Spring 2022

From Gene Expression to Physiology: A Study of Chronic Thermal Tolerance in the *Mytilus Edulis* Species Complex

Lindsey Cate Schwartz

FROM GENE EXPRESSION TO PHYSIOLOGY: A STUDY OF CHRONIC THERMAL TOLERANCE IN THE *MYTILUS EDULIS* SPECIES COMPLEX

by

Lindsey Cate Schwartz

Bachelor of Science
Muhlenberg College, 2016

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biological Sciences

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

Thomas J Hilbish, Major Professor

Daniel I Speiser, Committee Member

Jeffry L Dudycha, Committee Member

Yen-Yi Ho, Committee Member

Manuela Truebano, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

DEDICATION

This dissertation is dedicated, with all my love, to Marvin Streit, for sparking the passion that has fueled this great adventure.

ACKNOWLEDGEMENTS

I cannot begin in any other way but to thank my parents Linda and Howard Schwartz for their continued, unwavering support throughout this journey, and the rest of my life. I am twice grateful to my mother for drying every tear and listening to every frustration; I could not have come this far without you both. Another very special thank you to my partner Joshua Peacock, my support, cheerleader, and companion, for keeping me laughing even on the darkest days.

I thank my advisor Dr. Jerry Hilbish, my supervisor while at the University of Plymouth and committee member Dr. Manuela Truebano, and the rest of my committee, Dr. Daniel Speiser, Dr. Jeffry Dudycha, and Dr. Yen-Yi Ho. I cannot begin to express my gratitude for the training, support, and guidance I received from Dr. Truebano, who went well above and beyond for me over the last several years. Additionally, my sponsor at the National Museum of Natural History Dr. Ellen Strong, and my wonderful mentor Dr. Vanessa González. Lastly, I am ever grateful for the technical and emotional support offered by Marie Palmer while working in the lab at the University of Plymouth.

With much love, I also want to thank my friends here at the University of South Carolina for all their insights, moral and statistical support, and coffee runs: Rachel Steward, Rachel Schomaker, Trenton Agrelius, Maeve Snyder, and Blair Stokes. Lastly, I want to thank the scientists who mentored and inspired me long before I set foot on this campus; Ms. Joan Soon, Dr. Suzanne Fenstemaker, Dr. Erika Iyengar, and Dr. Jon Norenburg. Thank you all for the invaluable gifts of courage and curiosity.

ABSTRACT

Rising ocean temperatures are a severe and ever-present threat to marine life. With environmental temperature having such a large impact on organismal performance, understanding the mechanisms which contribute to the ability to survive at higher temperatures is a crucial research focus. Significant progress has been made in discovering these mechanisms on the cellular, biochemical, and physiological levels, but is it much less common for those to be examined together. Using the *Mytilus edulis* species complex as a model system, this dissertation takes a closer look at how prolonged exposure to sub-lethal high temperatures impacts marine organisms on multiple levels of biological organization. The *M. edulis* species complex is ideal for this work given the district thermal niches of the congeners, existence of persistent natural hybrid zones, and a reference genome.

Using a combination of molecular, physiological, and bioinformatic approaches, we were able to assess the population structure of the *M. edulis* x *M. galloprovincialis* hybrid zone in southwest England, compare the responses to chronic warming between the two parent species and their hybrids, and link gene expression patterns to physiological responses. We found that individuals collected from parent populations were mostly genetically pure, and hybrids showed a gradient of ancestry from mostly *M. galloprovincialis*-like to mostly *M. edulis*-like, with a skew toward *M. edulis* ancestry. Physiologically, our results support previous work that suggests feeding rate limits the

cool water *M. edulis* at high temperatures. Our work is the first to study this response in *M. edulis* x *M. galloprovincialis* hybrids, and we show that unexpectedly, they perform better energetically at high temperatures, and show a more *M. galloprovincialis*-like response. This translates well to the transcriptomic analysis which showed that in feeding rate (the limiting physiological factor in *M. edulis*), hybrids were more like the warm tolerant *M. galloprovincialis* than their cool water parent. Additionally, despite their close phylogenetic relationship, the three genetic groups deployed largely unique transcriptomic responses to chronic warming. This, coupled with the limited number of genes which exhibited a genotype-by-environment interaction, indicates there might be limited genetic variation for plasticity related to thermal stress in this species complex. Lastly, we found significant standing genetic variation in the populations but were unable to corelate it to either of the physiological phenotypes measured. With little genetic variation in plasticity and in traits shown to be limiting at high temperatures, the M. edulis species complex may be especially vulnerable to climate change. However, we did find that hybrids tended to exhibit greater plasticity in gene expression and were better able to cope physiologically with the high temperature treatment. Therefore, hybridization may be able to generate new combinations of genes that are more fit in the novel environments created by climate change and may offer a route to species persistence.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Environmental temperature is arguably the most important abiotic factor affecting the performance of organisms. This is especially true for ectotherms, whose body temperature is mediated by the environmental temperature. Organismal performance increases with temperature until reaching an optimum temperature where performance is maximized (T_{opt}); beyond T_{opt}, performance sharply declines (Huey and Stevenson 1979; Huey and Kingsolver 1993). The relationship between temperature and performance is visualized by plotting some metric of performance against temperature, creating a thermal performance curve (TPC). The curve is bounded by the critical thermal minimum and maximum (CT_{min} and CT_{max}), beyond which the organism cannot survive. The range of temperatures between these points is called the tolerance breadth. The shape of the TPC, CT_{min}, CT_{max}, and T_{opt}, varies widely between organisms, and depending on life stage, season, prior thermal environment, and even the trait measured (Martin and Huey 2008; Angilletta 2009; Dowd et al. 2015; Sinclair et al. 2016), but generally maintains this left-skewed shape.

The evolution, plasticity, and heritability of these traits are still active areas of research (Dowd et al. 2015; Gunderson and Stillman 2015; Sinclair et al. 2016; Kingsolver and Buckley 2017) and are particularly interesting considering climate

change. Climate change can alter the mean environmental temperature, the variance, or both. As mean temperatures rise, organisms are now facing temperatures near their thermal limits in places they did not previously. The increased frequency and severity of extreme weather events, including heat waves, puts organisms at increased risk of exposure to lethal conditions. These changes affect organisms differently, and both come with a unique set of challenges and adaptations. While many studies focus on acute stress, chronic, but non-lethal, thermal stress will also play an important role in reshaping communities as mean temperatures continue to rise (Huey and Kingsolver 2019).

Chronic high temperature can cause energetic strain, as metabolic demand increases with temperature. Increased metabolic demand must be met with increased energy intake, or survival becomes time limited. The theory of energy limited thermal tolerance posits that organisms are limited at high temperatures due to their inability to maintain a positive energy budget (i.e., the amount of energy available to an organism) (Sokolova 2013; Sokolova 2021). This "metabolic meltdown" can be caused by external food limitation or by mechanical or behavioral failure by the organism, but in either case, chronic thermal stress leads to an inability take in enough energy to meet increasing metabolic demand (Huey and Kingsolver 2019).

The blue mussel species complex (genus *Mytilus*) is a group of marine bivalves composed of the three congeners *Mytilus trossulus*, *M. edulis*, and *M. galloprovincialis*. Mytilids are important space occupiers in rocky intertidal and shallow subtidal ecosystems with global distributions (Koehn 1991). Members of the *M. edulis* species complex and other closely related Mytilids are also farmed and harvested, forming the basis of a multibillion-dollar industry worldwide (FAO 2019). The three congeners are

physiologically and genetically differentiated, each having a distinct thermal niche. *M. trossulus* has the lowest T_{opt}, inhabiting boreal waters, followed by *M. edulis,* and then *M. galloprovincialis*, which is native to the Mediterranean. Despite this differentiation, in places where their ranges overlap blue mussels can readily hybridize, creating persistent natural hybrid zones (Skibinski et al. 1978; Skibinski et al. 1983; Hilbish et al. 2002; Brannock et al. 2009).

Blue mussels, particularly *M. edulis*, have been used as a model system for ecophysiological studies for more than two decades, meaning there is a rich knowledge of their thermal physiology (Bayne 1976; Lockwood et al. 2015). Closely related congeners with distinct thermal physiologies make excellent model systems to study the effects of climate warming, and the mechanisms by which organisms may be limited by high temperature. Climate change has already begun to impact the *M. edulis* species complex, as range shifts have been documented and/or projected in all three congeners (Jones et al. 2010; Wethey et al. 2011; Fly et al. 2015). Previous evidence indicates that feeding rate may be the physiological process which limits survival at high temperatures in the cool-temperate *M. edulis* (Fly and Hilbish 2013). This would fit well with the theory of energy limited thermal tolerance, as failure at high temperature is caused by insufficient feeding.

Here, I address the effects of chronic high temperature using a variety of physiological, molecular, and computational approaches using *M. edulis* and *M. galloprovincialis* as model species. This work addresses four major objectives: **1)** to determine the mechanism by which *Mytilus edulis* is limited at high temperatures (Chapter 3), **2)** to corelate the physiological and molecular responses to chronic high temperature in *M. edulis, M. galloprovincialis*, and their hybrids (Chapter 3), **3)** to

categorize the response of *M. edulis* x *M. galloprovincialis* to chronic high temperature both physiologically (Chapter 3) and trascriptomically (Chapter 2), and **4)** to assess the population structure of collected individuals and relate genetic variation to physiology (Chapter 4).

CHAPTER 2

DIVERGENT TRANSCRIPTOMIC RESPONSES BY HYBRID AND PATRENTAL BLUE MUSSELS TO CHRONIC THERMAL STRESS

2.1 Introduction

With global climate change bringing increased average temperatures in both terrestrial and marine habitats (Collins et al. 2019; Oliver et al. 2021), ectotherms will continue to experience the physiological consequences of warming (Helmuth et al. 2006; Pinsky et al. 2020). Understanding the mechanisms underpinning thermal tolerance can help predict population level consequences of warming and help determine how communities will change, aiding in management of ecologically and economically important species (Buckley et al. 2010; Somero 2011; Collins et al. 2019). There has been a considerable focus on thermal extremes, and how organisms respond to acute, lethal, conditions (Barshis et al. 2013; Gunderson and Stillman 2015), but considering how organisms will respond to chronic, sub-lethal stress is also crucial (Woodin et al. 2013; Fly et al. 2015; Gunderson and Leal 2016). Elucidating the mechanisms behind chronic, sub-lethal, thermal stress tolerance will be especially important with rising average sea surface temperatures, and the increase in frequency and severity of longer term stress events (Collins et al. 2019).

In the face of rapidly changing conditions, the question of whether and how organisms will be able to persist is increasingly important, and the relative roles of phenotypic plasticity and evolution in this struggle remain unclear (Hoffmann and Sgrò 2011; Gunderson and Stillman 2015; Kingsolver and Buckley 2017; Oostra et al. 2018). Phenotypic plasticity is the ability of a single genotype (or genome) to produce multiple phenotypes in response to different environmental conditions (West-Eberhard 1989; Sommer 2020; Rivera et al. 2021). The genotype by environment (G x E) interaction can then be thought of as differential plasticity among genotypes, where the effect of the environment is different depending on the genotype of the organism. The G x E interaction is particularly important when considering hybrid zone dynamics. G x E interactions can contribute to the generation, persistence, and structure of hybrid populations via the differential fitness of genotypes under different environmental conditions (Campbell and Waser 2001; Shields et al. 2008; Thompson et al. 2021). Hybridization shuffles the genome and creates new combinations of phenotypes, which can be less, equally, or more fit than either parent, which can also depend on the environment. Transgressive phenotypes happen when hybrid phenotypes are outside of the range of phenotypes exhibited by the parent species, either positively (leading to heterosis, or hybrid vigor) or negatively (Rieseberg et al. 1999). If hybridization reshuffles genomes and create transgressive phenotypes for traits relevant to persisting in changing climates, then it may be a mechanism by which species can adapt, particularly if the phenotype(s) can contribute to niche specialization from the parent species (Rieseberg et al. 1999).

The blue mussel species complex (genus *Mytilus*) has been used as a model in ecophysiology for decades (Lockwood et al. 2015). Mytilids are both major substrate occupying members of rocky intertidal communities (Menge and Branch 2001), and are farmed and wild caught in fisheries worldwide (Goulletquer 2004). The *M. edulis* species complex consists of the three congeners *M. trossulus*, *M. edulis*, and *M. galloprovincialis*. The three species have different thermal niches, inhabiting primarily boreal, cold-temperate, and warm-temperate/Mediterranean climates, respectively. They are physiologically and genetically distinct (Hilbish et al. 1994; Fly and Hilbish 2013) but are still sufficiently closely related to hybridize and form persistent hybrid zones (Skibinski et al. 1978; Rawson et al. 1999; Hilbish et al. 2002; Brannock et al. 2009). While the physiological responses to temperature of both *M. edulis* and *M. galloprovincialis* are well characterized (Bayne 1976; Braby 2006; Anestis et al. 2007; Dimitriadis et al. 2012; Fly and Hilbish 2013), the response of their hybrids remains largely unexplored (but see Hilbish et al. 1994). Sea surface temperature (SST) has been shown to play a large role in structuring the biogeographic distribution of the complex, and changes in average SST have already lead to range shifts in some populations (Jones et al. 2010; Fly et al. 2015). Their physiological distinction despite close phylogenetic relatedness, the persistence of natural hybrid zones, and the fact that temperature is known to be an important mechanism for structuring their populations, make the *M. edulis* species complex particularly well suited to investigate the plasticity of gene expression under thermal stress.

In the present study we aim to explore the species-specific transcriptomic responses of *M. edulis* and *M. galloprovincialis* to a chronic, sub-lethal, thermal

challenge. In particular, we focus on genes that show genotype-specific responses to temperature (G x E), which are likely to contribute to species specific differences in thermal tolerance. We also aim to characterize the molecular response of *M. edulis* x *M. galloprovincialis* hybrids to this type of thermal challenge to investigate whether hybrids may differ in their transcriptomic response to thermal stress from either or both parents. Finally, we briefly investigate the potential functional roles of genes involved in mytilid thermal tolerance. This mechanistic understanding of the species-specific plasticity in gene expression to chronic high temperature provides important insight into the potential role of gene expression plasticity in structuring hybrid zones, and how they may change as the world continues to warm.

2.2 Materials and Methods

2.2.1 Sampling sites and pre-exposure conditions in the laboratory

Mytilus individuals were collected from four populations along the coast of Southwest England. The area is characterized by a mosaic hybrid zone, where sympatric and allopatric populations occur in succession along the coast of Cornwall (*M. edulis* east of Start Point and *Mytilus galloprovincialis* west of Saint Ives, hybrids between the two) (Hilbish et al. 2002). *M. edulis* were collected from an allopatric site in Sidmouth (SD), *M. galloprovincialis* from allopatric sites in Port Quin (PQ) and Trebarwith (TW), and putative hybrids from the sympatric site Whitsand Bay (WB). Only young adult mussels (25-35 mm shell length) were collected, as this size range has the most even mixture *of M. edulis* and *M. galloprovincialis* alleles at Whitsand Bay, according to the size

dependent allelic frequency of hybrids described in Hilbish et al. 2002. The narrow size range also helps to reduce size dependent variation in metabolism. Animals were transported to Plymouth University's Marine Biology and Ecology Research Center and kept in six 15L stock aquaria under constant conditions (temperature = 15°C, salinity = 32 ppm, 12h:12h L:D regime) for a pre-exposure period lasting a minimum of two weeks. During this time, mussels were fed *Isochrysis galbana* (Iso 1800™, Reed Mariculture) and full water changes were performed daily.

2.2.2 Temperature treatments

Experimental temperatures were selected based on current biogeographic distributions of *M. edulis* and *M. galloprovincialis*, and their physiological thresholds. The control temperature of 15°C represents average summer sea surface temperature (SST) in southwest England where the animals were collected. The warm temperature of 23°C is not only the thermal optimum for *M. galloprovincialis*, but also the maximum summer sea surface temperature (SST) which can sustain populations of *M. edulis*; there are no populations of *M. edulis* that persist where summer SST reaches 23°C (Fly and Hilbish 2013; Fly et al. 2015).

Eight closed system mesocosms were set up with water recirculating between a 32L main tank, where mussels were held, and a 60L overflow tank. All tanks were held in a 15°C controlled temperature room, and the warm treatment tanks were ramped from 15°C (holding temperature) to 23°C (experimental temperature) and maintained with aquarium heaters in the overflow tank. Partial water changes were conducted every other

day and the temperature was checked daily. All other conditions were kept as per the pre-exposure period. Mussels were fed the same *I. galbana* (200 µL per tank, twice daily) as in pre-exposure tanks.

2.2.3 Sample preparation and sequencing

At the end of the exposure period, mussels were dissected and one pair of gills from each animal was preserved in RNA*later*™ and shipped to the Laboratories of Analytical Biology (LAB) at the Smithsonian Institution's National Museum of Natural History for preparation for transcriptome sequencing. Total RNA was extracted from the 90 gill tissue samples (15 per treatment per genotype) using TRIzol™ Reagent (ThermoFisher Scientific) following the manufacturer's protocol, with a final elution volume of 30 µL. The RNA was quantified with a Qubit 2.0 fluorometer and integrity assessed via spectral analysis performed with an Epoch Microplate Spectrophotometer (BioTek®). Ninety cDNA Libraries were prepared using the KAPA mRNA HyperPrep Kit for Illumina (KAPA Biosystems) following the manufactures protocol, beginning with Poly-A selection using KAPA oligo dT beads. Poly-A selection was followed by 6 minutes of RNA fragmentation at 94 °C and 11 PCR cycles for library amplification. Libraries were indexed with KAPA Dual Indexed Adapters (diluted to 7 µM). Libraries were quantified using a Qubit 4.0 fluorometer and the size distribution was checked using an Agilent TapeStation 2200 with High Sensitivity D1000 Screen Tapes. Adapter dimer was removed from four libraries using a BluePippin (Sage Science), selecting for fragments longer than 200 bp. Libraries were diluted to 15 nM, pooled, and shipped to GENEWIZ Next Generation Sequencing (South Plainfield, NJ, USA) for sequencing.

The pool was sequenced (150 bp paired-end) across four lanes on an Illumina HiSeq 4000 sequencer.

2.2.4 Quality assessment and filtering

An average of 18.07 ± 3.96 million reads per sample were generated, with an average quality score of 36.4. Read quality was assessed with FastQC 0.11.8 (Andrews 2010) before and after trimming with Trimmomatic 0.39 (Bolger et al. 2014) using the parameters recommended by MacManes (2014), but with the keepBothReads option set to TRUE in the ILLUMINACLIP argument. After filtering, an average of 15.1 ± 3.6 million reads per sample were passed onto downstream analysis. FastQC reports indicated rRNA contamination, so the reads were processed with Bowtie2 (Langmead and Salzberg 2012) to remove rRNA sequences. Using a custom FASTA file of *Mytilus* nuclear and mitochondrial rRNA sequences compiled from publicly available GenBank data, an index was generated with bowtie2-build, and all reads that failed to align to *Mytilus* rRNA were passed to downstream analyses. The same method was used to filter out potential algal contamination using the *Emiliana huxleyi* genome as a substitute for *I. galbana*, for which there is no available genome.

