Summer 8-18-2023

Predicting Marine Teleost Responses to Ocean Warming and Pollution

Akila Harishchandra
University of Maine, kanahera.harishchandra@maine.edu

PREDICTING MARINE TELEOST RESPONSES TO

OCEAN WARMING AND POLLUTION

By

Akila Harishchandra

B.Sc. University of Ruhuna, 2011

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy

(in Marine Biology)

The Graduate School

The University of Maine

August 2023

Advisory Committee:

Dr. Nishad Jayasundara, Assistant Professor, Duke University (Co-chair)

Dr. Paul Rawson, Professor of Marine Biology, University of Maine (Co-chair)

Dr. Huijie Xue, Professor of Oceanography, Xiamen University

Dr. Jeffery Runge, Professor of Oceanography, University of Maine

Dr. Zheng Wei, Assistant Professor, Texas A&M University

PREDICTING MARINE TELEOST RESPONSES TO

OCEAN WARMING AND POLLUTION

By Akila Harishchandra

Dissertation Co-Advisors: Dr. Nishad Jayasundara and Dr. Paul Rawson

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Marine Biology)
August 2023

Ocean warming and pollution are two detrimental anthropogenic factors causing rapid marine ecosystem degradation recorded in the past decades. These factors alter the marine environment intolerable for many marine species, forcing them to either adapt or shift their contemporary habitat ranges to reduce the extinction risk embedded with environmental degradation. Estimating marine species' habitat range shifts, and their potential for developing adaptive mechanisms are critical for ecosystem conservation and management, human health risk assessment, and climate change vulnerability assessments. Given that, for the first chapter of this thesis, we focused on developing a species distribution model (SDM) integrating marine species temperature-sensitive physiological factor, into a bioclimate model to better predict future habitat patterns with warming. We integrated two omics datasets for the second and third chapters to determine the potential transcriptomic and epigenomic mechanisms underlying marine species' evolved resistance to extreme pollution.

We tested the new model to predict the future (the 2050s and 2080s) habitat ranges of the highly eurythermal intertidal minnow, Atlantic killifish (*Fundulus heteroclitus*), as a best-case scenario. Our SDM predicts complex and diverse habitat patterns for Atlantic killifish, including habitat fragmentation, migration between adjacent populations, and range contractions but no poleward range expansion. Our model predictions are quite unique compared to existing SDMs, mainly with the integration of thermal physiology into the model.

The molecular analysis in the second and third chapters posited the repeated desenstivity of the Aryl Hydrocarbon Receptor (AHR) pathway regulated through the downregulation of the *ahr2* gene. *ahr2* gene intron hypermethylation was also detected in a Polycyclic Aromatic Hydrocarbons (PAHs)-resistant killifish population, a potential novel molecular mechanism underlying killifish rapid adaptations to PAHs toxicity. Reduced lipid metabolism and mitochondrial respiration were also identified as other key molecular processes underlying the evolved PAHs resistance in Atlantic killifish.

Overall, the chapters of this thesis demonstrate the importance of integrating ectotherm physiology into SDMs to better predict their future habitat range shift patterns with ocean warming and the necessity of integrating different omics data to uncover the complex patterns of molecular mechanisms underlying marine organisms' evolved resistance to ubiquitous aquatic pollution

DEDICATION

This dissertation is dedicated to people working hard to build a better future for the world.

ACKNOWLEDGEMENTS

grateful to you for tolerating difficulties and sorrows and keeping my mind free for studies. You

all are awesome. I have no words to appreciate your dedication and commitment. Thank you

very much you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. AN INTRODUCTION TO THE OCEAN WARMING, POLLUTION, AND ITS IMPACT ON THE MARINE ECOSYSTEM

1.1   Ocean warming and pollution, past present, and the future.

The Earth and its ecosystems have been evolving continuously since their formation 4.5 billion years ago, but the rate of evolution has changed with modern human activities starting in the Cenozoic era (10000 years before the present, Wang et al., 2006), leading to Anthropocene (the most recent geological time which has been influenced by humans-based social activities). Regardless of several major climate cooling and warming events that happened in the Miocene (5.3-24 million years BP) and Pliocene (1.6-5.3 million years BP) epochs, global climate change mostly remained neutral (Wang et al., 2006) until the Anthropocene. The quality of the atmosphere, hydrosphere, and biosphere has been declining throughout the Anthropocene that started in the later 1700s or early 1800s (Crutzen, 2017; Steffen et al., 2017) leading to a range of climatic and non-climatic crises referred to as "global warming", "climate change", and "pollution".

Global warming is the gradual rise in the temperature of the Earth's atmosphere, due to the effect of increasing levels of greenhouse gases like water vapor, carbon dioxide, chlorofluorocarbons, and other pollutants like methane. Ice core-derived temperature records have found that the global temperature between the year 1000 to 1900 was below the 1961 -1990 average (Wang et al., 2006), witnessing the rapid warming that occurred in recent history. With the industrial revolution (IR) in the 18th century, the global energy requirement boosted several folds compared to before IR time, increasing the demand for fossil fuel (mainly coal and hydrocarbon) as a cheap and reliable energy source. Rapid combustion of fossil fuel caused the release of trillion tons of greenhouse gases into the Earth's atmosphere, contributing to rising the global mean surface

temperature 1°C above the pre-industrial period and is predicted to be 1.5°C above between 2030-2050 (Stocker et al., 2013).

The ocean that covers a large area of the Earth's surface acts as a buffer for global warming by absorbing a significant amount of excess greenhouse gases and heat but is highly vulnerable to increasing water temperature. The ocean covers nearly 71% of the Earth's surface and has absorbed around 90% of the total atmospheric energy (Bindoff et al., 2013; Rhein et al., 2013) during the last 5-6 decades (Levitus et al. 2012) and also stored 50 times more carbon than the atmosphere (Sabine et al., 2004), which is 30% of the total greenhouse gas emitted from fossil fuel combustion, deforestation, and other land-use activities (Mikaloff-Fletcher et al., 2006). Rapid fluctuations in the atmospheric radiative properties and heat budget (Bindoff et al., 2013), and observed, frequent climate anomalies (Rhein et al., 2013) are closely linked to the changes in the ocean temperature. The ocean surface (0-75m) and the deep-water column up to 700m had been warmed at alarming rates of 0.11 and 0.015 °C per decade respectively, between 1971-2010 (Rhein et al., 2013). The rate of ocean warming is strongly localized across different latitudes, longitudes, and depths, like the recorded differential rates in the Indian, Atlantic, and Pacific Oceans between 1950-2009 (0.65,0.41 and 0.31°C respectively, Rhein et al., 2013).

Ocean observation is critical to understand the past and present ocean temperature changes and as baselines for predicting the future states critical to improving preparedness. The Intergovernmental Panel for Climate Change (IPCC) is an international consortium responsible for predicting Earth's future climate by employing state-of-art coupled ocean-atmosphere-cryosphere-ecosystem models (ESMs). ESMs can simulate the future ocean states based on historical information on atmospheric forcing (i.e. greenhouse gases, aerosols, solar radiation, and volcanic eruptions) and nearly 30 ESMs have been employed to predict the Earth's climate

2

system at 2100 and beyond in the latest climate prediction project, Coupled Model Intercomparison Project phases 6 (CMIP6) that constructed on Shared Socioeconomic Pathways (SSP) scenarios. The Coupled Model Intercomparison Project phase 5 simulated the Earth's climate variability at 2100 and beyond, employing ESMs based on emission scenarios called Representative Concentration Pathways (RCPs). RCP2.6 is considered the "peak and decline" scenario where the radiative forcing reaches the peak of ~3 W/m^2 before 2100 and then declines. RCP4.5 and 6 are the intermediate "stabilization without overshoot" scenarios where the radiative forcing reaches 4.5 and 6 W/m^2 respectively but stabilizes after 2100. The worst-case 'business-as-usual' is the RCP8.5, the "rising" scenario where the radiative forcing exceeds 8.5 W/m^2 by 2100 and continues to rise for some amount of time (Moss et al. 2008).

The multi-ensemble means of the CMIP5 ESMs uniformly predict a rising trend of the ocean surface temperature in the near future (up to 2035) and the end of the 21st century compared to its reference period (1986-2005), (Stocker et al., 2013) which will be ~1 (RCP2.6) or >3°C (RCP8.5) warmer than the reference period (Collins et al., 2013). In addition, the water column between 0-1000m has also been predicted to warm 0.5°C (RCP2.6) or 1.5°C (RCP8.5) than the reference period at the end of the 21st century (Collins et al., 2013). The future predicted ocean temperatures also show a strong spatial heterogeneity, where the tropical to subtropical surface oceans and the Southern Ocean epipelagic water is predicted to warm faster than other regions (Collins et al., 2013). Overall, the average ocean temperature (surface to deep basins) will more likely be 1.8-3.3 °C warmer than the 1986-2005 average temperature by the end of the 21st century (Rhein et al., 2013).

Besides the climatic factors, non-climatic features like ocean pollution have also become a rising environmental crisis detrimental to the dynamics of the ocean environment and the inhabitants.

Ocean pollution, the change in the state of the ocean due to the release of hazardous anthropogenic wastes, unwanted materials, and liquids into the marine environment is considered a key environmental crisis in the modern world capable of destroying the marine ecosystem and hindering the ocean-human interconnectedness (Anderson et al., 2019). Ocean pollution is panoceanic, but the coastal zones, the transitional area between the land and the deep ocean are highly vulnerable to pollution (Vikas and Dwarakish, 2015). Coastal geo-morphologies like bays, estuaries, and embayment along the global coastlines have been polluted due to land-based wastewater discharges, industrial and agricultural wastes, and river runoffs (Landrigan et al., 2020).

Six broadly defined types of pollutants are considered key ocean pollution agents; plastic wastes, oil spills, mercury, manufactured chemicals, pesticides, and nutrients (Landrigan et al., 2020). The majority (>80%) of these pollutants have a terrestrial origin while some are marine-specific discharges (UNESCO, 2017). Unlike climate warming, which is at least partly an extension of a natural phenomenon, ocean pollution is an entirely human-induced disaster propelled in the Anthropocene (Crutzen, 2017; Steffen et al., 2017) that started to rise exponentially after the industrial age (Landrigan et al., 2020). Oil pollution, associated with the mining industry, was the dominant agent of ocean pollution during the 16th century (Crutzen, 2017; Steffen et al., 2017), but Persistent Organic Pollutants (POP) and radionuclides served as the key pollution agent in the later phase (in the mid-1900s, Alava, 2019). POP pollution was mainly adverse in temperate countries (the main hub of the industrial revolution) due to the excess manufacturing, storing, and marine disposal of POP (Alava, 2019; Blasius and Goodmanlowe,2008), like the extreme POP pollution recorded along the Northeast United States coastal sites, such as New Bradford Harbor, Newark, Elizabeth River, and British Point. These sites have extensively been

polluted by the components of POPs like Polycyclic Aromatic Hydrocarbons (PAHs), polychlorinated biphenyls (PCB), and Dioxin-like compounds (DCD) (Di Giulio and Clark, 2011; Reid et al., 2016; Whitehead et al., 2017) while the Elizabeth River, Virginia, USA recorded the world's highest sediment PAH concentrations (Di Giulio and Clark, 2015).

Past records showed that some specific chemicals dominated over others as the key pollution agent in different periods. After the POPs-dominant era, Organochlorine had emerged as the dominant ocean pollutant between the 1940s to 1980s, mainly due to poor disposal (Hoffman et al., 2002) and creating adverse risks of ecosystem and human health degradation (Carson et al., 2009). Given that, strict risk reduction and regulatory measures were amended to reduce organochlorine usage, but the legacy of the added organochlorine in the environment is still responsible for environmental deterioration and is expected to continue throughout this century (Alava,2019).

After the 'Organochlorine era', plastic has dominated marine pollution in the mid-19th century (Alava,2019). Plastic contamination in the marine environment was first observed in the 1970s (Moor, 2015), but rapidly become the dominant agent of marine pollution. Terrestrial plastics floating in the surface runoffs or in rivers end up on beaches, coastal environments, or deep oceans, and remain for decades or centuries. The annual average global plastic emission to the aquatic environment (rivers, lakes, and oceans) has ranged between 9-23 metric tons per year, since 2016 (Borelle et al., 2020; Liu et al., 2020). Natural (through the weathering-the process of plastic breaking down into smaller pieces called micro and nano plastics) or artificial removal of plastic are considered the potential solution for controlling plastic pollution. However, these methods are less effective in plastic removal, thus the plastic has become a 'poorly reversible' environmental pollutant (Macleod et al., 2021). Weathering breaks apart and transforms the

larger plastics into far-less observable tiny microscale plastics (micro-plastics), that can circulate through food webs, creating plastic pollution further 'irreversible' and adverse. Marine pollution occurs throughout the global oceans, but the ocean surface and the remote coastlines are much more vulnerable to plastic pollution, like in the north and south Pacific gyres, north and south Atlantic, and the Indian Ocean (Van Sebille et al., 2020). It has been estimated that nearly 0.3 million metric tons of plastics are floating with ocean currents (Van Sebille et al., 2020), and are weathered by extreme solar radiation, mechanical (wave and wind) disturbance, and temperature variability resulting in micro and nano plastics that sink down into the subsurface and depth (Macleod et al., 2021). Removing larger and micro-level plastics requires substantial monitoring and technical advances. With the rapid accumulation of plastic and the poor removal rates, the quantity of marine plastic debris in the natural environment has been rising exponentially over the past decades and is also predicted to be doubled by 2025 compared to the 2016 reference period in the 'business as usual' scenario (Macleod et al., 2021).

1.2 The ecological ramification of ocean warming and pollution on marine organisms

The physiological performance of marine organisms is directly linked to the physicochemical environment of the ocean, thus any changes in the ambient environment alter the organismal homeostasis. Ocean temperature is a key driver of marine organisms' physiology, ecology, evolutionary mechanisms (Castaneda, 2004), the biogeographic distribution (Somero et al., 2017) thus any change in the ocean temperature (more likely with ocean warming) directly impedes the organism's homeostasis and the biogeographical distribution. The alternations in the physicochemical homeostasis of the ocean pushes the inhabitant's environmental preference ranges towards the intolerable limits, resulting in adverse ecosystem impacts including species range shifts and extinctions.

1.2.1 Impacts of ocean warming on marine organisms

The rising ocean temperature is generally unfavorable to the whole marine ecosystem, but marine ectotherms are more vulnerable than endotherms due to their greater environmental temperature dependency. Marine ectotherms (species entirely depend on the environmental temperature to regulate the thermodynamics of the internal body cavity, Somero et al., 2017) are susceptible to ocean warming, but the degree of impact varies with the species' thermal history within and across generations. Some ectothermic species such as Atlantic killifish (Fangue et al., 2009) and Atlantic Cod (Righton et al., 2010) have a wider thermal tolerance range (thermal generalists/eurytherms) in contrast to species with narrower ranges (thermal specialists/ stenotherms) such as Atlantic notothenioids (Beers and Jayasundara,2015), and polar cod (Jacobsen and Ozhigin,2011). The breadth of the thermal tolerance is intra or inter-specific, depending on the population, size, age, phenotypic plasticity (Bowler and Terblanche 2008), and habitat-specific thermal variability (local adaptations) (Kuo and Sanford 2011).

The organismal performance as defined by fecundity, growth, metabolic rate, and mobility at the whole-organism level, heart rate, nerve conduction velocity, and enzyme activity at the physiological and biochemical level is highly correlated with the environmental temperature. Organismal performance increases gradually with the rising temperature, peaks at an intermediate temperature (physiologically optimum temperature T_{opt}) and then decreases with the further warming (Angilletta,2009) that is commonly visualized in the thermal performance curves (TPCs) (Shultz et al., 2011). With ocean warming, environmental temperature eventually exceeds the locally adapted T_{opt}, more likely leading to decrease the organism's performance, habitat range shifts or acclimation with the cost of tolerance (Jayasundara et al., 2017).

The shift in marine organisms' contemporary habitat ranges as a consequence of ocean warming (Aboim et al., 2010) is considered a key detrimental consequence of ocean warming and negatively affects the marine ecosystem dynamics and the fisheries. The surface isotherms have been migrating poleward with ocean warming, and the average boundaries of inhabiting marine species are likely to follow the direction and velocity of the isotherms (Burrows et al., 2011; Burrows et al., 2014) resulting in a poleward range shift of the contemporary biogeographical range. However, diverse range shifts patterns have been observed among marine species like the northward range shift of Humboldt squid (*Disidicus gigas*), American lobster (*Homarus americanus*), Cushion star (*Patiriella exigua*) (Pinksy et al., 2013) and the deepwater migration of North Sea demersal fish assemblages (Dulvy et al.2008) over the past few decades.

Several theories have been presented to explain at the physiological link between marine ectotherms' thermal dependence and geographic distribution, and the Oxygen and Capacity Limited Thermal Tolerance (OCLTT) (Pörtner,2010; Pörtner,2014; Pörtner & Farrel 2008; Pörtner & Knust 2007) is considered one of the leading theories. Theory predicts that marine ectotherms are evolutionary adapted to inhabit a geographical range that optimizes their maximum aerobic scope and any extreme change in the aquatic temperature will increase ectotherms' metabolic demand and decrease aerobic scope (Pörtner & Farrel 2008; Pörtner & Knust 2007) leading to a deficiency of oxygen supply to tissues (Deutsch et al., 2015, Pörtner & Knust 2007) creating sub-lethal conditions such as feeding restrictions altering the growth or reproduction (Gunderson & Leal 2016). Under such circumstances, ectotherms are required to acclimate to the increased habitat temperature to re-establish the optimum aerobic scope or shift the current habitat range seeking thermal refugia that optimize the evolutionary/locally adapted optimum aerobic scope limit. Despite several existing critiques that questioned the validity of

8

OCLTT (Clark et al., 2013), this theory provides a robust mechanistic framework to explain the geographical distribution of aquatic ectotherms.

However, it's important to note that marine organisms' responses to ocean warming are not limited to habitat range shifts. Many marine thermal generalist species have been able to acclimate their physiological functions to the increased habitat temperature, though physiological costs exist underlying the acclimation. The 'Temperature-size rule' (TSR) posits an inverse relationship between body size and the environmental temperature that has been uniquely observed in a wide range of ectotherms from unicellular bacteria to vertebrates (Atkinson,2004). It further predicts a potential size reduction of the global marine fish assemblage within a range of 14-24% by 2050 with ocean warming (Cheung et al., 2009; van Rijin et al., 2017), similar to the observed size reduction in the commercial fish species in the northern and Mediterranean oceans in the last few decades (Baudron et al., 2014; van Rijin et al., 2017).

1.2.2 Impacts of ocean pollution on marine organisms

Ocean pollution brings adverse impacts to the coastal community. Ocean pollutants like plastics, pesticides, agricultural effluents, and persistent organic pollutants have been actively perturbing marine species richness, and ecological dynamics directly and indirectly harming the ecosystem service provision to the adjacent human communities. Marine organisms' responses to pollution vary widely with the rate, intensity, and quantity of the pollution and the physiological, acclimatory and evolutionary capacity of the species in the community.

Mass mortality is a critical consequence of marine pollution. Historical records of fish mortality during the oil spills such as the Exxon Valdez oil spill (Peterson et al., 2003) and the Deepwater Horizon disaster (Crone and Tolstoy, 2010) are well-known examples that demonstrate the

9

ecological severity of marine oil pollution. The mortality rate of marine organisms is high around the vicinity of the oil spill, but the impact extends gradually as oil disperses with ocean currents. All forms and life stages of marine species are susceptible to oil pollution, but fish eggs and larvae are more vulnerable because of their small size, lower development of cell membranes and detoxification mechanisms, and high probability of direct exposure to contaminants during the free-floating stage at the surface (Langangen et al., 2017). The adversity of an oil spill depends on the compound and its concentration, even low concentrations of a highly toxic compound like polycyclic aromatic hydrocarbons (PAHs) can cause complete, sub-lethal (morphological abnormalities, reduced physiological rates, and increased risk of predation and starvation) or morality in fish eggs and larvae in the natural environment (Debruyn et al., 2007; Incardona et al., 2012) or controlled tank conditions (Hicken et al., 2011).

Some marine species show higher levels of resistance to pollution. Pollution resistance fundamentally depends on phenotypic plasticity, evolutionary and locally developed adaptations. A population capable of adjusting its phenotypic features (i.e.growth rate, metabolic rate, heart rate) to the pollution-modified environmental conditions can demonstrate higher survivorship. Phenotypic plasticity (PP) is a favorable, rapid firsthand organismal feature to cope with the extreme environmental fluctuations (Mundey et al., 2013). PP is the capacity of a given organism's genotype to deliver subordinate phenotypes with better feasibility to cope with the changing ambient environmental conditions (Donelson et al., 2019; West-Eberhard,2005). PP can produce morphological, physiological, or behavioral modifications in organisms with some potential to be transferred to the next generations (transgenerational plasticity). PP is an important protective mechanism for organisms for immediate survival to sustain the population in the changing environment, but cost and limitations always adjoin with PP.

Besides PP, organisms with adaptative capacities may have led to higher survival rates with ocean pollution. Adaptations develop through the inheritance of favorable phenotypes from one generation to another resulting in a greater fit of the population to the changing environment conferring shifts in the allele frequency (Donelson et al., 2019). Whether adaptations act for protection, the organismal capacity for adaptation is up for the debate with the current rates of global change (Reznick and Ghalambor, 2001), but some species exhibited rapid adaptations to chemical pollution in the marine environment like pest and pathogen adaptation to pesticides and pharmaceuticals (Reviewed in Palumbi, 2001; Reviewed in Whitehead et al., 2017). Atlantic killifish (*Fundulus heteroclitus*) rapid adaptations to extreme levels of Persistent Organic Pollutants (POPs) (Polycyclic Aromatic Hydrocarbons (PAHs) and Poly Chlorinated Biphenyls (PCBs)) contamination in the estuaries along the north-east of the United States (Reid et al., 2016; Whitehead et al., 2017) has also been considered a major scientific interest in the integrative environmental toxicology and human health risk assessment studies conducted over the past decades.

Bioaccumulation and biomagnification are considered two other key consequences of marine pollution. Toxic compounds enter the inhabitant's body tissues in different ways (i.e.through ingestion, through gills and skin) and can move up the trophic levels of the food web leading to bioaccumulation. It is considered a critical issue detrimental to higher tropic organisms including the human community. As pointed out above, there is a diverse range of ecological consequences of ocean warming and pollution existing under the current global change scenario. Some are extremely destructive while some pose a moderate risk.

1.3 Scope of the Study

The key focus of this research is to develop integrated methods to determine marine species' responses to ocean warming and pollution. The second chapter is dedicated to discussing a method to predict marine ectotherms' habitat range shifts with ocean warming. Here a novel species distribution model was developed integrating marine ectotherms' physiological responses to warming and predicted the range shifts of a highly eurythermal estuarine fish in the 2050s and 2080s. The third and fourth chapters were oriented toward developing an integrated approach to determine the molecular mechanisms underlying evolved pollution resistance in a highly-tolerant estuarine fish species. Two different multi-omics (transcriptomics-RNASeq, and methylomic-DNA-methylation), data were employed to investigate the molecular processes underlying the Atlantic killifish evolved pollution resistance in the last two chapters.

The key focus of the second chapter is to develop a species distribution model integrating marine ectotherms' physiological responses to better predict their future habitat distribution with ocean warming. The accurate prediction of potential habitat range shifts of marine species with ocean warming has become a timely necessity in several sectors like ecosystem management and conservation, and the fisheries industry. Species distribution models (SDMs) are the focal elements that have been developed and applied for predicting the future habitat range shifts of ecologically and economically significant marine species (Dhalke et al., 2015; Grieve et al., 2017; Kachner et al., 2019; Reygondeau & Beaugrand, 2011). Therefore, for the second chapter, we set up the overarching goal to develop a method to integrate an appropriate deterministic thermal physiological parameter into a conventional SDM and predict the future habitat range shifts of a highly eurythermal estuarine fish, Atlantic killifish (*Fundulus heteroclitus*, Linnaeus 1791) as the best-case scenario of marine ectotherms potential habitat range shifts with ocean

warming. Aerobic metabolism was selected as the most fitting physiological parameter to integrate into the bioclimate model.

The joint focus of the third and fourth chapters is to determine the modifications reflected at the molecular level underlying the pollution adaptations in Atlantic killifish inhabiting an extremely PAHs polluted estuarine environment on the northeast coast of the United States. Estimating the mechanisms underlying vertebrate pollution adaptation is critical for environmental conservation, assessing environmental and human health risks and societal vulnerability to pollution. For these two chapters, we integrated the results of genome-wide scans of transcriptomics, and methylomic signals reflected in pollution-resistant Atlantic killifish compared to a pollution-sensitive population.

With the development of cost-effective and high-precision sequencing techniques and access to high-performance computers, more opportunities have been emerging for in-depth analysis of molecular mechanisms underlying pollution adaptations in natural populations. Considering these advancements, for the third chapter, we sequenced the mRNA expression of a batch of Atlantic killifish subpopulation inhabited polycyclic Aromatic hydrocarbon (PAH)-polluted and non-polluted sites and compared to see the differential gene expression between two fish populations. For the fourth chapter, we sequenced sodium bisulfite-treated cellular DNA fragments from the same two fish batches and estimated the differential DNA methylation status between the two fish groups. The methylation regulation on gene differential expression was evaluated at last. In the fifth and last "Conclusion" chapter briefly summarizes the overarching findings of this analysis and discusses the way forward.

1.4 Model fish of the study
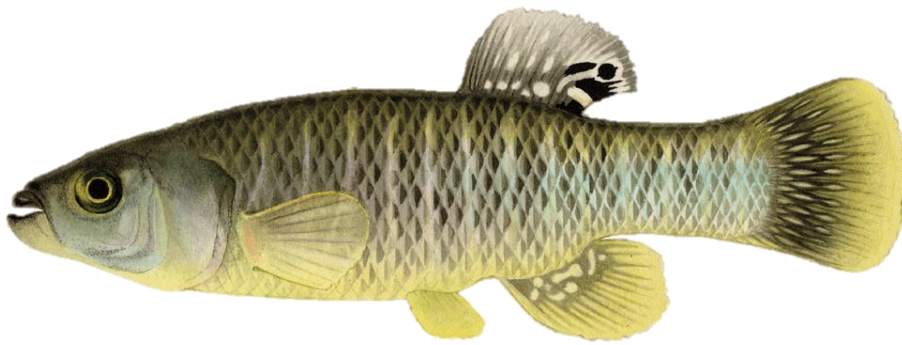
Fig. 1.1. Atlantic killifish (*Fundulus heteroclitus*)
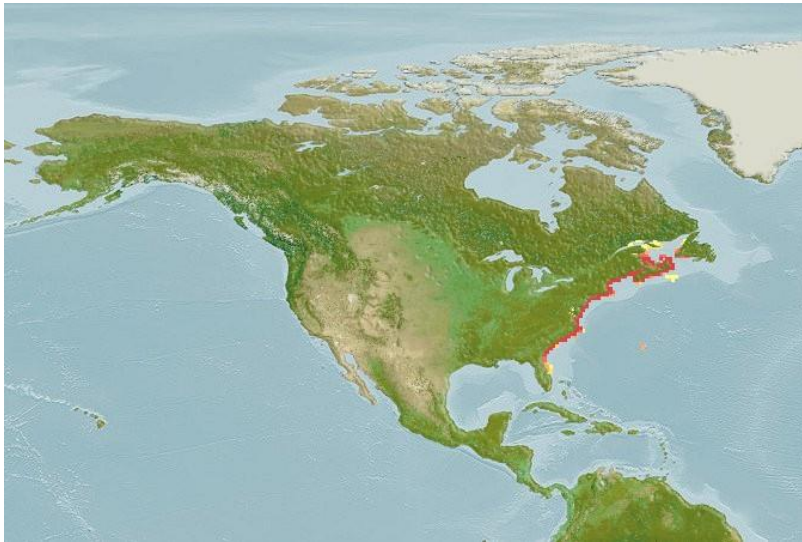
Fig. 1.2. Atlantic killifish native habitat range (Source- AquaMaps)

In this study, we used Atlantic killifish (referred to as common killifish, mummichog) (*Fundulus heteroclitus*) (Fig. 1.1) as the model marine ectotherm. Atlantic killifish is an intertidal minnow that inhabits a wide range of temperature range along the Atlantic coast of North America all the way up to Canada (New Brunswick) from northern Florida (Fangue et al., 2006; Dhillon and Schulte, 2011) (Fig. 1.2). Two Atlantic killifish subspecies have been identified, the northern subspecies, *Fundulus heteroclitus heteroclitus* inhabits mainly across a mean temperature range

between -1.4 to 21°C from New Brunswick (Canda) to Hudson Bay (USA), and *Fundulus heteroclitus macrolepidotus* inhabits the southern territorial range from Hudson Bay to Florida coast where the average ocean temperature ranged between 7-31°C (Fangue et al., 2006). In addition to the wide thermal tolerance, killifish are resistant to a range of abiotic environmental factors like salinity, oxygen concentration, and pH that fluctuate widely over the daily to seasonal cycles (reviewed in Di Giulio and Clark, 2015). Given the differential thermal regimes these subspecies experience, the acclimation to thermal fluctuations became more subspecies-specific, where the northern subpopulation exhibits higher acclimation capacity compared to its southern counterpart when exposed to acute or chronic temperature fluctuations (Healy and Schultz, 2012). Also, given its greater environmental tolerance, this species has been considered a crucial vertebrate model organism to examine the ectotherm's physiological mechanisms underlying wide environmental tolerance (reviewed in Di Giulio and Clark, 2015) and to predict marine ectotherms' potential physiological shifts with ocean warming.

