Project Description

1a. List of Participants, Roles, and Institutions

U.S. Participants

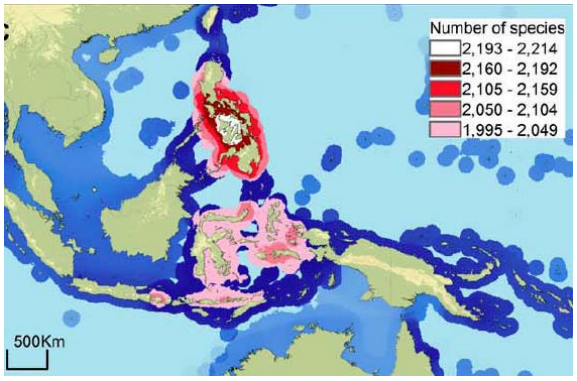
Kent E **Carpenter**, PI, Biological Sciences, Old Dominion University (ODU)
Daniel J **Barshis**, Co-PI, Biological Sciences, Old Dominion University
Christopher E **Bird**, Co-PI, Life Sciences, Texas A&M University - Corpus Christi (TAMUCC)
Malin L **Pinsky**, Co-PI, Ecology, Evolution & Natural Resources, Rutgers University (RU)
Beth A **Polidoro**, Co-PI, School of Mathematical & Natural Sciences, Arizona State University (ASU)
Jeffrey T **Williams**, Senior Personnel, Smithsonian National Museum of Natural History (NMNH)
Robin S **Waples**, Senior Personnel, Northwest Fisheries Science Center, NOAA Fisheries (NOAA)

Filipino Participants

Angel C. **Alcala**, Senior Personnel, Angelo King Center for Research & Environmental Management, Silliman University (SUAKCREM)
Mudjekeewis D **Santos**, Senior Personnel, National Fisheries Research & Development Institute (NFRDI)
Richard N **Muallil**, Collaborator, College of Technology & Oceanography, Mindanao State University Tawi Tawi (MSUTT)

1b. Research Plan

Background. Coastal development, pollution, and overexploitation of marine resources have caused widespread and severe habitat degradation and ecological impacts over recent centuries (Halpern *et al.* 2008, 2015; McCauley *et al.* 2015). What is less known, however, are the evolutionary consequences of these impacts and to what degree humans have already altered the evolutionary potential of marine species at the genomic level. Our ability to investigate ancient and historical changes in population genomics of marine species has been limited almost entirely to a few temperate species (e.g. Therkildsen *et al.* 2013a, Bonanomi *et al.* 2016, Moss *et al.* 2016, Nikulina *et al.* 2016). Our understanding of genetic changes in tropical species over long periods is even less known (Longenecker *et al*. 2015) despite heightened human impacts on these ecosystems and exceptionally high levels of biodiversity compared to temperate systems. An unprecedented opportunity exists, however, to unlock historical evolutionary impacts on tropical marine fishes because extensive historical collections in the Philippines (Smith & Williams 1999) are uniquely preserved to allow genomic investigation. Over the past century, marine populations in the Philippines have been severely impacted by intense fisheries exploitation and habitat degradation (Bryant *et al.* 1998, Roberts *et al*. 2002, Nañola *et al*. 2011, Lavides *et al.* 2016). The overarching research question of this study is: **what changes in evolution, connectivity, and community-composition have taken place in marine fishes of the Philippines over the past century during the course of substantial human impacts**? This proposed Partnerships for International Research and Education (PIRE) project will take advantage of natural phenomena unique to the Philippines, unparalleled U.S. research collections, and a proven track record of international collaboration in the region to address novel scientific questions about the evolutionary impacts of marine overexploitation and habitat loss.



The Philippines has more marine species per unit area than any other place on Earth, according to range overlap analyses (Carpenter & Springer 2005, Allen 2008, Sanciangco *et al.* 2013). This includes an epicenter of biodiversity in the central Philippines (Figure 1). The Philippines also ranks highest globally

Figure 1. Fish diversity across the Coral Triangle region, highlighting the Philippines as a center of biodiversity (Sanciangco *et al.* 2013). Blue shades range from 2048 overlapping ranges (dark blue) down to 96 (light blue).

for threats to marine biodiversity (Bryant *et al.* 1998, Roberts *et al.* 2002); for example, a hotspot of fisheries overexploitation also occurs in the central Philippines (Lavides *et al.* 2016, Nañola *et al.* 2011) where biodiversity concentration peaks. There are numerous threats to the biodiversity of coral reefs and other marine habitats in the Philippines from habitat destruction and climate change (Gomez *et al.* 1994, Gurney *et al.* 2013). However, as in other parts of the ocean, overexploitation is the most prominent threat to marine fishes (Collette *et al.* 2011, McClenachan *et al*. 2011,Sadovy de Mitcheson *et al.* 2012, Dulvy 2014).