2.2.5 Mapping, quantification, and differential expression

Trimmed and filtered reads were aligned to the *M. galloprovincialis* genome (Gerdol et al. 2020) using STAR (with default parameters plus --outFilterScoreMinOverLread and --outFilterMatchNminOverLread both set to 0.3 (Dobin

et al. 2013). Exons were quantified by geneid with featureCounts. The raw counts matrix containing each sample was imported into R for downstream analysis. More than half of the resulting features in the counts matrix were long non-coding RNAs (lncRNAs). Long non-coding RNAs are generally defined as non-coding RNA molecules greater than 200 nucleotides in length. They are generally shorter and differentially regulated from protein coding RNAs (Ma et al. 2013), so it was decided that they should be considered separately from protein coding genes. To do this, two counts matrices were generated and analyzed independently: one with only mRNAs, and one with only lncRNAs.

The two counts matrices were read into DESeq2 (Love et al. 2014) and analyzed, with the same parameters. The model used for the design matrix included genotype, temperature, and the interaction between the two. *Mytilus edulis* and control temperature (15°C) were set as the baselines for genotype and temperature, respectively. Using this model, comparisons were run for (i) the effect of temperature on *M. edulis*, (ii) the effect of temperature on *M. galloprovincialis*, (iii) the effect of temperature on hybrids, (iv) the interaction between genotype and temperature for *M. edulis* vs. *M. galloprovincialis*, (v) the interaction between genotype and temperature for *M. edulis* vs. hybrids, and (vi) the interaction between genotype and temperature for *M. galloprovincialis* vs. hybrids. Reads were normalized within DESeq2, and log fold change shrinkage was applied with the ashr adjustment (Stephens 2017). The shrunken estimates were used for all analyses. Features were considered significantly differentially expressed if they had an adjusted p-value of $p < 0.05$.

2.2.6 Hierarchical clustering and GO term enrichment

Gene Ontology (GO) terms were retrieved from the recently published genome (see Gerdol et al. 2020 for details). A GO enrichment analysis was performed with GOSeq (v. 1.44.0, Young et al. 2010) to examine functional enrichment of differentially expressed features. For each genotype, the differentially expressed features (first protein-coding genes and then lncRNAs) were split into upregulated and downregulated lists fed into GOSeq as the genes of interest. GO terms were considered significantly enriched if GOSeq returned a p-value < 0.05. All genes that responded significantly to the interaction between genotype and temperature (GxE genes; comparisons iv – vi) were combined for functional analysis. The GxE genes were clustered using complete linkage hierarchical clustering, and GOSeq was run on each cluster, and then the entire list. GxE genes were not split into up- and downregulated lists since those are not meaningful categories in the interaction term.

2.3 Results

3.1 Differential expression in response to temperature

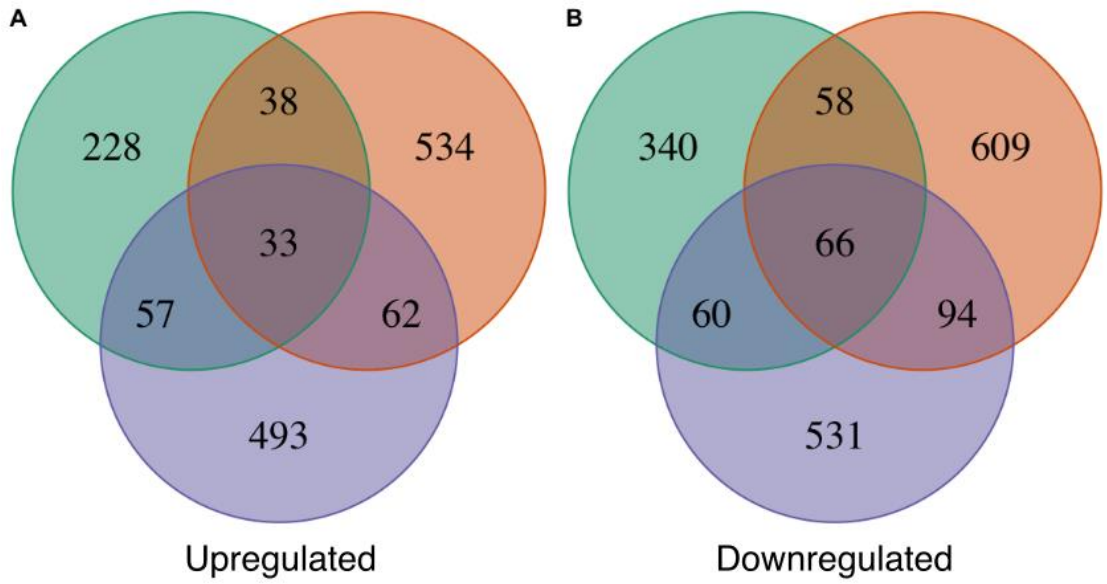

Figure 2.1: Genes that are significantly differentially expressed at 23°C compared to 15°C for each genotype groups. Differentially expressed genes are grouped into upregulated (A) and downregulated (B) genes, and color represents genotype group. *M. edulis* is in green, *M. galloprovincialis* in orange, and hybrids in purple.

In all three genotype groups, more genes were downregulated in response to high temperature than upregulated. In *M. edulis* there were 880 differentially expressed genes (DEGs) between the two temperatures, 356 upregulated (40%) and 524 downregulated (60%). In *M. galloprovincialis* 1,494 genes were differentially expressed, 667 upregulated (45%) and 827 (55%). In hybrids there were 1,396 DEGs, 645 upregulated (46%) and 751 downregulated (54%) (Fig. 2.1). Only 99 genes were differentially expressed in the same direction by all three groups. *M. edulis* had the lowest proportion of unique DEGs, *M. galloprovincialis* had the highest, and hybrids were intermediate. A principal components analysis (PCA) including the 500 most variable genes showed that individuals cluster by genotype along PC1, but there was no apparent separation with

respect to temperature treatment (Fig. 2.2). Along PC1 hybrids tend to be intermediate to the parent species, but overlap with *M. edulis* to a greater extent. Although many genes respond significantly to temperature within each group, when plotted individually, only hybrids show a mild separation by temperature in a PCA (Fig. A.1).

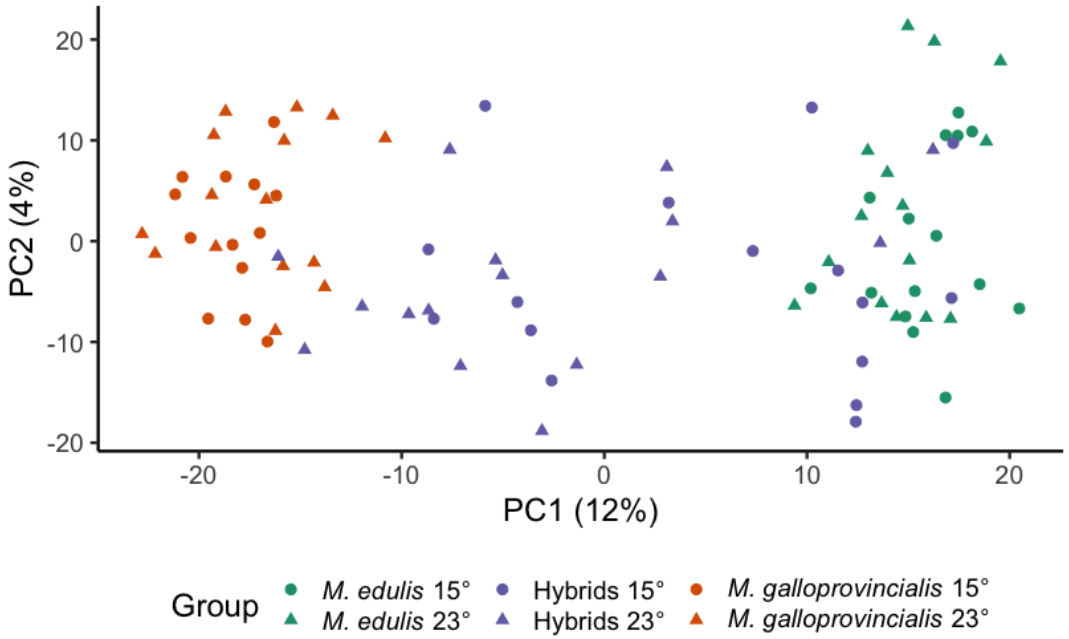

Figure 2.2: In principal component space, individuals sort based on genotype, but not temperature. Principal components analysis of the top 500 differentially expressed genes from all individuals. Circles are control temperature individuals (15°C), and triangles are heat treatment samples (23°C). Colors represent genotype groups, with *M. galloprovincialis* in orange, *M. edulis* in green, and hybrids in purple.

2.3.2 Functional Enrichment in protein coding genes

While the overlap in differentially expressed genes was limited, gene set enrichment analysis revealed significant overlap in the functions of genes that respond to temperature in the three groups. Major functional categories that appeared in all three groups include cytoskeletal organization, mitosis and meiosis, immune and inflammatory responses, and the cellular stress response (i.e. redox, protein quality control, DNA repair, and apoptosis). The importance (number of enriched terms) and direction

(positive, negative, or unspecified regulation) of the terms was different between the groups in several cases, indicating that though the processes involved in the thermal stress response overlap, they may be coordinated in different ways. There were also several functional categories that were important in only one or two groups, such as histone modification and methylation and the regulation of gene expression (enriched in downregulated genes only in *M. edulis* and hybrids), reinforcing the idea that the three genotype groups employ different transcriptional strategies under thermal stress.

2.3.3 The G x E interaction

The number of genes which exhibited an interaction between genotype and temperature (G x E) was very small (n = 147) compared to the effect of temperature in each group. Hierarchical clustering sorted those genes into six distinct clusters based on expression patterns (Fig. 2.3). In all but one cluster, the parent species showed opposite reactions to temperature, and the hybrids resembled one of the parents. In cluster two, there was only a mild genotype by environment interaction, with genes in this cluster being downregulated by *M. edulis* and hybrids, but unchanged in *M. galloprovincialis*. Genes in clusters one, five, and six, were upregulated in *M. edulis* and downregulated in *M. galloprovincialis*; genes in clusters three and four showed the opposite trend. Clusters four and six also showed an interesting pattern in which *M. edulis* and *M. galloprovincialis* had similar expression at 15°C but have opposite reactions to the increase in temperature.

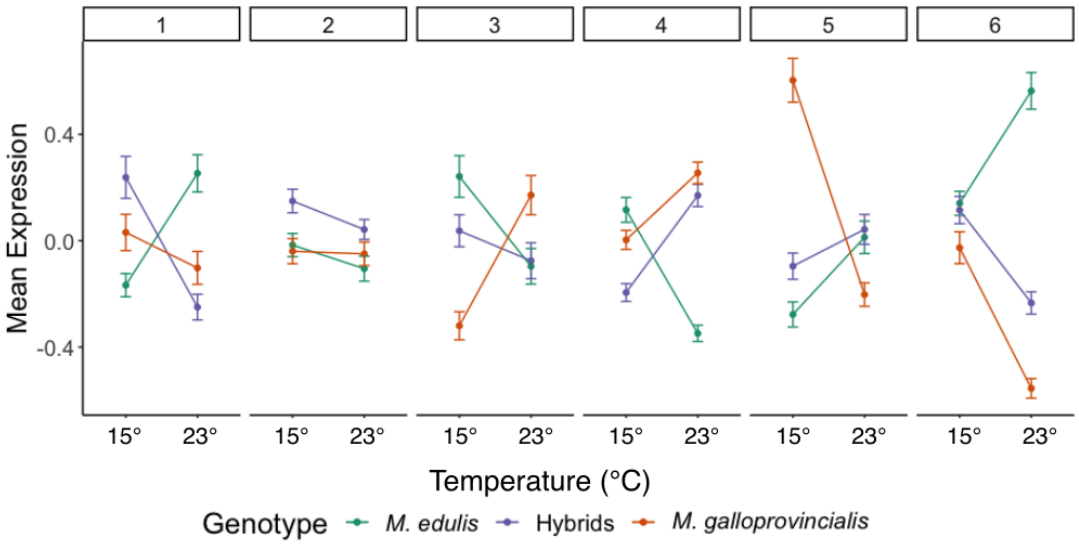

Figure 2.3: Reaction norms for the six clusters of genes that exhibited a statistically significant interaction between genotype and temperature (adjusted p-value < 0.05). The clusters contain 16, 34, 14, 42, 19, and 22 genes, respectively. Expresseion values were transformed using the variance stabilizing transformation (VST) in DESeq2, and those values were scaled and centered prior to clustering. Error bars represent the standard error. *Mytilus edulis* is in green, *M. galloprovincialis* in orange, and hybrids in purple.

A vast majority of the genes involved in the G x E interaction were not known proteins in the genome annotation, with 114 of 147 annotated as "hypothetical predicted protein" (~87%), making functional analysis challenging. A small number of the identified genes may be involved in the stress response, including two proteins involved in ubiquitination, one DNA repair protein, and one potential metabolic protein (serum amyloid A protein). There was a noticeable lack, however, of canonical stress response genes like heat shock proteins (HSPs). Functional enrichment for G x E genes was limited, given that most of the genes involved were not assigned gene ontology (GO) terms. Clusters one and five did not have any enriched biological process GO terms. Cluster two had the most significantly enriched GO terms, dealing mostly with immune functions and cell signaling, and some for apoptosis, meiosis, the cell cycle, and metabolism. Cluster three was also enriched mainly for immune response terms, along with some relating to apoptosis and ubiquitination. Cluster four was enriched for acute-

phase response, DNA replication, and macromolecular biosynthesis, and cluster six was enriched for the negative regulation of cellular processes. Lastly, cluster six was enriched only for the negative regulation of cellular processes, which is a vague, high-level GO term that does not provide much information on function.

We compared the gene expression plasticity of hybrids to both parents by comparing the log fold changes of DEGs (Fig. 2.4). We found that for many genes, plasticity in the hybrids was similar to either *M. edulis* or *M. galloprovincialis*. Alternatively, there were also many genes that displayed increased plasticity with respect to both parents, falling along the y = -x line (Fig. 2.4 dotted line). This indicates that overdominance in plasticity may be very prominent in this system. The unexpected number of genes that exhibit overdominance in plasticity supports the idea that hybrid intermediacy is actually the exception, not the rule (Rieseberg et al. 1999; Thompson et al. 2021), especially when paired with the relative lack of genes which are intermediately plastic in hybrids. This lack of intermediacy is also supported by the limited overlap in differentially expressed features between hybrids and either parent, both in protein coding genes (Fig. 2.2) and in long non-coding RNAs (Fig. 2.5).

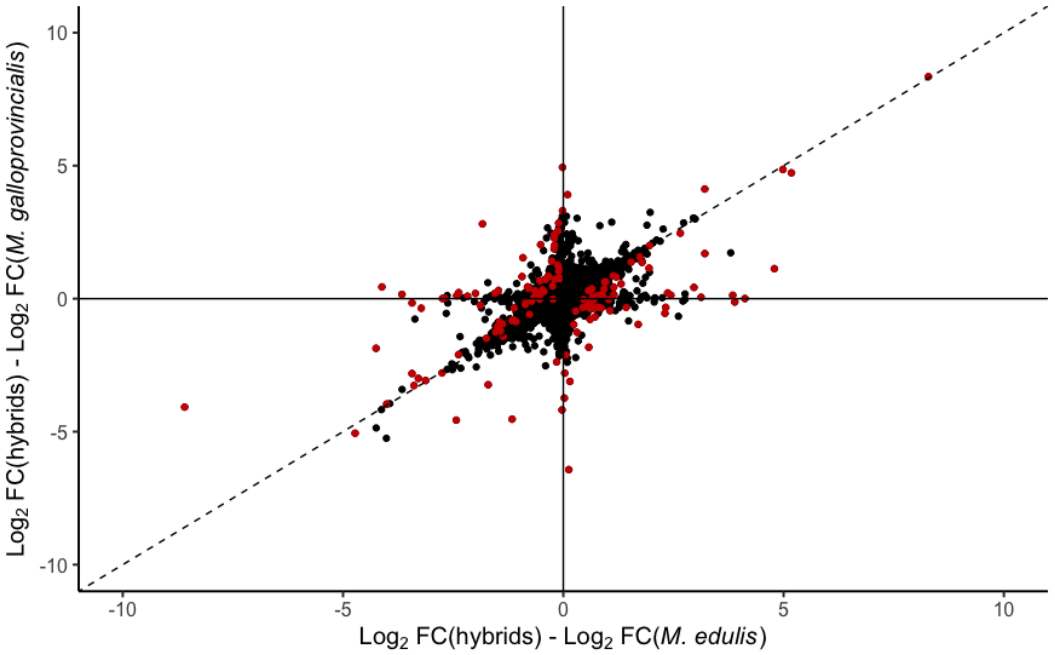

Figure 2.4: Differences in gene expression plasticity between the hybrids and each parent species shown as the difference in $\log_2$ fold change between hybrids and either parent plotted against each other. Plotted genes are the union of all differentially expressed genes (genes differentially expressed by temperature in each group plus the genes which exhibit an interaction between genotype and temperature). Genes in black are differentially expressed between temperatures in at least one genotype group, and genes in red are those which exhibit a genotype by environment interaction. Genes that are similarly plastic (i.e. similar changes in expression in magnitude and direction) would fall near the origin. Genes have similar plasticity to one parent, but not the other, would fall along either the x or y axis, depending on the similar parent. Genes which are intermediate in plasticity in hybrids (the magnitude and/or direction of change in hybrids is between that of the two parents) should plot in the (x, -y) or (-x, y) regions (i.e. along the y = -x line). On the other hand, genes that fall in the (x, y) and (-x, -y) regions of the graph (i.e. along the y = x line, the dotted line on this graph) indicate increased plasticity in the hybrids compared to either parent, or overdominance. For clarity, axis limits were restricted excluding five points from this plot; three overdominant points, one *M. galloprovincialis*-like point, and one *M. edulis*-like point.

