Introduction

Over the last several decades, climate change has driven the restructuring of ecosystems (Scheffers et al. 2016). Mainly, species distributions are one of the most impacted ecological processes across all ecosystems, as species distributions continue to track optimal (Chen et al. 2011; Jones and Cheung 2015; Scheffers et al. 2016). This is particularly prominent in marine species because marine ecosystems experience much greater climate change velocities and seasonal shifts than terrestrial systems (Burrows et al. 2011). Marine species are also better at tracking latitudinal isotherm shifts than terrestrial species (Pinsky and Palumbi 2014; Lenoir et al. 2020). The main reason for this is that marine organisms tend to live closer to their upper thermal limits (Pinsky et al. 2019) and have fewer dispersal and colonization barriers than terrestrial organisms (Sunday et al. 2012). **Because of the increasing rate of global climate change** (Burrows et al. 2011) **that is predicted to continue** (Loarie et al. 2009)**, more marine species are expected to experience range-shifts that will occur at unprecedented rates**.

The survival of populations undergoing rapid range expansion requires more than the ability to track a thermal optimum. The fitness of a population will depend on genetic, selective, and demographic processes that it will encounter in its newly expanded range. Populations at the edge of the expanding range are often characterized by having smaller population sizes and lower genetic diversity (Brow et al. 1995; Vucetich and Waite 2003; Slatkin and Excoffier 2012; Peter and Slatkin 2013; Pierce et al. 2017). Considering this, the persistence of newly expanded populations will be directly impacted by the dispersal limitation from neighboring populations (Hastings et al. 2005; Hargreaves and Eckert 2014), interspecific competition with resident species (Svenning et al. 2014; Louthan et al. 2015), and the populations' ability to adapt to the new environment (Holt and Barfield 2011; Polechová and Barton 2015).

One process that remains largely understudied during range expansions is the rapid accumulation of genetic mutations – and in particular deleterious mutations in edge populations (Fig1) (Edmonds et al. 2004). In general, range expansions occur by a series of colonization events in waves of a few dispersed individuals at a time. This leads to the establishment of a new population that originates from few individuals, resulting in a small non-random representation of the genetic variation from the original population – a process known as founder effect (Slatkin and Excoffier 2012). As a result of founder effects, a deleterious mutation that arises at the range edge disperses to a newly expanded habitat and can rapidly increase in frequency within the new population (i.e., expansion load) **(Fig 1)**, leading to a loss in fitness (Edmonds et al. 2004; Klopfstein et al. 2006; Gilbert et al. 2017, 2018; Perrier et al. 2017; Willi et al. 2018). This process is enhanced during range expansion due to the small population sizes and low genetic variation present to select against the new mutation (Edmonds et al. 2004; Klopfstein et al. 2006; Gilbert et al. 2018). **Thus, populations at the forefront of their range expansion may be disadvantaged by having a reduced fitness and a low adaptive potential (the ability for genetic material to evolve and drive changes in traits) in response to changing environmental conditions, but there are few empirical tests of this theory.**

Expansion loads have been mostly studied in theory using quantitative modeling (Edmonds et al. 2004; Gilbert and Whitlock 2017; Gilbert et al. 2018; Foutel-Rodier and Etheridge 2020), or empirically in a few terrestrial (Bosshard et al. 2017; Willi et al. 2018; Perrier et al. 2020) and fresh-water systems (Yoshida et al. 2016; Rougemont et al. 2020). Marine ecosystems experience much faster warming velocities than terrestrial systems (Burrows et al. 2011), and marine species are better at closely tracking a thermal optimum (Pinsky et al. 2013; Lenoir et al. 2020). Therefore, marine organisms represent a unique opportunity to understand the processes that drive the accumulation of deleterious mutations during range expansions.

Moreover, individuals at the expanded front may be faced with new environmental conditions, species assemblages, inter-specific competition, and changes in trophic food webs (Perry et al. 2005). **Understanding the evolution of species traits during range expansions is crucial to developing proper conservation and management strategies and to further our understanding on the effects of climate change on species re-distributions.** In general, theory predicts range edge individuals have enhanced dispersal and colonization traits than range center individuals. These traits include, but are not limited to, larger body sizes, better body conditions, faster sexual maturation, faster growth rates, and

having a more aggressive behavior at the range edge compared to center (Chuang and Peterson 2016), but there are also few empirical tests of this theory. A study on a freshwater fish, *Coregonus Albula,* reported faster sexual maturation at the range edge but shorter body length and life span (Amundsen et al. 2012). However, this study did not control for the confounding effects of temperature. Another study looking at range expansions of tropical fishes invading temperate ecosystems found that reduced temperatures decreased performance (growth) but maintained body condition (i.e., energy reserves) (Kingsbury et al. 2020).

