, 66-73.

Munguia-Vega A, Esquer-Garrigos Y, Rojas-Bracho L*, et al.* (2007) Genetic drift vs. natural selection in a long-term small isolated population: major histocompatibility complex class II variation in the Gulf of California endemic porpoise (*Phocoena sinus*). *Molecular Ecology* **16**, 4051-4065.

Murray B, Malik S, White B (1995) Sequence variation at the major histocompatibility complex locus *DQB* in beluga whales (*Delphinapterus leucas*). *Molecular Biology and Evolution* **12**, 582-593.

Murray B, White B (1998) Sequence variation at the major histocompatibility complex *DRB* loci in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*). *Immunogenetics* **48**, 242-252.

Murrell RN, Gibson JE (2009) Brevetoxins 2, 3, 6, and 9 show variability in potency and cause significant induction of DNA damage and apoptosis in Jurkat E6-1 cells. *Archives of Toxicology* **83**, 1009-1019.

Naar J, Flewelling L, Lenzi A*, et al.* (2007) Brevetoxins, like ciguatoxins, are potent ichthyotoxic neurotoxins that accumulate in fish. *Toxicon* **50**, 707-723.

Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics* **26**, 484-492.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**, 2841-2847.

National Marine Fisheries Service (2004) Interim Report on the Bottlenose Dolphin (*Tursiops truncatus*) Unusual Mortality Event along the Panhandle of Florida, March-April 2004, 36 pp. Available at: http://www.nmfs.noaa.gov/pr/pdfs/health/ume_bottlenose_2004.pdf

Nei M, Chesser R (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253-259.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418-426.

Nielsen EE, Hemmer-Hansen J, Poulsen NA, *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews: Genetics* **12**, 443-451.

Niranjan SK, Deb SM, Kumar S, *et al.* (2010) Allelic diversity at MHC class II DQ loci in buffalo (*Bubalus bubalis*): evidence for duplication. *Veterinary Immunology and Immunopathology* **138**, 206-212.

O'Shea TJ, Rathbun GB, Bonde RK, Buergelt CD, Odell DK (1991) An epizootic of Florida manatees associated with a dinoflagellate bloom. *Marine Mammal Science* **7**, 165-179.

Pabinger S, Dander A, Fischer M, *et al.* (2013) A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*. doi: 10.1093/bib/bbs086

Perry GH (2014) The promise and practicality of population genomics research with endangered species. *International Journal of Primatology* **35**, 55-70.

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology* **19**, 5332-5344.

Pflugmacher S, Wiegand C, Werner S, Schroder H, Kankaanpaa H (2005) Activity and substrate specificity of cytosolic and microsomal glutathione S-transferase in Australian black tiger prawns (*Penaeus monodon*) after exposure to cyanobacterial toxins. *Environmental Toxicology* **20**, 301-307.

Piertney SB, Webster LM (2010) Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica* **138**, 419-432.

Plummer N, Galt J, Jones J, *et al.* (1998) Exon organization, coding sequence, physical mapping, and polymorphic intragenic markers for the human neuronal sodium channel gene SCN8A. *Genomics* **54**, 287-296.

Poli M, Mende T, Baden D (1986) Brevetoxins, unique activators of voltage-sensitive sodium channels, bind to specific sites in rat brain synaptosomes. *Molecular Pharmacology* **30**, 129.

Powell JR, Wells RS (2011) Recreational fishing depredation and associated behaviors involving common bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida. *Marine Mammal Science* **27**, 111-129.

Pyenson ND, Gutstein CS, Parham JF, *et al.* (2014) Repeated mass strandings of Miocene marine mammals from Atacama Region of Chile point to sudden death at sea. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **281**, 20133316.

R Core Team (2013) *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria.

Rabert D, Koch B, Ilnicka M, *et al.* (1998) A tetrodotoxin-resistant voltage-gated sodium channel from human dorsal root ganglia, hPN3/SCN10A. *Pain* **78**, 107-114.

Radwan F, Ramsdell J (2006) Characterization of *in vitro* oxidative and conjugative metabolic pathways for brevetoxin (PbTx-2). *Toxicological Sciences* **89**, 57.

Radwan FFY, Wang ZH, Ramsdell JS (2005) Identification of a rapid detoxification mechanism for brevetoxin in rats. *Toxicological Sciences* **85**, 839-846.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Redlow T, Foley A, Singel K (2003) Sea turtle mortality associated with red tide events in Florida. In: *Proceedings of the 22nd Annual Symposium on Sea Turtle Biology and Conservation*, U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SEFSC-503, pp. 272.