2.3.4 Long non-coding RNAs

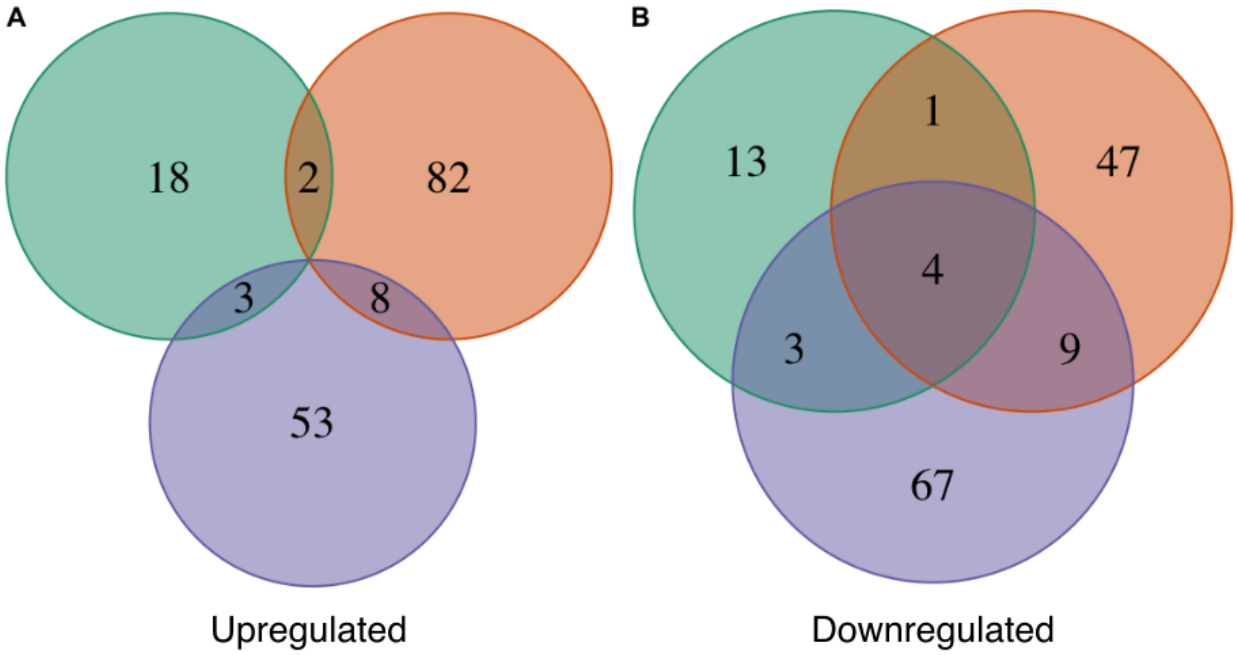

Figure 2.5: Venn diagrams showing the numbers of significantly upregulated (A) and downregulated (B) genes at 23°C compared to 15°C. *M. edulis* is in green, *M. galloprovincialis* in orange, and hybrids in purple.

In *M. edulis*, there were 44 significantly differentially expressed long non-coding RNAs (lncRNAs); 23 upregulated and 21 downregulated) (Fig. 2.5), and no significantly enriched GO terms. This is in contrast to both *M. galloprovincialis* and hybrids, which had more than three times as many differentially expressed lncRNAs (153 and 147, respectively) and several significantly enriched GO terms. In *M. galloprovincialis*, the 92 upregulated lncRNAs were enriched for oxidation-reduction processes, while the 61 downregulated lncRNAs were enriched for energy generation and transcription/ translation functions. Hybrids had 64 lncRNAs significantly upregulated in heat, which were enriched for calcium ion regulation, cell signaling, and regulation of respiratory gaseous exchange. The 83 downregulated lncRNAs were enriched for chromosome

condensation, RNA-templated transcription, and cell differentiation. Despite the much smaller number of differentially expressed features, there were also 21 lncRNAs which responded to the interaction between genotype and temperature. Hierarchical clustering sorted those lncRNAs into five clusters (Fig. 2.6).

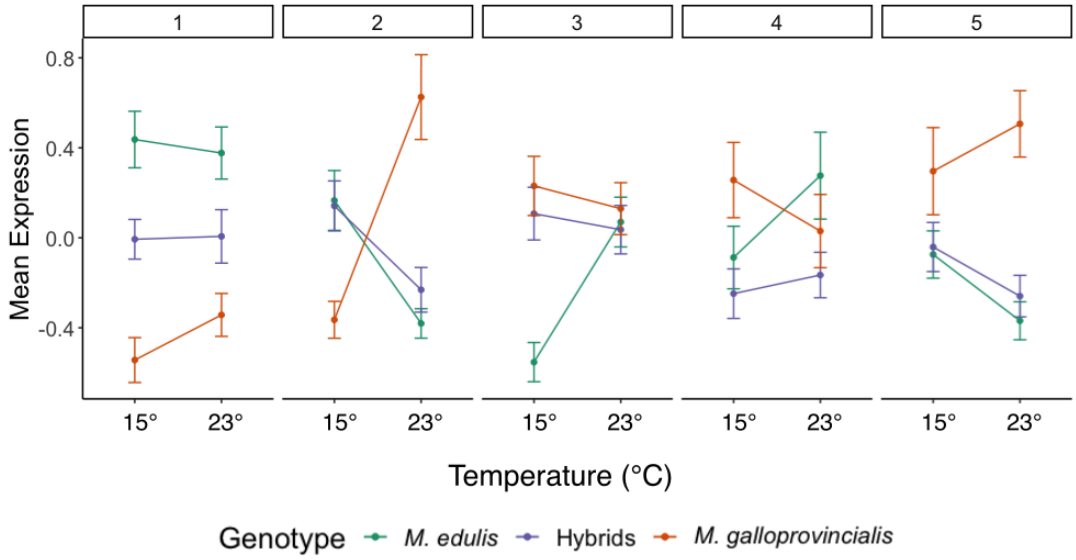

Figure 2.6: The five clusters of long non-coding RNAs (lcnRNAs) that exhibited a statistically significant interaction between genotype and temperature (adjusted p-value < 0.05) (A). Expression values and clustering were calculated with the same methods as Fig. 4. Clusters contain 5, 3, 4, 5, and 3 lncRNAs, respectively. *M. edulis* is in green, *M. galloprovincialis* in orange, and hybrids in purple.

2.4 Discussion

We subjected the cold-temperate *M. edulis*, the warm-temperate *M. galloprovincialis*, and their hybrids, to a two-week thermal challenge to determine their species-specific transcriptomic responses to chronic thermal stress. Our results show that while there are differentially expressed genes between genotypes and temperatures, genotype explains a greater proportion of the variance among samples, suggesting that in this complex the difference between genotypes outweighs the response to temperature.

We identified extensive interspecific differences in the transcriptomic response to high temperature, with each genotype group responding with largely unique strategies. This included hybrids, which were distinct from both parent species in terms of gene expression. We also found relatively few genes that exhibited a G x E interaction, suggesting that differences in response may arise from the expression of entirely different genes, rather than genotype specific differences in expression of the same genes.

2.4.1 Differential expression and the G x E interaction

Despite the close phylogenetic relationship between *M. edulis* and *M. galloprovincialis*, the genes responding to temperature are largely unique. It has recently been discovered that the *M. galloprovincialis* genome has substantial and widespread presence absence variation (Gerdol et al. 2020), which may contribute to the largely non-overlapping responses of the three groups. Surprisingly, hybrids also had a high proportion of uniquely differentially expressed genes, rather than having more overlap with one or both parent species. These differences indicate that despite close relatedness and continuing hybridization, the three groups have largely different transcriptional responses to chronic high temperature. These species-specific responses could prove advantageous in a changing climate, as certain responses will become more or less adaptive based on changes in environmental conditions. If a newly adaptive response is heritable, it could be a way for advantageous stress responses to introgress throughout the species complex.

The cold-temperate *M. edulis* had fewer DEGs under high temperature than either *M. galloprovincialis* or hybrids. Prior studies of thermal tolerance have shown that the more thermosensitive congener has a more robust response (i.e. more DEGs between temperature treatments) under chronic, but not acute, stressors (Narum and Campbell 2015; Veilleux et al. 2018). Under acute stress, the reverse is true (Barshis et al. 2013; Gleason and Burton 2015; Chen et al. 2019). This muted response in thermally sensitive strains/ populations may indicate limited expression plasticity for responding to thermal challenges (Veilleux et al. 2018). Alternatively, an increased response from more thermally tolerant strains/ populations may indicate a large and complicated set of genes having evolved to cope with high temperature (Narum and Campbell 2015). Without direct experimental evidence, however, it is impossible to say whether either (or perhaps both) causes this pattern.

The remarkably high proportion of genes exhibiting a G x E interaction annotated only as hypothetical predicted protein was unexpected. Genes which exhibit an interaction between genotype and temperature are more likely to play an important role in determining species-specific differences. Therefore, we anticipated these genes would be involved in important processes like metabolism, which tend to be conserved and well annotated. The results of this functional analysis also showed surprisingly little metabolic involvement in the response to chronic high temperature. This was unexpected given that metabolic demand increases with temperature (Kooijman 2010). Under the theory of energy limited thermal tolerance, the outpacing of metabolic capacity by metabolic demand contributes to organismal failure at high temperatures (Sokolova 2021). It has also been shown that metabolic genes play a large role in the thermal stress response

(Clark et al. 2013; Artigaud et al. 2015). Our functional analysis returned only a few enriched gene ontology terms, mainly dealing with broad functions like lipid biosynthesis and the response to glucose. Though it is possible that the functional analysis was not able to detect enrichment of metabolic terms in this non-model organism (see Doolittle 2018 for a discussion on assigning "functions" in non-model organisms), this is unlikely due to the fact that metabolic genes are generally highly conserved and well annotated. Also missing from the list of G x E genes were heat shock proteins and other molecular chaperones, antioxidant proteins, or other genes involved in the canonical thermal stress response among the G x E genes. This, however, is likely because the genes involved in chronic thermal stress are not the same as those involved in acute thermal stress, where one would expect to see immediate actors like heat shock proteins (Clark et al. 2013; Collins et al. 2021). Future studies should aim to identify more of these potentially important genes, giving us a better idea of what functions are involved in the genotype by temperature interaction in the species complex.

The small number of genes involved in the G x E interaction is consistent with earlier transcriptomic work on thermal stress in the *M. edulis* species complex (Lockwood et al. 2010) which showed that only a small proportion of genes measured responded differentially to temperature between *M. galloprovincialis* and *M. trossulus*. The limited number of G x E interaction genes suggests that variation in plasticity is limited within the species complex. This supports the conclusions of other recent studies that demonstrate little G x E interaction, despite the large individual effects of genotype and environment (Oostra et al. 2018; Sirovy et al. 2021). Both studies concluded (with the aid of additional genomic data) that genetic variation in plasticity was limited, and

that plasticity was largely parallel between environments. In the current study, however, responses to temperature were largely unique, rather than parallel, among genotype groups. This is likely at least partially because we are looking at a species complex, as opposed to different families within a species.

2.4.2 Gene expression plasticity

Gene expression, and the ability to alter the expression of key genes in response to environmental change, plays a major role in an organisms' ability to cope with stress. We show that gene expression plasticity was often transgressive in hybrids (Fig. 2.4). The log fold change in expression between temperatures in hybrids was much higher (i.e. expression of these genes was more plastic) in hybrids than in either parent for many differentially expressed genes. Heritable variation in expression plasticity can promote ecological speciation (and/or population persistence in novel environments) and increased stress tolerance (Pavey et al. 2010; Rivera et al. 2021). When a population(s) has significant variation in expression plasticity and selection is strong, the most adaptive response may be evolution in regulatory regions controlling the expression of key genes (Whitehead and Crawford 2006; Ghalambor et al. 2015; Campbell-Staton et al. 2020). Even variation in expression plasticity which is not heritable can be adaptive, either by maintaining plasticity throughout multiple generations via epigenetics, or by allowing populations to survive long enough for adaptive evolution to take place in other important, but less plastic, genes (Logan and Cox 2020).

Depending on the function of a gene and its baseline expression, increased stress tolerance can come from either increased or decreased expression plasticity (Rivera et al. 2021). The transgressive expression phenotypes in hybrids can then lead to differences in stress tolerance. Our work shows that hybridization is a potential avenue for population persistence in novel environments created by climate change. Future work should focus on identifying genes with increased plasticity under stress, to better understand in what functions expression plasticity is most important for stress tolerance. Understanding this would also help track the expression and plasticity patters of key genes through hybrid zones, or across active selection gradients.

2.4.3 Long non-coding RNAs

Long non-coding RNAs are involved in a wide variety of biological processes and can act as regulatory elements through a variety of mechanisms, including the regulation of transcription and translation (Ma et al. 2013; Vance and Ponting 2014; Yao et al. 2019). There is also evidence that lncRNAs play an important role in bivalve immunity (Sun and Feng 2018). Our results indicate that the differential expression of lncRNAs plays a role in regulating the thermal stress response, and that it may contribute to the species-specific differences in thermal tolerance. While there were proportionally fewer differentially expressed lncRNAs (which would support the fact that they are acting as regulatory elements), many of the patterns seen in differentially expressed protein coding genes help up. Like with protein coding genes, *M. edulis* had far fewer differentially expressed lncRNAs compared to both *M. galloprovincialis* and hybrids. In addition, differentially expressed lncRNAs were largely unique among the groups, with only four

differentially expressed (all downregulated) by *M. edulis*, *M. galloprovincialis*, and hybrids. lncRNAs can regulate genes in both cis- and trans-acting ways, and trans-acting lncRNAs have the capability to modulate large-scale changes in genes expression patterns (Vance and Ponting 2014). Without a chromosome-level genome assembly, however, it would be impossible to tell whether the lncRNAs identified here were in close genomic proximity to any of the differentially expressed protein coding genes.

2.4.4 Conclusions

Here we aimed to characterize the transcriptional response to sub-lethal thermal stress in the *Mytilus edulis* species complex. To our knowledge it is the first in-depth transcriptomic analysis of thermal tolerance in *M. edulis* x *M. galloprovincialis* adult hybrids. We found a significant number of genes responded to temperature in all three genotypes, and that those genes were largely unique to each group. In addition, there were comparatively few genes that exhibited an interaction between genotype and temperature. This lack of G x E interaction corroborates other recent results, which have suggested there may be limited standing genetic variation for plasticity when it comes to thermal tolerance (Oostra et al. 2018; Sirovy et al. 2021). Functional analysis showed that many of the same processes are involved in the response to temperature, with some key differences between species, which is consistent with previous work (Lockwood et al. 2010). We also show that long non-coding RNAs are a potentially important regulator of the thermal stress response, and that they may contribute to species specific differences in thermal tolerance. Future work should pursue this avenue with more directed experiments to further explore the potential role of lncRNAs in thermal performance. Finally, we

present that differences in gene expression and its plasticity hybridization may be an important pathway by which species can survive climate change.

CHAPTER 3

CORRELATING PHYSIOLOGICAL AND TRANSCRIPTOMIC PATTERNS UNDER CHRONIC THERMAL STRESS IN THE *MYTILUS EDULIS* SPECIES COMPLEX

3.1 Introduction

Environmental temperature plays a critical role in the performance and survival of all organisms, and anthropogenic climate change threatens these functions (Collins et al. 2019; Pinsky et al. 2020). The study of chronic heat stress is essential to predicting how species will respond to climate change (Jones et al. 2010; Somero 2011; Fly et al. 2015). Understanding thermal physiology can help us understand organisms' current biogeographic distributions and to forecast changes in distribution and performance with climate change (Jones et al. 2010; Sokolova et al. 2012; Woodin et al. 2013). It is equally important to think about the biological mechanisms which cause these physiological patterns, as they are important when considering the adaptive, or acclimatory, potential of important organisms (Chevin et al. 2010; Kingsolver and Buckley 2017).

Predicted climate warming will profoundly affect the energy balance of ectotherms. The energy budget considers energy intake through feeding, and energy expenditure through a variety of processes including homeostatic maintenance and fitness processes. Given metabolic demand scales with temperature, the cost of survival will increase in a warmer world. Energy limited stress tolerance has been proposed as a

framework to understanding how organisms cope with environmental stressors, in this case temperature. Under optimal conditions, organisms have sufficient aerobic scope to not only support basic metabolic function, but also devote energy to growth (including deposition of energy into reserves like glycogen and lipid stores) and reproduction (Kooijman 2010; Klepsatel et al. 2016). Under mild to moderate stress, organisms shift resources away from fitness related functions, toward their increased maintenance costs; aerobic growth is reduced, but may remain positive. As conditions worsen, aerobic scope approaches zero and there is only enough energy to maintain basal metabolic functions. Once aerobic scope becomes negative, energy is insufficient for covering metabolic costs, and the organism will eventually die (Sokolova 2013). Thus, populations are limited to areas in which they can maintain sufficient energy budget for reproduction and survival to reproductive age. Temperatures that are energetically demanding, but not yet lethal, are projected to become more frequent and more severe in the coming decades (Woodin et al. 2013; Fly et al. 2015; Collins et al. 2019), posing a major energetic challenge for ectotherms.

The blue mussel species complex (genus *Mytilus*) is the ideal study system for investigating the mechanisms contributing to differences in thermal tolerance. The complex consists of the three congeners *Mytilus trossulus*, *M. edulis*, and *M. galloprovincialis*, which have distinct thermal optima, inhabiting boreal, cool-temperate, and warm-temperate environments, respectively. Their thermal physiology has been studied extensively over the last several decades, particularly *M. edulis* (Bayne 1976; Fly and Hilbish 2013). Previous work has shown that southern range limits in this species complex are set by the temperature at which aerobic scope becomes negative (Jones et al.

2010; Fly and Hilbish 2013; Fly et al. 2015). This is especially true for European populations, where energy budget is thought to be especially important in setting range limits (Woodin et al. 2013). The rate at which mussels filter particles out of the water- clearance rate- has been proposed as the limiting factor at high temperatures (Hilbish et al. 1994; Fly and Hilbish 2013). As temperature increases, *M. edulis'* feeding rate cannot support its metabolism; this limits *M. edulis* geographically to places where summer sea surface temperature does not reach or exceed 23°C (Fly and Hilbish 2013; Fly et al. 2015).

The persistence of natural hybrid zones is another advantage of the *Mytilus* study system. Populations of *M. edulis* and *M. galloprovincialis* in southwestern England and northern France coexist and readily hybridize (Hilbish et al. 2002). Hybridization reshuffles the genome, potentially decoupling traits of interest and making it easier to identify their genomic basis. The same logic can be applied using gene expression data instead of genomic sequence data, to identify gene expression patterns that are associated with certain phenotypes. The thermal physiology of adult *M. edulis* x *M. galloprovincialis* hybrids is poorly understood (though see Hilbish et al. 1994). Hybrid phenotypes can be similar to one parent, intermediate, or transgressive (outside the range of the parent phenotypes) (Rieseberg et al. 1999). This is important in the context of climate change, as it may allow populations to persist via introgression of adaptive alleles or production of transgressive phenotypes that are more fit in novel environments (Rieseberg et al. 1999; Thompson et al. 2021). Prior research indicates that *M. edulis* x *M. galloprovincialis* hybrid phenotypes are highly variable, showing one parent-like, intermediate, or transgressive phenotypes based on the trait measured and the locus used

for grouping (Hilbish et al. 2002; Tolman et al. 2019). There is also evidence that the transcriptomic response of hybrids is often similar to one parent or transgressive, but rarely intermediate.

Here we address three questions: (i) How do *M. edulis* x *M. galloprovincialis* hybrids respond to chronic thermal stress, (ii) are there modules of similarly expressed genes that correlate with physiological traits under thermal stress, and (iii) can including hybrids help to identify those gene modules are related to physiological traits. *M. edulis*, *M. galloprovincialis*, and their hybrids were exposed to a two-week thermal challenge, and we measured physiological parameters that relate to energy balance and assessed full transcriptomic profiles from those same individuals. We then analyze those two data sets and combine them to investigate how gene expression correlates to physiological energetics in this species complex.