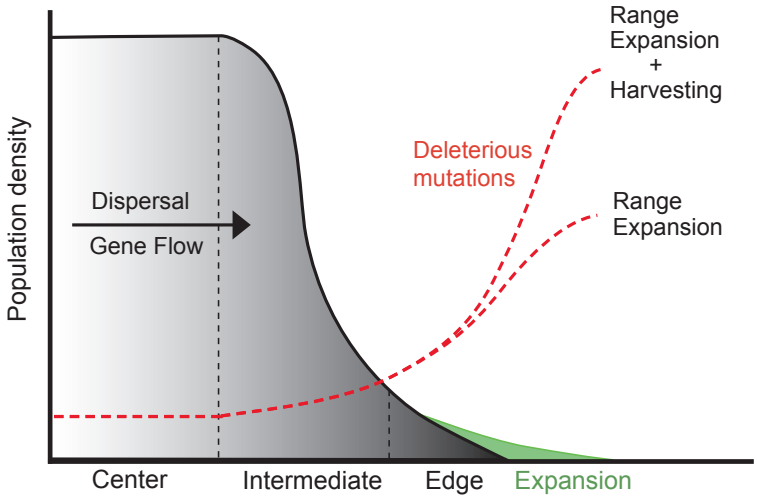

Fig 1. Range expansion demographics adapted from Chuang & Peterson (2016). Population size decreases as you transition from the range center to range edge. Range expansion tends to occur through colonization events of range edge individuals. The frequency of deleterious mutations (red dashed line) rapidly increases during range expansions (i.e., expansion load) and may be exacerbated when coupled with fishing pressure.

The Northeast U.S has been identified as one of the fastest warming marine ecosystems around the globe, with the sea surface temperature of the Gulf of Maine (GOM) increasing 99% faster than oceans worldwide (Pershing et al. 2015; Saba et al. 2016). This area is a highly productive temperate system characterized by its steep temperature gradient that has historically impeded many species from moving north (Pappalardo et al. 2015). However, warming has begun driving northward expansions in many temperate species (Nye et al. 2009; Kleisner et al. 2017), including important fisheries such as the Atlantic cod (Pershing et al. 2015) and red hake (Kleisner et al. 2017; Adams et al. 2018). Commercial species will not only need to keep up with accelerated rates of warming and distribution changes but will also need to survive existing fishing pressures (Adams et al. 2018). Overfishing can cause steep population declines, lower genetic diversity (Pinsky and Palumbi 2014), and increase the number of deleterious mutations (Rolland et al. 2020). **Moreover, large, long-lived, and vulnerable fishery species accumulate more deleterious mutations than smaller, short-lived species** (Rolland et al. 2020)**.** In marine systems, range expansions and harvesting may commonly occur simultaneously, yet, to the best of my knowledge, the coupled effect of both these forces have yet to be assessed. **Studies that focus *only* on range expansions or *only* the effects of harvesting may be underestimating their mutational load (Fig 1), which could be detrimental to marine species impacted by climate driven expansions.**

Overview

In this study, I propose to further our understanding of the evolutionary processes present in populations experiencing range expansions and fishing pressures. Black sea bass, *Centropristis striata*, is a commercially important temperate fish that has expanded more than 300 km north from their historic northern range in the last decade **(Fig 2)** (McMahan 2017). There are three recognized populations of *C. striata*: one in the Gulf of Mexico and two in the western Atlantic coast. Strong life history, morphometric, and genetic differences support these historic population designations (Mercer 1978; McGovern et al. 2002; Roy et al. 2012). Historically, the northern populations were only found up to Cape Cod and were rarely seen in the Gulf of Maine (GOM). However, their distribution continues to expand further into the GOM and were first reported in mid-coast Maine as bycatch by commercial lobstermen in 2012 (Fig. 2) (McMahan 2017). *C. striata* represent a promising alternative for the Atlantic Cod fishery, which has been slowly declining in the (GOM) due to warming (Pershing et al. 2015). **Because of this rapid expansion, *C. striata* provide an ideal study species to test the evolutionary**

theory of deleterious mutations in climate-driven range expansions that are simultaneously exposed to fishing pressures.

