FUNCTIONAL GENETICS OF EASTERN OYSTERS (*CRASSOSTREA VIRGINICA*) ACROSS SALINITY VARIATION IN A SINGLE ESTUARY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Laura Elizabeth Eierman

May 2014

FUNCTIONAL GENETICS OF EASTERN OYSTERS (*CRASSOSTREA VIRGINICA*) ACROSS SALINITY VARIATION IN A SINGLE ESTUARY

Laura Elizabeth Eierman, Ph. D.

Cornell University 2014

Understanding the interaction between selection and phenotypic plasticity is important for predicting species' persistence in rapidly changing environments. For sessile organisms, phenotypic plasticity is a typical mechanism for responding to environmental variation. However, the additional characteristics of high fecundity and widely dispersing offspring present an opportunity for selection to reshape the functional genetic composition of populations across habitat heterogeneities every generation. I tested for this pattern of recurrent viability selection in eastern oysters, *Crassostrea virginica,* using two experimental approaches. First, I made within-reef pair crosses from low, intermediate and high salinity source oysters and analyzed larval survival after a ten day exposure to 10 and 30 salinity treatments. The parental reef source-by-larval salinity treatment interaction term was a significant predictor of larval survival. Second, I sequenced and assembled the oyster transcriptome *de novo* in order to conduct RNA-seq to identify differential gene expression patterns in response to salinity treatment and oyster reef source. The samples for RNA-seq were twenty-four adult oysters collected from high and low salinity source reefs and acclimated for 9 weeks in 10 and 30 salinity common gardens. A total of 9,921 reference transcriptome contigs (reftigs; 23.6%) were significantly differentially expressed (DE), with 0.6% of all reftigs DE for the reef source, 18.9% DE for treatment salinity, and 13.9% DE for the reef-by-treatment factor. The reftigs responding to treatment and the reef-

by-treatment factors demonstrate a genomically pervasive pattern of plastic gene expression in response to salinity. Additionally, the abundant genotype-by-environment patterns suggest that the history of selection at each reef is generating different plastic responses after acclimation to the same osmotic condition. Overall, the reef-specific patterns of gene expression and larval survival indicate that oyster responses to habitat heterogeneity are shaped by both phenotypic plasticity and recurrent viability selection. Furthermore, the larval results suggest that the functional plasticity differences observed in the adults were heritable. Studies on the interaction between plasticity and evolutionary responses typically classify plasticity as acting within generations and selection as acting between generations. I suggest that for high dispersal species with type III survivorship, intra-generational selection can shape patterns of plasticity across habitats.

BIOGRAPHICAL SKETCH

Laura Elizabeth Eierman was born in Jarrettsville, MD and was raised by her parents, Frederick and Roberta, along with her sister Elaine. When she was two years old, her grandfather, Howard Vonderheide, spent an entire day coaxing her into the water at Ocean City, MD. After that first introduction to the waves, Laura fell in love with the ocean, and she decided early on to become a marine scientist. She attended Salisbury University in Salisbury, MD, graduating summa cum laude within the Bellavance Honor's program in 2002. She dual majored in Biology and Environmental Science with a concentration in Marine Science and a minor in Chemistry. Laura immediately began a Master of Science in Marine Biology at the University of Massachusetts, Dartmouth where she was advised by Dr. Richard Connor. She spent three years studying wild dolphin foraging behavior and habitat selection along with fish communities and acoustics at Turneffe Atoll, Belize. Following the completion of her MS degree, Laura proceeded to teach general science at Sparrows Point Middle School in Baltimore County, MD while earning a Master of Arts in Teaching at Towson University. She started a school-wide environmental club and recycling program. She was awarded the TABCO Rookie of the Year teaching award in 2006. During her time as a teacher, Laura was blessed to live across the street from her grandmother, Roberta Vonderheide. After teaching for two years, Laura began a PhD program in Natural Resources at Cornell University in the fall of 2007. During the past seven years, she has pursued both her passion for marine conservation through her research on eastern oysters and her love of teaching. After completing her degree, Laura plans to work in academia, teaching at the undergraduate level while providing her students with research opportunities.

Dedicated to all of the teachers in my life

ACKNOWLEDGMENTS

I could not have completed this dissertation without the support, guidance, patience and generosity of many people. First, I would like to thank my committee chair, Dr. Matthew Hare. I first met Matt at the University of Maryland, College Park in October, 2006. We discussed my research interests and the skills I wanted to improve through my PhD work. At the end of the meeting, Matt shook my hand and said, "It was a pleasure talking with you." Unlike other potential advisors I had met, Matt was sincere. I knew immediately that if he offered me a position as a graduate student, I would say yes. Over the course of my time at Cornell, Matt has met with me weekly, given me guidance, and asked me the hard questions to help me and this project be the best possible. He is a true mentor.

My committee members have also been a great support. Before my first semester began, I knocked on Dr. Patrick Sullivan's door and said, "I am going to need a lot of help with statistics." Due to Pat's guidance and extraordinary patience, I now have a thorough understanding of statistics. Additionally, Dr. Jim Morin and Dr. Brian Lazzaro have provided invaluable feedback, particularly regarding the scope of the second and third papers.

I would like to thank both of my research assistants, Hilary Kates (summer 2010) and Katie Kirk (summer 2011). They provided invaluable help during field work, and they valiantly made it through long nights of spawning oysters and counting larvae. They also made our time in the field enjoyable through trips to the beach, movie nights, and excursions to find ice cream.

Additionally, I would like to acknowledge the generosity of Dr. John Scarpa, the staff at Harbor Branch Oceanographic Institute, Dr. David Bushek and the staff at Haskin Shellfish Research Laboratory. Both Dr. Scarpa and Dr. Bushek kindly provided me with hatchery space at their respective facilities for my larval experiment, and they shared their expertise while I learned how to condition and spawn oysters.

Also, thank you to Dr. Willie Fetzer and Dr. Suzy Strickler. During a chat with Willie about data analysis, he brought up the 'survival' R package as a potential means for analyzing my larval survival data. Suzy graciously shared a custom pipeline for calculating dN/dS ratios between sequences and helped me troubleshoot the pipeline for use with my data.

Graduate programs are intense and the community of people in your life makes all of the difference. I am grateful for the camaraderie of the Hare Lab members. Harmony Borchardt-Wier provided guidance through the nitty gritty of lab protocol. The undergraduates, particularly Courtney Verscaj, were helpful in completing analyses. Finally, Catherine Sun has been one of the best lab mates anyone could ask for. We have brainstormed, vented, run half marathons, and consumed a lot of coffee together. Cat, thank you for everything. Additionally, the graduate students of DNR are a family, supporting and cheering each other on. I would like to specifically thank Laura Martin and Taza Schaming, two members of my cohort. Their friendship from the first day we started at Cornell University has kept me going all of these years.

Finally, a huge heart-felt thank you to my family and friends. Daisy Kaplan has been one of my closest friends since we met at UMASS Dartmouth in 2002. A fellow PhD student, Daisy has been the friend who "gets it" without having to explain, and has regularly served as a diversion with stories of her adventures in Manhattan. My childhood friends, April, Rebecca and Adrienne, have remained close, visiting me over the years and keeping me grounded. Thank you to my parents who have been my greatest supporters. They have stopped what they were doing for every phone call, and they have visited whenever I have needed company or help. In general, they have been the parents that every graduate student should have. Thank you to my sister and her family - Drew, Molly and Brody - who have been a refreshing break from school and research. Thank you to my grandma, who is always happy to listen and cheer me on. Finally, thank you to Scott Flickinger. Scott has been the most immediate person in my life during my

PhD: the person who sat on the couch next to me while I studied for my A exam and the person who got the first phone call when I passed my B exam. Scott, thank you for your patience while I pursued my PhD, and ultimately, thanks for making me laugh during the ride.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

To understand how species will persist in the face of anthropogenic environmental change, researchers must examine the mechanisms through which populations will respond to a shifting environment. Current models of population response to environmental variation focus on three mechanisms: (1) dispersal to new locations, (2) genetic evolution to new conditions within the current population range, and (3) phenotypic plasticity. Niche modeling predicts a species' ability to track the preferred habitat through space without including evolutionary change (Thomas et al. 2004) and therefore does not account for the genetic variation and phenotypic plasticity that may lead to population persistence without changing location (Chevin et al. 2010). Alternatively, evolutionary models examine the ability of a population to respond to environmental change within the existing population range through genetic adaptation (Lynch and Lande 1993, Burger and Lynch 1995, Willi and Hoffmann 2009), but these models typically do not include phenotypic plasticity or the impact of broad dispersal (reviewed in Hoffman and and Sgrò 2011; e.g. Hellmann and Pineda-Krch 2007). Recently, models have begun to incorporate phenotypic plasticity, concluding that phenotypic plasticity may increase the rate of genetic adaptation (Chevin et al. 2010, Schlichting and Wund 2014), but these models often do not consider the consequences of frequent dispersal among habitats. In general, phenotypic plasticity has been modeled primarily as a within-generation process whereas evolutionary adaptation has been modeled as a trans-generational process. While all of these models provide valuable insight, the consequences of the combined effects of phenotypic plasticity, high gene flow and strong selection from environmental variation remains an important question for understanding how many types of species may persist in a rapidly changing environment.

Organisms with high fecundity, a large percentage of early mortality, and high gene flow provide a valuable empirical context for testing the combined effects of phenotypic plasticity, dispersal and strong selection from environmental variation. First, the 'elm-oyster' model, proposed by Williams (1975), emphasized that for sexually-reproducing sedentary organisms with high fecundity and high early mortality, abundant genetic variation is critical for the species' success. Williams was trying to understand species whose cohorts regularly experience unpredictable and highly variable conditions during dispersal and after settlement. For such species, the likelihood of any given offspring having high fitness in the habitat where it settles is slim, but the overall success of a cohort is improved by genotypic diversity among individuals. By chance, at least a few offspring will settle in an environment in which they can survive to adulthood (Thorson 1950). The heterogeneous habitats within the dispersal distance of offspring generate selective sieves through which the diverse genotypes of offspring are filtered. Mortality due to low fitness in the post-settlement environment has been described as a phenotype-environment mismatch (Marshall et al. 2010). Every generation produces a new set of genetically variable migrants that disperse and settle across the gauntlet of environmental variation.

Second, for species with an 'elm-oyster' or similar life history, phenotypic plasticity is an important means by which sessile adults and juvenile migrants can maintain homeostasis in temporally and spatially variable conditions (Scheiner 1993). The limits and strength of plasticity, however, can vary by genotype, producing different plastic phenotypes for selection to act upon. Thus, organisms with high fecundity, type III survival, which is defined by high mortality as juveniles and low mortality as adults, and high gene flow may be significantly

impacted by intra-generational selection, with every cohort passing through an environmental selective sieve, despite broad phenotypic plasticity.

The estuarine eastern oyster (*Crassostrea virginica*) is well suited for testing the interplay of natural selection, phenotypic plasticity and genetic adaptation. The species is highly fecund with a single female producing millions of eggs per spawn (Davis and Chanley 1955). Additionally, oysters have high genetic diversity (Eierman and Hare in press; Zhang et al. 2014), such that resulting larvae produced from broadcast spawning include an enormous diversity of genotypes. With a planktonic period of 2-3 weeks, larvae have a dispersal potential of 10 to 100km (Kennedy 1996), a distance that covers the full length of a typical estuary. Although a strong genetic cline has been found between the Atlantic and Gulf of Mexico regions of the eastern oyster distribution (Reeb and Avise 1990, Karl and Avise 1992, Hare and Avise 1996), allele frequencies of most putatively neutral markers within the Atlantic and Gulf regions are homogenous, suggesting high gene flow within and among the estuaries of each region (Reeb and Avise 1990, Karl and Avise 1990, Burocker 1983, McDonald et al. 1996, He et al. 2012). Furthermore, the habitat of oysters is highly variable. Estuaries are dynamic environments that vary spatially from the freshwater upstream reaches to the oceanic mouths as well as temporally with changing tides. Thus, a well-mixed pool of dispersing larvae, as suggested by the evidence of high gene flow from neutral markers, can result in juvenile settlement in an environment very different from the natal source, leading to phenotype-environment mismatches and functional differentiation between oyster reefs.

However, eastern oysters are the quintessential example of phenotypic plasticity. Temporal variability in the environment, like that experienced by sessile juvenile and adult oysters, should favor physiological plasticity so that traits can track environmental change and

maintain homeostasis (Scheiner 1993). If plasticity is broad enough to match conditions across environmental extremes, phenotypic plasticity can mitigate the impact of environmental change through trait optimization (Chevin and Lande 2010, Chevin et al. 2010) and, under an assumption of a well-mixed larval pool, lead to homogeneous genetic variation in oysters throughout the estuary. The plastic responses observed in oysters occur in traits ranging from morphology to physiology. Shells can grow around physical obstacles and increase in thickness as a response to abiotic and biotic pressures such as predation (Newell et al. 2007, Johnson and Smee 2012, Lord and Whitlatch 2012, Robinson et al. 2014). Oyster physiology is also highly plastic. As osmoconformers, oysters regulate cell volume to maintain a near constant size despite frequent changes in environmental and therefore extracellular salinity. At the cellular level, cell volume homeostasis is an example phenotypic buffering (Reusch 2013). However, the up- and down-regulation of gene expression to move solutes, and therefore water, into or out of the cell in response to hyper- or hypoosmotic stress, is highly plastic.

The eastern oyster response to osmotic pressure across the salinity gradient of an estuary sets up a natural experiment for testing a model of intra-generational selection against a null expectation of broad phenotypic plasticity. For adult oysters, temporal variation in salinity due to tide and weather events requires the constant alteration of the osmotic bulk to maintain cell volume through plastic gene expression. In contrast to this temporal variation, newly settled juveniles that have dispersed along the salinity gradient of the estuary must acclimate to an entirely different salinity regime than the one experienced by their parents, setting up the potential for phenotype-environment mismatches over and above the capability of plastic physiological responses.

The process of osmoregulation by osmoconformers is an exquisite adaptation that requires the well-orchestrated interaction of plastic gene expression. Knowledge of the mechanisms used to manipulate cell volume through the movement of solutes and water into and out of cells provides insight into the required plasticity to maintain homeostasis as well as the genetic landscape on which selection may act. Inorganic solutes such as Na^+, K^+ and Cl^- account for a maximum of 500 to 600 mOsm of the total osmotic pressure in the cytoplasm of marine invertebrates (Kirschner 1991), while cytoplasm at the ambient salinity of seawater has an osmotic concentration of 1100 mOsm. The remaining osmotic bulk is made up of organic osmolytes consisting of free amino acids, primarily taurine, alanine, proline, glycine and glutamic acid, and quaternary amine compounds. While few studies have examined the regulation of inorganic ions, the role of organic osmolytes has been studied more thoroughly. The movement of organic osmotic bulk in response to osmotic pressure can be separated into two responses depending on whether environmental conditions are hyperosmotic or hypoosmotic to the cell osmolality.

1. Response to Hyperosmotic Conditions

Under hyperosmotic pressure, cells increase osmotic bulk in order to prevent the loss of water to the extracellular environment. The accumulation of organic osmolytes can be broken into three stages: (1) 24 hour response, (2) two week response, and (3) two month response.

In the first 24 hours of exposure to increased salinity, cells rapidly accumulate alanine and betaine (Baginski and Pierce 1978, Pierce et al. 1992, Bishop et al. 1994, Deaton 2001, Hosoi et al. 2003). In *Crassostrea gigas*, this response occurred within 0-2 hours of exposure to 100% seawater (Hosoi et al. 2003). A peak concentration of alanine is reached within 24 hours and then declines (Baginski and Pierce 1978, Hosoi et al. 2003). Alanine is a product of protein

catabolism followed by transamination (Bishop et al. 1994, Deaton 2009). Additionally, alanine may be synthesized from pyruvate (Baginski and Pierce 1978). In keeping with these metabolic sources, the control of alanine is associated with the regulation of pyruvate dehydrogenase and alanine transaminase (Bishop et al. 1994, Meng et al. 2013).

Under a constant hyperosmotic condition, the next stage is the accumulation of proline and glycine within the cell with proline increasing quickly over 6 days and glycine increasing gradually over two weeks (Baginski and Pierce 1978). Proline is synthesized through the glutamate metabolism pathway (Szabados and Savoure 2010) via Δ-1-pyrroline-5-carboxylate (Δ-1-P5C) and catalyzed by Δ-P5C synthase and P5C reductase (Meng et al. 2013). The increase in glycine is likely through a reversible process of biosynthesis from serine by serine hydroxymethyl transferase (Meng et al. 2013); however, other studies have suggested glycine may also be taken up from the environment (Anderson and Bedford 1973, Stephens and Virkar 1966).

In the final stage, the concentration of taurine increases over two months while the concentrations of proline and glycine decrease. Taurine is a non-protein amino acid and its sources are synthesis from cysteine already within the cell (Meng et al. 2013, Wellborn and Manahan 1995, Jacobsen and Smith 1968) and the uptake of taurine from the extracellular environment into the cell by a high-affinity transport system (Toyohara et al. 2005, Hosoi et al. 2007). Taurine synthesis begins with the oxidation of cysteine by cysteine dioxygenase to cysteine sulfinic acid. Cysteine sulfinic acid is decarboxylated by cysteine sulfinic acid decarboxylase to hypotaurine (Meng et al. 2013, Wellborn and Manahan 1995, Jacobsen and Smith 1968). The process by which hypotaurine converts to taurine is unknown. The uptake of taurine from the extracellular environment into the cell occurs through taurine transporter, a

well-documented transporter protein (Toyohara et al. 2005, Hosoi et al. 2007, Meng et al. 2013). The slow increase in taurine is likely due to a limited capacity by adults to synthesize taurine (Deaton 2009). *De novo* synthesis is an important source of taurine during larval development, but taurine appears to be a dietary requirement for adults (Wellborn and Manahan 1995, Simpson et al. 1959, Bishop et al. 1983).

During the gradual accumulation of taurine, proline is degraded by proline dehydrogenase and P5C dehydrogenase. Glycine is primarily broken down into ammonia and CO_2 by the glycine cleavage system, which is catalyzed by glycine dehydrogenase (Meng et al. 2013). Alternatively, glycine may be degraded to serine by serine hydroxymethyl (Meng et al. 2013). The shift to taurine as the major component of the osmotic bulk completes the free amino acid portion of the cellular response to hyperosmotic conditions.

In addition to the free amino acids, quaternary amines also increase in concentration under hyperosmotic pressure. The accumulation of glycine betaine has been studied specifically in *C. virginica* (Pierce et al. 1992). Glycine betaine is synthesized from choline in a two-step reaction. Choline oxidase converts choline to an aldehyde, which is then oxidized in a reaction catalyzed by betaine aldehyde dehydrogenase (Pierce et al. 1992). In a comparison of oysters from the Atlantic Ocean and from the Chesapeake Bay acclimated to the same salinity, Atlantic oysters synthesized glycine betaine four-fold faster than Chesapeake oysters, despite similar rates of choline uptake (Pierce et al. 1992). These results suggest regulation differences between the populations (Pierce et al. 1992).

2. Response to Hypoosmotic Conditions

Under hypoosmotic pressure, cells release both organic and inorganic osmolytes to reduce their osmolality and regulate cell volume. Cells initially swell due to influx of water and

then decrease in volume with the release of osmolytes (Deaton 2009). The chronology of events is not as well studied as for hyperosmotic responses and existing studies are contradictory across taxa (Deaton 2009). However, a recent study on *C. gigas* (Meng et al. 2013) found that genes from the Ca^{2+} signaling pathway were up-regulated in response to hypoosmotic conditions, corroborating the signaling model proposed for *Noetia ponderosa* (Amende and Pierce 1980a, Amende and Pierce 1980b, Pierce and Politis 1990). In this model, cell swelling activates a stretch-activated channel selective for calcium. As Ca^{2+} enters the cell, it binds to calmodulin and this complex in turn activates a kinase. The kinase then proceeds to phosphorylate various targets on the plasma membrane, leading to the efflux of amino acids.

The amino acids comprising the osmotic bulk are also likely degraded or their accumulation halted in response to hypoosmotic conditions. Meng et al. (2013) found a decrease in glycine content in response to low salinity. The predominant pathway for this decrease was degradation by the glycine cleavage system. Additionally, *C. gigas* was found to down-regulate the expression of cysteine dioxygenase, cysteine sulfinic acid and taurine transporter in response to decreased salinity, preventing further accumulation of taurine (Meng et al. 2013).

These trends lead to the prediction that osmoregulatory genes should show distinct patterns of expression in response to hyper- vs. hypoosmotic stress. For example, genes such as serine hydroxymethyl transferase, which synthesizes glycine, should be up-regulated in hyperosmotic conditions but down-regulated hypoosmotic conditions. Additionally, any genotypic differences that shape the limits and breadth of these genes' plasticity generate the phenotypic differences in osmoregulation upon which selection may act. These genotype-by-environment interactions can lead to phenotype-environment mismatches and result in intra-generational selection.

My dissertation consists of three manuscripts that examine the effect of intra-generational selection on the functional genetic variation of eastern oysters along an estuarine salinity gradient. The chapters are presented in order of completion. The first chapter investigates the impact of the parental selective history, as defined by the average salinity at the parental source reefs, on larval survival in different salinity treatments. The second chapter is the assembly and annotation of an eastern oyster transcriptome that serves as a reference for testing functional genetic differences among groups of oysters. The third chapter measures gene expression differences in response to salinity treatment between oysters from different reefs in order to test the limits of osmoregulatory plasticity and identify differences in these limits between oysters from different source reefs. The results from these studies demonstrate that plastic gene expression responses are genomically pervasive, while also implicating selection in generating reef-specific patterns of gene expression. Additionally, the larval survival experiment suggests that the functional differences in the plasticity of gene expression between the reefs are heritable, impacting larval survival in different salinity treatments. Together, these results support the inference of recurrent viability selection leading to functional genetic differences between oyster reefs within a single estuary. We suggest that for many species with high fecundity, type III survival and broad dispersal, intra-generational selection molds patterns of plasticity across habitats and needs to be taken into account to understand the diversity of reaction norms and their adaptive value.

REFERENCES

Amende LM, Pierce SK (1980a) Cellular volume regulation in salinity stressed molluscs: the response of *Noetia ponderosa* (Arcidae) red blood cells to osmotic variation. *J. Comp. Physiol.* 138, 283-289.

Amende LM, Pierce SK (1980b) Free amino acid mediated volume regulation of isolated *Noetia ponderosa* red blood cells: control by Ca^{2+} and ATP. *J. Comp. Physiol.* 138, 291-298.

Anderson JW, Bedford WB (1973) The physiological response of the estuarine clam, Rangia cuneata (Gray), to salinity. II. Uptake of glycine. *Biological Bulletin*, 144(2), 229-247.

Baginski RM, Pierce SK (1978) A comparison of amino acid accumulation during high salinity adaptation with anaerobic metabolism in the ribbed mussel, *Modiolus demissus demissus*. *Journal of Experimental Zoology*, 203(3), 419-428.

Bishop SH, Ellis LL, Burcham JM (1983) Amino acid metabolism in mollusks. In: *The Mollusca*, vol. 1: *Metabolic biochemistry and molecular biomechanics* (ed Hochachka PW), pp. 243-327. Academic Press, New York.

Bishop SH, Greenwalt DE, Kapper MA, Paynter KT, Ellis LL (1994) Metabolic regulation of proline, glycine, and alanine accumulation as intracellular osmolytes in ribbed mussel gill tissue. *Journal of Experimental Biology*, 268, 151-161.

Burger R, Lynch M (1995) Evolution and extinction in a changing environment: A quantitative-genetic analysis. *Evolution*, 49(1), 151-163.

Buroker NE (1983) Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. *Marine Biology*, 75, 99-112.

Carroll SP, Hendry AP, Reznick DN, Fox CW (2007) Evolution on ecological time-scales. *Functional Ecology*, 21, 387-393.

Chevin LM, Lande R (2010) When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population? *Evolution*, 64, 1143-1150.

Chevin LM, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *Plos Biology*, 8, e1000357.

Davis HC, Chanley PE (1955). Spawning and egg production of oysters and clams. *Biological Bulletin* (Woods Hole), 110, 117-128.

Deaton LE (2001) Hyperosmotic volume regulation in the gills of the ribbed mussel, *Geukensia demissa*: Rapid accumulation of betaine and alanine. *Journal of Experimental Marine Biology and Ecology*, 260(2), 185-197.

Deaton L (2009) Osmotic and ionic regulation in molluscs. In: *Osmotic and Ionic Regulation: Cells and Animals* (ed Evans DH), pp. 107-133. CRC Press, Boca Raton, FL.

Eierman LE, Hare MP (in press) Transcriptomic analysis of candidate osmoregulatory genes in the eastern oyster *Crassostrea virginica*. *BMC Genomics*.

Hare MP, Avise JC (1996) Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution,* 50, 2305-2315.

He Y, Ford SE, Bushek D, Powell EN, Bao Z, Guo X (2012) Effective population sizes of eastern oyster *Crassostrea virginica* (Gmelin) populations in Delaware Bay, USA. *Journal of Marine Research*, 70, 357-379.

Hellmann JJ, Pineda-Krch M (2007) Constraints and reinforcement on adaptation under climate change: selection of genetically correlated traits. *Biological Conservation*, 137, 599-609.

Hoffman AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, 470, 479-485.

Hosoi M, Kubota S, Toyoharo M, Toyoharo H, Hayashi I (2003) Effect of salinity change on free amino acid content in Pacific oyster. *Fisheries Science*, 69, 395-340.

Hosoi M, Shinzato C, Masaya T *et al.* (2007) Taurine transporter function from the giant Pacific oyster *Crassostrea gigas*: function and expression in response to hyper- and hypo-osmotic stress. *Fisheries Science*, 73, 385-394.

Jacobsen JG, Smith LH (1968) Biochemistry and physiology of taurine and taurine derivatives. *Physiological Reviews*, 48, 424-511.

Johnson KD, Smee DL (2012) Size matters for risk assessment and research allocation in bivalves. *Marine Ecology Progress Series*, 462, 103-110.

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, 256, 100-102.

Kennedy VS (1996) The biology of larvae and spat. In: *The eastern oyster: Crassostrea virginica* (eds Kennedy VS, Newell RI, Eble AF), pp. 371-422. Maryland Sea Grant, College Park, MD.

Kirschner LB (1991) Water and ions. In: *Environmental and Metabolic Animal Physiology* (ed Prosser CL), pp. 13-107. Wiley-Liss, New York.

Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, 22, 1435-1446.

Lord J, Whitlatch R (2012) Inducible defenses in the eastern oyster *Crassostrea virginica* Gmelin in response to the presence of the predatory oyster drill *Urosalpinx cinerea* Say in Long Island Sound. *Marine Biology*, 159, 1177-1182

Lynch M, Lande R (1993) Evolution and extinction in response to environmental change. In: *Biotic Interactions and Global Change* (eds Kareiva PM, Kingsolver JG, Huey RB), pp. 234-250. Sinauer, Sunderland, MA.

Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010) Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters*, 13, 128-140.

McDonald JH, Verrelli BC, Geyer LB (1996) Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. *Molecular Biology and Evolution*, 13, 1114-1118.

Meng J, Zhu Q, Zhang L, Li C *et al* (2013) Genome and transcriptome analyses provide insight into the euryhaline adaptation mechanism of *Crassostrea gigas*. *PLoS ONE*, 8, e58563.

Newell RE, Kennedy V, Shaw K (2007) Comparative vulnerability to predators, and induced defense responses, of eastern oysters *Crassostrea virginica* and non-native *Crassostrea ariakensis* oysters in the Chesapeake Bay. *Marine Biology*, 152, 449-460.

Pierce SK, Politis AD (1990) Ca^{2+}-activated cell volume recovery mechanisms. *Ann. Rev. Physiol.*, 52, 27-42.

Pierce SK, Rowland-Faux LM, O'Brien SM (1992) Different salinity tolerance mechanisms in Atlantic and Chesapeake Bay conspecific oysters: glycine betaine and amino acid pool variations. *Marine Biology*, 113, 107-115.

Raubenheimer D, Simpson SJ, Tait AH (2012) Match and mismatch: conservation physiology, nutritional ecology and the timescales of biological adaptation. *Phil. Trans. R. Soc. B*, 367, 1628-1646.

Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics*, 124, 397-406.

Reed TE, Schindler DE, Waples RS (2011) Interacting effects of phenotypic plasticity and evolution on population persistence in a changing climate. *Conservation Biology*, 25, 56-63.

Reusch TBH (2014) Climate change in the oceans: evolutionary versus phenotypically plastic responses of marine animals and plants. *Evolutionary Applications*, 7, 104-122.

Robinson EM, Lunt J, Marshall CD, Smee DL (2014) Eastern oysters *Crassostrea virginica* deer crab predators by altering their morphology in response to crab cues. *Molecular Ecology*, 20, 111-118.

Scheiner SM (1993) Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics*, 24, 35-68.

Schlichting CD, Wund MA (2014) Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution*, 68, 656-672.

Simpson JW, Allen K, Awapar J (1959) Free amino acids in some aquatic invertebrates. *Biol. Bull. Mar. Biol. Lab., Woods Hole*, 117, 371–381.

Stephens GC, Virkar RA (1966) Uptake of organic material by aquatic invertebrates. IV. The influence of salinity on the uptake of amino acids by the brittle star, Ophiactis arenosa. *Biol. Bull. Mar. Biol. Lab., Woods Hole*, 131, 172-185.

Szabados L, Savoure A (2010) Proline: a multifunctional amino acid. *Trends Plant Sci*, 15, 89-97.

Thomas CD, Cameron A, Green RE, *et al.* (2004) Extinction risk from climate change. *Nature*, 427, 145-148.

Thorson G (1950) Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews*, 25, 1–45.

Toyohora H, Yoshida M, Hosoi M, Hayashi I (2005) Expression of taurine transporter in response to hypo-osmotic stress in the mantle of Mediterranean blue mussel. *Fisheries Science*, 71, 356-360.

Wellborn JR, Manahan DT (1995) Taurine metabolism in larvae of marine mollusks (Bivalvia, Gastropda). *Journal of Experimental Biology*, 198, 1791-1799.

Willi Y, Hoffmann AA (2009) Demographic factors and genetic variation influence population persistence under environmental change. *Journal of Evolutionary Biology*, 29, 124-133.

Williams GC (1975) Sex and Evolution. Princeton University Press, Princeton, N.J.

Zhang L, Li L, Zhu Y, Zhang G, Guo X (2014) Transcriptome analysis reveals a rich gene set related to innate immunity in the eastern oyster (*Crassostrea virginica*). *Marine Biotechnology*, 16, 17-33.

CHAPTER 2

SURVIVAL OF OYSTER LARVAE IN DIFFERENT SALINITIES DEPENDS ON SOURCE POPULATION WITHIN AN ESTUARY[1]

Abstract

The role of environmental heterogeneity in limiting connectivity and shaping population structure continues to be a major question in evolutionary biology, particularly for high-dispersal species. Many marine species have a two part life cycle comprised of a sedentary adult phase and a dispersing larval phase. For estuarine species such as *Crassostrea virginica* (eastern oyster), larvae are often carried through very distinct water masses that can affect growth and survival prior to settlement, potentially impacting population connectivity. On the mesoscale of an estuary, gene flow may be a homogenizing force; however, for genomic regions experiencing strong differential selection along estuarine gradients, gene flow may be minimal if recurrent viability selection maintains functional genetic differentiation. Estuaries are defined by salinity gradients and many taxa rely on phenotypic plasticity to thrive there. Nonetheless, even euryhaline species like eastern oysters have their physiological limits, and this study tests whether survival of *C. virginica* larvae in different salinities depends on parental source reef and/or conditioning salinity. Oysters from high, intermediate and low salinities within Delaware Bay, New Jersey, were spawned in a common garden to test for differences in larval survival that have a genotypic basis. Under the null hypothesis of functional homogeneity among adult oyster populations we expected no difference in larval survival. Broodstock were conditioned in low and high salinity common gardens for 4-6 weeks before spawning. Larvae from 56 pair-cross

[1] This paper has been published with the citation as follows:
Eierman LE, Hare MP (2013) Survival of oyster larvae in different salinities depends on source population within an estuary. *Journal of Experimental Marine Biology and Ecology* 449, 61-68.

families were reared in low and high salinities for 13 days. Cox proportional hazard models were used to determine significant predictors of larval survival. Population source interacted with larval salinity treatments to significantly affect larval survival. This finding suggests that the larval pool of single estuaries contains abundant genetic variation for survival across different salinities, stemming in part from functional genetic differences among source reefs. Our results can help parameterize larval connectivity models that incorporate environment-dependent survivorship.

Keywords: *Crassostrea virginica*, salinity, larval survival, population differentiation, phenotype-environment mismatch

1 Introduction

Understanding the degree of population connectivity, ranging from "closed" systems characterized by persistent genetic differentiation between populations to "open" systems showing broad-scale homogeneity, is vital to fishery management, restoration design and the designation of marine reserves (Cowen et al., 2007). Many marine species, particularly invertebrates, have a bipartite life cycle comprised of a sedentary adult phase and a dispersing larval phase. Planktonic larval stages can persist long enough for organisms to travel hundreds to thousands of kilometers; however, the connectivity of marine populations is often more restricted than predicted by the dispersal capabilities of migrants and the known hydrographic barriers (Koehn et al., 1980; Lewis and Thorpe, 1994). Two plausible and not exclusive explanations are physical barriers, such as isoclines and hydrographic fronts (Pineda et al., 2007), and biological barriers (Gaines et al., 2007; Grosberg and Cunningham, 2001). Physical explanations such as barriers to circulation have successfully predicted patterns of larval transport (Gilg and Hilbish, 2003) but the effects of physical barriers are frequently hard to determine due to interactions with larval behavior (Shanks, 2009). Biological barriers may be particularly important in systems with environmental gradients or patchiness where strong selective pressures during and after dispersal both limit connectivity and shape population genetic variation among breeders. Salinity gradients from fresh to oceanic water define estuaries and provide an excellent system for measuring biological barriers to connectivity.

Biological barriers to connectivity can occur during both larval dispersal and post-settlement. A large percentage of mortality for high fecundity marine species occurs during dispersal (Thorson, 1950). Predation and starvation are spatially unpredictable circumstances for larvae leading to potentially high mortality rates over and above intrinsic factors stemming from

genetic load. In contrast, physiological stress as larvae disperse across abiotic gradients may account for spatially non-random mortality that could shape population differentiation. Apart from dispersal, the 'getting there' part of connectivity, post-settlement survivorship further determines realized connectivity between populations in terms of adult abundance, and only with successful reproduction do immigrants have an evolutionary impact. Phenotypic plasticity is a common adaptation to habitat heterogeneity, but every trait has tolerance thresholds beyond which plasticity is no longer sufficient to acclimate to the environment (reviewed in Auld et al., 2010). These thresholds define habitat use boundaries below the spatial scale of dispersal.

To the extent that habitat heterogeneity occurs at scales below that of dispersal, a proportion of dispersal constitutes 'migrants' across different microhabitats. Immigrants to non-parental microhabitats can experience a phenotype-environment mismatch and low relative fitness (Marshall et al., 2010) resulting in spatially balanced polymorphisms (Sanford and Kelly 2011). Along spatially stable environmental gradients, each generation of migrants will undergo recurrent viability selection resulting in persistent population differentiation among adults when the strength of selection is strong relative to N_em (gene flow as measured by effective population size (N_e) and migration rate (m)) (Alleaume-Benharira et al., 2006; Antonovics, 1968; Barton, 2001; Holt, 2003; Garcia-Ramos and Kirkpatrick, 1997; Kirkpatrick and Baron, 1997). For species with sedentary adults and proximity-dependent mating (e.g. broadcast spawners), the recurrent functional population differentiation among adults can be translated into greater functional diversity among larvae than expected under panmixia. Alternatively, where the strength of selection on a trait is less than N_em, but environmental stress is beyond plasticity thresholds, surviving immigrants can lower mean population fitness, constrain local adaptation and no functional population differentiation would be observable (Hendry and Taylor, 2004;

Nosil and Crespi, 2004; reviewed in Garant et al., 2007). The likelihood of these two outcomes depends on the degree of plasticity for a given trait, the strength of selection on that trait, and the distribution of gene effects underlying the trait (Yeaman and Whitlock, 2011). Thus, depending on the traits under investigation, populations compared across a gradient may exhibit different levels of connectivity and genetic differentiation related to these traits (Caillaud and Via, 2011).