3.2 Materials and Methods

To address these questions, *M. edulis*, *M. galloprovincialis*, and their hybrids were subjected to a two-week thermal challenge, after which respiration and clearance rate were measured, and gill tissue was preserved for transcriptome sequencing. The methods for this experiment are described in full in Schwartz et al 2022 and are only briefly reviewed here. *M. edulis* and *M. galloprovincialis* were collected from allopatric populations in southwest England, and putative hybrids were collected from a sympatric site. Mussels were acclimated to 15°C, and then moved into experimental treatments of either 15°C (the approximate average sea surface temperature at the time of collection) or

23°C (the upper thermal limit of population persistence for *M. edulis*, and the thermal optimum for *M. galloprovincialis*) and maintained for two weeks. After the exposure period, respiration rate and clearance rate were measured, and the animals were dissected. Upon dissection the blotted weight of each animal was recorded, and one pair of gills was preserved in RNAlater® for transcriptomic analysis.

3.2.1 Physiological Measurements

Respiration was measured via closed chamber respirometry using the PreSens Fibox 4 system. Individual mussels were placed in 250 mL jars which were sealed underwater and placed on a stir plate with a stir bar for the duration of the measurement. Percent air saturation was measured every 15 minutes until it decreased by approximately 20%, after which the mussel was placed back into the experimental tank. If an individual spawned during the trial, it was not included in analyses. Respiration was calculated by converting percent oxygen consumed into mL O_2 (corrected for atmospheric pressure with the Weiss coefficient) and expressed as an hourly rate.

Clearance rate was measured using the flow through chamber method (Riisgård 2001; Filgueira et al. 2006). Using a gravity-fed system, seawater with an average initial algal concentration of 10,000 cells mL^{-1} (Riisgård et al. 2003; Pascoe et al. 2009) of *Isochrysis galbana* flowed through each chamber at approximately 67 (± 4) mL min^{-1}. The chamber volume (300 mL) and radius (46.5 mm) were chosen to minimize the mixing of inflow and outflow water within the chamber. Chamber dimensions were selected based on the size of the mussels in the experiment according to the recommendations of Filgueira et al. 2006. Flow rates were chosen following Fly and

Hilbish 2013 to minimize the effect of flow rate on clearance rate and achieve an approximate 20% reduction in particles between inflow and outflow (though our actual reduction often exceeded 20%). Mussels were placed in their chambers and left to acclimate for a period of two hours, after which water samples were taken from the inflow and outflow of each chamber. No mussels spawned during the clearance rate trials. Particle concentration was measured using a Z2™ Coulter Counter® Analyzer (Beckman Coulter), and clearance rate was calculated as:

$$CR = f \; x \; \frac{C_i - C_o}{C_o} \quad (1)$$

where f is the flow rate in mL min^{-1} and C_i and C_o are the inflow and outflow concentrations, respectively, in cells mL^{-1} (Filgueira et al. 2006). Once all physiological measurements were complete, each mussel was dissected and the shell length and blotted wet weight were recorded.

3.2.2 Statistical Analysis

Normality was assessed for each physiological trait and traits were transformed as needed. Clearance rate and the $\log_{10}$ of respiration rate were used as the response variables. The initial model for both variables included blotted weight, genotype, and temperature, and all possible interactions. Models were reduced in a stepwise fashion until they had only significant predictors and/or interactions. ANOVA or ANCOVA (depending on whether or not weight was retained) were done using the Anova function in the car package (Fox and Weisberg 2019).

3.2.3 Weighted Gene Co-expression Network Analysis

RNASeq data generated from the gill tissue of 90 mussels (28 *M. galloprovincialis*, 30 *M. edulis*, and 32 hybrids) were filtered for algal and ribosomal contamination and mapped to the *M. galloprovincialis* genome (Gerdol et al. 2020). Reads were counted using featureCounts (Liao et al. 2014), and exported for further analysis in R. A full discussion of the methodology regarding transcriptome sequencing and processing can be found in Schwartz et al. 2022. Once imported into R, the protein coding genes were separated from the non-coding RNAs, and only the protein coding genes were passed into downstream analyses. Genes with fewer than 10 reads in more than 90% of the samples were filtered out to reduce noise, resulting in a matrix of 31,338 genes across the 90 samples. Read counts were then normalized using the VST transformation in the DESeq2 package (Love et al. 2014). Clusters (modules) of co-expressed genes were identified with the Weighted Gene Co-expression Network Analysis (WGCNA) package (Langfelder and Horvath 2008). A signed, blockwise, network was constructed with a soft thresholding power of 10 and a minimum module size of 30 genes. Modules were detected using the consensus approach in which the genotype groups were analyzed separately and recombined at the end. This approach removes heterogeneity from the data, making it more suitable for WGCNA, while still producing results that are comparable across groups.

To relate each gene expression module identified in the gene expression data to the physiological variables, a trait correlation analysis was performed with the WGCNA package. Clearance rate and weight corrected $\log_{10}$ respiration rate were correlated to each module of co-expressed genes. Log_{10} respiration rate was adjusted using the weight

and weight by genotype interaction coefficients from the ANCOVA. Correlations were calculated between the eigengene for each module and the physiological traits using Pearson's correlation coefficient and significance was calculated as Fisher's asymptotic p-value in the WGCNA package (fig. S1, S2). The WGCNA packages defines eigengene as the first principal component of the expression matrix for a given module. A positive correlation indicates that with increasing trait values, the "expression" of the eigengene increases, whereas a negative correlation indicates the opposite. Following this, consensus correlation coefficients and p-values were calculated between each genotype group pair. A consensus correlation refers to a module that is correlated with a physiological trait in the same direction in both genotype groups in a pair (i.e. a module that is positively associated with respiration in *M. edulis* and *M. galloprovincialis*). In these cases, consensus values can be calculated. If the module correlates to physiological variables in opposing directions, no consensus is reached, and the correlation is left as "NA".

3.3 Results

3.3.1 Oxygen Consumption

The best model for $\log_{10}$ respiration rate retained blotted weight, temperature, genotype, and the interaction between weight and genotype. Temperature ($p < 2.2e\text{-}16$), genotype (0.0012), and the interaction between weight and genotype ($p = 4.72e\text{-}4$) were significant predictors of $\log_{10}$ respiration rate, the main effect of weight (0.41) was not. The weight by genotype interaction is most likely due to the significant differences in blotted wet weight between genotypes. Hybrids were the largest, followed by *M. edulis*,

and *M. galloprovincialis* were the smallest (Tukey's HSD $p < 0.01$). While animals from a small size range were collected to minimize the effect of weight on respiration, weight can vary significantly even within a small range of shell lengths. All three genotype groups significantly increase $\log_{10}$ respiration rate after two weeks at 23°C (fig 1B), which is consistent with the increase in metabolic demand at high temperatures (Huey and Stevenson 1979; Angilletta 2009).

3.3.2 Clearance Rate

For clearance rate, the best model retained genotype, temperature, and the interaction between genotype and temperature. Only the main effect of genotype ($p = 0.043$) and the interaction between genotype and temperature ($p = 0.021$) were significant, the main effect of temperature was not ($p = 0.053$). The genotype by temperature interaction is clearly visible in the reaction norm depicting clearance rate (fig. 1B). In both *M. galloprovincialis* and hybrids, clearance rate increased from 15°C to 23°C, but in *M. edulis* it sharply decreased after two weeks at 23°C.

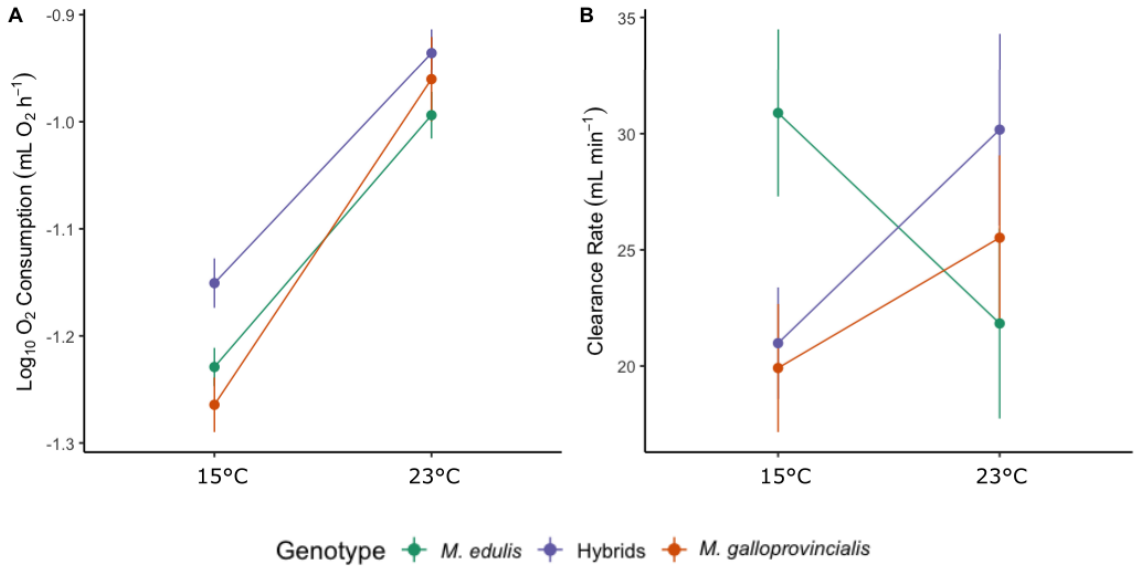

Figure 3.1: Reaction norms showing the change in oxygen consumption (A) and clearance rate (B) from 15°C to 23°C. Error bars represent mean standard error. *M. edulis* is shown in green, *M. galloprovincialis* in orange, and hybrids in purple.

3.3.4 Gene modules and their relationship to physiological traits

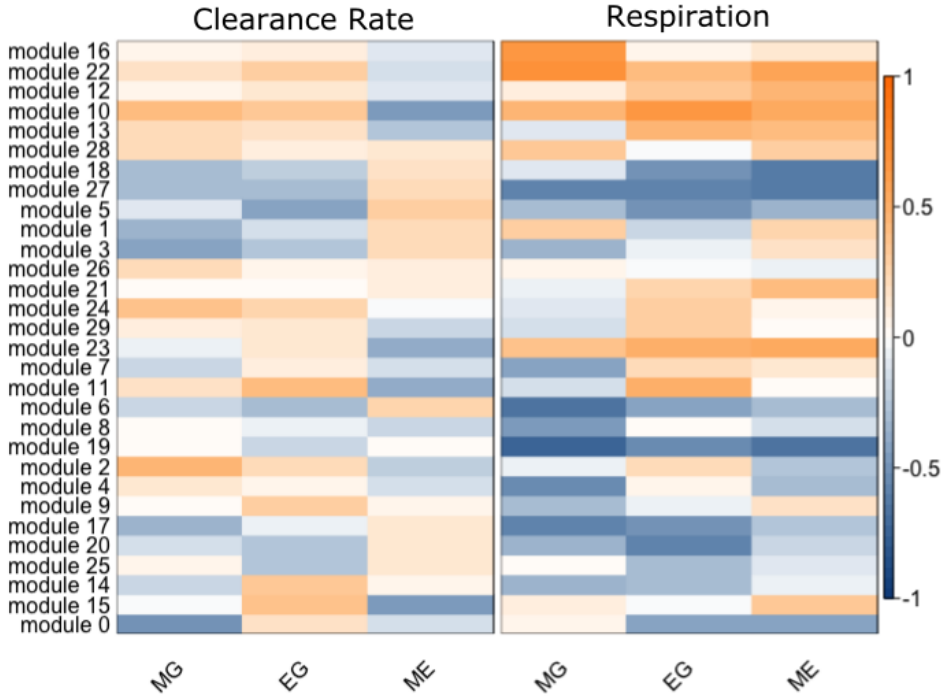

Figure 3.2: Heat map showing correlations between each gene module to clearance rate (left), or weight corrected log_{10} respiration (right). Saturation represents the strength of the correlation, so darker orange boxes have stronger positive correlations, and darker more blue boxes have stronger negative correlations. The column abbreviations indicate species: MG is *M. galloprovincialis*, EG are hybrids, ME is *M. edulis*.

29 modules of co-expressed genes were identified ranging in size from 58 to 7,262 genes, with a median module size of 241 genes, and a mean of 1,044 genes. There were also 7,262 genes not assigned to any module (module 0). Overall, respiration correlated much more strongly with gene expression modules than clearance rate. In *M. galloprovincialis* 18 modules (62.1%) had a correlation coefficient stronger that 0.3 with respiration; there were 13 (44.8%) such correlations in hybrids, and 16 (55.2%) in *M. edulis*. These include seven (24.1%), six (20.7%), and eight (27.6%) correlations, stronger than 0.5, respectively. With clearance rate, *M. galloprovincialis* had only nine (31%) modules with correlation coefficients greater than 0.3, hybrids had five (17.2%), and *M. edulis* had eight (27.6%). Only one module correlated with clearance rate with a coefficient stronger than 0.5, in *M. galloprovincialis*. Consensus correlations, in which two groups had a module correlated to respiration or clearance rate in the same direction, were also more abundant in respiration than in clearance rate (fig. 4). The consensus correlations also show that the gene expression modules associated with clearance rate were much more similar (i.e. more, stronger, consensus correlations) between *M. galloprovincialis* and hybrids than either was to *M. edulis*.

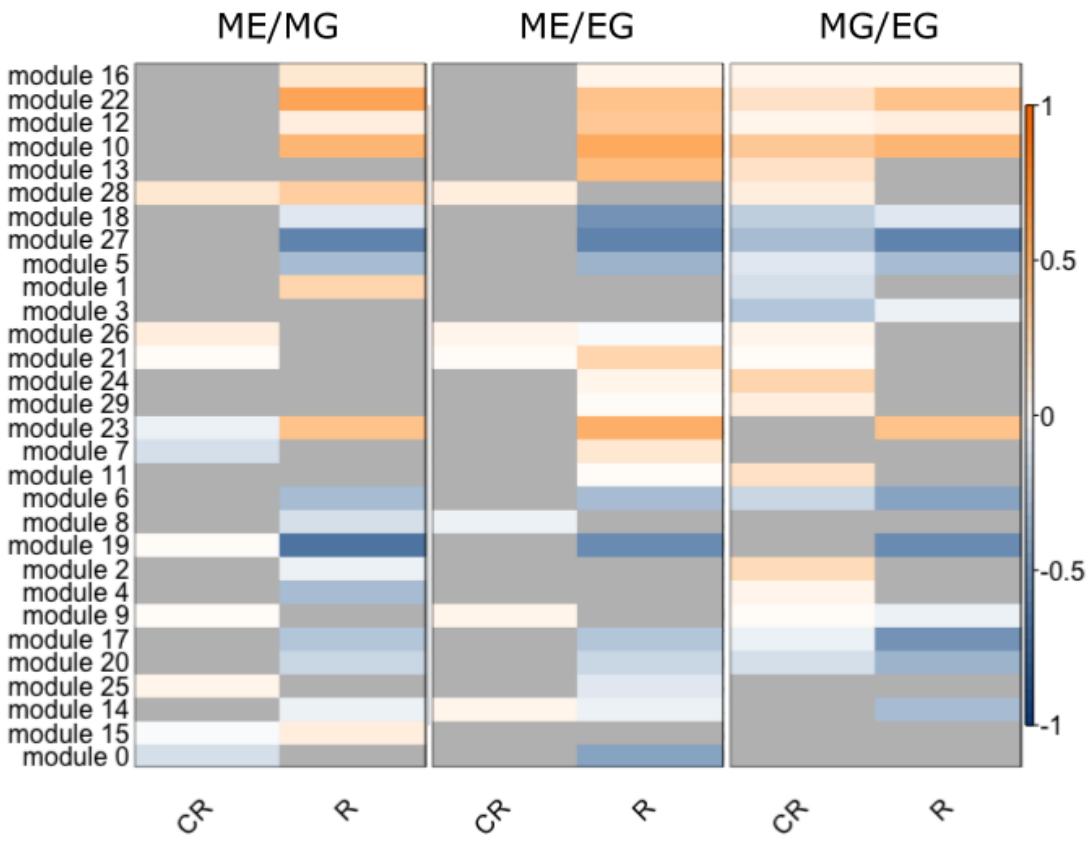

Figure 3.3: Heat maps of consensus correlations for each module and either clearance rate (left) or respiration (right) between genotype pairs. Consensus correlations are calculated when both genotype groups in a pair have a gene expression module correlate with either clearance rate or respiration in the same direction. Grey boxes indicate that the correlation was in opposite directions for the two genotype groups, and no consensus could be reached. Genotype pairs are *M. edulis/M. galloprovincialis* (left), *M. edulis*/hybrids (center), and *M. galloprovincialis*/hybrids (right). The column abbreviations indicate the trait measured: CR is for clearance rate, R is for weight corrected $\log_{10}$ respiration rate. Saturation represents the strength of the correlation, so darker orange boxes have stronger positive correlations, and darker more blue boxes have stronger negative correlations.

3.4 Discussion

Neither energy expenditure (in terms of oxygen consumption) nor energy intake (in terms of clearance rate) alone is enough to evaluate an animal's condition. At 15°C, *M. edulis* and *M. galloprovincialis* have equal rates of respiration, but *M. edulis* has a much higher clearance rate. This is consistent with previous work which reports that *M.*

edulis has a higher scope for growth at 15°C compared to *M. galloprovincialis* (Fly and Hilbish 2013). After two weeks at 23°C, all genotypes significantly increased respiration, but they do not all increase clearance rate. *M. galloprovincialis* and hybrids both increased their clearance rate at 23°C, but clearance rate in *M. edulis* declined drastically, potentially causing an energy deficit. This support previous work showing that in *Mytilus* mussels, clearance rate limits survival and persistence as temperature increases.

While hybrids had the highest $\log_{10}$ respiration rate at 15°C, the genotype by environment interaction was not significant, and therefore this difference is likely an artifact of the genotype by weight interaction. The interaction between genotype and temperature was significant for clearance rate, however, supporting that *M. galloprovincialis* and hybrids increased clearance rate at 23°C, but *M. edulis* decreased it. While hybrids appear to perform poorly at 15°C (high oxygen consumption and low energy intake), they seem to outperform even *M. galloprovincialis* after two weeks at 23°C, suggesting potential hybrid vigor at higher temperatures. This indicates a potential avenue for persistence as temperatures warm, and *M. edulis* struggles to maintain a positive energy budget.