Two large natural historical collections from the Philippines that reside at the Smithsonian National Museum of Natural History (NMNH) provide an unprecedented opportunity for long-term studies: one collection is ideal for genetic studies and the other is ideal for studying changes in species composition. The first collection is one of the greatest marine natural history collections in history: the expedition aboard the U.S. Research Vessel *Albatross* (Figure 2) from 1907 to 1909 (Smith & Williams 1999). This voyage resulted in 28,440 cataloged containers with over 91,000 specimens of fishes. An important characteristic of these collections is that nearly all were originally fixed and continue to be preserved in ethanol, thereby reducing the damage to DNA molecules compared to other preservation methods.

Figure 2. The Research Vessel *Albatross*.

The use of historical collections to address previously un-addressable research questions regarding phylogeny, phylogeography, taxonomy, demographic history and the sociality of organisms has gained recent interest from genetic studies, particularly with the advent of high throughput sequencing technologies (Bi *et al*. 2013, Fortes *et al*. 2016, Holmes *et al*. 2016, Suchan *et al.* 2016, Leonardi *et al.* 2017). For vertebrate organisms, however, opportunities to take advantage of historical collections are limited predominantly to dried tissues of mammals and birds (Wandeler *et al.* 2007, Mason *et al.* 2011, Rowe *et al.* 2011), because of the damage done to DNA in the predominantly formalin-fixed specimens of most taxa that have been collected over the past century (Chakraborty *et al*. 2006, Hykin *et al*. 2015). For fish, there are only a few species such as cod, trout, and salmon where fisheries studies have archived otoliths and scales, allowing examination of recent decadal-scale population genetic trends (Nielsen *et al.* 1999, Wandeler *et al.* 2007, Nielsen & Hansen 2008, Therkildsen *et al.* 2013a, Bonanomi *et al.* 2016). Archaeological collections of fish bones have elucidated population structure at millennial scales, but these are also limited to just a few temperate species, such as herring, sturgeon, cod, and one subtropical parrotfish (Speller *et al.* 2012, Longenecker *et al.* 2015, Ólafsdóttir *et al.* 2014, Moss *et al.* 2016, Nikulina *et al.* 2016). Unlike the formalin-fixed specimens that comprise the majority of specimens collected over the past 150 years (Hykin *et al.* 2015), ethanol fixed and preserved specimens such as the *Albatross* collections are much less prone to DNA damage. **The *Albatross* collection is unique among natural history collections worldwide in allowing for extensive 'genomic time travel' and an unprecedented opportunity to broadly examine, for the first time, the impacts of a hundred years of anthropogenic pressure on tropical marine fishes.**

Central Questions. Our overarching question is whether and to what extent genetic and species-level changes have taken place over the past century of intense fisheries exploitation and habitat degradation in the Philippines. Given our access to the unique NMNH historical tissue and community composition data from the Philippines, we will focus on three primary questions:

Q1. To what extent has genetic diversity been lost from marine species in the Philippines, and are some species more susceptible than others to loss of genetic diversity?

Q2. Have overfishing and habitat loss reduced gene flow between marine populations?

Q3. To what extent do genetic diversity and species diversity reveal similar changes over a century of human impacts?

Preliminary Data. We now have preliminary data from Restriction-site Associated DNA (RAD) analysis that clearly demonstrate *Albatross* specimens can be used to investigate population level genetic changes in fishes of the Philippines over the past century. Based on these data, we estimate that **75% of *Albatross* fish can be genotyped at 500 loci, given the observed sequencing depths and the addition of RAD capture** (see Ali *et al.* 2016). Preliminary traditional RAD sequencing (Miller *et al.* 2007) of six samples of *Caesio cuning* from the *Albatross* collections resulted in high-quality sequence data from five specimens with ~24,000 loci (RAD-tags) that were variable and genotyped in all individuals. Following this initial success, the DNA of 85 *Albatross* specimens from 64 lots (a *lot* is a jar of one or more specimens of a single species collected from a single locality) representing 42 species were isolated and sequenced following the newRAD protocol (Ali *et al*. 2016) that adds an enrichment reaction for RAD loci. Initially, all 85 specimens yielded DNA extractions with sufficient size and quantity for library prep and sequencing (see example traces in Figure 3). The large fragment sizes of samples 1 and 7 demonstrate the very high preservation quality of many of the Albatross samples (Figure 3). RAD data from these libraries netted an average of 873,000 reads/sample with a maximum of over 8 million reads/sample. Twenty-five of the specimens were re-sequenced to confirm the initial results and determine the potential for increased read depth through re-sequencing. Of the 25, 60% had more than 1 million reads/sample and 95% had more than 200,000 reads/sample (the number required to reliably genotype 500 captured RAD loci, based on Ali *et al.* 2016). We project that if all 85 samples were re-sequenced, then 48% would have more than 1 million reads/sample and 75% would have more than 200k reads/sample.