One way to test for balancing selection is to measure population differentiation among adults for loci that are likely to be responding to selection gradients. Finding the relevant loci makes this classical population genetic approach challenging in non-model organisms, but with some luck and rigorous subsequent experiments, dramatic patterns of small scale genetic differentiation have been shown to result from post-settlement selection in several estuarine systems. One example is clinal variation at the *Lap* locus of *Mytilus edilus* (Koehn et al., 1976, 1980; Koehn and Hilbish, 1987). Among adult populations in the Atlantic Ocean and Long Island Sound a *Lap* allele decreased in frequency from 0.55 to 0.15 over a 10 mile distance with decreasing salinity (Koehn et al., 1976). In estuarine cohorts the oceanic allele was common in newly settled juveniles and progressively declined to the characteristic frequency found in local adults, consistent with recurrent post-settlement selection. Similarly, a strong selection gradient across the intertidal zone filters genotypes from the mixed larval pool in *Semibalanus balanoides* and maintains a stable polymorphism (Schmidt et al., 2000). These examples, along with other studies (rainbow smelt: Saint-Laurent et al., 2003; three-spined stickleback: Hendry et al., 2002, McCairns and Bernatchez, 2010), demonstrate the impact of a strong selection gradient on population differentiation in high gene flow systems.

An alternative approach is to experimentally test for genetically-based differences in survival limits for larvae derived from breeding populations experiencing different environments

within a single estuary. If adults from different habitats are functionally differentiated as a result of recurrent selection then, after controlling for maternal effects, they should produce larval cohorts with distinct genotype-by-environment patterns of viability. Previous studies experimentally testing for genotype-by-environment effects on survival and growth of larvae have found phenotype-environment mismatches that suggest better survival and growth in the natal habitat than in other environments (eastern oyster: Newkirk et al., 1977; Newkirk, 1978; European oyster: Newkirk, 1986; hard clam: Knaub and Eversole, 1988; Manzi et al., 1991). In fact, larvae have been shown to have narrower physiological tolerances than adults in several bivalve species (Bayne et al., 1976) facilitating this experimental approach. The strength of this approach is that no *a priori* knowledge of candidate loci or markers for population differentiation is needed. Additionally, differentiation is identified directly at the phenotypic level after accounting for plasticity and maternal effects, explicitly demonstrating the extent of phenotype-environment mismatch at the dispersal stage.

The goal of this study was to identify functional differentiation in C*rassostrea virginica* (eastern oyster) adults along a salinity gradient within a single estuary by experimentally measuring the impact of source location and broodstock conditioning salinity on larval progeny survivorship at low and high salinity treatments. In western North Atlantic estuaries the eastern oysters are ecosystem engineers (Lenihan and Peterson, 1998) whose complex reef systems provide habitat for over 300 species (Beck et al., 2011). Due to its diverse ecosystem services (reviewed in Constanza et al., 1997), the oyster is considered a keystone estuarine species (Barnes et al., 2007; Coen et al.,1999). With historic loss of 90% of eastern oyster reefs in North America (Jackson et al., 2001; Kirby, 2004), restoration of oyster populations is needed to realize these ecosystem services again. Many states are engaged in efforts to restore oysters (Beck et al.,

2011), often through reef construction and planting of hatchery-produced oysters. It is this restoration objective that motivates a more rigorous examination of larval tolerances and the mechanisms that determine them.

A potentially valuable methodological advance in restoration planning is to couple hydrodynamic models with larval particle tracking and habitat heterogeneity to project the efficacy of different management and restoration procedures. The Oyster Restoration Optimization model (North et al., 2010) designed for the Chesapeake Bay and a model of oyster larval dispersal in the Delaware Bay (Narváez et al., 2012) are two such models. The integration of larval swimming behavior and environment-dependent mortality potentially increases the accuracy of source and sink relationships inferred from the models. By identifying sources and sinks, organizations can focus on the appropriate sites for their particular restoration goals such as constructing reefs at sink locations or enhancing stock at source locations. Currently, environment-dependent larval mortality is modeled based on species-specific thresholds. If functional genetic differentiation occurs among breeding oyster populations within single estuaries, and especially given that dispersal is predicted to be strongly asymmetric and downstream based on these models (North et al., 2010), then implementing population specific thresholds may improve the ability of models to accurately predict the realized connectivity resulting from differential larval and post-settlement survival.

2 Materials and Methods

2.1 Sample Collection

Two hundred adult oysters were collected from each of three sites with distinct salinity regimes within the Delaware Bay on April 18, 2011 (Fig. 2.1). Oysters from Cape Shore (39° 04.10' N, 74° 54.77' W; salinity range 20-25; Narváez et al., 2012) were hand collected from

intertidal reefs. Oysters from Arnolds reef (39° 23.055' N, 75° 27.002' W; salinity range 6.5-14.5; Bushek et al., 2012) and New Beds reef (39° 14.518' N, 75° 15.071' W; salinity range 9-16.5; Bushek et al. 2012) were collected by dredge from the NJ Fish and Wildlife vessel *Zephyrus*.

2.2 Adult Oyster Conditioning

The main objectives of adult oyster conditioning were to minimize the impact of maternal effects on larval survival and to have adults undergo gametogenesis under two different salinities (10 and 30). Half of the oysters were conditioned in recirculating tanks at Haskin Shellfish Research Laboratory (HSRL) of Rutgers University while the other half was conditioned in the field. For the tank-conditioned oysters at the hatchery, fifty de-fouled oysters from each of the three populations were placed in a tank of UV-irradiated 1mm filtered seawater (salinity 30) and fifty from each were in a separate tank with the seawater diluted to a salinity of 10 with distilled freshwater. Total tank volume was 500L. Temperature for the first three weeks of tank conditioning was 18°C. Water heaters were then used to slowly increase the temperature to 22°C and maintained there until spawning. The broodstocks were fed a 2:2:1 mixture of *Pavlova lutheri*, *Chaetoceros muelleri* and *Tetraselmis chui* twice a day. For a slow release of algae during feeding, a bucket of the mixture was siphoned via an airline into the tank. Broodstocks were conditioned in tanks for 14 weeks.

Previous experiences in an unpublished pilot study indicated that tank conditioning can be challenging with oysters collected from low-salinity wild stocks. Therefore, in this study half of the oysters were conditioned at field sites in the Delaware Bay. Fifty oysters from each source population were outplanted on racks in Cape May harbor (38° 56.73' N, 74° 53.98' W) for the high salinity conditioning (25-30; Narváez et al., 2012) and the remaining fifty oysters from each

population were outplanted in bags off of a dock in the Cohansey River (39° 22.75' N, 75° 21.32' W) for the low salinity conditioning (~5; Narváez et al., 2012). Because of the geographic distance between the two field sites, the locations may have differed in other biologically important ways such food availability. The oysters remained outplanted in the field for 10 weeks.

2.3 Oyster Strip Spawning

When field and tank broodstock became ripe, gametes were stripped from the oysters using standard hatchery methods (e.g. Allen and Bushek, 1992). Field oyster condition was monitored by microscopically examining subsamples of oyster gonads and when they appeared to be ripe, moving all oysters to a 10 salinity tank for Cohansey-conditioned oysters and a 30 salinity tank for Cape May-conditioned oysters on the day before the spawn. Females were considered ripe when their follicles were filled with large, round oocytes and males were considered ripe when the follicles were densely packed with moving spermatozoa. All ripe males and females were used for spawning. Eggs were passed through an 80µm sieve and retained on a 20µm sieve in order to clean the eggs. The eggs were then resuspended in 200ml filtered sea water at the conditioning salinity of the female. Sperm were passed through a 20µm sieve to remove gonad tissue. Sperm from a single male was then slowly added to the eggs of a single female to produce a single family. Sperm-egg mixtures were examined on a Sedgewick-Rafter slide to ensure that sperm density was approximately 7-10 sperm per egg (determined by counting the average number of sperm surrounding an egg) and sperm was added until this density was reached (similar to Eudeline et al., 2000). After two hours fertilization was confirmed based on observation of polar bodies.

2.4 Larval Cultures

Two hours after fertilization, embryos were added to 1000ml beakers of water at a density of 10 embryos/ml, one beaker per family per experimental treatment salinity. Embryos were maintained at the parental conditioning salinity until twenty four hours after fertilization. Then, to reduce osmoregulatory shock, the water was changed and salinities were adjusted to a midway salinity between conditioning and experimental treatment salinities. At 48 hours, 25 D-stage larvae from each beaker were transferred to separate small glass dishes with 50mL of water at experimental salinities of 10 and 30. Initiation of the treatment was counted as Day 1 and survival data were recorded every other day when all surviving larvae were pipetted into a new watchglass with clean water. Larvae were fed T-*Isochrysis galbana* initially and a 1:1 ratio with *Pavlova lutheri* starting Day 7 of the experiment. Feeding was daily and the quantity increased as the larvae grew. Larvae were kept in a temperature controlled room (temp =25°C) and eyed larvae began to develop on Day 11. Final counts were taken on Day 13.

2.5 Data Analysis

Data from tank and field-conditioned oysters were analyzed separately. Each combination of population (H , I or L) and conditioning salinities (H or L) was replicated by having multiple families. The twenty-five larvae from each family replicated the time to mortality for a family (each population and conditioning combination) at a specific salinity treatment (H or L). We attempted to make ten pair-cross families for each source population (H, I and L) at each conditioning salinity (H or L). Eyed larvae first appeared on day 11 and many families went extinct before day 13. Day 11 was therefore used as the time point for calculating mean number of surviving larvae and standard deviation as well as for comparisons between model-predicted reaction norms.

A Cox proportional hazards regression model (survival model) was used to compare survival of larvae over the course of the 13 day experiment based on population source (P), conditioning salinity (C) and treatment salinity (T). Coefficient subscripts distinguished their association to a particular factor and the subscript *i* indicated the individual larvae. The model, where $h_i(t)$ is the instantaneous risk of demise for individual larvae (*i*) at time *t*, leaves the baseline hazard distribution unspecified:

$$\log h_i(t) = \alpha(t) + \beta_j * P_i + \beta_k * C_i + \beta_l * T_i$$

Larvae from tank-conditioned and field-conditioned oysters were analyzed separately. In the survival model, time dependent variables are incorporated through a counting process that accounts for the time of mortality for each larva in the experiment or for survival until the end of the experiment (Table 2.1) (Andersen and Gill, 1982). The model was implemented with the *survival* package (Therneau, 2011) in R (R Development Core Team, 2011). The best models were identified using Akaike information criterion (AIC) to evaluate relative model fit in relation to the number of parameters and the model with the lowest AIC was selected.

3 Results

3.1 Oyster Conditioning and Spawning

The field-conditioned oysters developed mature gametes earlier than the tank-conditioned oysters as determined visually from gonad and gamete characteristics. Therefore, the spawning and larval culture of field-conditioned families preceded that of tank-conditioned oysters by four weeks. The high salinity population conditioned at the high salinity field site was spawned first, followed by the low and intermediate populations two weeks later and finally all three populations at the low salinity field site three weeks after the high salinity spawn. Additionally,

for tank-conditioned broodstock, gonad maturation and spawning was one week earlier for the high salinity population than for the low and intermediate populations.

Larval families were only successfully produced from a subset of the source populations. Oysters conditioned in tanks yielded multiple families from both high and intermediate salinity source populations (Table 2.1), but eggs did not fertilize in most pair crosses from the low salinity population. The gonads and gametes for low salinity source oysters appeared fully developed relative to oysters from intermediate and high salinity, but the lack of successful fertilization suggests the eggs were not fully mature. Oysters conditioned in the field yielded multiple families from both intermediate and low salinity source populations (Table 2.1). For oysters from the high salinity source, ten out of ten spawned pairs conditioned at the high salinity location resulted in successful fertilization but the embryos did not develop to the D-stage for unknown reasons. This was the earliest population in which spawning was attempted, so it is possible that eggs were not fully mature. High-salinity source oysters conditioned at the low salinity location had high mortality while outplanted and the surviving oysters did not develop gonads.

3.2 Survival Model

Analysis for the Cox proportional hazards models began on Day 5 when the first mortality event occurred. The survival curves predicted by the best models are presented in Figure 2.2 and reaction norms of the number of surviving larvae at Day 11, predicted from the models, are presented in Figure 2.3. For the tank-conditioned oysters from high and intermediate salinity, significant predictors of survivorship in the best regression model included population source (P), conditioning salinity (C), treatment salinity (T), and pairwise interactions of these factors, including the PxT interaction ($p=0.012$) (Table 2.2). The largest model coefficient was

population source (P: $p<0.0001$) and the strongest contrast in survivorship curves showed higher overall survivorship in the high salinity source families relative to intermediate source families regardless of conditioning and larval culture treatments (Fig. 2.2A & 2.2B).

For the field-conditioned oysters from low and intermediate salinity populations, statistically significant predictors of survival in the best model included C, T, C x T and P x C x T terms ($p<0.0001$ for each; Table 2.3). The low salinity source population larvae reared in the low salinity 'home' treatment had the greatest overall survival and lowest among-family coefficient of variation for survival (CV = 0.244) of any field experimental group (Fig. 2.2C & 2.2D, Fig. 2.3B, Table 2.1). In this field conditioning experiment the best survivorship model showed the location/salinity conditioning factor having the largest model coefficient ($p<0.0001$, Table 2.3). Both source populations showed a relatively large change in larval survivorship in response to conditioning treatments, best illustrated by the predicted reaction norms at Day 11 (Fig. 2.3B). Predicted reaction norms from the field conditioning experiment illustrate the significant treatment (T) effect as steep reaction norm slopes (Fig. 2.3B).

To determine what factors were driving the higher-order interaction term of PxCxT for the field-conditioned oysters, the two conditioning locations were analyzed separately for the four possible models: P, T, P+T and PxT. For the high salinity conditioning, the best model was P + T where the P term was not significant but the T term was highly significant ($p<0.001$) (Fig. 2.2C). For the low salinity conditioning, the best model was PxT where the main effect P term was not significant but both T ($p<0.0001$) and PxT ($p<0.0001$) were highly significant (Fig. 2.2D and dashed lines of Fig. 2.3B).

4 Discussion

Some recent studies have suggested that marine populations are not as homogenized as larval planktonic duration and potential dispersal distance would predict (Sanford and Kelly, 2011; Schmidt et al., 2008). Until the major mechanisms limiting gene flow are identified, it will remain difficult to generalize guidelines for spatially explicit management and restoration plans. One potentially important biological barrier to connectivity is phenotype-environment mismatches during dispersal or at settlement when habitat heterogeneity exceeds the tolerances and plasticity of individuals. For euryhaline species adapted to tidally variable estuaries, plasticity is the assumed primary mechanism by which individuals cope with both temporal and spatial variation in salinity. Phenotypic plasticity incurs an energetic cost, so species are expected to experience environmental margins where plasticity is stressful and beyond which environmental variation may be lethal, depending on genotype. For any particular species and environmental gradient there is presumably a zone of marginal habitat where differential viability selection becomes relatively important, relative to phenotypic plasticity, for population persistence. To the extent this is true, and mating is local, offspring from parents in marginal environments may have genotypes that are quite distinct from the species' norm. The demographic and evolutionary consequences of these marginal populations depend on their extent and patterns of connectivity. High fecundity and broad scale dispersal are life history traits that may jointly increase the likelihood that differential viability selection has spatially broad effects. Not only will broad dispersal make phenotype mismatches common, but the large effective population size associated with these life history traits will increase the efficacy of selection relative to drift so that more moderate habitat heterogeneities have consequences in terms of a selective cost.

This conceptual model requires quantitative theoretical development and empirical systems conducive to hypothesis testing. One fundamental prediction is that spatially proximate populations exchanging many migrants across a steep environmental gradient will show functional genetic differentiation. Previous studies have demonstrated that recurrent post-settlement viability selection produces spatially balanced polymorphisms at one or a few loci responding to fine-scale estuarine habitat gradients (Day, 1990; Koehn et al., 1976, 1980; Johannesson et al., 1995; Schmidt et al. 2000). Transplant experiments have also been used to demonstrate local adaptation at various scales for marine species with larval dispersal (reviewed in Sanford and Kelly, 2011).

Here we took an alternative approach to testing for phenotype-environment mismatches that could generate a biological barrier to dispersal. We experimentally tested whether limits to salinity tolerance differentially affect survivorship for larvae derived from a 'common garden' of local populations that had settled and survived in different salinity regimes. The approach taken here focuses on the complex physiological phenotype of salinity tolerance, crudely in terms of survivorship, and measures the fitness impact of phenotype-environment mismatches within a single estuary where larvae are well mixed (Milbury et al., 2008).

Under the null hypothesis of phenotypic plasticity, reaction norms should have no slope and there should be no PxT interaction effects. Alternatively, if broodstock populations are genetically adapted to the salinity regime of their home reef, then their larvae should survive better at that 'home' salinity relative to larvae from other broodstock source locations (population [P] effect), and better relative to cultures from the same population reared at non-natal salinities (treatment [T] effect) for a combined PxT effect. This effect was clear in the survival model for larvae from the more experimentally-controlled tank-conditioned oysters collected at high and

intermediate salinities within Delaware Bay. For tank-conditioning, the PxT term had a significant effect on larval survivorship ($p=0.012$), although it was not the most significant predictor (Population: $p<0.0001$). The significance of the population main effect was driven by overall low survival in the intermediate population (the population that was not environmentally matched by either 'high' or 'low' conditioning or treatments).

For the field-conditioned oysters, conditioning location/salinity had a large effect, perhaps not surprisingly given the potential for confounding with other environmental factors such as temperature and primary productivity that may have co-varied with salinity between the conditioning locations. A PxT effect was apparent in the larvae from broodstock conditioned up-bay at low salinity (Fig. 2.2D), and was likely subsumed in the overall field experiment model within the higher-order interaction of PxCxT ($p<0.0001$). When the two conditioning locations were analyzed separately a highly significant PxT effect ($p<0.0001$) was indeed found for larvae from low and intermediate source broodstock conditioned up-bay at low salinity (Fig. 2.2D). With low-salinity field conditioning, larvae from the low salinity population had as much as four-fold better survival in low salinity versus high salinity cultures, whereas intermediate-source larvae showed only a two-fold difference (Fig. 2.2D).

Intermediate source broodstock were not conditioned nor were their larvae cultured under 'home' conditions in either experiment so predictions were ambiguous. Nonetheless, all else being equal we expected larvae from the intermediate broodstock to respond to salinity treatments similarly in the tank and field experiments. This was true with the reaction norm slope produced by the intermediate source larvae after low-salinity conditioning, suggesting comparability of results across the experiments, but not for high-salinity conditioning (compare Fig. 2.3A and B). If the two experiments are comparable they imply that low salinity broodstock

populations are more genetically differentiated with respect to alleles influencing osmoregulatory tolerance and produce larvae with narrower phenotypic plasticity than intermediate or high salinity populations under most conditions.

Across these analyses a population by treatment effect is evident both statistically and graphically, either alone or in interaction with conditioning location/salinity. An unfortunate constellation of experimental factors eliminated our ability to make many of the direct comparisons sought, complicating interpretation of the results. However, the experimentally cleanest (tank) experiment generated results consistent with a significant PxT interaction with respect to high and intermediate salinity source populations. Interestingly, the prediction of home-environment advantage, testable in this experiment only with the high population source larvae tested under 'home' and 'away' conditions, was only seen with conditioning at high (home) salinity (Fig. 2.2A). Home-environment advantage was also found for larvae from low-salinity source broodstock (Fig. 2.2C & D) with a steep reaction norm no matter what conditioning location/salinity was experience by broodstock (Fig. 2.3B). Overall these results provide tentative support for a model in which selection across the salinity gradient in Delaware Bay was strong enough to generate functional genetic differences among low, intermediate and high salinity adults such that they produced larvae with different survival probabilities at different salinities.

An intriguing pattern that emerges from these results is that the low salinity population has much greater survival in low salinity than in high salinity treatments whereas the high salinity population shows more similar survival rates across salinity treatments. Some biophysical models, such as the Oyster Restoration Model (North et al. 2010), suggest a predominantly downstream movement of larvae from low salinity regions to high salinity regions

of the estuary. Our results suggest a greater potential for genetic differentiation in the upstream reaches of the estuary and this is consistent with limited up-estuary dispersal. Furthermore, the more plastic phenotypic response (lower reaction norm slope) observed in families from high salinity regions was potentially due to downstream transport generating recruitment from a more diverse larval pool. Further research should investigate the degree of asymmetric gene flow within estuaries and its consequences for functional genetic differentiation.

To draw inferences about genotypic differentiation, we minimized maternal effects on larval survival with the one exception of conditioning salinity, a variable that in principle could be manipulated in the hatchery if there were strong justification to do so. To minimize general maternal effects two approaches were taken. In the more controlled experiment using tank-conditioned broodstock to generate larvae, broodstock were collected early in gametogenesis and maintained in common garden tanks where temperature, water volume and water change frequency were uniform and only salinity differed. Oysters in each tank were also fed equal densities of algae relative to the mass of oysters in the tank. The second experiment using field-conditioned broodstock was an attempt to test larvae based on a common garden broodstock design, while using a more natural 'garden.' However, environmental conditions other than salinity may have differed between the two conditioning locations in Delaware Bay. Production of and experimentation with F2 progeny from the original broodstock is a more thorough method of controlling maternal effects, but captive propagation of oysters typically entails reductions in genotypic diversity (Boudry et al. 2002) and invites inadvertent artificial selection (Christie et al. 2012). Many maternal effects are expected to wane during larval development, particularly after metamorphosis to a feeding veliger. Newkirk et al. (1977) reported that significant maternal effects on the survival of *C. virginica* larvae ended after day 6. Thus, in this study general

maternal effects were further minimized by starting the larval survival experiment 48 hours after fertilization and measuring larval survival out to 13 days.

Our experimental focus on larvae allows us to directly relate results to the dispersing phase of oysters and concomitant selection in the plankton, and as such will help parameterize dispersal and recruitment models. The relevance of these results to post-settlement selection is less clear, especially given that salinity tolerances are somewhat narrower in oyster larvae than in adults (Kennedy, 1996). Nonetheless, given other examples of strong post-settlement selection (e.g., Koehn et al., 1976, 1980; Schmidt et al., 2008) we can expect that functional genetic differentiation among adults from different salinity regimes was produced by a combination of pre- and post-settlement selection.

The genetic patterns demonstrated here lead to several recommendations for restoration practice and modeling. For hatchery-based restoration methods, survival of outplanted juveniles may be improved by collecting broodstock from the region of the estuary where the outplanting will occur, or from an environmentally similar region within the estuary. Additionally, larval survival in the hatchery can be maximized by conditioning broodstock at a salinity that falls within the source location range of variation. Environmental matching between broodstock source location and outplant site can maximize post-outplant survival but does not necessarily improve the success of subsequent larvae. Success of subsequent larval cohorts will depend on their dispersal patterns relative to salinity gradients, among other factors. For modeling, realized dispersal may be more accurately estimated if larval survivorship is parameterized as a function of parental environment. Because broadcast spawning enforces local mating at the scale of individual reefs (Levitan et al., 1991), our results imply that the larval pool is not just a product of generalist parents, but includes contributions from assortative mating among physiological

specialist genotypes along the habitat margins. Depending on the demographic extent of contributions, this interpretation of the larval pool suggests a dramatically different source-sink dynamic for oyster recruitment than would be presumed for a homogeneous habitat.

Our study has demonstrated greater larval survival at salinities that more closely match the parental source salinity, consistent with pre-settlement selection contributing to functional genetic differentiation of osmoregulatory genes in adults spanning the estuarine salinity gradient. In order to quantify the combined pre- and post-settlement effects of selection on functional connectivity, future research should compare estimates of neutral marker gene flow to that realized in functional genes under selective pressure across habitat heterogeneities.

Acknowledgements

We thank J. Scarpa and D. Bushek for providing access to hatchery facilities and information on oyster aquaculture, B.J. Landau for assistance in algal culture, and H. Kates and K. Kirk for research assistance. We would also like to thank P.J. Sullivan and W. Fetzer for statistical consultation.

Role of the funding source

Funding for the research was provided by BioOce 0830547, Sigma Xi Grant-in-Aid of Research, Department of Natural Resources and Cornell funds, and a Cornell Travel Grant. These sources had no role in the study design; in data collection, analysis, and interpretation; in the writing of the report; or in the decision to submit the paper for publication.

Table 2.1. Number of independent pair-cross families and day 11 surviving larvae per family in the larval survival experiment. P is the population source and C is the conditioning salinity. The % mor. is the percentage of mortality experienced by adult oysters during conditioning, not including random oysters killed to inspect gonad and gamete condition. The n_1 value is the number of surviving day 11 families. Mean (mean number of surviving larvae) and sd (standard deviation) are for the families at Day 11. The n_2 value is the sample size (number of individual larvae) for the Cox proportional hazards regression model. The * indicates a family not used in the analysis.

				Treatment							
				High				**Low**			
Location	**P**	**C**	**% mor.**	**n_1**	**mean**	**sd**	**n_2**	**n_1**	**mean**	**sd**	**n_2**
Tank	High	High	3.7	7	14.00	7.66	175	7	10.29	5.74	175
		Low	4.0	10	12.20	6.09	250	10	16.1	5.70	250
	Intermediate	High	5.1	6	1.83	2.71	150	6	0.50	0.84	150
		Low	4.6	9	7.22	3.99	225	9	4.44	5.27	225
	Low	High	32.0	0	--	--	--	0	--	--	--
		Low	8.0	1*	--	--	--	1*	--	--	--
Field	High	High	8.5	0	--	--	--	0	--	--	--
		Low	50.0	0	--	--	--	0	--	--	--
	Intermediate	High	2.0	4	9.75	7.76	100	4	16.75	6.18	100
		Low	0.0	6	0.33	0.52	150	6	5.67	3.27	150
	Low	High	8.2	5	7.80	3.11	125	5	19.80	3.11	125
		Low	5.5	10	1.78	3.35	250	10	15.78	5.21	250

Table 2.2. Cox proportional hazards regression results for tank-conditioned experiment involving broodstock from high and intermediate salinity populations (P), conditioned at low and high salinities (C), with survival estimated for larvae at low and high treatment salinities (T). The Δ AIC is the difference in Akaike's information criterion between the best-performing model (top row) and the model being compared. Significant model coefficients (α=0.05, df=6) are indicated in bold italics. Negative values increase survival and positive values decrease survival relative to that of larvae from the intermediate salinity population conditioned at high salinity and reared in the high salinity treatment.

	Model Coefficient							
Model	P	C	T	P*C	P*T	C*T	P*C*T	Δ AIC
P*C+P*T+C*T	***-1.32***	***-0.64***	***0.32***	***0.74***	***-0.28***	***-0.37***		0
P*C*T	***-1.46***	***-0.75***	0.19	***0.97***	0.00	-0.15	-0.47	7.62
C+P*T	***-0.91***	***-0.46***	0.09		***-0.23***			51.21
C+P+T	***-1.02***	***-0.46***	-0.02					53.76
P*C	***-1.45***	***-0.82***		***0.73***				11.51
P*T	***-0.80***		0.08		***-0.26***			112.2
C*T		0.10	0.15			***-0.37***		349.5
P	***-0.93***							114.4
C		***-0.28***						358.3
T			-0.08					380.1

Table 2.3. Cox proportional hazards regression results for tank-conditioned experiment involving broodstock from low and intermediate salinity populations (P), conditioned at low and high salinities (C), with survival estimated for larvae at low and high treatment salinities (T). The Δ AIC is the difference in Akaike's information criterion between the best-performing model and the model being compared. Significant model coefficients (α=0.05, df=7) are indicated in bold italics. Negative values increase survival time and positive values decrease survival time relative to that of larvae from the low salinity population conditioned at high salinity and reared in the high salinity treatment.

	Model Coefficient							
Model	P	C	T	P*C	P*T	C*T	P*C*T	Δ AIC
P*C*T	1.22	***3.27***	***0.36***	0.94	1.02	***0.29***	***2.89***	0
P*C+P*T+C*T	0.93	***2.73***	***0.26***	***1.46***	***2.00***	***0.47***		13.17
C+P*T	1.16	***2.43***	***0.15***		***2.19***			41.88
C+P+T	***1.58***	***2.38***	***0.22***					75.53
P*C	1.22	***1.36***		***1.61***				525.1
P*T	1.11		***0.20***		***1.94***			209.6
C*T		***3.11***	***0.37***			***0.44***		87.9
P	***1.53***							599.6
C		***1.60***						590.1
T			***0.27***					263.8

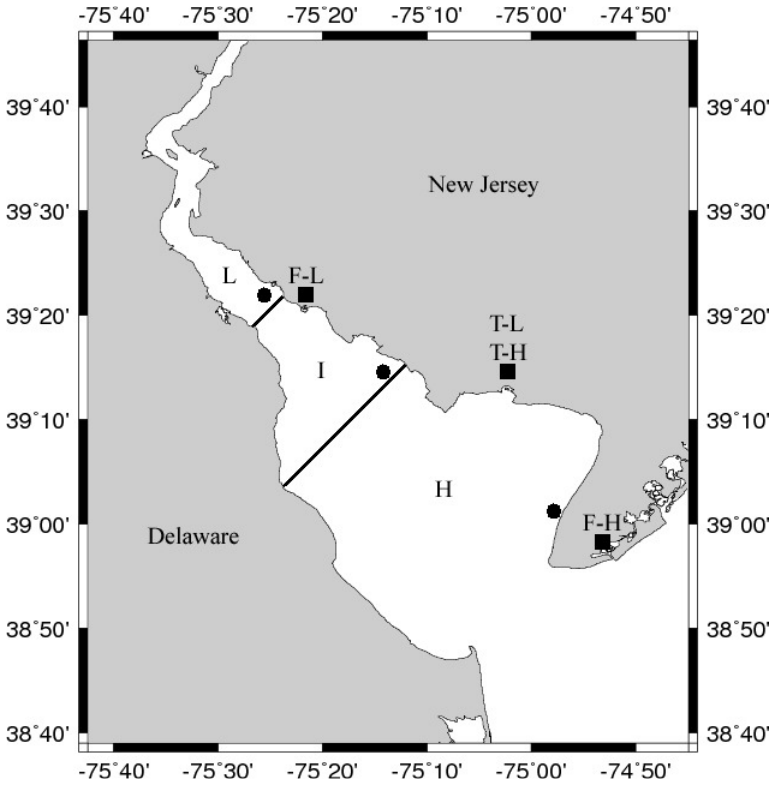

Figure 2.1. The Delaware Bay is divided into three salinity regimes: L -low salinity (6.5-14.5), I - intermediate salinity (9-16.5), H - high salinity (20-25). Circles represent 3 oyster broodstock collection sites. Squares are conditioning locations with T representing hatchery conditioning in tanks with low (L, 10) or high (H, 30) salinity and F representing field conditioning in open water field sites with low (L, ~5) and high (H, ~25-30) salinity.

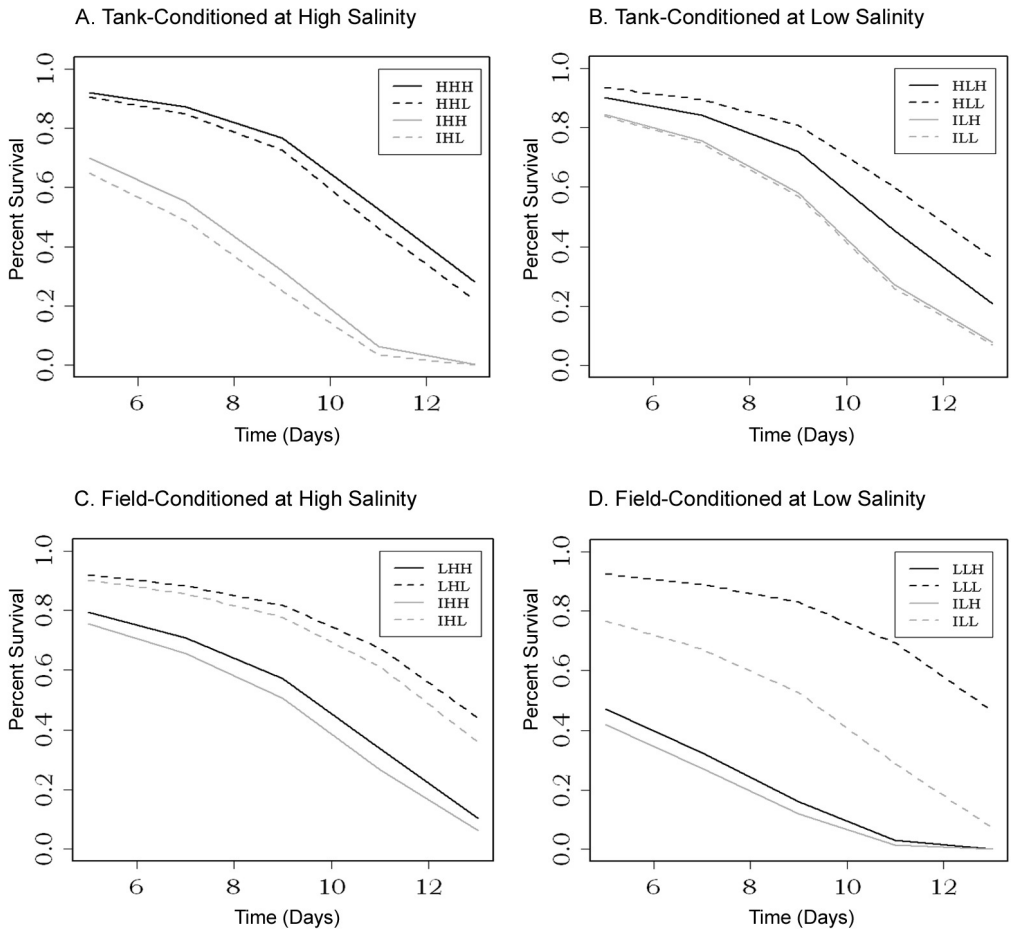

Figure 2.2. Survival curves for tank-conditioned (A and B) and field-conditioned (C and D) oyster larvae over time predicted from two independent Cox proportional hazards regression models. The three letter legend abbreviations indicate the population source with the first letter, conditioning salinity with the second and treatment salinity with the third. Dashed lines are low (L) salinity treatments and solid lines are high (H) salinity treatments. Black and grey lines indicate population source with gray always indicating intermediate (I) population source, whereas black lines represent results from high source oysters with tank-conditioning in graphs A and B and represent Low source oysters with field-conditioning in graphs C and D.