The results of the weighted gene co-expression network support the physiology. There was a similar level of overlap in modules that correlate with respiration between genotype groups (figs. 2, 3). Likewise, all three genotype groups had increased respiration rate at 23°C, with no significant effect of genotype. This is markedly different from the clearance rate correlation data, which shows very little overlap between *M. edulis* and the other groups, but a significant consensus between *M. galloprovincialis* and hybrids. The consensus correlation values are generally weaker, but correlations are more

similar more than when either is compared to *M. edulis*. This mirrors the clearance rate data in which *M. galloprovincialis* and hybrids showed the same pattern of increase after two weeks at 23°C, while *M. edulis* did the opposite. Correlations with modules of co-expressed genes are likely weaker with clearance rate due to the high variation in clearance rate measurements, especially when compared to variation in oxygen consumption measurements. The presence of correlations despite the high variance, and that those correlations match the patterns seen in the clearance rate data, suggests that the pattern is a real biological phenomenon, despite the lack of statistical significance.

Our results support the hypothesis that clearance rate is the limiting factor for *M. edulis* at high temperatures. We also show here that the hybrids seemingly outperform both parent species at high temperatures. The weighted gene co-expression network analysis was a highly successful method to identify modules of similarly expressed genes and correlate them with physiological traits of interest. The patterns in the physiology are mirrored in the trait correlations, which strengthens the evidence that genes in those modules may contribute to difference in thermal tolerance. Our study supports the use of gene network correlations as a method to identify suites of genes that may be involved in phenotype of interest without relying on functional annotation, which can be difficult in non-model organisms (Doolittle 2018). Future research should focus on expanding this type of analysis to other physiological traits of interest to further explore the relationship between gene expression profiles and physiological phenotypes that will be important determinants of success under climate change.

CHAPTER 4

GENETIC VARIANTS ASSOCIATED WITH COMPLEX PHENOTYPES REMAIN ELUSIVE IN THE *MYTILUS EDULIS* SPECIES COMPLEX

4.1 Introduction

Climate change continually threatens marine species with rising temperatures, increased frequency of severe weather events, and more variable temperatures. With higher variation, organisms are more likely be exposed to novel thermal conditions. To survive in these conditions, organisms can acclimate (i.e. via phenotypic plasticity) or migrate. Over successive generations, species can also adapt to new conditions, or move into new ranges via larval dispersal. Standing genetic variation is the primary basis for adaptation, so variation in traits that relate to persistence under climate change is crucial for long term survival (Barrett and Schluter 2008; Bitter et al. 2019). Variation in important phenotypes allows for selection phenotypes which may become adaptive in novel environmental conditions, and can also ameliorate climate impacts by potentially limiting range contractions caused by climate change (Razgour et al. 2019).

As temperatures rise and become more variable, it is likely to become more difficult for sessile marine organisms to maintain a positive energy budget. An organism's energy budget comprises the amount of energy available for growth and

reproduction, and is highly dependent on temperature. Maintenance costs (the energy required to sustain basal metabolic and cellular functionality such as protein turnover, etc.) drastically increase with temperature (Nisbet et al. 2000; Kooijman 2010; Sokolova et al. 2012). Many different physiological functions contribute to energy budget, including respiration rate, energy intake, energy absorption, and excretion. Given that these processes are directly related to fitness and vary with temperature, standing genetic variation in genes contributing to them is necessary for adaptation to climate change.

Studying the genetic variation associated with metabolic traits can aid in both selection of genotypes in aquaculture species that are more likely to withstand climate change and assessing the potential of wild populations to adapt to that change. Hybrid zones have long been referred to as natural laboratories for studies of evolution (Hewitt 1988), and are an especially useful tool in assessing adaptive potential to climate change. Hybridization can maintain, and even increase, genetic diversity, and the reshuffling of genotypes can lead to new combinations, which may be adaptive in novel climates (Seehausen 2013; Scriber 2014). Hybridization can therefore lead to "genetic rescue" of populations which would otherwise be threatened via the introgression of adaptive alleles (Tallmon et al. 2004; Scriber 2014).

The *Mytilus edulis* species complex is an excellent model system for studying genetic variation in metabolic traits, as they naturally hybridize, the congeners occupy distinct thermal niches, and they have a published reference genome. This group of marine mussels has significant economic and ecological importance as the foundation of aquaculture industries and rocky intertidal communities. Composed of the three congeners *Mytilus trossulus*, *M. edulis*, and *M. galloprovincialis*, the three species are

physiologically and genetically distinct, but produce viable hybrid offspring, generating persistent mosaic hybrid zones (Koehn 1991; Hilbish et al. 2002; Vendrami et al. 2020). Over the last 50 years, extensive research has transformed the *M. edulis* species complex into a model system in ecophysiology and thermal biology (Bayne 1976; Koehn 1991; Lockwood et al. 2015). Temperature is especially important in determining the range limits for these species, Energy budget, specifically clearance rate (i.e. feeding rate- the rate at which a suspension feeder filters particles from the water), limits the ability of *M. edulis* to persist at higher temperatures (Chapter 2, Fly and Hilbish 2013). Inability to increase energy intake to meet rising metabolic costs causes an energy deficit, which limits population persistence above a given temperature. Upper thermal limits correlate strongly with range limits for these species, and those range limits are changing as average summer sea surface temperatures continue to rise (Jones et al. 2010; Fly et al. 2015).

Here we aim to investigate the level of standing genetic variation related to thermal tolerance traits related to energy budget. Using the *M. edulis* x *M. galloprovincialis* hybrid zone in southwest England as a model system we ask (i) what is the extent of hybridization and introgression among sites along the hybrid zone, and (ii) are there genetic variants associated with physiological traits related to the energy budget? Given that the two species have distinct thermal niches, and that survival at high temperature seems to be dependent on feeding rate and the ability to maintain a positive energy budget, we predict there will be several genetic variants associated with metabolic functions, particularly clearance rate.

4.2 Materials and Methods

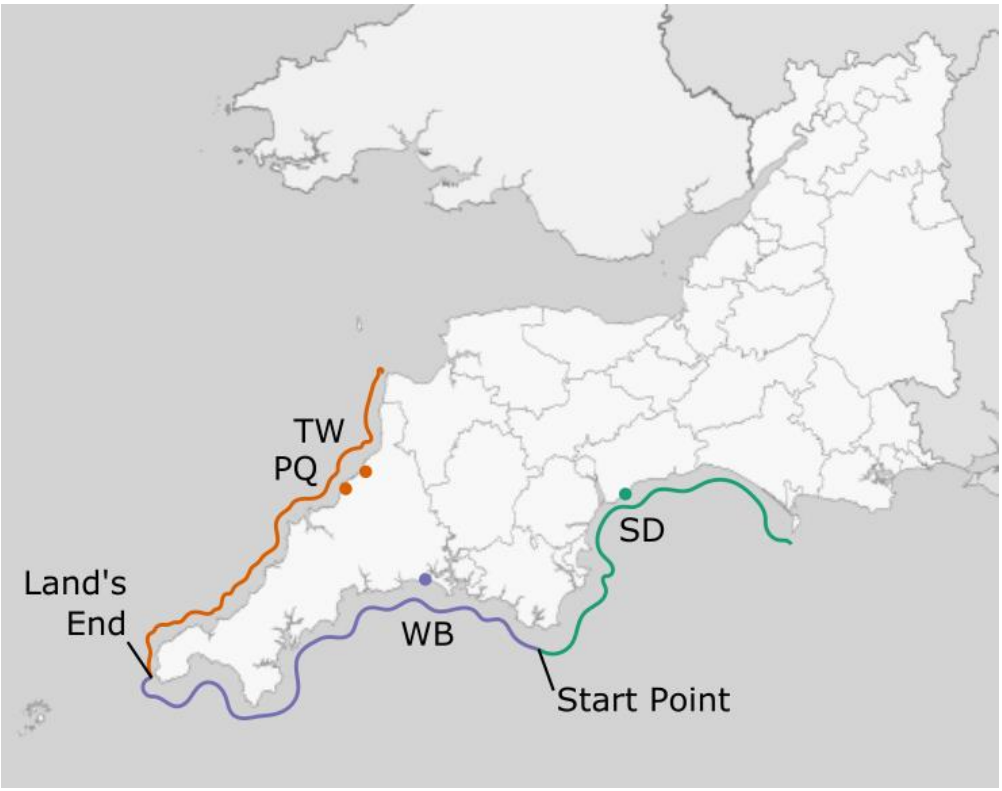

Figure 4.1: Map of collection sites in southwest England. Colors correspond to genetic regions: the region north of Land's End (orange) has *Mytilus galloprovincialis*, the region east of Start Point (green) has *M. edulis*, and the region between the two (purple) is a hybrid zone (Hilbish et al. 2002). *M. galloprovincialis* were collected from Port Quin (PQ) and Trebarwith (TW), *M. edulis* were collected from Sidmouth (SD), and putative hybrids were collected at Whitsand Bay (WB).

Mytilids in southwest England exist in both sympatry and allopatry. Mussels between Land's End (to the west) and Start Point (to the east) are putative hybrids, those north of Land's End are *M. galloprovincialis*, and those east of Start Point are *M. edulis* (Hilbish et al. 2002). Mussels for this study were collected from four sites: *M. galloprovincialis* (n = 28) were collected from Port Quinn (PQ) and Trebarwith (TW), *M. edulis* (n = 32) were collected from Sidmouth (SD), and putative hybrids (n = 30) from Whitsand Bay (WB) (Figure 1). Mussels were acclimated to laboratory conditions (salinity = 32 ppt, 15°C) and then subjected to a two-week thermal challenge at either 15°C or 23°C. The control temperature of 15°C represents the average summer sea

surface temperature (SST) in the region the mussels were collected. 23°C is both the thermal optimum for *M. galloprovincialis* and the upper thermal limit for population persistence in *M. edulis*. Few populations of *M. edulis* exist where summer SST exceeds 20°C, and none where it exceeds 23°C (Fly et al. 2015). Following the two-week acclimation period, respiration was measures via closed chamber respirometry and clearance rate was measure using the flow-through method (Filgueira et al. 2006). The complete experimental procedure can be found in chapter one of this dissertation.

A full description of the methods and parameters for RNA extraction through read filtering can be found in chapter one of this dissertation. Briefly, transcriptomes were generated from gill tissue of the 90 individuals and sequenced on an Illumina HiSeq 4000. The 150 bp paired end reads were quality filtered and mapped to the *M. galloprovincialis* reference genome (Gerdol et al. 2020) using STAR (Dobin et al. 2013) in 2-pass mode. Duplicates were marked with Picard (http://roadinstitue.github.io/picard) and BAM files were re-indexed accordingly. SNPs were called for each individual using the GATK haplotype caller (Van der Auwera and O'Connor 2020), and the resulting VCF files were merged. Only biallelic SNPs with no more than 70% missing data were retained using bcftools and vcftools (Danecek et al. 2011; Danecek et al. 2021). A high missing data allowance was chosen given the fact that SNPs were called from RNASeq data rather than RADSeq, meaning the presence/ absence of variants could potentially depend on expression. 70% was chosen to retain markers present in only one of the three genotype groups. SNPs were also filtered by depth, with a minimum depth of 10 and a maximum of twice the mean depth to filter out repetitive regions and multi-mappers.

Variants were then filtered with plink2 (https://www.cog-genomics.org/plink/2.0/) on minor allele count (20) and minor allele frequency (0.01) as recommended by plink2.

Ancestry coefficients were calculated using fastStructure (Raj et al. 2014) with 2-10 underlying populations, and the best model was chosen with their choose k tool. SNP associations were performed with plink2's GLM procedure, with the top two principal components (Price et al. 2006) and temperature as covariates. Respiration was $\log_{10}$ transformed to meet the assumption of normality, so the $\log_{10}$ of respiration and the untransformed clearance rate were used as the phenotypes (response variables). The resulting p-values were adjusted for multiple comparisons using --adjust-file in plink2.

4.3 Results

After quality and missing data filtering, 10,833 biallelic SNPs were retained. fastStructre detected two underlying populations, corresponding to an *M. galloprovincialis* group and an *M. edulis* group (Figure 2). Genetically, the two *M. galloprovincialis* sites were indistinguishable both in terms of ancestry assignment and in principal component space (Figures S1, S2), so they were visualized as one population.

4.3.1 Population structure

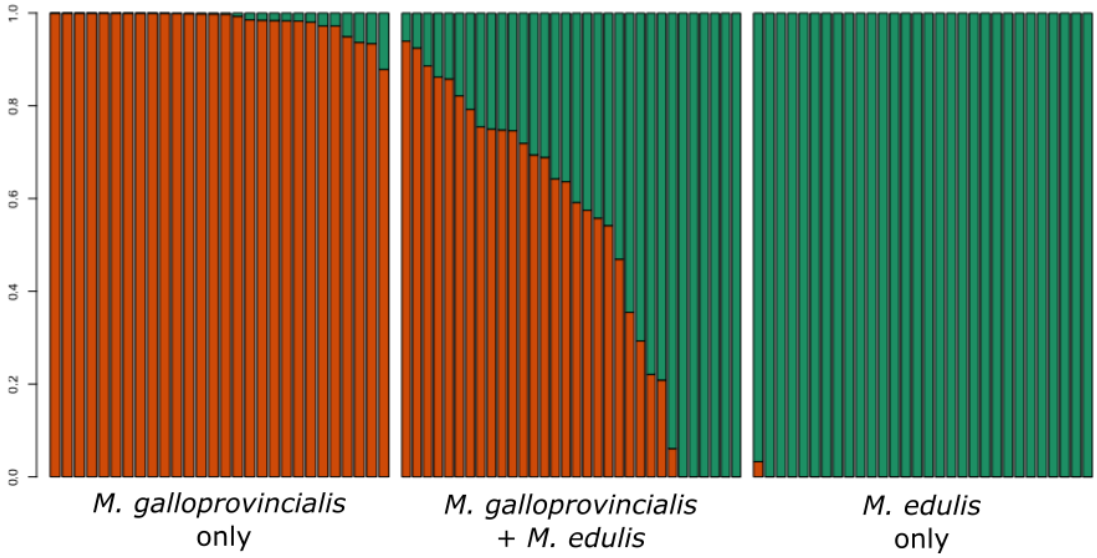

Figure 4.2: Ancestry coefficients by collection region. The proportion of either *M. edulis* or M. galloprovincialis ancestry assigned to each individual by collection region (Figure 1). *M. edulis* ancestry is represented in green, and *M. galloprovincialis* ancestry is represented in orange.

The fastStructure analysis separated individuals from *M. galloprovincialis* and *M. edulis* well, with only one individual from either species assigned more than 10% ancestry of the other. Putative hybrids had mostly mixed ancestry, though there were seven individuals with higher than 90% *M. edulis* ancestry, and two with >90% *M. galloprovincialis* ancestry (Fig. 2). *M. galloprovincialis* individuals were also separated from *M. edulis* in principal component space along PC1, which explained 62.4% of the total variance (Fig. 3). Hybrids were intermediate along PC 1, but more concentrated near *M. galloprovincialis* (except for the majority *M. edulis* individuals, which fall within the *M. edulis* cluster). PC 2, which explains 6.61% of the variance, separated *M. galloprovincialis* from the hybrids.

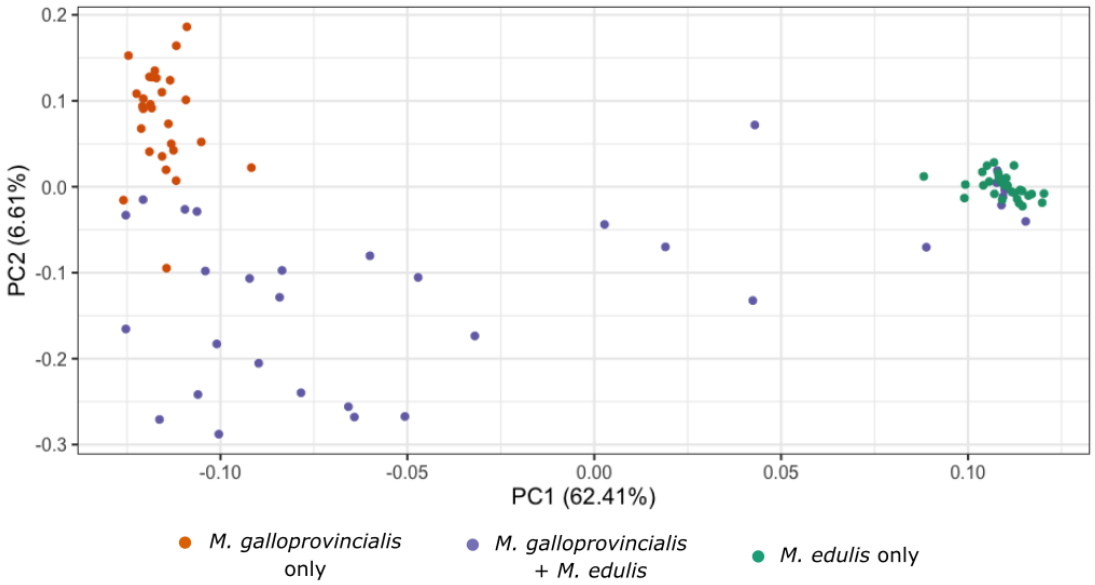

Figure 4.3: A principal components analysis (PCA) of the 10,833 SNPs, with individuals colored by collection region. Individuals collected from *M. edulis* sites (SD) cluster together very tightly and are separated from those collected from *M. galloprovincialis* sites along PC1 which explains 62.41% of the total variance. Individuals collected from sympatric sites (WB) were intermediate but concentrated closer to the *M. galloprovincialis* cluster. They are separated from the parent populations along PC2, which explains 6.61% or the variance.

4.3.2 SNP associations

After correcting for multiple comparisons, there were no SNPs significantly associated with either $\log_{10}$ respiration or clearance rate when analyzed using plink2's GLM. SNP association studies can suffer from low power due to the high number of comparisons, and corrections like Bonferroni are generally considered overly conservative (Chen et al. 2021). A significance threshold of $p = 5 \times 10^{-8}$ has become standard in GWAS studies to circumvent issues caused by multiple comparisons and linkage disequilibrium (Chen et al. 2021), however there were also no SNPs in our results that met this criterion for significance. Due to the exploratory nature of this study, we went forward with the uncorrected p-values to identify potential candidates, rather than strict hypothesis testing. In the uncorrected data, there were 116 SNPs associated with

respiration at the $\alpha = 0.01$ level, and 14 at the $\alpha = 0.001$ level. For clearance rate, there were 124 associations at the $\alpha = 0.01$ level and 12 at the $\alpha = 0.001$ level (Appendix C).