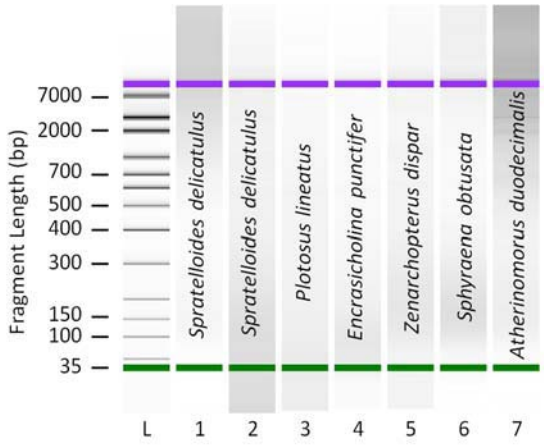


Figure 3. Bioanalyzer "gel image" visualization of *Albatross* samples showing sufficient quantity (darker = more) and quality (longer = better) of DNA for sequencing based upon our preliminary data. 100% of samples yielded DNA >7000 bp (n=175), with 23% of samples resembling lanes 1 and 7 that exhibit a high degree of preservation. Lanes resembling 2-6 also yielded adequate data. Overall 75% of samples can be genotyped.

We anticipate that the proportion of genotyped samples and reads from RAD loci can be further improved not only by the addition of RAD capture (Ali *et al.* 2016), but also via addition of normalization steps to the newRAD protocol to ensure even depth of coverage among samples and optimized ligation conditions (improvements are described in detail in "*Genotyping* methods" below). As a whole, given the historical age of the *Albatross* collections, we believe these pilot sequencing results demonstrate a very reasonable sequencing success rate that can be further optimized during additional protocol development as part of this study.

Methods.

Genetic Samples

A strength of this proposal lies in the diversity of species we will be able to analyze from the *Albatross* collection, greatly expanding our understanding of the consequences of human-driven fisheries exploitation beyond our current knowledge of a few temperate fish species (Wandeler *et al.* 2007, Bonanomi *et al.* 2016). **We will analyze single-locality populations of at least 20 species, comparing over a century of genetic changes**. This sampling is possible as the *Albatross* specimens include over 79,000 specimens in 23,877 lots from the Philippines (NMNH online database). Given our projected genotyping success rate of 75% (which we expect to increase following protocol improvement) and our target of 30-50 specimens per species (Waples & Do 2010, Wang 2016), there are more than enough lots available for meaningful hypothesis testing.

The museum records of all Philippine *Albatross* samples reveal that 81 lots contain more than 80 specimens. Of these, 48 lots have species that we are confident can be currently collected from local fish markets and in the field, based on our team's extended studies in the Philippines (Carpenter *et al*. 2011, Nañola *et a*l. 2011, Carpenter *et al.* 2017). These 48 lots include 29 species that will be the primary focus

of Central Question 1. These species have especially abundant material available, with 150 specimens per site or more present in the *Albatross* collection. These 29 species span a variety of habitats, distinctly different locations, and exploitation levels in the Philippines (Central Question 1), as well as 15 Families (Ambassidae, Apogonidae, Atherinidae, Caesionidae, Carangidae, Clupeidae, Leiognathidae, Engraulidae, Gerreidae, Gobiidae, Plotosidae, Pomacentridae, Serranidae, Siganidae, Sphyraenidae). Of these species, 8 are from coral reefs, 5 are shallow coastal semi-pelagic, 5 are pelagic, 4 are estuarine, 3 are soft-bottom benthic, 3 are found in seagrass beds and 1 is rocky intertidal. In addition, five of these species were collected at multiple locations and will be the focus of Central Question 2 (one species at 7 distinct localities, two at 4 localities, and one at 3 localities). The remaining 24 species collected from only a single locality will be used for tests of Central Question 1. All these distinct localities span areas of the Philippines previously demonstrated to exhibit population level genetic differences (Carpenter *et al.* 2011; Raynal *et al.* 2014; DeBoer *et al.* 2014a,b; Arriesgado *et al.* 2016; Guo *et al.* 2016).

The levels of exploitation and habitat destruction in the Philippines are such that all marine species were likely affected by these factors. In this sense, there is no pristine 'control' available. However, qualitative comparisons can be made between species based on the extent of influence of overfishing and habitat destruction. For example, coral reef habitats are more heavily impacted than nearshore coastal and soft-bottom habitats, in part because trawling is banned within 15 km of shore in the Philippines and other methods of fishing soft-bottom habitats are less efficient. Furthermore, the extinction risk of each of the species in this study will be thoroughly evaluated at the level of the Philippines using IUCN Red List Criteria by experts familiar with their biology and local population status (see "Broader Impacts" section) to allow further qualitative comparisons.