Besides the wide environmental tolerance, Atlantic killifish is considered a marine vertebrate with rapid adaptability to extreme organic pollution in several highly industrialized estuarine environments (reviewed in Di Giulio and Clark,2015; Reid et al., 2016; Whitehead et al., 2016; 2017). Several Atlantic killifish populations have shown rapid and repeated adaptations to highly toxic environmental pollutants (like polycyclic aromatic hydrocarbons; PAHs, halogenated aromatic hydrocarbons like polychlorinated biphenyls; PCBs, 2,3,7,8-tetrachlorodibenzo-p-dioxin; TCDD-'dioxin', and other dioxin-like compounds; DLC) superfund sites along the coast of Northeast United States (Bello et al., 2001; Nacci, 2010). To this end, the pollution-resistant Atlantic killifish populations have been tested in a range of physiological and ecotoxicological experiments to uncover mechanisms underlying organic pollution adaptations

and the genomic basis of the evolutionary responses (Whitehead et al., 2017). For the third and fourth chapters of this study, we compared molecular level data from an Atlantic killifish population inhabiting the Elizabeth River (ER) system in southeastern Virginia that evolved resistance to extreme PAHs pollution that exceeded the recommended levels by several hundred times (sediment PAH concentration is 2200mg/kg dry weight; Greaves,1990; Alden,1995). The ER system has a history of military and industrial development with excess wood treatment processes applying PAHs-rich oil-derived Creosote (Di Giulio and Clark, 2015) until the recent past. PAHs are one of the most widespread groups of natural/anthropogenic persistent organic pollutants (POP) with pyrogenic (originated by fuel combustion), biogenic (organism-based), and petrogenic origins and comprised higher risks of human mutagens, carcinogens, transplacental carcinogens and tumor formation in liver, lung, and lymphatic tissues (see Bozinovic et al., 2021). Given these considerations, Atlantic killifish inhabiting the ER system has been tested in a range of experiments targeting to uncover the impact of PAHs on vertebrate evolutionary adaptations, and molecular, physiological, and behavioral modifications (Aluru et al., 2015; Jayasundara et al., 2017; Nacci et al., 2010; Reid et al., 2016; Whitehead et al., 2017).

CHAPTER 2. PREDICTING HABITAT RANGE SHIFTS OF A MARINE ECTOTHERM WITH OCEAN WARMING- PHYSIOLOGY INTEGRATED BIOCLIMATE MODELLING

2.1 Abstract

Species distribution models (SDMs) are computational programs applied to predict organisms'
current and future habitat range shifts with changing environmental conditions. SDMs employ
the contemporary environmental niche of a given species to the projected climatic conditions to
predict the future distribution, assuming the contemporary and projected niches are necessarily
unique. A range of SDM techniques has been employed to predict species distribution with
climate change, from bioclimate envelope to machine learning methods, but the predicted future
distributions with ocean warming primarily display a poleward migration for most marine
ectotherms. A key limitation of current species distribution models (SDM) is that they do not
account for population-specific heterogeneity in physiological responses to temperature change
resulting from local adaptations and acclimatization, underlies these unique habitat predictions.
To address this gap, in this study, we developed a novel SDM, the Physiology Integrated
BioClimate Model (PIBCM) that combines habitat-specific metabolic thermal physiological
tolerance of a marine ectothermic species into a bioclimate envelope model. We used the
contemporary coastal sea surface temperature (SST) data to determine the thermal niche of the
concurrent geographical distribution and integrated it with fine-resolution future SST data,
downscaled from three IPCC climate models to estimate habitat distribution of a highly
eurythermal intertidal minnow, the Atlantic killifish (*Fundulus heteroclitus*) for the 2050s and
2080s, a best-case scenario for coastal vertebrates. In this *in silico* study, we demonstrate that the

killifish northern boundary can potentially be shifted southwards, and distinct habitat fragmentation can occur in the southern sub-population (due to the migration of adjacent fish populations to the nearest metabolically optimal thermal habitat) in the future, under predicted warming scenarios. When compared to current SDMs (e.g., AquaMaps), our results emphasize the need for thermal physiology integrated range shift models and indicate that habitat fragmentation for coastal fishes may reshape nursery habitats for many commercially and ecologically important species.

2.2 Introduction

Rising global ocean temperature is a global pressing concern that is potentially shifting habitat distribution patterns for aquatic ectotherms profoundly. This has generated significant interest in predictive species distribution models (SDMs) (Grieve et al., 2017; Reygondeau & Beaugrand, 2011) to infer habitat suitability and range by combining species' spatial extent with environmental data (Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000). A range of different SDM techniques has been tested in the current ecological and conservation context, and Climate envelope modeling (Kaschner et al., 2006) is one standard SDM technique that uses a species' present geographical distribution to determine its environmental niche to predict the re-distribution patterns by directly mapping the niche to the future climatic conditions (Pearson & Dawson, 2003).

Two key limitations exist in correlative species distribution models when predicting habitat shifts driven by warming. First, most models consider the minimum and maximum habitat temperatures of the entire population range as the upper and lower thermal limits of a given population, without accounting for potential local adaptation of sub-populations (Buckley, 2008; Kolbe et al., 2010). Secondly, existing models do not consider species' capacity to modify their

thermal physiological properties, i.e. acclimation or acclimatization, to optimize fitness in their local environment (Kuo & Sanford, 2009; Somero et al., 2017). Given these limitations, the predicted habitat ranges for marine ectotherms uniquely demonstrate a mere poleward range expansion, which is considered oversimplified and overestimated. Recently developed hybrid correlative models have addressed some of these limitations (Buckley et al., 2011; Gamliel et al., 2020; Kearney et al., 2010) by integrating thermally dependent phenotypic parameters (e.g., development threshold for larvae (Buckley et al., 2011), growth rates, and clearance rate (Gamliel et al., 2020) into model predictions, but failures to incorporate within-species physiological variation, like differences in local thermal variation and organismal physiological acclimation capacity, at a high spatial resolution when predicting distribution patterns are still existing.

Here, we posit that using fundamental cellular physiological processes (e.g., energy metabolic rates) can yield more accurate predictions of a population's response to changing thermal habitats than whole-organism parameters like growth rate. Aerobic scope—the absolute difference between the maximum metabolic rate (MMR) and routine metabolic rate (RMR)—is a highly thermally-dependent parameter, that governs a species' reproductive success, growth, survival, and biogeography (Brett, 1971; Donelson et al., 2012; Pörtner & Farrell, 2008; Pörtner & Knust, 2007; Somero et al., 2017). This notion is captured in the oxygen and capacity-limited thermal tolerance (OCLTT) framework (Pörtner, 2001; Pörtner, 2010; Pörtner & Farrell, 2008; Pörtner & Knust, 2007), which indicates that declining aerobic performance measured in terms of aerobic scope (AS) results in a thermal limitation of ectotherms at both ends of the thermal envelope. Despite criticism of the overall framework (Clark et al., 2013), the highest level of aerobic scope is considered to be tightly linked to optimum performance (Farrell, 2013). While

19

studies have integrated these concepts to predict the future metabolic changes in terrestrial and aquatic ectotherms with global warming (Deutsch et al., 2015; Dillon et al., 2010), attempts to integrate metabolic concepts in predicting species range shifts are lacking. Therefore, in the current study, we propose a novel framework to integrate the thermal plasticity of aerobic metabolism into a bioclimate envelope model and predict range shifts of marine ectotherms at a high spatial resolution.

Here, we propose a novel framework to integrate the thermal plasticity of aerobic metabolism into a bioclimate envelope model and predict range shifts of marine ectotherms at a high spatial resolution. The central premise of our model is that a marine ectotherm maintains its routine metabolic rates within a certain habitat temperature range (at the climatological averaged minimum and maximum habitat temperature of the given location, i.e., T1 and T2 in Fig.2.1) in its current thermal environment, where it can sustain the highest possible aerobic scope at that location. In other words, we consider routine rate range as a proxy for aerobic scope, which is considered a key physiological determinant of organismal fitness. Changes in the current thermal environment (e.g., due to ocean warming) could reduce the optimum aerobic scope (explained in the OCLTT theory), leading to a habitat range shift unless the organism acclimates to the rising habitat temperatures. In this study, we hypothesized that the routine metabolic rates range (the range between routine metabolic rates calculated at the minimum and maximum habitat temperatures; defined as Metabolic Rate Range -MRR, Fig.2.1) is a representative parameter of the optimum aerobic scope of the considered marine ectotherm in the given location. We determined MRR for a given ectotherm in a given location by calculating the absolute difference between routine metabolic rates at the climatological maximum and minimum habitat temperatures in that location over the last 37 years (Fig. 2.1). For this, we use the Metabolic

20

Theory of Ecology equation (MTE) (Gillooly et al.2001, Brown et al.2004), which explains the thermal dependence on aerobic metabolism accounting for body size and environmental temperature (equation 1).

$$B = b_o M^{3/4} e^{\frac{-E}{KT}} \quad \text{equation (1)}$$

Here, B is resting/routine metabolic rate, b_o is an empirically derived taxon-specific normalization constant ($8.617 \times 10\text{-}5$ eV/K), M is body mass, E is the averaged enzyme activation energy in eV, K is Boltzmann constant, and T is the habitat temperature in Kelvin. Importantly, the averaged enzyme activation energy (E) is the thermally dependent physiological variable that can be modified through acclimation, developmental, transgenerational plasticity, and/or local adaptation. Although it was initially assumed a universal temperature dependence for E, indicating that it is constant within a taxonomic group (e.g., 0.43 for fish) (Gillooly et al.2001), ectotherms may modulate their E value (Clark,2004, Clark and Fraser,2004). Accordingly, here we posit that under thermal stress ectotherms may modify E to maintain a constant MRR in the current habitat.
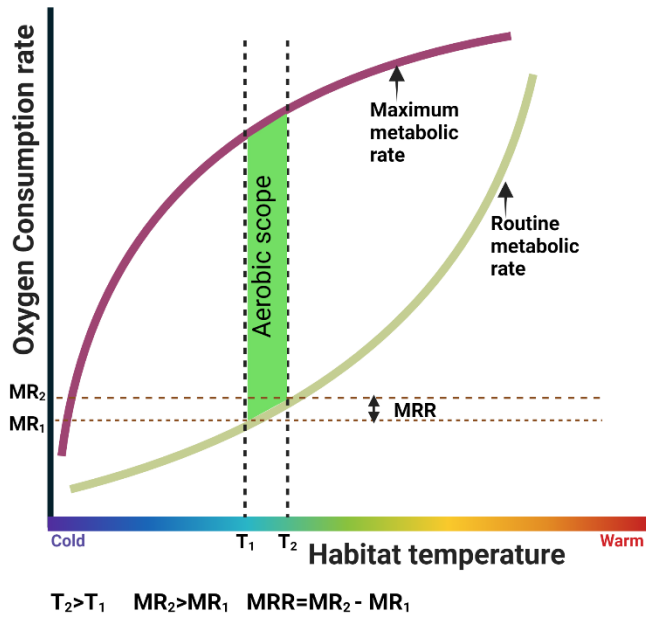
Fig. 2.1. The physiological theory underpinning physiology integrated bioclimate envelope model. Conceptual depiction of metabolic rate range (MRR) used to integrate Atlantic killifish thermal tolerance into the physiology integrated bio-climate model. Theory predicts that the routine metabolic rate (blue line) of an ectothermic teleost exponentially increases with the increasing temperature while the maximum metabolic rate (brown line) plateaus with increasing temperature and that the ideal thermal window (T1 –T2) is where aerobic scope (difference between the maximum and routine metabolic rates) is maximized. We postulate that to maintain maximum aerobic scope following an increase in temperature, a fish from a given habitat will modify their thermal physiological properties to stay within a certain routine metabolic rate range (MRR - the difference in routine metabolic rates between T1 and T2) (Created with BioRender.com)

To develop our physiology-integrated bio-climate model (PIBCM), we first tested the species specificity of the E value across a range of fish species. We show that this value is highly species-dependent and sensitive to the population's local thermal envelope. We then simulated

metabolic rate sensitivity to E in the MTE equation to demonstrate that metabolic rates and the

metabolic rate ranges shift with E value changes. Then we quantified the habitat range shift

between 2040-2069 and 2070-2099 for a highly eurythermal intertidal minnow, the Atlantic

killifish (*Fundulus heteroclitus)*. We calculated the current MRR using climatological monthly

minimum and maximum sea surface temperature (SST) data over the last 37 years for killifish in

their current local habitat at a resolution of 0.5o. We parameterized the model for the two

subspecies of Atlantic killifish that live along the North American East Coast (Dhillon and

Shulte, 2011; Fangu et al., 2006) (Fig. 2.2). The southern population is more heat tolerant (7-

31℃ range) than the northern population (-1.4 to 21℃ range) (Fangu et al., 2006), and the daily

and seasonal temperature fluctuations vary widely between populations (Fangu et al., 2006,

Fangu et al., 2009). Thus, we calculated E values for a range of acclimation temperatures for

northern and southern killifish based on data from an extensive killifish acclimation experiment

(Healy and Schulte, 2012). To calculate MRR values associated with future projects, we

substituted contemporary SSTs with predicted SSTs based on a novel coastal SST dataset we

developed for the 2050s and 2100s. We implemented a downscaling approach built on averaging

three coupled model intercomparison project phase 5 (CMIP5) model outputs (Representative

Concentration Pathways (RCP)2.6,4.5,8.5; Supplementary Table S1). Subsequently, the model

allowed a fish population in a given habitat (0.5o grid cells) to remain in the same grid if the fish

can modulate the E value at the predicted future ocean temperature. If not, the model will find

the nearest grid location with the same or lesser MRR as the best habitat for that given fish

population to relocate with the ocean warming.
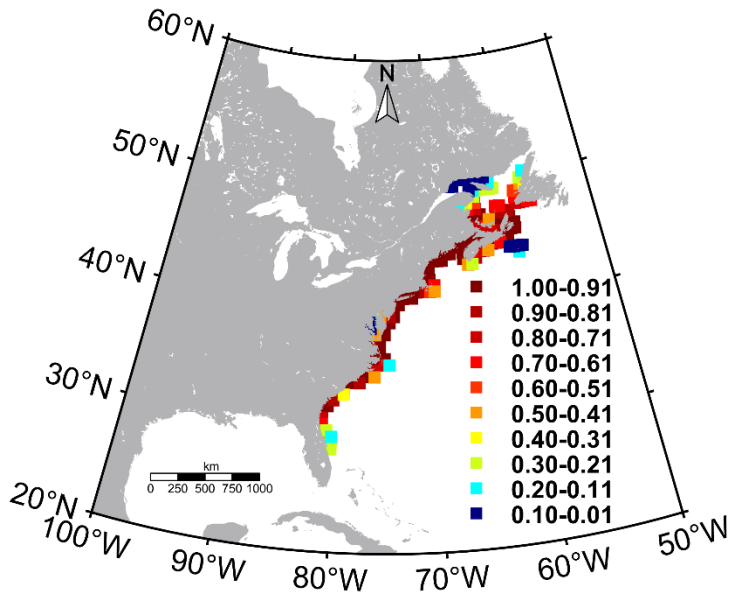
Fig. 2.2. Atlantic killifish probability habitat distribution within the native range
(Source: AquaMaps).

To our knowledge, this is the first study to integrate thermal physiological acclimation into
SDMs, and our use of an eurythermal species like the killifish could represent a best-case
scenario for other coastal and estuarine fishes in response to future warming.

2.3 Materials and Methods

2.3.1 Species-specific E value calculation

Species-specific E values were calculated using collected fish routine or standard metabolic rate
data published between 2000-2018. Data available in the University of Maine Aquatic Science
and Fisheries Abstract and Web of Science databases were used in this calculation. Metabolic
rate data of fish treated in three or more acute/acclimatory temperature treatments (Lear et al.,
2019) were only considered in this study. Metabolic rate data collected as a function of
temperature, CO_2, or Salinity were considered in the study after removing metabolic rate data

collected at projected scenarios. Respective authors were contacted personally to obtain

metabolic rate data unless data weren't available online. Species-specific E values were

calculated using the Arrhenius-Boltzmann (Clarke & Johnston, 1999) (Equation 2) relationship

considering its statistical thermodynamic robustness for scaling metabolic rate variability with

temperature (Clarke & Johnston, 1999).

$$R_b = A_o e^{-E/KT} \text{ (Equation 2)}$$

R_b is the RMR/SMR, Ao is a constant, E is the averaged activation energy (eV), K is the

Boltzmann constant, and T is the temperature (Kelvin). The relationship between ln (R_b) and

1/T was estimated using the least square regression method which preserves the statistical

thermodynamic robustness (Clarke & Johnston, 1999) and the slope is equal to -E/K.

2.3.2 Sensitivity analysis

A sensitivity analysis was conducted to test the effect of temperature, body mass, and E value on

the metabolic rate and metabolic rate range (MRR). The metabolic theory of ecology equation

was used to calculate hypothetical metabolic rates for mass values ranging between 0.001-10kg

with 100g intervals, E values from 0.01-1 with 0.1 intervals, and temperature values ranging

between 0-40°C with 5°C intervals. 14.47 was used as the metabolic scaling component (bo¬)

(Gillooly et al., 2001). Metabolic rates were recalculated for the same mass and E values, but a

range of temperature values between 0-40°C with 2°C intervals to estimate the effect of

temperature on MRR. MRR was calculated as the difference between metabolic rates at

consecutive temperatures (e.g. 0-2°C, 2-4°C) at a given E and mass.

2.3.3 Atlantic killifish habitat distribution

Atlantic Killifish probability distribution data were downloaded from AquaMaps (www.aquamaps.org). AquaMaps generates Atlantic Killifish native range probability distribution (0-1) in 0.5º×0.5º grid resolution (Fig. 2.2).

2.3.4 Contemporary coastal Sea Surface Temperature (SST) data

High-resolution (0.05º×0.05º×daily) reprocessed global SST data (Good et al., 2020) between 1982-2018 were downloaded from Marine Copernicus (https://marine.copernicus.eu) server for the model domain of 20º-60ºN and 100º-54ºW. Atlantic Killifish mostly inhabits estuaries and salt marshes (Fangue et al., 2006). Therefore, we assume that Atlantic killifish inhabits within the 0-10m depth range and extracted SST data between 0-10m depth contours using very high-resolution ETOPO2v2 bathymetry data (National Geophysical Data Center) (https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO2/ETOPO2v2-2006/ETOPO2v2/netCDF/). Extracted SST data were upscaled to 0.5º to match the model resolution (0.5º). Climatological monthly maximum and minimum SST per grid point were calculated by averaging 37 years of SST data. Grid-specific SST range was calculated as the difference between the maximum and minimum climatological monthly mean SSTs.

2.3.5 Downscaling the climate model projected SST data

Historical (1961-1990) and the future (the 2050s (2040-2069), and the 2080s (2070-2099)) SST data were downloaded from three climate models of the Climate Model Intercomparison Project Phase 5 (CMIP5) and the Intergovernmental Panel for Climate Change (IPCC) from ESGF data distribution center at (https://esgf-node.llnl.gov/projects/cmip5/) representing three Representative Concentration Pathways (RCP 2.6,4.5, and 8.5). RCP considers three different

emission scenarios, strong greenhouse gas mitigation (~2.6Wm-2) by 2100 (RCP2.6), intermediate mitigation (~4.5 Wm-2, RCP4.5), and no mitigation (~8.5Wm-2, RCP8.5) (Stocker et al.2013). SST is expected to rise respectively in each RCP scenario which is ~1°C relative to 2006 in RCP2.6 and 2.4°C compared to 2006 in RCP8.5. Most of the CMIP5 projects are based on ensemble calculations that have different initialization methods, initial states, and physical details (Stocker et al.2013). In the CMIP5, ensemble nomenclature is based on the rip nomenclature (r – realization, i- initialization, and p-physics). Among the available ensembles, we always used the first model ensemble (r1i1p1) data in this study to reduce the complexity of the computational process.

The coarser resolution of the downloaded SST predictions hindered comprehensive estimation of the future SST variability along the North American coastline. Therefore, we downscaled the individual climate model data using the Delta method (Navarro-Racines et al., 2020) to a 0.5°×0.5° grid interval. This method takes the difference (Delta change) between the historical and future climatological SST of a given climate model, re-grid the difference to the same resolution of contemporary climatology, and adds it up to contemporary climatology. Here we developed contemporary climatology by calculating climatological monthly means of 0.5°×0.5° gridded Copernicus SST data. Thin Plate Spline Interpolation (Frank, 1982) was applied to re-grid the delta change to the contemporary climatology resolution. Delta change was calculated as the difference between the climatological monthly means of 2040-2069 (2070-2099) and the climatological monthly means of the historical (1961-1990) period of the same climate model. At last, downscaled individual model SST products were averaged to get the model mean SST prediction. The final climatological monthly mean SST product was used to calculate the projected mean annual, monthly minimum, and maximum SST per grid cell.

2.3.6 Metabolic rate range (MRR) calculation

We assumed that Atlantic Killifish in each 0.5º grid is a unique population that is locally adapted to maintain a constant MRR depending on the grid thermal envelope and the thermal envelope-specific E value. We used Atlantic Killifish routine metabolic rate (RMR) data published in Healy and Schulte (Healy & Schulte, 2012) to calculate the thermal envelope-specific E values. Healy and Schulte (Healy & Schulte, 2012), recorded RMR data of seven groups of Atlantic Killifish from both north and south subpopulations, acclimated to a temperature range of 5º,10º,15º,20º,25º,30º, and 33℃. For this, the maximum and minimum SST (current and future projected) values in each grid were rounded up to the nearest 5℃. The experimental RMR values within the rounded temperature range were extracted from Healy and Schulte (Healy & Schulte, 2012) and plotted as a function of assay temperature to determine the slope of the regression to calculate grid-specific E value as per the Arrhenius-Boltzmann equation. The maximum and minimum metabolic rates in each grid were determined (Fig. 2.3) using the MTE equation and the grid-specific E value to calculate the MRR using equation 3. All calculations were conducted assuming the fish body weight as 10g.

MRR= maximum metabolic rate-minimum metabolic rate (Equation 3)

We set the 40ºN as the geographical cline between Northern and Southern Atlantic killifish populations. Therefore, experimental RMR data collected using Northern Atlantic killifish subspecies were used to calculate E values in grids at or above 40ºN and vice-versa. Healy and Schulte (Healy & Schulte, 2012) explained that both fish subpopulations reached their peak metabolic performance at 30℃ and dropped at 33℃. Accordingly, we selected 32℃ as the physiologically optimum break temperature.
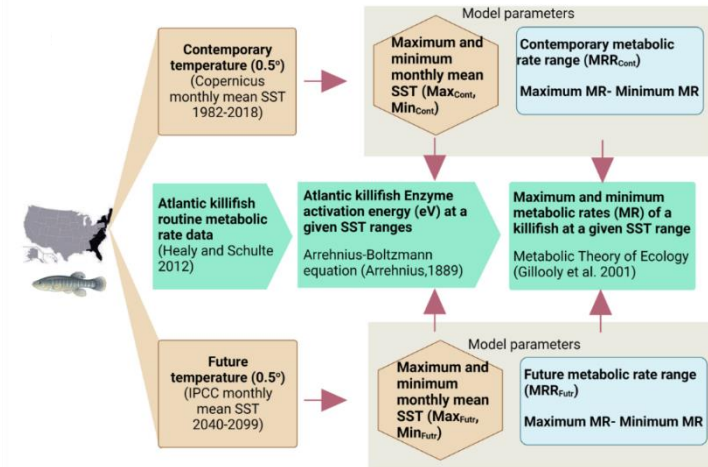
Fig. 2.3. Flow chart of the PIBCM data processing. Pale blue, green, and orange boxes respectively represent the physiological data source, physiological theories used in the study, and the data that feed into the PIBCM algorithm (Created with BioRender.com)

2.3.7 Predicting Fundulus heteroclitus future distribution using AquaMaps

AquaMaps model is regularly used as a reliable bio-climate envelope model to predict around 33000 species' current and future habitat distributions, and its operational process is more efficient than the other statistical or machine learning species distribution models. Therefore, we chose AquaMaps as a baseline climate envelope model result to compare with our PIBCM model results. SST climate envelopes used in AquaMaps (Kaschner et al., 2019) were used to predict the Atlantic Killifish distribution in the 2050s and 2080s. AquaMaps uses a Relative Environmental Suitability (RES)(Kaschner et al., 2006) method to predict species distribution considering several bio-climate indices with 5.63º, 7.74º, 21.97º, and 27.05°C as the respective minimum, preferred minimum, preferred maximum, and maximum SST envelopes (Kaschner et al., 2006; Kaschner et al., 2019) to predict the distribution. Climatological mean SSTs from the three downscaled RCP scenarios were used to generate habitat predictions.

2.3.8 Physiology integrated bioclimate model (PIBCM) development

We assume Atlantic Killifish habitat probabilities determined (predicted probabilities by AquaMaps) in a single grid cell have been locally adapted to maintain constant MRR during the past 37 years. The fish population remains in the same location if the grid MRR in the 2050s or the 2080s is lower or equal to the historical MRR even though the grid-specific future SST exceeds its respective contemporary temperature. So, we implemented three basic model conditions to simulate Atlantic killifish future habitat ranges (Fig. 2.4).

Model condition 1: if maximum SST (Tmax) ≤ 32°C and Tmax (future) ≥ Tmax (contemporary) and MRR (future) ≤ MRR (contemporary) in a particular grid, the inhabited fish population stays in the same grid (no habitat range shift).

Model condition 2: if Tmax ≤ 32°C and Tmax (future) ≤ Tmax (contemporary) or MRR (future) ≤ MRR (contemporary), the inhabited fish population remains in the same grid.

Model condition 3: If Tmax >32°C or Tmax (future) ≥ Tmax (contemporary) and MRR (future) ≥ MRR (contemporary), the inhabited fish population moves to the nearest grid location where MRR (contemporary) ≤ MRR (future).
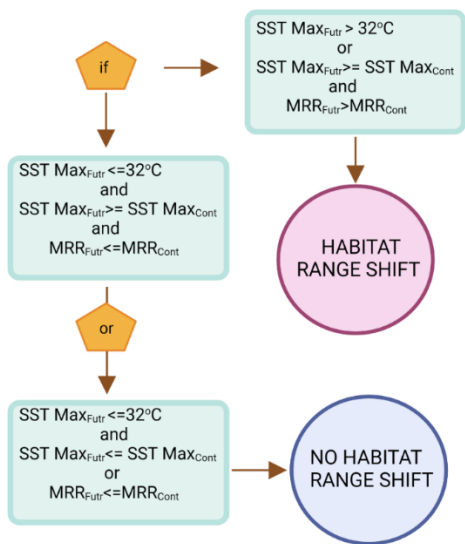
Fig. 2.4. The PIBCM model algorithm (Created with BioRender.com)

Model building, running, and mapping (m_map toolbox v1.4 (Pawlowicz, 2020)) were completed using the MATLAB R2020a version (University of Maine academic license 358411). Cartesian climate data grids were re-gridded into regular latitude and longitude grids using area-weighted bilinear interpolation in Climate Data Operators (CDO v.1.9.9 (Schulzweida et al., 2006)). Fish E value calculation, statistical analysis, and thin-plate spline interpolation (fields library v.11.6 (Nychka et al., 2017)) were conducted in R v 4.0.3 (www.r-project.org)

2.3.9 Statistical analysis

The Arrhenius-Boltzmann relationship (Clarke & Johnston, 1999; Equation 2) was used to model the relationship between fish routine metabolic rates with the absolute temperature. The plot of natural logarithm transformed routine metabolic rates (ln (metabolic rate)) with the inverse temperature (T-1) yields a slope that is equal to -E/K where E is the enzyme activation energy, and K is the Boltzmann constant. To estimate this relationship, we used the least-square regression (model I) in this study assuming its statistical robustness of scaling metabolic rates

with independent variables such as body mass and temperature (Clarke & Johnston, 1999). E values with probability (P) <0.05 were considered significant.

2.4 Results

2.4.1 Species-specific E value

Species-specific E values for 19 teleost fishes (five freshwater, seven marine, four brackish, and three euryhaline fish) were calculated using published standard or routine metabolic rates (SMR/RMR) (Table 1 and Fig.3). E values were highly species-specific and ranged from 0.23 (mosquitofish, Gambusia affinis) to 0.96 eV (Nile tilapia, Oreochromis niloticus), and depends on acclimation temperature (Table 1 and Fig.2.5). The mean E value calculated in our study (0.55eV ±0.17) was lower than previous estimations for all taxa (0.62eV) and higher than fish-specific E value (0.433eV) (Gillooly et al., 2001). Within-species differences in E value were clearly reflected in comparisons between northern and southern killifish subpopulations, confirming that acclimation and local adaptions significantly alter these values (Healy and Shult, 2002), (Table 1). Acclimated and acutely exposed northern Atlantic killifish E values were lower than that of the Southern populations. Overall, these data demonstrate the species specificity of the E value and its dependence on the population's thermal history.

Table 2.1. Summary of the species-specific E values calculated for 19 fish species. Fish from the same species may have different E values depending on the experimental conditions used to estimate the fish's routine metabolic rates. Least square regression was applied to regress the natural log-transformed fish routine or standard metabolic rates with the inverse temperature.