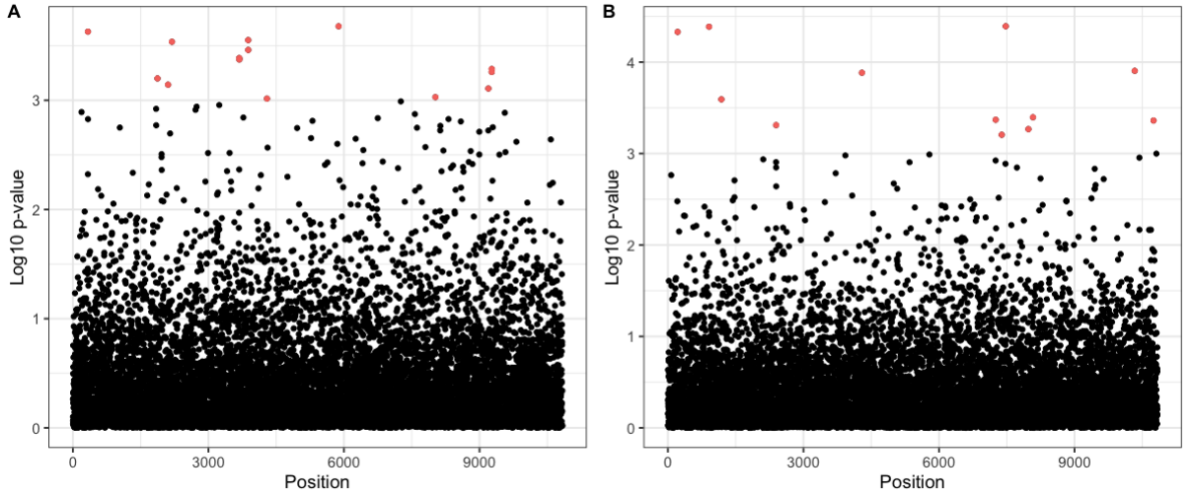

Figure 4.4: Manhattan-like plots showing SNP associations for $\log_{10}$ respiration (A) and clearance rate (B). The genome assembly for *M. galloprovincialis* is at the scaffold, not chromosome level, so SNPs are ordered by scaffold, and their position (in bp) along that scaffold. Points highlighted in red represent SNPs where $p < 0.001$ with the corresponding phenotype in the uncorrected GLM results.

All but one of the associated SNPs at $p < 0.001$ were located in coding regions; one SNP associated with respiration (UYJE01003802.1; 166295) was located just before the start of a coding sequence. Nine of the remaining SNPs (~35%) were in unidentified proteins (annotated as "hypothetical predicted protein" in NCBI). Sixteen SNPs (62%) across both respiration and clearance rate that were in coding regions of annotated proteins (Table 4.1).

Table 4.1: Twenty-six SNPs were associated with either $\log_{10}$ respiration or clearance rate with $p < 0.001$. The scaffold, the position of the SNP on that scaffold, the gene product, and the uncorrected p-value for each SNP are listed here. The full list of SNPs with $p < 0.01$can be found in table Appendix C.

Respiration

Scaffold	Position	p-value	Gene product
UYJE01000305.1	189974	2.35E-04	ATP-binding cassette, subfamily A (ABC1), member 3
UYJE01001871.1	15972	6.31E-04	all-trans-retinol dehydrogenase (NAD+)
UYJE01002067.1	218348	7.19E-04	A-kinase anchor protein 13/18/28
UYJE01002171.1	35685	2.91E-04	COP9 signalosome complex subunit 6
UYJE01003645.1	50152	4.10E-04	Hypothetical predicted protein
UYJE01003664.1	47887	4.23E-04	FERM, RhoGEF and pleckstrin domain protein 2
UYJE01003802.1	166295	2.81E-04	Hypothetical predicted protein
UYJE01003802.1	166440	3.46E-04	Hypothetical predicted protein
UYJE01004296.1	172183	9.66E-04	Hypothetical predicted protein
UYJE01005897.1	145024	2.10E-04	calcium/calmodulin-dependent protein kinase IV
UYJE01007938.1	123828	9.36E-04	solute carrier family 35 (GDP-fucose transporter)
UYJE01009018.1	61348	7.81E-04	Hypothetical predicted protein
UYJE01009100.1	116689	5.50E-04	period circadian protein
UYJE01009100.1	116932	5.16E-04	period circadian protein

Clearance Rate

Scaffold	Position	p-value	Gene product
UYJE01000179.1	84642	4.68E-05	cell cycle related kinase
UYJE01000973.1	160115	4.11E-05	Hypothetical predicted protein

UYJE01001196.1	63227	2.56E-04	proto-oncogene tyrosine-protein kinase Met
UYJE01002404.1	368500	4.89E-04	heterogeneous nuclear ribonucleoprotein U-like protein 1
UYJE01004296.1	169650	1.31E-04	N/A; repeat region
UYJE01007298.1	62712	4.27E-04	Hypothetical predicted protein
UYJE01007398.1	12486	6.23E-04	Hypothetical predicted protein
UYJE01007463.1	242045	4.05E-05	RAS protein activator-like 2
UYJE01007893.1	139482	5.41E-04	protein O-GlcNAcase / histone acetyltransferase
UYJE01007989.1	196418	4.01E-04	Hypothetical predicted protein
UYJE01010078.1	127106	1.24E-04	SAP domain-containing ribonucleoprotein
UYJE01010474.1	31655	4.36E-04	inositol 1,4,5-triphosphate receptor-associated cGMP kinase substrate

4.4 Discussion

In this work we aimed to (i) analyze the population structure of the *M. edulis* x *M. galloprovincialis* hybrid zone in southwest England, and (ii) detect genetic variants associated with traits contributing to energy budget. We found that divergence between individuals on either side of the hybrid zone is strong, with little evidence of alleles moving through the hybrid zone into either parent population. We detected few SNPs that were associated with either respiration or clearance rate, and none that were statistically significant after correcting for multiple comparisons.

Individuals collected from sites north of Land's End (PQ and TW; Figure 1) were assigned mostly *M. galloprovincialis* ancestry, and those collected from east of Start Point (SD) were nearly entirely *M. edulis*. This supports results from another recent

study, which showed that a population from Exmouth (east of Start Point, near SD) had little introgression of *M. galloprovincialis* alleles (Vendrami et al. 2020). This study included populations of pure *M. edulis* and *M. galloprovincialis* collected from their respective core ranges and far from potential hybrid zones, making it a good comparison to ground-truth our results. The putative hybrids had a range of ancestries, from very *M. galloprovincialis*-like to entirely *M. edulis* like (Figure 3). Overall, the ancestry analysis suggests minor introgression of *M. edulis* alleles through the hybrid zone and into the *M. galloprovincialis* region, but hardly, if any, of the reverse. This may be because this population is at the northern edge of *M. galloprovincialis*'s range limit, and *M. edulis* performs better at the temperatures currently experienced in this region (Chapter 2, Fly and Hilbish 2013). Other factors, including immunity (Tolman et al. 2019), are more likely to explain the persistence of *M. galloprovincialis* and hybrid populations in this region.

The SNPs associated with respiration and clearance are linked to a variety of functions, including cells signaling, cytoskeletal processes, innate immunity, regulating circadian rhythm, cell cycle control, growth and metabolism, and transcriptional regulation. Although many of the identified proteins may be related to either physiological process (i.e. period circadian rhythm and its role in day/night cycles, or SAP domain-containing ribonucleoprotein, which is associated with transcription and mRNA processing), it is important to note that these annotations are based on sequence homology to model organisms, often human, mouse, or *Drosophila*. This makes functional prediction difficult, as their function in *Mytilus* is unknown, and many proteins are well-studied only in the context of human disease.

GWA studies in wild populations of non-model organisms come with a host of challenges, including trouble with reproducibility, accounting for natural population structure, accounting for linkage in multiple testing correction, and difficulty identifying loci of small effect size (Santure and Garant 2018). The *M. galloprovincialis* reference genome only has scaffold-level, rather than chromosome-level, resolution, making finding and accounting for linkage nearly impossible. Coupled with the fact that complex, whole-organism level phenotypes are most often polygenic, with many contributing loci of small effect sizes (Ingvarsson and Street 2011; Healy et al. 2018; Santure and Garant 2018), it is not surprising that we were unable to detect SNPs significantly associated with either $\log_{10}$ respiration rate or clearance rate. So, while there are a significant number of SNPs, it remains unclear whether which, if any, are associated with the analyzed phenotypes that may aid in adaptation to rapidly changing conditions.

Lastly, it could be argued that modeling both populations together is inappropriate given the strong population structure shown by the PCA (Figure 3), and the sharp difference in variability between *M. edulis* (little within-population differentiation) and *M. galloprovincialis* (higher within-population differentiation). However, population structure was accounted for by including the top principal components in the model, and by using a high missing data threshold. Additionally, we assessed this possibility by splitting individuals by ancestry (individuals with > 90% ancestry for either *M. edulis* or *M. galloprovincialis* were analyzed separately). This method, along with using lower missing data thresholds for the full and partial data sets (data not shown) all failed to detect variants significantly associated with either phenotype at the standard GWAS significance threshold, even with as few as 527 variants. Thus, the failure to detect

significance is likely due to the independent and polygenic nature of complex traits mentioned above, rather than the chosen parameters or loss of power from multiple comparisons. Additionally, our power is limited by sample size, which is constrained by resources when measuring complex phenotypes like respiration. Future work should focus on increasing detection power by utilizing more individuals, and exploring other methods of SNP detection, such as restriction-site associated DNA sequencing (RADSeq) as seen in Vendrami et al. 2020.

CHAPTER 5

CONCLUSION

In this work I addressed four major objectives: **1)** to determine the mechanism by which *Mytilus edulis* is limited at high temperatures, **2)** to correlate the physiological and molecular responses to chronic high temperature in the *M. edulis* species complex, **3)** to categorize the response of *M. edulis* x *M. galloprovincialis* to chronic high temperature both physiologically and trascriptomically, and **4)** to assess the population structure of collected individuals and relate genetic variation to physiology.

While there was some evidence in North American populations of *Mytilus edulis* that clearance rate was the limiting factor at high temperatures, we now have strong evidence for this hypothesis in European populations as well. This is significant, as different thermal properties govern species ranges in North American and European population of these mussels. *M. edulis*, *M. galloprovincialis*, and their hybrids all consume more oxygen at 23°C than at 15°C, which was expected since metabolic demand increases with temperature (Figure 3.1). Only *M. galloprovincialis* and hybrids, however, increased their clearance rates to support their higher metabolisms (Figure 3.1). *M. edulis* decreased its clearance rate (Figure 3.1), meaning they would not have enough

energy available to persist at this temperature long-term. This supports the idea that clearance rate limits population persistence in *M. edulis* at high temperatures.

By using a weighted gene correlation network analysis, we were able to correlate physiological phenotypes with gene expression profiles (Figure 3.2). We showed that overall, groups of similarly expressed genes (modules) were correlated more strongly with respiration than with clearance rate. More importantly, however, we were able to show that gene modules were correlated with clearance rate in the same direction more often between *M. galloprovincialis* and hybrids than either with *M. edulis* (Figure 3.3). This result mirrors the physiological results in which *M. galloprovincialis* and hybrids increase their clearance rates from 15°C to 23°C, while *M. edulis* decreases theirs (Figure 3.1).

When evaluating the response of *M. edulis* x *M. galloprovincialis* to chronic high temperature, we found that their differentially expressed genes were largely unique from either parent species, and no more similar to either parent than the parent were to each other (Figure 2.2). Physiologically, hybrids had a more *M. galloprovincialis*-like response when looking at clearance rate, though they had a much higher clearance rate at 23°C than even the Mediterranean *M. galloprovincialis* (Figure 3.1). This transgressive response was also seen in the gene expression data, where hybrids had gene expression plasticity values well outside the range of the two parent species. Intermediate plasticity, on the other hand, was hardly observed (Figure 2.4). This evidence of hybrid vigor suggests that hybridization may pose a potential route for species persistence under climate change via creation of new phenotypes.

Lastly, we were able to show that in the sampled populations, regionally based genotype group assignments were nearly always accurate. Some individuals from the Whitsand Bay hybrid site were purely *M. edulis*, excluding those, individuals showed a gradient of ancestry makeup from mostly *M. galloprovincialis*-like to mostly *M. edulis*-like. We failed to associate any genetic variants with the metabolic phenotypes measured, but this was likely due to the complex nature of respiration and clearance rate. Complex phenotypes like these are generally highly polygenic, contributed to by many loci of small effect sizes, rather than determined by one or a few loci with large effect sizes. This makes detecting SNPs that are significantly associated with either phenotype statistically very difficult.

REFERENCES

Andrews S. 2010. FastQC: A Quality Control tool for High Throughput Sequence Data. Babraham Bioinformatics. [accessed 2020 Jan 23]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anestis A, Lazou A, Pörtner HO, Michaelidis B. 2007. Behavioral, metabolic, and molecular stress responses of marine bivalve *Mytilus galloprovincialis* during long-term acclimation at increasing ambient temperature. Am J Physiol Reg I. 293(2):R911–R921. doi:10.1152/ajpregu.00124.2007.

Angilletta M. 2009. Thermal Adaptation: A Theoretical and Empirical Synthesis. Oxford University Press. [accessed 2019 Mar 28]. https://asu.pure.elsevier.com/en/publications/thermal-adaptation-a-theoretical-and-empirical-synthesis.

Artigaud S, Richard J, Thorne MA, Lavaud R, Flye-Sainte-Marie J, Jean F, Peck LS, Clark MS, Pichereau V. 2015. Deciphering the molecular adaptation of the king scallop (*Pecten maximus*) to heat stress using transcriptomics and proteomics. BMC Genomics. 16(1). doi:10.1186/s12864-015-2132-x. http://www.biomedcentral.com/1471-2164/16/988.

Barrett R, Schluter D. 2008. Adaptation from standing genetic variation. Trends Ecol Evol. 23(1):38–44. doi:10.1016/j.tree.2007.09.008.

Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR. 2013. Genomic basis for coral resilience to climate change. PNAS. 110(4):1387–1392. doi:10.1073/pnas.1210224110.

Bayne BL. 1976. Marine Mussels: Their Ecology and Physiology. Cambridge University Press.

Bitter MC, Kapsenberg L, Gattuso J-P, Pfister CA. 2019. Standing genetic variation fuels rapid adaptation to ocean acidification. Nat Commun. 10(1):5821. doi:10.1038/s41467-019-13767-1.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Braby CE. 2006. Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus *Mytilus*). J Exp Biol. 209(13):2554–2566. doi:10.1242/jeb.02259.

Brannock P, Wethey D, Hilbish T. 2009. Extensive hybridization with minimal introgression in *Mytilus galloprovincialis* and *M. trossulus* in Hokkaido, Japan. Mar Ecol Prog Ser. 383:161–171. doi:10.3354/meps07995.

Buckley LB, Urban MC, Angilletta MJ, Crozier LG, Rissler LJ, Sears MW. 2010. Can mechanism inform species' distribution models? Ecol Lett. 13(8):1041–1054. doi:10.1111/j.1461-0248.2010.01479.x.

Campbell DR, Waser NM. 2001. Genotype-by-environment interaction and the fitness of plant hybrids in the wild. Evolution. 55(4):669–676.

Campbell-Staton SC, Link to external site this link will open in a new window, Winchell KM, Rochette NC, Fredette J, Inbar M, Schweizer RM, Julian C, Link to external site this link will open in a new window. 2020. Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. Nat Ecol Evol. 4(4):652–658. doi:http://dx.doi.org/10.1038/s41559-020-1131-8.

Chen N, Huang Z, Lu C, Shen Y, Luo X, Ke C, You W. 2019. Different transcriptomic responses to thermal stress in heat-tolerant and heat-sensitive Pacific abalones indicated by cardiac performance. Front Physiol. 9(1895). doi:10.3389/fphys.2018.01895. [accessed 2021 Jul 21]. https://www.frontiersin.org/articles/10.3389/fphys.2018.01895/full.

Chen Z, Boehnke M, Wen X, Mukherjee B. 2021. Revisiting the genome-wide significance threshold for common variant GWAS. De Koning D-J, editor. G3. 11(2). doi:10.1093/g3journal/jkaa056. [accessed 2022 Jan 28]. https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkaa056/6080665.

Chevin L-M, Lande R, Mace GM. 2010. Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. Kingsolver JG, editor. PLoS Biol. 8(4):e1000357. doi:10.1371/journal.pbio.1000357.

Clark MS, Thorne MAS, Amaral A, Vieira F, Batista FM, Reis J, Power DM. 2013. Identification of molecular and physiological responses to chronic environmental challenge in an invasive species: the Pacific oyster, *Crassostrea gigas*. Ecol Evol. 3(10):3283–3297. doi:https://doi.org/10.1002/ece3.719.

Collins M, Clark MS, Spicer JI, Truebano M. 2021. Transcriptional frontloading contributes to cross-tolerance between stressors. Evolutionary Applications. 14(2):577–587. doi:https://doi.org/10.1111/eva.13142.

Collins M, Sutherland M, Bouwer L, Cheong S-M, Frölicher T, Jacot Des Combes H, Koll Roxy M, Losada I, Ratter B, Rivera-Arriaga E, et al. 2019. Extremes, abrupt

changes and managing risk. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. p. 591–655.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2):giab008. doi:10.1093/gigascience/giab008.

Dimitriadis VK, Gougoula C, Anestis A, Pörtner HO, Michaelidis B. 2012. Monitoring the biochemical and cellular responses of marine bivalves during thermal stress by using biomarkers. Mar Environ Res. 73:70–77. doi:10.1016/j.marenvres.2011.11.004.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15–21. doi:10.1093/bioinformatics/bts635.

Doolittle WF. 2018. We simply cannot go on being so vague about 'function.' Genome Biol. 19(1):223. doi:10.1186/s13059-018-1600-4.

Dowd WW, King FA, Denny MW. 2015. Thermal variation, thermal extremes and the physiological performance of individuals. J Exp Biol. 218(12):1956–1967. doi:10.1242/jeb.114926.

Filgueira R, Labarta U, Fernandez-Reiriz MJ. 2006. Flow-through chamber method for clearance rate measurements in bivalves: design and validation of individual chambers and mesocosm: CR protocol for validation. Limnol Oceanogr Methods. 4(8):284–292. doi:10.4319/lom.2006.4.284.

Fly EK, Hilbish TJ. 2013. Physiological energetics and biogeographic range limits of three congeneric mussel species. Oecologia. 172(1):35–46. doi:10.1007/s00442-012-2486-6.

Fly EK, Hilbish TJ, Wethey DS, Rognstad RL. 2015. Physiology and biogeography: the response of European mussels (*Mytilus spp.*) to climate change. Am Malacol Bull. 33(1):136–149. doi:10.4003/006.033.0111.

Fox J, Weisberg S. 2019. An {R} Companion to Applied Regression. Third. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, Venier P, Naranjo-Ortiz MA, Murgarella M, Greco S, et al. 2020. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. Genome Biol. 21(1):275. doi:10.1186/s13059-020-02180-3.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature. 525(7569):372–375. doi:10.1038/nature15256.

Gleason LU, Burton RS. 2015. RNA-seq reveals regional differences in transcriptome response to heat stress in the marine snail *Chlorostoma funebralis*. Mol Ecol. 24(3):610–627. doi:10.1111/mec.13047.

Goulletquer P. 2004. Cultured Aquatic Species Information Programme: *Mytilus edulis* (Linnaeus, 1758).

Gunderson AR, Leal M. 2016. A conceptual framework for understanding thermal constraints on ectotherm activity with implications for predicting responses to global change. Ecol Lett. 19(2):111–120. doi:10.1111/ele.12552.