I will use a combination of genome-wide data, fitness tests, and quantitative modeling approaches to test multiple hypotheses of the evolutionary theory of deleterious mutational loads in a rapidly expanding marine temperate fish, *C. striata*. Understanding the frequency and effect of deleterious mutations across a species range will offer insight into how expanding marine populations will respond to selective forces of global warming and harvesting. Specifically, using a comprehensive approach I will test the evolutionary theory of deleterious mutations by: **(Obj 1)** quantifying expansion load across the range of a rapidly expanding and harvested marine fish species, *C. striata*; **(Obj 2)** evaluating fitness of *C. striata* through a series of proxy tests (i.e., body condition); **(Obj 3)** running projection simulations to assess the evolutionary impact of range expansions and fishing pressure on marine fishes. With this in mind, my study will test the following hypotheses:

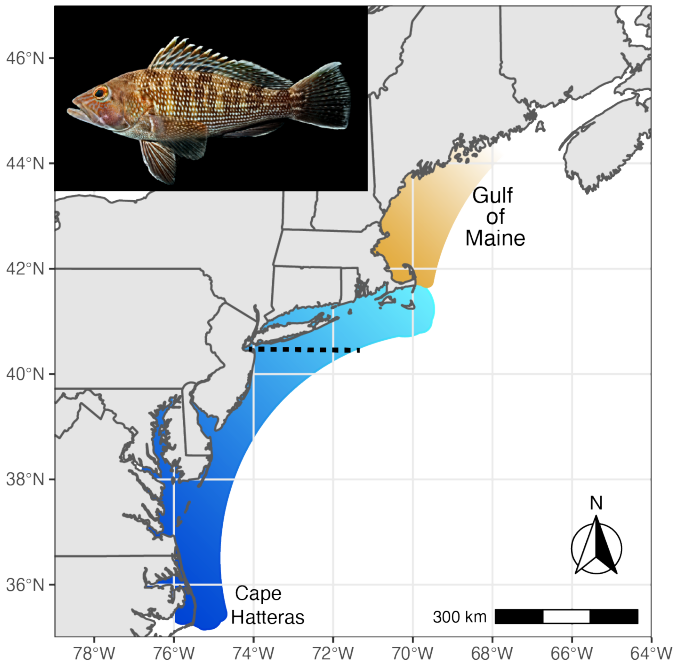

Fig 2. Northern distribution of black sea bass (top left) showing the historical range in blue and the newly expanded range in orange (Gulf of Maine). The dotted black line indicates my designated cutoff between the historical range center (Cape Hatteras to Long Island; dark blue) and the historical range edge (Long Island to south of Cape Cod; light blue).

- **H1: (Higher load at range edge):** *C. striata* will show an increase in deleterious mutations in coding regions at the range expansion front compared to the historical range center and edge populations.
    - **In Objective 1**, I will test H1 by using genome-wide coding region data (EecSeq) to obtain estimates of population structure, effective population size, genetic diversity, migration rates, and quantify deleterious mutations.
- **H2: (Decreased fitness at range edge)** Range expanding populations with a greater expansion load will show evidence of decreased fitness-related traits.
    - **In Objective 2**, I will test H2 by quantifying fitness-related body condition traits (length:weight ratio, percent protein, and C:N ratio) to test whether deleterious mutations affect body condition (i.e., fitness), while accounting for the potentially confounding effects of temperature, age, and diet quality.
- **H3: (Load exacerbated by fishing)** Overfishing will exacerbate the accumulation of deleterious mutational load in expanding populations.
    - **In Objective 3,** I will test H3 by using simulations to extend theory into a parameter space relevant to marine species with a high dispersal capacity that undertake range expansion and harvesting at the expanded range.

Host Organization and Sponsoring Scientist

My proposed research organization is the Marine Science Center (MSC) at Northeastern University (NU), in Nahant, Massachusetts. The MSC is a state-of-the-art facility with an interdisciplinary science community, where I will have access to a molecular laboratory, genomics core lab, flow-through water

system, and a high-performance computer cluster. Additionally, the MSC is devoted to enhancing community outreach and provides an excellent platform to reach underrepresented students.

My proposed sponsoring mentor is Dr. Kathleen Lotterhos, an expert in eco-evolutionary genomics, marine biology, and seascape genetics. Her research focuses on how climate has shaped biodiversity in the past and how a rapidly accelerating climate will affect biodiversity in the future. Dr. Lotterhos is my ideal project mentor because she uses a holistic approach in her own research by combining field surveys, experiments, mathematical modeling, genomics, and bioinformatics to address how climate has and will shape marine biodiversity. Additionally, after speaking with her lab members, I was impressed by how devoted she was to mentorship and to the professional development of her students and members of her lab. Dr. Lotterhos provides the knowledge, expertise, and mentorship skills to ensure the success of this research project and to further my professional development and career in science. She will oversee the overall project, provide lab space, and work closely with me to develop and improve my scientific skills.