We will follow a multi-level iteration of *Albatross* lot selection starting with stringent species identification using the best available taxonomic keys (e.g. Carpenter & Niem, 2001) and expert consultation if necessary. DNA extraction using Qiagen DNeasy kits will be optimized for each species by testing the concentration (using fluorescent quantification, Biotium AccuBlue kit, Spectramax M3 plate reader) and fragment length distribution (via an Advanced Analytical Fragment Analyzer) of DNA yielded by different tissue types (e.g., liver, muscle, gill, fin), quantities (e.g., 10, 25, 50, 100, 150, 200 mg), and times (1907-1909 and present). The best combinations of tissue and mass will be determined using a generalized linear model. The species-specific, optimized DNA extraction protocols will be used to test subsets of individuals from the 48 candidate lots to get an indication of the variation in DNA quality and quantity within lots. The highest quality lots will then be genotyped according to the strategy described below.

While these tests are ongoing, we will begin contemporary collections. Contemporary sampling will include visits to fish landing sites and local village fish markets when provenance can be verified by vendors (Ackiss *et al*. 2013, Stockwell *et al*. 2016, Carpenter *et al*. 2017). We will avoid regional markets as these fish may have come from more distant locations. In-water sampling will also be conducted using methods such as spearing, ichthyocides, and trawling (permits have and will be procured according to Philippine law) that have proven successful in the Philippines (McManus *et al*. 1981, Gill & Williams 2011, Greenfield & Jewett 2011, Raynal *et al*. 2014, Hanson 2015).

Genotyping

We propose to directly compare genetic differences between *Albatross* era populations and present day populations using a combination of RAD capture and Illumina sequencing. This type of RAD data cost-effectively provides single nucleotide polymorphisms in sufficient quantity to accurately and precisely detect fishing-induced declines in genetic diversity (Pinsky & Palumbi 2014). In this section on genotyping, we first describe the processing of samples in general, then we emphasize specific procedures and modifications to the Ali *et al*. (2016) methodology that will ensure success in this project.

DNA Isolation- DNA will be extracted using Qiagen DNeasy kits using the optimal tissue type and mass determined during sample selection (as detailed in "*Genetic Samples*" above). Given the propensity of older samples to contain low molecular weight DNA (Figure 3), extracted DNA will be enriched for high molecular weight fragments, where appropriate, using Beckman-Coulter SPRI-Select paramagnetic beads. Size selection of DNA will be modulated with respect to the frequency distribution of fragment lengths in each sample so as to avoid removing all DNA. The concentration of all DNA samples will be quantified in triplicate on a Spectramax M3 fluorescent plate reader using the Biotium AccuBlue kit, and all samples will be normalized.

RAD Library Prep- Following Ali *et al*. (2016), DNA will be digested with New England Biolabs SbfI-HF restriction enzyme. A biotinylated, inline barcode will be ligated to the digested DNA prior to sonication with a Diagenode Bioruptor to adjust the average DNA fragment size to 300bp. The target, barcoded, biotinylated DNA will be isolated using Thermo Fisher Scientific M-280 Streptavidin Dynabeads, then a second SbfI digestion will be performed to remove the biotin-Dynabead complex. Illumina adapters will be ligated to the samples using the KAPA Biosystems Hyper Plus DNA prep kit, as in ezRAD (Toonen *et al.* 2013). The DNA concentration of each library will be quantified using a KAPA qPCR library quantification kit on an Applied Biosystems Incorporated StepONEplus real-time thermocycler. Pooled libraries within a species will, themselves, be normalized and pooled prior to capture.