Gunderson AR, Stillman JH. 2015. Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. Proc R Soc B. 282(1808):20150401. doi:10.1098/rspb.2015.0401.

Healy TM, Brennan RS, Whitehead A, Schulte PM. 2018. Tolerance traits related to climate change resilience are independent and polygenic. Glob Chang Biol. 24(11):5348–5360. doi:10.1111/gcb.14386.

Helmuth B, Mieszkowska N, Moore P, Hawkins SJ. 2006. Living on the Edge of Two Changing Worlds: Forecasting the Responses of Rocky Intertidal Ecosystems to Climate Change. Annu Rev Ecol Evol Syst. 37(1):373–404. doi:10.1146/annurev.ecolsys.37.091305.110149.

Hewitt GM. 1988. Hybrid zones-natural laboratories for evolutionary studies. Trends Ecol Evol. 3(7):158–167. doi:10.1016/0169-5347(88)90033-X.

Hilbish T, Carson E, Plante J, Weaver L, Gilg M. 2002. Distribution of *Mytilus edulis, M . galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. Mar Biol. 140(1):137–142. doi:10.1007/s002270100631.

Hilbish TJ, Bayne BL, Day A. 1994. Genetics of physiological differentiation within the marine mussel genus *Mytilus*. Evolution. 48(2):267–286. doi:10.1111/j.1558-5646.1994.tb01311.x.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature. 470(7335):479–485. doi:10.1038/nature09670.

Huey RB, Kingsolver JG. 1993. Evolution of resistance to high temperature in ectotherms. Am Nat. 142:S21–S46. doi:10.1086/285521.

Huey RB, Kingsolver JG. 2019. Climate warming, resource availability, and the metabolic meltdown of ectotherms. Am Nat. 194(6). doi:10.1086/705679. [accessed 2019 Nov 7]. https://www.journals.uchicago.edu/doi/10.1086/705679.

Huey RB, Stevenson RD. 1979. Integrating Thermal Physiology and Ecology of Ectotherms: A Discussion of Approaches. Am Zool. 19(1):357–366. doi:10.1093/icb/19.1.357.

Ingvarsson PK, Street NR. 2011. Association genetics of complex traits in plants. New Phytol. 189(4):909–922. doi:10.1111/j.1469-8137.2010.03593.x.

Jones SJ, Lima FP, Wethey DS. 2010. Rising environmental temperatures and biogeography: poleward range contraction of the blue mussel, *Mytilus edulis* L., in the western Atlantic. J Biogeogr. 37(12):2243–2259. doi:10.1111/j.1365-2699.2010.02386.x.

Kingsolver JG, Buckley LB. 2017. Evolution of plasticity and adaptive responses to climate change along climate gradients. Proc R Soc B. 284(1860):20170386. doi:10.1098/rspb.2017.0386.

Klepsatel P, Gáliková M, Xu Y, Kühnlein RP. 2016. Thermal stress depletes energy reserves in *Drosophila*. Sci Rep. 6(1):33667. doi:10.1038/srep33667.

Koehn RK. 1991. The genetics and taxonomy of species in the genus *Mytilus*. Aquaculture. 94(2–3):125–145. doi:10.1016/0044-8486(91)90114-M.

Kooijman SALM. 2010. Dynamic Energy Budget Theory for Metabolic Organisation. 3rd ed. Cambridge University Press.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 9(1):559. doi:10.1186/1471-2105-9-559.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357–359. doi:10.1038/nmeth.1923.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30(7):923–930. doi:10.1093/bioinformatics/btt656.

Lockwood BL, Connor KM, Gracey AY. 2015. The environmentally tuned transcriptomes of *Mytilus* mussels. J Exp Biol. 218(12):1822–1833. doi:10.1242/jeb.118190.

Lockwood BL, Sanders JG, Somero GN. 2010. Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. J Exp Biol. 213(20):3548–3558. doi:10.1242/jeb.046094.

Logan ML, Cox CL. 2020. Genetic Constraints, Transcriptome Plasticity, and the Evolutionary Response to Climate Change. Front Genet. 11:538226. doi:10.3389/fgene.2020.538226.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550. doi:10.1186/s13059-014-0550-8.

Ma L, Bajic VB, Zhang Z. 2013. On the classification of long non-coding RNAs. RNA Biol. 10(6):924–933. doi:10.4161/rna.24604.

MacManes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. Front Genet. 5(13). doi:10.3389/fgene.2014.00013. [accessed 2020 Jan 23]. https://www.frontiersin.org/articles/10.3389/fgene.2014.00013/full.

Martin TL, Huey RB. 2008. Why "suboptimal" is optimal: Jensen's inequality and ectotherm thermal preferences. Am Nat. 171(3):E102–E118. doi:10.1086/527502.

Menge BA, Branch G. 2001. Rocky Intertidal Communities. In: Marine Community Ecology. Sinaur Associates. p. 221–251.

Narum SR, Campbell NR. 2015. Transcriptomic response to heat stress among ecologically divergent populations of redband trout. BMC Genomics. 16(1):103. doi:10.1186/s12864-015-1246-5.

Nisbet RM, Muller EB, Lika K, Kooijman S a. LM. 2000. From molecules to ecosystems through dynamic energy budget models. J Anim Ecol. 69(6):913–926. doi:10.1111/j.1365-2656.2000.00448.x.

Oliver ECJ, Benthuysen JA, Darmaraki S, Donat MG, Hobday AJ, Holbrook NJ, Schlegel RW, Sen Gupta A. 2021. Marine heatwaves. Annu Rev Mar Sci. 13(1):313–342. doi:10.1146/annurev-marine-032720-095144.

Oostra V, Saastamoinen M, Zwaan BJ, Wheat CW. 2018. Strong phenotypic plasticity limits potential for evolutionary responses to climate change. Nat Commun. 9(1):1005. doi:10.1038/s41467-018-03384-9.

Pascoe P, Parry H, Hawkins A. 2009. Observations on the measurement and interpretation of clearance rate variations in suspension-feeding bivalve shellfish. Aquat Biol. 6:181–190. doi:10.3354/ab00123.

Pavey SA, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. Ann NY Acad Sci. 1206:110–129. doi:10.1111/j.1749-6632.2010.05765.x.

Pinsky ML, Selden RL, Kitchel ZJ. 2020. Climate-driven shifts in marine species ranges: scaling from organisms to communities. Annu Rev Mar Sci. 12(1):153–179. doi:10.1146/annurev-marine-010419-010916.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38(8):904–909. doi:10.1038/ng1847.

Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics. 197(2):573–589. doi:10.1534/genetics.114.164350.

Rawson PD, Agrawal V, Hilbish T. 1999. Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. Mar Biol. 134:201–211.

Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, Puechmaille SJ, Novella-Fernandez R, Alberdi A, Manel S. 2019. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. Proc Natl Acad Sci USA. 116(21):10418–10423. doi:10.1073/pnas.1820663116.

Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and speciation. Heredity. 83(4):363–372. doi:10.1038/sj.hdy.6886170.

Riisgård HU. 2001. On measurement of filtration rates in bivalves—the stony road to reliable data: review and interpretation. Mar Ecol Prog Ser. 211:275–291.

Riisgård HU, Kittner C, Seerup DF. 2003. Regulation of opening state and filtration rate in filter-feeding bivalves (Cardium edule, Mytilus edulis, Mya arenaria) in response to low algal concentration. J Exp Mar Biol Ecol. 284(1):105–127. doi:10.1016/S0022-0981(02)00496-3.

Rivera HE, Aichelman HE, Fifer JE, Kriefall NG, Wuitchik DM, Wuitchik SJS, Davies SW. 2021. A framework for understanding gene expression plasticity and its influence on stress tolerance. Mol Ecol. 30(6):1381–1397. doi:https://doi.org/10.1111/mec.15820.

Santure AW, Garant D. 2018. Wild GWAS—association mapping in natural populations. Molecular Ecology Resources. 18(4):729–738. doi:10.1111/1755-0998.12901.

Scriber JM. 2014. Climate-driven reshuffling of species and genes: potential conservation roles for species translocations and recombinant hybrid genotypes. Insects. 5:1–61.

Seehausen O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. J Evolution Biol. 26(2):279–281. doi:10.1111/jeb.12026.

Shields JL, Barnes P, Heath DD. 2008. Growth and survival differences among native, introduced and hybrid blue mussels (*Mytilus* spp.): genotype, environment and interaction effects. Mar Biol. 154(5):919–928. doi:10.1007/s00227-008-0985-0.

Sinclair BJ, Marshall KE, Sewell MA, Levesque DL, Willett CS, Slotsbo S, Dong Y, Harley CDG, Marshall DJ, Helmuth BS, et al. 2016. Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? Ecol Lett. 19(11):1372–1385. doi:10.1111/ele.12686.

Sirovy KA, Johnson KM, Casas SM, Peyre JFL, Kelly MW. 2021. Lack of genotype-by-environment interaction suggests limited potential for evolutionary changes in plasticity

in the eastern oyster, *Crassostrea virginica*. Mol Ecol. 30(22):5721–5734. doi:10.1111/mec.16156.

Skibinski D, Beardmore J, Cross T. 1983. Aspects of the population genetics of *Mytilus* (Mytilidae: Mollusca) in the British Isles. Biol J Linn Soc. 19(2):137–183.

Skibinski F, Ahmad M, Beardmore JA. 1978. Genetic evidence for naturally occurring hybrids between *Mytilus edulis* and *Mytilus galloprovincialis*. Evolution. 32(2):354–364.

Sokolova I. 2021. Bioenergetics in environmental adaptation and stress tolerance of aquatic ectotherms: linking physiology and ecology in a multi-stressor landscape. J Exp Biol. 224(Suppl 1):jeb236802. doi:10.1242/jeb.236802.

Sokolova IM. 2013. Energy-Limited Tolerance to Stress as a Conceptual Framework to Integrate the Effects of Multiple Stressors. Integr Comp Biol. 53(4):597–608. doi:10.1093/icb/ict028.

Sokolova IM, Frederich M, Bagwe R, Lannig G, Sukhotin AA. 2012. Energy homeostasis as an integrative tool for assessing limits of environmental stress tolerance in aquatic invertebrates. Mar Environ Res. 79:1–15. doi:10.1016/j.marenvres.2012.04.003.

Somero GN. 2011. Comparative physiology: a "crystal ball" for predicting consequences of global change. Am J Physiol Reg I. 301(1):R1–R14. doi:10.1152/ajpregu.00719.2010.

Sommer RJ. 2020. Phenotypic plasticity: from theory and genetics to current and future challenges. Genetics. 215(1):1–13. doi:10.1534/genetics.120.303163.

Stephens M. 2017. False discovery rates: a new deal. Biostat. 18(2):275–294. doi:10.1093/biostatistics/kxw041.

Sun W, Feng J. 2018. Differential lncRNA expression profiles reveal the potential roles of lncRNAs in antiviral immune response of *Crassostrea gigas*. Fish Shellfish Immun. 81:233–241. doi:10.1016/j.fsi.2018.07.032.

Tallmon D, Luikart G, Waples R. 2004. The alluring simplicity and complex reality of genetic rescue. Trends in Ecology & Evolution. 19(9):489–496. doi:10.1016/j.tree.2004.07.003.

Thompson KA, Urquhart-Cronish M, Whitney KD, Rieseberg LH, Schluter D. 2021. Patterns, predictors, and consequences of dominance in hybrids. Am Nat. 197(3):17.

Tolman D, Wood HL, Skibinski DOF, Truebano M. 2019. Differential immunity as a factor influencing mussel hybrid zone structure. Mar Biol. 166(12):151. doi:10.1007/s00227-019-3604-3.

Van der Auwera G, O'Connor B. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition).

Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends in Genetics. 30(8):348–355. doi:10.1016/j.tig.2014.06.001.

Veilleux HD, Ryu T, Donelson JM, Ravasi T, Munday PL. 2018. Molecular response to extreme summer temperatures differs between two genetically differentiated populations of a coral reef fish. Front Mar Sci. 5(349). doi:10.3389/fmars.2018.00349. [accessed 2021 Jul 21]. https://www.frontiersin.org/articles/10.3389/fmars.2018.00349/full.

Vendrami DLJ, De Noia M, Telesca L, Brodte E, Hoffman JI. 2020. Genome-wide insights into introgression and its consequences for genome-wide heterozygosity in the *Mytilus* species complex across Europe. Evol Appl. 13(8):2130–2142. doi:10.1111/eva.12974.

West-Eberhard M. 1989. Phenotypic plasticity and the origins of diversity. Annu Rev Ecol Evol Syst. 20:249–278.

Wethey DS, Woodin SA, Hilbish TJ, Jones SJ, Lima FP, Brannock PM. 2011. Response of intertidal populations to climate: Effects of extreme events versus long term change. Journal of Experimental Marine Biology and Ecology. 400(1–2):132–144. doi:10.1016/j.jembe.2011.02.008.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. Molecular Ecology. 15(5):1197–1211. doi:10.1111/j.1365-294X.2006.02868.x.

Woodin SA, Hilbish TJ, Helmuth B, Jones SJ, Wethey DS. 2013. Climate change, species distribution models, and physiological performance metrics: predicting when biogeographic models are likely to fail. Ecol Evol. 3(10):3334–3346. doi:10.1002/ece3.680.

Yao R-W, Wang Y, Chen L-L. 2019. Cellular functions of long noncoding RNAs. Nat Cell Biol. 21(5):542–551. doi:10.1038/s41556-019-0311-8.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11(R14):12

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 1

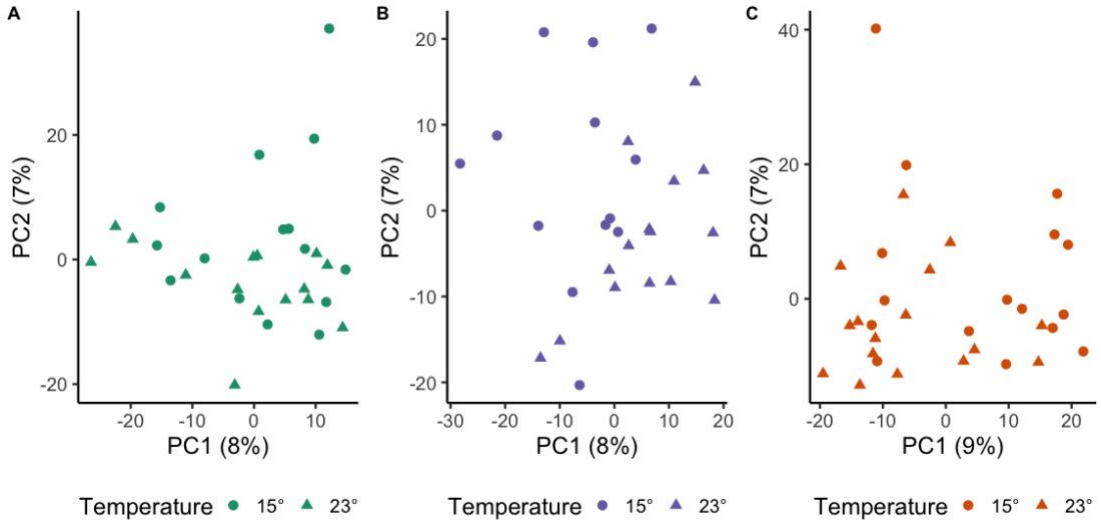

Figure A.1: Gene expression in the *M. edulis* species complex in response to chronic high temperature by genotype. Principal components analysis of the top 500 differentially expressed genes in (A) *M. edulis*, (B) hybrids, and (C) *M. galloprovincialis*. Circle points represent animals kept at 15°C, triangle points represent animals kept at 23°C.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

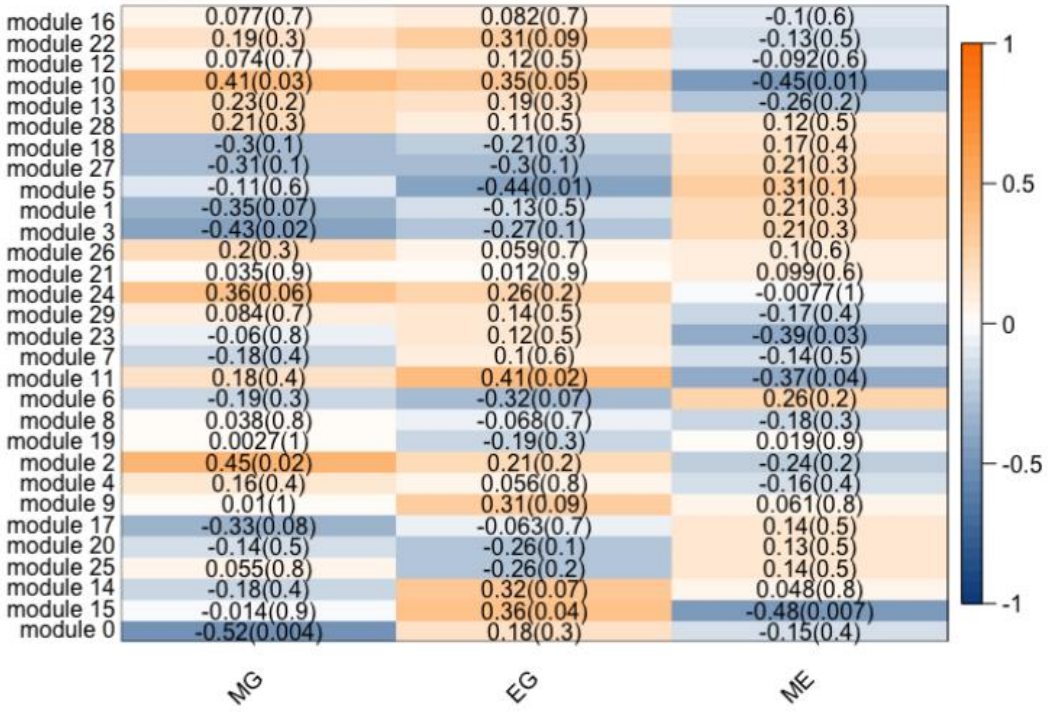

Figure B.1: Trait correlations for the eigengene of each gene module with clearance rate with correlation coefficients and p-values. The correlation coefficient is listed for each module, followed by the Fisher's asymptotic p-value. The column abbreviations indicate species: MG is *M. galloprovincialis*, EG are hybrids, ME is *M. edulis*. Darker orange boxes have stronger positive correlations, darker more blue boxes have stronger negative correlations.