Reimand J, Arak T, Vilo J (2011) g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Research* **39**, W307-W315.

Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (2013) Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PloS One* **8**, e80422.

Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112-113**, 183-198.

Rice W (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Rooney A, Merritt D, Derr J (1999) Microsatellite diversity in captive bottlenose dolphins (*Tursiops truncatus*). *Journal of Heredity* **90**, 228-231.

Rosel P, Hansen L, Hohn A (2009) Restricted dispersal in a continuously distributed marine species: common bottlenose dolphins *Tursiops truncatus* in coastal waters of the western North Atlantic. *Molecular Ecology* **18**, 5030-5045.

Rosel PE, Block B (1996) Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Marine Biology* **125**, 11-22.

Rosel PE, Forgetta V, Dewar K (2005) Isolation and characterization of twelve polymorphic microsatellite markers in bottlenose dolphins (*Tursiops truncatus*). *Molecular Ecology Notes* **5**, 830-833.

Rosel PE, France S, Wang J, Kocher T (1999) Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. *Molecular Ecology* **8**, S41-S54.

Roselli F, Livrea P, Jirillo E (2006) Voltage-gated sodium channel blockers as immunomodulators. *Recent Patents on CNS Drug Discovery* **1**, 83-91.

SAS Institute Inc. (2014) JMP® 11 Fitting Linear Models. SAS Institute Inc., Cary, NC.

Sas KM, Baatz JE (2010) Brevetoxin-2 induces an inflammatory response in an alveolar macrophage cell line. *International Journal of Hygiene and Environmental Health* **213**, 352-358.

Satta Y, O'Huigin C, Takahata N, Klein J (1994) Intensity of natural selection at the major histocompatibility complex. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 7184-7188.

Sayer A, Hu Q, Bourdelais A, Baden D, Gibson J (2005) The effect of brevenal on brevetoxin-induced DNA damage in human lymphocytes. *Archives of Toxicology* **79**, 683-688.

Sboner A, Mu X, Greenbaum D, Auerbach R, Gerstein M (2011) The real cost of sequencing: higher than you think! *Genome Biology* **12**, 125.

Scott M, Wells R, Irvine WA (1990) A Long-Term Study of Bottlenose Dolphins on the West Coast of Florida. In: *The Bottlenose Dolphin* (eds. Leatherwood S, Reeves RR), pp. 235-244. Academic Press, Inc., San Diego.

Seeb JE, Carvalho G, Hauser L, *et al.* (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources* **11 Suppl 1**, 1-8.

Sellas AB, Wells RS, Rosel PE (2005) Mitochondrial and nuclear DNA analyses reveal fine scale geographic structure in bottlenose dolphins (*Tursiops truncatus*) in the Gulf of Mexico. *Conservation Genetics* **6**, 715-728.

Shannon W, Culverhouse R, Duncan J (2003) Analyzing microarray data using cluster analysis. *Pharmacogenomics* **4**, 41-52.

Sherry S, Ward M, Kholodov M, *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308-311.

Singer L, Lee T, Rosen K, Baden D, Abraham W (1998) Inhaled Florida red tide toxins induce bronchoconstriction (BC) and airway hyperresponsiveness (AHR) in sheep. *American Journal of Respiratory and Critical Care Medicine* **157**, A158.

Slade R (1992) Limited MHC polymorphism in the southern elephant seal: Implications for MHC evolution and marine mammal population biology. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **249**, 163-171.

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**, 1742-9994.

Spurgin LG, van Oosterhout C, Illera JC, *et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**, 5213-5225.

Stahl E, Bishop J (2000) Plant-pathogen arms races at the molecular level. *Current Opinion in Plant Biology* **3**, 299-304.

Steidinger K (1975) Basic factors influencing red tides. In: *Proceedings of the First International Conference on Toxic Dinoflagellate Blooms* (ed. LoCicero V), pp. 153-162.

Steidinger K (2009) Historical perspective on *Karenia brevis* red tide research in the Gulf of Mexico. *Harmful Algae* **8**, 549-561.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution* **18**, 94-101.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Sudarsanam S, Virca GD, March CJ, Srinivasan S (1992) An approach to computer-aided inhibitor design: application to cathepsin L. *Journal of Computer-Aided Molecular Design* **6**, 223-233.

Sylvestre J, Tasaka S (1985) On the intergeneric hybrids in cetaceans. *Aquatic Mammals* **11**, 101-108.