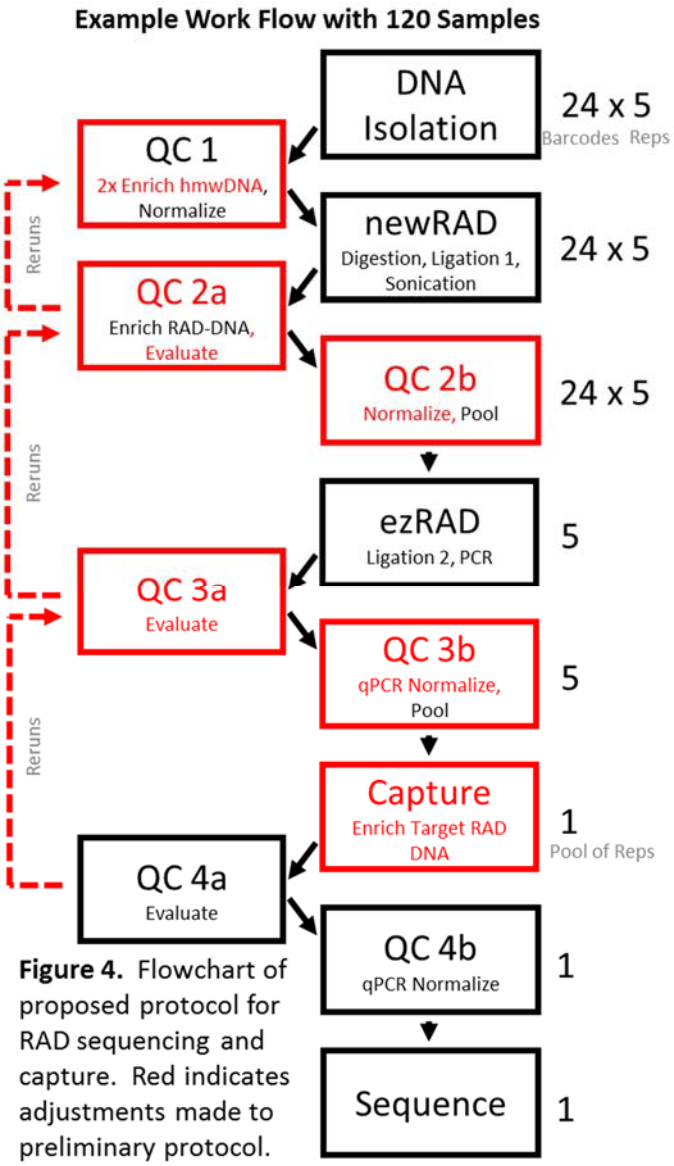


Figure 4. Flowchart of proposed protocol for RAD sequencing and capture. Red indicates adjustments made to preliminary protocol.

Capture- As explained above, our preliminary results and those of Ali *et al.* (2016) indicate that we can genotype 75% of the *Albatross* samples at 500 captured loci with 200,000 reads/fish without protocol modifications. **We aim to increase our reads/fish to 2 million and enrich libraries from each species for 5,000 loci** using custom 120bp MYcroarray MYbaits kits, where every nucleotide in each RAD locus is targeted by an average of two baits, as recommended by MYcroarray. Each kit contains custom biotinylated capture baits for one species (20 kits). Our success in reducing read depth variation among fish (see "*Keys to Success*" below) will dictate the precise number of loci, above 500, that we target with capture baits.

Prior to capture, bait sequences will be generated *de novo*. In doing so, we aim to balance the quality of the capture baits and their cost. To sample as much diversity as possible during bait sequence generation, all individuals of a given species and time period (1907-1909 and present day) will be pooled (2 time periods * 20 spp. = 40 pooled libraries) and processed as described above. To improve the selection of valid RAD loci that pass stringent filtering, nine libraries per species and time period will also be constructed as described above, with each library containing one individual (9 individuals * 2 time

periods * 20 species = 360 libraries). All libraries will be sequenced on an Illumina sequencer (2 x 150 bp paired-end; 4.5 million reads per library * 400 libraries = 1,800 million reads; 1,800 million reads / 450 million reads per lane = 4 lanes of sequencing to develop bait sequences).

Final Quality Control & Sequencing- The size of fragments in the final libraries will be selected using a Sage Science BluePippin pulsed-field electrophoresis rig, and the DNA concentration will be quantified using a KAPA qPCR library quantification kit. All libraries will be sequenced using an Illumina HiSeq 4000 (or a NovaSeq 6000 depending upon availability) at a target depth of 3 million reads per individual (paired-end 2 x 150bp; 3 million reads per fish * 3000 fish = 9,000 million reads; 9,000 million reads / 450 million reads per lane = 20 lanes for production-scale genotyping).

A very recent pre-print (Sinha *et al*. in review) has identified an issue with Illumina HiSeq 4000 sequencers where indexes can be erroneously assigned to clusters in 5-10% of sequence reads - "index switching". If this issue is confirmed by peer review and not resolved by Illumina, we will follow the prescription of Sinha *et al*. (in review) and employ a dual-indexed adapter strategy to identify and remove reads affected by "index switching". We will also spike our RAD sequencing libraries with an indexed shotgun genome library, rather than the default PhiX library, which Sinha *et al*. (in review) recommend to better estimate the prevalence of "index switching".