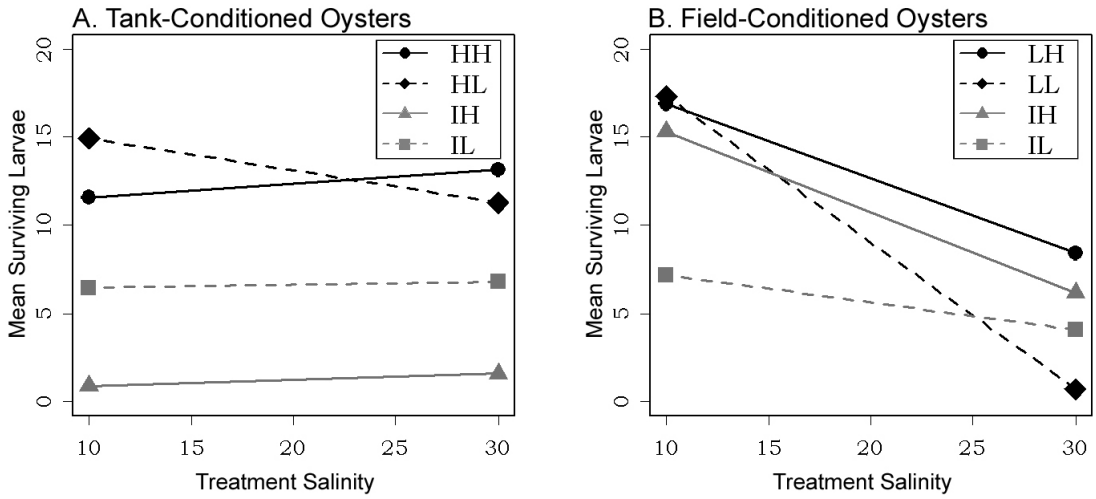

Figure 2.3. Reaction norms of predicted larval survival from the Cox proportional hazards models for each experimental group: (A) Tank-conditioned oysters and (B) Field-conditioned oysters. The two letter legend abbreviations indicate population source (low, intermediate or high salinity) with the first letter and the conditioning salinity (low ~ 10, high ~ 30) with the second letter. For tank-conditioned broodstock (A) the only larvae experiencing 'home' conditions were those from the high salinity source at treatment salinity of 30 in which case mean survival was slightly better than at a salinity of 10, but only when conditioned at high salinity (HH). Mean survival was better at salinity of 10 when conditioned at low salinity (HL). With field conditioning (B) the only larvae experiencing ‘home’ conditions were those from low salinity source oysters. The L population larvae survived better at 10 than 30, regardless of conditioning.

REFERENCES

Alleaume-Benharira, M., Pen, I.R., Ronce, O., 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J. Evol. Biol. 19, 203–215.

Allen Jr., S.K., Bushek, D., 1992. Large scale production of triploid *Crassostrea virginica* (Gmelin) using "stripped" gametes. Aquaculture 103, 241-251.

Andersen, P.K., Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. Annals of Statistics 10, 1100-1120.

Antonovics, J., 1968. Evolution in closely adjacent plant populations. VI. Manifold effects of gene flow. Heredity 23, 507-524.

Auld, J. R., Agrawal, A.A., Relyea, R.A., 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proc. R. Soc. B 277,503–511.

Barnes, T.K., Volety, A.K., Chartier, K., Mazzotti, F.J., Pearlstine, L., 2007. A habitat suitability index model for the eastern oyster (*Crassostrea virginica*), a tool for restoration of the Caloosahatchee Estuary, Florida. J. Shellfish Res 26(4), 949-959.

Barton, N., 2001. Adaptation at the edge of a species' range, in: Silvertown, J., Antonovics, J. (Eds.), Integrating Ecology and Evolution in a Spatial Context. Blackwell, London, pp. 365–92.

Bayne, B.L., Thompson, R.J., Widdows, J., 1976. Physiology I, in: Bayne, B.L. (Ed.), Marine Mussels: Their Ecology and Physiology. Cambridge Scientific Press, UK, pp. 121–206.

Beck, M.W., Brumbaugh, R.D., Airoldi, L., Carranza, A., Coen, L.D., Crawford, C., Defeo, O., Edgar, G.J., Hancock, B., Kay, M.C., Lenihan, H.S., Luckenbach, M.W., Toropova, C.L., Zhang, G., Guo, X., 2011. Oyster reefs at risk and recommendations for conservation, restoration, and management. BioScience 61(2), 107-116.

Boudry, P., Collet, B., Cornette, F., Hervouet, V., Bonhomme, F., 2002. High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. Aquaculture 204, 283–296.

Bushek, D., Ford, S.E., Burt, I., 2012. Long-term patterns of an estuarine pathogen along a salinity gradient. Journal of Marine Research 70, 225-251.

Caillaud, M. C., Via, S., 2012. Quantitative genetics of feeding behavior in two ecological races of the pea aphid, *Acyrthosiphon pisum*. Heredity 108, 211–218.

Christie, M.R., Marine, M.L., French, R.A., Blouin, M.S., 2012. Genetic adaptation to captivity can occur in a single generation. Proc Natl Acad Sci USA 109, 238–242.

Coen, L.D., Luckenbach, M.W., Breitburg, D.L., 1999. The role of oyster reefs as essential fish habitat: a review of current knowledge and some new perspectives, in: Benaka, L.R. (Ed.), Fish habitat: essential fish habitat and rehabilitation. American Fisheries Society, Symposium 22, Bethesda, MD, pp. 438-454.

Constanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., Van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260.

Cowen, R.K., Gawarkiewicz, G., Pineda, J., Thorrold, S.R., Werner, F.E., 2007. Population connectivity in marine systems. Oceanography 20, 14–21.

Day, A.J., 1990. Microgeographic variation in allozyme frequencies in relation to the degree of wave action in the dogwhelk *Nucella lapillus* (L.) (Prosobranchia: Muricacea). Biol. J. Linn. Soc. 40, 245–261.

Eudeline, B., Allen Jr., S.K., Guo, X., 2000. Optimization of tetraploid induction in Pacific oysters, *Crassostrea gigas*, using first polar body as a natural indicator. Aquaculture 187, 73-84.

Gaines, S.D., Gaylord, B., Gerber, L.R., Hastings, A., Kinlan, B.P., 2007. Connecting places: the ecological consequences of dispersal in the sea. Oceanography 20, 90–99.

Garant, D., Forde, S.E., Hendry, A.P., 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. Functional Ecology 2, 434-443.

García-Ramos, G., Kirkpatrick, M., 1997. Genetic models of adaptation and gene flow in peripheral populations. Evolution 51, 21-28.

Gilg, M.R., Hilbish, T.J., 2003. Patterns of larval dispersal and their effect on the maintenance of a blue mussel hybrid zone in Southwest England. Evolution 57, 1061-1077.

Grosberg, R. K., Cunningham, C.W., 2001. Genetic structure in the sea: from populations to communities, in: Bertness, M.D., Gaines, S., Hay, M.E. (Eds.), Marine Community Ecology. Sinauer Associates, Sunderland, MA, pp 61-84.

Hendry, A.P., Taylor, E.B., McPhail, J.D., 2002. Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. Evolution 56, 1199-1216.

Hendry, A.P., Taylor, E.B., 2004. How much of the variation in adaptive divergence can be explained by gene flow: an evaluation using lake/stream stickleback. Evolution 58, 2319–2331.

Holt, R.D., 2003. On the evolutionary ecology of species' ranges. Evol. Ecol. Res. 5, 159–178.

Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange,

C.B., Lenihan, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J., Warner, R.R., 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293,629–638.

Johannesson, K., Rolan-Alvarez, E., Ekendahl, A., 1995. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. Evolution 49, 1180-1190.

Kennedy, V.S., 1996. The biology of larvae and spat, in: Kennedy, V.S., Newell, R.I., Eble, A.F. (Eds.), The eastern oyster: Crassostrea virginica. Maryland Sea Grant, College Park, MD, pp. 371-422.

Kirby, M.X., 2004. Fishing down the coast: historical expansion and collapse of oyster fisheries along the continental margins. Proc Natl Acad Sci USA 101, 13096–13099.

Kirkpatrick, M., Barton, N.H., 1997. Evolution of a species' range. Am. Nat. 150, 1–2.

Knaub, R.S., Eversole, A.G., 1988. Reproduction of different stocks of *Mercenaria mercenaria.* J. Shellfish Res. 7, 371-376.

Koehn, R.K., Hilbish, T.J., 1987. The adaptive importance of genetic variation. American Scientist 75, 134-141.

Koehn, R.K., Milkman, R., Mitton, J.B., 1976. Population genetics of marine pelecypods. IV. Selection, migration, and genetic differentiation in the blue mussel *Mytilus edilus*. Evolution 30, 2-30.

Koehn, R.K., Newell, R.I.E., Immermann, F., 1980. Maintenance of an aminopeptidase allele frequency cline by natural selection. Proc. Natl. Acad. Sci. USA 77(9), 5385-5389.

Lenihan, H.S., Peterson, C.H., 1998. How habitat degradation through fishery disturbance enhances effects of hypoxia on oyster reefs. Ecological Applications 8, 128-140.

Levitan, D.R., Sewell, M.A., Chia, F., 1991. Kinetics of fertilization in the sea urchin *Strongylocentrotus franciscanus*: Interaction of gamete dilution, age, and contact time. Biol. Bull. 181, 371-378.

Lewis, R.I., Thorpe, J.P., 1994. Temporal stability of gene frequencies within genetically heterogeneous populations of the queen scallop *Aequipecten* (Chlamys) *opercularis*. Mar. Biol. 121, 117–126.

Manzi, J.J., Hadley, N.H., Dillon Jr., R.T., 1991. Hard clam, *Mercenaria mercenaria* broodstocks: growth of selected hatchery stocks and their reciprocal crosses. Aquaculture 94, 17-26.

Marshall, D.J., Monro, K., Bode, M., Keough, M.J., Swearer, S., 2010. Phenotype-environment mismatches reduce connectivity in the sea. Ecology Letters 13(1), 128-140.

McCairns, R.J.S., Bernatchez, L., 2010. Adaptive divergence between freshwater and marine sticklebacks: Insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. Evolution 64(4), 1029-1047.

Milbury, C., Guo, X., Bushek, D., Ford, S.E., 2008. Spatial population structure in Delaware Bay oysters. J. Shellfish Res. 27(4), 1033. (abstract)

Narváez, D.A., Klink, J.M., Powell, E.N., Hofmann, E.E., Wilkin, J., Haidvogel, D.B., 2012. Modeling the dispersal of eastern oyster (*Crassostrea virginica*) larvae in Delaware Bay. Journal of Marine Research 70, 381-409.

Newkirk, G.F., 1978. Interaction of genotype and salinity in larvae of the oyster *Crassostrea virginica*. Marine Biology 48, 227-234.

Newkirk, G.F., 1986. Controlled mating of the European oyster, *Ostrea edulis*. Aquaculture 57, 111-116.

Newkirk, G.F., Waugh, D.L., Haley, F.E., 1977. Genetics of larval tolerance to reduced salinities in two populations of oysters, *Crassostrea virginica*. J Fish Res Bd Canada 34, 384-387.

North, E.W., King, D.M., Xu, J., Hood, R.R., Newell, R.I.E., Paynter, K.T., Kellogg, M.L., Liddel, M.K., Boesch, D.F., 2010. Linking optimization and ecological models in a decision support tool for oyster restoration and management. Ecological Applications 20, 851–866.

Nosil, P., Crespi, B.J., 2004. Does gene flow constrain trait divergence or vice-versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. Evolution 58, 101-112.

Pineda, J., Hare, J.A., Sponaugle, S., 2007. Larval transport and dispersal in the coastal ocean and consequences for population connectivity. Oceanography 20, 22–39.

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Saint-Laurent, R., Legault, M., Bernatchez, L., 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordex* Mitchill). Molecular Ecology 12, 315-330.

Sanford, E., Kelly, M.W., 2011. Local adaptation in marine invertebrates. Annu. Rev. Mar. Sci. 3, 509–535.

Schmidt, P.S., Bertness, M.D., Rand, D.M., 2000. Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. Proc. R. Soc. Lond. B 267, 379-384.

Schmidt, P.S., Serrão, E.A., Pearson, G.A., Riginos, C., Rawson, P.D., Hilbish, T.J., Brawley, S.H., Trussell, G.C., Carrington, E., Wethey, D.S., Grahame, J.W., Bonhomme, F., Rand, D.M., 2008. Ecological genetics in the North Atlantic environmental gradients and adaptation at specific loci. Ecology 89(11) Supplement, S91-S107.

Shanks, A.L., 2009. Pelagic larval duration and dispersal distance revisited. Biological Bulletin 216, 373–385.

Therneau, T., and original Splus->R port by Thomas Lumley (2011). survival: Survival analysis, including penalised likelihood. R package version 2.36-9. http://CRAN.R-project.org/package=survival

Thorson, G. 1950. Reproductive and larval ecology of marine bottom invertebrates. Biological Reviews 25, 1–45.

Yeaman, S., Whitlock, M.C. 2011. The genetic architecture of adaptation under migration selection balance. Evolution 65, 1897-1911.

CHAPTER 3

TRANSCRIPTOMIC ANALYSIS OF CANDIDATE OSMOREGULATORY GENES IN THE EASTERN OYSTER *CRASSOSTREA VIRGINICA*[2]

Abstract

Background

The eastern oyster, *Crassostrea virginica*, is a euryhaline species that can thrive across a wide range of salinities (5-35). As with all estuarine species, individual oysters must be able to regulate their osmotic balance in response to constant temporal variation in salinity. At the population level, recurrent viability selection may be an additional mechanism shaping adaptive osmoregulatory phenotypes at the margins of oyster salinity tolerance. To identify candidate genes for osmoregulation, we sequenced, assembled and annotated the transcriptome of wild juvenile eastern oysters from 'high' and 'low' salinity regimes and used normalized libraries and 454 sequencing technology. Annotations and candidates were mostly based on the Pacific oyster (*Crassostrea gigas*) genome sequence so osmoregulatory relevance in *C. virginica* was explored by testing functional enrichment of genes showing spatially discrete patterns of expression and by quantifying coding sequence divergence.

Results

The assembly of sequence reads from the two populations resulted in 157,022 reftigs (contigs and singletons) that collapsed to 98,729 reftigs using permissive clustering parameters in order to combine oversplit alleles. Annotations were applied to 50,736 reftigs and one fifth of these belonged to a set of candidate osmoregulatory genes identified from the *C. gigas* genome. A total

[2] This paper is accepted pending slight revision at *BMC Genomics* with the authors LE Eierman & MP Hare

of 218,777 SNPs were identified in annotated reftigs of *C. virginica*, corresponding to an average SNP density of 0.0185 per bp. Amino acid divergence between translations of *C. virginica* annotated reftigs and *C. gigas* coding sequence averaged 23.2 % with an average dN/dS ratio of 0.074, suggesting purifying selection on protein sequences. Consistent with known molecular mechanisms for osmotic regulation of cell volume in molluscs, the set of annotated genes found to be uniquely expressed in the low salinity population was significantly enriched for 'integral to membrane' and 'intrinsic to membrane' cellular component gene ontologies while the high salinity population was enriched for 'extracellular region'. Similarly, the low salinity population was significantly enriched for molecular functions related to the hydrolysis of peptides and the phosphorylation of plasma membrane proteins while the high salinity population was enriched for gated and ion-gated channel activity.

Conclusions

Most of the osmoregulatory gene candidates experimentally identified in *C. gigas* are present in this *C. virginica* transcriptome. In general these congeners show coding sequence divergence too high to make the *C. gigas* genome a useful reference for *C. virginica* bioinformatics. On the other hand, strong purifying selection is characteristic of the osmoregulatory candidates so functional annotations are likely to correspond. An initial examination of *C. virginica* expression patterns across the salinity gradient in a single estuary suggests that many of these candidates have expression patterns that co-vary with salinity, consistent with osmoregulatory function in *C. virginica*.

Keywords: *Crassostrea virginica*, osmoregulation, transcriptome, dN/dS, gene enrichment, SNP, Cd-hit

Background

The eastern oyster (*Crassostrea virginica*) builds reefs that support productive estuarine communities and provide important ecosystem services [1, 2]. However, overfishing, disease pressure and environmental stress have led to the loss of approximately 90% of biomass across the eastern oyster's home range since the early 1900's [3, 4, 5]. Two important topics in oyster biology and restoration are the mechanisms by which oysters respond to stress [e.g. 6, 7] and the ability of oyster populations to either acclimate to stress through phenotypic plasticity or adapt via selection. The majority of work on eastern oysters has focused on immune response to pathogens [egg. 8, 9, 10, 11] with a few observational studies on other environmental stressors [e.g. 6, 12]. Spatial and temporal variation in salinity is a given for estuaries, and phenotypic buffering of cell volume through osmolyte control is an essential adaptation for all organisms that live there.

Eastern oysters are found along salinity gradients ranging from near freshwater conditions (salinity of 5) to oceanic salinities (salinity of 35) [13, 14, 15].Their greatest abundance is typically at intermediate salinities, with the adult physiological optimum posited to be as narrow as salinities of 15 - 18 [15]. At the margins of this environmental envelope, recent results suggest that post-settlement viability selection is one important process for sustaining adult populations [16]. While the genes involved in osmoregulation have not been well characterized in the eastern oyster, recent studies on the Pacific oyster (*Crassostrea gigas*) [17, 18, 19] provide valuable tools for investigating the genetics of osmoregulation. Generating genome-scale resources such as transcriptome sequences for *C. virginica* can facilitate studies of gene expression and the physiology of osmoregulation in order to better understand responses to osmotic stress at the individual and population levels.

Oysters regulate cell volume in response to changing salinity through multiple mechanisms. Oysters are osmoconformers with no ability to osmoregulate their extracellular fluid [reviewed in 20]. Salinity fluctuations therefore result in energetically costly processes to maintain isoosmostic balance by accumulating or releasing osmotically active solutes (osmolytes) [20]. These osmolytes include both inorganic ions such as N^+, K^+ and Ca^{2+} and organic substances such as free amino acids (FAA) and quaternary amines [reviewed in 21]. Oysters, like many organisms when under great osmotic stress, primarily use organic osmolytes such as taurine, alanine, aspartic acid, glycine and betaine [21, 22, 23]. Organic osmolytes are able to provide osmotic bulk under high osmotic stress without the direct physiological trade-offs that inorganic ions would have [24]. Furthermore, organic osmolytes can stabilize proteins and protect cells from oxidative stress [20, 24]. A variety of functional classes of enzymes are likely involved in osmoregulation, including peptidases to catalyze the hydrolysis of peptides into amino acids, kinases to phosphorylate plasma membrane proteins and transporters to move molecules across cell membranes [20].

Most molecular physiological studies of osmoregulation in oysters have focused on the products of single genes [e.g. taurine transporter: 25, 26]. The availability of the *C. gigas* genome sequence and the initial evaluation of gene expression between salinity treatments [17] demonstrated differential expression for hundreds of genes. Genomic studies of *C. virginica* gene expression across natural salinity gradients have also shown many genes responding to this environmental gradient [6]. To enable more focused future studies on osmoregulation in *C. virginica*, a first step is the identification of candidate genes involved in this core physiological process.

Our objective was to identify genes putatively involved in osmoregulation in the eastern oyster by sequencing, assembling, and annotating the transcriptome from low- and high-salinity source populations of juvenile oysters by using 454 sequencing technology. Using annotations and differential expression data from *C. gigas,* we identified *C. virginica* transcripts that are candidates for osmoregulatory function. Given that these congeners shared a common ancestor more than 82 Mya [27], we explored the functional appropriateness of these annotations in two ways. We quantified the distribution of coding sequence divergence and estimated the strength of purifying selection maintaining similar polypeptide sequences in the two species. Second, we tested for predicted expression patterns in normalized cDNA libraries from low- and high-salinity wild oysters. Specifically, we predicted that transcripts found in one salinity population but not the other would be enriched for candidate osmoregulatory genes and for osmoregulation-related gene ontology terms (GO; www.geneontology.org). Our evaluation of this transcriptome and results of these associated analyses provide some confidence that these candidate genes are a comprehensive starting point for experiments investigating the physiological and evolutionary responses of eastern oysters to osmoregulatory challenges in their estuarine environment.

Materials and Methods

I. Sample Collection and Archiving

Shell substrate was deployed at a "high" salinity field site (27°10'58.2"N 80°12'22.2"W; mean salinity = 15.9, max=33.5, min=4.6) and at a "low" salinity field site (27°13'11.2"N 80°13'38.9"W; mean salinity = 8.0, max=18.2, min=1.0) in the St Lucie River, Florida, on June 2, 2010. Water temperature, salinity, and percent dissolved oxygen were recorded every hour at both sites from March 23, 2010 until July 1, 2010 with a Sonde (YSI 600OMS V2). Over this time interval these two sites were significantly different in salinity (Fig. 3.1, t=-38.6, df=1397,

p<0.001). Mean water temperature was 26.5°C at the high salinity site and 27.4°C at the low salinity site. Mean dissolved oxygen was 83.1% at the high salinity site and 86.6% at the low salinity site. The temperature and dissolved oxygen did not differ significantly between the sites (p=0.062, p=0.091). Juvenile oysters (spat; 4 – 10 mm total length) were collected from the shell substrate on July 1, 2010. All soft tissue, including gill, mantle and adductor muscle, was archived for each individual in RNALater® (Ambion) after removing the visible digestive system. Within two weeks, the RNALater® was drained and the samples were archived at -80°C.

II. RNA Extraction

Approximately 30µg tissue from each of four individuals per site was used for individual RNA extractions using Qiagen RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. Total RNA from each sample was quantified using NanoDrop (ThermoScientific) and 5µg from each of four individuals per site was pooled.

III. 454 Library Prep and Sequencing

The construction of two normalized cDNA libraries and 454 pyrosequencing was carried out at the W.M. Keck Center for Comparative and Functional Genomics, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. We chose to normalize the libraries in order to increase the likelihood that rare transcripts would be sequenced, leading to a more complete transcriptome with limited sequencing effort. For each library, messenger RNA was isolated from 10µg of pooled total RNA with the Oligotex kit (Qiagen, Valencia, CA). The messenger RNA was then converted to a primary cDNA library with adaptors compatible with the 454 system using Multiplex Identifier (MID) tags to distinguish the two population pools [28]. The libraries were diluted to $1x10^6$ molecules/µL, pooled, and sequenced on a full plate using the 454 Genome Sequencer FLX+ system according to the manufacturer's instructions

(454 Life Sciences, Branford, CT). Signal processing and base calling were performed using the bundled 454 Data Analysis Software v2.6.

IV. Transcriptome Assembly and Clustering

The two barcoded sets of reads were independently trimmed prior to assembly. Reads were trimmed from each end using a phred-scale quality score of 20 with fastq_quality_trimmer (FASTXToolkit). When the trimmer encountered a base pair with a quality less than 20, the closest read end was trimmed up to that base. Reads with less than 70% of the original length remaining were discarded. Trimmed reads were then imported into Newbler (gsAssembler, 454 Life Sciences, Roche Diagnostics). Any remaining adapters were trimmed and reads were filtered against an *E. coli* database to remove contaminants. Reads were then assembled *de novo* using the default settings except for a minimum overlap length of 30bp (default 40bp). The quality of the initial assembly was evaluated by comparing assembly statistics to other published molluscan transcriptomes from 454 sequencing. Newbler reports consensus "contigs" using the overlap-layout-census (OLC) approach, which merges reads into contigs when their alignments overlap. Reads with no alignment overlap with other reads are denoted as singletons. Because singletons may belong to unique genes that were not highly expressed, they were included in further analysis. We defined a reference transcriptome as the combined set of 200bp or longer contigs (consensus from multiple overlapping reads) and singletons and hereafter refer to these as "reftigs" (reference transcriptome sequences).

Large indels, highly polymorphic sequences and other *de novo* assembly challenges can often lead to redundancy in sequences between singletons and contigs [29]. Particularly in highly polymorphic species such as oysters, alleles will often assemble into separate contigs or remain as singletons. This redundancy complicates downstream applications of the resulting

transcriptome, e.g. for gene expression analyses [30], because reads from some single copy genes will not map uniquely within the transcriptome and will be discarded. We assumed that the reftigs from the Newbler assembly included many oversplit loci, so to improve the transcriptome we consolidated redundant reftigs by clustering. Cd-hit [31] was used to cluster reftigs with various sequence identity thresholds ranging from permissive clustering at 80% to conservative clustering at 99% using a k-mer word size of 5 to 10 increasing incrementally with the threshold (e.g. word size 5 with threshold 80%).

Several approaches were used to evaluate whether clustering improved the transcriptome. For the 80% and 95% clustering results we compared statistics bearing on transcriptome quality including the percent of reftigs that were annotated, the distribution of annotation between contigs and singletons, and the number of osmoregulatory candidates (identified in *C. gigas*) recovered.

Additionally, we evaluated the two clustering results by comparing the proportion of Illumina reads from a barcoded individual that uniquely mapped to annotated reftigs based on a pilot RNA-seq experiment. The barcoded individual was from an oyster reef in Delaware Bay with a salinity regime ranging from 6.5 to 14.5. The mRNA from 30mg of gill tissue was extracted using Dynabeads® mRNA DIRECT™ kit (Life Technologies). The library was prepared with NEBNext® mRNA Library Pep Reagent Set for Illumina® (New England BioLabs Inc.). The library constituted 16.25% of a single 100 bp Hi-Seq Illumina lane and was sequenced at the Biotechnology Resource Center Genomics Facility of Cornell University. The resulting reads were trimmed following the same procedure as the 454 reads and any remaining adapters were clipped using fastx_clipper (FASTXToolkit). The remaining reads were then

mapped to the annotated reftigs using BWA [32] with a mismatch edit distance of 0.005 and SAMtools [33] with only uniquely mapped reads retained.

V. Annotation

To annotate the *de novo C. virginica* transcriptome assembly, reftigs were compared to NCBI's non-redundant (nr) protein sequence database that included the annotated proteins deduced from the *C. gigas* genome (May 2013), plus the Swiss-Prot and TrEMBL databases from the Uniprot protein knowledge base, using the BLASTx algorithm with an e-value cut-off of 10^{-5}. Gene Ontology (GO; www.geneontology.org) annotation was retrieved from Uniprot. The annotated and unannotated reftigs were then compared with respect to proportion of contigs and singletons as well as GC content in order to explore if unannotated reftigs may represent non-oyster contamination in the 454 sequences. The number of unique genes represented by the transcriptome was then identified by grouping reftigs that shared the same GenBank gene identifier.

We considered genes as osmoregulatory candidates if they were included in the 1241 annotated genes found to be differentially expressed in *C. gigas* adults in response to six different salinity treatments when compared to a control salinity of 30 [Table S21 in 17]. Additionally, we quantified the number of genes in the normalized libraries that were uniquely represented in one *C. virginica* population sample or the other by mapping the trimmed and filtered 454 read pools from ‘low’ and ‘high’ salinity samples back to the annotated reftigs from the 80%-clustered transcriptome using GSMapper (454 Life Sciences, Roche Diagnostics) with default settings. Enrichment of functional classes was tested at the level of genes, based on reftig annotation results described below, for two subsets compared to the entire annotated transcriptome: (1) all osmoregulatory candidates and (2) genes unique to each population.

Enrichment tests used a Fisher's exact test as implemented in TopGO from Bioconductor [34]. Genes that were unique to one of the two populations were identified as "asymmetric." Results from enrichment tests were depicted in the context of the hierarchical structure of gene ontology terms in order to visualize the degree of functional integration among the most significantly enriched genes.

VI. Sequence Comparisons with *C. gigas*

Simple sequence repeats and low complexity regions of the annotated reftigs were masked with RepeatMasker [35], using the rmblast search engine. Reftigs with masked regions were removed from analysis. The coding sequence reading frame for each remaining reftig was then predicted using ESTscan [36]. ESTscan was trained using the EMBL, RefSeq and UniGene clusters from the mollusk *Aplysia californica,* the most closely related species for which a full set of references were available at the time of this study. The matrices from this training were then used to predict coding sequences for the reftigs using a hidden Markov model [36]. The predicted coding sequence for each reftig was then used to analyze sequence divergence from *C. gigas*.

A local directory of *C. gigas* coding sequence for predicted proteins from the *C. gigas* genome was downloaded from http://gigadb.org/dataset/view/id/100030/sort/size and clustered with Cd-hit using the same parameters as for reftig clustering (sequence identity threshold=0.8, word size=5). Coding sequences that clustered were assumed to be paralogs and removed from analysis to reduce the bias that would occur with comparison of paralogs between *C. gigas* and *C. virginica*. The *C. virginica* reftigs were then compared against the *C. gigas* coding sequences using tBLASTx with intron linking disabled and an e-value cutoff of 10^{-5}. Best hits were interpreted as putative ortholog pairs for analysis. Ortholog pairs were then run through a custom

pipeline to align sequences using ClustalW [37] and calculate dN/dS ratios using the codeml function of paml [38]. The distribution of dN/dS values relative to ClustalW alignment length was evaluated before choosing to remove alignments less than 60% of the total reftig length.

VII. SNP Discovery

The mapped 454 reads from both population samples were combined and aligned against the masked, annotated reftigs with mpileup of SAMtools [33]. SNPs were then identified using SNAPE-pooled [39]with a base quality average of 37 or greater, theta of 0.01, divergence of 0.1, flat prior and folded spectrum, and the SNP density for each contig was calculated.

Results and Discussion

I. Assembly and Clustering Results

A total of 1,256,652 raw 454 reads included 718,009 from the high salinity population and 538,553 from the low salinity population. The raw reads are available through the National Center for Biotechnology Information Short Read Archive under accession numbers SRR1029667 for the high salinity population and SRR1029668 for the low salinity population. After trimming and filtering, 1,182,107 reads remained and were assembled into 28,939 contigs that contained 86.7% of the reads. The 128,083 unassembled reads were designated as singletons and included in further analysis. The combined contig and singleton set consisted of 157,022 reftigs. The assembly size for contigs alone was approximately 18,202,631 nucleotide bases, similar to other molluscan transcriptome assemblies based on 454 sequences (Table 3.1), and had an average contig length of 629.1 bases (N50 = 500 bases) and maximum contig length of 7,512 bases. The total transcriptome (contigs and singletons, 157,022 reftigs) was 51,918,466 nucleotides with an average length of 453.0 bases (N50 = 381 bases).

Given the high polymorphism of oysters [17], clustering of assembled reftigs was explored as a method of consolidating alleles that remained apart after assembly. Consolidating alleles is an important consideration before using a transcriptome assembly as a reference for RNAseq expression analyses because oversplit alleles in the assembly will decrease the number of reads that map uniquely. A total of 136,000 reftigs (86.7%) were longer than 200bp and used for clustering at different sequence identity thresholds. Comparing the change in total reftig number resulting from increasingly permissive clustering, the rate of reftig consolidation was initially rapid based on thresholds from 99% to 95%; then the rate of change slowed and was nearly constant between 95% and 80% (Fig. 3.2). As the sequence identity threshold decreased, the ratio of contigs to singletons increased as expected if singletons were being clustered with greater frequency than contigs (Table 3.2).

Transcriptomes resulting from both the 95% and 80% sequence identity thresholds were annotated for comparison. Both transcriptomes had a similar percentage of reftigs successfully annotated, with a similar distribution of contigs and singletons. Likewise, 1014 osmoregulatory candidates (see below) were obtained with the 95% threshold and this dropped by only seven candidate genes at the 80% threshold (Table 3.2). The large percentage (99.4%) of candidate genes that remain in the transcriptome at the 80% threshold compared to the 95% threshold suggests that any potential paralog clustering resulted in a minimal loss in the number of uniquely annotated genes, particularly osmoregulatory candidates.

We mapped 100 bp Illumina RNAseq reads from a single individual to the 95% and 80% transcriptomes to test whether oversplit alleles were consolidated by clustering. Relative to the 95% transcriptome, the 80% transcriptome had a higher percentage of annotated reftigs with mapped reads but the effect was small (Table 3.2). In terms of the proportion of Illumina reads

that mapped, two percent more reads mapped uniquely to the 80% transcriptome than the 95% transcriptome. The increase in the percent of uniquely mapped reads in the 80% transcriptome suggests that consolidation of allelic reftigs was achieved by clustering, resulting in more reads mapping uniquely. Based on these results the 80% threshold transcriptome was chosen for further analysis.

In one cluster examined in more detail, an original contig annotated as Heat Shock 70 kDA Protein 12 was ultimately clustered with three singletons. Two singletons clustered at the 95% threshold. The third singleton (330bp) was unannotated at the 95% threshold. With a similarity of 84.55% estimated by Cd-hit, it was clustered with the contig and the other two singletons at the 80% threshold. An alignment between this third singleton and the original contig showed five indels ranging in size from 1 to 17bp and two polymorphisms as the cause for the 84.55% sequence identity. We suspect these indels represent 454 sequencing error because they were partially shared across the three singletons, most of them would disrupt the reading frame, and they occurred within simple nucleotide repeats and low complexity sequence. Some of the SNPs present in the singletons may also be sequencing error but not obviously so: most were not adjacent to indels and they were already represented in the original contig. Therefore, clustering provided two distinct transcriptome improvements; (1) oversplit alleles were consolidated, facilitating downstream mapping to the transcriptome for RNAseq expression analysis, and (2) more of the 454 sequencing coverage was used to call SNPs.

The optimum balance between consolidating oversplit alleles and clustering paralogs or sequence errors is impossible to know because it depends on the distribution of allelic sequence differences relative to paralog differences in any particular species as well as the sequencing error rate. The comparative approach used here was ad hoc and took advantage of computational

efficiencies when clustering consensus sequences from an assembly rather than exploring parameter values in separate assemblies. When no reference genome is available this comparative empirical approach can be a valuable method of improving transcriptome quality.

II. Annotation Results

The BLASTx search against multiple databases provided annotation for 50,736 reftigs (51.4%) representing 20,249 unique GenBank accessions and 16,392 distinct putative proteins that we will refer to as genes. Only 0.05% of the annotations were achieved with a database other than GenBank nr (Fig. 3.3). Reftigs that did not have a BLASTx match with an e-value smaller than 10^{-5} from any database was designated as unannotated.

Of 16,392 distinct genes, 8,161 are represented by a single reftig. The number of reftigs per gene ranges from 1 to 470 with only 13 genes represented by 100 or more reftigs. Gene duplication and large gene families, particularly in the *C. gigas* genome from which 89% of our annotations were identified, are the primary reasons for the large number of reftigs per "gene". For example, the 456 reftigs identified as the gene "tripartite motif-containing protein 2" from *C. gigas* were annotated from 201 unique GenBank accessions. For *C. gigas*, these different GenBank accessions represent different coding sequence locations within the genome. Therefore, we define a "gene" here as a protein product, which often represents large gene families.

Gene Ontology (GO) terms were assigned to 36,924 of the annotated reftigs, representing 11,583 putative genes, based on sequence similarity to known proteins in the UniProt databases. Annotated reftig sequences have been archived and are accessible through FigShare (http://dx.doi.org/10.6084/m9.figshare.873865) [40].

A consequence of whole animal extractions and normalizing the libraries was the increased potential to sequence non-oyster transcripts, such as bacteria and algae. Singletons

made up ~74% of the reftigs in the full (annotated and unannotated) 80% transcriptome and less than half of the singletons were successfully annotated. Other studies using 454 sequencing also have described singletons as comprising a large proportion of their transcriptome (e.g. 81.7% [41]; 58.5% [29]; 55.3% [42]) with a subset getting annotated. Singletons are the inevitable consequence of assembling transcripts with low coverage, so they are not necessarily indicative of contamination. However, here the mean GC content for annotated reftigs was 44% (SD=5.4), very similar to the 45.2% (SD=4.3) average for *C. gigas* coding sequences (Fig. 3.4). In contrast, the unannotated portion of the *C. virginica* transcriptome had a significantly different mean GC content of 34.5% (SD=5.8%) (Fig. 3.4; $t=266$, $df=973455$, $p<0.001$). The difference in GC content provides very strong evidence that many of the unannotated reftigs (both contigs and singletons) came from other organisms such as prokaryotes or protozoa. Without the benefit of the *C. gigas* reference genome for annotation and GC content comparison, *de novo* analysis of the eastern oyster transcriptome generated here would be highly compromised by contaminants.

III. Osmoregulation Candidate Genes

Of the 1241 osmoregulatory candidate genes identified in *C. gigas* [17], 1007 (81.2%) were identified in the *C. virginica* transcriptome based on 9307 reftigs (18.3% of all annotated reftigs) (Supplementary Table 3.1). The *C. gigas* candidates were identified experimentally based on differentially expressed genes between different salinity treatments of adult oysters [17], while we obtained these transcripts in wild juveniles. Thus, life stage is one factor that could help account for the *C. gigas* candidates that were not obtained in our samples. Additionally, *C. virginica* and *C. gigas* have mostly overlapping but slightly different salinity tolerances with *C. virginica* having lower mortality rates than *C. gigas* at low salinity and vice versa at higher salinities [43].