	E value (eV)	R²	P	Experiment temperature (°C)	Mean Weight (g)	Environment	Reference
Redband Trout	0.59	0.57	<0.001		1.14	F*	Chen et al. 2018
	0.58	0.53	<0.001	12,15,18,21,24	2.41		
	0.56	0.49	<0.001		3.4		
Goby	0.32	0.28	<0.001	15,22,28,33,36	1.15	M¶, BΓ	da Silva et al.2019
	0.61	0.3	<0.001		0.73		
Black sea bass	0.69	0.74	<0.001	24,27,30	265.6	M	Slesinger et al.2019
	0.56	0.87	<0.001	12,17,22,27,30	375		
Mosquito fish	0.31	0.04	0.0052	24.8,29.6,34.9,35.5,37.1	50.57	F, B	Moffett et al.2018
	0.23	0.02	0.062	19.2,22.4,23.7,27.6,29.9	70.38		
Nile Tilapia	0.96	0.95	<0.001	19,22,25,28,31	50	F, B	Turker, 2011
	0.86		<0.001		200		
Atlantic Killifish	0.6	0.85	<0.001		4.58	M, B, F	Healy and Schult,2012
	0.73	0.81	<0.001	5,10,15,20,25,30,33	6.14		
	0.51	0.9	<0.001		6.26		
	0.6	0.91	<0.001		6.77		
Atlantic Salmon	0.55	0.9	<0.001	3,8,13,18,23	443.77	M, B, F	Hvas et al. 2017
Lump fish	0.37	0.54	<0.001	3,9,15	310.21	M	

Table 2.1 continued

Polar cod	0.33	0.42	<0.001	0,3,6,8	16	M	
Atlantic cod	0.4	0.68	<0.001	3,8,12,16	40.76	M	**Kunz et al. 2016**
Atlantic cod	0.46	0.63	<0.001	8 20	73	M	Norin et al. 2019
	0.37	0.51	<0.001	12 23	64.6		
Bone fish	0.37	0.1	0.013	22 35		M	Nowel et al. 2015
Quingbo	0.4	0.71	<0.001	10,15,20,25,30	2.87	F	Pang et al. 2015
Thorny Skate	0.7	0.33	0.0018	5,9,13	1.56	M	
Clearnose Skate	0.44	0.26	0.01	20,24,28	1.22	M	Schweiterman et al. 2019
Snakehead	0.81	0.95	<0.001	15,20,25,30,35	4.4	F	Xie et al. 2017
	0.72	0.91	<0.001		4.42		
Ballan Wrass	**0.62**	**0.14**	**<0.001**	**5,10,15,20,23**	**162.59**	**F**	
Roach	0.6	0.24	<0.001	5,10,15,20,23	63.76	B, F	
Vendace	0.64	0.4	<0.001	4,8,15	30.78	M, B, F	**Ohlberger et al. 2012**
Stechlin cisco	0.54	0.44	<0.001	4,8,15	16.15	F	

*F- Fresh water, Γ B-Brackish water, ¶ M-Marine
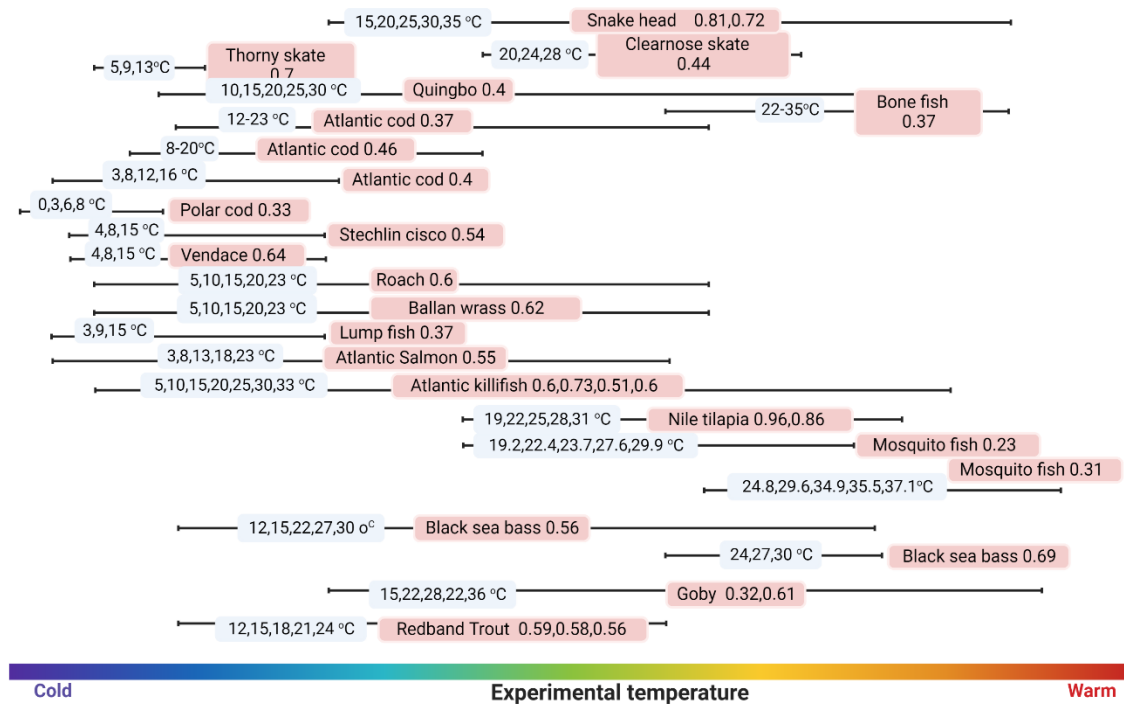
Fig. 2.5. Summary of the species-specific E values calculated for 19 fish species. Same fish species may have different E values depending on the experimental conditions and temperature ranges. The solid lines depict the experimental temperature range, and the blue legends show the temperature values. Fish species name and the estimated E values (in eV) are displayed in the pale red boxes (Created with BioRender.com).

2.4.2 Sensitivity analysis

To determine the effects of different E values on the metabolic rate and metabolic rate range of a given organism, we examined the relationship between the E value and metabolic rate based on a sensitivity analysis of the MTE equation. Hypothetical metabolic rates were calculated as a function of temperature (0-40°C) and body mass (1-1000 g) for E values between 0.01-1 eV. Metabolic rate can range from -10.95 to 9.6 (log10 mW). As expected from the MTE equation, a small variation (e.g., 0.1 eV) in the E value resulted in a 70-40-fold change in metabolic rate

across the temperature range used for metabolic rate calculations (0-40℃) (Fig. 2.6). Metabolic

rates generally increased with increasing temperature and body size as expected, but the changes

in the E value had the greatest impact on metabolic rates (Fig. 2.6). The effect of size (weight) on

metabolic rate was highest at very small size ranges (~1g -100g), and this effect decreased as the

animals got larger (Fig. 2.6). As such, the impact of a 0.1eV change in the E value on metabolic

rate is equivalent to a seventy-fold change in size for a given organism. For a 10 g fish, the effect

of temperature on metabolic rate was minimal at higher E values but increased with increasing E

values, i.e., the slope of the line between metabolic rate and body mass decreased with

decreasing E values (Fig. 2.6). Calculated metabolic rate ranges (MRR) between two consecutive

temperatures (e.g., 0-2℃, 2-4℃, 4-6℃) increased linearly (Fig. 2.6), except for the MRR

calculated at the lowest E value (0.01eV).
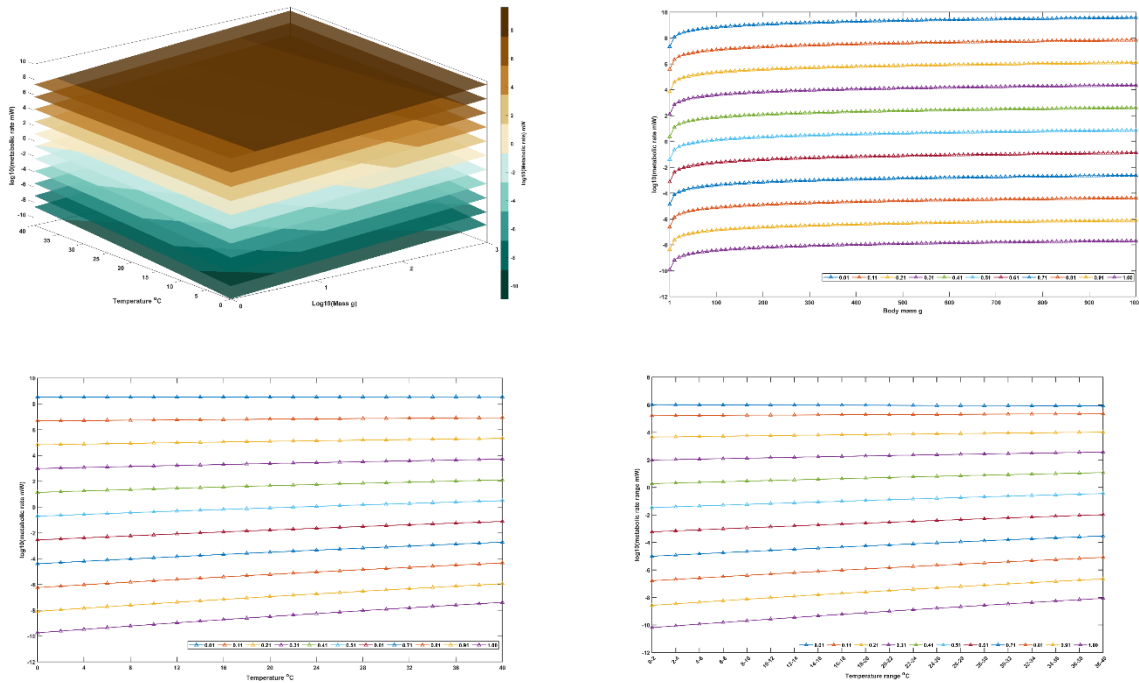
Fig. 2.6. Sensitivity analysis calculates hypothetical metabolic rates (mW) for a range of temperature (0-40 °C) (top left), body mass (1-1000g) (top right), and E values (0.01-1eV) (bottom left) and metabolic rate range (bottom right) using the Metabolic Theory of Ecology equation. Different color lines in figures B, C, and D indicate the respective E value (displayed in the legend box) used for the calculation.

2.4.3 Historical and future coastal SST distribution

As expected, climatological mean (1982-2018) SST along the North American coastline showed a distinct latitudinal gradient (Fig. 2.7). The highest and lowest mean SSTs (28.19 and -1°C) were recorded for 20.25N (Florida Keys) and 59.75N (Newfoundland and Labrador coast) respectively. Within the native Atlantic killifish habitat distribution (28-52N), climatological mean SST extended from 1.54 to 24.29°C. Location-specific climatological mean SST range (the difference between the maximum and minimum mean SST for a given location) spanned between 8-24.3°C in the current killifish habitats (between 28N and 52N), indicating a potential

preference for thermally variable environments (Fig. 2.7). The highest SST range (24.29°C) was recorded in the Chesapeake Bay, while the SST range experienced by killifish was generally higher for the populations in the Delaware Bay region, and the southern Gulf of St. Lawrence coastal region.

Our novel downscale approach to obtaining predicted SST showed a distinct latitudinal thermal gradient similar to the contemporary pattern and predicted a positive mean SST anomaly between contemporary and future periods (the 2050s and 2080s) under all RCP scenarios (Fig. 2.7, C and D). The location-specific SST range (the difference between the maximum and minimum climatological mean SST for a given location) in the 2050s and 2080s also showed a similar pattern to its contemporary distribution (Fig. 2.7). For the 2050s (RCP 2.6) and 2080s (RCP 8.5), the predicted highest SST range was 25.52°C and 27.34°C respectively, and was recorded in the Chesapeake Bay region (38.75N,76.25W) (Fig 2.7). Overall, our downscaled model output predicted an increasing SST range.
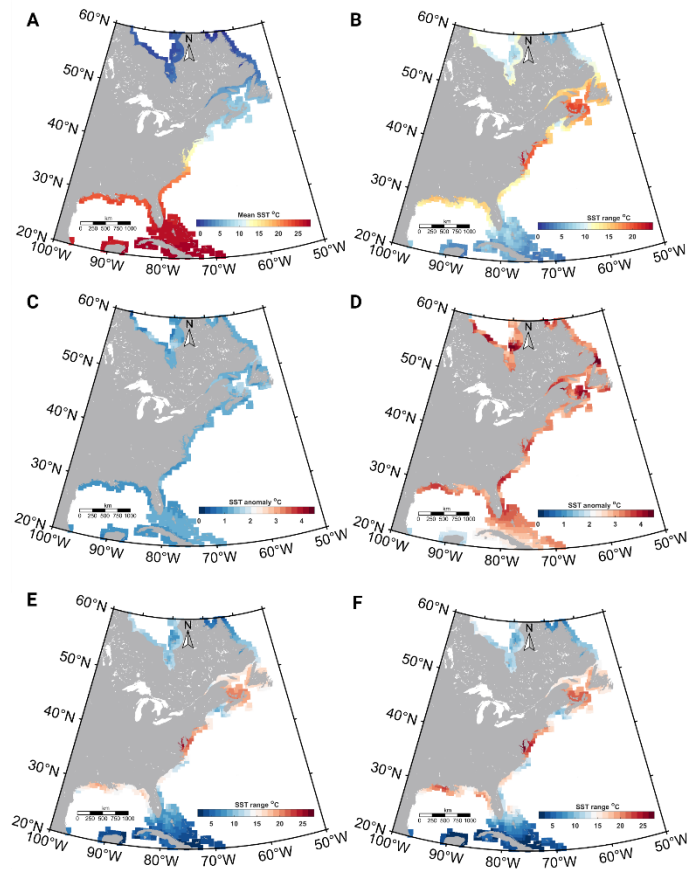
Fig.2.7. The contemporary and the climate model projected future sea surface temperature (SST) variability along the East Coast of North America (A) Climatological mean SST distribution along the North American East coastline during the contemporary (1982-2018) period (Copernicus data). (B) The SST range (Difference between maximum and minimum SST) along the same region and period. (C) SST anomaly (difference between the future and contemporary SST distribution) in the 2050s (RCP 2.6) (D) SST anomaly in the 2080s (RCP 8.5) (E) The SST range along the North American coastline during the 2050s (RCP 2.6) and (F) 2080s (RCP 8.5).

2.4.4 Atlantic killifish thermal envelope-specific E values

Based on location-specific SST ranges, we determined thermal habitat envelopes for killifish. The killifish metabolic rates we adopted for our study were estimated at a temperature array within 5°C intervals (5,10,15,20,25,30 and 33°C, (Healy and Schult,2012). Therefore, we rounded up the long-term-averaged minimum and maximum habitat temperatures to the nearest 5°C to define all the possible killifish thermal envelopes. Essentially a given thermal envelope reflects the maximum and minimum temperatures of a given location along the current killifish habitat. We estimated nine thermal envelopes for the Northern subpopulations' habitat range, and six thermal envelopes for the Southern subpopulation and found three thermal envelopes (5-25°C, 5-30°C, and 10-30°C) to be common for both populations. Six thermal envelopes (5-10°C, 5-15°C, 5-20°C, 10-15°C, 10-20°C, and 10-25°C) were unique to the Northern population range, while 15-30°C, 20-30°C, and 25-30°C were unique to the Southern population range. E values were estimated for each of the thermal envelopes, and we found E values for the common thermal envelopes were higher in the Southern subpopulation than in the Northern subpopulation, further confirming that Southern subspecies are more thermally sensitive than the Northern subspecies. We then used the same approach to calculate thermal envelopes and their respective E values based on the predicted future temperatures. Overall, the metabolic rates, and respective metabolic rate ranges in each grid varied as a function of the thermal envelope and the envelope-specific E values. The yearly E value variance across the Atlantic killifish native range during the contemporary period showed a higher variance at the northern and southern ends (Fig.2.8) and an evenly lower variance in the middle part. Notably, the southern habitat range showed the highest yearly E variance (Fig.2.8) and the habitat fragmentation during the 2080s as predicted in our model.
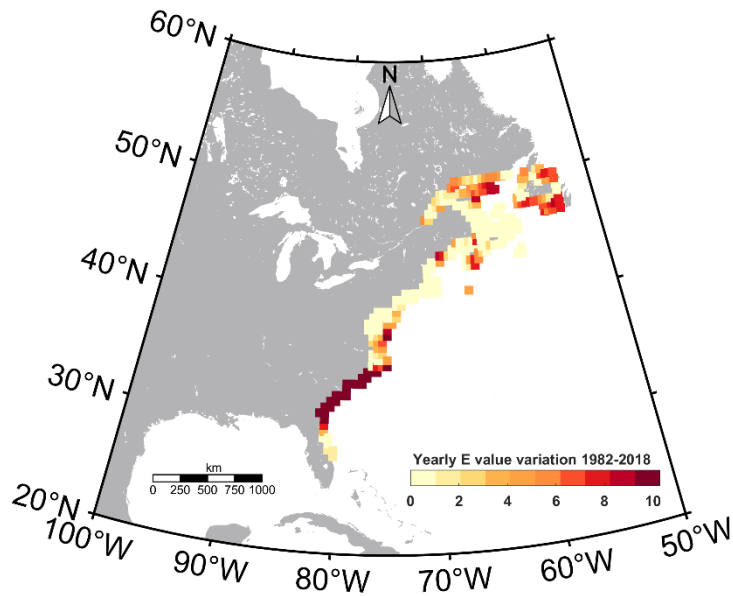
Fig. 2.8. Yearly Atlantic killifish E value variation (103) during the 1982-2018 period. Respective E values (eV) for each year's maximum and minimum SSTs along the Atlantic killifish native habitat range were calculated using Healy and Schulte,2012 data.

2.4.5 Predicted, historical, and future MRR distribution

MRR, which is determined as a function of the population-specific E-value and contemporary or predicted future temperature change, is the deterministic physiological parameter that was in the climate envelope model. MRR within the Atlantic Killifish's contemporary habitats ranged from -7 to -2 log10 mW. The highest killifish MRR was observed in the Northern limit of the contemporary habitat boundary around Nova Scotia and the surrounding coast (44 to 52N). In addition, patches of high MRR (~ -2 log10 mW) were also observed within the contemporary habitat range. Overall, our downscaled SST data predicts three distinct clusters of MRR that can be categorized into low (-8.5 to -7 log10 mW), moderate (-6 to -5 log10 mW) and high (-4 to -3 log10 mW) MRR regions. With increasing temperatures in the 2050s (RCP 2.6) scenario, we

observe a clear shift in these clusters, indicating an expansion of high MRR regions. This is particularly prominent in Nova Scotia, where the high MRR region expands its spatial extent. The coastal zone between 52-60N is predicted to become a region with minimum MRR (~ -10 to -7 log10 mW) under future RCP scenarios.

2.4.6 Atlantic killifish habitat range shifts predicted using PIBCM

Three model conditions were implemented to predict the Atlantic killifish future habitat ranges (see method section). The first criterion is that if the maximum habitat temperature is less than 32°C (physiological break temperature used for killifish in this model) and if the future maximum temperature is higher than the contemporary maximum temperature but future MRR is lower than current MRR, the fish population will remain in the same grid. Accordingly, PIBCM predicts that fish populations in several regions, including Newfoundland, Nova Scotia, New Brunswick, Cape Cod, Cape Hatteras, and the northern Florida coast will continue to stay in their current grids (Fig. 2.9). In the Nova Scotia region (just south of Newfoundland), the number of killifish populations that met this criterion increased with time and the severity of RCP scenarios (Fig. 2.9). This outcome is a result of the potential capacity of killifish to modify MRR by regulating their E value, even though the future maximum temperatures are higher than the contemporary maximum temperatures. The second criterion is that if the maximum habitat temperature is less than 32°C and the future maximum temperature or MRR is lower than their contemporary values, the fish population will remain in the same grid. However, no killifish grid location followed this condition. The third criterion was that if the future maximum temperature exceeds 32°C or if the future maximum temperature and future MRR exceed contemporary values, then the killifish population will move to the nearest grid locations with equal or lower MRR relative to their current habitat. ~78% of the grid cells representing killifish habitats

followed this model condition. Under RCP 8.5, 8% of habitat grid cells exceeded the break

temperature of 32˚C while 1% did so under RCP 4.5 in the 2080s.
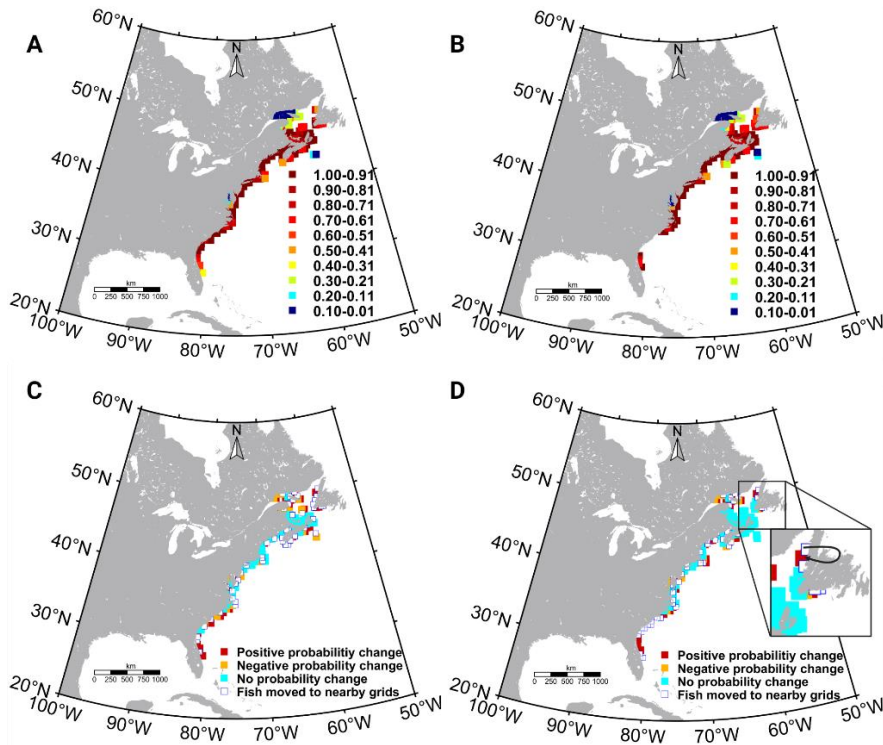
Fig. 2.9. Physiology integrated model predicted killifish future distributions for two climate

projections (representative concentration pathway (RCP)) and the comparison with the

contemporary distribution. (A) Atlantic killifish distribution (probability) in the 2050s (RCP2.6)

and (B) 2080s (RCP8.5). (C) The comparison between new model predicted Atlantic killifish

distribution for the 2050s (RCP2.6) and (D) the 2080s (RCP8.5) with the native habitat range.

The dark arrows in the figure D shows the southward migration of the killifish populations

inhabiting the northern most grid with ocean warming

We observed a combination of outcomes such as population shifts to their adjacent grid

locations, habitat range contractions, and habitat fragmentations (Fig. 2.9). We did not detect a

clear northward range shift, and in fact, the northernmost killifish populations may shift to

43

nearby southern grid locations (Fig. 2.9). Under the RCP 2.6 2050s scenario, small-scale habitat fragmentations were predicted along the coastline (mainly due to the exceeding future MRR than the contemporary values), with the most pronounced changes in the southern part of the Gulf of Maine and the Cape Cod coast. The size of fragmentations widened in the 2080s and under different RCP scenarios. In particular, profound habitat fragmentations were observed for the Southern killifish subpopulations (Fig. 2.9). This is a result of predicted habitat temperatures exceeding 32°C, where some Southernmost killifish populations may aggregate around ~28N-30N seeking thermal refugia.

To quantify the negative and positive predicted aggregations to a given grid location, we calculated the difference in probability of a given grid cell being occupied by killifish under future scenarios compared to the contemporary distribution (Fig. 2.10). Some grid cells that served as thermal refugia for nearby fish populations exceeded the cumulative probability when the contemporary grid probabilities shifted with the ocean warming and the model was set to adjust the exceeded probability value to 1. As such, Fig. 2.10 shows that in the 2050s (RCP 2.6) and 2080s (RCP 8.5), the probability of killifish inhabiting a given site remains unimpacted for some sites along the east coast, while some sites show reduced probability with habitat fragmentations.
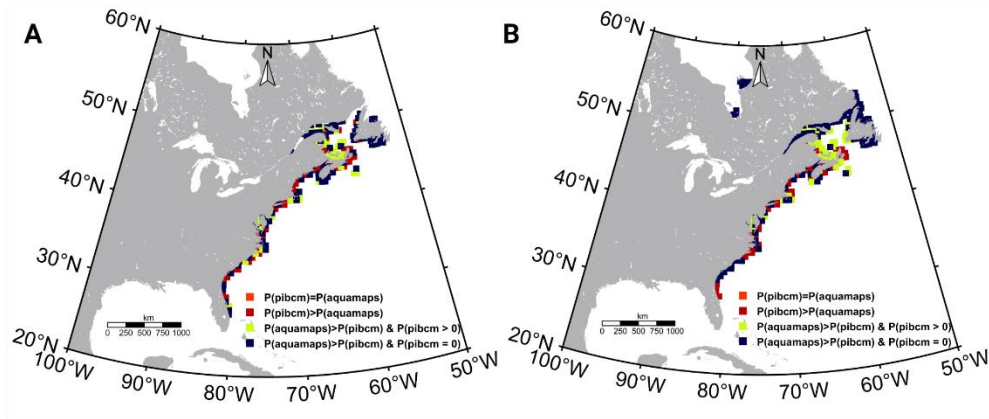
Fig. 2.10. Comparison of Atlantic killifish habitat probabilities predicted in physiology integrated bio-climate model and AquaMaps model for two representative concentration pathway (RCP) scenarios. (A) Differences in the habitat probabilities predicted by the physiology integrated model and the traditional species distribution model (AquaMap) during the 2050s(RCP2.6) and 2080s (B) (RCP8.5). P(pibcm) is the predicted Atlantic killifish habitat probabilities by the physiology integrated model and P(aquamaps) means the predicted habitat probabilities by the AquaMaps model.

2.4.7 Comparison between PIBCM and AquaMaps predictions

To compare the PIBCM predictions with a traditional bio-climate envelope model, we simulated an AquaMaps model for killifish distribution using climatological mean SST as the single environmental driver. According to AquaMaps predictions, the Northern and Southern boundaries of the Atlantic killifish native habitat will be expanded during the 2050s and 2080s (all RCP scenarios). The minimum northward range expansion was predicted for the 2050s (RCP2.6) and the maximum was in the 2080s (RCP8.5). The maximum range expansion was limited to the Canadian east coast at around 55ºN. This contrasts with our PIBCM predicted habitat distributions, which show little northward expansion. When comparing the predicted

population probability distribution by AquaMaps and PIBCM for a given RCP scenario, results showed contrasting likelihoods of killifish existence in a given habitat. Notably, as described earlier, the PIBCM model predicted habitat fragmentation, while AquaMaps did not.

2.5 Discussion

Our PIBCM indicates a complex habitat pattern shift that may lead to a series of disconnected killifish populations along the North American coastline in the future (Fig. 2.9). Notably, in contrast to the killifish habitat poleward expansion predicted in the traditional bio-climate model (e.g., AquaMaps), the PIBCM predicted aggregations to thermally optimal sites including shifts by the Northernmost killifish populations to their nearest Southern grids (Fig. 2.9). These predicted future habitat distribution patterns are likely to occur both over multiple generations as well as potentially within the lifetime of a fish.

Our model emphasizes the importance of integrating physiological responses into species distribution models. Most of the current habitat modeling techniques match organisms' contemporary environmental niche with the future environmental conditions (Grieve et al., 2017; Reygondeau and Beaugrand,2011) to predict future habitats but ignore organismal capacity to acclimatize. For example, a Generalized Additive Model that used SST as the primary parameter predicted an erroneous northward migration of *Calanus helgolandicus* (Chust et al., 2014), inferring the likely importance of integrating physiological responses. To this end, our approach provides a novel framework for integrating organismal physiological responses into species distribution models. Here, metabolic rate range (MRR) was considered the fundamental predictive unit that decides the physiological affinity of a given Atlantic killifish population to its geographical range. The central premise of our model is that a marine ectotherm will maintain its routine metabolic rates within a certain range in its current thermal environment, to sustain the

highest possible aerobic scope at that location. Instead of using aerobic scope directly, we used

MRR for our calculations for several reasons. At the fundamental level of the model building, we

considered a fish population in a given grid, maintains its optimum aerobic scope to inhabit the

grid and the MRR (the difference between routine metabolic rates calculated at the maximum

and minimum habitat temperatures) is a proxy of the maintained aerobic scope. Aerobic scope,

the difference between the maximum and routine metabolic rates at a given temperature is

considered a key physiological determinant of organismal fitness in a given habitat, and in this

study, we postulated that the Atlantic killifish habitat probability in a given grid is linked to the

grid-specific aerobic scope. However, the numerical calculation of aerobic scope is challenging

across the temperature range we consider in our study, since there is no established mathematical

equation to calculate maximum metabolic rates as a function of temperature. Therefore, we use

the routine rate ranges calculated based on the MTE equation as a proxy for aerobic scope. This

also enabled us to integrate the E value, which is indicative of the population-level biochemical

capacity of ectotherms to acclimate to rising ocean temperatures. Additionally, although it is

considered that organisms maintain their highest AS at the optimum temperature (Topt), the

laboratory-measured T_{opt} to some extent may not be correlated with the organism's preferred

functional habit temperature(Norin et al., 2004), and in several tropical fish species,' optimum

performance temperatures reflect more closely at the warmer end of their habitat temperature

range (Payne et al., 2016), that showed the ambiguities underlying the accurate estimation of the

most appropriate temperature to determine the optimum habitat AS. Therefore, we anticipated

that MRR provides a more robust physiological parameter to use as an indicator of optimum

thermal range. Collectively, we posit here that, MRR serves as a useful deterministic

physiological parameter, but further experiments are encouraged to prove its predictive capacity of species' habitat distribution patterns.

We assumed that even with increasing SST, a fish population would remain in the same grid cell if future MRR is lower or equal to the current MRR if the fish can increase its E value through acclimation, developmental or transgenerational plasticity. Indeed, we show a wide range of E values for different teleost ectotherms, inferring the critical role of thermal acclimation capacity and local adaptations likely to play in determining thermal optimums (Raffel et al., 2013; Sinclair et al., 2016). Accordingly, the MRR calculations can serve as an informative physiological parameter that can be integrated into other bioclimate models as well as statistical species distribution modeling approaches such as generalized additive or linear models (GAM or GLM) (Dhalke et al., 2018; Grieve et al., 2017).

Our model relies on several key assumptions that warrant further experimental analyses. For example, the metabolic theory of ecology (MTE Gillooly et al., 2001, Brown et al., 2004) equation and the oxygen capacity-limited thermal tolerance theory (OCLTT – (Portner and Knust,2007; Portner and Farell,2008; Portner,2001, Portner,2010) are critical to our model. There are significant debates on OCLTT (Clark et al., 2013; Portner and Giomi,2013; Portner,2014) and MTE (Clark,2004; Clark and Fraser,2004) frameworks. One widely debated aspect of MTE is the scaling exponent. Here, we used 0.75 as the mass scaling factor (Kleiber,1932), although this value may depend on lifestyle, swimming mode, ambient temperature, and other environmental variables (Killen et al., 2010, Norin Gamperl,2018). Previous studies have shown that the killifish scaling exponent is ~0.75 (Jayasundara et al., 2018), validating our approach. While species-specific scaling exponents should be considered in PIBCM in future studies, bioenergetic outcomes of acclimation and local adaptation to abiotic

environmental variables including temperature, are likely to be reflected via changes in the E value as also depicted by our sensitivity analysis. Despite the ambiguities around MTE and OCLTT frameworks, the two key attributes we focused on, aerobic scope (AS) optimization within a thermal window and the kinetics of the E value, are generally established. Here, we assumed that a given population optimizes AS by maintaining a routine metabolic rate within a certain range by regulating thermodynamic effects on enzyme activation energy (E). Importantly, we deviated from the MTE equation by postulating that the E value is species-specific and depends on the thermal history of a given population. For example, Kinnison et al. (Kinnison et al., 2003) demonstrated fluctuations in E values of locally adapted chinook salmon resulting in a counter-gradient variation between these populations. This is further inferred in our analysis of data from published literature on 19 different fish species shows that the E value ranges between 0.23 eV to 0.96eV. Moreover, our sensitivity analysis confirmed that a 0.1eV change in the E value can have a significant effect on metabolic rate output. We posit that changes in the E value reflect biochemical factors sensitive to environmental temperature changes. Therefore, acquired thermal tolerance following thermal acclimation or developmental plasticity can alter the E value, providing a quantifiable mechanistic parameter for our model. However, there is limited information on precise physiological factors driving the E value (Clark and Johnston,1999) and further research is warranted to examine a causal relationship between enzyme kinetic energy changes (e.g., enzymes involved in oxidative phosphorylation) and the whole organismal E value.