Keys to Success- The ultimate key to the success of this proposal is to successfully genotype DNA from 110 year old samples, which will require the processing of samples with fragmented DNA, as well as from present day samples. While this project can be completed with the 75% genotyping efficiency that we estimate given our preliminary genotyping, the proposed protocol has a number of key adjustments from our preliminary protocol in order to increase efficiency and reduce the overall cost of the project. The particular adjustments are highlighted here. **(1)** The quantity and rigor of sample and pool normalization will be increased to **ensure equal representation of each individual**, which varied in our preliminary trial and has been reported using the same protocol by Ali *et al.* (2016). Variation in read depth can be traced to the assumption of Ali *et al*. (2016) that DNA concentration does not have to be re-normalized prior to pooling. Fluorescent quantification of DNA concentration will be performed in triplicate, after sonication and Dynabead enrichment, but prior to pooling to ensure that each individual is equally represented. Prior to the pooling of pools for capture, and again prior to sequencing, the library pools will be normalized following the quantification of the concentration of targeted loci using the KAPA qPCR kit. **(2) Enriching libraries for targeted RAD loci using MYcroarray capture baits** will substantially increase the proportion of reads from targeted RAD loci (Ali *et al*. 2016) and decrease overall sequencing costs by reducing non-target reads. The libraries will be constructed by the Genomics Core Laboratory Coordinator at TAMUCC (supervised by co-PI Bird), who has four years of experience troubleshooting all varieties of RAD. **(3)** We will employ a two-step process to **enrich samples for high molecular weight DNA**, which will improve the first ligation step and the proportion of sequence reads from targeted loci. During the elution step of DNA extraction, smaller DNA fragments are differentially eluted first; thus, the second elution will have a higher proportion of high molecular weight DNA than the first. The best of three elutions per individual will be selected for further enrichment using SPRI-select beads, as described above. **(4)** In order to **minimize DNA loss during reaction cleanup**, we will employ the "with-bead" procedure of Fisher *et al.* (2011), where paramagnetic beads are retained within the samples until a plate transfer is necessary. The beads can be reactivated downstream by adding a 3M NaCl 20% PEG solution. This will improve both ligation steps and reduce adapter dimers. **(5)** Following the first ligation step, libraries will not be pooled so that **the quality of each library can be assessed independently after the Dynabead enrichment,** prior to the second ligation. This will facilitate the identification of problematic libraries prior to sequencing and allow libraries to be re-made early in the protocol. **(6)** Within a species, **samples from different "treatments" will be randomized** with respect to pooling to ensure that library preparation and sequencing are not confounded with treatment. It is important to note that the DNA from *Albatross* specimens do not resemble ancient DNA. Ancient DNA requires special ultra-clean procedures because there is so little of it that its signal can easily be swamped by contamination. While we will take appropriate measures to minimize cross-contamination, the larger concern is to avoid methodological confounds. It is well known that sequencing the same libraries on two different lanes of an Illumina HiSeq can yield "genetic structure" where there was none (Mastretta-Yanes *et al*. 2015).

Bioinformatics

We are proposing to discover target loci using 20 RAD libraries per species (see "*Genotyping*: *Capture*" section above), then conduct RAD capture on all samples using the discovered loci. First we describe, in detail, the discovery of target RAD loci, which will involve stringent filtering to identify high quality RAD loci. Then, we describe the processing of RAD capture data.

Marker discovery- Following sequencing, sequence reads will be re-associated with each sample using the process_radtags function in STACKS (Catchen *et al*. 2013). The newest version of the dDocent pipeline (Puritz *et al.* 2014a) will be used to trim adapters and low quality reads from sequences using Trimmomatic (Bolger 2014), assemble *de novo* reference "genomes" for each species using Rainbow (Chong *et al*. 2012) because none of them have draft genomes available, map reads to the reference using bwa (Li and Durbin 2009), filter improper pairs and PCR clones using samtools (Li *et al*. 2009), and genotype samples using FreeBayes (Garrison 2010). There are many genotyping packages available, and we intend to revisit the choice of a genotyper. Co-PI Bird contributes to the coding of dDocent, and it will be modified as appropriate to incorporate new advancements - also highlighting the strength of using dDocent here.

We will employ an analysis of both a subset of individuals (most stringent filtering) and pools of all individuals (best representation of genetic variation). Prior to filtering, FreeBayes will be rerun to genotype both pools and individuals at the same loci. At this point, the goal will be to remove non-target loci. The "dDocent filtering pipeline", which relies heavily on vcftools (Danecek *et al*. 2011) and vcflib (Garrison 2012), will be utilized to filter loci and samples for minimum alternate read depth and frequency, minimum nucleotide and mapping quality score, minimum mean read depth, missing data, and PCR clones. Data from fish(es) are then aggregated by location and time for "sample aware" filtering of loci and samples based upon missing data, reference allele frequency in heterozygotes (Puritz *et al*. 2014b), strand bias, imbalanced proportions of forward and reverse reads, imbalanced mapping quality between allelic states, proper pairing, deflated locus quality scores (Li 2014), maximum mean read depth, and Hardy-Weinberg equilibrium. Haplotypes for each RAD locus will be assembled using rad_haplotyper, which additionally filters loci for paralogs, missing data, low depth of coverage, genotyping errors, and excess haplotypes. The loci filtered by rad-haplotyper will also be filtered from the curated VCF file, and SNPs with more than two allelic states removed. The curated data from both individuals (SNPs and haplotypes) and pools (haplotypes) will be compared for congruence. Loci that were only filtered in one of the two sets will be scrutinized to determine if these loci should be removed. We expect to remove loci from the pools that were filtered in the individuals because the filtering process was more stringent (e.g. paralog filtering). However, it is possible that some loci will be filtered from individuals (e.g., minor allele depth) that pass the filter in the pools that harbor the whole sample. From the final filtered data set, 5000 loci will be selected at random for capture bait design.