I have also assembled a team of collaborators who will help collect tissue samples of *C. striata* and communicate results with fishery management. Dr. Jonathan Grabowski is a Professor at NU with expertise in the ecological and social dynamics of New England fisheries. Over the last two decades, he has established a collaborative network of fishers who will help secure my field collections. Dr. Marissa McMahan is the Fisheries Division Director at Manomet, a sustainability nonprofit in New England. Her dissertation focused on the ecological and socioeconomic impacts of black sea bass populations expanding into the GOM, making her one of the most knowledgeable people on the subject. Finally, Dr. Kiersten Curti is a Research Fishery Biologist in the Population Dynamics branch at NOAA Northeast Fisheries Center (NFSC). She is the secondary NOAA scientist on the black sea bass stock assessment committee for the Atlantic States Marine Fisheries Commission (ASMFC).

Objective 1. Quantify expansion load

1.1) Background

Studies that approach the evolutionary theory of the accumulation of deleterious mutations during range expansions (i.e., expansion load) have relied heavily on quantitative modelling and terrestrial systems. *C. striata* is a marine species that has recently undergone a rapid range expansion and provides a unique opportunity to address theoretical assumptions in a real marine dataset.

My study will build upon an existing research project of Dr. Lotterhos, carried out in collaboration with Dr. Jonathan Grabowski from Northeastern University, that aims to evaluate the ecological, socioeconomic, and population structure of *C. striata* expanding into the GOM. Their existing project uses double-digest restriction-site associated sequencing (ddRADseq) on a small number of samples to obtain an initial assessment of the population structure of *C. striata*. It is important to note that this sparse data is not appropriate to study the evolutionary theory of deleterious mutations, which requires larger sample sizes for statistical power (Lotterhos and Whitlock 2015) and to obtain more extensive coverage across the coding regions of the genome. However, their previous specimen collections will be re-sequenced and added to my total collection pool (Table 1). Also, I would like to mention that Dr. Lotterhos and I have applied for funding to Revive and Restore's "Wild Genomes: Marine" call (currently pending). Our project proposes to assemble and annotate the genome of black sea bass and sequence a limited number of samples. I want to clarify that if awarded it would complement this OCE-PRF proposal and not replace it. Any project timeline overlap would be managed accordingly. Revive and Restore's budgets would also be reassessed to avoid any postdoctoral salary overlap and redirected to fund undergraduate and graduate students' mentorship.

1.2) Approach

Sampling design: Deleterious mutations are primarily recessive, suggesting that large sample sizes are needed to detect low-frequency deleterious mutations (Li et al. 2010). For this project, we have

determined that 150 individuals per population will allow us to detect deleterious mutations above a frequency of 0.003 (1/(150 ind.*2 chrom./ind.)) in each population. I will sequence the complete coding exons of three populations for a total of 450 individuals.

Data collection: *C. striata* samples will be collected from recreational and commercial fishers between July and November in 2021 and 2022 from three main regions: the historical range center (Cape Hatteras, NC to Long Island), historical range edge (Long Island to south of Cape Cod) and the newly expanded range (GOM) (Fig 2). *C. striata* are a highly migratory species. In order to maximize the probability that they collected individuals originated close to the sampling location, only individuals less than 30 cm (standard length) will be sampled. A number *C. striata* samples have been previously collected in these regions in the past two years and are in possession of Dr. Lotterhos. I will collect the remaining samples to obtain 150 individuals from each region, for a total of 450 samples (Table 1). Date, location, depth, and habitat type will be recorded at each sampling site. Temperature will be obtained from the closest NOAA logger to the sampling site. Each individual will be processed by measuring standard length and weight, identifying sex and maturity, and by collecting tissue samples (muscle, fin, and gill) that will be immediately preserved in 95% ethanol and stored at -20ºC for further processing.

Table 1. *Centropristis striata* tissue sample summary.

	Samples in hand	Samples to collect	Total samples
Northern Range – Center	50	100	150
Northern Range – Edge	40	110	150
Expanded Range	48	102	150