Fifty-nine osmoregulatory candidate genes were identified from only the high salinity population and 56 were identified from only low salinity, together representing 11.4% of the candidate genes. These asymmetrically expressed candidate genes were mostly cases with 1x coverage (1:0 asymmetry), but 3 genes (2.6%) had an asymmetry ratio of 5:0 or greater.

Ignoring candidate status, 4,053 of the 16,392 annotated *C. virginica* genes (24.7%) were identified from only one of the two populations. Of these, 2,185 were found only in the high-salinity population, including 1,431 genes (8.7%) with 1:0 asymmetry and 74 genes (0.5%) with 5:0 or greater asymmetry. An additional 1,868 genes were found only in the low-salinity population, including 1,355 (8.3%) with a 1:0 asymmetry and 31 genes (0.2%) with 5:0 or greater asymmetry. A total of 105 genes (2.6%) from the two populations had 5:0 or greater asymmetry.

The fact that 24.7% of all genes showed asymmetry, while only 11% of osmoregulatory candidates did so, suggests that there may be many biological processes leading to population-specific expression in addition to the stochasticity expected with low-expression genes. Also, given that buffering against osmotic stress is a chronic physiological need for oysters, lower asymmetry among osmoregulatory candidates might reflect the proportion of genes within this functional category that have constant but variable expression across salinities.

It is difficult to know how much asymmetry to expect by chance for genes with a given level of expression in normalized libraries. However, enrichment of functional categories within the set of population-specific genes is not expected from stochastic variation in library normalization or read coverage. Our prediction was that among asymmetric genes, annotations related to osmoregulatory function should be the most highly enriched relative to the frequency of functional ontologies in the overall annotated transcriptome. We initially built a frame of

reference by testing for functional enrichment among the entire set of osmoregulatory candidate genes in *C. virginica* and found 12 cellular component GO terms and 86 molecular function GO terms significantly enriched compared to the complete annotated gene set (Fig. 3.5 and 3.6, Supplementary Tables 3.2 and 3.3). For cellular components, the 'extracellular region', 'plasma membrane' and 'membrane' components were among the significantly enriched terms (Fig. 3.5 and Supplementary Table 3.2). At the level of molecular function, 'catalytic' activities, 'binding' functions, 'electron carrier' activities, 'transporter' activities and 'molecular transducer' activities were among the significantly enriched terms (Fig. 3.6, Supplementary Table 3.3). These enriched GO terms serve to functionally characterize the osmoregulatory candidates on the whole and therefore might be indicators of osmoregulatory function in additional enrichment tests when found in concert.

As predicted, the overall group of asymmetric genes (24.7% of all genes) showed significantly enriched ontologies relating to osmoregulatory function, as indicated by similarities with enriched GO terms in the total set of osmoregulatory candidate genes. Interestingly, the enriched GO terms were only partially overlapping in the low versus high salinity population and relative enrichment magnitudes shifted among GO terms. In the low salinity population at the level of cellular components, the strongest result among 12 significantly enriched ontologies included 'integral to membrane' (GO:0016021, p= 0.00037) and 'intrinsic to membrane' (GO:0031224, p= 0.00058), two ontologies nested within 'membrane' components, the level-three GO term enriched among osmoregulatory candidates (Supplementary Table 3.4). Transmembrane channels are important in maintaining cell volume in response to hypoosmotic stress. For example, Ca^{2+} channels are upregulated in hypoosmotic stress and osmolytes such as taurine are taken up through high affinity transport systems that may involve transmembrane

proteins [18]. Additional terms such as ‘cell periphery’ (GO:0071944, p=0.0033), ‘plasma membrane part’ (GO:0044459, p=0.00624) and ‘plasma membrane’ (GO:0005886 , p=0.00815) were terms significantly enriched both in the asymmetric low salinity genes and the full candidate gene set (Supplementary Tables 3.2 and 3.4). In general, however, the cellular component terms most strongly enriched in the full set of osmoregulatory candidates, extracellular region and its ‘children’ terms, were not enriched in genes expressed solely at low salinity in *C. virginica*.

In contrast to the low population, the most significant functional enrichment at the level of cellular components in the high population was ‘extracellular region’ (GO:0005576, p=6.4e-08). This term refers to the gene products that are secreted from the cell but retained in the interstitial fluid or hemolymph and it was also the most significantly enriched for the full osmoregulatory candidate gene set (Supplementary Table 3.2, Fig. 3.5). While this parent GO term had the highest level of enrichment among the *C. gigas* genes experimentally associated with salinity treatments [19] (Supplementary Table 3.2), it is also likely to include immune response genes responding to the larger disease burden found in oysters from high salinity [19, 44]. Several additional GO terms were significantly enriched both in the asymmetric high salinity genes and in the full candidate gene set including ‘intrinsic to membrane’ (GO:0031224, p=6.4e-06) and ‘plasma membrane’ (GO:0005886, p=1.3e-05) (Supplementary Tables 3.2 and 3.5).

At the molecular functions GO level both the high and low salinity populations showed the strongest significant enrichments related to DNA replication and transcription/translation (Supplementary Tables 3.6 and 3.7). For the low salinity population, many of the other significantly enriched molecular function ontologies (Supplementary Table 3.6) were ‘children’

terms of those significantly enriched both for osmoregulatory genes and for unique low salinity genes (Supplementary Tables 3.3 and 3.6, Fig. 3.6). For example, 'G-protein coupled receptor activity' (GO:0004930, p=6.6e-07) is a 'child' term of 'receptor activity' and 'aspartic-type peptidase activity' (GO:0004190, p=2.2e-06) is a 'child' term of 'hydrolase activity.' These enriched functions match predictions that the phosphorylation of plasma membrane proteins and the hydrolysis of peptides are part of the physiological response to osmotic stress. For the high salinity population, significant enrichment was found for potential osmoregulatory terms related to 'substrate-specific transporter' and 'transmembrane transporter' activities such as 'receptor activity' (GO:0004872, p=7.8e-06), 'gated channel activity' (GO:0022839, p=0.00022), and 'ion gated channel activity' (GO:0022839, p=0.0022) (Supplementary Table 3.7). These enrichment results are consistent with expectations for differential expression of osmoregulatory genes by juvenile eastern oysters from different salinity regimes. Furthermore, it confirms the functional relevance in *C. virginica* of osmoregulatory candidates identified in *C. gigas*.

At the level of reftigs, rather than genes, among those with annotations linked to osmoregulatory function in *C. gigas* (9703 reftigs), 57.3% showed expression in only one of the two populations. This high frequency of asymmetry is in striking contrast to the 11% asymmetry measured at the gene level among osmoregulatory candidates. One possible explanation for this pattern is that asymmetric reftigs represent differentially expressed splice variants of genes expressed by both populations. This hypothesis will be testable with the benefit of this transcriptome as a reference for RNA-seq analyses.

IV. SNP Discovery and dN/dS Ratio with C. gigas

The transcriptome we present here provides the most comprehensive estimate of polymorphism to date for *C. virginica*. Among 13,108 annotated contigs, there was 12,355,575

bp of aligned sequence within which 218,777 SNPs were identified. Average SNP density was 0.0185 per base pair with a standard deviation of 0.0238 (Fig. 3.7). Ninety percent of contigs had at least one predicted SNP. This SNP density falls within the range previously reported for the eastern oyster. Quilang et al. [45] found a rate of 0.0059 SNPs/bp from 4,688 EST sequences. In contrast, Zhang and Guo [46] estimated 0.042 SNPs/bp based on resequencing 6.8kb of ESTs. Similarly, a single gene study of serine protease inhibitor reported an overall SNP frequency of 0.044/bp [47]. For comparison, SNP density averaged across all exons in wild-caught *C. gigas* was 0.0102 per bp [17]. Our finding is therefore consistent with previous estimates of nucleotide heterozygosity in *C. virginica* and tentatively supports the contention that this species is more polymorphic than *C. gigas* [46].

Quantifying genomic patterns of divergence between *C. virginica* and *C. gigas* can help assess the relevance of discoveries in one species with respect to the other. Also, the ratio of substitution rates at nonsynonymous and synonymous sites can help to identify genes undergoing positive selection. After various filtering steps to remove potential artifacts and paralog gene pairs, 26,102 annotated reftigs from *C. virginica* were paired with an ortholog from *C. gigas*. Estimates for the number of nonsynonymous substitutions per nonsynonymous site ranged from near 0 to 0.012/bp per ortholog gene pair (Fig. 3.8A) while the number of synonymous substitutions per synonymous site ranged from 0.0003 to 0.66/bp (Fig. 3.8B). The mean dN/dS ratio of 0.074 (SD = 0.066, Fig. 3.8C) indicates a pervasive role for purifying selection maintaining similar amino acid sequences. The mean protein similarity was 76.8% and the mean nucleotide similarity was 74.2%. It is possible that these divergence estimates between the oyster congeners are biased downward because filtering steps (see Methods) inevitably removed some more divergent ortholog pairs.

This degree of purifying selection provides some confidence that functional candidate genes identified in *C. gigas* will often be applicable to *C. virginica*, at least as a starting point. At the same time, transcriptomes in these two species are probably too diverged to expect *C. gigas* genomic reference sequences to help with *C. virginica* bioinformatics. A simulation study by Vijay et al. [48] demonstrated that reference genomes with average nucleotide sequence divergence up to 15% can help improve transcriptome assemblies while with greater divergence there was no improvement over a *de novo* assembly. Similarly, the potential for a heterologous reference genome to provide improved RNA-seq analyses, relative to a *de novo* transcriptome assembly, was determined to be at nucleotide sequence divergences less than 15% [48].

Conclusions

The goal of our study was to assemble and annotate the *C. virginica* transcriptome with particular focus on potential osmoregulatory genes. Largely with the benefit of the Pacific oyster genome, we assigned provisional annotations to 50,736 reftigs representing over 16,000 putative proteins. More than 80% of the osmoregulatory gene candidates identified in *C. gigas* experiments with adults were identified here in wild juvenile samples from different salinities. The low dN/dS between *C. virginica* and *C. gigas* indicates purifying selection in the coding regions of orthologous genes and provides justification that genes identified as osmoregulatory in *C. gigas* are likely to maintain the same function in *C. virginica.* Even stronger justification is reported for a subset of osmoregulatory candidates that were expressed in only one of the two different salinity populations. Genes with an asymmetric expression pattern across the salinity gradient were significantly enriched for functions that may be related to osmoregulation, consistent with these genes having osmoregulatory functions in *C. virginica*.

Additionally, we have demonstrated that permissive clustering of contig and singleton sequences may improve downstream applications of assembled transcriptomes. In some *de novo* transcriptome assembly studies the singleton reftigs are discarded and only the contigs are analyzed. Such a stringent filter, if applied here to *C. virginica*, would have eliminated 37,717 singletons that were successfully annotated. The goal of clustering is to keep the singletons and reduce redundancy across reftigs that can result from *de novo* assembly challenges due to factors such as sequencing error and high levels of polymorphism. Several studies employing programs such as Cd-hit to cluster sequences based on similarity used a threshold of 95% similarity [7, 49]. We explored a range of threshold values from 100% down to the lower limit of the algorithm at 80%. The improvement in uniquely mapped reads may be beneficial for downstream applications, depending on experimental goals. For RNA-seq experiments, a greater number of uniquely mapped reads means that a greater percentage of data can be retained for the estimation of expression. Future development of these clustering procedures should focus on evaluating trade-offs, particularly with respect to the incorporation of sequencing error at more permissive clustering thresholds.

Finally, we have provided a valuable set of resources for eastern oyster research. We have annotated 50,736 reftigs, doubling the 48,183 *C. virginica* transcriptome contig sequences provided by Zhang et al. [7]. After careful filtering of these reftigs we identified 218,777 candidate SNPs for use in genetic mapping or for population analyses. The 1007 candidate genes for osmoregulation identified here will provide a reference for future studies on the molecular basis of osmoregulation in *C. virginica*, phenotypically plastic responses to salinity stress, and patterns of selective differentiation across heterogeneous environments.

Competing Interests

The authors declare that they have no competing interests.

Authors' contributions

Both authors conceived of and designed the study. LE collected field samples and water quality data, extracted RNA, performed the assembly, annotation and analysis of the transcriptome sequences, and drafted the manuscript. MP advised in the assembly, annotation and analysis of the transcriptome sequences, and was involved in the development of the manuscript. Both authors edited the manuscript and approve of the final paper.

Acknowledgements

We thank Dr. Qi Sun and colleagues at Life Sciences Core Laboratories Computational Biology Service Unit (CBSU) of Cornell University for bioinformatics consultation. We are indebted to Dr. Suzy Strickler for a custom pipeline for dN/dS calculation. Thanks to Dr. Ximing Guo for generously providing *C. gigas* genomic resources. We also thank Dr. Colleen Burge and members of the Hare Lab for comments on this manuscript. We also express gratitude to Sigma Xi Grant-in-Aid of Research and Cornell University for funding.

Table 3.1. Assembly comparison to other molluscan transcriptomes sequenced using 454 technology.
The assemblies are order in decreasing size of estimated total assembly. This calculation is from the mean contig length and number of assembled contigs and is provided as a means to compare transcriptome size.

Species	Normalized	Mean unfiltered read length (bp)	Unfiltered Reads (n)	Assembler	% of filtered reads assembled	Contigs (n)	Mean contig length (bp)	Estimated Total Assembly (bp)*	Reference
Mytilus edulis	No	279	2,393,441	Celera, Cap3	92.0	74,622	645	48,131,190	[50]
Bathymodiolus azoricus	Yes	283	778,996	MIRA	74.8	75,407	509	38,382,163	[51]
Hyriopsis cumingii	No	296	981,302	Cap3	70.5	47,812	634	30,312,808	[52]
Meretrix meretrix	No	413	751,970	Cap3	87.3	35,205	679	23,904,195	[53]
Patinopecten yessoensis	Yes/No	313	970,422	Cap3	86.7	32,590	618	20,140,620	[54]
Crassostrea virginica	Yes	343	1,256,652	Newbler	86.8	28,939	629	18,202,631	Present study
Ruditapes philippinarum	Yes	--	457,717	MIRA3	--	32,606	546	17,802,876	[55]
Chamelea gallina	Yes	210	298,494	MIRA	--	39,750	352	13,992,000	[56]
Laternula elliptica	No	369	1,034,155	Newbler	33.9	18,290	535	9,785,150	[57]
Crassostrea angulata	No	309	555,215	Newbler	79.9	10,462	723	1,057,026	[58]
Pinctada martensii	No	349	434,650	Newbler	--	3,574	--	--	[49]
Pinctada margaritifera	No	234	276,738	TGICL	79.2	76,790	--	--	[59]

Table 3.2. Comparison of quality statistics for transcriptome assembly at unclustered, 95% and 80% sequence identity thresholds.
With more permissive clustering, the number of unique sequences in the transcriptome (reftigs) decreases and the proportion of non-singleton reftigs (contigs) increases. The clustering threshold has a trivial effect on the proportion of reftigs annotated and the number of osmoregulatory candidates recovered. With 42.8% of Illumina reads uniquely mapping, transcriptomes based on more permissive clustering have a greater percentage of annotated reftigs with reads mapped to them as well as a greater percentage of uniquely mapped reads

	Unclustered	95% Sequence Identity Threshold	80% Sequence Identity Threshold
Composition			
Total # Reftigs	136,000	114,716	98,729
% Contigs	16.4%	18.9%	20.5%
% Singletons	83.6%	81.1%	79.5%
Annotation			
Total Reftigs Annotated		58,811	50,736
% Reftigs Annotated		51.3%	51.4%
% of Contigs		62.6%	64.3%
% of Singletons		48.6%	48.1%
# of Osmoregulation Candidates		1014	1007
Sample Illumina Reads Mapped to Annotated Transcriptome			
% Reftigs with Mapped Reads		86.7%	88.3%
% of Reads Mapped Uniquely		40.8%	42.8%

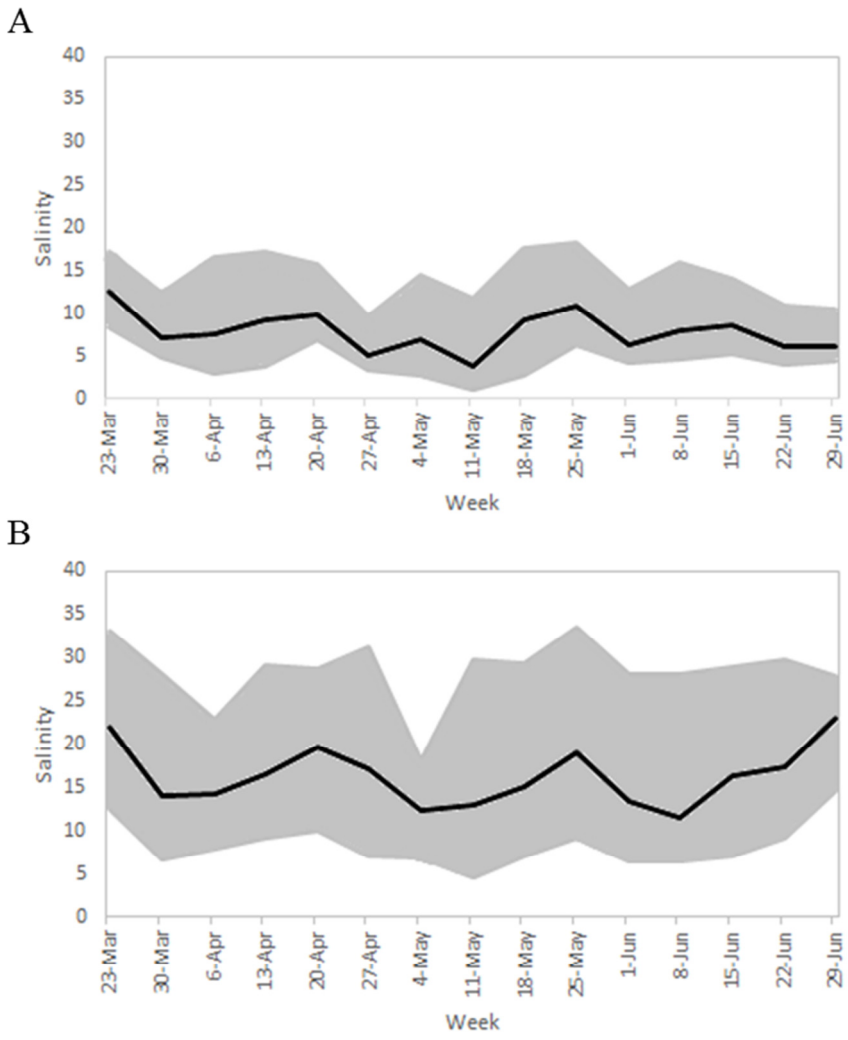

Figure 3.1. Salinity at the 'high' and 'low' wild juvenile oyster collections sites. A) Maximum mean and minimum salinity at low salinity site. B) Maximum, mean, and minimum salinity at high salinity site

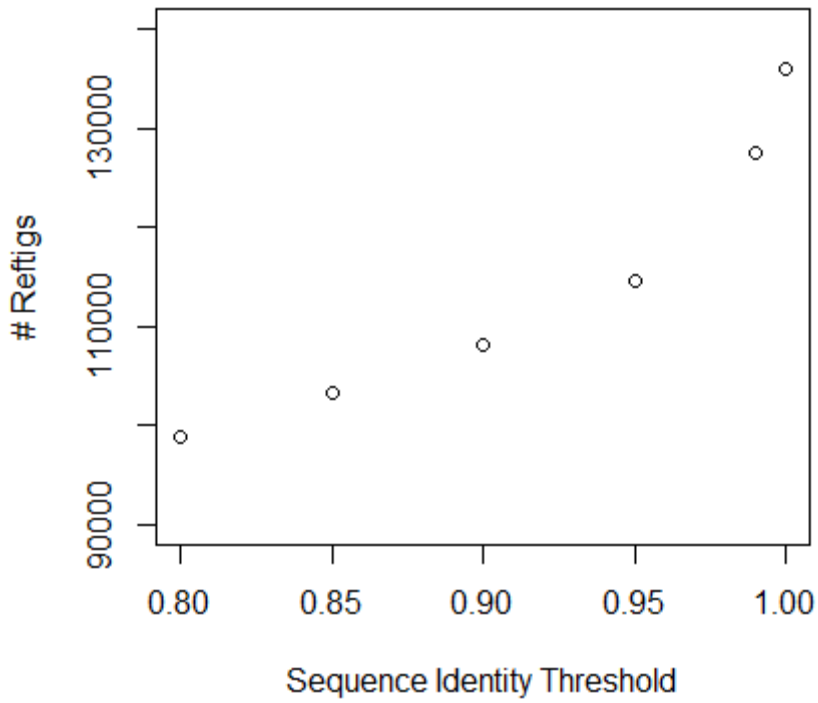

Figure 3.2. Collapsing of reftigs with decreasing sequence identity threshold in Cd-hit. As sequence identity threshold decreases in Cd-hit clustering, the number of reftigs decreases. The rate of reftig consolidation is highest from 1 to 0.95 but remains consistent from 0.95 to 0.90.

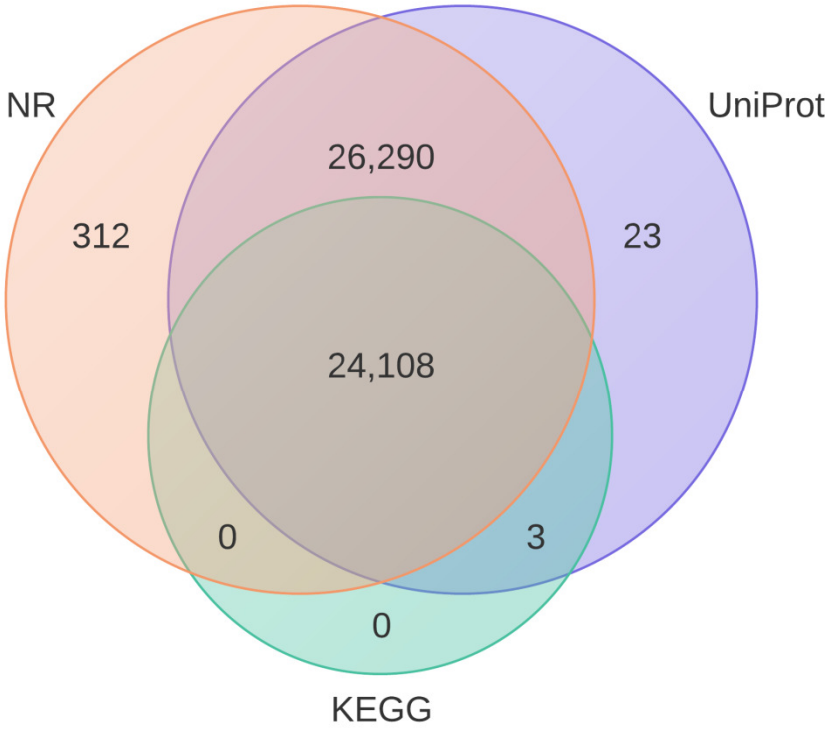

Figure 3.3. Number of reftigs annotated by each database out of 50,736 total reftigs. The GenBank nr database supplied 99.95% of annotations.

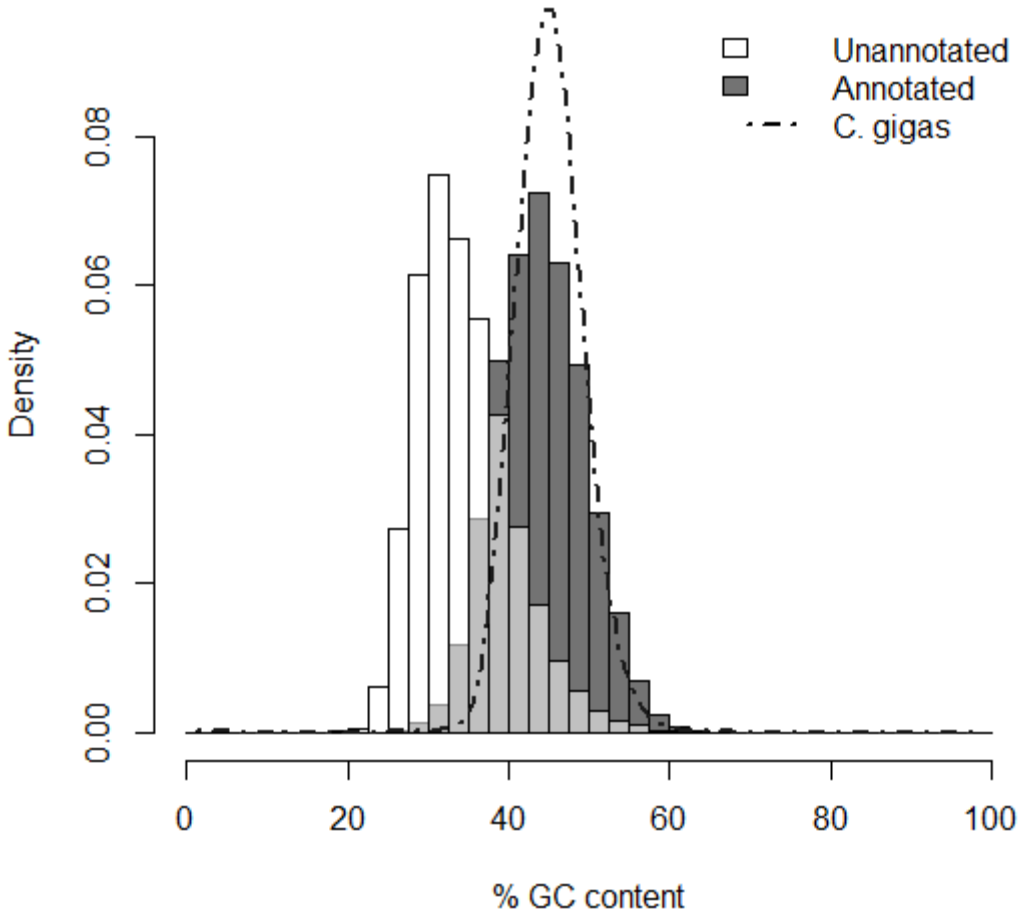

Figure 3.4. GC content of annotated and unannotated portions of transcriptome and *C. gigas*.
GC content of the annotated portion of the transcriptome is higher than in the unannotated portion, suggesting potential contamination from other species or the presence of other RNA types in the sequencing. The GC content of the annotated portion mirrors the GC content found in the coding sequences of *C. gigas*.

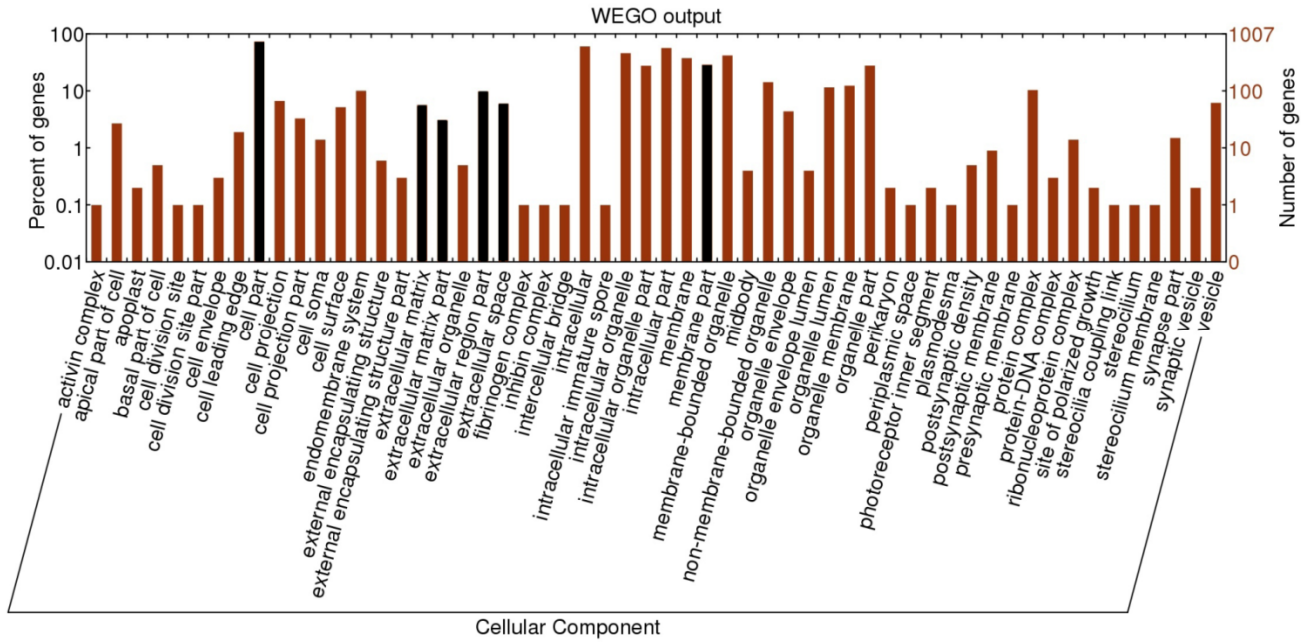

Figure 3.5. Distribution of level 3 Cellular Component GO terms for the osmoregulatory candidate genes.
Black bars indicate the terms or the parents of GO terms that are significantly enriched in the osmoregulatory candidate genes compared to the complete set of annotated genes, while the red bars indicate GO terms that are not significantly different between the candidates and the complete set of annotated genes.

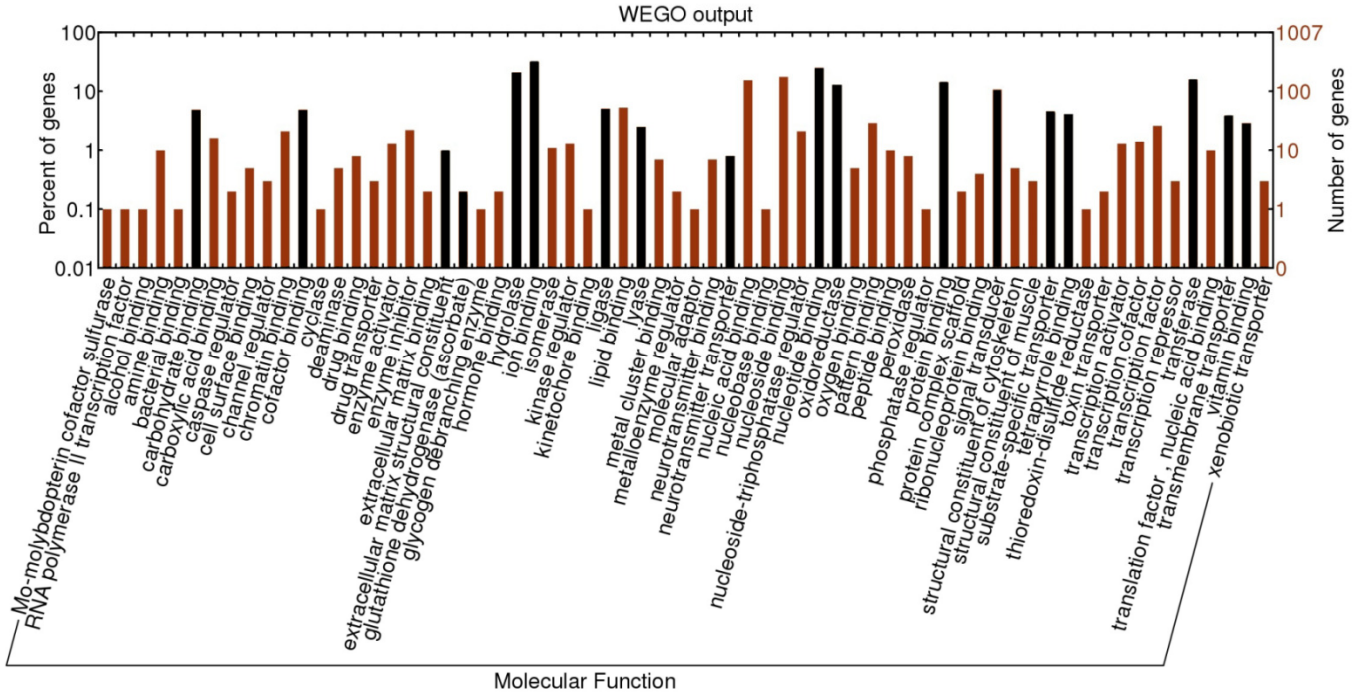

Figure 3.6. Distribution of level 3 Molecular Function GO terms for the osmoregulatory candidate genes.
Black bars indicate the terms or the parents of GO terms that are significantly enriched in the osmoregulatory candidate genes compared to the complete set of annotated genes, while the red bars indicate GO terms that are not significantly different between the candidates and the complete set of annotated genes.

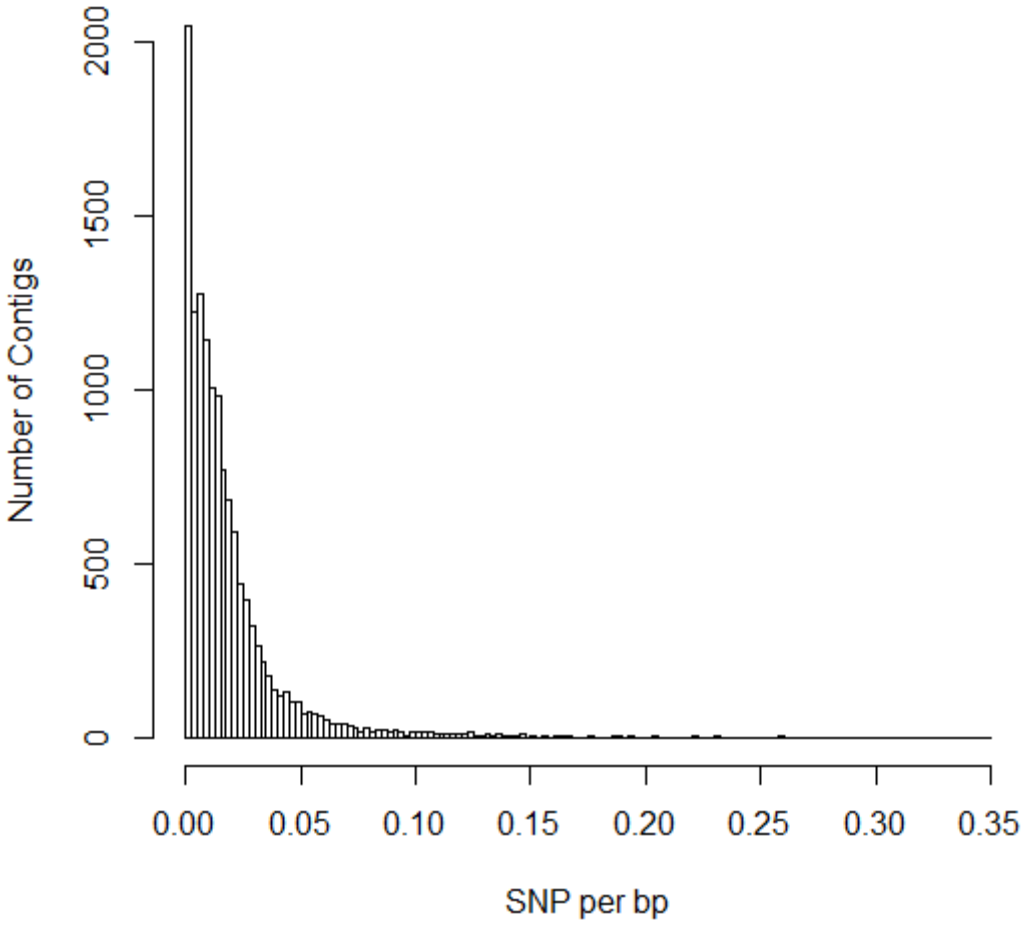

Figure 3.7. The distribution of SNP density per base pair within annotated contigs from the 80% clustered transcriptome.