Based on E value calculations for killifish acclimated to 5-33°C with 5°C intervals (Healy and Schulte,2012), we predicted 12 different thermal envelopes for Northern and Southern killifish populations. This thermal envelope calculation allowed us to consider several

49

thermal ranges for the E-value calculations, thus enabling us to examine potential variation in the E value along the current habitat distribution while minimizing the effects of noise to some degree. This is likely an underestimation of the number of thermal envelopes, given that we rounded the climatological minimum and maximum temperatures to the nearest 5°C to match the acclimation data. Further, a major dividing zone between the Northern and Southern Atlantic Killifish subpopulations was identified at approximately around 40N, where the Hudson River enters the Atlantic Ocean (Duvernell et al., 2008). Studies also indicate a clear hybrid zone for the Northern and Southern subspecies; however, we did not account for this variation and demarcated 40N as the dividing line between the two. To account for the different thermal physiological capacities of fish in the hybrid zone, we simulated the model using several latitudinal clinal values (39-41N), but the final result for this region remained unchanged.

Inherent in our calculation is the assumption that killifish in a given thermal range are acclimated or locally adapted to their thermal range to optimize aerobic scope. Considering our calculations are based on lab-acclimated fish in 5°C increments, further studies at a higher thermal resolution will help to determine more accurate thermal envelopes. Also, our model assumes that the E value of a given subpopulation is consistent across all Northern or Southern fish within the native range and remains an important area of research to explore. While these studies will increase the resolution of the habitat shift patterns, the overall predictions from our study on limited poleward movement and habitat fragmentation are likely to remain consistent.

The downscaled data from the IPCC CMIP5 project is novel and our high-resolution coastal thermal variability addresses an important gap in our understanding of future temperature variability along the North American East Coast. We used a climate data interpolation method (Delta method (Nawarro-Racines et al., 2020) to downscale the coarser-resolution CMIP5 SST

data. This method requires a higher resolution baseline climatology to add up the thin plate splined (Frank,1982) delta values. World Ocean Atlas (WOA ~0.25o×0.25o, (Levitus et al., 2012) climatology data is a widely used baseline climatology, but coverage along the coastal zones is insufficient. Therefore, we developed a new climatology upscaling 0.05o×0.05o Copernicus SST data to increase the data resolution along the coastline and reduce the data source biasness. As expected, downscaled data from CMIP5 showed that SST along the killifish habitat range will be higher than its contemporary values (Fig. 2.7) and the anomaly increases with time and the RCP scenario. Dillon et al. (Dillon et al., 2010) concluded the importance of calculating metabolic rates using high-frequency temperature data (daily temperatures) to reduce the fallacy of the averages. Here we used annual maximum and minimum SST data to calculate the location-specific metabolic rates by averaging the daily variations and the effect on the final model prediction is negligible. Our downscaled data also showed several coastal habitats that are likely to experience wide thermal fluctuations in the future compared to the contemporary period. For example, Chesapeake Bay is predicted to experience the highest SST variation during future periods as depicted in the downscaled SST product. Further, our predicted values on increased thermal variability in several important coastal habitats (e.g., the Chesapeake Bay, Southern Gulf of Maine), infer at-risk regions for inhabiting ectotherms.

SST is the only environmental parameter used in our model, while AquaMaps (Kaschner et al., 2019) uses five environmental parameters (SST, sea surface salinity, sea ice content, primary productivity, and distance from the coast), and other SDMs have used a number of environmental parameters. Nonetheless, our PIBCM model can be extended to include other variables that are also determinants of MRR. Oxygen limitation from increasing oceanic hypoxic events may directly impact AS (Portner, 2001; Portner and Knust,2007; Portner and

Farrell,2008) and the Metabolic Index (Φ) (Deutsch et al.2015) and mitochondrial respiration and mitochondria health index (Jayasundara,2017)could serve as parameters to integrate changing DO and/or temperature levels into the PIBCM. Further, the inclusion of key parameters such as food availability, current fish density, and fishing pressures is critical for model accuracy and remains an important limitation in the current model.

Despite the limitations and important assumptions, our analysis on killifish presents a likely best-case scenario for coastal ectotherms and emphasizes the important role of acclimation and local adaptations that may play in determining the habitat shift patterns (Kuo and Sanford,2009; Somero et al., 2017). As the warming global oceans are predicted to push the species' preferred thermal envelopes northward (Chust et al., 2014), we demonstrate that aside from strict stenotherm (e.g., some Antarctic notothenioids (Beers and Jayasundara,2016), acclimation and local adaptions likely drive habitat fragmentations as well as aggregations, inter-subpopulation mixing, and genetic hybridizations for a coastal ectotherm. Species inhabiting lower latitudes experience habitat temperatures closer to their maximum tolerance limit, thus more likely to shift their current habitat ranges with the ocean warming (e.g., the wide habitat fragmentation predicted in this study for southern subpopulations of Atlantic killifish) than their northern conspecific species. Accordingly, certain regions of the North American coastline (e.g., Southern Gulf of Maine and Cape Cod coast) may serve as critical nursery habitats in the future while others may see significant declines in fish populations. Our PIBCM approach can estimate such habitats for economically important pelagic spices and indicate the critical need to incorporate thermal plasticity in population distribution models.

CHAPTER 3. TRANSCRIPTOMIC MODIFICATIONS UNDERLYING THE ATLANTIC KILLIFISH RAPID ADAPTATIONS TO EXTREME POLYCYCLIC AROMATIC HYDROCARBONS (PAHs) POLLUTION

3.1 Abstract

Several vertebrate species have rapidly adapted to extreme aquatic pollution and, thus serve as important model organisms to discover potential mechanisms underlying the pollution-tolerance. Atlantic killifish (*Fundulus heteroclitus*) is a popular vertebrate fish model that shows rapid adaptions to extreme levels of persistent organic pollutants (POPs) along the northeast coast of the United States, thus has been tested in various types of physiological and ecotoxicological experiments. In this study, we examined the transcriptome of Atlantic killifish to determine the molecular mechanisms underlying killifish evolved pollution tolerance, by comparing global transcriptomic data from 22 fish (11 from each site) from a highly polycyclic aromatic hydrocarbon (PAHs) polluted site in the Elizabeth River, Virginia (Republic (RP) site) and a clean site (Kings Creek (KC) site).

We observed the desensitivity of a key xenobiotic metabolic pathway, the aryl hydrocarbon receptor pathway (AHR pathway) in the RP fish, as a key molecular mechanism underlying killifish pollution tolerance. The detected downregulation of the *ahr2* gene in the RP fish postulates that the repeated AHR desensitivity could potentially be regulated through the *ahr2* gene downregulation in RP fish. There were 1607 genes differentially expressed (FDR <0.05) between the two groups, 514 genes were upregulated and 717 were downregulated in the PAH-resistant fish samples. The downregulated genes were uniquely enriched in processes like lipid metabolic processes, highlighting the reduced lipid metabolic processes in RP fish. Besides, the genes related to mitochondrial processes showed reduced plasticity of expression, drawing

another important conclusion that the mitochondrial process was reduced as a response to pollution tolerance. Lastly, we identified the importance of *cyp51*, *msmo1*, and *dhcr24* as potential novel biomarkers of vertebrate PAH adaptations. The results of the conducted weighted gene co-expression analysis (WGCNA) confirmed these results, thus the confidence in accepting these results is high in this study. Further controlled experiments are suggested to warrant the findings of this study, but this study emphasizes the importance of global transcriptomics for uncovering potential pollution adaptation mechanisms.

3.2 Introduction

Ocean pollution (the addition of anthropogenic wastes (predominantly hazardous compounds) into the ocean environment) has been intensifying at an alarming rate over the past couple of decades (Whitemee et al., 2015) causing a significant biodiversity loss (Landrigan et al., 2020), habitat modifications (by actively removing pollution-sensitive populations) (Landrigan et al., 2020) and destabilization of the interactions between the trophic levels across the marine food webs (Abbate et al., 2015; Bauer-Ceviello et al., 2019). Pollution occurs everywhere over the world's oceans, but the open ocean and interfaces between the deep sea and adjacent land mass (i.e., coastlines, bays, harbors, and estuaries) are more vulnerable to pollution by industrial effluents and agricultural runoffs (UNESCO-2019). With the spread of polluting agents in the natural environment, the inhabiting organisms respond independently, and the responses depend on the type, magnitude, rate of pollution, and species' physiological and evolutionary mechanisms. Some species are evolutionary more tolerant to pollution, compared to pollution-sensitive groups, thus starting to dominate the environment, resulting in a disruption in the existing ecosystem dynamics.

Understanding the species' responses to marine pollution is critical in several aspects, including fisheries management, conservation, and ecosystem and human health risk assessments. Given the species' differential evolutionary capacity for pollution, several species have been identified as important model organisms to determine the physiological, molecular, and evolutionary mechanisms underlying evolved pollution tolerance. Atlantic killifish (*Fundulus heteroclitus*) has evolved resistance to extreme levels of Persistent Organic Pollutants (POP) in the estuarine environments along the northeast coast of the United States and has been considered an important vertebrate model organism to study the vertebrate pollution tolerance mechanisms. Given the recent attention the Atlantic killifish received as a better vertebrate model, it has been tested in a range of physiological, behavioral, physiological, and evolutionary studies to discover mechanisms underlying vertebrates' rapid pollution adaptations (Aluru et al., 2015; Jayasundara et al., 2017; Nacci et al., 2010; Reid et al., 2016; Whitehead et al., 2017).

The enormous number of experiments conducted over the past couple of decades was able to establish a sound knowledgebase of the molecular processes underlying pollution adaptation. The desenstivity of the Aryl Hydrocarbon Receptor signaling pathway (AHR pathway, a ligand-activated transcription factor sensitive to PAHs, PCBs, and TCDD) is considered a key molecular mechanism underlying Atlantic killifish rapid PAH resistance (Aluru et al., 2015; Brown et al., 2016; reviewed in De Guilio and Clark,2015; Nacci et al., 2010; Oleksiak et al. 2011).

3.2.1 Aryl Hydrocarbon Receptor Pathway (AHR pathway)

The AHR pathway is a major regulatory mechanism underlying vertebrate xenobiotic metabolism (Aluru et al., 2011;2015; Di Giulio and Clark, 2015; Nacci et al. 2010). The AHR is a ligand-activated transcription factor active in mediating xenobiotic toxicity by inducing

xenobiotic enzyme activation to metabolize the dioxin (TCDD-2,3,7,8-tetrachlorodibenzo-p-dioxin) and other AHR-ligand (PAHs, PCBs, and dioxin-like-chemicals-DLCs) toxicity.   It has long been consistently identified that the Atlantic killifish POP adaptation is mainly with the global desenstivity of the AHR signaling pathway (Whitehead et al., 2017). Controlled experiments conducted using embryos of fish from common garden conditions found the effects of AHR activation with POP exposure, including altered gene expression, malformations, and decreased survivorship (Aluru et al. 2011;2015; Nacci et al. 2010), thus the repeated desenstivity of this pathway is considered an essential modification underlying adaptive pollution tolerance in Atlantic killifish (Aluru et al., 2011;2015; Di Giulio and Clark, 2015; Nacci et al. 2010).

The experiments found that the canonical mechanism of the AHR pathway initiated with cellular exposure to ligands like TCCD, PCB126, or PAHs, which dissociate the chaperone complex of hsp90 and AHR-interacting protein (AIP), in the cytosol (Fig. 3.1). Next, the AHR moves into the nucleus and dimerizes with AHR nuclear translocator (ARNT) in the nucleus, and further binds with Xenobiotic response elements (XRE) to activate a range of xenobiotic metabolizing enzymes including cyp1a, and cyp1b1, etc… to initiate the biotransformation and excretion of toxic compounds. However, exposure to poorly metabolized xenobiotics or constant exposure to exogenous xenobiotics leads to cellular toxicity, thus negative feedback mechanism of the AHR receptor protein (AHRR) activates next to reduce the AHR pathway processes. The AHRR provides negative feedback to the pathway by decreasing the activity of several chaperone proteins like XAP2, p23, and hsp90.

The repeated desenstivity of the AHR pathway observed in PAHs-tolerant Atlantic killifish is expected to reduce the cellular toxicity with continuous exposure to the excess amount of

exogenous PAHs, thus the majority of the AHR pathway genes are expected to maintain reduced

activity during the development phases.
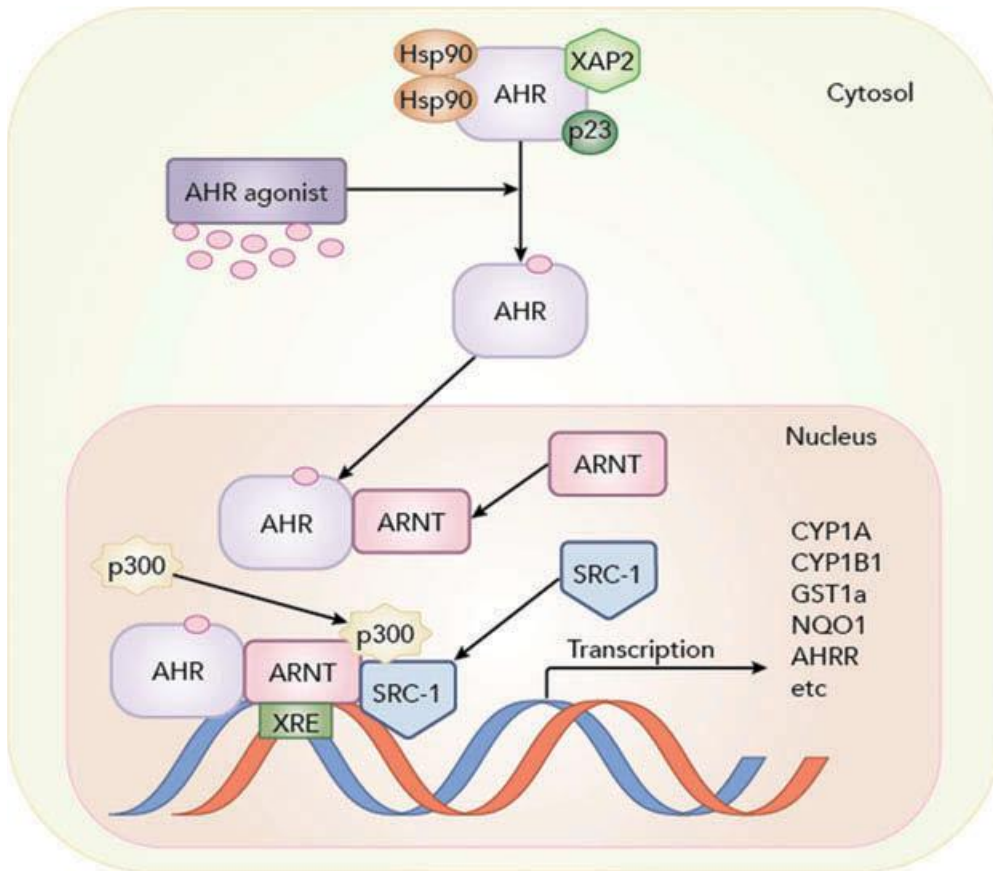
Fig. 3.1. The schematic diagram of the Aryl Hydrocarbon Receptor (AHR) pathway. Copied

from Di Giulio and Clark, 2015 (Respective authors have received necessary permissions to use

this figure from the original publisher)

 In this chapter, we focused on estimating the global gene expression in the Atlantic killifish

population that developed rapid adaptations to an extreme level of PAH toxicity in the Republic

Site (RP) Elizabeth River, Virginia. The core hypothesis we wanted test in this chapter was the

molecular mechanism underlying Atlantic killifish PAHs adaptations reflected in the fish global

transcriptomics profile. With the rapid advancement of cost-effective, hi-resolution sequencing

techniques, sequencing a large number of samples became more straightforward within a short period of time, so we used next-generation sequencing techniques to scan the killifish transcriptome.

3.2.2. Ribonucleic acid sequencing (RNA-Seq)

RNA-Sequencing is an experimental approach being used in current health and ecological contexts. The core concept of RNA-Seq analysis is determining the cellular ribonucleic acid (RNA) transcript abundance under a given condition compared to a controlled condition and determining any differential expression of genes/ transcripts between two conditions. The current high-throughput RNA-Seq technique is a result of a continuous effort of developing methods to characterize the changes in cellular RNA composition. Initially, gene expression studied via Expressed Sequence Tags (EST) method (Adams et al., 1992) that analyses the partial complementary DNA (cDNA) molecules, but with the identification of constraints in each step, the methodologies were advanced further, moving into serial analysis of gene expression (SAGE, Velculescu et al., 1995), microarray techniques and at last, the RNA-Seq techniques. RNA-Seq methods become extremely popular among the scientific community with the advent of high-throughput next-generation sequencing(NGS) methods that generate single-base resolution read sequences, which are critically important for identifying gene expression and detecting novel RNA transcripts, and better performed compared to its predecessor microarray techniques (Hrdlickova et al., 2017). With compared to microarray-based methods, RNA-Seq sequences are less noisy and applicable for a wide range of detections with direct sequence identity crucial for novel gene and transcript identification (Hrdlickova et al., 2017). With the technological and methodological advancement of RNA-Seq analyzing analyses provides greater opportunities to

study the quantitative and qualitative aspects of transcriptomic biology of a wide range of organisms from procaryotes to eucaryotes have been opened up.

Eukaryotes genotypes generate phenotypes through gene expression. The genetic information in the genome is transcribed and translated into proteins through a series of processes, which is considered the fundamental step of developing phenotypes through genotypic interaction. At the most fundamental level, three RNA molecules, messenger, ribosomal, and transporter RNA are involved in this process, but the synthesis and maturation of the RNA molecules are extremely controlled and highly networked to initiate the biological process. The regulation of gene expression is heavily dependent on the cellular conditions, but primarily diseases (i.e., cancers, diabetics) or environmental changes (i.e., thermal stress, pollution exposure) can modify the expression, thus the interest of RNA-Seq studies mainly contained to comparing RNA transcript profiles between two conditions (i.e. cancer and non-cancer patients). The RNA-Sequencing process can mainly be subdivided into three steps, library preparation, adapter ligation, and next-generation sequencing (Marguerat and Bahler, 2010). During the library preparation step, cellular RNA is extracted through a series of lab protocols and generated respective cDNA fragments for each RNA library. In the next step, adapters are ligated into each cDNA fragment as preparation for sequencing. It is considered that the library preparation is the critical step in the RNA-Seq process, which determines how closely the cDNA sequences represent the total RNA pool wanted to examine in the samples (Marguerat and Bahler, 2010). In the final step, next-generation sequencers like Illumina, PacificBio, Roche 454, Helicos Bioscience, and technologies sequence the prepared libraries for further downstream analysis to get the anticipated gene/transcript expression counts.

3.3 Materials and Methods

We followed three main steps to prepare samples, generate data, bioinformatics analysis, and treat the data statistically to obtain the differentially expressed and differentially methylated genes (Fig. 3.2), prior to the subsequent biological and molecular processes analysis (Gene Ontology enrichment and protein-protein interaction analysis). Sample preparation was carried out in the wet lab, followed by next-generation sequencing and bioinformatics analysis for quality controlling and mapping of the sequenced data, and extracting information for statistical analysis.
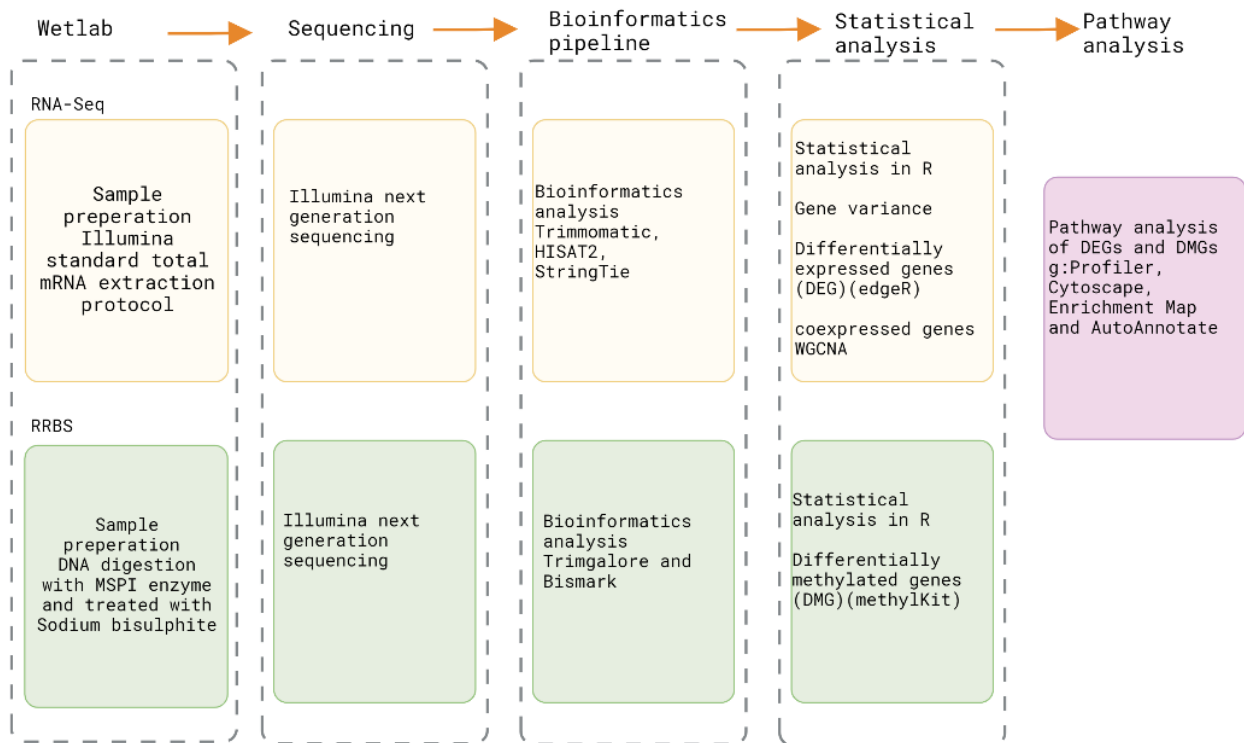
Fig. 3.2. The schematic diagram of the complete experimental protocol followed in chapters 3 and 4 data analysis. The first row represents the sequence of processes followed in the RNA-Seq analysis (chapter 3) and the bottom row processes were followed for the DNA-methylation analysis conducted to develop chapter 4.

3.3.1 Fish sampling sites

The ER is a tidal estuary with four main branches (east, west, south, and Lafayette river) that flow across Chesapeake, Norfolk, Portsmouth, and Virginia beach (Fig. 3.3). This river is historically considered an important commercial hub connecting the trade between Europe, the US, and Canada, and later a tactical naval shipyard that played a major role as a shipbuilding site during world war I and II. However, with rapid industrialization, urbanization, and later military activities, estuarine health started to degrade dramatically, recording an extraordinarily high sedimentary PAH concentration (2200mg/kg dry weight) (Alden,1995; Greaves,1990). Extensive use of PAHs-rich oil-derived Creosote in the naval industry is a major causative factor underlying the extreme PAHs contamination in the ER, and the legacies of the contamination are still active even after implementing extensive environmental remediations (Reviewed in Di Giulio and Clark, 2015). The Atlantic killifish subpopulation inhabiting the ER is considered an ideal study system to determine the molecular mechanisms underlying the rapid adaptations to extreme PAH toxicity. Republic (RP) site is one key extremely polluted site in the ER, with an abundant pollution-tolerant killifish population, thus the physiological, biochemical, and molecular properties of this fish subpopulation have usually been compared to the respective properties of fish in a relatively cleaner environment (Kings Creek) located north of the Republic Site (Fig. 3.3).
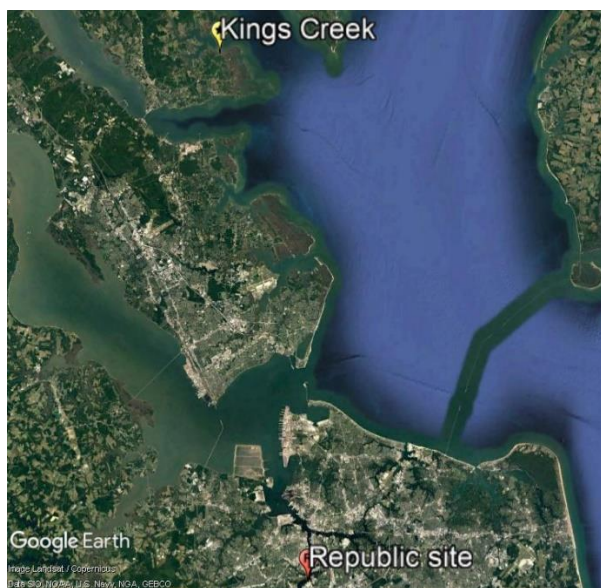
Fig. 3.3. Atlantic killifish sampling site for this study. Fish from Kings Creek (KC) was considered pollution-sensitive and fish from the Republic (RP) site were considered pollution-tolerant populations

Twelve female Atlantic killifish from each site (KC and RP site), were caught deploying a fish trap and bought into the laboratory at the Nicholas School of the Environment, Duke University, Durham for sample preparation. Fish with nearly equal lengths and weights were selected to reduce the biological variance in the statistical analysis. Fish were euthanized in ice and livers were extracted for sample preparation. Fish handling and sampling were conducted under the Duke University Institutional Animal Care and Use Committee (IACUC) approval A184-13-07.

3.3.2. Fish tissue sample preparation

The total hepatic RNA was extracted following a standard protocol that uses The RNeasy Plus Mini Kit (Qiagen, USA). The samples were homogenized using rotor–stator homogenizers with 500 µL of Buffer RLT (contains 1% of β-ME) and the homogenates were centrifuged for 30 seconds at ≥8000 x g by adding ethanol and then spun down with Buffer RW1 and Buffer RPE at

8000 x g for 30 seconds and 2 min respectively. At the end of the procedure, the RNA was eluted

in 25 μl nuclease-free water. The amount of RNA was quantified (per μl) using Nanodrop ND-

1000 (Thermo Scientific). Eleven biological replicates were conducted per treatment group.

3.3.3 Bioinformatics pipeline development

109 and 120 million paired-end sequence reads (2 × 126 bp) from KC and RP fish respectively

were sequenced in the sequencing machine. However, due to the lower sequence read counts in a

fish sample from each group, we decided to use data from 22 fish from both populations in the

subsequent downstream pipelines.

At the beginning of the pipeline, we trimmed out the Illumina TruSeq3 adapters, and poor-

quality (Q<20) reads in raw sequences using Trimomatic v0.36 (Bolger et al., 2014). Trimmed

reads were aligned to the Atlantic killifish reference genome v3.0.2 (downloaded from Ensembl

genome browser http://ftp.ensembl.org/pub/release104/fasta/fundulus_heteroclitus/dna/) using

HISAT2 v2.2.0 (Kim et al. 2019). The reference genome was indexed for splice sites and exons

coordinates available in the Atlantic killifish reference genome version 3.0.2 annotation file

(available in Ensemble genome browser http://ftp.ensembl.org/pub/release-

104/gtf/fundulus_heteroclitus/), employing HISAT2 scriptlets 'extract_splice_sites.py' and

'extract_exons.py'. Generated SAM files in the mapping process were sorted and converted into

BAM format using Samtools v1.15.1(Li et al., 2009) (http://www.htslib.org/). Mapped reads

were assembled into transcripts, quantified, and merged using StringTie v2.1.5 (Kovaka et al.,

2019). HISAT and StringTie software are part of the new memory-efficient and faster iterations,

the Tuxedo protocol (Pertea et al., 2015)) for sequence read alignment (HISAT) and assemble

(StringTie) into the reference genome. After the read alignment, we employed StringTie and the

python scriplet 'prepDE.py' (available on the StringTie online support page

) with the '-e' option to generate edgeR (the statistical package for downstream differential gene expression estimation) readable read count data. The bioinformatics pipeline development, testing, and processing were conducted in the University of Maine high-performance computer facility.

3.3.4 Analysis of transcript variation across the samples

We were interested in determining any specific patterns of transcript variation across the samples underlying the killifish PAH resistance. Therefore, we calculated the coefficients of variation (CV) of each transcript across the KC and RP fish samples. For that, we estimated the standard deviation of each transcript across each fish population and divided it from the mean (equation 1)

$$CV_{(transcript)} = Standard\ deviation_{(transcript)} / mean_{(transcript)} \qquad equation\ 1$$

We detected four distinctive groups of transcripts based on the calculated CV values, transcripts that did not detect in both groups (CV=0 in both groups), transcripts only detected in KC fish, transcripts only detected in RP, and transcripts detected in both groups. To determine the variation of expression between transcripts detected in both KC and RP groups, we calculated the ratio of CVs between KC and RP following equation 2 and considered the transcripts with higher ratio values had higher expression in KC populations and vice versa.

$$Ratio_{(CV)} = CV_{(KC)} / CV_{(RP)} \qquad equation\ 2$$

3.3.5 Statistical analysis for the gene differential expression

Transcript counts of the individual genes across the samples were statistically treated using Bioconductor package edgeR v3.38 (Chen et al., 2014) in R statistical environment (http://www.R-project.org/). edgeR estimates the differential gene expression by applying the

Fisher exact test (classic method) or generalized linear model (complex design) for the processed read counts. In edgeR analysis, we first filtered the low-expressed transcripts (if the row sum is zero). Filtered read counts were normalized to remove the technical effects in the samples (Evans et al. 2018) using the Transcripts Per Million Reads (TMM) normalization method and adjusted to remove the sample dispersion using 'estimateDisp' function before testing for the differential expression. Then the remaining transcripts were annotated to their respective gene name and tested for differential expression (between 11 fish from each population) by applying the 'exactTest' function (edgeR classic mode). Genes with an adjusted p-value (FDR) less than 0.05 (FDR<0.05) and an absolute log2 fold change higher than one were considered differentially regulated genes for this analysis. The site (KC and RP) was considered as the covariate parameter.