Molecular methods: To address my first hypothesis (H1: Higher load at range edge) I require a comprehensive representation of the genetic diversity in coding regions across the historical northern range center, range edge, and the newly expanded range (Figure 2). I will re-sequence 150 individuals per population for a total of 450 individuals (Table 1, some of the samples are in-hand). This large sample size will allow us to detect deleterious mutations with sufficient statistical power (Lotterhos and Whitlock 2015). Expressed exome capture (EecSeq), developed at Northeastern University by my sponsoring scientist Dr Lotterhos (Lotterhos and Whitlock 2015), will be used as a cost-effective genomic tool ($25/sample) to assess population structure, detect deleterious mutations, and study genomic signatures of selection across populations. Unlike traditional exome capture, EecSeq probes are made directly from *C. striata* RNA and save months of time in bioinformatics and probe development. We need high-quality DNA for the EecSeq exome probes, thus, tissue samples from a subset of 20 fish across the range (muscle, brain, liver, gonads, and heart) will be frozen in liquid nitrogen and stored at -80ºC for processing. Probe development, genomic library preparation, and exome capture will follow the protocol outlined in Puritz and Lotterhos (2018). Unique barcodes will be ligated to each individual and pooled into one single library prior to exome capture. EecSeq libraries will be sequenced on a NovaSeq S1 300 flowcell (150 bp paired-end reads) to obtain an average of 20X coverage across the exome per individual.
Bioinformatics. Once raw sequences are obtained, single-nucleotide polymorphisms (SNPs) will be detected across the exome to compare across populations following the bioinformatics pipeline outlined in Puritz and Lotterhos (2018). A *de novo* transcriptome assembly will be carried out with the RNA sequences using Trinity (Grabherr et al. 2011) following the pipeline outlined in Konrad et al (2021).
Data analysis: Theory predicts the accumulation of deleterious mutation is driven by population bottlenecks that reduce standing genetic variation and the efficacy of purifying selection (Kirkpatrick and Jarne 2000; Comeron 2017). My first step will be to assess the effective population size (N_e), genetic diversity, and gene flow between populations. **Population structure** will be determined to understand whether the expanded population is genetically differentiated from its historical northern populations using Bayesian assignment and clustering techniques with the program STRUCTURE (Falush et al. 2003; Raj et al. 2014). This will be further examined by running a principal components analysis (PCA) to depict genetic structure among individuals without *a priori* population assumptions using the program

ADEGENET (Jombart 2008; Jombart and Ahmed 2011). After the population clusters have been identified by STRUCTURE and the PCA, I will run an analysis of molecular variance (AMOVA) and pairwise site comparisons using F_{ST} and D_{est_Chao} to estimate population differentiation between clusters with the program GenoDive (Meirmans 2020). **Genetic diversity** metrics (heterozygosity and allelic richness) will be estimated using vcftools (Danecek et al. 2011) and Hierfstat (Goudet 2005), and will be plotted against distance (km) to the furthest southern sites. I expect genetic diversity to decrease the further populations are from southern ancestral populations, as predicted by theory Kirkpatrick and Jarne 2000). **Migration rates and effective population sizes** (N_e) will then be determined across space and time using a method called Migration and Population-size Surfaces (MAPS) (Al-Asadi et al. 2019). MAPS uses a coalescent model that was built upon the previous estimated effective migration surfaces method (EEMS) (Petkova et al. 2016).

After obtaining a better understanding of the population dynamics of *C. striata*, I will focus on quantifying the accumulation of deleterious mutations. The mutation of a nucleotide that results in the change of an amino acid sequence of a protein is referred to as a nonsynonymous mutation. However, not all nonsynonymous mutations are equal, each protein variation may lead to a variety of fitness effects. Neutral variants tend to persist in the population while deleterious amino acid mutations tend to be removed by purifying selection (Li 1997). In range expansion populations, purifying selection may become compromised leading to an increase in frequency of these deleterious mutations. To detect deleterious mutations, we will first search for these nonsynonymous SNPs using three programs: PROVEAN (Choi 2012; Choi and Chan 2015), MutPred2 (Pejaver et al. 2020), and SIFT (Sim et al. 2012). Only deleterious mutations detected across all three programs will be considered. Each program assigns unique scores associated with biological activity and can indicate the degree of functional impact of a given protein variation. For example, PROVEAN classifies them in two categories: deleterious (< -2.5), having a strong phenotypic effect, or neutral (> -2.5), with no phenotypic effect.

The above results will be used to calculate a relative frequency of deleterious alleles per individual p_D (Yoshida et al. 2020). The p_D for each individual is calculated as the proportion of total nonsynonymous mutations that are identified as deleterious. In threespine stickleback, this proportion varied approximately between 0.05 and 0.1 in different populations (Yoshida et al. 2020).

H1 (Higher load at range edge) will be tested with the following statistical model: $p_D \sim$ Population + 1|Site + 1|Year, where Population is a fixed factor with 3 levels (expanded range, range edge, range center) and 1|Site and 1|Year are random variables indicating the location (nested within population) and time of catch, respectively. Since p_D is proportional data, this will be modeled as a logistic generalized linear mixed model. Note that our sample collection is limited to fish < 30 cm (e.g., juveniles) to maximize the probability of catching individuals who originated near the catch site.

H1 will be supported if p_D expanded range > p_D range edge > p_D center range. If H1 is supported by the data, results could indicate population vulnerability at the expanded range and should be considered to develop fishery management strategies for *C. striata*. In a broader context, this could open new areas of opportunity to study a largely neglected phenomenon in other marine species. However, if H1 is rejected, it suggests that marine systems are unique and might be driven by different processes than terrestrial systems, possibly due to their high dispersal potential and lack of true barriers. Simulations in Obj 3 will give insight into what processes might drive these results.