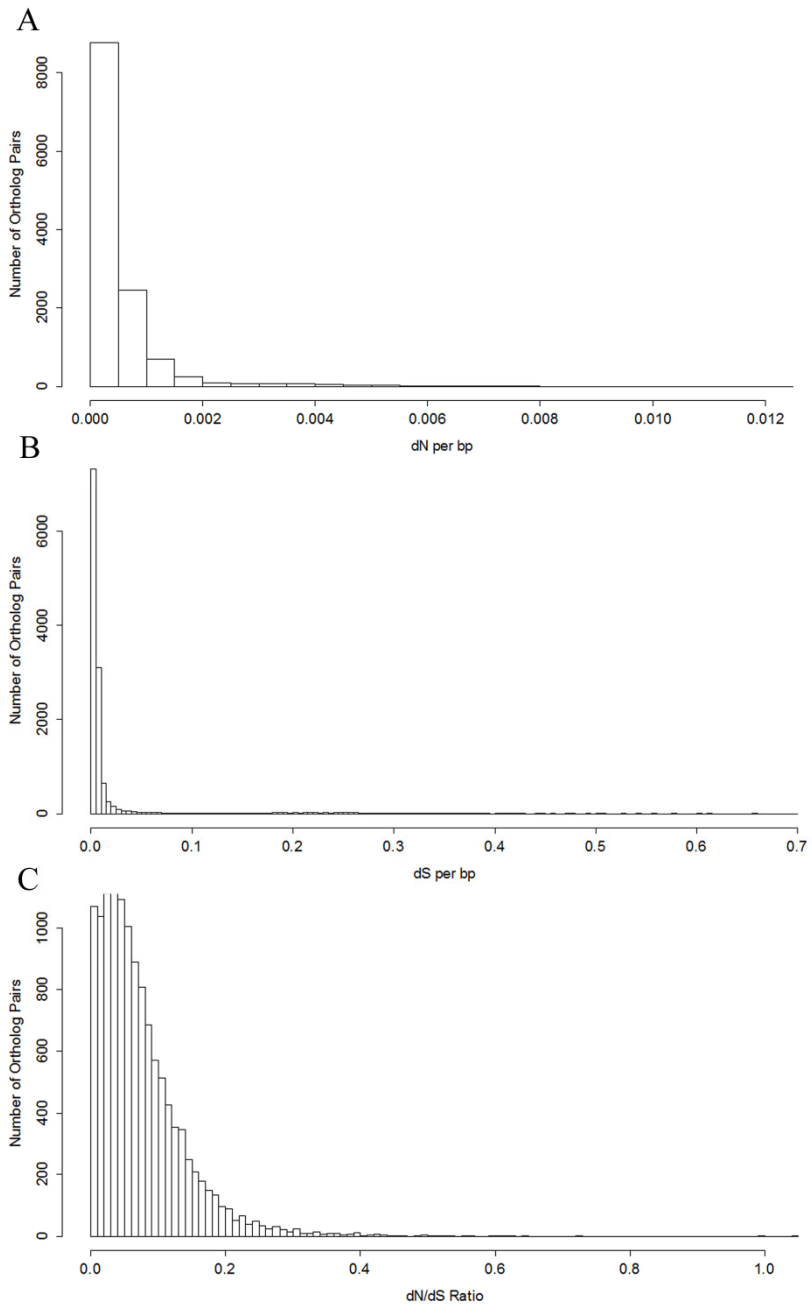

Figure 3.8. Distribution of (A) dN, (B) dS and (C) dN/dS ratio values.
dN and dS show similar distribution shapes with the number of synonymous substitutions much larger than the number of nonsynonymous substitutions. Most ortholog pairs had a dN/dS ratio below 0.2 indicating a strong role for purifying selection on oyster peptide sequences.

REFERENCES

1. Coen LD, Luckenbach MW, Breitburg DL: **The role of oyster reefs as essential fish habitat: a review of current knowledge and some new perspectives.** *Am Fish Soc Symp* 1999, **34**:303–307.
2. Coen LD, Brumbaugh RD, Bushek D, Grizzle R, Luckenback MW, Posey MH, Powers SP, Tolley SG: **Ecosystem services related to oyster restoration.** *Marine Ecology Progress Series* 2007, **341**:303–307.
3. zu Ermgassen PSE, Spalding MD, Grizzle RE, Brumbaugh RD: **Quantifying the loss of a marine ecosystem service: Filtration by the eastern oyster in US estuaries.** *Estuaries and Coasts* 2013, **36**:36-43.
4. Beck MW, Brumbaugh RD, Airoldi L, Carranza A, Coen LD, Crawford C, Defeo O, Edgar GJ, Hancock B, Kay M, Lenihan HS, Luckenbach MW, Toropova CL, Zhang G, Guo X: **Oyster reefs at risk and recommendations for conservation, restoration and management.** *BioScience* 2011, **61**: 107–116.
5. Willberg MJ, Livings ME, Barkman JS, Morris BT, Robinson JM: **Overfishing, disease, habitat loss, and potential extirpation of oysters in the upper Chesapeake Bay.** *Marine Ecology Progress Series* 2011, **436**:131-144.
6. Chapman RW, Mancia A, Beal M, Veloso A, Rathburn C, Blair A, Holland AF, Warr GW, Didinato G, Sokolova IM, Wirth EF, Duffy E, Sanger D: **The transcriptomic response of eastern oyster, *Crassostrea virginica*, to environmental conditions.** *Molecular Ecology* 2011, **20**(7):1431-49.

7. Zhang L, Li L, Zhu Y, Zhang G, Guo X: **Transcriptome analysis reveals a rich gene set related to innate immunity in the eastern oyster (*Crassostrea virginica*).** *Marine Biotechnology* 2013, **2013**: 1-17.
8. Goedken M, Morsey B, Sunila I, de Guise S: **Immunomodulation of *Crassostrea gigas* and *Crassostrea virginica* cellular defense mechanisms by *Perkinsus marinus*.** *Journal of Shellfish Research* 2005, **24**(2):487-496.
9. Tanguy A, Guo X, Ford SE: **Discovery of genes expressed in response to *Perkinsus marinus* challenge in Eastern (*Crassostrea virginica*) and Pacific (*Crassostrea gigas*) oysters.** *Gene* 2004, **338**(1):121-131.
10. Jenny MJ, Ringwood AH, Lacy ER, Lewitus AJ, Kempton JW, Gross PS, Warr GW, Chapman RW: **Potential indicators of stress response identified by expressed sequence tag analysis of hemocytes and embryos from the American oyster, *Crassostrea virginica*.** *Marine Biotechnology* 2002, **4**(1): 81-93.
11. Anderson RS, Burreson EM, Paynter KT: **Defense responses of hemocytes withdrawn from *Crassostrea virginica* infected with *Perkinsus marinus*.** *Journal of Invertebrate Pathology* 1995, **66**:82-89.
12. Ivanina AV, Taylor C, Sokolova IM: **Effects of elevated temperature and cadmium exposure on stress protein response in eastern oysters *Crassostrea virginica* (Gmelin).** *Aquatic Toxicology* 2009, **91**(3):245-254.
13. Loosanoff VL: **Behavior of oysters in waters of low salinity.** *Proceedings of the National Shellfisheries Association* 1953, **1952**:135–151.
14. Shaw WN: **Oyster setting in two adjacent tributaries of Chesapeake Bay.** *ASB Bulletin* 1966, **13**(2):45.

15. Shumway SE: **Natural environmental factors.** In: *The Eastern Oyster Crassostrea virginica.* Edited by Kennedy VS, Newell RIE, Eble AF. College Park, Maryland: Maryland Sea Grant College Publication; 1996:467–513.

16. Eierman LE, Hare MP: **Survival of oyster larvae in different salinities depends on source population within an estuary.** *Journal of Experimental Marine Biology and Ecology* 2013a, **449**:61-68.

17. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, et al: **The oyster genome reveals stress adaptation and complexity of shell formation.** *Nature* 2012, **490**:49–54.

18. Meng J, Zhu Qihui, Zhang L, Li C, Li L, She, Z, Huang B, Zhang G: **Genome and transcriptome analyses provide insight into the euryhaline adaptation mechanism of *Crassostrea gigas.*** *PLoS ONE* 2013, **8**(3):e58563 doi:10.1371/journal.pone.0058563.

19. Zhao Z, Yu H, Kong K, Li Q: **Transcriptomic responses to salinity stress in the Pacific oyster *Crassostrea gigas*.** *PLoS One* 2012, **7**(9):e46244 doi: 10.1371/journal.pone.0046244.

20. Evans DH: Osmotic and Ionic Regulation: Cells and Animals. Boca Raton: CRC Press; 2009.

21. Pierce SK: Invertebrate cell volume control mechanisms: **A coordinated use of intracellular amino acids and inorganic ions as osmotic solute.** *Biological Bulletin* 1982, **163**: 405–419. http://dx.doi.org/10.2307/1541452.

22. Hosoi M, Kubota S, Toyohora M, Toyohora H, Hayashi I: **Effect of salinity change on free amino acid content in Pacific oyster.** *Fisheries Science* 2003, **69**(2):395-400.

23. Pierce SK, Rowland-Faux LM, O'Brien SM: **Different salinity tolerance mechanisms in Atlantic and Chesapeake Bay conspecific osyters: glycine betaine and amino acid pool variations.** *Marine Biology* 1992, **113**:107-115.

24. Yancey PH: **Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.** *Journal of Experimental Biology* 2005, **208**:2819-2830.

25. Toyohora H, Yoshida M, Hosoi M, Hayashi I: **Expression of taurine transporter in response to hypo-osmotic stress in the mantle of Mediterranean blue mussel.** *Fisheries Science* 2005, **71**: 356-360.

26. Hosoi M, Shinzato C, Masaya T, Hosoi-Tanabe S, Sawada H, Terasawa E, Toyohara H: **Taurine transporter function from the giant Pacific oyster *Crassostrea gigas*: function and expression in response to hyper- and hypo-osmotic stress.** *Fisheries Science* 2007, **73**:385-394.

27. Ren J, Liu X, Jiang F, Guo X, Liu B: **Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia.** *BMC Evolutionary Biology* 2010, **10**:394.

28. Lambert JD, Chan XY, Spiecker B, Sweet HC: **Characterizing the embryonic transcriptome of the snail *Ilyanassa*.** *Integr Comp Biol* 2010, **50**(5): 768-777.

29. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV: **Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx.** *BMC Genomics* 2009, **10**:219.

30. Wolf JBW, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D: **Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow.** *Molecular Ecology* 2010, **19**:162-175.

31. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22**(13):1658-1659.

32. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler Transform.** Bioinformatics 2009, **25**:1754-60.

33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.

34. Alexa A, Rahnenfuhrer J: **topGO: Enrichment analysis for Gene Ontology.** 2010, R package version 2.12.0.

35. Smit AFA, Hubley G, Green P: **RepeatMasker Open-3.0.** 1996-2010. http://www.repeatmasker.org

36. Iseli C, Jongeneel CV, Bucher P: **ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** *Proc Int Conf Intell Syst Mol Biol* 1999,**1999**:138-148.

37. Larkin MA, Blackshields G, Brown NP,Chenna R, McGettigan PA, McWilliams H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947-2948.

38. Yang Z: **PAML 4:Phylogenetic Analysis by Maximum Likelihood.** *Molecular Biology and Evolution* 2007, **24**(8): 1586-1591.

39. Raineri E, Ferretti L, Esteve-Codina A, Nevado B, Heath S, Pérez-Enciso M: **SNP calling by sequencing pooled sample.** *BMC Bioinformatics* 2012, **13**:239.

40. Eierman LE, Hare MP: ***Crassostrea virginica*** **transcriptome.** *FigShare* 2013b, Available online at http://dx.doi.org/10.6084/m9.figshare.873865

41. Cheung F, Haas BJ,Goldberg SMD, May GD, Xiao Y, Town CD: **Sequencing *Medicago truncatula* expressed sequence tags using 454 Life Science technology.** *BMC Genomics* 2006, **7**:272.

42. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Molecular Ecology* 2008, **17**(7):1636-1647.

43. Calvo GW, Luckenback MW, Allen Jr. SK, Burreson EM: **Comparative field study of *Crassostrea gigas* and *Crassostrea virginica* in relation to salinity in Virginia.** *Journal of Shellfish Research* 1999, **18**:465-473.

44. Chu FE, La Peyre JF, Burreson CS: ***Perkinsus marinus*** **infection and potential defense-related activities in eastern oysters, *Crassostrea virginica*: Salinity effects.** *Journal of Invertebrate Pathology* 1993, **62**(3):226-232.

45. Quilang J, Wang S, Li P, Abernathy J, Peatman E, Wang Y, Wang L, Shi Y, Wallace R, Guo X, Liu Z: **Generation and analysis of ESTs from the eastern oyster, Crassostrea virginica Gmelin and identification of microsatellite and SNP markers.** *BMS Genomics* 2007, **8**:157.

46. Zhang L, **Guo X**: **Development and validation of single nucleotide polymorphism markers in the eastern oyster Crassostrea virginica Gmelin by mining ESTs and resequencing.** *Aquaculture* 2010, **302**:124-129.

47. Yu H, He Y, Wang X, Zhang Q, Bao Z, Guo X: **Polymorphism in a serine protease inhibitor gene and its association with disease resistance in the eastern oyster (Crassostrea virginica Gmelin).** *Fish Shellfish Immunol* 2011, **30**(3):757-762.

48. Vijay N, Poelstra JW, Künstner A, Wolf JBW: **Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments.** *Molecular Ecology* 2013, **22**:620-634.

49. Shi Y, Yu C, Gu Z, Zhan X, Wang Y, Wang A: **Characterization of the Pearl Oyster (*Pinctada martensii*) mantle transcriptome unravels biomineralization genes.** *Marine Biotechnology* 2013, **15**:175-189.

50. Philipp EER, Kraemer L, Melzner F, Poustka AJ, Thieme S, Findeisen U, Schreiber S, Rosenstiel P: **Massively Parallel RNA Sequencing Identifies a Complex Immune Gene Repertoire in the lophotrochozoan Mytilus edulis.** *PLoS ONE* 2012, **7**(3): e33091. doi:10.1371/journal.pone.0033091

51. Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS: **High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*.***BMC Genomics* 2010, **11**:559.

52. Bai Z, Zheng H, Lin J, Wang G, Li J: **Comparative Analysis of the Transcriptome in Tissues Secreting Purple and White Nacre in the Pearl Mussel Hyriopsis cumingii.** *PLoS ONE* 2013, **8**(1): e53617. doi:10.1371/journal.pone.0053617

53. Huan P, Wang H, Liu B: **Transcriptomic analysis of the clam *Meretrix meretrix* on different larval stages.** *Marine Biotechnology* 2012, **14**:69-78.

54. Hou R, Bao Z, Wang S, Su H, Li Y, Du H, Hu J, Wang S, Hu X: **Transcriptome Sequencing and De Novo Analysis for Yesso Scallop (Patinopecten yessoensis) Using 454 GS FLX.** *PLoS ONE* 2011, **6**(6): e21560. doi:10.1371/journal.pone.0021560

55. Milan M, Coppe A, Reinhardt R, Cancela LM, Leite RB, Saavedra C, Ciofi C, Chelazzi G, Patarnello T, Bortoluzzi S, Bargelloni L: **Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring.** *BMC Genomics* 2011, **12**:234.

56. Coppe A, Bortoluzzi S, Murari G, Marino IAM, Zane L, Papetti C: **Sequencing and Characterization of Striped Venus Transcriptome Expand Resources for Clam Fishery Genetics**. *PLoS ONE* 2012, **7**(9): e44185. doi:10.1371/journal.pone.0044185

57. Clark MS, Thorne MAS, Vieira FA, Cardoso JCR, Power DM, Peck LS: **Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing.** *BMC Genomics* 2010, **11**:362.

58. Qin J, Huang Z, Chen J, Zou Q, You W, Ke C: **Sequencing and *de novo* Analysis of *Crassostrea angulata* (Fujian Oyster) from 8 Different Developing Phases Using 454 GSFlx.** *PLoS ONE* 2012, **7**(8): e43653. doi:10.1371/journal.pone.0043653

59. Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y, Montagnani C: **Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization**. *BMC Genomics* 2010, **11**:613.

Supplementary Table 3.1. Annotation of osmoregulatory candidate transcriptome reftigs from *C. virginica*
(Microsoft Excel spreadsheet)

Each row is an osmoregulatory candidate reftig identified from the *C. virginica* transcriptome. The identification as an osmoregulatory candidate is from the match between the GenBank nr description (column 3) of the annotated reftig to the description of an osmoregulatory candidate identified by Zhang et al. [17] in the *C. gigas* genome. Information provided in the table for each reftig are the reftig length, GenBank nr description, nr e-value, UniProt match, UniProt e-value, UniProt ID, KEGG ID and nucleotide sequence. The table is available at http://dx.doi.org/10.6084/m9.figshare.873865.

Supplementary Table 3.2. Significantly enriched gene ontologies for cellular components in osmoregulatory candidate genes.

Osmoregulatory candidate genes were compared to the complete set of annotated genes, ordered by functional category. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor. Indentations represent the 'parent':'child' tiered relationship of GO terms with deeper indentations representing more specific terminology relative to the boldface level-three 'parent' terms shown as enriched in Fig. 3.5.

GO ID	Term	p-value
GO:0005576	extracellular region	4.6e-16
GO:0044421	**extracellular region part**	1.9e-15
GO:0031012	**extracellular matrix**	7.9e-14
GO:0044420	**extracellular matrix part**	1.7e-14
GO:0005578	proteinaceous extracellular matrix	1.1e-12
GO:0005581	Collagen	1.2e-11
GO:0005604	basement membrane	0.0033
GO:0005615	**extracellular space**	7.4e-07
GO:0005623	Cell	---------
GO:0044464	**cell part**	---------
GO:0071944	cell periphery	0.0011
GO:0005886	plasma membrane	0.0017
GO:0044459	plasma membrane part	0.0032
GO:0016020	Membrane	---------
GO:0044424	**membrane part**	---------
GO:0031224	intrinsic to membrane	---------
GO:0031225	anchored to membrane	0.0076

Supplementary Table 3.3. Significantly enriched gene ontologies for molecular functions in osmoregulatory candidate genes.

Osmoregulatory candidate genes were compared to the complete set of annotated genes, ordered by functional category. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor. Indentations represent the 'parent':'child' tiered relationship of GO terms with deeper indentations representing more specific terminology relative to the boldface level-three 'parent' terms shown as enriched in Fig. 3.6.

GO.ID	Term	p-value
GO:0003824	catalytic activity	6.3e-12
GO:0016491	**oxidoreductase activity**	1.2e-16
GO:0004497	monooxygenase activity	1.6e-14
GO:0016705	oxidoreductase activity acting on paired donors with incorporation or reduction of molecular oxygen	5.3e-14
GO:0016713	oxidoreductase activity acting on paired donors, with incorporation or reduction of molecular oxygen, reduced iron-sulfur protein as one donor, and incorporation of one atom of oxygen	0.00068
GO:0018685	alkane 1-monooxygenase activity	0.00068
GO:0031545	peptidyl-proline 4-dioxygenase activity	0.00028
GO:0031543	peptidyl-proline dioxygenase activity	0.00077
GO:0008392	arachidonic acid epoxygenase activity	0.00771
GO:0016712	oxidoreductase activity acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	3.5e-07
GO:0070330	aromatase activity	1.1e-06
GO:0016715	oxidoreductase activity acting on paired donors, with incorporation or reduction of molecular oxygen, reduced ascorbate as one donor, and incorporation of one atom of oxygen	0.00014
GO:0004500	dopamine beta-monooxygenase activity	0.00028
GO:0016627	oxidoreductase activity acting on CH-CH group of donors	---------
GO:0017150	tRNA dihydrouridine synthase activity	0.00253
GO:0016614	oxidoreductase activity acting on CH-OH group of donors	0.00335
GO:0016618	hydroxypyruvate reductase activity	0.00771
GO:0030267	glyoxylate reductase (NADP) activity	0.00771
GO:0030613	oxidoreductase activity acting on phosphorus or arsenic in donors	0.00771
GO:0030614	oxidoreductase activity acting on phosphorus or arsenic as donors, disulfide as acceptor	0.00771
GO:0050610	methylarsonate reductase activity	0.00771
GO:0016641	oxidoreductase activity acting on CH-NH2 group of donors, oxygen as acceptor	---------

Supplementary Table 3.3 continued

GO:0052597	diamine oxidase activity	0.00771
GO:0052598	histamine oxidase activity	0.00771
GO:0052599	methylputrescine oxidase activity	0.00771
GO:0052600	propane-1,3-diamine oxidase activity	0.18
GO:0016682	oxidoreductase activity acting on diphenols and related substances as donors, oxygen as acceptor	0.00807
GO:0051213	dioxygenase activity	0.00919
GO:0016787	**hydrolase activity**	---
GO:0004725	protein tyrosine phosphatase activity	3.5e-06
GO:0016791	phosphatase activity	1.8e-05
GO:0042578	phosphoric ester hydrolase activity	0.00023
GO:0004721	phosphoprotein phosphatase activity	0.00107
GO:0008833	deoxyribonuclease IV (phage-T4-induced) activity	0.00771
GO:0016740	**transferase activity**	---
GO:0016763	transferase activity transferring pentosyl groups	1.5e-05
GO:0003950	NAD+ ADP-ribosyltransferase activity	5.5e-05
GO:0016757	transferase activity transferring glycosyl groups	0.00232
GO:0047273	galactosylgalactosylglucosylceramide beta-D-acetylgalactosaminyltransferase activity	0.00771
GO:0016769	transferase activity transferring nitrogenous groups	0.00451
GO:0008483	transaminase activity	0.00451
GO:0016874	**ligase activity**	---
GO:0004812	aminoacyl-tRNA ligase activity	0.00085
GO:0016875	ligase activity forming carbon-oxygen bonds	0.00085
GO:0016876	ligase activity forming aminoacyl-tRNA and related compounds	0.00085
GO:0016829	**lyase activity**	---
GO:0004794	L-threonine ammonia-lyase activity	0.00771
GO:0005488	Binding	---
GO:0046906	**tetrapyrrole bonding**	3.3e-12

Supplementary Table 3.3 continued

GO:0020037	heme binding	1.7e-12
GO:0043167	**ion binding**	0.00012
GO:0043169	cation binding	---------
GO:0005506	iron ion binding	1.1e-10
GO:0005507	copper ion binding	4.3e-07
GO:0043168	anion binding	8.7e-05
GO:0030170	pyridoxal phosphate binding	0.00023
GO:0070403	NAD+ binding	0.00113
GO:0031406	carboxylic acid binding	0.00327
GO:0031418	L-ascorbic acid binding	0.00451
GO:0048037	**cofactor binding**	3.4e-05
GO:0030246	**carbohydrate binding**	0.00020
GO:0048029	monosaccharide binding	0.00802
GO:0036094	**small molecule binding**	0.00120
GO:0019842	vitamin binding	0.00116
GO:0000166	nucleotide binding	0.00527
GO:0005515	**receptor binding**	---------
GO:0070696	transmembrane receptor protein serine/threonine kinase binding	0.00253
GO:0033612	receptor serine/threonine kinase binding	0.00590
GO:0097159	**organic cyclic compound binding**	---------
GO:1901265	nucleoside phosphate binding	0.00527
GO:0005515	**protein binding**	---------
GO:0070697	activin receptor binding	0.00771
GO:0070699	type II activin receptor binding	0.00771
GO:0009055	electron carrier activity	2.9e-08
GO:0005215	transporter activity	---------
GO:0022892	**substrate-specific transporter activity**	---------
GO:0015370	solute:sodium symporter activity	1.5e-07
GO:0015294	solute:cation symporter activity	1.8e-05
GO:0005343	organic acid:sodium symporter activity	0.00013

Supplementary Table 3.3 continued

GO:0017153	sodium:dicarboxylate symporter activity	0.00048
GO:0015081	sodium ion transmembrane transporter activity	0.00079
GO:0005310	dicarboxylic acid transmembrane transporter activity	0.00735
GO:0015501	glutamate:sodium symporter activity	0.00771
GO:0005326	**neurotransmitter transporter activity**	0.00035
GO:0005328	neurotransmitter:sodium symporter activity	0.00035
GO:0022857	**transmembrane transporter activity**	---------
GO:0015293	symporter activity	0.00926
GO:0015296	anion:cation symporter activity	0.00994
GO:0060089	molecular transducer activity	---------
GO:0004871	**signal transducer activity**	---------
GO:0004872	receptor activity	0.00210
GO:0005001	transmembrane receptor protein tyrosine phosphatase activity	0.00026
GO:0019198	transmembrane receptor protein phosphatase activity	0.00026
GO:0005044	scavenger receptor activity	0.00131
GO:0004955	prostaglandin receptor activity	0.00253
GO:0004888	transmembrane signaling receptor activity	0.00298
GO:0038024	cargo receptor activity	0.00320
GO:0004953	icosanoid receptor activity	0.00590
GO:0004954	prostanoid receptor activity	0.00590
GO:0016209	antioxidant activity	0.00075
GO:0045174	**glutathione dehydrogenase (ascorbate) activity**	0.00771
GO:0005198	structural molecule activity	---------
GO:0005201	**extracellular matrix structural constituent**	0.00275

Supplementary Table 3.4. Significantly enriched gene ontologies for cellular components from the low salinity population.

Significantly enriched gene ontologies in 1:0 asymmetric genes from the low salinity population are ordered by p-value. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor.

GO ID	Term	p-value
GO:0016021	integral to membrane	**0.00037**
GO:0031224	intrinsic to membrane	**0.00058**
GO:0000795	synaptonemal complex	**0.00135**
GO:0005887	integral to plasma membrane	**0.00305**
GO:0071944	cell periphery	**0.00333**
GO:0042383	Sarcolemma	**0.00434**
GO:0031226	intrinsic to plasma membrane	**0.00490**
GO:0044459	plasma membrane part	**0.00624**
GO:0044425	membrane part	**0.00663**
GO:0000794	condensed nuclear chromosome	**0.00729**
GO:0043025	neuronal cell body	**0.00802**
GO:0005886	plasma membrane	**0.00815**

Supplementary Table 3.5. Significantly enriched gene ontologies for cellular components from the high salinity population.

Significantly enriched gene ontologies in 1:0 asymmetric genes from the high salinity population are ordered by p-value. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor.

GO ID	Term	p-value
GO:0005576	extracellular region	6.4e-08
GO:0031224	intrinsic to membrane	6.4e-06
GO:0005886	plasma membrane	1.3e-05
GO:0016021	integral to membrane	3.1e-05
GO:0071944	cell periphery	3.1e-05
GO:0044425	membrane part	0.00013
GO:0016020	Membrane	0.00055
GO:0031012	extracellular matrix	0.00061
GO:0005578	proteinaceous extracellular matrix	0.00082
GO:0044421	extracellular region part	0.00182
GO:0097060	synaptic membrane	0.00430
GO:0005615	extracellular space	0.00450
GO:0032992	protein-carbohydrate complex	0.00532
GO:0071666	Slit-Robo signaling complex	0.00532

Supplementary Table 3.6. Significantly enriched gene ontologies for molecular function from the low salinity population.

Significantly enriched gene ontologies in 1:0 asymmetric genes from the low salinity population are ordered by p-value. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor.

GOID	Term	p-value
GO:0034061	DNA polymerase activity	1.7e-21
GO:0003964	RNA-directed DNA polymerase activity	5.9e-18
GO:0016779	nucleotidyltransferase activity	1.2e-16
GO:0004518	nuclease activity	3.2e-13
GO:0003676	nucleic acid binding	2.9e-10
GO:0004519	endonuclease activity	8.4e-09
GO:0003887	DNA-directed DNA polymerase activity	6.4e-07
GO:0004930	G-protein coupled receptor activity	6.6e-07
GO:0004190	aspartic-type endopeptidase activity	1.0e-06
GO:0070001	aspartic-type peptidase activity	2.2e-06
GO:0016772	transferase activity, transferring phosphate-containing groups	8.5e-06
GO:0016788	hydrolase activity, acting on ester bonds	8.2e-05
GO:0038023	signaling receptor activity	0.00010
GO:0004872	receptor activity	0.00011
GO:0004888	transmembrane signaling receptor activity	0.00014
GO:0003677	DNA binding	0.00014
GO:0016787	hydrolase activity	0.00048
GO:0000405	bubble DNA binding	0.00049
GO:0004386	helicase activity	0.00152
GO:0004871	signal transducer activity	0.00200
GO:0060089	molecular transducer activity	0.00200
GO:1901363	heterocyclic compound binding	0.00208
GO:0097159	organic cyclic compound binding	0.00254
GO:0004527	exonuclease activity	0.00287
GO:0046914	transition metal ion binding	0.00332
GO:0008270	zinc ion binding	0.00349
GO:0004854	xanthine dehydrogenase activity	0.00431
GO:0004855	xanthine oxidase activity	0.00431

Supplementary Table 3.6 continued

GO:0016726	oxidoreductase activity, acting on CH or CH2 groups, NAD or NADP as acceptor	0.00431
GO:0016727	oxidoreductase activity: acting on CH or CH2 groups, oxygen as acceptor	0.00431
GO:0016725	oxidoreductase activity: acting on CH or CH2 groups	0.00460
GO:0008265	Mo-molybdopterin cofactor sulfurase activity	0.00619
GO:0009378	four-way junction helicase activity	0.00619
GO:0035312	5'-3' exodeoxyribonuclease activity	0.00619
GO:0042302	structural constituent of cuticle	0.00619
GO:0043176	amine binding	0.00619
GO:0045145	single-stranded DNA specific 5'-3' exodeoxyribonuclease activity	0.00619
GO:0046873	metal ion transmembrane transporter activity	0.00666
GO:0005262	calcium channel activity	0.00867

Supplementary Table 3.7. Significantly enriched gene ontologies for molecular function from the high salinity population.

Significantly enriched gene ontologies in 1:0 asymmetric genes from the high salinity population are ordered by p-value. The p-value is derived from a Fisher's exact test implemented in topGO from Bioconductor.

GO ID	Term	p-value
GO:0003964	RNA-directed DNA polymerase activity	1.5e-07
GO:0034061	DNA polymerase activity	1.2e-06
GO:0004872	receptor activity	7.8e-06
GO:0016779	nucleotidyltransferase activity	7.1e-05
GO:0022836	gated channel activity	0.00022
GO:0022839	ion gated channel activity	0.00022
GO:0038023	signaling receptor activity	0.00036
GO:0004888	transmembrane signaling receptor activity	0.00040
GO:0015276	ligand-gated ion channel activity	0.00042
GO:0022834	ligand-gated channel activity	0.00042
GO:0008270	zinc ion binding	0.00112
GO:0005230	extracellular ligand-gated ion channel activity	0.00148
GO:0004930	G-protein coupled receptor activity	0.00178
GO:0005216	ion channel activity	0.00248
GO:0022838	substrate-specific channel activity	0.00311
GO:0015267	channel activity	0.00444
GO:0022803	passive transmembrane transporter activity	0.00444
GO:0001786	phosphatidylserine binding	0.00525
GO:0005231	excitatory extracellular ligand-gated ion channel activity	0.00615
GO:0004970	ionotropic glutamate receptor activity	0.00618
GO:0046914	transition metal ion binding	0.00696
GO:0004890	GABA-A receptor activity	0.00710
GO:0005234	extracellular-glutamate-gated ion channel activity	0.00837

CHAPTER 4

REEF SPECIFIC PATTERNS OF GENE EXPRESSION PLASTICITY IN EASTERN OYSTERS (*CRASSOSTREA VIRGINICA*) AFTER OSMOTIC ACCLIMATION IN HIGH AND LOW SALINITY COMMON GARDENS[3]

Abstract

Environmental variation that causes differential fitness among genotypes results in natural selection, particularly beyond the limits of phenotypic plasticity. Understanding the interaction between these two processes is important for predicting species' persistence at range margins in rapidly changing environments. In species with sedentary adults, many traits typically show broad phenotypic plasticity. However, when their life history includes high fecundity and broadly dispersing offspring, there also is a large capacity for selection to reshape the functional genetic composition of populations across small scale habitat heterogeneity every generation. We tested for this pattern of selection using RNA-seq methods to study functional genetic variation controlling gene expression in adult eastern oysters, *Crassostrea virginica,* from different reefs after acclimation to salinity in common gardens. We analyzed 24 adult oysters collected from high and low salinity source reefs within a single estuary after they were acclimated for nine weeks to two common gardens: a 10 salinity tank and a 30 salinity tank. The oysters had significantly different expression (DE) for 9,921 transcriptome reference sequences (reftigs; 23.6%). The number of reftigs with DE in response to salinity treatments was 7,936 (18.9%), demonstrating extensive plasticity in gene expression in order to maintain homeostasis. A total of 5,858 reftigs (13.9%) were differentially expressed with respect to the reef by treatment interaction factor, indicative of a genotype by environment (GxE) response. Differential expression in response to treatment effects was eight times more frequent in oysters from the

3 This paper is currently in preparation with a plan for submission to *Molecular Ecology* with the authors of LE Eierman & MP Hare

high salinity reef compared to those from the low salinity reef. Contrasting the two treatment environments, significant reef effects were fifteen times more frequent after acclimated to low salinity compared with high salinity. These two asymmetries reinforce the prevalence of GxE effects. From a total of 79,660 SNPs, those with inter-reef Fst in the top 1% of the distribution ranged in Fst from 0.29 to 0.73. Reftigs containing these "outlier" SNPs were significantly enriched for annotations relating to free amino acid metabolism, a major mechanism for osmoregulation in marine invertebrates. Overall, we conclude that the reef-specific patterns of gene expression indicate that oyster responses to habitat heterogeneity are shaped both by phenotypic plasticity and recurrent selection on each generation of migrants. The presence of outlier SNPs in osmoregulatory genes predicted to be under differential viability selection across the estuarine salinity gradient further supports this conclusion.

Introduction

The mechanisms by which populations respond to environmental variation are influenced by both phenotypic plasticity and functional genetic variation. Often, the relative spatial and temporal scale of environmental heterogeneity determines the relative importance of plasticity or adaptive genetic differentiation (Levins 1968, Baythavong 2011). When the environmental gradient is at a finer spatial scale than the movement of individuals, phenotypic plasticity is predicted to be the primary response mechanism (Bradshaw 1965, Scheiner 1993, Gomez-Mestre and Jovani 2013, Scheiner 2013). Likewise, temporal variability in the environment should favor physiological plasticity so that traits at one level track environment change and maintain homeostasis at a high level of organization (Scheiner 1993). Plastic physiological responses, while maximizing organism performance across a range of environments, have thresholds beyond which performance declines (reviewed in Hofmann and Todgham 2010). Defining the limits of a species' physiological plasticity in response to the environment is an important objective to better define its niche. More specifically, in light of rapid anthropogenic environmental change, identifying the niche margin with respect to plasticity may be a useful first step for investigating the environmental and demographic contexts under which adaptive mechanisms become increasingly important for the evolution of highly plastic traits (Hofmann and Todgham 2010, Nicotra et al. 2010, Evans and Hofmann 2012).

In addition to plasticity, genetic variation within a species provides the capacity for an evolutionary response to a changing environment. Genetic variation can generate different phenotypically plastic responses (reaction norms) among populations or among genotypes within a population, leading to genotype-by-environment (GxE) interactions (Scheiner 1993, Falconer and Mackay 1996). Whether at the level of a population or among genotypes, variation in

reaction norms for a physiological response is likely to result in different environmental limits of plasticity. These genetic-based differences provide variation on which selection can act and by which the patterns and ranges of plasticity can evolve (Pigliucci 2005). As the environment changes, a reservoir of standing genetic variation facilitates the persistence of a species through the differential survival and reproduction of genotypes best suited for the new conditions. On the other hand, the factors that historically constrained the niche margin, and other potential factors, may limit evolutionary responses (Hoffmann and Willi 2008). To assess the likelihood of a species' or population's persistence in a changing environment, both the consequences of selection and the modulating effect of phenotypic plasticity must be considered (Reed et al. 2011).