3.3.6 Gene co-expression analysis

To detect the Atlantic killifish genes co-expresed between the PAH-resistant and sensitive populations, we conducted a Weighted Gene Co-expression Network Analysis using the WGCNA package (Langfelder and Horvath, 2008) in the R statistical environment. WGCNA is a popular network analysis for detecting genes with similar expression patterns, grouping, and summarizing into principal components called module eigengenes (MEs). Prior to conducting the network analysis, we filtered out all the transcripts with a cumulative transcript count of less than 50 across all the KC and RP fish samples considering reducing the sample noise. The remaining transcripts were normalized using the Variance Stabilizing Transformation (vst) function in the DEGSeq2 R package (Love et al., 2014) next. Determining the threshold value is a critical step to accurately buildup the network tree, thus we employed the 'pickSoftThreshold' WGCNA function to define the threshold, which was 20 in this study. To create the dendrogram and MEs,

we used the 'blockwiseModules' function in the WGCNA package with 30 for the minimum

module size, 23749 for the maximum block size, and 'signed' for the topographic network. The

co-expression analysis groups genes into distinctive principal components (MEs) depending on

the gene expression, and genes are tightly correlated and each ME has a set of highly correlated

genes called hub genes that are more likely involved in specific biological functions. The

definition of hub genes is arbitrary, thus in this study, we defined the genes as hub genes, if the

absolute correlation is higher than 0.8 (|correlation|>0.8).

3.3.7 Protein-Protein interaction (PPI) network analysis for identified groups of genes

We conducted protein-protein interaction (PPI) network analysis for the up and down-regulated

differentially expressed genes and for the hub genes identified in the module eigengenes except

the gray ME using STRING v11.5 (https://string-db.org/) web-based tool. We set the species to

'*Fundulus heteroclites*' and the interaction score to high confidence (0.7) and construct the PPI

network. The resulting networks were visualized using Cytoscape http://www.cytoscape.org/

(Shanon et al., 2003) and the highly connected networks were clustered employing the MCODE

package with the max depth parameter set to 5 and more score cutoff to 0.2. Among all of the

MEs, only ME2,3, and 9 hub genes showed considerable connections among each other, thus we

focused on conducting further analysis targeting the hub genes in those three MEs.

3.3.8 Enrichment analysis for transcriptomic analysis data.

Gene functions and regulatory mechanisms of the detected killifish genes were determined using

the web-based platform g:Pofiler (https://biit.cs.ut.ee/gprofiler/gost) (Raudver et al., 2019). It is a

regularly updated popular platform for checking gene ontologies (GO) of the enriched genes

from various sources of omics data sets. In this study, we were interested in uncovering the

molecular, cellular, and biological functions of the up and down-regulated genes, and the hub

genes in ME2,3, and 9. Accordingly, we determined the GO enrichment of ordered lists of

differentially expressed up and down-regulated genes, and hub genes (not in order) in the

selected MEs using g:Profiler.

We set the g: SCS (g: Profiler-specific p-value correction algorithm) to 0.05 to estimate the

significantly enriched GO terms. The resulting GO term networks were visualized using

Enrichmentmap in Cytoscape by applying the q-value cutoff to 0.05 and the 0.375 similarity

index between two nodes. At last, we were interested in extracting the most described GO

processes in each network, so we used AutoAnnotate app in Cytoscape setting the number of

frequent words for each cluster to three.

3.4 Results

3.4.1 Transcript variability across samples

We used a total of 109 and 120 million sequenced mRNA reads from the respective KC and RP

fish samples and the trimming off of low-quality reads (Q<20) resulted in 88 and 107 million

clean reads (Table 3.1). The clean read distribution trend was nearly unique across the samples,

thus reads were mapped to the Atlantic killifish reference genome (version 3.0.2) without

additional filtrations and normalizations. The reads were mapped to the reference genome with

average mapping efficiencies of   81.3% and 81.5% in respective KC and RP samples (Table 3.1)

and annotated into the available 23469 Atlantic killifish transcripts available in the Atlantic

killifish annotation version 3.0.2.

To estimate the phenotypical gene expression variability between KC and RP fish samples, we

calculated the coefficient of variation (CV) for each transcript across KC and RP fish samples.

We identified 2996 transcripts with zero CVs in both populations (transcripts that weren't detected in both fish populations), 1508 transcripts that were only detected in KC fish samples (zero CV in RP fish), 1399 transcripts detected in RP fish only (zero CV in KC fish) and 17566 transcripts with estimated CVs in both groups (transcripts detected in both groups) (Fig. 3.3). The calculation of ratios of CVs of the detected transcript in both groups found 14469 transcripts had higher CV in KC populations compared to RP populations and 2937 transcripts vice versa (Fig. 3.4). To examine the functions of the transcripts that were not detected in both populations (transcripts that were potentially unresponsive to the ambient environmental conditions), we conducted a GO terms analysis using g:Profiler online annotation software and found a considerable portion of genes was enriched in the Atlantic killifish cellular functions like 'membrane' (GO:0016020), 'integral component of membrane'(GO:0016021), and 'intrinsic component of membrane' (GO:0031224) (Fig. 3.5). The top five significant GO terms enriched in the gene that wasn't detected in both fish populations were molecular functions (MF) (G protein-coupled receptor activity- GO:0004930, transmembrane signaling receptor activity- GO:0004888, molecular transducer activity- GO:0060089, signaling receptor activity- GO:0038023, olfactory receptor activity- GO:0004984).

Table 3.1. Summary of the RNA-Seq bioinformatics analysis

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
KC fish	**Sequence Depth (millions)**	10.35	10.96	9.34	6.96	8.5	8.17	7.09	8.26	9.16	9.84	10.76
	Reads after trimming (millions)	9.85	10.15	9.04	6.47	6.56	6.61	6.59	6.62	6.62	6.67	6.57
	Percentage %	95.19	92.66	96.75	92.91	77.19	80.9	92.97	80.07	72.28	67.82	61.06
	Mean alignment rate	70.45	64.71	56.53	47.01	63.22	60.59	61.53	66.14	68.69	59.55	72.71
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
RP fish	**Sequence Depth (millions)**	10	12.57	10.75	11.08	12.04	15.2	8.97	11.19	12.45	13.67	11.96
	Reads after trimming (millions)	6.55	11.97	10.07	10.49	11.2	8.99	7.3	10.59	11.77	12.96	11.42
	Percentage %	65.51	95.24	93.65	94.7	93.04	59.13	81.36	94.68	94.55	94.78	95.5
	Mean alignment rate	67.22	72.59	72.02	77.42	73.04	71.87	74.68	62.83	71.36	74.08	70.11
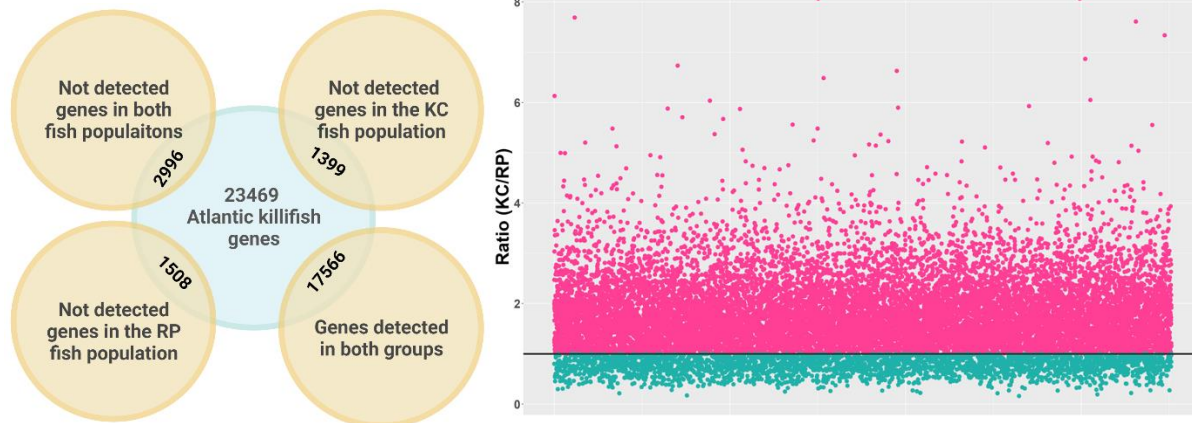
Fig. 3.4. Genes detected in two populations (left) and the ratio of gene variance in KC fish compared to the RP population (right) (The dark line is the ratio=1 line)
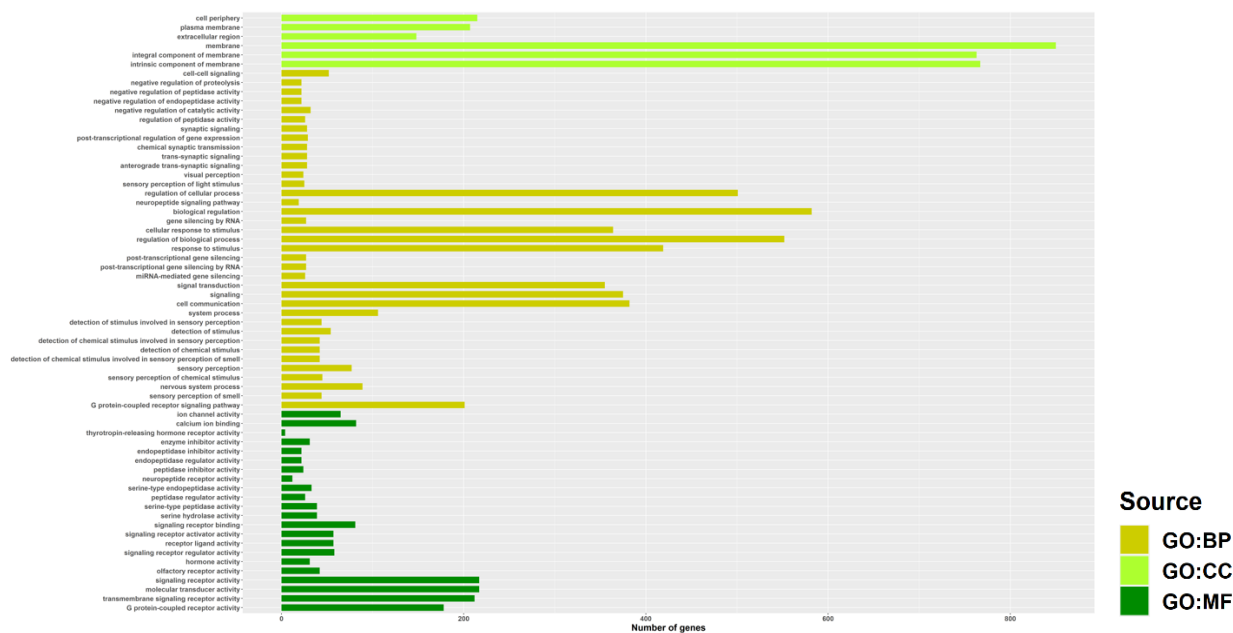
Fig. 3.5. The GO enrichment terms of the not-detected genes in both populations.

3.4.2 Differential gene expression between KC and RP fish

We used the transcript counts count data resulting in the bioinformatics analysis to determine the gene differential expression between KC and RP fish populations. Prior to the downstream statistical analysis, we filtered out transcripts with zero counts across all the samples (row

sum=0) (transcripts that were not detected in both groups) and the remaining 20473 transcript data were used in edgeR to calculate the expression differences using 'exactTest' function (edgeR classic mode). The differential expression analysis found 1607 transcripts with adjusted p-value (FDR) less than 0.05 (FDR<0.05) with 717 downregulated (log2 fold change <1) transcripts and 514 upregulated (log2 fold change >1) in RP fish compared to KC fish (Fig 3.6). Among the differentially expressed genes, the aryl hydrocarbon receptor-like (*ahr2*) gene (a key gene in the AHR pathway regulating the pollution resistance mechanisms in Atlantic killifish) was downregulated in the RP fish (log2 FC= -1.29), but no other AHR pathway gene was found differentially expressed. Repeated de-sensitivity of the cytochrome P450 1 family genes, more specifically the cytochrome P450 1A (*cyp1a*) de-sensitivity has been considered a biomarker of Atlantic killifish PAHs recalcitrance, but none of the CYP1 gene family including *cyp1a* was detected expressed differentially in this study. On the contrary, several genes in another isoform member of the Cytochrome family (cyp2), like cytochrome P450 family 27 subfamily C member 1(*cyp27c1*), cytochrome P450 family 2 subfamily R member 1(*cyp2r1*), cytochrome P450 2J6-like (*cyp2n13*) were upregulated in the RP fish and cytochrome P450, family 51 (*cyp51*) gene was downregulated in the RP population.
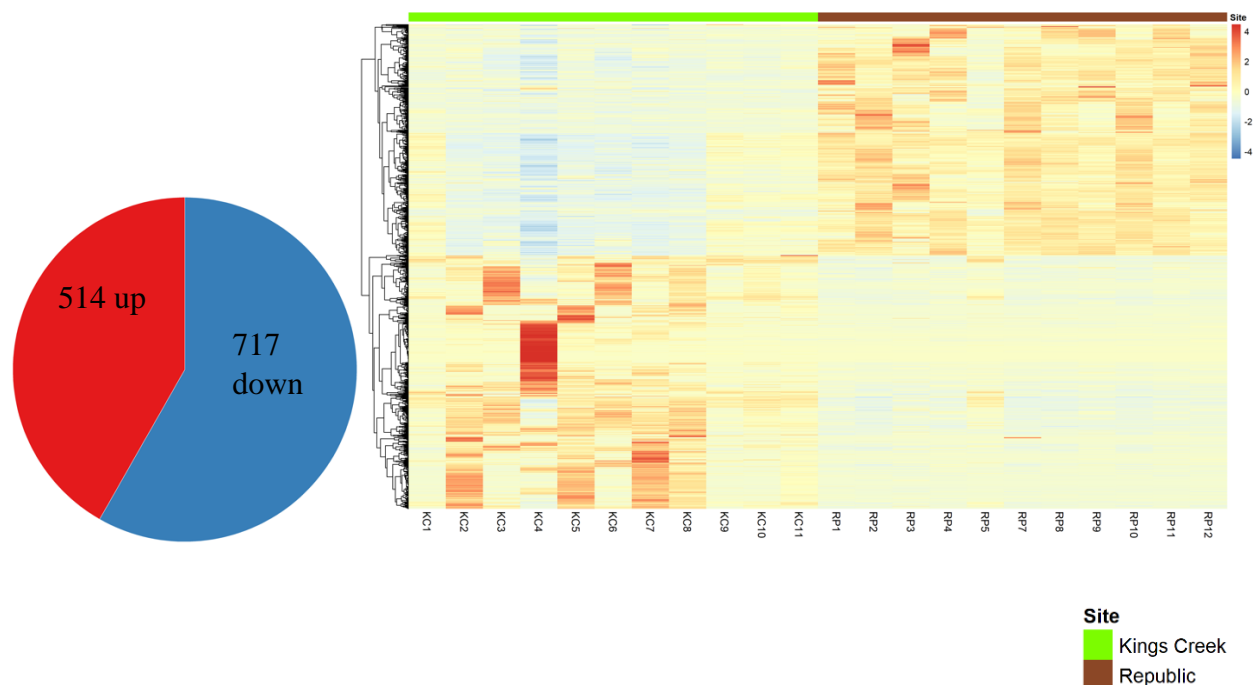
Fig. 3.6. Pie distribution (left) and the heat map of the differentially expressed genes

To better understand the functions of the differentially expressed genes, we conducted a gene

ontology enrichment analysis for each up and down-regulated gene separately, using g:Profiler

web server. The enrichment analysis for the downregulated genes found 22 significantly

enriched (g: SCS <0.05) biological processes (BP) and 12 molecular functions (MF) (Fig. 3.6).

The top five most significant downregulated biological processes were lipid metabolic process

(GO:0006629), sterol biosynthetic process (GO:0016126), steroid metabolic process

(GO:0008202), sterol metabolic process(GO:0016125) and lipid biosynthetic

process(GO:0008610) (Fig.3.6) and acyltransferase activity, acyl groups converted into alkyl on

transfer (GO:0046912), triplet codon-amino acid adaptor activity (GO:0030533), intramolecular

oxidoreductase activity, transposing C=C bonds(GO:0016863), symporter

activity(GO:0015293), and oxidoreductase activity(GO:0016491) were the top most significantly

enriched molecular functions of the downregulated genes (Fig.3.7). Upregulated genes in the RP

killifish were most significantly enriched in the biological processes like small molecule metabolic process(GO:0044281), carboxylic acid metabolic process(GO:0019752), oxoacid metabolic process(GO:0043436), organic acid metabolic process(GO:0006082), and alpha-amino acid metabolic process (GO:1901605), and in the molecular processes like transaminase activity (GO:0008483), transferase activity, transferring nitrogenous groups (GO:0016769), vitamin binding (GO:0019842), oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616) and in oxidoreductase activity, acting on CH-OH group of donors (GO:0016614) (Fig. 3.8)
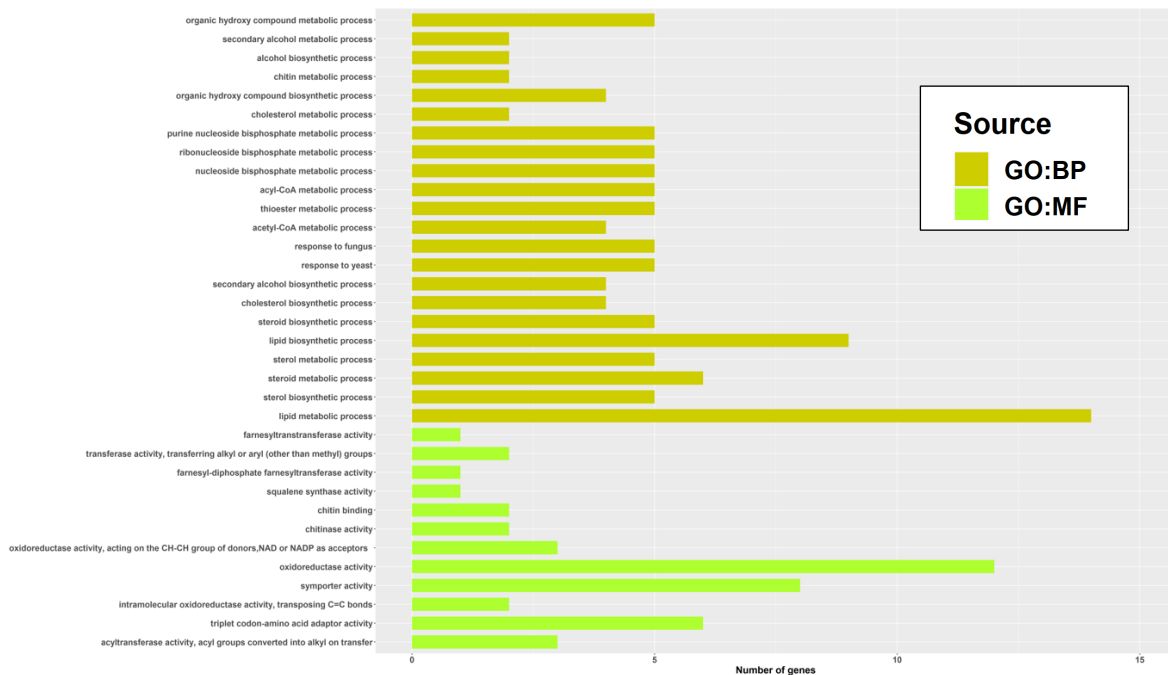
Fig. 3.7. GO enrichment of the downregulated genes in the RP population
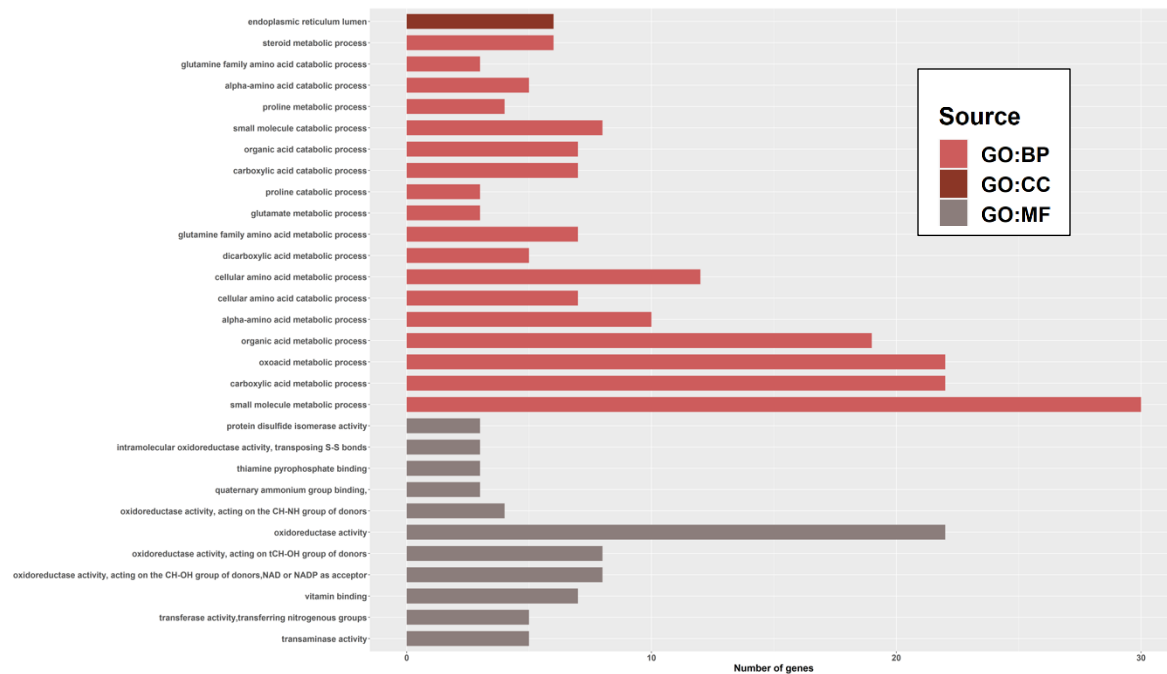
Fig. 3.8. GO enrichment of the upregulated genes in the RP population

The MCODE clustering in the Cytoscape categorized the majority of enriched GO terms of the downregulated genes into two main clusters, lipid biosynthetic processes and metabolic processes (Fig. 3.9) while most of the upregulated GO terms were related to molecule catabolic and metabolic processes (Fig. 3.9). The Protein-Protein Interaction (PPI) map of the DEGs identified several functionally related genes clustered and the most distinct and tightly connected network was comprised of nodes cytochrome P450, family 51 (*cyp51*), methylsterol monooxygenase 1 (*msmo1*), regulator of G-protein signaling 5 (*mvd*), NAD(P) dependent steroid dehydrogenase-like (*nsdhl*), transmembrane 7 superfamily member 2(*tm7sf2*), 24-dehydrocholesterol reductase (*dhcr24*), phosphoethanolamine N-methyltransferase 3 (ENSFHEP00000002111) and hydroxymethylglutaryl-CoA synthase, cytoplasmic-like (ENSFHEP00000007962) proteins (Fig. 3.10), which were entirely downregulated in the RP fish population.

74
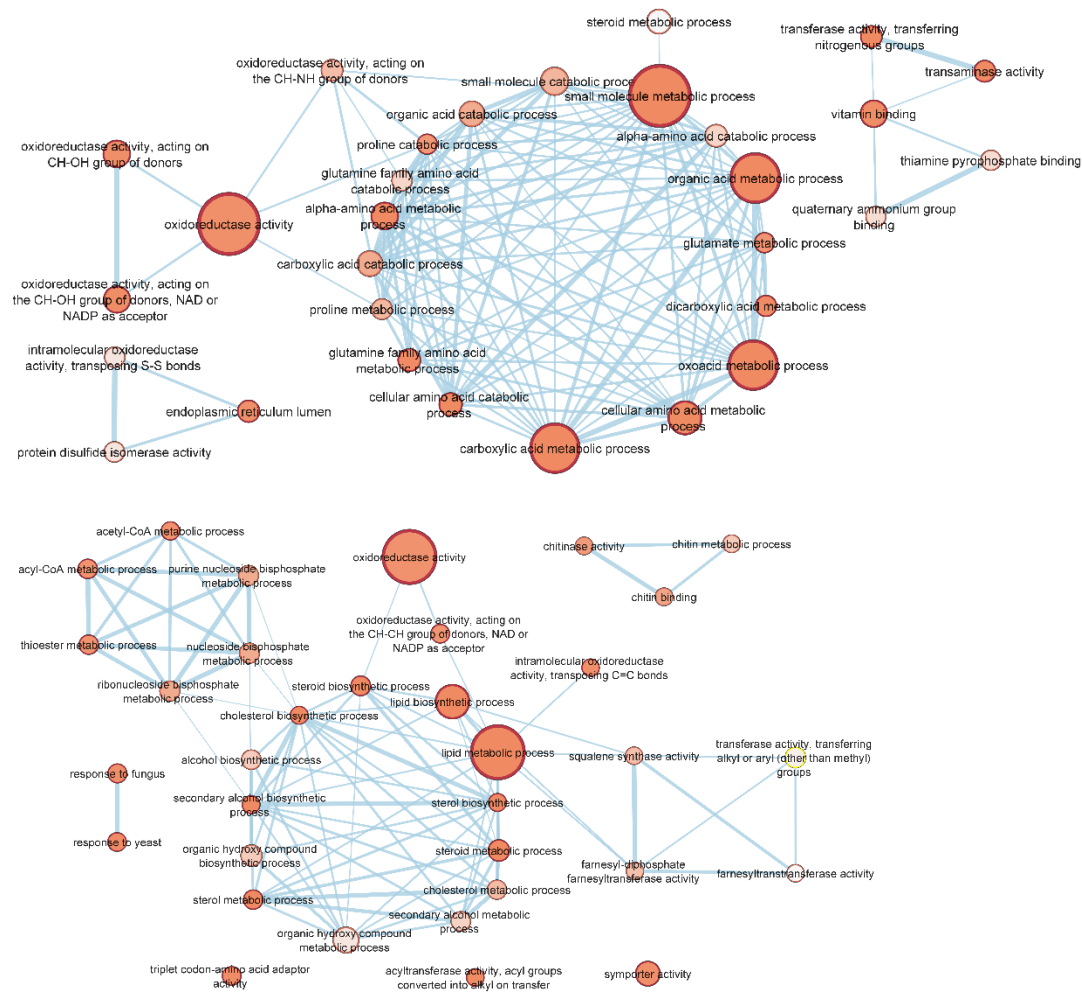
Fig. 3.9. Cytoscape clusters of the GO enrichment of up (top) and down regulated (bottom) genes
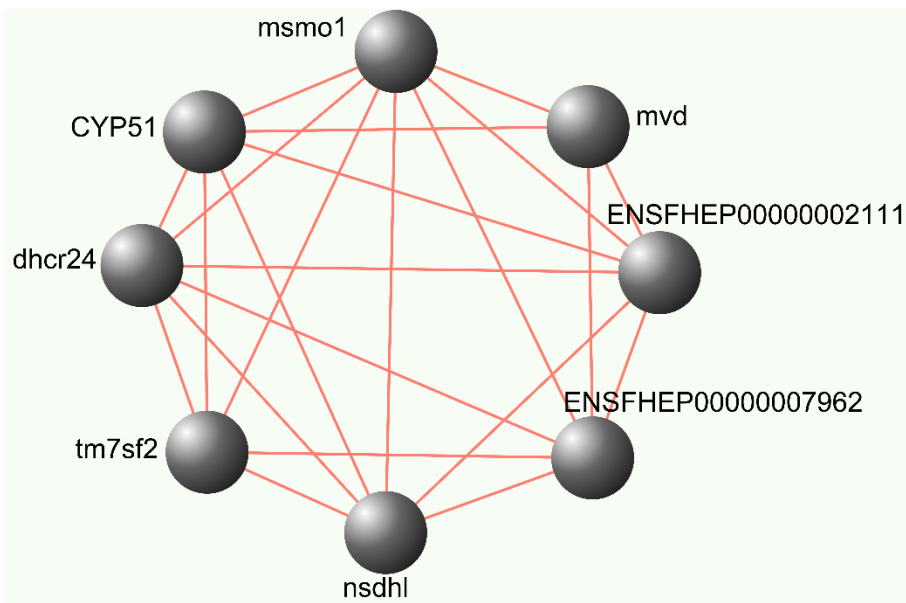
Fig. 3.10. String protein interaction map of the downregulated genes

3.4.3. Gene co-expression between two fish groups

One of the main focuses of this analysis was to determine the killifish molecular processes

expressed simultaneously supporting the evolved resistance to PAHs toxicity in the RP site. To

this end, we implemented a WGCNA analysis using 23469 mapped Atlantic killifish transcripts

that detected 14 gene clusters with similar expression patterns (Fig.3.11). The gray module

contained 7210 genes (non-co-expressed genes) and the turquoise module contained the most

highly co-expressed 4205 genes in this study (Fig. 3.11). Tan and Salmon modules contained 34

genes, which were the least number of co-expressed genes detected in the Atlantic killifish

samples.
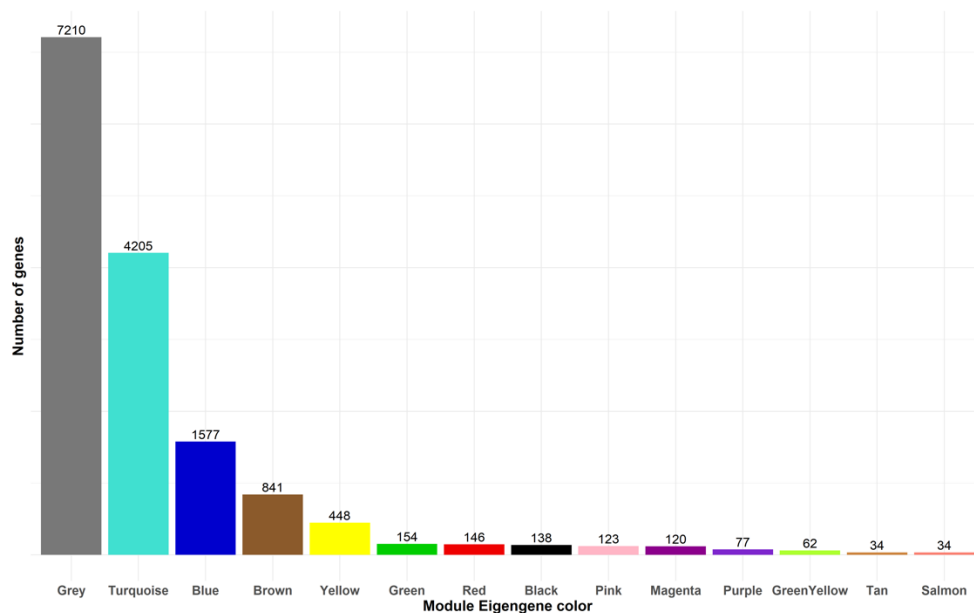
Fig. 3.11. WGCNA clustered co-expressed genes in identified 14 gene modules