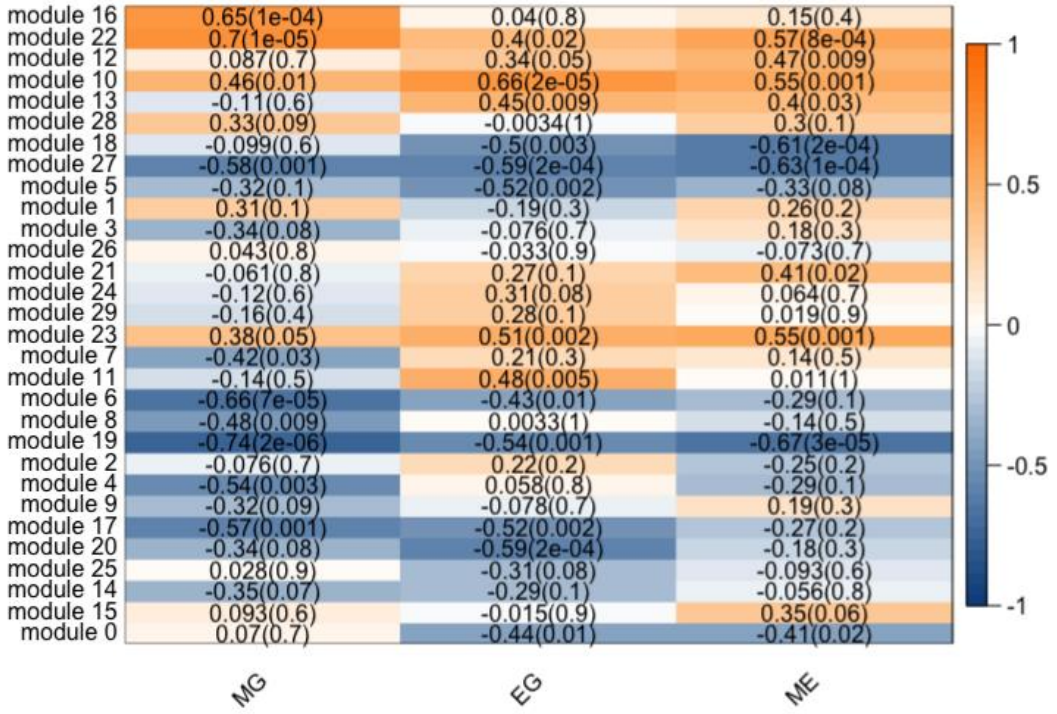

Figure B.2: Trait correlations for the eigengene of each gene module with the weight corrected $\log_{10}$ respiration rate with correlation coefficients and p-values. The correlation coefficient is listed for each module, followed by the Fisher's asymptotic p-value. The column abbreviations indicate species: MG is *M. galloprovincialis*, EG are hybrids, ME is *M. edulis*. Saturation represents the strength of the correlation, so darker orange boxes have stronger positive correlations, and darker more blue boxes have stronger negative correlations

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

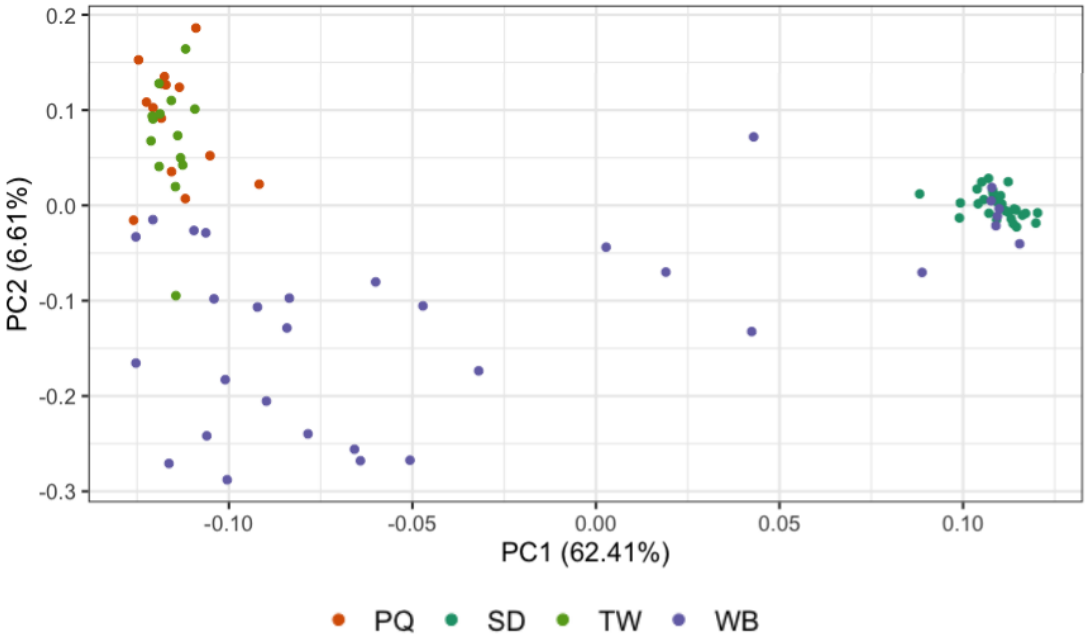

Figure C.1: Principal component analysis of the 10,833 biallelic SNPs colored by collection site. Port Quin (PQ) and Trebarwith (TW) are north of Land's End, where *M. galloprovincialis* occurs alone, Sidmouth (SD) is east of Start Point, where only *M. edulis* occurs, and Whitsand Bay (WB) is between the two, in the hybrid zon

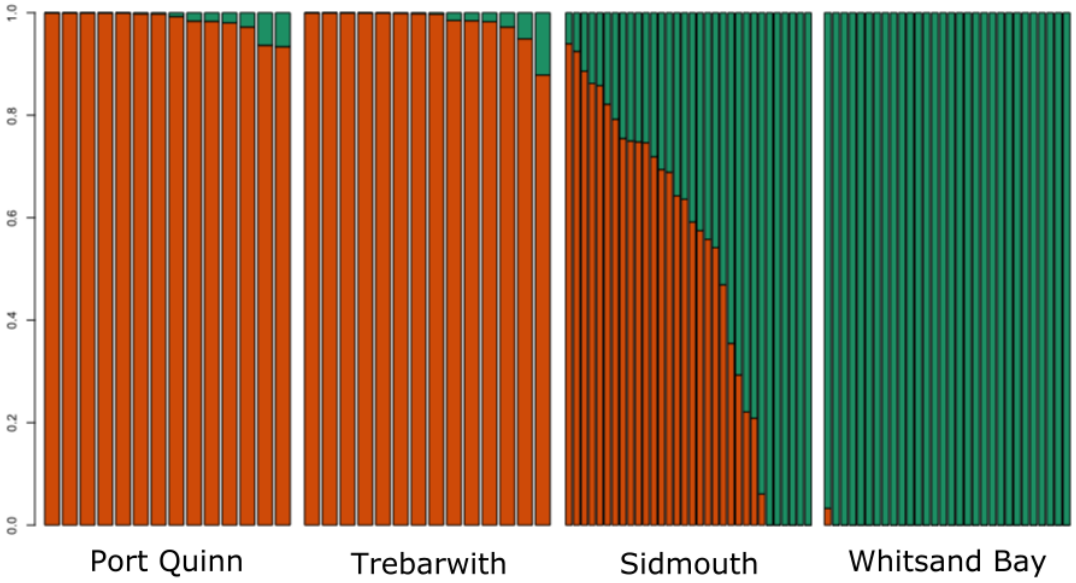

Figure C.2: Ancestry coefficients for each individual separated by collection site. *M. galloprovincialis* ancestry is shown in orange, and *M. edulis* ancestry is shown in green.

Table C.1: Complete list of the 244 SNPs association with respiration or clearance rate at the α = 0.01 level.

Respiration

Scaffold	Position	p-value
UYJE01000172.1	132157	1.28E-03
UYJE01000305.1	186768	4.75E-03
UYJE01000305.1	189974	2.35E-04
UYJE01000305.1	190016	1.49E-03
UYJE01000555.1	101678	6.52E-03
UYJE01000641.1	107959	7.50E-03
UYJE01000972.1	99355	8.85E-03
UYJE01001097.1	177615	1.78E-03
UYJE01001363.1	33331	4.61E-03
UYJE01001616.1	42412	7.44E-03
UYJE01001653.1	120838	5.90E-03
UYJE01001843.1	8354	1.20E-03
UYJE01001843.1	9361	1.69E-03
UYJE01001871.1	15972	6.31E-04
UYJE01001947.1	148816	3.32E-03
UYJE01001947.1	148894	3.11E-03
UYJE01001947.1	148903	8.30E-03
UYJE01001947.1	149547	4.35E-03
UYJE01001984.1	79899	8.41E-03
UYJE01002027.1	48582	7.31E-03
UYJE01002067.1	218348	7.19E-04
UYJE01002115.1	130474	2.01E-03
UYJE01002171.1	35685	2.91E-04
UYJE01002301.1	38857	6.38E-03
UYJE01002457.1	104718	8.24E-03
UYJE01002739.1	262653	1.22E-03
UYJE01002756.1	15090	1.14E-03
UYJE01002917.1	13146	5.55E-03
UYJE01002968.1	25762	3.04E-03
UYJE01003202.1	548362	7.32E-03
UYJE01003210.1	4801	7.00E-03
UYJE01003222.1	51898	1.11E-03
UYJE01003385.1	94098	4.45E-03
UYJE01003464.1	106745	3.03E-03
UYJE01003509.1	55155	5.56E-03
UYJE01003538.1	45039	6.67E-03
UYJE01003645.1	50056	4.30E-03
UYJE01003645.1	50152	4.10E-04

UYJE01003664.1	47887	4.23E-04
UYJE01003748.1	537368	1.44E-03
UYJE01003802.1	166295	2.81E-04
UYJE01003802.1	166440	3.46E-04
UYJE01003887.1	49930	9.24E-03
UYJE01004091.1	76858	6.41E-03
UYJE01004149.1	58509	4.85E-03
UYJE01004296.1	172183	9.66E-04
UYJE01004298.1	180235	2.71E-03
UYJE01004761.1	371110	5.02E-03
UYJE01004971.1	181745	1.80E-03
UYJE01005022.1	171219	8.95E-03
UYJE01005280.1	39411	2.22E-03
UYJE01005322.1	17809	1.54E-03
UYJE01005569.1	28696	3.90E-03
UYJE01005643.1	350856	3.71E-03
UYJE01005856.1	210692	2.51E-03
UYJE01005897.1	145024	2.10E-04
UYJE01005906.1	80531	5.40E-03
UYJE01005981.1	191300	6.25E-03
UYJE01006149.1	202270	8.97E-03
UYJE01006253.1	16959	2.25E-03
UYJE01006318.1	169894	9.09E-03
UYJE01006368.1	77315	7.61E-03
UYJE01006452.1	31642	3.78E-03
UYJE01006485.1	312934	2.86E-03
UYJE01006577.1	204702	9.16E-03
UYJE01006673.1	109092	9.76E-03
UYJE01006696.1	351781	7.60E-03
UYJE01006745.1	206511	6.38E-03
UYJE01006788.1	15378	8.25E-03
UYJE01006822.1	117490	9.91E-03
UYJE01006826.1	110017	1.46E-03
UYJE01006913.1	277489	3.64E-03
UYJE01007273.1	40341	4.19E-03
UYJE01007312.1	45104	1.02E-03
UYJE01007377.1	154964	8.50E-03
UYJE01007513.1	109089	1.34E-03
UYJE01007516.1	12149	6.74E-03
UYJE01007552.1	190523	1.78E-03
UYJE01007597.1	54233	6.26E-03
UYJE01007641.1	110470	8.54E-03
UYJE01007724.1	47350	2.68E-03
UYJE01007938.1	123828	9.36E-04
UYJE01008063.1	39751	1.89E-03
UYJE01008063.1	338700	1.71E-03
UYJE01008080.1	69870	4.47E-03
UYJE01008114.1	113760	2.89E-03
UYJE01008196.1	147963	1.49E-03
UYJE01008238.1	94975	4.16E-03
UYJE01008278.1	420823	3.95E-03
UYJE01008364.1	174732	6.62E-03
UYJE01008468.1	71940	1.56E-03
UYJE01008493.1	22969	6.30E-03
UYJE01008499.1	80160	8.02E-03
UYJE01008510.1	73868	4.09E-03
UYJE01008536.1	319357	4.51E-03
UYJE01008623.1	259411	3.67E-03
UYJE01008667.1	71537	2.92E-03
UYJE01008708.1	126193	3.84E-03
UYJE01008813.1	29598	3.14E-03
UYJE01008813.1	31345	1.94E-03
UYJE01009018.1	58895	1.89E-03
UYJE01009018.1	61348	7.81E-04
UYJE01009059.1	305342	7.96E-03
UYJE01009100.1	116689	5.50E-04
UYJE01009100.1	116932	5.16E-04
UYJE01009117.1	50217	5.45E-03
UYJE01009139.1	111083	1.77E-03
UYJE01009267.1	160488	3.15E-03
UYJE01009371.1	86637	1.30E-03
UYJE01009396.1	164698	2.98E-03
UYJE01009652.1	25955	2.40E-03
UYJE01009863.1	275881	8.66E-03
UYJE01010286.1	188058	5.96E-03
UYJE01010297.1	122077	2.28E-03
UYJE01010330.1	117001	5.70E-03
UYJE01010536.1	210564	8.61E-03

Clearance Rate		
Scaffold	**Position**	**p-value**
UYJE01000046.1	93126	1.72E-03
UYJE01000173.1	207080	3.32E-03
UYJE01000179.1	84642	4.68E-05
UYJE01000221.1	77195	7.12E-03
UYJE01000309.1	34604	4.78E-03
UYJE01000339.1	142018	4.81E-03
UYJE01000559.1	108063	6.37E-03
UYJE01000641.1	107959	6.16E-03
UYJE01000886.1	142646	3.81E-03
UYJE01000951.1	25631	7.66E-03
UYJE01000973.1	159802	4.38E-03
UYJE01000973.1	160115	4.11E-05
UYJE01000973.1	161651	4.81E-03
UYJE01001148.1	117042	9.11E-03
UYJE01001196.1	63227	2.56E-04
UYJE01001308.1	19901	5.64E-03
UYJE01001367.1	81318	6.59E-03
UYJE01001418.1	24781	3.25E-03
UYJE01001450.1	19820	3.01E-03
UYJE01001450.1	19871	1.97E-03
UYJE01001479.1	93527	5.04E-03
UYJE01001505.1	19294	8.88E-03
UYJE01001915.1	73897	3.86E-03
UYJE01002068.1	56553	1.16E-03
UYJE01002193.1	430790	6.71E-03
UYJE01002363.1	189625	8.50E-03
UYJE01002404.1	367947	6.57E-03
UYJE01002404.1	368091	1.24E-03
UYJE01002404.1	368127	1.42E-03
UYJE01002404.1	368500	4.89E-04
UYJE01002404.1	368988	2.28E-03
UYJE01002573.1	162347	6.47E-03
UYJE01002655.1	21118	7.20E-03
UYJE01002724.1	13462	3.57E-03
UYJE01002738.1	371297	3.83E-03
UYJE01002919.1	19639	8.83E-03
UYJE01002991.1	444755	4.13E-03
UYJE01002993.1	133834	5.36E-03
UYJE01003474.1	24142	3.40E-03
UYJE01003539.1	423239	8.63E-03
UYJE01003585.1	89042	7.54E-03
UYJE01003675.1	53937	1.64E-03
UYJE01003869.1	20790	1.05E-03
UYJE01004051.1	106001	2.88E-03
UYJE01004296.1	169650	1.31E-04
UYJE01004388.1	235506	9.47E-03
UYJE01004512.1	110076	4.53E-03
UYJE01004593.1	147934	8.94E-03
UYJE01004644.1	83152	6.68E-03
UYJE01004846.1	105832	8.01E-03
UYJE01005016.1	154951	2.13E-03
UYJE01005100.1	117915	2.42E-03
UYJE01005298.1	24383	7.34E-03
UYJE01005356.1	5058	1.24E-03
UYJE01005391.1	37625	6.59E-03
UYJE01005444.1	13751	3.89E-03
UYJE01005522.1	37236	8.28E-03
UYJE01005643.1	539782	8.49E-03
UYJE01005744.1	20739	5.48E-03
UYJE01005772.1	41955	1.02E-03
UYJE01005983.1	218560	6.97E-03
UYJE01006064.1	335207	5.19E-03
UYJE01006064.1	346635	3.68E-03
UYJE01006113.1	101924	3.76E-03
UYJE01006121.1	56505	3.89E-03
UYJE01006151.1	134818	5.05E-03
UYJE01006401.1	109350	5.19E-03
UYJE01006405.1	181751	8.88E-03
UYJE01006500.1	152643	8.79E-03
UYJE01006534.1	54580	9.28E-03
UYJE01006536.1	92260	3.79E-03
UYJE01006569.1	398454	8.32E-03
UYJE01006584.1	124083	8.46E-03
UYJE01006621.1	106920	8.88E-03
UYJE01006780.1	217512	3.17E-03
UYJE01006823.1	226719	4.01E-03

UYJE01006830.1	298397	3.62E-03	**UYJE01008672.1**	65362	3.30E-03
UYJE01006865.1	233882	5.45E-03	**UYJE01008672.1**	65422	3.30E-03
UYJE01006874.1	110340	4.98E-03	**UYJE01008672.1**	65986	3.30E-03
UYJE01007120.1	211545	5.00E-03	**UYJE01008672.1**	66025	3.30E-03
UYJE01007206.1	230056	8.70E-03	**UYJE01008732.1**	8671	4.51E-03
UYJE01007298.1	62673	1.19E-03	**UYJE01008810.1**	23360	9.99E-03
UYJE01007298.1	62712	4.27E-04	**UYJE01008897.1**	87694	8.14E-03
UYJE01007317.1	40234	6.58E-03	**UYJE01009183.1**	74116	3.09E-03
UYJE01007348.1	49503	3.05E-03	**UYJE01009232.1**	56669	6.58E-03
UYJE01007398.1	12486	6.23E-04	**UYJE01009246.1**	18541	8.56E-03
UYJE01007463.1	239345	1.29E-03	**UYJE01009268.1**	80780	2.43E-03
UYJE01007463.1	242045	4.05E-05	**UYJE01009268.1**	80834	1.47E-03
UYJE01007549.1	20691	6.05E-03	**UYJE01009268.1**	86818	2.43E-03
UYJE01007641.1	105079	1.42E-03	**UYJE01009269.1**	263745	2.21E-03
UYJE01007724.1	47414	9.24E-03	**UYJE01009461.1**	30588	1.90E-03
UYJE01007788.1	31718	5.41E-03	**UYJE01009724.1**	9084	8.43E-03
UYJE01007893.1	139482	5.41E-04	**UYJE01009949.1**	332620	6.06E-03
UYJE01007989.1	196418	4.01E-04	**UYJE01010078.1**	127106	1.24E-04
UYJE01008116.1	92444	4.18E-03	**UYJE01010180.1**	58498	1.11E-03
UYJE01008119.1	15405	1.88E-03	**UYJE01010180.1**	58539	6.81E-03
UYJE01008155.1	153015	3.65E-03	**UYJE01010372.1**	205770	6.84E-03
UYJE01008239.1	18281	7.60E-03	**UYJE01010421.1**	286658	6.82E-03
UYJE01008510.1	77008	9.34E-03	**UYJE01010474.1**	31655	4.36E-04
UYJE01008665.1	436213	9.87E-03	**UYJE01010538.1**	174698	1.00E-03