A fundamental distinction in the way organisms respond to environmental variation hinges on movement; physiological plasticity and its limits are particularly important for sedentary and sessile organisms such as plants and many benthic marine species. The "elm-oyster" model proposed by George Williams (1975) recognized that for sexually-reproducing sedentary organisms, high fecundity and early mortality of offspring favors abundant individual genetic variation in fitness because of the uncertain and variable conditions experienced by a cohort during dispersal and after settlement. Species' ranges often cover a range of conditions such that few of the many dispersing genotypes have high fitness in the environment where they (more or less) randomly settle. This spatial dynamic sets up local selective sieves through which many genotypes are filtered. More recently this dynamic has been described as a phenotype-environment mismatch, where migrants die shortly after settlement and reduce connectivity (Marshall et al. 2010). Recent studies with Scots pines (*Pinus sylvestris*) highlight the importance of investigating both phenotypic plasticity and genetic variation in light of climate

change (Berg and Ellers 2010, Savolainen et al. 2011, Alberto et al. 2013), with drought resistance exemplifying a complex phenotype shaped both by plasticity in seedling traits and genotypic variation in resistance for any particular environment (Richter et al. 2012).

The majority of benthic marine species, both invertebrates and fish, are sedentary or sessile as adults and disperse primarily as larvae with weak swimming abilities relative to tides and currents. Many of these species are also highly fecund and experience high mortality in the larval and early post-settlement stages (Morgan 1995). Thus, for a huge proportion of marine biodiversity, their life history implies an evolutionary dynamic – recurrent selection every generation through micro-environmental selective sieves – quite distinct from classical intergenerational natural selection and phenotypic plasticity models (Raubenheimer et al. 2012, Lande 2009, Carroll et al. 2007). A fundamental prediction of this recurrent selection model, assuming the simplest case of complete population mixing during larval dispersal, is that local differential viability selection will produce fine scale functional genetic variation among adults. Single gene examples of this have been well studied in marine invertebrates including acorn barnacles (*Semibalanus balanoides*) (Schmidt and Rand 2001, Schmidt and Rand 1999, Schmidt et al. 2000) and blue mussels (*Mytilus edulis*) (Koehn et al. 1976, Koehn et al. 1980, Koehn and Hilbish 1987). These examples describe single gene products contributing to physiological homeostasis across narrow environmental gradients relative to larval dispersal distances, and illustrate single generation evolutionary responses beyond the limits of plasticity. The persistent conundrum about these examples has been the lack of geographic reproducibility (Schmidt et al. 2008). In both these species the alleles found to be favored by certain environments in one geographic context were not favored in similar environmental gradients elsewhere (Väinölä and Hvilsom 1991, Rand et al. 2002). This difference implies that organismal responses involve

genetically complex traits and that the interplay between plastic and genetic responses will be a complex function of the particular genotypic diversity going into the local sieve. One way to begin testing these hypotheses is to expand the analysis to a genomic scale.

Several aspects of the eastern oyster (*Crassostrea virginica*) make them particularly interesting subjects for testing the degree to which a physiologically plastic trait experiences spatially adaptive divergence at scales smaller than average dispersal distance. Eastern oysters have high genetic diversity (Eierman and Hare in press; Zhang et al. 2014), so females spawning millions of eggs have the potential to test many genotypes against unpredictable environmental conditions in the next generation. Additionally, eastern oysters display a high degree of phenotypic plasticity observed in traits ranging from morphology to physiology. Oyster shell morphology is a highly plastic trait, as anyone who has eaten oysters can attest, and is also responsive to environmental conditions. For example, under increased predation pressure, oysters produce thicker, heavier shells (Newell et al. 2007, Johnson and Smee 2012, Lord and Whitlatch 2012, Robinson et al. 2014). Oyster physiology is also highly plastic. As osmoconformers, osmoregulation in oysters resembles phenotypic buffering (Reusch 2014) with cell volume maintained at a constant level despite frequently changing extracellular osmolarity. At the gene expression level, the up- or down-regulation of genes in response to hyper- or hypoosmotic stress controls the movement of solutes, and therefore water, either into or out of the cell. Buffering of cell volume across varying osmotic conditions is an exquisite adaptation that emerges from the orchestrated plasticity of gene expression.

Next-generation RNA sequencing technology (RNA-seq) allows for the quantification of gene expression and genetic variation across thousands of genes without previous genomic information, a valuable aspect for studying species that lack a reference genome such as the

eastern oyster. The use of common-gardens to acclimate individual oysters from spatially and environmentally distinct reefs and to compare them at several treatment salinities allows for the identification of differential gene expression in response to osmotic pressure (phenotypic plasticity) and in interaction with reef-source (GxE). RNA-seq has the additional benefit of providing sequence information for single nucleotide polymorphism (SNP) identification. These SNPs provide gene-specific information on genetic differentiation among reefs.

In this study, we conditioned adult *C. virginica* from a downstream, high salinity oyster reef and an upstream, low salinity oyster reef in the Delaware Bay for 9-10 weeks in high and low salinity common-garden tanks at a research hatchery. Using RNA-seq, we measured adult expression across 42,072 transcriptome sequences. Our objectives were to (1) measure differential expression between low and high salinity common-garden treatments in order to identify candidate genes for osmoregulation, (2) test for differences in expression between adult reefs after acclimation in common gardens to identify reef-specific patterns of phenotypic plasticity, (3) test for SNP outliers indicating non-neutral genetic differentiation between oyster reefs, and (4) explore the distribution of CpG (occurrence of cytosine nucleotide next to a guanine nucleotide) observed vs. expected ratio within genes as an indicator of methylation and potential epigenetic mechanisms.

Materials and Methods

I. Sample Collection

Adult oyster collections in the Delaware Bay occurred on April 18, 2011. Two hundred oysters were hand collected in a shallow subtidal region of Cape Shore (39° 04.10' N, 74° 54.77' W; salinity range 20-25 based on model in Narváez et al. 2012) and two hundred were collected by dredge from the NJ Fish and Wildlife vessel *Zephyrus* from Arnolds reef (39° 23.055' N, 75°

27.002' W; salinity range 6.5-14.5ppt as reported in Bushek et al. 2012). Oysters were then conditioned in tanks at the Haskin Shellfish Research Laboratory of Rutgers University (see Appendix A).

II. Common Garden Conditioning

At the hatchery, the oysters were de-fouled and then split into two recirculating tanks so that each tank had fifty oysters from each of the two collection localities. Each 500L tank contained UV-irradiated 1mm filtered seawater with salinity either maintained at 30 or diluted with distilled freshwater to a salinity of 10. Water temperature increased from 18°C to 22°C during the conditioning period and the oysters were fed *ad libitum* as described in Eierman and Hare (2013).

The oysters were also used in a larval survival experiment detailed in Eierman and Hare (2013). Immediately after each adult oyster was shucked for strip-spawning at room temperature, a piece of gill tissue was removed and placed in RNALater® (Ambion). Within two weeks, the RNALater® solution was drained from the tubes and the samples were archived at -80°C. The resulting genetic samples were from 51 high salinity reef source oysters acclimated to high salinity (3% mortality), 48 high salinity reef source oysters acclimated to low salinity (4% mortality), 32 low salinity reef source oysters acclimated to high salinity (32% mortality), and 46 low salinity source oysters acclimated to low salinity (8% mortality).

III. RNA-seq Library Preparation and Sequencing

Six adult tissue samples were randomly chosen from each of the four experimental groups. To extract mRNA, approximately 30mg of archived tissue from each individual was ground using liquid nitrogen with a mortar and pestle. The frozen powder was then lysed in a microcentrifuge tube using the Dynabeads® mRNA DIRECT™ Purification Kit (Ambion®) and

following the manufacturer's instructions for a standard size extraction. The lysate was passed through a 21 gauge needle to shear DNA and a 4th round of washing was used to eliminate rRNA contamination. The extracted mRNA was purified using the RNeasy® MinElute® Cleanup Kit (Qiagen®) following the manufacturer's instructions for starting volumes less than 100 µl. The mRNA quality was assessed using an Agilent 2100 Bioanalyzer.

The NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England BioLabs® Inc.) was used to prepare double-stranded cDNA libraries for sequencing. The manufacturer's protocol B instructions, starting with purified mRNA, were followed. Sera-Mag™ Magnetic SpeedBeads™ (Thermo Scientific™) were used for all purification and cleaning steps. Each adult individual was individually barcoded using NEBNext® Multiplex Oligos for Illumina® (New England BioLabs® Inc.). The PCR step was replaced with KAPA Real-Time Library Amplification Kit (KAPABiosystems). Each sample was removed from the Applied Biosystems Viia7 Detection system at a consistent level of exponential amplification, requiring from 10 to 13 cycles. Library quality for each sample was assessed using an Agilent 2100 Bioanalyzer.

The barcoded samples were then multiplexed in a design aimed to get uniform read coverage for each individual and for each pool. The multiplexed libraries were sequenced on five lanes of 100 bp Hi-Seq Illumina 2000 at the Biotechnology Resource Center Genomics Facility of Cornell University.

IV. Reference Transcriptome, Read Filtering, and Mapping

Given the availability of two published transcriptomes for *C. virginica* at the time of this work (Eierman and Hare in press, Zhang et al. 2014), each developed from different life stages, we chose to combine them to gain a more comprehensive reference. From Eierman and Hare (in press), 50,736 contigs were assembled and successfully annotated from 454 sequencing of cDNA

from mixed tissue types of eight juvenile oysters from Florida lagoons south of Cape Canaveral. From Zhang et al. (2014), 66,229 contigs were assembled from Illumina sequencing of cDNA from the hemolymph, gill, digestive gland, mantle and adductor muscle of a single adult oyster.

The annotated Eierman and Hare (in press) transcriptome and complete (annotated and non-annotated) Zhang et al. (2014) transcriptome were combined using Cd-hit-est v.3 (Li and Godzik 2006). The goal was to reduce the redundancy of sequences in the two transcriptomes without reassembling from raw reads. Following the recommendations of Eierman and Hare (in press), Cd-hit-est was implemented using a sequence identity threshold of 0.8 and a word size of 5, collapsing the total 116,965 sequences into 73,220 clusters. The combined sequences were then re-annotated as described in De Wit et al. (2012), using the *Crassostrea gigas* protein set downloaded from NCBI (on 1/15/2014). The full NCBI nr database was not used because 89.2% of the annotated sequences from Eierman and Hare (in press) and 99.2% of the 48,562 sequences annotated by Zhang et al. (2014) were annotated from *C. gigas* proteins. Based on previous annotation experience (Eierman and Hare in press), using this targeted database approach dramatically reduced computational time and more exhaustive comparisons would yield few results. Reference transcriptome sequences are referred to as "reftigs" hereafter. Reftig stands for reference contig and refers to representative consensus sequences from both assembled contigs and unassembled reads (singletons).

Reads for the 24 adult samples were separated based on barcodes, clipped with fastz_clipper, and trimmed from either end up to any phred-scale quality score of <33 with fastq_quality_trimmer (FASTXToolkit). The quality and length distribution of the reads were examined via quality score boxplots and nucleotide distribution charts generated using the

Galaxy web server. Any reads less than 90bp were discarded, resulting in all reads ranging in length from 90 to 102bp.

We chose Burrows-Wheeler Aligner (Li and Durbin 2009) parameters that maximized the number of reads that mapped uniquely and the number of reference sequences with mapped reads while minimizing the number of reads that mapped to multiple reference sequences. We did this by exploring the consequences of a range of values for the maximum edit distance (8-30), maximum number of gap opens (1-6), maximum number of gap extensions (1-5), maximum edit distance in the seed (2-6), and gap open penalty (6-11). We first explored maximum edit distance in the seed with all other values constant. We then explored maximum edit distance and maximum gap opens in a full matrix. With those values set constant, we then varied the gap open penalty and maximum number of gap extensions. Based on this exploration, we chose a maximum edit distance (-n) of 20 to correspond with the clustering similarity applied to join the two published transcriptomes, and a maximum edit distance in the seed (-k) of 3. The trade off in mapping time to further increase –k was too great for the small gains in mapped reads. All other parameters were used at their default values and all 24 samples were individually mapped against the reference transcriptome.

V. Differential Expression

Differentially expressed genes were identified using the edgeR package (Robinson et al. 2010) for R (R Development Core Team, 2011). A file of read counts for each sample was generated using custom script from De Wit et al. (2012). Each sample was identified as belonging to a reef (high or low) and treatment salinity (high or low).

Read counts were normalized in order to account for differences in total read counts between samples using the trimmed means of M as implemented in edgeR. A common

dispersion and "tagwise" dispersions (variance in read counts for each reftig) were estimated in edgeR. The following GLM log-linear model was fit to each reftig:

$$\log \mu_{gi} = \mathrm{x}_i^T \beta_g + \log N_i$$

The μ_{gi} is normalized read counts per reftig (g) and sample (i) and follows a negative binomial distribution. The x_i term is a vector of covariates that specify the reef and treatment conditions of sample i, and β_g is a vector of the regression coefficients (for reef, treatment and reef by treatment interaction) for the covariates of reftig g. The N_i term is the total number of mapped reads for sample i and is used as an offset. The model for each reftig was fit using an iteratively reweighted least squares method for maximum likelihood estimation. To identify differentially expressed reftigs and the factor to which they were responding, the fitted negative binomial GLM models for each reftig were tested using a generalized linear model likelihood ratio test where each coefficient was dropped out one at a time to generate a null model and then compared to the full model (Robinson et al. 2010).

Three initial comparisons were made, one for each coefficient: (1) reef, (2) treatment, and (3) reef by treatment interaction (RxT). To further test for any reef-specific patterns of differential expression in response to treatment, we completed four contrasts, comparing (4) high vs low treatment in the high salinity reef, (5) high vs. low treatment in the low salinity reef, (6) high vs low salinity reef source oysters in the low salinity tank and (7) high vs. low salinity reef source oysters in the high salinity tank for a total of seven likelihood ratio tests. The resulting p-values from all seven sets of likelihood ratio test comparisons were corrected for false discovery rate (FDR) and reftigs with a FDR <0.05 were determined to be significantly differentially expressed.

We tested for functional enrichment of differentially expressed reftigs using custom-made functional groups from pertinent literature and our reftig annotations. First, we generated an "osmoregulatory" functional group of 1,241 candidate proteins based on the overlap between *C. gigas* osmoregulation candidates identified experimentally (Zhang et al. 2012). We then developed a second subset of critical osmoregulatory genes within the "osmoregulatory" group from 22 free amino acid (FAA) metabolism proteins studied by Meng et al. (2013) that we call the "FAA" functional group. The metabolic pathways of taurine, glycine, arginine, proline, alanine and beta-alanine for use as osmolytes in bivalve osmoregulation are well described in the literature (Meng et al. 2013, Bishop et al. 1994, Toyohara et al. 2005, Hosoi et al. 2007, Perrino and Pierce 2000a, Perrino and Pierce 2000b). This particular subset of proteins was identified by Meng et al. (2013) due to their roles in metabolic maps charted using the KEGG database and from quantitative real-time PCR validation of differential gene expression between salinity treatment groups that followed the same design as that in Zhang et al. (2012). For several of the free amino acids, the regulation of the mRNA expression for enzymes involved in their metabolic pathways is related to whether the current salinity conditions are hypo- or hyperosmotic to the cellular osmolarity. In the taurine pathway, for example, the enzymes cysteine dioxygenase (CDO), cysteine sulfinic acid decarboxylase (CSAD) and taurine transporter (TAUT) are all down-regulated in hypoosmotic conditions while both CSAD and TAUT are up-regulated in hyperosmotic conditions (Meng et al. 2013).

Another functional group, "stress," includes 112 stress proteins identified as mortality gene expression signatures prior to mass mortality events in *C. gigas* (Chaney and Gracey 2011). The "stress" functional group and the "osmoregulatory" functional group (but not the "FAA" subgroup) overlap by 30 proteins identified by both Zhang et al. (2012) in response to salinity

treatment and by Chaney and Gracey (2011) as a signature of mortality. Differentially expressed reftigs for each factor were tested for enrichment for each of these functional groups using Fisher's exact test in comparison to the complete reference transcriptome (n=42,072).

VI. SNP ID and Analysis

The Genome Analysis Toolkit (McKenna et al. 2010, DePristo et al. 2011) was used to detect SNPs, following the best practice protocol of the Broad Institute (http://www.broadinstitute.org/gatk/guide/best-practices) along with recommendations from De Wit et al. (2012). The Base Quality Score Recalibration step was omitted because we did not have known variant sites for input. A variant call set was first obtained using a phred-scale SNP quality threshold of 30 along with the filtering recommendations from the Broad institute with the exception of the cluster-window size = 10. This variant call set was used to train the Variant Quality Score Recalibration model in recalibrating a call set using a variant quality score threshold of 4 with all other settings as recommended by the Broad Institute. Genotypes from the final set of SNPs were then filtered based on a phred-scale genotype quality threshold of 20.

Next, SNPs were restricted to those for which at least 20 individuals had confident genotypes. The SNPs were then further filtered to reduce false positives caused by paralogs mapping to the same reftig. Reftigs with a SNP density greater than 0.05 SNP/bp and significantly excess heterozygosity (greater than 70%) compared to Hardy-Weinberg expectations were removed, under the hypothesis that this pattern was caused by paralogs. Initial statistics and neutrality tests on these filtered SNPs were conducted using vcftools (Danecek et al. 2011). These statistics included allele frequency, nucleotide diversity, Hardy-Weinberg equilibrium, and Tajima's D using a sliding window approach. The SNPs were then further filtered to remove SNPs with a minor allele frequency less than 0.25, as suggested by Roesti et

al. (2012). The Fst values and linkage disequilibrium for the minor allele filtered SNPs were then calculated using vcftools. To test for SNPs under selection based on Fst, we used BayeScan v.2.0 (Foll and Gaggiotti 2008) using default parameter settings. BayeScan compares two models, a neutral model and a selection model. The selection model breaks apart a locus-reef Fst coefficient into a reef component (Beta) and a locus-specific component (alpha). Departure from the neutral model is concluded when the alpha value is needed to explain the pattern of diversity observed in the locus. A reversible-jump MCMC is used to estimate the posterior probability of each model, and these probabilities are then compared via a Bayes factor to determine which model is favored over the other.

The SNPs with the highest 1% Fst values were identified and hereafter are referred to as "outliers." The reftigs containing these SNPs were tested for functional enrichment for the "osmoregulatory", "stress" and "FAA" groups previously described using a Fisher's exact test. The reftigs containing outlier SNPs were also tested for enrichment of differentially expressed reftigs by each factor (reef, treatment, and reef by treatment).

Results

I. Reference Transcriptome

The consolidation of the two transcriptomes by clustering resulted in 73,220 clusters (reftigs) with a maximum length of 19,334 bp, N50 of 433 bp, and total length of 58,225,860 bp. These clusters were annotated using the *C. gigas* protein nr database from NCBI along with Swiss-Prot and TrEMBL databases resulting in 42,072 annotated reftigs. Of these, 41,006 were annotated through the *C. gigas* database.

Each of the 41,006 reftigs was assigned a top hit from the BLASTx search, resulting in a match to a particular accession number and protein product from *C. gigas*. The annotations of the

reftigs can be discussed at two different levels. First, each reftig had a top hit to a particular accession number that represents a unique position within the *C. gigas* genome. The annotations of the 42,006 reftigs represent 16,623 unique genome positions, referred to hereafter as "genes." At the gene level, 7,892 genes were represented by a single reftig. The remaining 8,731 genes had an average of 4 reftigs per gene, possibly due to splice variants or assembly difficulties (see example in Appendix B). Second, each reftig was identified by a protein name from the catalog of names used at KEGG. Some protein names have multiple accession numbers because *C. gigas* has numerous gene duplications (Zhang et al. 2012) and because some of the protein names only identify a protein family. The annotations of the reftigs represent 12,865 different proteins. All functional enrichment studies take place at the protein level of annotation.

Non-annotated reftigs were not used as reference sequences for two reasons. First, the lack of annotation is due to an absence of similar sequences in the *C. gigas* reference genome and Swiss-Prot and TrEMBL databases. Second, non-annotated reftigs collectively had a significantly lower GC content than annotated reftigs ($t=194.85$, p-value<0.0001). This lower GC content could represent contamination by non-oyster sequences such as from microalgae or bacteria filtered in the gills of the oysters (Eierman and Hare in press). The useful portion of the transcriptome (42,072 annotated reftigs) had a maximum reftig length of 19,334 bp, N50 of 593 bp, and total length of 45,267,490 bp.

Of the 1,241 osmoregulatory candidate proteins identified in *C. gigas*, defined here as the "osmoregulatory" functional group, 1,036 were represented by 9,785 reftigs in our combined reference transcriptome. Of the 22 proteins for the "FAA" group defined in Meng et al. (2013), 15 were represented by 58 reftigs. Finally, of the 112 proteins for the "stress" functional group (Chaney and Gracey 2011), 65 were represented by 402 reftigs (Chaney and Gracey 2011).

II. Illumina Read Mapping

The Illumina sequencing resulted in an average 28,143,975 reads per barcoded individual sample. After clipping adapters and trimming for quality, an average of 18,684,557 (66.4%) reads per barcoded sample remained (range: 7.2 – 35.7 million), and these reads were mapped to the reference transcriptome. The average depth of coverage per sample after unique mapping, including zero depths, was 18.3 reads per bp. The average proportion of reads mapped uniquely was 41.3% (SD=8.1%).

III. Differential Expression

A total of 9,921 reftigs (23.6% of total, n= 42,072) were identified as being significantly differentially expressed (DE), with a FDR<0.05, based on logfold change in read counts fitted to a negative binomial model (Fig. 4.1). Examining DE reftigs by factor in a model with reef source and salinity treatment factors as well as their interaction, the smallest number of reftigs was significant for the reef factor (Reef) while the largest number was significant for the treatment factor (Treat) (Fig. 4.1). Of the 252 reftigs responding to reef, only 21 were significantly DE solely in response to reef and not in response to any other factor. In comparison, 4,039 reftigs were significantly DE solely in response to treatment, with no response to the other factors. The reef-only DE reftigs represent a classic "genotype" response in their reaction norms (Fig. 4.2A), whereas the treatment-only DE reftigs represent a classic "environment" response in their reaction norms (Fig. 4.2B). In addition, a total of 5,848 reftigs showed a degree of contrast between reef-specific reaction norms across the two salinity treatments that the reef by treatment interaction factor was significant (Fig. 4.1, Fig. 4.3A). These reftigs represent the classic "GxE" reaction norm (Fig. 4.3B) and indicate reef-specific patterns of expression after acclimatization to low and high osmotic pressures.

We further explored the treatment-only DE reftigs by comparing the treatment response in oysters from the high salinity reef to that in oysters from the low salinity reef using specific contrasts of the full interaction generalized linear models (GLM). Only 234 reftigs showed a significant treatment response in both sets of oysters (Fig. 4.4A (blue)). The reaction norms for these 234 reftigs show the average treatment response to be similar from both reefs, a classic "plastic" response to environmental variation (Fig. 4.4B). The remaining 3,805 DE reftigs that were significant only for the treatment factor in the full model, responded significantly to treatment in only one of the two reefs. This reef-specific response did not generate a GxE response, and the patterns of the reaction norms follow the normal "plastic" response, with a steep slope for reftigs identified from the reef with significant differential expression and a similar but shallower slope for those from the reef with non-significant differential expression (Figs. 4.4C and 4.4D). However, a strong asymmetry suggests a fundamental difference in how these two reefs responded to salinity (Fig. 4.4A (grey and black)). First, when only one reef showed a significant treatment effect it was the high salinity reef most of the time (88.8%). Oysters from the high salinity reef responded to salinity with significant DE in 3,379 reftigs whereas only 426 reftigs were significant from the low reef (Fig. 4.4A and Table 4.1). Reaction norm slopes were generally steeper for oysters from the high salinity reef (Fig. 4.5) and they had up-regulated and down-regulated patterns of expression for a similar number of reftigs in the high treatment compared to the low treatment (Fig. 4.4A and Table 4.1). In contrast, the response of the low reef oysters was skewed with a larger percentage of significant DE reftigs being up-regulated in response to high salinity compared to low salinity as opposed to down-regulated (Fig. 4.4A and Table 4.1).

We also quantified reef-specific responses to osmotic pressure using within-treatment contrasts between oysters from the two reef sources and found nearly 15 times more DE between reef sources in the low salinity treatment. In the low salinity treatment, 3,675 reftigs were differentially expressed between the low salinity reef source oysters and the high salinity reef source oysters. Of these DE reftigs, 2,030 reftigs (55%) were more highly expressed by the low salinity reef source oysters than by the high salinity reef source oysters. In the high salinity treatment, 252 reftigs were differentially expressed between the low salinity reef source oysters and the high salinity reef source oysters. Of these DE reftigs, 202 reftigs (80%) were more highly expressed by the low salinity reef source oysters than by the high salinity reef source oysters. The majority of the reftigs (77%) identified as significantly differentially expressed in these contrasts were previously identified as responding to the reef factor and/or the reef-by-treatment interaction term in the full model. Only 268 of these reftigs were previously identified as responding to only the treatment factor in the full model. Therefore, the reftigs identified as DE by the within-reef contrasts (see above) and the reftigs identified by the within-treatment contrasts are nearly mutually exclusive groups of reftigs.

Statistical tests for functional enrichment within groups of significantly differentially expressed reftigs were generally not significant. The only evidence of functional enrichment was for the “stress” group in reftigs differentially expressed for the treatment factor ($p=0.00023$, odds ratio=1.72). Looking at treatment-only reftigs unique to each reef, both the low reef oysters ($p=0.0086$, odds ratio=2.52) and the high reef oysters ($p=0.0097$, odds ratio=1.52) were enriched for “stress” gene expression. For the high reef, 26 reftigs had higher expression in low salinity than in high salinity, and 22 reftigs had higher expression in high salinity than in low salinity, indicating that no particular treatment caused an increase in the “stress” reftigs. In contrast, for

the low reef, 10 “stress” reftigs had higher expression in the high salinity “away” treatment than the low salinity “home” treatment, and only one had higher expression in the low salinity “home” treatment, suggesting that the high salinity treatment induced an increase in the expression of “stress” reftigs. To check whether this pattern held true across all DE reftigs for the low reef, the reef by treatment reftigs identified as part of the “stress” group were also explored. In this case, 29 reftigs followed the same pattern of higher expression in high salinity “away” treatment but a similar number of reftigs, 22, had higher expression in the low salinity “home” treatment.

No reftigs differentially expressed for the reef factor were identified as coding for “stress” proteins, and the reftigs differentially expressed for the reef by treatment factor were not significantly enriched (p=0.94). No set of reftigs were functionally enriched for the “FAA” group (reef: no reftigs, treatment: p=0.82, reef by treatment: p=0.52). No set of reftigs were functionally enriched for the “osmoregulatory” group (reef: p=0.12, treatment: p=0.43, reef by treatment: p=0.50).

All reftigs that were differentially expressed only in response to the treatment factor (4,039 reftigs) were selected to be archived as osmoregulatory candidates. We chose to archive these reftigs because they responded to the salinity treatment and were not confounded with a GxE effect, suggesting the clearest functional relationship with osmoregulation. Annotations for these reftigs included 2,807 different protein products and 405 (14%) of these proteins were also identified experimentally as osmoregulatory candidates in *C. gigas* (Zhang et al. 2012).

IV. SNPs, Fst, and Linkage Disequilibrium

A total of 1,345,639 single nucleotide polymorphisms were identified using GATK Unified Genotyper. After filtering to remove SNPs that were monomorphic for non-reference

alleles and that were high density within reftigs and high heterozygosity, suggesting mapping of paralogs, 409,763 remaining SNPs from 15,644 reftigs were genotyped for at least 20 of the 24 samples. The mean SNP density was 0.016 SNPs/bp (Fig. 4.6) and the mean minor allele frequency was 0.13 (Fig. 4.7A). The nucleotide diversity was 0.203 for the low reef, 0.199 for the high reef and 0.202 overall (Fig. 4.7B). Tajima's D averaged -0.34 for the low reef and -0.41 for the high reef, suggesting most sites were neutral or under weak purifying selection as expected within coding sequences (Fig. 4.7C).

The mean Fst value across all 409,736 SNPs was 0.0027. After filtering to remove SNPs with a minor allele frequency less than 0.25, 79,660 SNPs remained for 12,240 reftigs and the mean Fst was 0.0025 (Fig. 4.8A). In general, higher gene flow provides greater power to detect outlier loci under selection. Here, BayeScan outlier tests found no SNPs with a significantly improved model with selection included relative to the neutral hypothesis. Nonetheless, there were many highly differentiated loci. The 797 SNPs above the 99th percentile, defined as the outlier SNPs, had Fst values ranging from 0.29 to 0.73 (Fig. 4.8B).

In order to determine if outlier SNPs were disproportionately associated with significantly DE reftigs, we tested for the enrichment of DE reftigs within the set of reftigs with outlier (top 1% Fst magnitude) SNPs, with each factor and their interaction term tested separately (reef: n=9, treatment: n=2006, reef by treatment: n=922). The outlier SNP reftigs and the remainder of the SNP reftigs did not differ in frequency of DE for the reef by treatment interaction factor (p=0.82) or the treatment-only factor (p=0.20). None of the outlier SNPs was in the reef-only reftigs.

Of the 12,240 reftigs with identified SNPs, 140 were annotated as part of the "stress" functional group. The outlier SNP reftigs were not enriched for the "stress" group (p=0.56).

Furthermore, 2,170 SNP reftigs were annotated as part of the "osmoregulatory" group. The outlier SNP reftigs were not enriched for this group either (p=0.053). Twenty-one of the total SNP reftigs were identified as members of the "FAA" functional group. The outlier SNP reftigs were significantly enriched for the "FAA" group (p=0.020, odds ratio=4.38). Four of the 21 FAA reftigs were also reftigs with outlier SNPs and were the source of the significant enrichment. These reftigs were annotated as aldehyde dehydrogenase family 3 member B1, cysteine dioxygenase type 1, argininosuccinate lyase and pyrroline-5-carboxylate reductase 2. The aldehyde dehydrogenase family 3 member B1 reftig, in addition to having high Fst between the reef samples, showed significant differential expression for the reef by treatment interaction factor.

The linkage disequilibrium analysis was restricted to reftigs larger than 8000bp (n=99) (Fig. 4.9). The overall distribution of r^2 values indicates a decay in linkage disequilibrium to $r^2 <$ 0.5 for SNP comparisons more than 2000bp apart within long reftigs. Because reftigs do not account for introns the decay of LD is somewhat larger than this.

V. CpG Observed vs. Expected

The distribution of CpG observed versus expected ratio (CpG O/E) for all reftigs is bimodal, suggesting a difference in the extent of methylation among the reftigs (Fig. 4.10A). In experimental data on methylation patterns from pacific oysters and bees, a low intragenic CpG O/E was associated with a relatively high level of methylation associated with germline methylation (Roberts and Gavery 2013). In contrast, these invertebrate studies found that intragenic regions with a high CpG O/E ratio were sparsely methylated and had annotations implying inducible or tissue specific expression (Roberts and Gavery 2013). Interestingly, in this study the eastern oyster reftigs that were differentially expressed only in response to the salinity

treatment, and only for both reefs (i.e., classic plasticity) have a CpG O/E distribution skewed toward higher values (Komogorov-Smirnov, p=1.363e-13; Fig. 4.10B). All other subsets of DE loci identified above have CpG O/E distributions no different than the total distribution (data not shown).

Discussion

Our results show highly plastic gene expression by *C. virginica* after acclimation to different osmotic pressures. A total of 7,936 reftigs, representing 5,669 genes and coding for 4,818 proteins, were differentially expressed for the treatment factor. The related *C. gigas* also showed plasticity in response to osmotic pressure with a combined 1,761 genes coding for 1,241 proteins differentially expressed in pairwise comparisons of salinities ranging from 5 to 40 (Zhang et al. 2012). Similarly, Zhao et al. (2012) identified 3,480 *C. gigas* reftigs that responded to salinity. Indeed gene expression plasticity involving many genes has repeatedly been shown across taxa and environmental parameters where physiological homeostasis is maintained across environments, whether in response to temperature as with the rainbow fish *Melanotaenia dboulayi* (Smith et al. 2013) or to variation in oxygen partial pressure as with the mouse *Peromyscus maniculatus* across altitudes (Cheviron et al. 2014), as examples.

The greater number of reftigs, and therefore genes, differentially expressed in our study compared to the *C. gigas* salinity treatments in Zhang et al. (2012) may be driven by differences in experimental design and statistical approach. First, our study examined the response across two different sets of oysters collected from the wild. The *C. gigas* study used oysters purchased from an aquacultural farm. We sampled six individuals from each treatment group whereas the *C. gigas* study sampled three individuals. Aquaculture procedures typically result in a reduction of genetic diversity (Boudry et al. 2002) and potentially cause artificial selection in response to

hatchery conditions (Christie et al. 2012). This inadvertent selection and reduction in genotypic diversity may have led to fewer genes with significant expression differences in *C. gigas* across salinity treatments. Additionally, studies have shown that increasing the number of biological replicates is critical to increasing power in RNA-seq experiments (Liu et al. 2013, Auer and Doerge 2010); therefore, by doubling the number of individuals for each treatment, in comparison to the *C. gigas* study, we likely increased power to detect expression differences. Additionally, our analysis fitted a negative binomial model to the read count data, a method shown to better model the dispersion of counts mapped to each reference sequence than the Poisson model used for the *C. gigas* work (review in Rapaport et al. 2013). The program used here, edgeR, has consistently been found to provide accurate results across both low and high expression levels when tested with simulated and validated data, although it is more likely to have false positives than the very conservative DESeq package (Rapaport et al. 2013, Soneson and Delorenzi 2013, Kvam et al. 2012).

I. Reef-Specific Gene Expression

In addition to a plastic response to osmotic pressure after acclimation, the oysters collected from high and low salinity reefs in the same estuary also had opposite gene expression patterns at 5,898 reftigs, representing 4,427 genes and coding for 3,811 proteins, indicative of a significant GxE interaction. If the individuals taken from both reefs were a random sample from an overall homogeneous, though genetically diverse, population within the estuary, then no GxE or reef effect should be detected beyond that from sampling error. The large number of reftigs showing a GxE interaction after acclimation suggests that the samples collected from high and low salinity reefs represent different genotypes. While almost all genotypes were able to survive for 9 weeks (presumably maintaining cell volume homeostasis, but see below), the gene

expression patterns that accomplished this homeostasis were very different. This difference in gene expression patterns is consistent with physiologically divergent phenotypes that resulted from recurrent differential viability selection across the estuarine salinity gradient.

Oysters from high and low salinity reefs not only differed in the direction of gene expression change for a subset of genes but they also differed in the magnitude of gene expression. First, oysters from the low salinity reef responded to salinity treatments with smaller logfold changes in the differentially expressed reftigs on average, resulting in shallower slopes for reaction norms compared to the high salinity reef oysters. Second, they had a smaller number of reftigs that were significant for differential expression across treatments, in which case the treatment effect driving the significant GxE stemmed more often from the oysters collected at high salinity (e.g. Fig. 4.3B). Finally, the oysters collected from low salinity up-regulated three times as many reftigs as they down-regulated in comparisons between the high and low salinity treatments whereas the oysters collected from high salinity up-regulated and down-regulated a nearly equal number. Even without knowing the mechanisms underlying these asymmetries (see below), they add support to the conclusion that the oysters from the two reefs had different ways of being plastic in response to salinity, and this strongly suggests functional genetic differentiation.

Intriguingly, the significant differential expression between oysters from different reefs was primarily found under one osmotic environment, not symmetrically in both high and low salinity environments. In the low salinity treatment contrast, nearly 15 times as many reftigs were differentially expressed compared to the high salinity contrast between reef sources (3,675 versus 252 reftigs). A possible explanation is that the oysters from the two reef sources respond to hyperosmotic conditions through similar processes but respond to hypo-osmotic conditions in

distinct ways, resulting in reef-specific expression patterns primarily in the low salinity tank (but see below for potential confounding factor).