Each ME has a set of highly correlated genes called module members (MM) and in this study, we defined MM if a gene absolute correlation exceeds 0.8 (|correlation|>0.8). Accordingly, we found a range of MMs in each ME (Table 2) and we were interested in determining the connectivity of these genes into a specific biological, molecular, or cellular process. So, we first checked the protein-protein interaction network of MMs in each ME and if MMs were connected sufficiently, we then analyzed the MMs GO enrichment terms subsequently. In the PPI network analysis, we found MMs in modules 2,3 and 9 only showed a higher degree of connectivity to other nodes (Fig. 3.12), thus we conducted GO term analysis for those three sets of hub genes. The GO term analysis of the MMs in ME2 detected 40 GO terms enriched in the tested hub genes (Fig. 3.12). The top five most significant biological processes of the ME2 hub genes were, the generation of precursor metabolites, and energy (GO:0006091), small molecule metabolic process (GO:0044281) organic acid metabolic process (GO:0006082), energy derivation by oxidation of organic compounds (GO:0015980), and carboxylic acid metabolic process

(GO:0019752) (Fig. 3.12). The MCODE network analysis in the Cytoscape detected the 40 GO terms enriched in ME2 hub genes belonged to a unique common functional cluster of 'metabolic process/activity (Fig. 3.12). In contrast to the identified 40 GO terms enriched in the second ME hub genes, the third ME hub genes were enriched only in 11 GO terms (Fig. 3.12) and the top five biological processes were cellular macromolecule biosynthetic process (GO:0034645) translation (GO:0006412), peptide biosynthetic process (GO:0043043), amide biosynthetic process (GO:0043604), and peptide metabolic process (GO:0006518) (Fig. 3.12). The MCODE clustering showed us that all these GO terms belonged to a common group of terms related to peptide biosynthetic processes (Fig. 3.12). The developed PPI network of the ME9 hub genes was closely related to the downregulated gene PPI network. The top five enriched GO terms (out of nine total GO terms) of the ME9 hub genes were lipid biosynthetic process (GO:0008610), lipid metabolic process (GO:0006629), steroid metabolic process (GO:0008202), sterol biosynthetic process (GO:0016126), and sterol metabolic process (GO:0016125) (Fig. 3.12). The MCODE analysis confirmed that all of these GO terms had a shared origin of 'metabolic/oxidoreductase process' (Fig. 3.12) and several genes like *CYP51*, *msmo1*, and *dhcr24* were identical with the genes identified in the distinct PPI network of the differentially expressed genes. The identification of similar gene networks in the DEGs and ME9 hub genes was compelling.
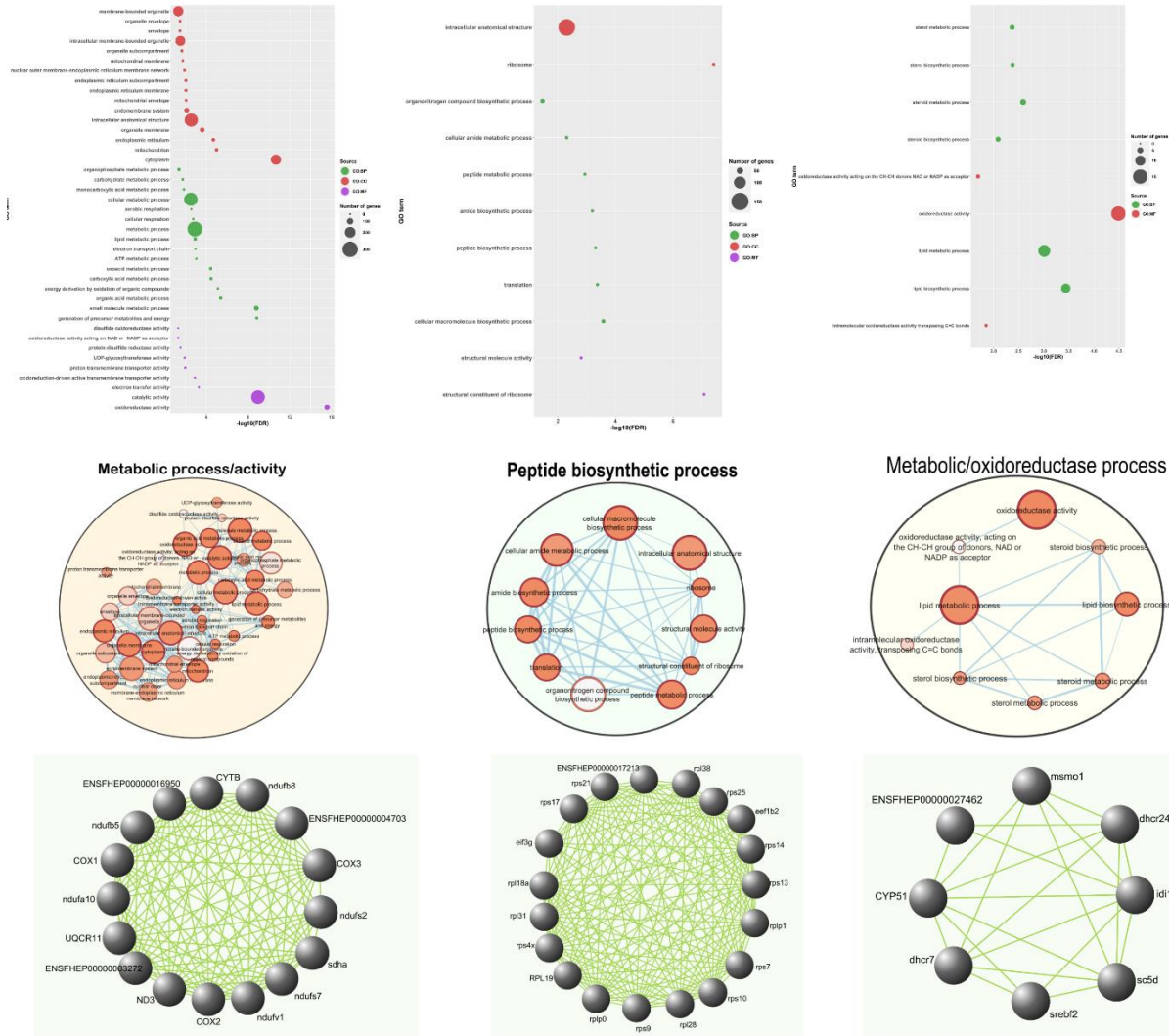
Fig. 3.12 The respective GO enrichment, Cytoscape and PPI interaction maps of the most connected genes (hub genes) of module eigengene 2 (3 figures in the first raw), 3 (second raw) and 9 (last raw).

3.5 Discussion

The molecular and evolutionary mechanisms underlying the Atlantic killifish evolved pollution tolerance have not been fully understood due to the complexity of the mechanisms and the sampling and technical limitations. However, the recent rapid advancement of sequencing

techniques and bioinformatics has permitted us to compare the whole-genome modifications of fish from multiple populations to explore the diverse molecular and evolutionary mechanisms that evolved underlying killifish resistance to environmental pollutants. Existing data suggest that repeated recalcitrance of the aryl hydrocarbon receptor (AHR) pathway (Aluru et al., 2015; Nacci et al., 2010; reviewed in De Guilio and Clark, 2015) through natural selection (Osterberg et al., 2018; Reid et al., 2016;2017, reviewed in Whitehead et al., 2017) is considered a key molecular mechanism underlying the killifish pollution resistance. Besides these identified mechanisms, more molecular mechanisms could potentially involve underlying the killifish pollution tolerance but remained to be uncovered. We hypothesized that there could be more novel mechanisms underlying killifish pollution tolerance reflected at the transcriptome level, thus in this study, we scanned the global transcriptomics of pollution-tolerant and sensitive fish to compare the gene expression profiles and identified several mechanisms inherent to pollution resistance. Our study confirms that the desensitizing of the AHR pathway, potentially through the downregulation of the *ahr2* gene is fundamental for killifish evolved pollution tolerance. Reduced activity of lipid metabolic processes and the shrinking plasticity of genes related to mitochondrial aerobic respiration were detected in the pollution-tolerant fish, which we consider novel and important molecular processes underlying killifish PAHs tolerance.

3.5.1 Gene variance across the two populations.

In the transcriptomic analysis, we found 2996 Atlantic killifish genes that weren't expressed in both populations (Fig. 3.4), and most of these genes were enriched in several cellular functions (e.g., cell periphery, membrane, and integral component of membrane), molecular functions related to ion metabolism and biological processes related to organismal sensory performance (Fig. 3.5). This observation could be an indication of reduced cell membrane and sensory

processes in wild Atlantic killifish, but further experiments are required prior to generalizing this finding across all killifish populations. However, we did not detect a direct involvement of epigenetic mechanisms underlying observed gene silencing in this study. These findings further emphasize the importance of testing wild killifish populations, in the context of creating a complete understanding of the molecular and evolutionary processes underlying killifish population tolerance, which has also been discussed in several earlier studies (Brauer et al., 2017; Reid et al. 2016; Whitehead et al., 2010) but with extreme caution when interpreting experimental outcomes (Pavey et al., 2012). The comparison between the genes detected in both fish populations (17566), found a relatively low number of genes (2937) with higher variance in the RP population compared to the KC fish (Fig. 3.4) (rest had a higher variance of expression in KC fish compared to RP fish). We consider this reduced plasticity in the pollution-tolerant killifish population result in due to the eroded genetic diversity (Bijlsma & Loeschcke, 2012), and adaptations to ambient environmental toxicity (del Mar Ortega-Villalzian et al., 2011). The gene pool of a population may have weaker plasticity, with the evolved adaptations to the local conditions, thus the gene expression variance could be limited (Bijlsma & Loeschcke, 2012), or natural selection can selectively favor several genotypes that help to survive in the stressful environment (Chevin & Lande, 2011; Paschke et al.2003). The central notion of the observed reduced gene expression plasticity in the PAHs-tolerant Atlantic killifish population clearly overlaps with the first explanations, thus we posit that the gene pool of the Republic killifish has eroded with the evolved PAHs tolerance. Twenty reduced variance genes shared evolutionarily selected loci, but the functionality of these genes supporting killifish pollution tolerance wasn't clear. The genes with a variance of expression higher than two between two populations were mainly enriched in aerobic respiration and mitochondrial processes (Fig. 3.5), which is leading to

a conclusion that Atlantic killifish has a potential cost underlying their evolved resistance to PAHs, reduced mitochondrial aerobic respiration. The phenotypic and physiological constraints of the reduced aerobic respiration processes in wild killifish populations should be evaluated in future experiments.

3.5.2 Gene differential expression

The gene expression analysis detected 1607 differentially expressed genes (FDR<0.05) where 514 were upregulated (log2 fold change >1) and 717 downregulated (log2 fold change <1) in the pollution-tolerant fish compared to pollution-sensitive population (Fig. 3.6). The downregulated genes were enriched in several biological processes related to lipid metabolic biosynthesis (Fig. 3.6), so we posit that the hepatic lipid metabolism was decreased in the PAHs-tolerant Atlantic killifish. The fish liver is the key organ that deals with the metabolizing of incoming toxic compounds (including PAHs) from the environment, so modifications are required for killifish to survive in an environment with extreme toxicity. Riley et al. pointed out that the PAHs tolerant Atlantic killifish liver is adapted to store PAHs globules in the liver to avoid PAHs catabolism, potentially causing ultimate liver lesions. As we found in our study, the reduced liver lipid metabolism in PAHs tolerant killifish could be a protective mechanism to avoid the deposited PAHs being dissolved in the liver tissues. Given that, our finding of reduced hepatic lipid metabolism processes further strengthens Riley et al. conclusion, but further experiments are required to warrant it. The Protein-Protein Interaction network of the differentially expressed genes indicates a few distinct gene networks, and the main network was comprised of highly connected downregulated genes (Fig. 3.10), which were mainly involved in the lipid biosynthesis process (*cyp51*, *msmol*, *dhcr24,* and *nsdh1*). Given this network study, we conclude that the reduced lipid biosynthesis was regulated by downregulating the lipid metabolic genes, which

were functionally connected to each other. *msmo1* and *nsdh1* were two highly connected nodes in the main PPI network (Fig. 3.10), which are functionally regulating cholesterol biosynthesis, and it was found that the co-expression of these two genes regulates adipogenesis (Bai et al., 2018; Bauer et al., 2018). The observed downregulation of these two genes in the RP killifish could be signaling the reduced adipogenesis underlying the evolved PAHs tolerance in Atlantic killifish. However, as we emphasize above, these findings should be properly validated to reduce any potential confounding signals reflected in the gene structure of fish from wild populations. The downregulation of *cyp51* in the PAHs tolerant killifish may be a response underlying the evolved PAHs resistance or may be reflecting the contamination of *cyp51* inhibitors like Triadimefon (TDF) that has been used in recent agricultural practices as a fungicide (Liu et al., 2017), in the RP site, thus the chemical contamination in the Republic site should be properly evaluated to confirm the gene-chemical interaction.

3.5.3 Gene co-expression across both populations

We conducted a whole gene co-expression network analysis for the annotated transcripts and checked the functional level connectivity of the most connected genes (hub genes; the genes with absolute correlation > 0.8) in the identified gene modules using the String protein network. Except for gene modules 2,3, and 9, (turquoise, blue, and red modules in Fig. 3.11) hub genes in the other modules did not show connectivity at the protein level, thus we excluded these modules from our further analysis. The PPI network of module 9 hub genes consisted of eight downregulated genes in the PAH-resistant fish population, thus we can infer these genes were negatively correlated but co-expressed between the two populations. There were three common genes (*cyp51*, *msmo1*, and *dhcr24*) functional for lipid metabolism, identified in the co-expressed and differentially expressed genes PPI networks, leading us to describe the role of these three

genes as biomarkers of reduced lipid metabolism underlying the evolved PAHs tolerance in

Atlantic killifish. The biological processes resulting in the GO enrichment analysis of the red

module hub genes strongly agreed that these genes are participatory genes of lipid metabolism

(Fig. 3.12), which confirms the notion that lipid metabolic processes were downregulated as well

as co-expressed in the PAHs tolerant population, underlying the evolved resistance. The most

connected genes of module 2 (turquoise module), i.e., *cox1*, *cox2*, *cox3*, *ndusf2*, *ND3* (Fig. 3.12),

were closely enriched in several biological processes like cellular respiration (i.e., aerobic

respiration, cellular respiration), cellular components like membrane functions (organelle

envelope, envelope) and several molecular functions like acid metabolic process (Fig. 3.12), but

none of these genes were differentially expressed in this study. On the contrary, the module 2

hub genes variance of expression was lower in the republic fish compared to the King's creek

population, indicating convergent expression of a group of co-expressed genes resulting in

reduced activity of the Republic fish mitochondrial aerobic respiration processes but the

selection targets or epigenetic significance underlying these genes was not clear.

The PPI network of the module 3 hub genes was mainly comprised of ribosomal protein genes

(i.e., *rpl18a*, *rpl19*, *rpl28*, *rpl31*, *rps9*, and *rp10*) (Fig. 3.12) which functionally regulate

biological processes like peptide metabolic and biosynthetic processes, amide biosynthetic

processes, and cellular functions like intercellular anatomical structure (Fig. 3.12).   None of

these genes were differentially expressed, the directionality of the gene expression variance

wasn't clear, and did not share pathways with natural selection or epigenomic processes in this

study. Several ribosomal genes like *rps9* (ribosomal protein S9) have been identified as potential

biomarkers of metal contamination in mollusks (Joris et al., 2006), thus we conclude that the role

of ribosomal genes underpinning Atlantic killifish PAHs adaptation could also be evaluated

further.

CHAPTER 4. DETERMINING THE EPIGENOMIC MECHANISMS DRIVING THE TRANSCRIPTOMIC CHANGES UNDERLYING ATLANTIC KILLIFISH RAPID ADAPTIONS TO PAHS POLLUTION IN THE ELIZABETH RIVER, VIRGINIA

4.1 Abstract

Understanding the mechanisms underlying Atlantic killifish adaptative success to extreme organic pollution in several sites along the northeast coast of the United States has long been a great scientific interest, and the conducted experiments have found that evolutionary mechanisms (natural selection, mutations, and gene flow) are involved in mainly underlying the evolved pollution adaptions. Though a deep understanding of molecular mechanisms underlying the evolved pollution resistance critical, most of these remain poorly identified. The potential of epigenetics to regulate the killifish pollution tolerance is considered, but experimental attempts to test the epigenetic-pollution tolerance interconnection are greatly lacking. Several target gene methylation profiling analyses found weak methylation regulation on killifish pollution resistance, but detailed genome-wide-methylation profiling studies on the killifish context are not available yet.

To overcome this knowledge gap, in this study we compared the global methylome of two Atlantic killifish populations (pollution-tolerant and pollution-sensitive) to determine the differential methylation status underlying the evolved pollution tolerance. However, we limited the study to estimate the reduced differential methylation by actively selecting the methylation in the CpG islands. To this end, we estimated the genome-wide methylation at the nucleotide, regional (300 bp regions-differentially methylated regions-DMRs), and gene level (differentially methylated genes-DMGs). Our study found 767 differentially methylated cytosines in 488 DMRs overlapped in 308 gene regions (DMGs). The DMRs were distributed among 29 exons and 252

intron regions and two key AHR pathway genes, *ahr1*, and *ahr2* were found among the DMGs where there was intron hypermethylation. We found three intron-hypermethylated genes (*eml1*, *phactr1*, and *WDR47*) were also downregulated in the RP killifish population, which we consider a potential methylation-induced regulatory mechanism underlying the killifish rapid adaptablity to PAHs contamination, which showed a similar expression-methylation pattern in human exposure to airborne PAHs contamination. In this study we identified several epigenomic mechanisms potentially important for developing pollution tolerance in Atlantic killifish, but further controlled experiments and genome-wide methylome profiling are encouraged for the next level.

4.2 Introduction

The climatic and non-climatic environment has been changing rapidly in the Anthropocene, generating adverse impacts on the ecosystem, and on the human community. Organisms exhibit differential responses to environmental changes, several species adapt rapidly, while some become more susceptible to changes. The ocean environment has also been changing rapidly over the past couple of decades, thus the inhabitants are needed to develop protective mechanisms more effectively to face the rapid changes.

Organisms' environmental tolerance heavily relies on their adaptative capacity based on the evolutionary history that makes changes in the individual genomes, like gene drift, natural selection, and mutations (Reid et al., 2016; Whitehead et al., 2017). However, it has been found that, despite evolutionary mechanisms, several other mechanisms like phenotypic plasticity (PP) help greatly to increase environmental tolerance (Donelson et al., 2019; Mundey et al., 2013). It has been observed marine species' phenotypic plasticity (PP) plays a large role in the rapid changes in environmental cues, but triggered another question, what are the mechanisms driving

PP in marine organisms? is it genomic (like point mutations, deletions, insertions) or other factors? Several mechanisms have been considered, and among them, epigenetic modifications have long been considered advantageous mechanisms underlying the PP changes in many marine organisms encountering climatic and non-climatic environmental changes (Ryu et al., 2018). Several studies have found the impact of rising ocean temperatures on changing marine ectotherms' global methylation profile (Anastadia et al., 2017; Ryu et al., 2018), resulting in favorable resistive phenotypes. There are several epigenomic mechanisms identified, but histone modifications and DNA methylation are considered more common and imperative in regulating gene expression (Akalin et al., 2012). The core concept of this chapter is to study the DNA-methylation drivers regulating the gene expression underlying Atlantic killifish rapid adaptations to PAHs pollution in the Elizabeth River, Virginia, so the remaining sections of this chapter will mostly be confined to discussing about the DNA methylation.

4.2.1 DNA methylation

Awareness of DNA methylation has a long history, as quite old as the discovery of DNA as the genetic material (Avery et al., 1944; McCarty and Avery, 1946). Mammalian DNA methylation was the widespread research interest at the beginning (Li and Beard,1999; Smith et al., 2012) but later expanded over to other species too. It is considered that the DNA profile of a malignant cell is either hyper (increased) or hypo (decreased) methylated compared to the normal cellular DNA profile, and often hyper-methylation is associated with gene upregulation and vice versa (Akalin et al., 2012). DNA methylation is one of the key epigenomic mechanisms in organisms and the most prominent regulatory epigenomic mechanism of gene expression (Akalin et al., 2012). The cytosine methylation in cytosine-guanine (C-G) rich regions (CpG islands) that are mostly overlying within the gene promoter regions is more frequent compared to methylation in other

gene regions (Moore et al., 2013). Gene promoter hypermethylation can effectively block the

binding of transcription factors, leading to reduce transcription that leads to a reduced the gene

expression (Bird and Wolffe, 1999; Hendrich and Bird, 1998). DNA methylation could also be

genetic imprinting, which is the inheritance of DNA methylation patterns throughout the

generations, and also associates with the developmental plasticity of organisms' early life stages

(Li and Beard,1999; Smith et al., 2012).

DNA methylation is an energetic process that requires an enzymatic process that involves a

family of DNA methyltransferase (Dnmts) to transfer a methyl group from S-adenyl methionine

(SAM) to the fifth carbon of the cytosine, which is then considered 5-methyl-Cytosine (5mC) (Li

and Beard,1999; Morre et al., 2013; Smith et al. 2012). Dnmt3a and Dnmt3b enzymes are

involved in the novel methylation in unmodified cytocines and Dnmt1 functions on copying the

information during the replication process (Moore et al., 2013) (Fig. 4.1).
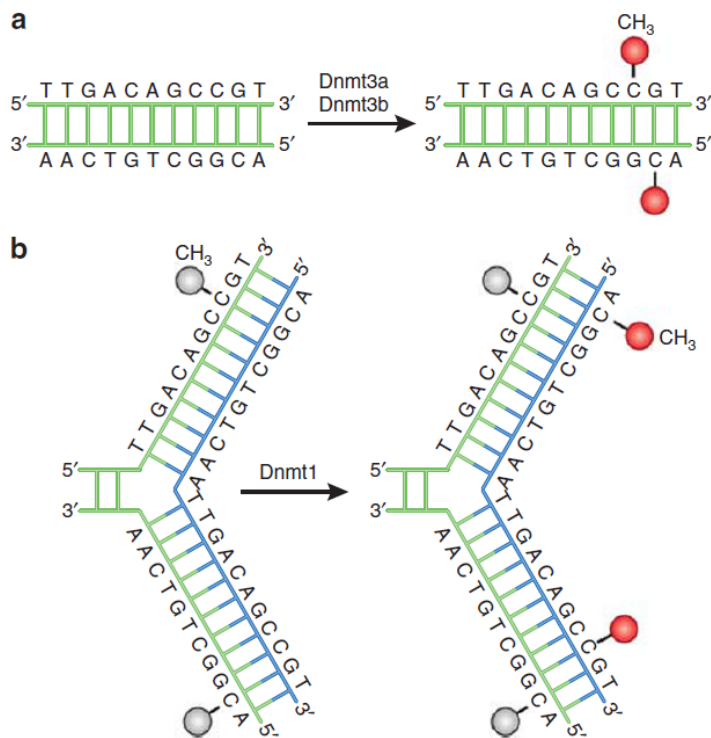
Fig. 4.1. DNA methylation process and the enzymes involved in each step

DNA methylation is highly tissue-specific (i.e. more methylation in the mammalian brain than other tissues) and genome location-specific (Moore et al., 2013). Intergenic regions (regions between genes), CpG sites (nearly 1000 base pair long sequences with higher C-G sites compared to the other regions), and gene body can be methylated due to external cues like temperature fluctuations (Anastadia et al., 2017; Moore et al., 2013) or diseases like cancers (Li and Beard,1999; Moore et al., 2013; Smith et al., 2012)or with aging (Anastadia et al., 2018). Promoter regions are generally associated with the CpG islands, and often are evolutionarily conserved to remain unmethylated (Aluru et al., 2011; Gardiner-Garden and Frommer, 1987; Moore et al., 2013), but any unfavorable condition could methylate the CpG regions, decreasing or inactive the gene transcription (Aluru et al., 2011; Moore et al., 2013), thus the promoter methylation has commonly be targeted in a large number of experiments.

Cellular DNA-methylation measuring techniques have long been developed and resulted in a process that combined wet lab, high throughput sequencing, high-performance computing, and statistical processes. The most prominent methodological approach to determining DNA methylation in the current context is bisulfite sequencing (Akalin et al. 2012; Meissner et al., 2008). In this method, extracted DNA is treated with sodium bisulfite, which transforms unmethylated cytocines to uracil, but leaves the methylated cytosine residue unaffected (Akalin, et al., 2012). The novel and more advanced highthouput sequencing that is capable of scanning the genome at single-nucleotide level measures the genome-wide nucleotide count and percentage methylation then calculate by counting the ratio of Cytocine/(Cytocine+Thiamin) at each base (Akalin, et al., 2012). Multiple bisulfite sequencing methods have been developed depending on the computational and sequencing facility investment, like reduced rapid bisulfite sequencing (Meissner et al., 2008) that selectively determines the cytosine methylation in the CpG sites, whole genome shotgun bisulfite sequencing (BS-Seq) (Cokus et al., 2008), methyl-Seq (Lister et al., 2008), and target capture bisulfite sequencing (Ball et al.2009), and the modified methods of bisulfite sequencing have also been used to detect the DNA methylation (Anastadia et al., 2017) and 5-hydroxymethyl cytosine (Booth et al., 2012).

The potential DNA-methylation mechanisms underlying the Atlantic killifish evolved resistance to extreme levels of organic pollutants have been considered for a long period of time, but the experimental attention to determine the mechanisms is still poor. Aluru et al. attempted to test the hypothesis that hypermethylation in *ahr1* and *ahr2* promoters are associated with the recalcitrance of the AHR pathway underlying the Atlantic killifish evolved resistance to PCBs in New Bedford Harbor in Boston, Macceshushets. The study found some tissue-level differences in promoter hypermethylation and gene expression, but a clear link between pollution adaptation,

gene expression, and methylation wasn't found. Despite several attempts to determine the link between target gene regulation and DNA methylation, experiments assessing the global DNA methylation profile changes underlying the PAHs pollution resistance in Atlantic killifish are not available yet. To fill this gap, to our best knowledge for the first time, in this study we conduct a genome-wide scanning of the methylation profile in the Atlantic killifish evolve rapid adaptations to PAHs pollution in the ER, Virgina and compare it with the methylation profiles of a killifish population in a nearby reference site (Kings Creek) (Fig. 3.3). The key focus of this study is to reveal potential DNA methylation mechanisms driving the transcriptome changes in Atlantic killifish to develop rapid adaptations to extreme PAHs toxicity in the RP site.

4.3 Materials and Methods

4.3.1 Sample preparation

The liver of each killifish was extracted and preserved in ice until the sample preparation. In this study, we used the reduced rapid bisulfite sequencing (RRBS) method, which is a cost-effective DNA-methylation detection method relative to the whole genome bisulfite sequencing (WGBS). In the RRBS method, a digestive enzyme (MspI) is utilized to selectively cut the genome at MspI cutting sites (C^CGG) in CpG islands (Boyle et al., 2012), thus the resulting in reads with restricted sites for downstream analysis are several-fold lower than the read counts of the WGBS analysis.

For our study, we used 25 mg of liver sample from each fish and extracted the DNA following the manufacturer (DNeasy Blood and Tissue kit-QIAGEN) sample preparation protocol described in Lea et al., 2015. Next, we took 10-30 ng of extracted DNA in each sample and

prepared barcoded libraries. Then the batch of samples was treated with sodium bisulfite for bisulfite conversion using the EpiTect Bisulfite Conversion kit (QIAGEN) and the bisulfite-treated reads were amplified using PCR prior to sequencing on the Illumina HiSeq 2000 platform. Sequence data were subjected to a series of bioinformatics processes to get the differentially methylated cytocines, regions, and genes in the genome of the pollution-tolerant killifish compared to the pollution-sensitive population.

4.3.2 Bioinformatics analysis

Ninety-one and eighty-nine million paired-end sequenced reads (2×126 bp) of bisulfite-treated liver DNA data from KC and RP fish respectively were used for the bioinformatics analysis. me. We first trimmed out the law quality (Q<20) reads and adapters (ligated during the DNA digestion process with MspI enzyme) using '--rrbs' option in Trim galore v0.6.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

Next, filtered reads were mapped to the killifish reference genome version 3.0.2 using Bismark v0.22.3 (Kruger et al., 2011) to extract methylation information. The reference genome was converted into bisulfite-treated status (C to T and G to A transformed) by the 'bismark_genome_preparation' scriptlet. Default '--directional' mapping was used to map trimmed reads to the bisulfite-treated genome using the 'bismark' scriptlet. Bismark calculates the methylation status per a specific cytosine site following this formula; averaged methylation level= (counts of methylated reads/counts of total reads) $\times$ 100. Generated BAM files in the mapping process were sorted using Samtools v1.15.1(Li et al., 2009). A methylation bias (M-bias) was observed at the first base pair of the forward read and the last base pair of the reverse read in each mapped read. The bias was removed by applying the '--ignore_r2' and '--ignore_3prime_r2' options simultaneously with the per cytosine methylation status extraction

process using 'bismark_methylation_extractor'. Steps from trimming to mapping were completed in the University of Maine cluster computer facility. The coverage files (.cov.gz) comprised of extracted methylation information were statistically tested to determine the differential methylation status of the methylated cytosines using a Bioconductor package, methylKit 1.22.0 (Akalin et al., 2012).