Captured data processing- Captured data processing will follow the process described for marker discovery except that capture data will consist of only individuals (no pools) and the assembly of a reference genome will not be necessary because it will consist of the 5000 loci identified during marker discovery.

Genetic Diversity

Q1. To what extent has genetic diversity been lost from marine fishes in the Philippines, and are some species more susceptible than others?

The first use of the genetic data will be to test whether species have experienced fundamental genetic changes over the past century. Changes in the genetic makeup of populations are expected, given the intense level of exploitation of fisheries worldwide and in the Philippines in particular (Allendorf *et al.* 2014, Lavides *et al.* 2016). In addition, there has been substantial loss in habitat quantity and quality, even for those species not fished intensively (Gomez *et al.* 1994, Roberts *et al*. 2002, Nañola *et al*. 2011). Presently, over-exploited populations worldwide have about 2% lower heterozygosity and 12% lower allelic richness than less exploited populations, though these results are highly skewed towards temperate species in the US, Canada, and Europe (Pinsky & Palumbi 2014). In addition, theory predicts and empirical data confirm that genetic diversity within coral reef animals is highest in areas of high coral cover, and lowest in areas of low coral cover (Antonovics 1976, Selkoe *et al.* 2016).

However, such comparisons between related populations with divergent exploitation histories or geographic settings remain a highly indirect way for detecting loss of genetic diversity and rely on a number of assumptions, leaving open substantial questions about the magnitude and consistency of genetic changes. Comparisons in the same population over time provide a much more direct and accurate method for detecting change with fewer assumptions. The *Albatross* samples provide an opportunity to, in effect, travel back in time to the year 1907 to reconstruct evolutionary history over the past century.

As population sizes decline from overfishing or habitat loss, the strength of genetic drift and the rate of neutral evolution increases because a lower number of effective individuals contribute to the next generation (lower effective population size, N_e). The effects of increased drift are most immediately apparent through the loss of rare alleles from the population, a process that leads to a rapid decline in allelic richness (total number of alleles at a locus) (Maruyama & Fuerst 1985). Heterozygosity is the probability of any two alleles in a population being different. Over time, heterozygosity declines as well, but more slowly than allelic richness because it is less sensitive to rare alleles (Maruyama & Fuerst 1985, Ryman *et al.* 1995). The interplay between starting population sizes, magnitude of population size decline, and length of time at low abundance determine the extent to which allelic richness and heterozygosity decline (Pinsky & Palumbi 2014). Other characteristic signatures are also useful for detecting recent genetic bottlenecks, including skews in the relative frequencies of rare and common variants (Gattepaille *et al.* 2013).

In addition to stronger genetic drift, increased mortality from fishing and changing habitats may exert selection for particular phenotypes. When the variation of these phenotypes has at least a partial genetic basis, selection can drive evolutionary adaptation. Selection for early maturity at small sizes, increased reproductive investment, and slower growth have been proposed as common evolutionary responses to intense fishing, although the extent and degree to which this has occurred across species remains unclear (Marty *et al.* 2015, Belgrano & Fowler 2013, Stokes & Law 2000, Andersen *et al.* 2009, Eikeset *et al.* 2016). Habitat fragmentation also imposes novel selective forces, including selection for reduced dispersal (Baskett *et al*. 2007, Travis & Dytham 1999, Dytham 2003, Parvinen 2004, Stockwell *et al*. 2003). Strong selection reduces N_e, increasing the variance in reproductive success among individuals, thereby causing effects similar to a genetic bottleneck (Hare *et al*. 2011, Kuparinen *et al*. 2016). In addition, however, positive selection on loci underlying beneficial traits can lead to rapid allele frequency changes at particular loci, which appear as “outlier loci” with high genetic divergence through time (Foll & Gaggiotti 2006, Therkildsen *et al*. 2013b, Foll *et al*. 2015).

If they have occurred, these genetic changes have demographic and ecological consequences as well. Reduced heterozygosity leads to inbreeding depression, while the loss of alleles can reduce the evolutionary capacity of a population to cope with environmental change (Willi *et al*. 2006). Evolution in phenotypic traits can leave populations mal-adapted when fishing and other stresses are reduced (Jørgensen *et al*. 2007).