After exploring the GxE patterns in terms of differential expression between treatments for a single reef source as well as differential expression between reefs within a single treatment, we are left with two tantalizing asymmetric patterns. The response between treatments driving the significant GxE stemmed mostly from the high salinity reef source oysters whereas the genotypic effects between oyster sources were mostly seen in the low salinity treatment. Further investigation is needed to understand the mechanisms generating the asymmetric expression patterns observed in distinct groups of reftigs as well as the functional impact of these patterns.

One possible confounding factor influencing some of the asymmetric patterns is the differential mortality that occurred in the tanks during acclimation, particularly the 32% mortality in the low salinity reef source oysters acclimated in the high salinity tank. While the consequence of this mortality on our expression results is difficult to predict, a narrowed focus on only the results that do not include this group of oysters reaffirms the overall conclusions. From the high salinity reef source contrast between salinity treatments, 3,379 reftigs were differentially expressed. There was little difference in mortality between the two groups of high salinity reef source oysters acclimated to the two common gardens. Thus, the between-treatment contrast that has the least potential influence from tank mortality showed a strong treatment effect, supporting the conclusion of highly plastic gene expression in response to osmotic pressure. From the contrast between reefs in the low salinity treatment, where there was little difference in mortality between the two groups of oysters, 3,695 reftigs were significantly differentially expressed between the two reef sources. Thus, the between-reef contrast that has the least potential influence from tank mortality showed a strong reef-specific effect and

therefore supports our interpretation of recurrent viability selection shaping reef-specific patterns of differential expression.

Natural selection may have acted to generate these GxE patterns in gene expression through several different, but not mutually exclusive processes. First, osmoregulation is a complex response that involves numerous metabolic pathways (Meng et al. 2013). As osmoconformers, oysters maintain isoosmostic condition by accumulating or releasing solutes (Evans 2009). Oysters primarily use organic osmolytes such as free amino acids (FAA) and quaternary amines but can also use inorganic ions such as N^+, K^+ and Ca^{++} to maintain cell volume. The metabolism of the major free amino acids, such as taurine and glycine, can occur through multiple mechanisms. Glycine, for example, can be degraded to ammonia and CO_2 through a glycine cleavage system that is catalyzed by glycine dehydrogenase and aminomethyltransferase. It can also be converted to serine in a reversible reaction catalyzed by serine hydroxymethyl transferase. Reftigs from four key enzymes of the taurine, proline, arginine, and beta-alanine metabolic pathways identified by Meng et al. (2013), had outlier SNPs. Further investigation into whether these SNPs are nonsynonymous changes may strengthen the inference of selection.

Additionally, the GxE interactions may be the result of pleiotropy. First, 66% of the osmoregulatory candidates identified from Zhang et al. (2012) were also differentially expressed in response to other independently applied stressors, including air exposure, temperature, and heavy metals. With this overlap in response, it is likely that a large subset of osmoregulatory genes in oysters are pleiotropic. For oysters studied here, the selective agent in Delaware Bay that resulted in reef-specific differences in gene expression may include variables that co-vary with salinity. In the Delaware Bay, the temperature of the water varies spatially with warmer

waters upstream and colder waters downstream (Narváez et al. 2012). Thus, using field collected oysters, it is impossible to distinguish the causative agent of the functional genetic differentiation.

Finally, the observed differences in gene expression patterns may be the result of a general stress response. As noted above, two thirds of the osmoregulatory genes identified by Zhang et al. (2012) also responded to other environmental pressures. In a different study by Chaney and Gracey (2011), 402 reftigs were identified as "stress" genes based on a pre-mortality gene expression signature identified in *C. gigas*. For the GxE reftigs observed here, the two sets of oysters did not increase the expression of "stress" reftigs more so in one treatment than the other. However, examining the treatment-only DE reftigs unique to each reef, the low salinity oysters did increase the expression of a subset of "stress" reftigs in response to the high salinity treatment. This response is not surprising given that the low salinity reef oysters showed a general pattern of higher expression in high salinity than in low salinity, but this pattern may also reflect an overall stress response to high salinity.

Different patterns of gene expression plasticity in oysters from low and high salinity reefs may also relate to larval dispersal patterns. Models of larval dispersal in the Delaware Bay (Narváez et al. 2012) and Chesapeake Bay (North et al. 2010) predict an asymmetry in the movement of larvae such that the upstream regions get few larvae migrating from downstream reefs while the downstream regions get migrants from a more well-mixed pool. As the result of a low immigration rate and high self-recruitment, genotypes in the low salinity reef may have experienced a lower average immigration rate and some measure of trans-generationl adaptation to hypoosmotic conditions. In contrast, the genotypes settling in the downstream reaches come from many source populations, perhaps accumulating a greater amount of functional genetic

diversity, and selection under such high gene flow will only spatially alter allele frequencies within generations, mostly after settlement. Based on a Delaware Bay larval dispersal model (Narváez et al. 2012), larval movements from the low salinity reef used in our study is ~10% upstream dispersal, ~50% downstream dispersal and ~10% self-recruitment, with the remaining 30% predicted to be "unsuccessful". In contrast, dispersal estimates for the lower Bay regions (high salinity) are ~1% for upstream dispersal and 94% for self-recruitment, with ~5% "unsuccessful" (Narváez et al. 2012). It will be important to model plausible dispersal asymmetries under different quantitative trait models for plasticity to explore the feasibility of a plasticity gradient.

II. Genetic and Epigenetic Mechanisms

The genetic differences underlying reef-specific plasticity are potentially caused by both genetic and epigenetic mechanisms. If genetic, then the life history and dispersal biology of this species implicates strong differential viability selection. However, out of 79,660 conservatively filtered SNPs in these transcribed sequences, we did not find a single one for which neutral evolution could be rejected. In fact, the Fst distribution was skewed towards negative values (Fig. 4.8A). However, Weir & Cockerham's Fst value will often be negative when the true value is near zero and/or when the intraclass correlation is negative (Weir 1990, Cockerham 1973). Intraclass correlation is typically negative when individuals avoid self-mating and do not mate with closely-related kin, when there is high gene flow preventing the isolation of inbreeding subpopulations and when there is no fitness cost to heterozygosity (Cockerham 1973). Given the demography of oysters, a large proportion of negative Fst values is expected. With only two populations (reefs), 12 individuals per reef and a mean Fst of 0.0025, we did not have enough power to detect selection (Foll and Gaggiotti 2008), despite Fst values ranging from 0.29 to 0.73

in the highest 1% of the distribution ("outliers"). However, there were 107 reftigs that contained more than one outlier SNP. Work is in progress to estimate the differentiation of haplotype frequencies in these reftigs and a set of reftigs from housekeeping genes and test them against a null model for drift variance across loci. While the total set of 627 reftigs with outlier SNPs was not enriched for osmoregulatory candidates, either from *C. gigas* (Zhang et al. 2012) or based on DE genes identified here, they were enriched for the small "FAA" subset identified from Meng et al. (2013). While these reftigs may provide support for selection, particularly in a few genes essential for osmoregulation, they do not clarify whether selection is the dominant mechanism generating overall observed GxE expression patterns.

Of course, coding regions are not the most likely location for regulatory sequences that could modulate gene expression patterns. Studies identifying expression QTL (eQTL), loci associated with gene expression, are typically finding the greatest support for cis-regulatory eQTL (Battle et al. 2014, Lappalainen et al. 2013). In a review by Grishkevich and Yanai (2013), the GxE interactions in gene expression patterns identified in model species were disproportionality more likely to be associated with long-distance regulatory complexes than were other patterns of gene expression. For example, in *C. elegans*, most of the genes with GxE patterns of expression were related to distant QTLs (Grishkevich and Yanai 2013). This pattern suggests that the genetic differences driving GxE patterns of expression may be located in trans-regulatory elements or distant cis-regulatory elements. Although our study provides indirect evidence for a genetic or epigenetic (below) divergence between oysters from high and low salinity reefs within a single estuary, future studies focused on genomic DNA are needed to identify targets of selection.

Trans-generational epigenetic effects could be causing reef specific patterns, especially if the low and high salinity reefs receive larvae from different source pools. Reviewed for marine systems in Reusch (2014), the effect of non-genetic carry over to the next generation can greatly impact offspring tolerance. Examples of improved offspring fitness due to non-genetic parental effects range from temperature acclimation in tropic fish (Salinas and Munch 2012) to ocean acidification in Pacific rock oysters (Parker et al. 2012). Additionally, these non-genetic effects can be transferred across multiple generations (Schmitz and Ecker 2012). Since we examined a single generation sampled directly from the wild, trans-generational epigenetic effects cannot be ruled out as the mechanism for the observed divergence in plastic gene expression patterns.

In contrast to trans-generation epigenetic effects, we found patterns of CpG variation consistent with non-inherited DNA methylation in reftigs from both reefs that had similar expression patterns across salinity treatments. The bimodal distribution of CpG observed versus expected ratios is consistent with findings in *C. gigas* by Roberts and Gavery (2012). They suggest that this distribution correlates to methylation, with low CpG observed versus expected ratios indicative of highly methylated genes. These genes are often critical to proper function and development and therefore are ubiquitously expressed. The high CpG observed versus expected ratios occur more often in genes associated with inducible expression, often related to stress and environmental response. These genes may have increased flexibility in expression due to transient methylation. In our study, the treatment-only reftigs shared by both reefs, and therefore representing the purest plasticity pattern seen in this study, have a high CpG observed to expected ratio. This result matches the prediction set forth by Roberts and Gavery (2012) and provides a tantalizing suggestion that different mechanisms of plasticity lead to different plastic response patterns in species like oysters.

III. Conclusions and Future Directions

Overall, the findings of high plasticity along with GxE effects have implications for how organisms with elm/oyster life histories will acclimate and adapt to a rapidly changing environment. The two sets of oysters examined here display divergence in their gene expression patterns after acclimation to different salinities that represent the breadth of osmotic pressure found in most estuaries. In another study (Eierman and Hare 2013), larvae from a high salinity reef showed equally high survival across salinities of 10 and 30 whereas larvae from a low salinity reef had high survival rates in a low salinity treatment but greatly reduced survival in a high salinity treatment, which supports an inference of functional divergence between oyster reefs. This divergence provides evidence for intra-generational selection that creates a patchwork of adult genotypes that may widen a species' response to environmental variation. Studies on the interaction between plasticity and environmental change typically regard selection as acting across multiple generations as populations adapt to their local conditions (e.g. Reed et al. 2011, Raubenheimer et al. 2012, Lande 2009, Carroll et al. 2007). While this model is valid for many species, we suggest that for many plant and marine benthic species intra-generational selection molds patterns of plasticity across habitats and needs to be taken into account to understand the diversity of reaction norms and their adaptive value.

Future investigations to further explore the interplay between phenotypic plastic, genetic variation and environmental change in the elm/oyster life history should include several additional components to follow up on our findings. First, the oysters reared in common gardens of different salinities should be monitored by repeated noninvasive sampling of the hemolymph. This sampling would allow for two critical measurements. First, the osmolality of the hemolymph would give a measure of the extracellular fluid salinity and therefore the internal

ionic concentration of the oyster. Second, RNA extracted from the hemolymph would provide a time series of gene expression changes during acclimation. In addition, phenotypic measurements such as condition index and respiration rate would provide a context for assessing the acclimation success of the oysters in relation to gene expression patterns and stress indices. Finally, genetic variation should be measured in cis-regulatory regions of genes in order to investigate genetic differences related to gene regulation.

Table 4.1. Treatment-only DE reftigs from each reef that are significant because of large expression differences from a single reef source. The values indicate the number of reftigs that had higher expression in the associated tank compared to the alternative tank.

	High Salinity Reef (n=3379)	Low Salinity Reef (n=426)
Low Salinity Tank	1825	98
High Salinity Tank	1554	328

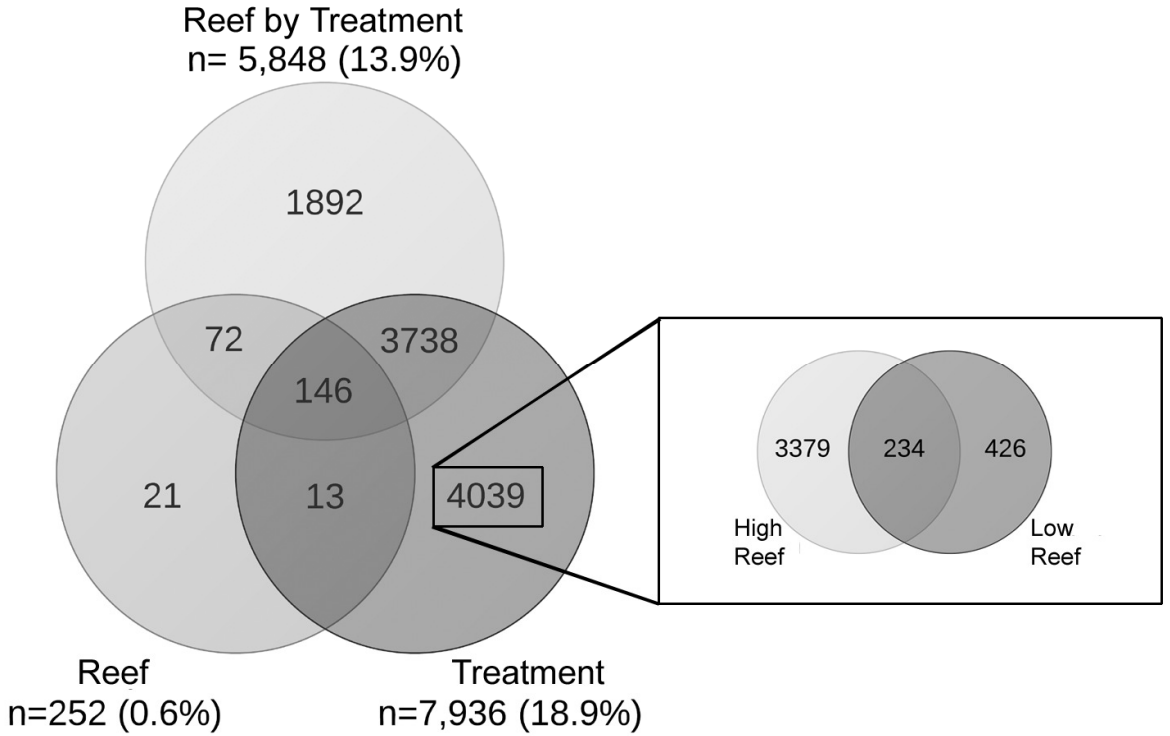

Figure 4.1. Venn diagram of the number of reftigs that were differentially expressed for each factor out of 42,072 reftigs. Overlaps indicate that the gene was differentially expressed in response to multiple factors. The center indicates that 146 reftigs were differentially expressed in response to reef source (Reef), treatment salinity (Treatment) and the interaction between reef source and treatment salinity (reef by treatment). The treatment-only reftigs were further investigated by contrasts within each reef and the boxed Venn diagram shows these results. Of these 4,039 DE reftigs, 234 were differentially expressed in response to treatment by both reefs whereas 3,379 were significant for a treatment response only in the high reef and 426 were significant only in the low reef.

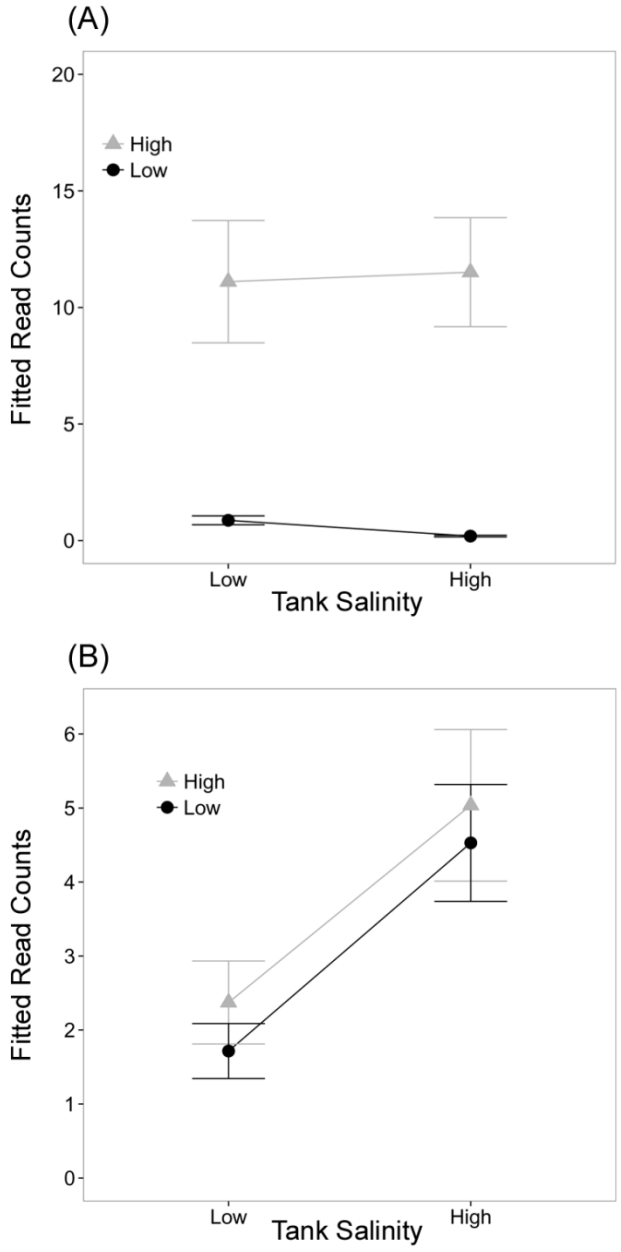

Figure 4.2. Reaction norms of read counts that have been normalized and fit to a negative binomial model. Mean read counts are plotted separately for each reef source. (A) Reef factor: Illustrative reaction norm of example reftig that is differentially expressed for the reef (Reef) factor, representing a classical "genotype" reaction norm. (B) Treatment factor: Illustrative reaction norm of example reftig that is differentially expressed for the treatment (Treat) factor, representing reaction norms with a strong "plasticity" pattern. Error bars are standard error.

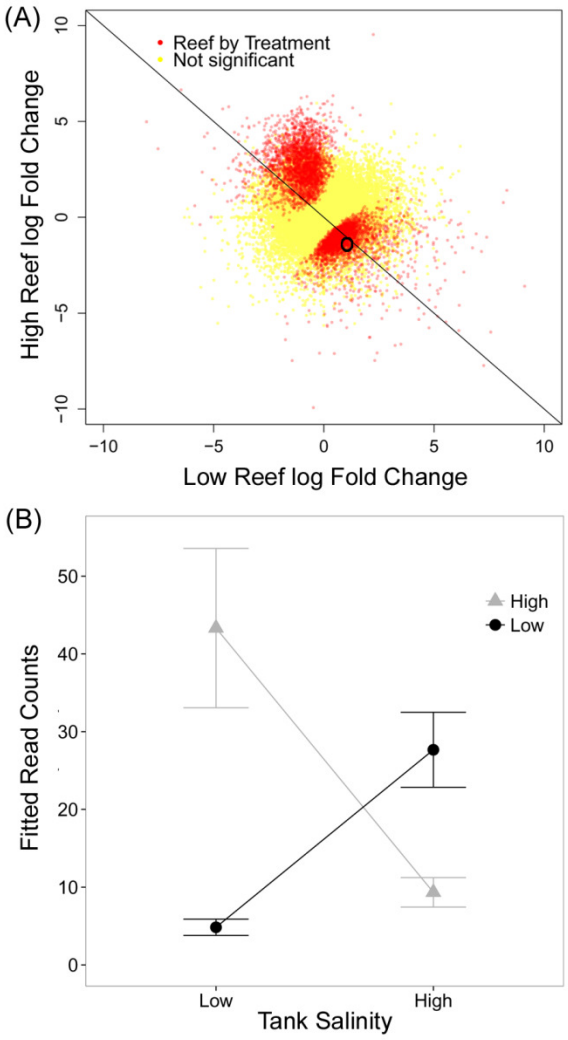

Figure 4.3. Differentially expressed reftigs responding only to the RxT factor. (A) Red points are reftigs that are responding only to RxT factor. Yellow points are reftigs that are not differentially expressed for any factor. Each point represents a single reftig. Each axis indicates the log fold change in expression in response to the high salinity common garden compared to the low salinity common garden for each reef. The black line represents the expectation for an RxT response. The reaction norms are of read counts that have been normalized and fit to a negative binomial model. Mean read counts are plotted separately for each reef source. Error bars are standard error. (B) Illustrative reaction norm of the mean read counts that have been normalized and fit to a negative binomial model. Mean read counts are plotted based separately for each reef source. Error bars are standard error. The reaction norm is for a representative reftig to illustrate the reaction norms for reftigs that are differentially expressed in response to RxT. The circled point in (A) corresponds to the same reftig used for the illustrative reaction norm in (B).

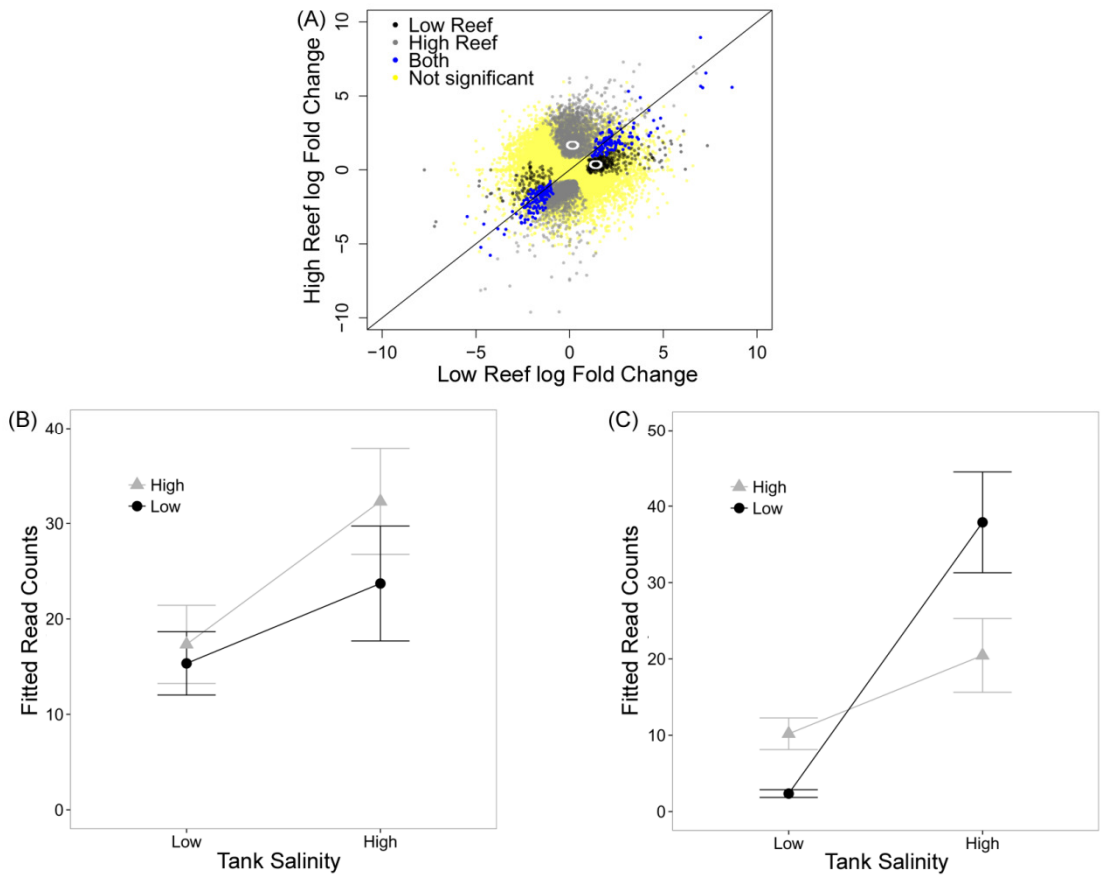

Figure 4.4. Differentially expressed reftigs responding only to the treatment factor. (A) Significantly differentially expressed reftigs in response to solely the treatment factor: those significant for both reefs in blue, significant for the low reef in black and significant for the high reef in gray. Yellow points represent reftigs that are not differentially expressed for any factor. Each point represents a single reftig. Each axis indicates the log fold change in expression in response to the high salinity common garden compared to the low salinity common garden for each reef. The black line represents the expectation for a treatment (Treat) response. The white circles each correspond to a reftig that are illustrated by reaction norms in (B) for the gray point and (C) for the black point. The reaction norms are of read counts that have been normalized and fit to a negative binomial model. Mean read counts are plotted separately for each reef source. Error bars are standard error. (B) Illustrative reaction norm representing reftigs that are differentially expressed in response to treatment for the high reef. (C) Illustrative reaction norm representing reftigs that are differentially expressed in response to treatment for the low reef. Error bars are standard error.

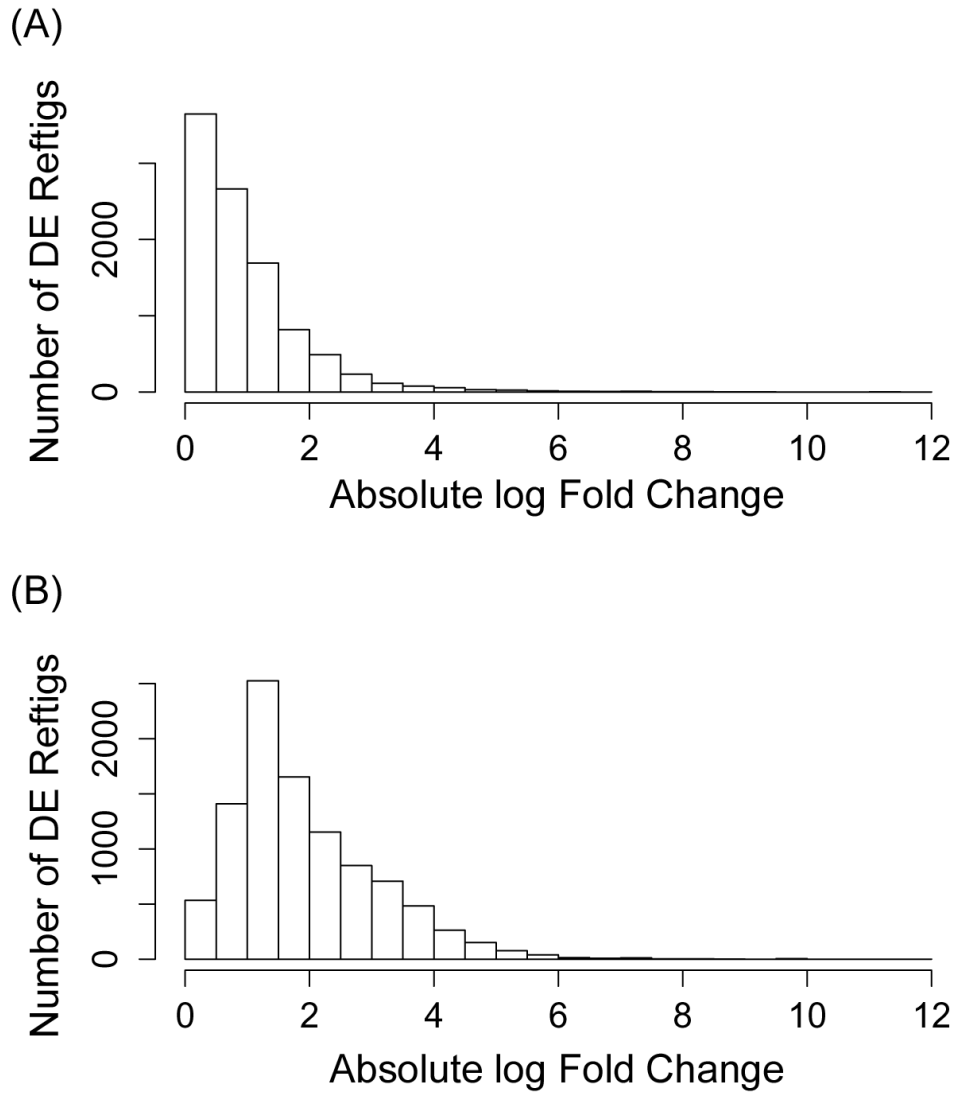

Figure 4.5. Distributions for the absolute values of the log fold changes for all differentially expressed reftigs for the treatment and reef by treatment interaction factors. The logfold change is proportional to the slope of a line in a reaction norm between the low and high treatment salinities. A steeper slope in a reaction norm is a larger logfold change. (A) The distribution for the low salinity reef oysters. (B) The distribution for the high salinity reef oysters. The two distributions show that the high salinity reef oysters have a pattern of larger logfold changes in RE reftigs than do the low salinity reef oysters.

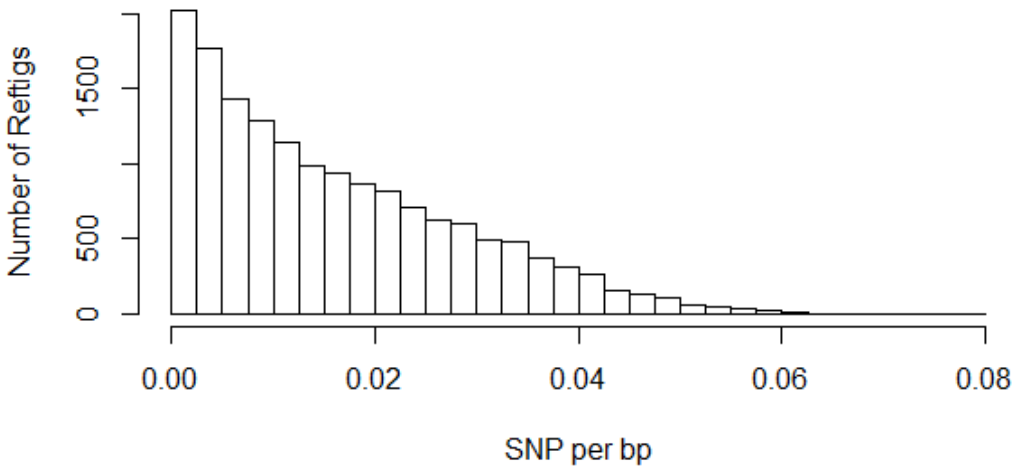

Figure 4.6. Distribution of SNP density in each reftig measured in SNP per bp.

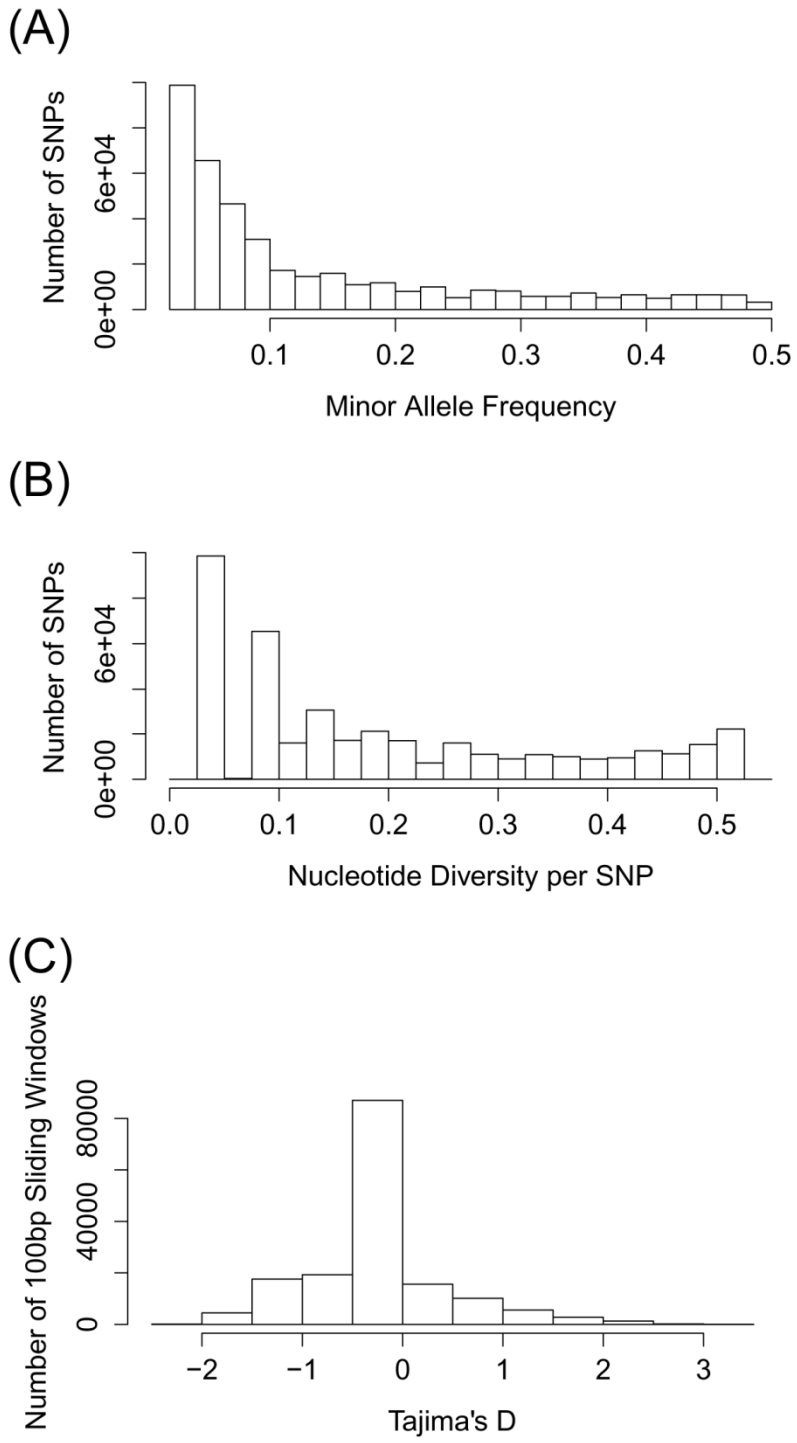

Figure 4.7. Summary information of the 409,763 SNPs genotyped for at least 20 out of the 24 individuals (A) Minor allele frequency of SNPs. (B) Nucleotide diversity of SNPs (C) Tajima's D in sliding windows for the SNPs

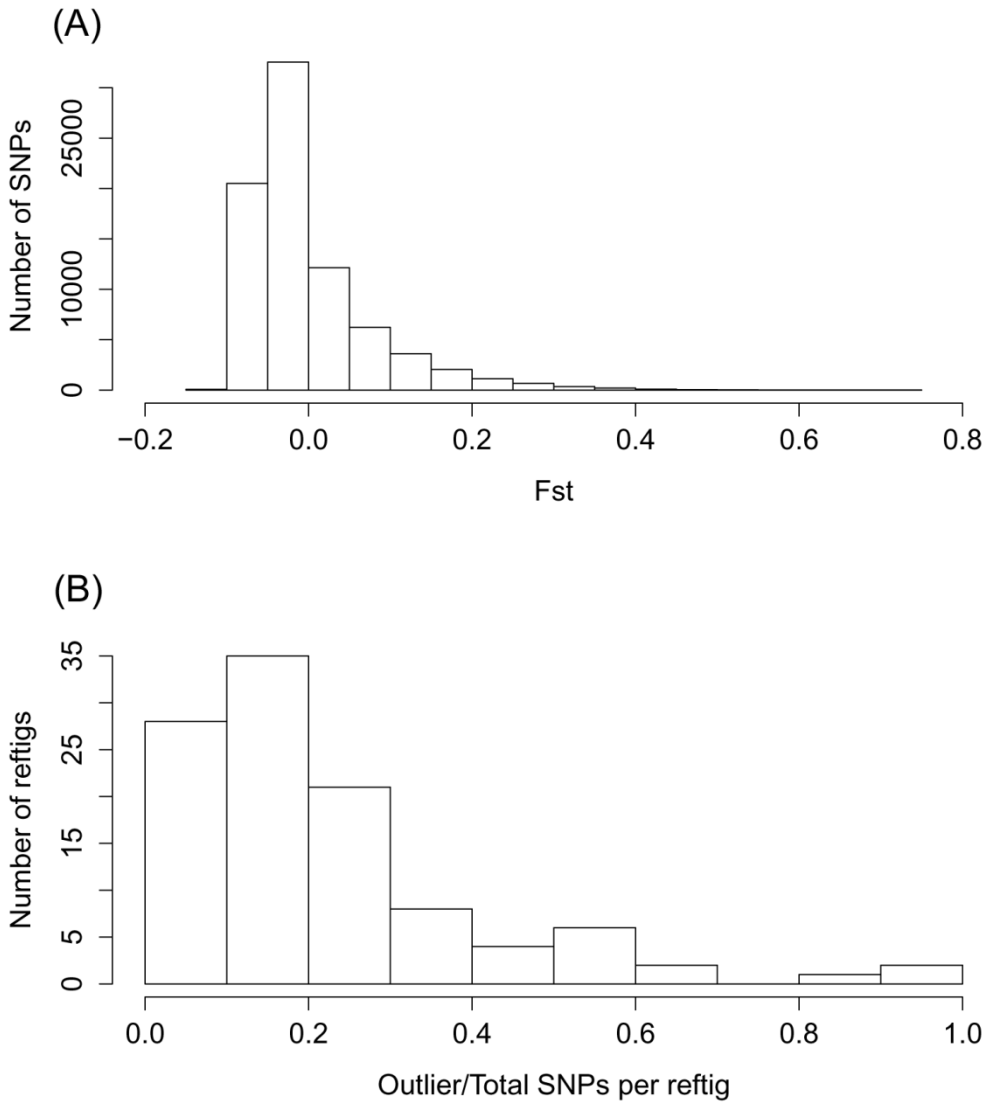

Figure 4.8. Distribution of Fst values. (A) Distribution of Fst values for each SNP after filtering for minor allele frequenices below 0.25. (B) Distribution of the ratio of outlier Fst SNPs to the total SNPs within each reftig containing more than 1 outlier SNP.