Objective 2. Quantify fitness-related traits

2.1) Background

Overall, fitness can be hard to quantify, especially for long-lived species that are not easily manipulated in a laboratory setting. In fish, researchers have used *in situ* body condition as a proxy for fitness and population success, because increased body condition has been linked to survival, growth, and reproduction (Irons et al. 2007; Beveren et al. 2014).

Theory predicts that range edge individuals have enhanced dispersal and colonization traits, such as larger body sizes, better body conditions, faster sexual maturation, and faster growth rates (Chuang and

Peterson 2016). A study by McMahan et al (2020), looking at life-history differences of *C. striata* across its northern distribution found that expanded populations reached maturity at a younger age, consistent with theory, but exhibited lower body condition (length:weight ratio), contrary to expected (Chuang and Peterson 2016). In this study, diversity and quality of prey decreased with latitude, directly impacting body condition of *C. striata*. However, these observed differences may have also arisen from the latitudinal variation in temperature (Yamahira and Conover 2002). **The goal of this objective is to statistically test for an association between fitness-related traits (body length: weight ratio, percent protein, and C:N ratio) and p_D, the relative frequency of deleterious alleles per individual, while controlling for the potentially confounding effects of the environment (temperature and diet quality).** It is important to note that changes in traits across a species range could be partially due to plasticity (Ghalambor et al. 2007), which is why my statistical model will incorporate variables that drive plastic changes in fish condition (temperature and diet quality). Note that logistical difficulties involved in transporting and maintaining live in aquaria over long distances prohibit experiments in many commercially important marine fish, including *C. striata*. Although I will use a field-based approach, it is designed with a comprehensive data collection for each individual that will be used to control for potentially confounding effects in a statistical model.

2.2) Approach

Phenotypes for all individuals collected in Obj. 1 (Table1) will be quantified at the time of collection. After euthanization (following IACUC approved protocols), weighed (wet weight), and measured (standard length), a small piece of white muscle will be removed and immediately frozen on ice. Body condition will be used as a proxy for fitness and population success (Irons et al. 2007; Beveren et al. 2014). Hence, body condition of each fish will be assessed using three approaches: Fulton's condition index, percent protein, and atomic mass ratio of carbon to nitrogen (C:N ratio). Fulton's condition index (K) is a weight:length ratio that relates to the energy content of muscle and liver (Lambert and Dutil 1997a). It has previously been used as an indicator of fecundity in mature adults (Lambert and Dutil 2000) and as an indicator of an individual's ability to survive in increased water temperature (Robinson et al. 2008). To obtain percent protein and C:N ratio, the muscle tissue will be freeze-dried for a minimum of 36 hours and ground to a fine powder using a ball mill. Samples will then be weighed using a Nu Instruments Horizon Continuous Flow IRMS to obtain percent nitrogen and C:N ratio. Percent nitrogen is a proxy for percent protein and will be calculated following Manthey-Karl et al (2016). The C:N ratio indicates lipid content and can be used to assess nutritional quality. Lipids are mostly composed of carbon and little nitrogen, so; high C:N levels correspond to high nutritional values (Fagan et al. 2011).

Ocean temperature, age, and diet quality will be assessed as confounding variables that could be affecting the observed condition phenotype in each fish. Temperature will be obtained from the closest NOAA logger to the sampling site. Aging will be carried out using otoliths following techniques described in McMahan et al (2020). Diet quality will be estimated using stomach fullness (Hyslop 1980) and the relative liver weight to body size ratio (Lambert and Dutil 1997b), which is thought to be a reliable indicator of the energy reserves of a fish.

H2 (Decreased fitness at range edge) will be tested using the following statistical model: Phenotype ~ p_D + Temp + Age + Diet quality. This is a multiple regression model, where phenotype is a continuous variable representing one of the three condition traits tested: body length:weight ratio, percent protein, or C:N ratio. p_D refers to the relative frequency of deleterious alleles per individual, calculated from Obj 1. This model will control for the confounding variables of ocean temperature of the collection site, age, and diet quality per individual. I will test for model assumptions (e.g., collinearity among predictors, linearity between predictors and explanatory variables, (Whitlock and Schluter 2014) and will explore other models (e.g., general additive models or random forests) if assumptions are violated. **H2 will be supported if the regression coefficient for p_D is negative for each response phenotype (e.g., fish with lower condition that also have higher deleterious mutation load after controlling for confounding variables).** If H2 is supported, this would be the first study to show an association between expansion load and a loss in a fitness-related trait in a marine species. For *C. striata* this would have

direct implications for fishery management, while in a broader context, these results would shed light on the importance of quantifying expansion loads in marine species.