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews: Genetics* **3**, 391-A396.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-depedent selection and polymorphism of major histocompatibility complex. *Genetics* **124**, 967-978.

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Moelcular Biology and Evolution* **9**, 678-687.

Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Tautz D, Ellegren H, Weigel D (2010) Next generation molecular ecology. *Molecular Ecology* **19 Suppl 1**, 1-3.

Thierry-Mieg D, Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biology* **7**, S12.

Thompson JN (1998) Rapid evolution as an ecological process. *Trends in Ecology & Evolution* **13**, 329-332.

Ting J, Trowsdale J (2002) Genetic control of MHC class II expression. *Cell* **109**, S21-S33.

Trowsdale J, Groves V, Arnason A (1989) Limited MHC polymorphism in whales. *Immunogenetics* **29**, 19-24.

Tsuji K, Aizawa M, Sasazuki T (1992) *HLA 1991: Proceedings of the eleventh international histocompatibility workshop and conference* Oxford University Press.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* **3**, e285.

Twiner MJ, Fire S, Schwacke L, *et al.* (2011) Concurrent exposure of bottlenose dolphins (*Tursiops truncatus*) to multiple algal toxins in Sarasota Bay, Florida, USA. *PloS One* **6**, e17394.

Twiner MJ, Flewelling LJ, Fire SE, *et al.* (2012) Comparative analysis of three brevetoxin-associated bottlenose dolphin (*Tursiops truncatus*) mortality events in the Florida Panhandle region (USA). *PloS One* **7**, e42974.

Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Molecular Ecology* **5**, 151-156.

van der Poel J, Groenen M, Dijkhof R, Ruyter D, Giphart M (1990) The nucleotide seuqence of the bovine MHC class II alpha genes: *DRA, DQA,* and *DYA*. *Immunogenetics* **31**, 29-36.

Van Dolah F (2000) Marine algal toxins: origins, health effects, and their increased occurrence. *Environmental Health Perspectives* **108**, 133.

Van Dolah FM, Doucette GJ, Gulland FMD, Rowles TL, Bossart GD (2003) Impacts of algal toxins on marine mammals. In: *Toxicology of marine mammals* (eds. Vos J, Bossart G, Fournier M, O'Shea TJ), pp. 247-269. Taylor & Francis, London.

Van Tassell CP, Smith TP, Matukumalli LK, *et al.* (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**, 247-252.

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

Vassilakos D, Natoli A, Dahlheim M, Hoelzel AR (2009) Balancing and directional selection at exon-2 of the MHC DQB1 locus among populations of odontocete cetaceans. *Molecular Biology and Evolution* **26**, 681.

Villanueva-Noriega MJ, Baker CS, Medrano-Gonzalez L (2013) Evolution of the MHC-DQB exon 2 in marine and terrestrial mammals. *Immunogenetics* **65**, 47-61.

Walsh CJ, Leggett SR, Carter BJ, Colle C (2010) Effects of brevetoxin exposure on the immune system of loggerhead sea turtles. *Aquatic Toxicology* **97**, 293-303.

Walsh CJ, Leggett SR, Henry MS, *et al.* (2009) Cellular metabolism of brevetoxin (PbTx-2) by a monocyte cell line (U-937). *Toxicon* **53**, 135-145.

Walsh CJ, Luer CA, Noyes DR (2005) Effects of environmental stressors on lymphocyte proliferation in Florida manatees, *Trichechus manatus latirostris*. *Veterinary Immunology and Immunopathology* **103**, 247-256.

Walsh J, Jolliff J, Darrow B, *et al.* (2006) Red tides in the Gulf of Mexico: Where, when, and why? *Journal of geophysical research* **111**, 1.

Wang Q, Li Z, Shen J, Keating M (1996) Genomic organization of the human SCN5A gene encoding the cardiac sodium channel. *Genomics* **34**, 9-16.

Waring G, Josephson E, Maze-Foley K, Rosel PE, editors. (2013) U.S. Atlantic and Gulf of Mexico Marine Mammal Stock Assessments--2012. NOAA Technical Memorandum NMFS NE 223, 419 pp. Available at: http://www.nefsc.noaa.gov/publications/tm/tm223/

Washburn BS, Baden DG, Gassman NJ, Walsh PJ (1994) Brevetoxin: tissue distribution and effect on cytochrome P450 enzymes in fish. *Toxicon* **32**, 799-805.