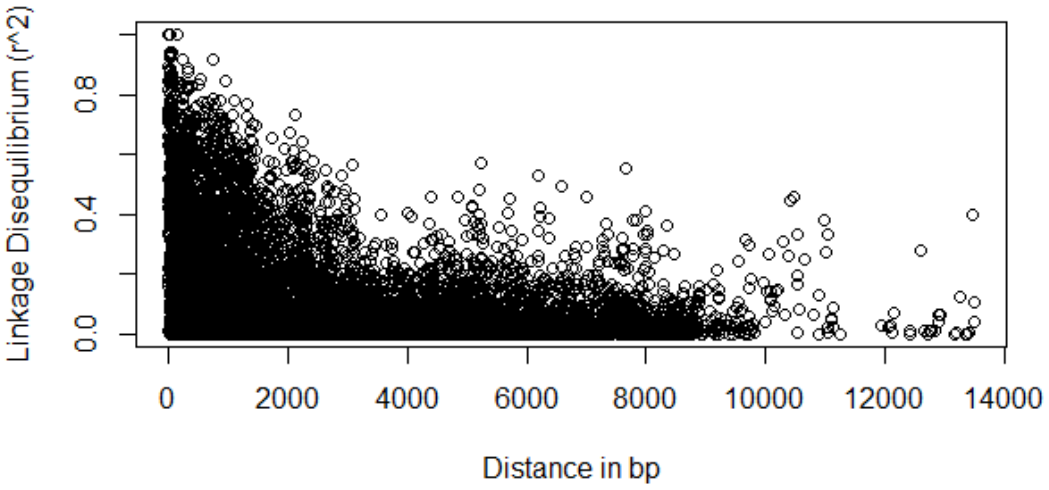

Figure 4.9. Linkage disequilibrium in pairwise SNP comparisons for 99 reftigs longer than 8000bp.

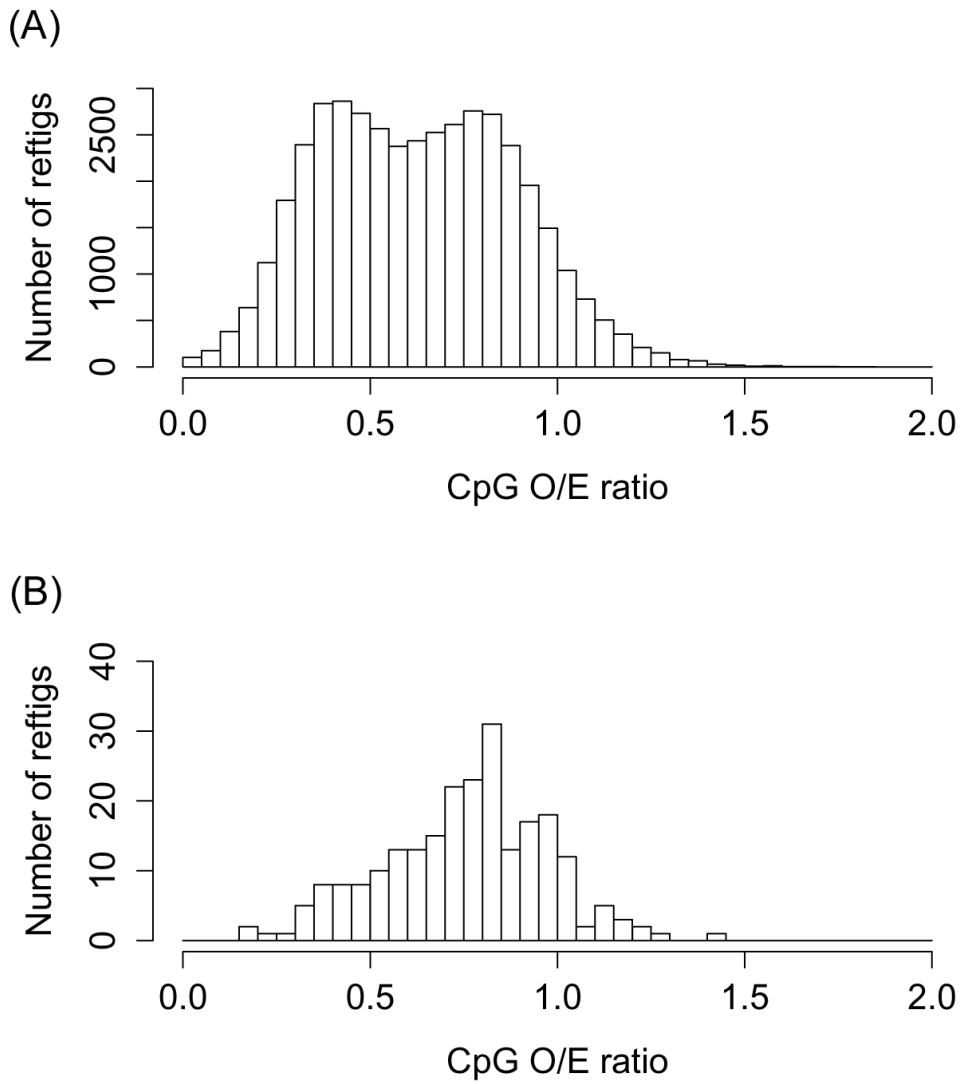

Figure 4.10. Distribution of CpG observed versus expected ratio. (A) distribution for the total 42,072 reftigs and (B) distribution for the 234 treatment-only reftigs differentially expressed by both reefs.

REFERENCES

Alberto FJ, Aitken SN, Alía R *et al.* (2013) Potential for evolutionary responses to climate change – evidence from tree population. *Global Change Biology*, 19, 1645-1661.

Auer PL, Doerge RW (2010) Statistical design and analysis of RNA sequencing data. *Genetics*, 185, 405-416.

Battle A, Mostafavi S, Zhi X *et al.* (2013) Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research*, doi:10.1101/gr.155192.113.

Baythavong BS (2011) Linking the spatial scale of environmental variation and the evolution of phenotypic plasticity: Selection favors adaptive plasticity in fine-grained environments. *The American Naturalist*, 178, 75-87.

Berg MP, Ellers J (2010) Trait plasticity in species interactions: a driving force of community dynamics. *Evolutionary Ecology*, 24, 617-629.

Bishop SH, Greenwalt DE, Kapper MA, Paynter KT, Ellis LL (1994) Metabolic regulation of proline, glycine, and alanine accumulation as intracellular osmolytes in ribbed mussel gill tissue. *J. Exp. Bio*, 268, 151-161.

Boudry P, Collet B, Cornette F, Hervouet V, Bonhomme F(2002) High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture*, 204, 283–296.

Bradshaw AD (1965) Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, 13, 115-155.

Bushek D, Ford SE, Burt I (2012) Long-term patterns of an estuarine pathogen along a salinity gradient. *Journal of Marine Research*, 70, 225-251.

Carroll SP, Hendry AP, Reznick DN, Fox CW (2007) Evolution on ecological time-scales. *Functional Ecology*, 21, 387-393.

Chaney ML, Gracey AY (2011) Mass mortality in Pacific oysters is associated with a specific gene expression signature. *Molecular Ecology*, 20, 2942-2954.

Cheviron ZA, Connaty AD, McClelland GB, Storz JF (2014) Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: Transcriptomic plasticity and thermogenic performance. *Evolution*, 68, 48-62.

Christie MR, Marine ML, French RA, Blouin MS (2012) Genetic adaptation to captivity can occur in a single generation. *Proc Natl Acad Sci USA*, 109, 238–242.

Danecek P, Auton A, Abecasis G, *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156-2158.

De Wit P, Pespeni MH, Ladner JT *et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, 12, 1058-1067.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491-498.

Eierman LE, Hare MP (2013) Survival of oyster larvae in different salinities depends on source population within an estuary. *Journal of Experimental Marine Biology and Ecology*, 449, 61-68.

Eierman LE, Hare MP (in press) Transcriptomic analysis of candidate osmoregulatory genes in the eastern oyster *Crassostrea virginica*. *BMC Genomics*.

Evans DH (2009) Osmotic and Ionic Regulation: Cells and Animals. Boca Raton: CRC Press.

Evans TG, Hofmann GE (2012) Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Phil. Trans. R. Soc. B.*, 367, 1733-1745.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, Ed. 4. Longmans Green, Harlow, Essex, UK.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian Perspective. *Genetics*, 180, 977-993.

Gomez-Mestre I, Jovani R (2013)A heuristic model on the role of plasticity in adaptive evolution: plasticity increases adaptation, population viability and genetic variation. *Proceedings of the Royal Society B*, 280, 20131869.

Grishkevich V, Yanai I (2013) The genomic determinants of genotype x environment interactions in gene expression. *Trends in Genetics*, 29, 479-487.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, 9, 421-432.

Hofmann GE, Todgham AE (2010) Living in the now: Physiological mechanisms to tolerate a rapidly changing environment. *Ann. Rev. Physiol.*, 72, 127-145.

Hosoi M, Shinzato C, Masaya T *et al.* (2007) Taurine transporter function from the giant Pacific oyster *Crassostrea gigas*: function and expression in response to hyper- and hypo-osmotic stress. *Fisheries Science*, 73, 385-394.

Johnson KD, Smee DL (2012) Size matters for risk assessment and research allocation in bivalves. *Marine Ecology Progress Series*, 462, 103-110.

Koehn RK, Milkman R, Mitton JB (1976) Population genetics of marine

pelecypods. IV. Selection, migration, and genetic differentiation in the blue mussel *Mytilus edilus*. *Evolution*, 30, 2-30.

Koehn RK, Newell RIE, Immermann F (1980) Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proc. Natl. Acad. Sci. USA*, 77, 5385-5389.

Koehn RK, Hilbish TJ (1987) The adaptive importance of genetic variation. *American Scientist*, 75, 134-141.

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.*, 99, 248-256.

Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, 22, 1435-1446.

Lappalainen T, Sammeth M, Friedländer MR *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501, 506-511.

Levins R (1968) Evolution in changing environments. Princeton University Press, Princeton, N.J.

Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 2009, 25:1754-60.

Li W, Godzik A: Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006, 22(13):1658-1659.

Liu Y, Zhou J, White KP (2013) RNA-seq differential expression studies: more sequence, or more replication? *Bioinformatics*, 30, 301-304.

Lord J, Whitlatch R (2012) Inducible defenses in the eastern oyster *Crassostrea virginica* Gmelin in response to the presence of the predatory oyster drill *Urosalpinx cinerea* Say in Long Island Sound. *Marine Biology*, 159, 1177-1182.

Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010) Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters*, 13, 128-140.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297-1303.

Meng J, Zhu Q, Zhang L, Li C *et al* (2013) Genome and transcriptome analyses provide insight into the euryhaline adaptation mechanism of *Crassostrea gigas*. *PLoS ONE*, 8, e58563. doi: 10.1371/journal.pone.0058563

Morgan SG (1995) Life and death in the plankton: larval mortality and adaptation. Pp. 279-321 L. McEdward, ed. *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, Florida (USA)

Narváez DA, Klink JM, Powell EN, Hofmann EE, Wilkin J, Haidvogel DB (2012) Modeling the dispersal of eastern oyster (*Crassostrea virginica*) larvae in Delaware Bay. *Journal of Marine Research*, 70, 381-409.

Newell RE, Kennedy V, Shaw K (2007) Comparative vulnerability to predators, and induced defense responses, of eastern oysters *Crassostrea virginica* and non-native *Crassostrea ariakensis* oysters in the Chesapeake Bay. *Marine Biology*, 152, 449-460.

Nicotra AB, Atkin OK, Bonser SP *et al.* (2010) Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, 15, 684-692.

North EW, King DM, Xu J *et al.* (2010) Linking optimization and ecological models in a decision support tool for oyster restoration and management. *Ecological Applications*, 20, 851–866.

Parker LM, Ross PM, O'Connor WA, Borysko L, Raftos DA, Pörtner HO (2012) Adult exposure influences offspring response to ocean acidification in oysters. *Global Change Biology*, 18, 82-92.

Perrino LA, Pierce, SK (2000a) Choline dehydrogenase kinetics contribute to glycine betaine regulation differences in Chesapeake Bay and Atlantic oysters. *J. Exp. Bio*, 286, 250-261.

Perrino LA, Pierce SK (2000b) Betaine aldehyde dehydrogenase kinetics partially account for oyster population differences in glycine betaine synthesis. *J. Exp. Bio*, 286, 238-249.

Pierce SK, Rowland-Faux LM, O'Brien SM (1992) Different salinity tolerance mechanisms in Atlantic and Chesapeake Bay conspecific oysters: glycine betaine and amino acid pool variations. *Marine Biology*, 113, 107-115.

Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology and Evolution*, 20, 481-486.

R Development Core Team 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rand DM, Spaeth PS, Sackton TB, Schmidt PS (2002) Ecological genetics of Mpi and Gpi polymorphisms in the acorn barnacle and the spatial scale of neutral and non-neutral variation. *Integrative and Comparative Biology*, 42, 825-836.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D (2013) *Genome Biology*, 14, R95.

Raubenheimer D, Simpson SJ, Tait AH (2012) Match and mismatch: conservation physiology, nutritional ecology and the timescales of biological adaptation. *Phil. Trans. R. Soc. B*, 367, 1628-1646.

Reed TE, Schindler DE, Waples RS (2011) Interacting effects of phenotypic plasticity and evolution on population persistence in a changing climate. *Conservation Biology*, 25, 56-63.

Reusch TBH (2014) Climate change in the oceans: evolutionary versus phenotypically plastic responses of marine animals and plants. *Evolutionary Applications*, 7, 104-122.

Richter S, Kipfer T, Wohlgemuth TW, Guerrero CC, Ghazoul J, Moser B (2012) Phenotypic plasticity facilitates resistance to climate change in a highly variable environment. *Oecologia*, 169, 269-279.

Roberts SB, Gavery, MR (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in Physiology*, 2, 116.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital expression data. *Bioinformatics*, 26, 139-140.

Robinson EM, Lunt J, Marshall CD, Smee DL (2014) Eastern oysters *Crassostrea virginica* deer crab predators by altering their morphology in response to crab cues. *Molecular Ecology*, 20, 111-118.

Salinas S, Munch SB (2012) Thermal legacies: transgenerational effects of temperature on growth in a vertebrate. *Ecology Letters*, 15, 159-163.

Savolainen O, Kujala ST, Sokol C *et al.* (2011) Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *Journal of Heredity*, 102, 526-536.

Scheiner SM (1993) Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics*, 24, 35-68.

Scheiner SM (2013) The genetics of phenotypic plasticity. XII. Temporal and spatial heterogeneity. *Ecology and Evolution*, 3, 4596-4609.

Schmidt PS, Bertness MD, Rand DM (2000) Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proc. R. Soc. Lond. B*, 267, 379-384.

Schmidt PS, Rand DM (2001) Intertidal microhabitat and selection at *MPI*: Interlocus contrasts in the Northern acorn barnacle, *Semibalanus balanoides*. *Evolution*, 53, 135-146.

Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of *MPI* genotypes. *Evolution*, 55, 1336-1344.

Schmidt PS, Serrão EA, Pearson GA *et al.* (2008) Ecological genetics in the North Atlantic environmental gradients and adaptation at specific loci. *Ecology*, 89, S91-S107.

Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends in Plant Science*, 17,149–154.

Smith S, Bernatchez L, Beheregaray LB (2013) RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics*, 14, 375.

Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 91.

Toyohora H, Yoshida M, Hosoi M, Hayashi I (2005) Expression of taurine transporter in response to hypo-osmotic stress in the mantle of Mediterranean blue mussel. *Fisheries Science*, 71, 356-360.

Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biol. J. Linn. Soc.*, 43, 127-148.

Williams GC (1975) Sex and Evolution. Princeton University Press, Princeton, N.J.

Zhang G, Fang X, Guo X *et al.* (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490, 49-54.

Zhang L, Li L, Zhu Y, Zhang G, Guo X (2014) Transcriptome analysis reveals a rich gene set related to innate immunity in the eastern oyster (*Crassostrea virginica*). *Marine Biotechnology*, 16, 17-33.

Zhao X, Yu H, Kong L, Li Q (2012) Transcriptomic responses to salinity stress in the Pacific Oyster *Crassostrea gigas*. *PLOS One*, 7, e46244.

APPENDIX A

OYSTER FIELD COLLECTION AND BROODSTOCK CONDITIONING/ACCLIMATION

Field Collection

I collected oysters from the Delaware Bay on April 17 and April 18, 2011. A large rainstorm on April 16 and continuing rain showers on the 17th and 18th had dropped the salinity throughout the Delaware Bay from the normal salinity ranges. The high salinity site was intertidal at Cape Shore, and 200 oysters were collected by hand on April 17 (Table A.1). The oysters were collected during the lowest point of a Spring low tide. Tidal amplitude at the location is approximately 0.6m. The Cape Shore oysters do not form a continuous bed and are not harvested. This location has shell bags set out to promote wild recruitment, and the oysters were collected off of these artificial reef structures. Most of the reef is exposed during low tide. I collected oysters that were still submerged below the water. Both the low salinity site and intermediate salinity site were collected off of the New Jersey Fish and Wildlife vessel *Zephyr*. The Arnolds oyster bed was the low salinity site, and 200 live oysters were collected from a single dredge (Table A.1). The New Beds oyster bed was the intermediate salinity site, and three dredges were required to collect 200 living oysters (Table A.1).

Outplant Conditions

On April 19, 2011, all of the field collected oysters were temporarily outplanted for 3 weeks at one of two field sites. The hatchery tanks could not be set-up until May 9 so oysters were maintained in the wild until then. Oyster from each reef source were outplanted at the field site that most closely matched their reef source salinity regime. The Arnolds low salinity reef source oysters and the New Beds intermediate salinity reef source oysters were outplanted at the low salinity outplant site in the Cohansey River (39° 22.75' N, 75° 21.32' W). At the time of

outplant, the salinity was 3.1 and the temperature was 11.9°C. Oysters were outplanted in mesh bags and hung from a dock. The bottoms of the bags were ~0.5m above the bottom of the river. The bags were completely submerged during all tidal states. The Cape Shore high salinity reef source oysters were outplanted at the high salinity outplant site in the Cape May harbor (38° 56.73' N, 74° 53.98' W). At the time of outplant, the salinity was 26.6 and the temperature was 12.6°C. Oysters were outplanted in mesh cages and secured to metal racks that were ~1m above the bottom of the harbor and were completely submerged during all tidal states.

On May 10, 2011, the outplanted oysters were collected and redistributed. One hundred oysters from each reef source (high, intermediate and low) were removed from the field locations and brought to Haskin Shellfish Research Laboratory of Rutgers University. The remaining 100 oysters from each reef source were then divided so that 50 oysters from each reef were outplanted in each of the two field outplant locations. These oysters remained at the field outplant sites for acclimation and conditioning in natural waters. At the time of retrieval on May 10, 98% of the original oysters were still alive. The low salinity site had a salinity of 7.1 and a temperature of 16.1°C during retrieval. The high salinity site had a salinity of 28.2 and a temperature of 17.2°C during retrieval.

Field-conditioned oysters were subsampled using 1-2 oysters per reef source per week from June 5 until July 4 to check for gonad condition. Once ripe, oysters were brought into the hatchery for spawning. The high salinity field-conditioned oysters were removed on June 18, 2011. The salinity was 28.4 and the temperature was 20.3°C. The low salinity field-conditioned oysters were removed on July 4, 2011. The salinity was 10.1 and the temperature was 19.1°C.

Tank Conditions

On May 10, 100 oysters from the high salinity reef source were recollected from the Cape May outplant site. On the same day, 100 oysters from the low salinity reef source and 100 oysters from the intermediate salinity reef source were recollected from the Cohansey River outplant site. These oysters were brought to the Haskin Shellfish Research Laboratory. After being cleaned, the oysters from each reef source were divided into two groups and placed in one of two common gardens for acclimation and conditioning. One tank was maintained at a salinity of 10 and the other tank was maintained at a salinity of 30. Each tank consisted of a flow-through trough and an additional tank for the recirculation of water back into the flow-through trough. Each flow-through trough held 225L of water and each recirculating tank held 275L of water for a total volume of 500L of water. Each recirculating tank contained a biofilter. The high salinity tank was filled with UV-irradiated 1mm filtered seawater (salinity 30) and the low salinity tank was filled with the seawater diluted to a salinity of 10 with distilled freshwater.

Tank temperatures were controlled via two heaters per tank, one in the flow-through trough and one in the recirculating tank, and temperature was recorded during the morning feeding. The initial tank temperatures were ~18°C and the temperature was slowly increased to ~20°C to induce gametogenesis (Fig. A.1). After four weeks at ~20°C with no gonad development, the temperature was again slowly increased to ~22°C, starting on June 23, 2011 (Fig. A.1).

The oysters were fed a 2:2:1 mixture of *Pavlova lutheri*, *Chaetoceros muelleri* and *Tetraselmis chui* twice a day, with the total number of cells dependent on the number of oysters and estimated dry weight of oysters in each tank. I initially fed oysters at a 3% ration of algae dry weight to oyster dry weight. Average dry weight per oyster for each reef source was calculated

by drying the meat of two oysters from each reef source and averaging the dried weights. When I began to increase the temperature again on June 23, 2011, I also increased the feeding ration to 4%. For a slow release of algae during feeding, a bucket of the mixture was siphoned via an airline into each tank. The tanks were cleaned once a week, and a partial water change was done every third day. During a partial water change, the flow-through tank was stoppered and left completely filled. The recirculating tank was emptied, refilled and adjusted to the appropriate salinity. The temperature was adjusted with heaters to match the temperature of the flow-through tank before tank circulation was resumed.

The high salinity reef source (Cape Shore) oysters were removed from the tanks for spawning and to obtain a genetic sample on July 13, 2011. The low salinity reef source (Arnolds) and intermediate salinity reef source (New Beds) were removed from the tanks for spawning and to obtain a genetic sample on July 20, 2011.

Table A.1. Oyster field collection locations and environmental conditions at the time of collection.

Oyster Reef	Latitude	Longitude	Salinity	Temperature	Depth	Collection Method
Arnolds	39° 23.06' N	74° 27.00' W	3.6	12.1°C	~5m	Dredge
New Beds	39° 14.52' N 39° 14.67' N 39° 14.51' N	74° 15.07' W 74° 15.23' W 74° 15.14' W	10.9	12.3°C	~6m	Dredge
Cape Shore	39° 04.10' N	74° 54.77' W	18.3	12.8°C	~0.25m	By hand

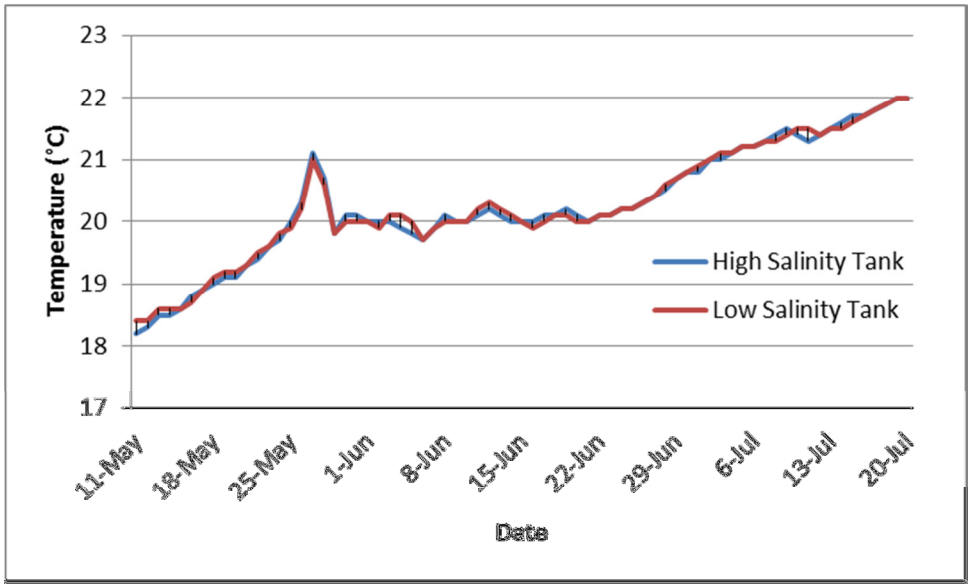

Figure A.1. Tank temperature during adult oyster acclimation and conditioning.

APPENDIX B

GENE-REFTIG CASE STUDY

Methods

Of the 16,623 genes identified in the reference transcriptome, 7,892 genes were represented by a single reftig, and the remaining 8,731 genes had an average of four reftigs per gene. In the differential expression results, a total of 3,926 genes had reftigs with conflicting results in that at least one reftig was significantly differentially expressed for a given factor and at least one reftig was not significantly differentially expressed for that same factor. An additional 42 genes had significant reftigs with opposite expression patterns as indicated by the direction of the log fold change. I completed a case study of a representative gene in order to explore the cause of these contradictory results.

I selected a gene with 4 representative reftigs, in accordance with the average number of reftigs per gene. The reftigs were annotated as coding for the protein MBT domain-containing protein 1, found on scaffold 1723 of the *C. gigas* genome. One reftig was significantly differentially expressed for the reef by treatment interaction and the remaining three reftigs were not. I mapped these four reftigs and the *C. gigas* coding sequence to the complete *C. gigas* genome using GMAP (Wu and Watanabe 2005). I then examined the alignment of exons in the sample reftigs to the *C. gigas* coding sequence to determine the positional relationship between reftigs sharing an annotation. I used the fitted read counts from the negative binomial GLM model to create reaction norms for each of the four reftigs in order to characterize and compare the reftig expression patterns.

Results

The four reftigs were successfully mapped to the correct scaffold of the *C. gigas* genome, and each reftig shared exons with the *C. gigas* coding sequence. However, two of the reftigs included adjacent regions of the *C. gigas* genome that were not identified as part of the *C. gigas* coding sequence for MBT domain-containing protein 1 (Fig. B.1). Additionally, two of the reftigs had exons that differed in length compared to those in the *C. gigas* model (Fig. B.1). These patterns suggest that the reftigs may represent alternative splice variants and/or partial transcripts. The patterns of expression are similar for all four reftigs (Fig. B.2) although only Reftig 4 (Fig. B.1D and Fig. B.2D) was significantly differentially expressed. The four reftigs differ in the relative variance of expression levels, which may affect the power to detect differential expression. These results suggest that non-significant reftigs that have the same annotation as reftigs that are significant for differential expression may be due to reduced power from high variance in expression across individuals.

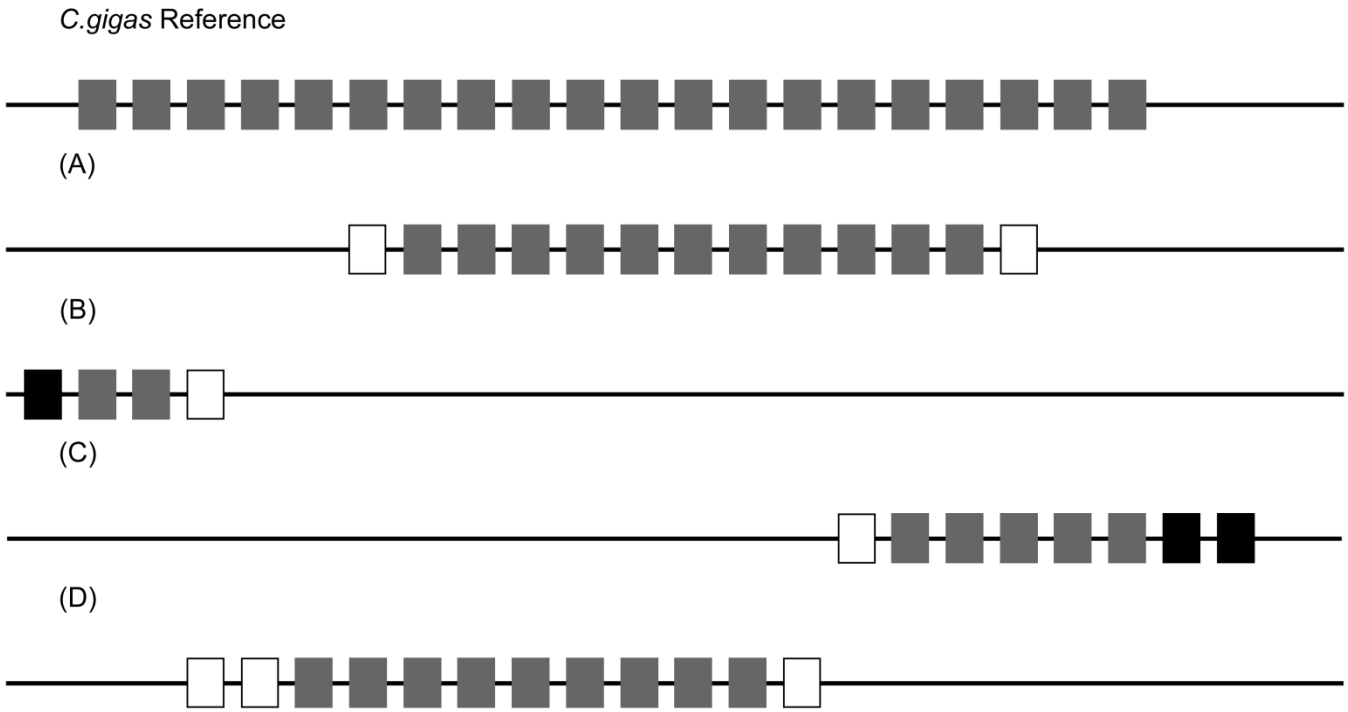

Figure B.1. Comparison of *C. gigas* reference gene and four reftigs annotated from the gene. The reftigs were mapped to the *C. gigas* genome using GMAP. The reference gene is MBT domain-containing protein 1 [Crassostrea gigas]. The rectangles represent exons. (A-D) represent the four reftigs mapped to the *C. gigas* genome. Grey rectangles are exons that are the same length as the reference exon. White rectangles are exons that differ in length from the reference exon. Black rectangles are exons that are not identified as part of the reference gene but that align to the neighboring stretch of sequence in the *C. gigas* genome.

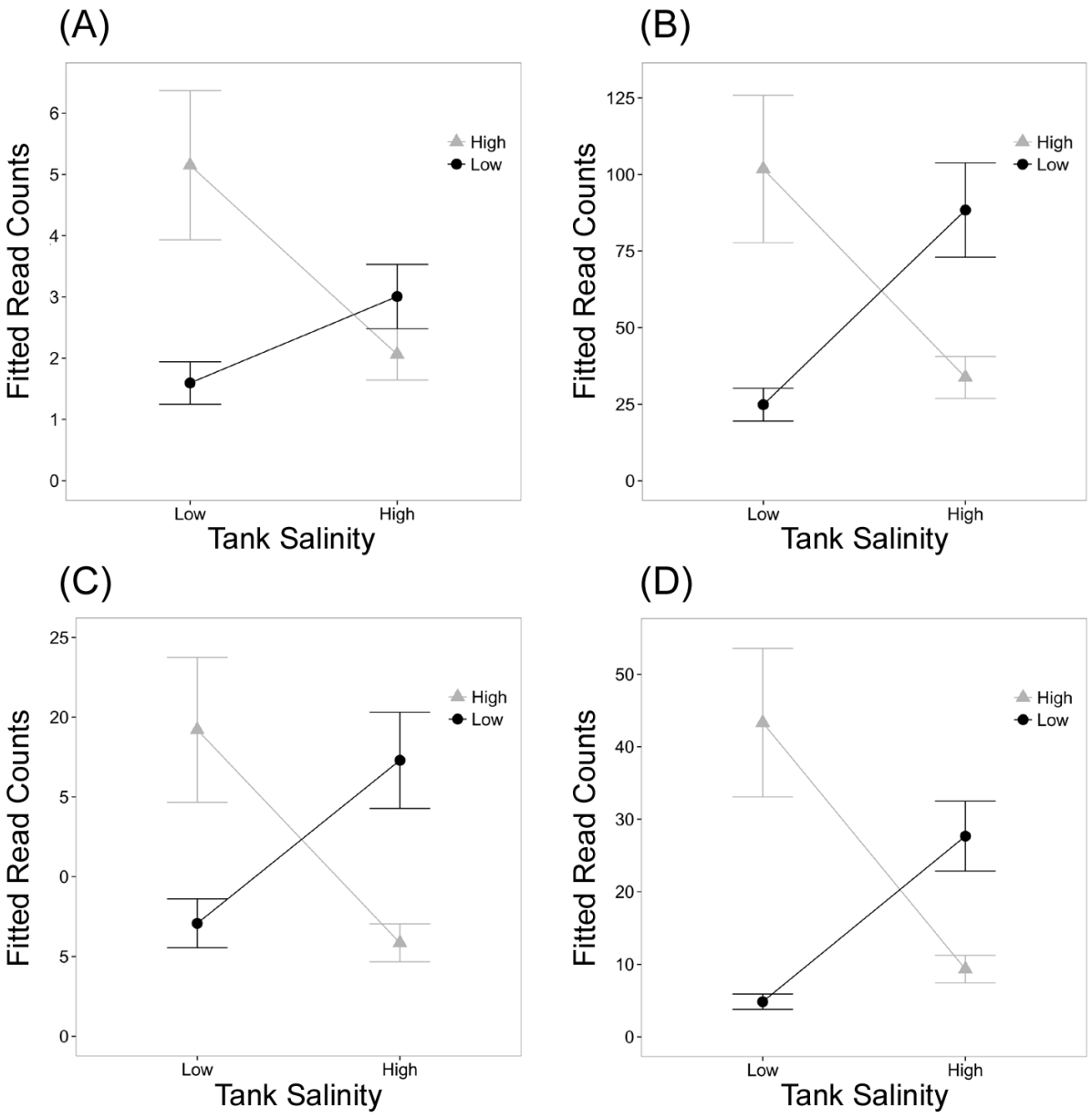

Figure B.2. Reaction norms for each of the four case study reftigs. The reaction norms are of read counts that have been normalized and fit to a negative binomial model. Mean read counts are plotted separately for each reef source. Error bars are standard error. (A-D) correspond to (A-D) in Figure 1. (A-C) were not significantly differentially expressed whereas (D) was significantly differentially expressed in response to the reef by treatment interaction term. Note that the y-axis changes in each panel of the figure.