4.3.3 Statistical analysis for the differentially methylated cytosines and regions.

Cytocines with poor read coverages (> 10) were removed before the downstream statistical analysis. The differentially expressed cytosines in the RP fish (12 fish samples) compared to the KC fish (12 fish samples) were estimated by fitting a logistic regression (Akalin et al., 2012) in the methylKit package. The calculated p-values were corrected for q-values using the Sliding Linear Multi Modle (SLMM) method (Wang et al., 2011) and the cytosines with q-value <0.05 and absolute methylation difference > 25% were considered differentially methylated.

In the next step, we pooled the individual read data into 300 bp tilling windows to estimate the differential methylation between the regions (Differentially Methylated Regions, DMRs). The optimum length of the region was tested numerically by testing a series of different tilling window sizes (100,200,300,500, and 1000 bp). We continued the DMR analysis using the 300 bp tilling window size since it resulted in the highest DMR count. Similar to the DMC estimation, logistic regression was applied to estimate the differential methylation between the 300 bp segments. Segments with a q-value less than 0.05 (FDR<0.05) and a methylation difference higher than 25% were considered differentially methylated regions. For the last step, we employed a custom R code (KillifishDMRAnnotations.R available at the GitHub repository) to annotate the estimated DMRs into different Atlantic killifish gene regions listed in the Ensemble Atlantic killifish gene transfer file (.gtf). We matched the DMRs into exons, introns, 5' and 3'

UTR, intergenic, and gene promoters (2000 bp upstream and 200 bp downstream of the gene Transcription Start Site) using the custom code. Genes comprised of at least one annotated DMR within the limits of promoter and gene body regions were considered differentially methylated genes (DMG) and the DMRs that weren't matched into gene regions were considered 'intergenic'.

4.3.4 Transcriptomic and epigenomic integrated analysis

To integrate the transcriptional and epigenetic information for this study, we detected the overlapping genes between differentially expressed genes and differentially methylated genes. The functions and the coexpression of the overlapping genes were tested using StringTie protein network and running a WGCNA.

4.4 Results

4.4.1 Differential methylation patterns between KC and RP fish

To determine the potential epigenetic mechanisms underpinning the detected transcriptome changes, we analyzed the reduced rapid bisulfite sequenced (RRBS) DNA sequence data generated from the same fish liver samples used for transcriptomic analysis. Ninety-one and seventy-seven million paired-end MSPI enzyme digested DNA sequence reads resulted in the respective KC and RP fish samples (Table 4.1 and 4.2). After trimming off the low-quality (Q<20) reads and adapters in the subsequent step, 89 and 57 million clean reads were resulted in and mapped to the Atlantic killifish reference genome (V. 3.0.2). At the mapping step, 45 and 38 million reads were uniquely mapped to the reference genome with average mapping efficiencies of 50% and 48% in the respective KC and RP populations (Table 4.1 and 4.2). The methylation of cytosine analysis found a total of 1543 and 1202 million methylated cytosines, and the

95

majority (185 and 151 million) of the methylated cytosines in KC and RP fish were densely

packed in the CpG sites which were average of 65% (Table 4.1 and 4.2), but the methylated

cytosine in CHG and CHH sites (Any of A, T, or C could be in the 'H' of the CHG and CHH)

were considerably low (average of 0.3% and 0.5% respectively), in the RP and KC fish samples.

Therefore, we limited the differential methylation analysis only to CpG sites where the strongest

methylation signal was observed.

The employed logistic regression analysis in the meythylKit package in R statistical

environments statistically compared the methylation status in KC fish with RP fish and detected

767 differentially methylated cytosines (Table 4.3), while 425 were hypermethylated

(methylation difference >25%) and 342 were hypomethylated (Table 4.3). Estimation of the

differential methylation between Atlantic killifish genomic regions was the next step of this

analysis, thus we determined the optimum tile size empirically, testing five tile sizes (100,200,

300, 500, and 1000 bp) to detect the tile size that generates the highest possible number of

differentially methylated regions (DMRs). The empirical test detected 300 bp as the tile range

capable of segregating the killifish genome to produce the highest amount of DMRs, thus we

continued our analysis using 300 bp tiles that detected 480 DMRs (FDR<0.05) with 214

hypermethylated and 274 hypomethylated (Table 4.3) regions. To find the overlapping gene

regions of the detected DMRs, we used a custom R code ('killifishAnnoatation.R') and

annotated the DMRs into the killifish genes that resulted in 308 DMGs (Table 4.3) with 29 exon

DMRs, 18 promoter DMRs, 1 in 5' UTR, 8 in 3' UTR, and 252 in introns (Table 4.3). The R

code parameters were set to annotate a DMR into a given genomic region, only if the considered

DMR fully overlapped within the gene regions (partial overlapping were not considered

Table 4.1. Summary of statistics of the methylation status of each fish from the reference (KC) site

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Sequence depth (million reads)	3.8	5.4	14.31	10.71	8.34	8.17	4.73	2.77	6.06	7.62	10.29	**9**
Reads after trimming	3.65	5.22	14.05	10.56	8.15	8.02	4.56	2.7	5.85	7.42	10.17	8.83
Uniquely mapped reads	1.79	2.43	7.11	5.35	4.13	4.05	2.25	1.37	2.91	3.68	5.14	4.53
Mapping efficiency %	48.89	46.48	50.58	50.65	50.67	50.49	49.31	50.7	49.7	49.53	50.59	51.34
Total Cs analyzed (million)	61.48	85.58	239.64	190.11	142.48	143.43	80.85	47.77	98.55	125.67	173.93	153.51
Methylated Cs in CpG sites (million)	7.52	11.13	16.92	24.42	18.53	18.45	10.29	6.06	12.37	16.56	22.7	19.88
% methylation in CpG sites	63.5	65.2	64.9	65.6	65.5	65.3	64.8	64.4	63.9	65.5	65.3	65.3
Methylated Cs in CHG sites (million)	0.08	0.12	0.34	0.26	0.2	0.2	0.11	0.07	0.13	0.18	0.25	0.21
% methylation in CHG sites	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 4.1 continued

Methylated Cs in CHH sites (million)	0.11	0.15	0.42	0.34	0.25	0.26	0.14	0.08	0.17	0.22	0.31	0.27
% methylation in CHH sites	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Table 4.2. Summary of statistics of the methylation status of each fish from the pollute (RP) site

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Sequence depth (million reads)	5.3	7.13	10.01	4.24	3.21	8.55	5.7	4.03	8.32	13.77	6.96
Reads after trimming	5.14	5.85	5.89	4.09	3.13	5.86	5.51	3.87	5.84	5.88	5.88
Uniquely mapped reads	2.38	3.5	4.89	2.01	1.52	4.3	2.68	1.90	4.1	**7**	3.38
Mapping efficiency %	47	50.8	50.2	49.8	49.1	51.7	49.2	50.1	51.2	52.5	50.1
Total Cs analyzed (million)	83.65	119.11	173.75	71.17	52.99	2.12	152.03	94.01	67.17	143.06	242.52
Methylated Cs in CpG sites (million)	10.8	14.38	22.19	8.48	6.69	0.27	19.18	11.83	8.05	18.08	30.74
% methylation in CpG sites	66.7	65.6	65.8	63.7	66.4	63.8	65.7	66.3	63.5	65.3	66.1
Methylated Cs in CHG sites (million)	22.17	0.15	0.24	0.09	0.07	0.003	0.21	0.13	0.09	0.19	0.33
% methylation in CHG sites	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 4.2 continued

Methylated Cs in CHH sites (million)	45.02	0.22	0.32	0.13	0.1	0.004	0.28	0.17	0.12	0.26	0.44
% methylation in CHH sites	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Table 4.3 Summary of the differential methylation detected in RP fish compared to the KC population

Feature	Count
DMCs in CpG sites	767
Hypermethylated DMCs	425
Hypomethylated DMCs	342
DMRs (300 bp)	488
Hypermethylated DMRs	214
Hypomethylated DMRs	274
Annotated DMRs to DMGs	308
Exon	29
Intron	252
5'UTR	1
3'UTR	8
Promoter	18
Intergenic	170

We then checked for the methylation status of any significant gene defined as key regulatory genes underlying Atlantic killifish pollution resistance. Interestingly, we found that the two key genes in the AHR pathway, *ahr1*, and *ahr2b* genes had hypermethylated regions in the gene intronic regions (Fig. 4.4), but none of the other AHR pathway genes didn't show a signal of differential methylation. In addition, several potassium ion channel regulatory genes like potassium voltage-gated channel modifier subfamily G member 1 (*kcng1*), potassium voltage-

gated channel subfamily H member 6 (*kcnh6*), potassium voltage-gated channel subfamily Q member 3 (*kcnq3*), and potassium voltage-gated channel, Shab-related subfamily, member 3 (*kcnb3*) were differentially methylated in the RP fish compared to the KC population. Interestingly, only two genes (kinesin family member 17 (*kif17*) and a novel unannotated killifish gene ENSFHEG00000001998) were enriched in the single detected GO term (cellular component) in the enrichment analysis, ciliary basal body (GO:0036064).
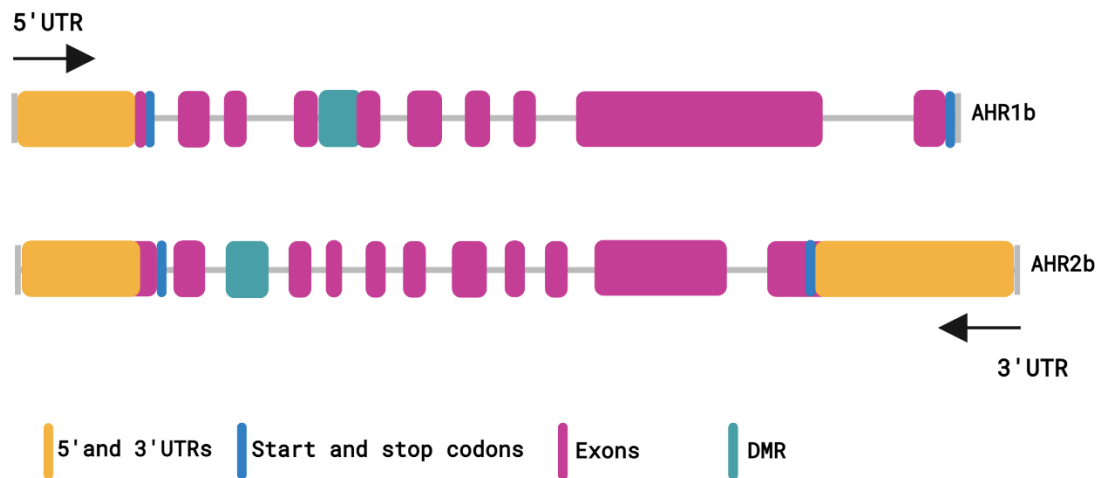
Fig. 4.2. The schematic representation of intron hypermethylation in *ahr1b* and *ahr2b* genes detected in RP killifish

4.4.2 Transcriptomic, and methylomic integrated analysis

To achieve the key aim of this study, determining the molecular processes underlying the Atlantic killifish's rapid adaptations to extreme PAH contamination and their underpinning epigenetic mechanisms, we integrated the gene expression data with methylation information to determine the shared pathways indicating the methylation drivers of the gene expression underlying killifish pollution resistance.

The integration of DMGs and gene expression variance data found no clear link between the two datasets, which indicates a weak methylation regulation on the gene expression variance underlying killifish PAHs resistance. However, the DEGs and DMGs found several key DNA methylation processes with the potential to regulate the gene expression underlying the killifish PAHs adaptations. We found 13 genes shared in both groups (Fig. 4.3), and all the genes comprised intron methylations. Ten hypermethylated genes were downregulated, three were downregulated but hypomethylated, and one was upregulated with intron hypomethylation (Fig. 4.4). Interestingly, the *ahr2* gene, which is considered a key regulatory gene in the antagonism of the AHR pathway underlying Atlantic killifish pollution adaptations was downregulated and intron hypermethylated in this study. (Fig. 4.2). We checked any potential molecular, cellular, or biological processes of the shared genes between DEGs and DMGs, but the conducted GO term analysis and protein-protein network analysis found no specific process.
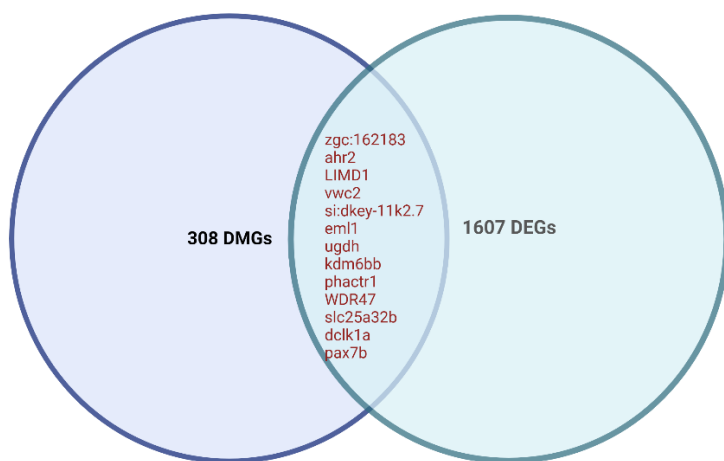
Fig. 4.3. Differentially expressed and methylated overlapping genes
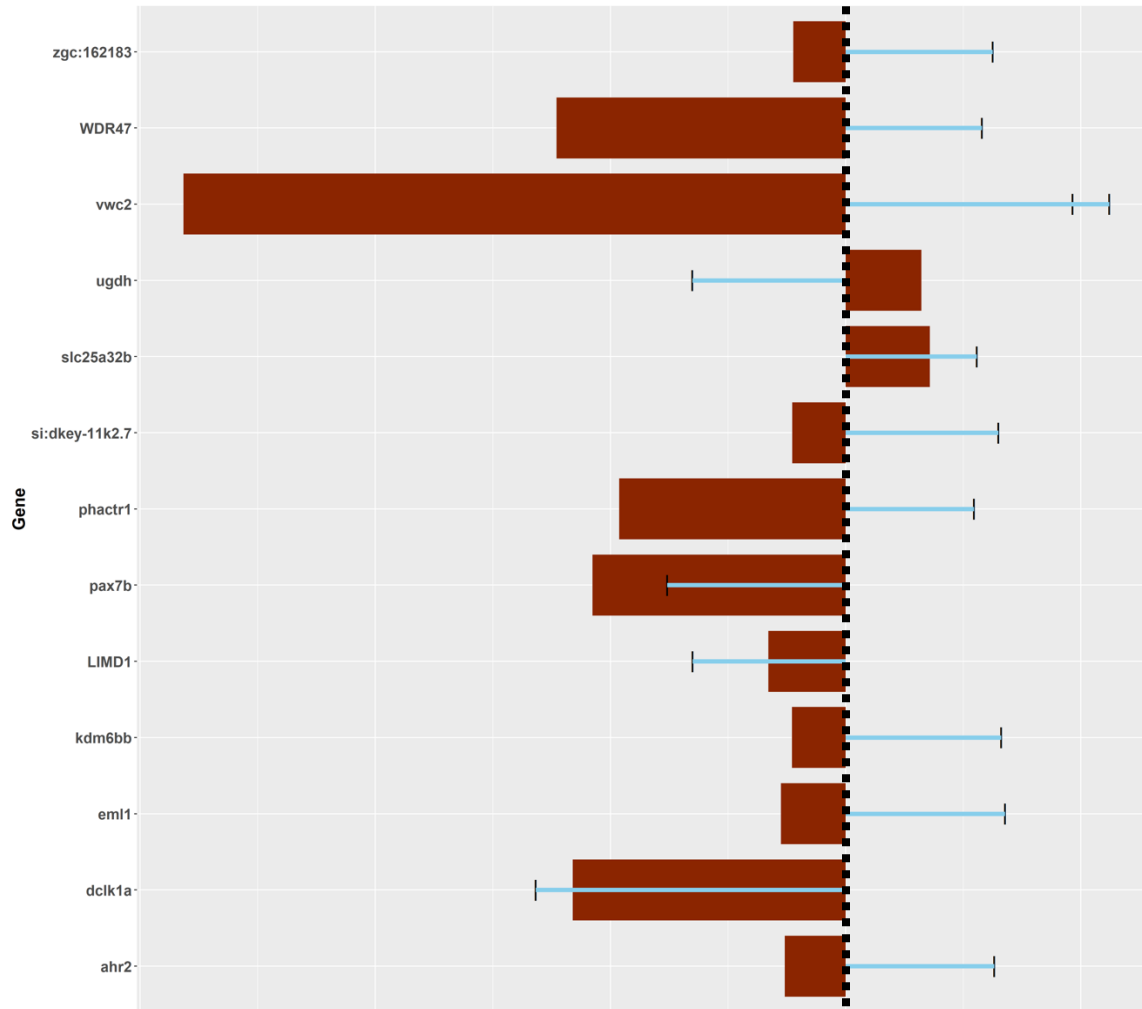
Fig. 4.4. The schematic representation of the differential expression and methylation of the
overlapping genes (red bars represent the gene expression and the blue lines represent the
methylation status). The dark dashed line is the zero line. Left to the dark line are

the downregulation and hypomethylation and vice versa.

4.5 Discussion

The integration of differentially expressed and differentially methylated genes found 13 shared

genes (*zgc:162183*, *ahr2*, *limd1*, *vwc2*, *si:dkey-11k2.7*, *eml1*, *ugdh*, *kdm6bb*, *phactr1*, *WDR47*,

slc25a32b, *dclk1a*, and *pax7b*) between two gene lists (*Fig. 6C*). Although we found 308 genes

with differential patterns of methylation across the gene regions, all shared genes consisted of intron methylation, which we found interesting in this study. The downregulated *ahr2* gene in the Republic fish showed intron hypermethylation, thus we see a potential epigenetic process working underlying the evolved killifish PAHs tolerance. *ahr2* is considered the key gene regulating the AHR de-sensitivity facing the AHR agonists, thus the downregulation of *ahr2* desensitized the AHR pathway. So, as we observed in this study, the intron hypermethylation and downregulation of the *ahr2* gene can be considered a potential mechanism underlying the killifish-evolved PAHs resistance. In general, it is expected that the promoter hypermethylation role in regulation the of gene expression (Aluru et al., 2011), but the interest in determining the intron methylation-driven gene expression changes has also been rising over the past years (as Nam et al. 2021 described the potential linkage between intron methylation of mammalian gene regulation responsible for cancers). However, to our knowledge, no data available on identified intron methylation-driven *ahr2* downregulation, thus this study could be considered the first such mechanism. In addition to *ahr2*, six other genes overlapped between the two groups ((*limd1*, *vwc2*, *eml1*, *ugdh*, *phactr1*, *wdr47*) had been identified as potential targets of intron methylation-driven variable gene expression with PAHs exposure. Human *eml1*, *phactr1*, and *wdr47* genes showed methylation variability that governs the respective gene expression under atmospheric PAH exposure (Jiang et al., 2017; Tryndyak et al., 2018) and *limd1*, *vwc2*, *ugdh*, and *wdr47* had 3 and 5'UTR, exon, and promoter methylation-drive altered gene expression in the human exposed to aerial PAHs. Our findings are overlapped with this existing theory, so we recommend further target gene experiments to dig deep into these mechanisms.

In this study, we integrated three different types of global genomic information of wild Atlantic killifish from the pollution-tolerant and pollution-sensitive populations and found clues that

confirm the existing theory and also some novel mechanisms underlying the evolved PAHs tolerance. This study emphasizes the importance of integrating different omics data for understanding the mechanisms underpinning the evolved pollution tolerance, and the importance of studying the wild fish population is also highlighted.

CHAPTER 5: CONCLUSION

The overarching goal of this dissertation is to predict one fish species' responses to two critical environmental crises, ocean warming, and pollution. Ocean has been warming and getting polluted for a long time, but their rates have abruptly increased during the last couple of decades, mainly due to intense human activities. The marine ecosystem and its inhabitants have also been threatened by the synergistic effects of ocean warming and pollution. Determining how marine organisms respond to ocean warming and pollution is critical for conservation and ecosystem health assessments. It would also be important in assessing the potential threats to human health conflicts embedded with ocean warming and pollution. To this end, in this dissertation, I focused on determining how marine teleost responds to the rapid rates of ocean warming and pollution and the future consequences.

In the second chapter of this dissertation, a computational approach was framed to predict the marine teleost future habitat changes of killifish with the concurrent ocean warming rates. The range shift for a wide diversity of species with rising ocean temperatures has been identified over the past couple of decades and computational approaches called species distribution models (SDMs) have been developed to predict the potential future range with the continuous ocean warming. However, the majority of SDMs have been developed to match the contemporary environmental range of a given species with the future predicted environmental conditions to predict its future geographical distribution. In other words, the current SDMs do not consider organisms' capacity for tolerating the upcoming environmental changes, thus the results in habitat predictions are considered oversimplified and overestimated. To overcome this limitation, in the first chapter of the dissertation, I developed a novel computational program that can integrate marine teleost thermal tolerance capacity when predicting their future geographical

distribution with ocean warming. I tested the method to predict the geographical distribution of a highly thermal tolerant intertidal minnow-Atlantic killifish (*Fundulus heteroclitus*) in the 2050s and 2080s, as a best-case scenario. Our model predicted Atlantic killifish future (the 2080s, if the concurrent ocean warming rate continues until the 2080s) habitat distributions showed diverse range shift patterns, including the mixing of killifish subpopulations between each other, isolated populations in the southern end of the killifish distribution, a southward movement of northernmost killifish subpopulation, but no northward movement. Given the predictions of the modeling study, I emphasize that the species' thermal tolerance capacity should be integrated into SDMs when predicting their future geographical distribution underlying the rising ocean warming scenario. However, a key limitation of our model is that it doesn't consider the ecological interactions of the target species when predicting its future habitat ranges. Thus, I encourage further experiments to develop models that integrate marine teleost environmental tolerance, and ecological interactions to predict their future habitat ranges with changing environments, more accurately.

For the third and fourth chapters of this dissertation, I focused on estimating teleost responses to ocean pollution. Several Atlantic killifish subpopulations have evolved rapid adaptations to extreme levels of environmental toxicity in several sites along the northeast coast of the United States. The inhabiting killifish were tested in various experiments to uncover the molecular and evolutionary mechanism underlying killifish's rapid adaptability to toxicants, which resulted in a piece of extensive knowledge underlying killifish (common among vertebrates) pollution tolerance. However, there could be more mechanisms that remained undetected, thus, I integrated two high-resolution molecular data sets to detect potential novel molecular mechanisms underlying killifish pollution tolerance. The comparison of total gene expressions

between pollution-tolerant and pollution-sensitive Atlantic killifish populations found some genes that can serve as biomarkers of pollution adaptation in Atlantic killifish. Besides, potential above the genome (epigenome) processes regulating the expression of several key genes of Atlantic killifish pollution tolerance were also detected in this study. This integrated, in-silico molecular study confirms that marine teleost are continuously evolving for environmental changes, so novel approaches should be employed to understand these adaptative mechanisms. It is critical to evaluate their survivability in the rapidly changing environment and to adopt conservation measures to protect them from destruction.

In the second chapter, I highlighted the importance of integrating marine teleost's thermal tolerance capacity in SDMs when predicting their future habitat range shifts with ocean warming. Besides, in the third and fourth chapters, I showed that marine teleost are evolving continuously to ocean pollution, thus integrated molecular methods should be employed to uncover novel adaptations. It is conclusive that, marine teleost's responses to multi-stressors (i.e., warming, pollution, acidification) should be considered when predicting their future habitat ranges. Given that, at the end of this dissertation, I would like to emphasize the importance of experiments to estimate the multi-stressor tolerance capacity of marine teleost, and develop *in silico* methods integrating tolerance capacity to predict the future geographical distribution with predicted environmental conditions. I believe the method developed in the second chapter would serve as a fundamental framework for, how to integrate marine teleost multi-environmental stressor tolerance into SDMs, i.e., the molecular mechanisms underlying the Atlantic killifish evolved pollution adaptations, detected in chapters three and four can be examined further to determine the results in phenotypic changes, and incorporated into SDMs to predict Atlantic killifish future geographical range shift with ocean warming and pollution.

BIBLIOGRAPHY

Abbate, M. C. L., Molinero, J. C., Guinder, V. A., Dutto, M. S., de Cao, M. S. B., Etcheverry, L. A. R., Pettigrosso, R. E., Carcedo, M. C., & Hoffmeyer, M. S. (2015). Microplankton dynamics under heavy anthropogenic pressure. The case of the Bahía Blanca Estuary, southwestern Atlantic Ocean. *Marine Pollution Bulletin*, *95*(1), 305–314.

Aboim, M. A., Mavárez, J., Bernatchez, L., & Coelho, M. M. (2010). Introgressive hybridization between two Iberian endemic cyprinid fish: A comparison between two independent hybrid zones. *Journal of Evolutionary Biology*, *23*(4), 817–828.

Achard-Joris, M., Gonzalez, P., Marie, V., Baudrimont, M., & Bourdineaud, J.-P. (2006). cDNA cloning and gene expression of ribosomal S9 protein gene in the mollusk Corbicula fluminea: A new potential biomarker of metal contamination up-regulated by cadmium and repressed by zinc. *Environmental Toxicology and Chemistry: An International Journal*, *25*(2), 527–533.

Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Fields, C., & Venter, J. C. (1992). Sequence identification of 2,375 human brain genes. *Nature*, *355*(6361), 632–634.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), 1–9.

Alava, J. J. (2019). Ocean pollution and warming oceans: Toward ocean solutions and natural marine bioremediation. In *Predicting Future Oceans* (pp. 495–518). Elsevier.

Alden, R. W. (1995). JGW. Defining the Problem: The Elizabeth River, A Region of Concern. *Regional Action Plan for the Elizabeth River. Rep Submitt to Virginia Dep Environ Qual*.

Aluru, N., Karchner, S. I., Franks, D. G., Nacci, D., Champlin, D., & Hahn, M. E. (2015). Targeted mutagenesis of aryl hydrocarbon receptor 2a and 2b genes in Atlantic killifish (Fundulus heteroclitus). *Aquatic Toxicology*, *158*, 192–201.

Aluru, N., Karchner, S. I., & Hahn, M. E. (2011). Role of DNA methylation of AHR1 and AHR2 promoters in differential sensitivity to PCBs in Atlantic Killifish, Fundulus heteroclitus. *Aquatic Toxicology*, *101*(1), 288–294.

Anastasiadi, D., Díaz, N., & Piferrer, F. (2017). Small ocean temperature increases elicit stage-dependent changes in DNA methylation and gene expression in a fish, the European sea bass. *Scientific Reports*, *7*(1), 12401.

Andersen, J. H., Harvey, T., Murray, C., Green, N., & Reker, J. (2019). Contaminants in Europe's Seas. Moving towards a Clean, Non Toxic Marine Environment. *European Environment Agency: Copenhagen, Denmark*.

Angilletta, M. J. (2009). *Thermal adaptation: A theoretical and empirical synthesis*.

Atkinson, D. (1994). Temperature and organism size-a biological law for ectotherms? *Adv. Ecol. Res.*, *25*, 1–58.

Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *Die Entdeckung Der Doppelhelix*, 97.

Bai, W., Zhang, C., & Chen, H. (2018). Transcriptomic analysis of Momordica charantia polysaccharide on streptozotocin-induced diabetic rats. *Gene*, *675*, 208–216.

Ball, M. P., Li, J. B., Gao, Y., Lee, J.-H., LeProust, E. M., Park, I.-H., Xie, B., Daley, G. Q., & Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, *27*(4), 361–368.

Baudron, A. R., Needle, C. L., Rijnsdorp, A. D., & Tara Marshall, C. (2014). Warming temperatures and smaller body sizes: Synchronous changes in growth of North Sea fishes. *Global Change Biology*, *20*(4), 1023–1031.

Bauer-Civiello, A., Critchell, K., Hoogenboom, M., & Hamann, M. (2019). Input of plastic debris in an urban tropical river system. *Marine Pollution Bulletin*, *144*, 235–242.

Beers, J. M., & Jayasundara, N. (2015). Antarctic notothenioid fish: What are the future consequences of 'losses' and 'gains' acquired during long-term evolution at cold and stable temperatures? *The Journal of Experimental Biology*, *218*(12), 1834–1845.

Bello, S. M., Franks, D. G., Stegeman, J. J., & Hahn, M. E. (2001). Acquired resistance to Ah receptor agonists in a population of Atlantic killifish (Fundulus heteroclitus) inhabiting a marine superfund site: In vivo and in vitro studies on the inducibility of xenobiotic metabolizing enzymes. *Toxicological Sciences*, *60*(1), 77–91.

Bijlsma, R., & Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Applications*, *5*(2), 117–129.

Bindoff, N. L., Cheung, W. W., Kairo, J. G., Arstegui, J., Guinder, V. A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M. S., & Levin, L. (2019). Changing ocean, marine ecosystems, and dependent communities. In *IPCC special report on the ocean and cryosphere in a changing climate*.

Bird, A. P., & Wolffe, A. P. (1999). Methylation-induced repression—Belts, braces, and chromatin. *Cell*, *99*(5), 451–454.

Blasius, M. E., & Goodmanlowe, G. D. (2008). Contaminants still high in top-level carnivores in the Southern California Bight: Levels of DDT and PCBs in resident and transient pinnipeds. *Marine Pollution Bulletin*, *56*(12), 1973–1982.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., & Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, *336*(6083), 934–937.

Borrelle, S. B., Ringma, J., Law, K. L., Monnahan, C. C., Lebreton, L., McGivern, A., Murphy, E., Jambeck, J., Leonard, G. H., & Hilleary, M. A. (2020). Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. *Science*, *369*(6510), 1515–1518.

Bowler, K., & Terblanche, J. S. (2008). Insect thermal tolerance: What is the role of ontogeny, ageing and senescence? *Biological Reviews*, *83*(3), 339–355.

Boyle, P., Clement, K., Gu, H., Smith, Z. D., Ziller, M., Fostel, J. L., Holmes, L., Meldrim, J., Kelley, F., & Gnirke, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biology*, *13*(10), 1–10.

Bozinovic, G., Shea, D., Feng, Z., Hinton, D., Sit, T., & Oleksiak, M. F. (2021). PAH-pollution effects on sensitive and resistant embryos: Integrating structure and function with gene expression. *PloS One*, *16*(4), e0249432.

Brauer, C. J., Unmack, P. J., & Beheregaray, L. B. (2017). Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened fish. *Molecular Ecology*, *26*(24), 6841–6856.