Washburn BS, Vines C, Baden DG, Hinton D, Walsh PJ (1996) Differential effects of brevetoxin and B-naphthoflavone on xenobiotic metbaolizing enzymes in striped bass (*Morone saxatilis*). *Aquatic Toxicology* **35**, 1-10.

Watkins S, Reich A, Fleming L, Hammond R (2008) Neurotoxic shellfish poisoning. *Marine Drugs* **6**, 431.

Weir B, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Widmark J, Sundstrom G, Ocampo Daza D, Larhammar D (2011) Differential evolution of voltage-gated sodium channels in tetrapods and teleost fishes. *Molecular Biology and Evolution* **28**, 859-871.

Wilkening S, Chen B, Bermejo JL, Canzian F (2009) Is there still a need for candidate gene approaches in the era of genome-wide association studies? *Genomics* **93**, 415-419.

Xu S, Chen B, Zhou K, Yang G (2008) High similarity at three MHC loci between the baiji and finless porpoise: trans-species or convergent evolution? *Molecular Phylogenetics and Evolution* **47**, 36-44.

Xu S, Ren W, Zhou X, Zhou K, Yang G (2010) Sequence polymorphism and geographical variation at a positively selected MHC-DRB gene in the finless porpoise (*Neophocaena phocaenoides*): implication for recent differentiation of the Yangtze finless porpoise? *Journal of Molecular Evolution* **71**, 6-22.

Xu S, Sun P, Zhou K, Yang G (2007) Sequence variability at three MHC loci of finless porpoises (*Neophocaena phocaenoides*). *Immunogenetics* **59**, 581-592.

Xu SX, Ren WH, Li SZ, *et al.* (2009) Sequence polymorphism and evolution of three cetacean MHC genes. *Journal of Molecular Evolution* **69**, 260-275.

Yang WC, Chou LS, Hu JM (2007) Molecular characterization of major histocompatibility complex class II DQB and DRB genes in bottlenose dolphins (*Tursiops truncatus* and *T. aduncus*) from the Western Pacific. *Zoological Studies* **46**, 664-679.

Yang WC, Hu JM, Chou LS (2008) Sequence variation of MHC Class II DQB gene in bottlenose dolphins (*Tursiops truncatus*) from Taiwanese waters. *Taiwania* **53**, 42-50.

Yu F, Catterall W (2003) Overview of the voltage-gated sodium channel family. *Genome Biology* **4**, 207.

Zavodna M, Grueber CE, Gemmell NJ (2013) Parallel tagged next-generation sequencing on pooled samples - a new approach for population genetics in ecology and conservation. *PloS One* **8**, e61471.

BIOGRAPHY

Kristina Marstrand Cammen
Born: May 12, 1985, Damariscotta, Maine

Education

Doctor of Philosophy, Ecology (2014)
Certificate in College Teaching (2014)
Duke University, Nicholas School of the Environment, Durham, North Carolina, USA

Master of Philosophy, Zoology (2008)
University of Cambridge, Cambridge, United Kingdom

Bachelor of Science, General Biology & Psychology (2007)
University of Maryland, College Park, Maryland, USA

Publications

Cammen, K., Hoffman, J.I., Knapp, L.A., Harwood, J., Amos, W. (2011). Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Molecular Ecology, 20*, 740-752.

Soulen, B.K.*, Cammen, K.*, Schultz, T.F., Johnston, D.W. (2013). Factors affecting harp seal (*Pagophilus groenlandicus*) strandings in the Northwest Atlantic. *PLoS ONE 8(7):* e68779. doi:10.1371/journal.pone.0068779.
*authors contributed equally to this manuscript

Bossart, G., Peden-Adams, M., Romano, T., Rice, C., Fair, P., Goldstein, J., Cammen, K., Reif, J. (2008). Hematological, biochemical and immunological findings in Atlantic bottlenose dolphins (*Tursiops truncatus*) with orogenital papillomas. *Aquatic Mammals, 34*, 166-177.

Fellowships and Awards

Katherine Goodman Stern Fellowship, Duke University Graduate School (2013)
Special Event Award, American Genetic Association (2013)
Student Poster Award, US National HAB Committee (2011)
Steven Berkeley Conservation Fellowship, American Fisheries Society (2010)
PADI Foundation Grant (2010)
Kaupe Research Grant, Duke University Marine Lab (2009)
Duke University Scholars Program Fellowship (2008)
James B. Duke Fellowship, Duke University (2008-2012)
Cambridge University Overseas Trust (2007)

Professional Affiliations

American Genetic Association, Society for Marine Mammalogy, American Association for the Advancement of Science