Objective 3. Simulation models: range expansion + harvesting

3.1) Background

Here, I propose to develop a general model to study the mechanisms in marine range expansions that give qualitatively similar patterns as those observed in Obj 1 and 2, and assess the interactive effect of expansion and harvesting. I hypothesize (H3) that expansion load will be exacerbated when range expansion and harvesting are both present (e.g., Fig 1). Effective population size and connectivity of neighboring populations play an important role in range expansions (Gilbert et al. 2018). Marine systems are particularly characterized by having large population sizes and high gene flow, driven by the lack of true barriers in the ocean. However, current models are based on terrestrial organisms and may misrepresent the role of expansion load in marine species with high migration rates and large *Ne*. By developing this unique simulation model described below, I will test the evolutionary effects of range expansions and overfishing in marine species, which has not been done before.

3.2) Approach

Building on code published in Gilbert et al (2018), I will model forward-time range expansions of individuals over a 1-dimensional space using a discrete stepping-stone model to mimic a coastline. This code is set up to simulate range expansion by releasing a total of 1000 loci with deleterious or beneficial effects that impact fitness (Gilbert et al. 2018). However, Gilbert et al (2018) did not consider high dispersal capabilities and large population sizes that are present in marine systems. In order to more accurately represent marine species, I will modify the simulation framework by Gilbert et al (2018) by increasing population sizes, carrying capacity, and migration between populations using non-Wright Fisher models in SLiM (Haller and Messer 2019). Parameters of the model will be informed by data collected in Obj 1 and Obj 2 (e.g., N_e, migration rates, genetic diversity, fitness). Two simulation scenarios will be carried out to test the effect of (1) expansion *only* and (2) expansion *with* harvesting, on the accumulation of deleterious mutations and resulting fitness in the range edge compared to the range center. For the first scenario, I will model range expansions across a wide parameter space that results in minimal to maximal load in the expanding range, being sure to encompass patterns that qualitatively resemble data from Objs. 1 and 2. For scenario 2, direct impact of harvesting will be represented by the removal of individuals from a population. I will build off scenario 1 and will remove between 10% to 50% (in 10% increments) of individuals per generation in the range expanded population. Each scenario will be plotted to show its effect on fitness and expansion load through time (generations). I plan to work directly with Dr. Lotterhos to develop and run these simulation models, as her lab has ample experience running simulation models of this sort (Lotterhos and Whitlock 2014, 2015; Lotterhos 2019).

H3 (Load exacerbated by harvesting) will be supported if the deleterious mutation load and fitness values are significantly higher when modeling range expansion and harvesting, rather than range expansion alone. Since I will be simulating a wide parameter space, this research will advance knowledge of the conditions under which H3 does or does not hold. The goal of this objective is not to model *C. striata* specifically, but to develop a general model that will lead to a deeper understanding of (i) the mechanisms that may be driving the patterns we observe in Obj 1 and 2 (e.g. if H1 and H2 are not supported, for example), and (ii) general conditions that could lead to a rapid fishery collapse during range expansions of marine species.

Intellectual merit

My comprehensive approach will test conventional knowledge on the evolutionary theory that predicts higher expansion load of deleterious mutations in individuals at the expanding-front of a species, leading to a loss of fitness, which could have detrimental effects on these populations. In addition, this theory remains largely understudied in marine systems, where range expansion are rapid and species are

commonly subject to harvesting. Therefore, *C. striata* provides a unique opportunity to test evolutionary theory in an empirical setting where both range expansion and harvesting are present. If the results support my hypotheses (H1: higher load at range edge; H2: lower fitness of range edge; H3: load exacerbated by harvesting) they would have broad implications to develop proper conservation and management strategies in marine populations experiencing range expansions. With climate change rapidly accelerating, considering the genetic impacts of range expansions may be vital to predict the persistence of many marine species. However, if the hypotheses are not supported, this research will challenge existing evolutionary theory and suggest that evolutionary dynamics may play out differently in marine systems characterized by large populations sizes and high migration rates. The simulations will provide a unifying framework to interpret these potentially different outcomes in *C. striata* and other species.

Broader Impacts and Broadening Participation in Science

Results of this study have broad implications on the conservation and management of biodiversity in the face of global climate change. Assessing the genetic impacts on range expansions could be crucial to predicting the persistence of many marine species. Genomic insight from this research will address a fundamental gap on whether *C. striata*'s expanded population of the GOM should be managed separately from range center populations. With help from my collaborators, Dr. Grabowski, Dr. McMahan, and Dr. Curti (see *Project Summary*), results will be presented and shared with the local fishing community (e.g., Maine Fishermen's Forum), Manomet, NOAA Northeast Fisheries Science Center, the relevant committees (e.g., Science and Statistical Committees, Research Committees) of the New England and Mid-Atlantic Fishery Management Councils (NEFMC and MAFMC), and the black sea bass stock assessment committee of the Atlantic States Marine Fisheries Commission (ASFMC).