Brett, J. R. (1971). Energetic responses of salmon to temperature. A study of some thermal relations in the physiology and freshwater ecology of sockeye salmon (Oncorhynchus nerkd). *American Zoologist*, *11*(1), 99–113.

Brown, D. R., Clark, B. W., Garner, L. V. T., & Di Giulio, R. T. (2016). Embryonic cardiotoxicity of weak aryl hydrocarbon receptor agonists and CYP1A inhibitor fluoranthene in the Atlantic killifish (Fundulus heteroclitus). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *188*, 45–51.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). TOWARD A METABOLIC THEORY OF ECOLOGY. *Ecology*, *85*(7), Article 7. https://doi.org/10.1890/03-9000

Buckley, L. B. (2008). Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. *The American Naturalist*, *171*(1), Article 1.

Buckley, L. B., Waaser, S. A., MacLean, H. J., & Fox, R. (2011). Does including physiology improve species distribution model predictions of responses to recent climate change? *Ecology*, *92*(12), Article 12.

Burrows, M. T., Schoeman, D. S., Buckley, L. B., Moore, P., Poloczanska, E. S., Brander, K. M., Brown, C., Bruno, J. F., Duarte, C. M., & Halpern, B. S. (2011). The pace of shifting climate in marine and terrestrial ecosystems. *Science*, *334*(6056), 652–655.

Burrows, M. T., Schoeman, D. S., Richardson, A. J., Molinos, J. G., Hoffmann, A., Buckley, L. B., Moore, P. J., Brown, C. J., Bruno, J. F., & Duarte, C. M. (2014). Geographical limits to species-range shifts are suggested by climate velocity. *Nature*, *507*(7493), 492–495.

Carson, R. (2009). *Silent spring. 1962*. Getty Publications Los Angeles, CA, USA.

Castañeda, L. (2004). *Adaptive latitudinal shift in the thermal physiology of a terrestrial isopod*.

Chen, Y., Lun, A. T., & Smyth, G. K. (2014). Differential expression analysis of complex RNA-seq experiments using edgeR. *Statistical Analysis of next Generation Sequencing Data*, 51–74.

Chen, Z., Farrell, A. P., Matala, A., & Narum, S. R. (2018). Mechanisms of thermal adaptation and evolutionary potential of conspecific populations to changing environments. *Molecular Ecology*, *27*(3), 659–674.

Cheung, W. W., Lam, V. W., Sarmiento, J. L., Kearney, K., Watson, R., & Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, *10*(3), 235–251.

Chevin, L.-M., & Lande, R. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology*, *24*(7), 1462–1476.

Chust, G., Castellani, C., Licandro, P., Ibaibarriaga, L., Sagarminaga, Y., & Irigoien, X. (2014). Are Calanus spp. shifting poleward in the North Atlantic? A habitat modelling approach. *ICES Journal of Marine Science*, *71*(2), Article 2.

Clark, T. D., Sandblom, E., & Jutfelt, F. (2013). Response to Farrell and to Pörtner and Giomi. *Journal of Experimental Biology*, *216*(23), Article 23. https://doi.org/10.1242/jeb.096313

Clarke, A. (2004). Is there a universal temperature dependence of metabolism? *Functional Ecology*, *18*(2), Article 2.

Clarke, A., & Fraser, K. P. P. (2004). Why does metabolism scale with temperature? *Functional Ecology*, *18*(2), Article 2.

Clarke, A., & Johnston, N. M. (1999). Scaling of metabolic rate with body mass and temperature in teleost fish. *Journal of Animal Ecology*, *68*(5), Article 5.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., & Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, *452*(7184), 215–219.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., & Krinner, G. (2013). *Long-term climate change: Projections, commitments and irreversibility*.

Crone, T. J., & Tolstoy, M. (2010). Magnitude of the 2010 Gulf of Mexico oil leak. *Science*, *330*(6004), 634–634.

Crutzen, P. J. (2016). Geology of mankind. *Paul J. Crutzen: A Pioneer on Atmospheric Chemistry and Climate Change in the Anthropocene*, 211–215.

da Silva, C. R. B., Riginos, C., & Wilson, R. S. (2019). An intertidal fish shows thermal acclimation despite living in a rapidly fluctuating environment. *Journal of Comparative Physiology B*, *189*(3–4), 385–398.

Dahlke, F. T., Butzin, M., Nahrgang, J., Puvanendran, V., Mortensen, A., Pörtner, H.-O., & Storch, D. (2018). Northern cod species face spawning habitat losses if global warming exceeds 1.5°C. *Science Advances*, *4*(11), Article 11. https://doi.org/10.1126/sciadv.aas8821

Debruyn, A. M., Wernick, B. G., Stefura, C., McDonald, B. G., Rudolph, B.-L., Patterson, L., & Chapman, P. M. (2007). In situ experimental assessment of lake whitefish development following a freshwater oil spill. *Environmental Science & Technology*, *41*(20), 6983–6989.

del Mar Ortega-Villaizan, M., Noguchi, D., & Taniguchi, N. (2011). Minimization of genetic diversity loss of endangered fish species captive broodstocks by means of minimal kinship selective crossbreeding. *Aquaculture*, *318*(1–2), 239–243.

Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H.-O., & Huey, R. B. (2015). Climate change tightens a metabolic constraint on marine habitats. *Science*, *348*(6239), Article 6239.

Dhillon, R. S., & Schulte, P. M. (2011). Intraspecific variation in the thermal plasticity of mitochondria in killifish. *Journal of Experimental Biology*, *214*(21), 3639–3648.

Di Giulio, R. T., & Clark, B. W. (2015). The Elizabeth River story: A case study in evolutionary toxicology. *Journal of Toxicology and Environmental Health, Part B*, *18*(6), Article 6.

Dillon, M. E., Wang, G., & Huey, R. B. (2010). Global metabolic impacts of recent climate warming. *Nature*, *467*(7316), Article 7316. https://doi.org/10.1038/nature09407

Donelson, J. M., Sunday, J. M., Figueira, W. F., Gaitán-Espitia, J. D., Hobday, A. J., Johnson, C. R., Leis, J. M., Ling, S. D., Marshall, D., & Pandolfi, J. M. (2019). Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. *Philosophical Transactions of the Royal Society B*, *374*(1768), 20180186.

Dulvy, N. K., Rogers, S. I., Jennings, S., Stelzenmüller, V., Dye, S. R., & Skjoldal, H. R. (2008). Climate change and deepening of the North Sea fish assemblage: A biotic indicator of warming seas. *Journal of Applied Ecology*, *45*(4), 1029–1039.

Duvernell, D. D., Lindmeier, J. B., Faust, K. E., & Whitehead, A. (2008). Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, Fundulus heteroclitus. *Molecular Ecology*, *17*(5), Article 5.

Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792.

Fangue, N. A., Hofmeister, M., & Schulte, P. M. (2006). Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, Fundulus heteroclitus. *Journal of Experimental Biology*, *209*(15), Article 15.

Fangue, N. A., Podrabsky, J. E., Crawshaw, L. I., & Schulte, P. M. (2009). Countergradient variation in temperature preference in populations of killifish Fundulus heteroclitus. *Physiological and Biochemical Zoology*, *82*(6), 776–786.

Farrell, A. P. (2013). Aerobic scope and its optimum temperature: Clarifying their usefulness and limitations–correspondence on J. Exp. Biol. 216, 2771-2782. *Journal of Experimental Biology*, *216*(23), 4493–4494.

Franke, R. (1982). Scattered data interpolation: Tests of some methods. *Mathematics of Computation*, *38*(157), 181–200.

Gamliel, I., Buba, Y., Guy-Haim, T., Garval, T., Willette, D., Rilov, G., & Belmaker, J. (2020). Incorporating physiology into species distribution models moderates the projected impact of warming on selected Mediterranean marine species. *Ecography*, *43*(7), Article 7.

Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of Molecular Biology*, *196*(2), 261–282.

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of Size and Temperature on Metabolic Rate. *Science*, *293*(5538), Article 5538. https://doi.org/10.1126/science.1061967

Greaves, J. (1990). *Analysis of organic pollutants in sediments and blue crab (Callinectes sapidus) tissues: Final report to Virginia State Water Control Board*.

Grieve, B. D., Hare, J. A., & Saba, V. S. (2017). Projecting the effects of climate change on Calanus finmarchicus distribution within the U.S. Northeast Continental Shelf. *Scientific Reports*, *7*(1), 6264. https://doi.org/10.1038/s41598-017-06524-1

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), Article 9.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2–3), Article 2–3.

Gunderson, A. R., & Leal, M. (2016). A conceptual framework for understanding thermal constraints on ectotherm activity with implications for predicting responses to global change. *Ecology Letters*, *19*(2), 111–120.

Healy, T. M., & Schulte, P. M. (2012). Thermal acclimation is not necessary to maintain a wide thermal breadth of aerobic scope in the common killifish (Fundulus heteroclitus). *Physiological and Biochemical Zoology*, *85*(2), Article 2.

Hendrich, B., & Bird, A. (1998). Identification and characterization of a family of mammalian methyl CpG-binding proteins. *Genetics Research*, *72*(1), 59–72.

Hicken, C. E., Linbo, T. L., Baldwin, D. H., Willis, M. L., Myers, M. S., Holland, L., Larsen, M., Stekoll, M. S., Rice, S. D., & Collier, T. K. (2011). Sublethal exposure to crude oil during embryonic development alters cardiac morphology and reduces aerobic capacity in adult fish. *Proceedings of the National Academy of Sciences*, *108*(17), 7086–7090.

Hoffman, D. J., Rattner, B. A., Burton Jr, G. A., & Cairns Jr, J. (2002). *Handbook of ecotoxicology*. CRC press.

Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*, *8*(1), e1364.

Hvas, M., Folkedal, O., Imsland, A., & Oppedal, F. (2017). The effect of thermal acclimation on aerobic scope and critical swimming speed in Atlantic salmon, Salmo salar. *Journal of Experimental Biology*, *220*(15), 2757–2764.

Incardona, J. P., Vines, C. A., Anulacion, B. F., Baldwin, D. H., Day, H. L., French, B. L., Labenia, J. S., Linbo, T. L., Myers, M. S., & Olson, O. P. (2012). Unexpectedly high mortality in Pacific herring embryos exposed to the 2007 Cosco Busan oil spill in San Francisco Bay. *Proceedings of the National Academy of Sciences*, *109*(2), E51–E58.

IPCC. (2013). IPCC 2013: Summary for policymakers. In Climate Change 2013: The Physical Science Basis. In *Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–33). Cambridge University Press New York, NY, USA.

Jakobsen, T., & Ozhigin, V. K. (2011). *The Barents Sea: Ecosystem, resources, management-Half a century of Russian-Norwegian cooperation*.

Jayasundara, N. (2017). Ecological significance of mitochondrial toxicants. *Toxicology*, *391*, 64–74.

Jayasundara, N., Fernando, P. W., Osterberg, J. S., Cammen, K. M., Schultz, T. F., & Di Giulio, R. T. (2017). Cost of tolerance: Physiological consequences of evolved resistance to inhabit a polluted environment in teleost fish Fundulus heteroclitus. *Environmental Science & Technology*, *51*(15), 8763–8772.

Jayasundara, N., Kozal, J. S., Arnold, M. C., Chan, S. S. L., & Giulio, R. T. D. (2015). High-Throughput Tissue Bioenergetics Analysis Reveals Identical Metabolic Allometric Scaling for Teleost Hearts and Whole Organisms. *PLOS ONE*, *10*(9), Article 9. https://doi.org/10.1371/journal.pone.0137710

Kaschner, K., Rius-Barile, J., Kesner-Reyes, K., Garilao, C., Kullander, S. O., Rees, T., & Froese, R. (2019). *AquaMaps: Predicted Range Maps for Aquatic Species. Worldwide web electronic publication*.

Kaschner, K., Watson, R., Trites, A. W., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, *316*, 285–310.

Kearney, M. R., Wintle, B. A., & Porter, W. P. (2010). Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, *3*(3), Article 3.

Killen, S. S., Atkinson, D., & Glazier, D. S. (2010). The intraspecific scaling of metabolic rate with body mass in fishes depends on lifestyle and temperature. *Ecology Letters*, *13*(2), Article 2. https://doi.org/10.1111/j.1461-0248.2009.01415.x

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915.

Kinnison, M. T., Unwin, M. J., & Quinn, T. P. (2003). Migratory costs and contemporary evolution of reproductive allocation in male chinook salmon. *Journal of Evolutionary Biology*, *16*(6), Article 6.

Kleiber, M. (1932). Body size and metabolism. *Hilgardia*, *6*(11), Article 11.

Kolbe, J. J., Kearney, M., & Shine, R. (2010). Modeling the consequences of thermal trait variation for the cane toad invasion of Australia. *Ecological Applications*, *20*(8), Article 8.

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, *20*(1), 1–13.

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572.

Kunz, K. L., Frickenhaus, S., Hardenberg, S., Johansen, T., Leo, E., Pörtner, H.-O., Schmidt, M., Windisch, H. S., Knust, R., & Mark, F. C. (2016). New encounters in Arctic waters: A comparison of metabolism and performance of polar cod (Boreogadus saida) and Atlantic cod (Gadus morhua) under ocean acidification and warming. *Polar Biology*, *39*, 1137–1153.

Kuo, E. S., & Sanford, E. (2009a). Geographic variation in the upper thermal limits of an intertidal snail: Implications for climate envelope models. *Marine Ecology Progress Series*, *388*, 137–146.

Kuo, E. S., & Sanford, E. (2009b). Geographic variation in the upper thermal limits of an intertidal snail: Implications for climate envelope models. *Marine Ecology Progress Series*, *388*, 137–146.

Landrigan, P. J., Stegeman, J. J., Fleming, L. E., Allemand, D., Anderson, D. M., Backer, L. C., Brucker-Davis, F., Chevalier, N., Corra, L., & Czerucka, D. (2020). Human health and ocean pollution. *Annals of Global Health*, *86*(1).

Langangen, Ø., Olsen, E., Stige, L. C., Ohlberger, J., Yaragina, N. A., Vikebø, F. B., Bogstad, B., Stenseth, N. C., & Hjermann, D. Ø. (2017). The effects of oil spills on marine fish: Implications of spatial variation in natural mortality. *Marine Pollution Bulletin*, *119*(1), 102–109.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 1–13.

Lau, W. W., Shiran, Y., Bailey, R. M., Cook, E., Stuchtey, M. R., Koskella, J., Velis, C. A., Godfrey, L., Boucher, J., & Murphy, M. B. (2020). Evaluating scenarios toward zero plastic pollution. *Science*, *369*(6510), 1455–1461.

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2016). Resource base influences genome-wide DNA methylation levels in wild baboons (Papio cynocephalus). *Molecular Ecology*, *25*(8), 1681–1696.

Lear, K. O., Whitney, N. M., Morgan, D. L., Brewster, L. R., Whitty, J. M., Poulakis, G. R., Scharer, R. M., Guttridge, T. L., & Gleiss, A. C. (2019). Thermal performance responses in free-ranging elasmobranchs depend on habitat use and body size. *Oecologia*, *191*(4), 829–842.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Seidov, D., & Yarosh, E. S. (2012). World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, *39*(10), Article 10.

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, *366*(6453), 362–365.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009a). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009b). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Liu, D., Wu, S., Xu, H., Zhang, Q., Zhang, S., Shi, L., Yao, C., Liu, Y., & Cheng, J. (2017). Distribution and bioaccumulation of endocrine disrupting chemicals in water, sediment and fishes in a shallow Chinese freshwater lake: Implications for ecological and human health risks. *Ecotoxicology and Environmental Safety*, *140*, 222–229.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21.

MacLeod, M., Arp, H. P. H., Tekman, M. B., & Jahnke, A. (2021). The global threat from plastic pollution. *Science*, *373*(6550), 61–65.

Marguerat, S., & Bähler, J. (2010). RNA-seq: From technology to biology. *Cellular and Molecular Life Sciences*, *67*, 569–579.

McCarty, M., & Avery, O. T. (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: II. Effect of desoxyribonuclease on the biological activity of the transforming substance. *The Journal of Experimental Medicine*, *83*(2), 89.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., & Jaffe, D. B. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770.

Meyer, J., & Di Giulio, R. (2002). Patterns of heritability of decreased EROD activity and resistance to PCB 126-induced teratogenesis in laboratory-reared offspring of killifish (Fundulus heteroclitus) from a creosote-contaminated site in the Elizabeth River, VA, USA. *Marine Environmental Research*, *54*(3–5), 621–626.

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., & Menemenlis, D. (2006). Inverse estimates of anthropogenic CO2 uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, *20*(2).

Moffett, E. R., Fryxell, D. C., Palkovacs, E. P., Kinnison, M. T., & Simon, K. S. (2018). Local adaptation reduces the metabolic cost of environmental warming. *Ecology*, *99*(10), 2318–2326.

Moore, C. J. (2015). How much plastic is in the ocean? You tell me. *Mar. Pollut. Bull*, *92*, 1–3.

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23–38.

Moss, R. H., Babiker, M., Brinkman, S., Calvo, E., Carter, T., Edmonds, J. A., Elgizouli, I., Emori, S., Lin, E., & Hibbard, K. (2008). *Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies*. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Munday, P. L., Warner, R. R., Monro, K., Pandolfi, J. M., & Marshall, D. J. (2013). Predicting evolutionary responses to climate change in the sea. *Ecology Letters*, *16*(12), 1488–1500.

Nacci, D. E., Champlin, D., & Jayaraman, S. (2010). Adaptation of the estuarine fish Fundulus heteroclitus (Atlantic killifish) to polychlorinated biphenyls (PCBs). *Estuaries and Coasts*, *33*, 853–864.

Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A., & Ramirez-Villegas, J. (2020). High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific Data*, *7*(1), Article 1.

Norin, T., Bailey, J. A., & Gamperl, A. K. (2019). Thermal biology and swimming performance of Atlantic cod (Gadus morhua) and haddock (Melanogrammus aeglefinus). *PeerJ*, *7*, e7784.

Norin, T., & Gamperl, A. K. (2018). Metabolic scaling of individuals vs. populations: Evidence for variation in scaling exponents at different hierarchical levels. *Functional Ecology*, *32*(2), Article 2. https://doi.org/10.1111/1365-2435.12996

Nowell, L. B., Brownscombe, J. W., Gutowsky, L. F., Murchie, K. J., Suski, C. D., Danylchuk, A. J., Shultz, A., & Cooke, S. J. (2015). Swimming energetics and thermal ecology of adult bonefish (Albula vulpes): A combined laboratory and field study in Eleuthera, The Bahamas. *Environmental Biology of Fishes*, *98*, 2133–2146.

Nychka, D., Furrer, R., Paige, J., & Sain, S. (2017). *fields: Tools for spatial data. R Package Version 11.6*.

Ohlberger, J., Mehner, T., Staaks, G., & Hölker, F. (2012). Intraspecific temperature dependence of the scaling of metabolic rate with body mass in fishes and its ecological implications. *Oikos*, *121*(2), 245–251.

Oleksiak, M. F., Karchner, S. I., Jenny, M. J., Franks, D. G., Mark Welch, D. B., & Hahn, M. E. (2011). Transcriptomic assessment of resistance to effects of an aryl hydrocarbon receptor (AHR) agonist in embryos of Atlantic killifish (Fundulus heteroclitus) from a marine Superfund site. *BMC Genomics*, *12*, 1–18.

Osterberg, J. S., Cammen, K. M., Schultz, T. F., Clark, B. W., & Di Giulio, R. T. (2018). Genome-wide scan reveals signatures of selection related to pollution adaptation in non-model estuarine Atlantic killifish (Fundulus heteroclitus). *Aquatic Toxicology*, *200*, 73–82.

Palumbi, S. R. (2001). Humans as the world's greatest evolutionary force. *Science*, *293*(5536), 1786–1790.

Pang, X., Yuan, X.-Z., Cao, Z.-D., Zhang, Y.-G., & Fu, S.-J. (2015). The effect of temperature on repeat swimming performance in juvenile qingbo (Spinibarbus sinensis). *Fish Physiology and Biochemistry*, *41*, 19–29.

Paschke, M., Bernasconi, G., & Schmid, B. (2003). Population size and identity influence the reaction norm of the rare, endemic plant Cochlearia bavarica across a gradient of environmental stress. *Evolution*, *57*(3), 496–508.

Pavey, S. A., Bernatchez, L., Aubin-Horth, N., & Landry, C. R. (2012). What is needed for next-generation ecological and evolutionary genomics? *Trends in Ecology & Evolution*, *27*(12), 673–678.

Pawlowicz, R. (2020). M_Map: A mapping package for MATLAB, version 1.4 m. *Computer Software, UBC EOAS, Available at: Https://Www. Eoas. Ubc. ca/∼ Rich/Map. Html, Last Access*, *29*.

Payne, N. L., Smith, J. A., van der Meulen, D. E., Taylor, M. D., Watanabe, Y. Y., Takahashi, A., Marzullo, T. A., Gray, C. A., Cadiou, G., & Suthers, I. M. (2016). Temperature dependence of fish performance in the wild: Links with species biogeography and physiological thermal tolerance. *Functional Ecology*, *30*(6), 903–912.

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography*, *12*(5), Article 5.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667.

Peterson, C. H., Rice, S. D., Short, J. W., Esler, D., Bodkin, J. L., Ballachey, B. E., & Irons, D. B. (2003). Long-term ecosystem response to the Exxon Valdez oil spill. *Science*, *302*(5653), 2082–2086.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine taxa track local climate velocities. *Science*, *341*(6151), 1239–1242.

Pörtner, H. (2001). Climate change and temperature-dependent biogeography: Oxygen limitation of thermal tolerance in animals. *Naturwissenschaften*, *88*(4), Article 4.

Pörtner, H. O., & Farrell, A. P. (2008). Physiology and climate change. *Science*, *322*(5902), Article 5902.

Pörtner, H. O., & Knust, R. (2007). Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, *315*(5808), Article 5808.

Pörtner, H.-O. (2010). Oxygen-and capacity-limitation of thermal tolerance: A matrix for integrating climate-related stressor effects in marine ecosystems. *Journal of Experimental Biology*, *213*(6), Article 6.

Pörtner, H.-O. (2014). How and how not to investigate the oxygen and capacity limitation of thermal tolerance (OCLTT) and aerobic scope – remarks on the article by Gräns et al. *Journal of Experimental Biology*, *217*(24), Article 24. https://doi.org/10.1242/jeb.114181

Pörtner, H.-O., & Giomi, F. (2013). Nothing in experimental biology makes sense except in the light of ecology and evolution – correspondence on J. Exp. Biol. 216, 2771-2782. *Journal of Experimental Biology*, *216*(23), Article 23. https://doi.org/10.1242/jeb.095232

Raffel, T. R., Romansic, J. M., Halstead, N. T., McMahon, T. A., Venesky, M. D., & Rohr, J. R. (2013). Disease and thermal acclimation in a more variable and unpredictable climate. *Nature Climate Change*, *3*(2), 146–151. https://doi.org/10.1038/nclimate1659

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., Karchner, S. I., Hahn, M. E., Nacci, D., & Oleksiak, M. F. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, *354*(6317), 1305–1308.

Reygondeau, G., & Beaugrand, G. (2011). Future climate-driven shifts in distribution of Calanus finmarchicus. *Global Change Biology*, *17*(2), Article 2.

Reznick, D. N., & Ghalambor, C. K. (2001). The population ecology of contemporary adaptations: What empirical studies reveal about the conditions that promote adaptive evolution. *Microevolution Rate, Pattern, Process*, 183–198.

Rhein, M., Rintoul, S. R., Aoki, S., Campos, E., Chambers, D., Feely, R. A., Gulev, S., Johnson, G. C., Josey, S. A., & Kostianoy, A. (2013). Observations: Ocean. In 'Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.' *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Righton, D. A., Andersen, K. H., Neat, F., Thorsteinsson, V., Steingrund, P., Svedäng, H., Michalsen, K., Hinrichsen, H.-H., Bendall, V., & Neuenfeldt, S. (2010). Thermal niche of Atlantic cod Gadus morhua: Limits, tolerance and optima. *Marine Ecology Progress Series*, *420*, 1–13.

Riley, A. K., Chernick, M., Brown, D. R., Hinton, D. E., & Di Giulio, R. T. (2016). Hepatic responses of juvenile Fundulus heteroclitus from pollution-adapted and nonadapted populations exposed to Elizabeth River sediment extract. *Toxicologic Pathology*, *44*(5), 738–748.

Ryu, T., Veilleux, H. D., Donelson, J. M., Munday, P. L., & Ravasi, T. (2018). The epigenetic landscape of transgenerational acclimation to ocean warming. *Nature Climate Change*, *8*(6), 504–509.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S. L., Wallace, D. W., & Tilbrook, B. (2004). The oceanic sink for anthropogenic CO2. *Science*, *305*(5682), 367–371.

Schulte, P. M., Healy, T. M., & Fangue, N. A. (2011). Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure. *Integrative and Comparative Biology*, *51*(5), 691–702.

Schulzweida, U., Kornblueh, L., & Quast, R. (2006). CDO user's guide. *Climate Data Operators, Version*, *1*(6), Article 6.

Schwieterman, G. D., Crear, D. P., Anderson, B. N., Lavoie, D. R., Sulikowski, J. A., Bushnell, P. G., & Brill, R. W. (2019). *Metabolic Rates and Hypoxia Tolerences of clearnose skate (Rostaraja eglanteria), summer flounder (Paralichthys dentatus), and thorny skate (Amblyraja radiata)*.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.

Sinclair, B. J., Marshall, K. E., Sewell, M. A., Levesque, D. L., Willett, C. S., Slotsbo, S., Dong, Y., Harley, C. D. G., Marshall, D. J., Helmuth, B. S., & Huey, R. B. (2016). Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? *Ecology Letters*, *19*(11), Article 11. https://doi.org/10.1111/ele.12686

Slesinger, E., Andres, A., Young, R., Seibel, B., Saba, V., Phelan, B., Rosendale, J., Wieczorek, D., & Saba, G. (2019). The effect of ocean warming on black sea bass (Centropristis striata) aerobic scope and hypoxia tolerance. *PLoS One*, *14*(6), e0218390.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, *484*(7394), 339–344.

Somero, G. N., Lockwood, B. L., & Tomanek, L. (2017). *Biochemical adaptation: Response to environmental challenges, from life's origins to the Anthropocene*. Sinauer Associates, Incorporated Publishers.

Steffen, W., Crutzen, P. J., & McNeill, J. R. (2007). The Anthropocene: Are humans now overwhelming the great forces of nature. *Ambio-Journal of Human Environment Research and Management*, *36*(8), 614–621.

Stocker, T. (2014). *Climate change 2013: The physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Türker, H. (2011). The effect of water temperature on standard and routine metabolic rate in two different sizes of Nile tilapia. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*.

UNESCO, U. F. (n.d.). *Figures on Marine Pollution, 2017*.

van Rijn, I., Buba, Y., DeLong, J., Kiflawi, M., & Belmaker, J. (2017). Large but uneven reduction in fish size across species in relation to changing sea temperatures. *Global Change Biology*, *23*(9), 3667–3674.

Van Sebille, E., Aliani, S., Law, K. L., Maximenko, N., Alsina, J. M., Bagaev, A., Bergmann, M., Chapron, B., Chubarenko, I., & Cózar, A. (2020). The physical oceanography of the transport of floating marine debris. *Environmental Research Letters*, *15*(2), 023003.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270*(5235), 484–487.

Vikas, M., & Dwarakish, G. S. (2015). Coastal pollution: A review. *Aquatic Procedia*, *4*, 381–388.

Wang, H.-Q., Tuominen, L. K., & Tsai, C.-J. (2011). SLIM: A sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics*, *27*(2), 225–231.

Wang, X., & Bhandari, R. K. (2020). DNA methylation reprogramming in medaka fish, a promising animal model for environmental epigenetics research. *Environmental Epigenetics*, *6*(1), dvaa008.

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences*, *102*(suppl_1), 6543–6549.

Whitehead, A., Clark, B. W., Reid, N. M., Hahn, M. E., & Nacci, D. (2017). When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (Fundulus heteroclitus) populations. *Evolutionary Applications*, *10*(8), Article 8.

Whitehead, A., Triant, D. A., Champlin, D., & Nacci, D. (2010). Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology*, *19*(23), 5186–5203.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A. G., de Souza Dias, B. F., Ezeh, A., Frumkin, H., Gong, P., & Head, P. (2015). Safeguarding human health in the Anthropocene epoch: Report of The Rockefeller Foundation–Lancet Commission on planetary health. *The Lancet*, *386*(10007), 1973–2028.

Xie, H., Lü, X., Zhou, J., Shi, C., Li, Y., Duan, T., Li, G., & Luo, Y. (2017). Effects of acute temperature change and temperature acclimation on the respiratory metabolism of the snakehead. *Turkish Journal of Fisheries and Aquatic Sciences*, *17*(3), 535–542.

BIOGRAPHY OF THE AUTHOR

Akila Harishchandra was born in a coastal village in the Galle district, Southern Sri Lanka. He grew up in the same region and earned the university entry qualifications from Nalanda College, Colombo, the capital city of Sri Lanka. Then he joined as an undergraduate of the first batch of the Faculty of Fisheries and Marine Sciences and Technology, University of Ruhuna, Matara, Sri Lanka, and graduated with a bachelor's degree in Oceanography and Marine Geology (honest stream) in 2011. After a short period of working as a temporary teaching assistant in his faculty, in the same year, he joined as a research scientist in the Oceanography Division of the National Aquatic Resources Research and Development Agency (NARA), Sri Lanka, the research arm of the Ministry of Fisheries and Aquatic Resources Development in Sri Lanka.

During his seven years of service in NARA, he participated in a range of ocean data collection programs around the country and earned extensive knowledge on data analyzing techniques as well. He is a PADI-qualified Rescue diver and a trained research diver too. In 2019, he joined as a graduate student in Dr. Nishad Jayasundara's lab at the School of Marine Sciences, University of Maine.

Chapter 2 has been published as:

Harishchandra, A., Xue, H., Salinas, S., & Jayasundara, N. (2022). Thermal physiology integrated species distribution model predicts profound habitat fragmentation for estuarine fish with ocean warming. Scientific Reports, 12(1), 21781.

Chapters 3 and 4 have been combined into a single manuscript and are being prepared for submission.

He is a candidate for the Doctor of Philosophy degree in Marine Biology from the University of Maine in August, 2023.