Despite ongoing initiatives over the last 40 years in the scientific community to enhance diversity and inclusion in our field, multiple racially marginalized groups (e.g., Black, Hispanic, Native American) remain largely underrepresented in STEM (Bernard and Cooperdock 2018). Recognizing biased racist history and barriers present in STEM fields is the first step to overcome these boundaries**.** As a Hispanic scientist, I have witnessed first-hand how being exposed early on to scientists and hands-on experiences can motivate one's decision to pursue a career in STEM. Using direct-action and evidence-based approaches, I plan to increase retention and inclusion of students, focusing on historically marginalized groups, in ecology, evolutionary, and conservation biology (EECB). **My efforts will focus on three main objectives: (1) undergraduate student mentorship, (2) outreach seminars for students in local community colleges, and (3) outreach with the local community.** These groups engage students at two critical stages in their lives: when they are first getting excited about science (K-12) and when they decide they want to pursue a career in science (undergraduates).

(1) I will actively recruit and mentor undergraduate students through Northeastern University's (NU) Marine and Environmental Sciences program. Students with the desire to get research experience will help throughout my project in field collections, DNA/RNA extractions, phenotypic measurements, and data analysis. I will develop clear goals alongside each student to develop their scientific skills with hands-on training in lab techniques, computational analyses, and by delivering constructive feedback on their written products. I will also focus on non-academic skills necessary for their career success (e.g., professional communication, cultural values, networking, and work-life balance in academia). During my PhD, I have successfully mentored two undergraduate students and helped train graduate students in a broad range of evolutionary biology topics and the use of molecular techniques and bioinformatics to develop their own research projects.

(2) As a postdoctoral fellow, I plan to broaden my reach and approaches to target underrepresented groups in EECB. **I will put together a series of outreach seminars to talk about the hidden curriculum in academia, targeting undergraduates from local community colleges and universities, such as North Shore Community College and Salem State.** Seminar topics will focus on the unspoken norms in science and what you need to know to get research experience, jobs, internships, and to get into grad school.

(3) I will use NEU's Marine Science Center's (MSC) outreach platform to reach local K-12 students and design activity-based lesson plans to demonstrate how climate change can drive range expansions and what this means for marine biodiversity and conservation. The MSC is devoted to providing resources and programs to students and the broad community to learn about the ocean. Some of their ongoing programs include the Beach Sisters program at Girls Inc in Lynn, the Annual High School Marine Science Symposium, and the Science Cafe, attracting more than 300 students per year in day- or quarter-long events. In collaboration with the MSC outreach program I will develop two teaching modules based on this research, climate change, and range expansions on marine species.

Long-term career goals and the Role of this Experience

As a professional scientist, I aim to build and maintain my own research program that focuses on understanding the processes that drive biodiversity and local adaptation patterns in marine organisms and how these are impacted under a rapidly changing climate. I aim to produce results that will challenge evolutionary theory and be applied to marine conservation and management efforts. Concurrently, I look forward to providing students, teachers, and the broader public with opportunities to engage in research and learn about the ocean. For my Master's and PhD, I successfully conducted population genetic studies in a variety of marine organisms. My research jumped from using a handful of genetic loci (microsatellite markers) (Gatins et al. 2018), to a small representation of the genome (RADseq), to assembling whole genomes. Through my research, I became increasingly interested in understanding how climate change will affect dispersal patterns in marine systems and the evolutionary consequences of these processes. **The NSF Ocean Science Postdoctoral fellowship (OCE-PRF) will allow me to expand my research from describing divergence patterns to actively testing the evolutionary consequences of a rapidly changing climate, considering migration and demographic effects.** My proposed project will train me in new molecular techniques (EecSeq), designing statistical models, developing computational genomics skills, and mentorship skills. This study will also expand my study focus to include temperate species, which expands my experience beyond tropical reef organisms. By dealing with a commercial important fishery species, I will get the chance to improve my scientific communication skills to translate science into conservation management. Overall, the OCE-PRF will further my career goals by providing broad opportunities for scientific advancement and professional development, which will prepare me to successfully establish a diverse and productive research program.

Eligibility of Fellowship. I affirm that I am a US citizen and a PhD Candidate at the University of California Santa Cruz. I will defend my dissertation at the end of May 2021.

Timeline