Our approach is designed to examine whether and to what extent these genetic changes have occurred via a direct comparison between historical genetic diversity from *Albatross* collections (1907-1909) and contemporary collections obtained during the course of this study (2018-2020). This approach avoids problems with previous indirect approaches to detecting change (e.g., Pinsky & Palumbi 2014), while expanding our understanding of genetic change into a severely understudied group of taxa (nearshore tropical fishes, including coral reef fishes) in an understudied region of the world (Southeast Asia). Given that marine fish generation times usually range from a few years to a decade, this time-span represents a dozen to even a few dozen generations (Pinsky & Palumbi 2014, Hutchings *et al.* 2012), which increases our statistical power to detect genetic changes. We will organize our research on temporal genetic patterns (Central Question 1) into a series of six specific questions:

<u>Q1.1. How much have allele frequencies, heterozygosity, and allelic richness changed over time?</u>

Answering this question will involve a direct comparison of within-population allele frequencies, heterozygosity, and allele richness from *Albatross* and modern samples. The magnitude of allele frequency change provides a direct measure of the strength of genetic drift, with the variance in allele frequency change from one generation to the next increasing with smaller population size (N_e) according

to $V_q = \frac{pq}{2N_e}$, where *p* and *q* are initial allele frequencies. We will measure both expected and observed heterozygosity to detect changes. Allelic richness will be calculated for each RAD locus (~300 bp), since these loci typically have 1-2 SNPs and therefore 2-4 haplotypes at each. We will rarefy the calculations of allelic richness back to an expected number of alleles at the smallest sample size to prevent differences in sampling effort from biasing our results (Leberg 2002). Our ***hypothesis*** is that we will find substantial allele frequency changes, a small reduction in heterozygosity, and a larger reduction in allelic richness. This would be the first demonstration of these effects in coral reef fishes, to our knowledge.

Past studies testing for temporal losses of genetic diversity have found mixed results, with some studies reporting strong declines (Smith *et al.* 1991, Hauser *et al*. 2002, Hutchinson *et al*. 2003) and others reporting none (Ruzzante *et al.* 2001, Therkildsen *et al*. 2010, Jakobsdottir *et al*. 2011). However, a power analysis revealed that most previous tests had low statistical power to detect a change, given the relatively few loci sequenced (Pinsky & Palumbi 2014). A simulation study revealed that at least 20 microsatellite loci were required to detect a statistically significant decline in heterozygosity or allelic richness (i.e., 80% statistical power), approximately equal to 200 SNPs (Hoban *et al*. 2014, Pinsky & Palumbi 2014). In contrast, we will have 5000 loci containing 1-2 SNPs each through RAD sequencing (see section on *Genotyping*), greatly enhancing our power to detect true reductions in diversity over time.

Q1.2. Have effective population sizes (N_e) changed over time?

As discussed above, stronger genetic drift and loss of genetic diversity may be caused by a reduction in N_e, either through fishing or through habitat loss. We will test for this explicitly by estimating N_e in the *Albatross* specimens and in the modern samples using single-sample methods. Linkage disequilibrium among physically unlinked loci is higher in smaller populations, providing a powerful tool for estimating N_e (Waples & Do 2008). We will apply this method as implemented in the NeEstimator V2.01 software package (Do *et al.* 2014). For comparison, we will also use the degree of allele frequency change to independently estimate average effective population size across the century using temporal moment-based and maximum-likelihood methods (Waples 1989, Wang & Whitlock 2003).

Our ***hypothesis*** is that N_e will be lower in the modern samples than in the *Albatross* samples. While N_e estimates for marine species with large population sizes can have wide confidence intervals, a strong advantage of our study design is that we will have 20 species to compare. Consistently lower N_e across many of the species will provide strong evidence of a decline. We will use resampling-based statistical tests to account for the non-normal confidence intervals of N_e estimates.

All estimates of N_e are somewhat sensitive to gene flow from surrounding populations, although linkage disequilibrium-based methods accurately estimate local N_e unless gene flow is relatively high (> 5-10%) (Neel *et al*. 2013, Waples & England 2011). With high gene flow, these methods estimate a metapopulation-wide N_e. In the high gene flow case, a reduction in N_e could be caused by habitat fragmentation and an interruption of gene flow. Whether caused by lower population size or reduced dispersal, lower estimates of N_e indicate stronger drift, which is our primary focus.

Q1.3. Are there signatures of a genetic bottleneck?

Recent declines in N_e leave characteristic signatures of a genetic bottleneck, though the large population sizes of marine fishes can limit these effects (Pinsky & Palumbi 2014). To complement our analyses of heterozygosity, allelic richness, and N_e through time, we will also test specifically for bottleneck signatures in our modern samples. We will apply the same analyses to the *Albatross* samples to test whether these bottleneck signatures are new, and not indicative of events before the 1900s. Signatures include the characteristic 'tick mark' pattern in the Site Frequency Spectrum (SFS), positive Tajima's *D*, and heterozygosity excess (Gattepaille *et al*. 2013, Nunziata *et al*. 2017, Luikart & Cornuet 1998). Our ***hypothesis*** is that these signatures will be present in the modern samples, but will not be present, or will be present at much lower rates, in the *Albatross* samples (as based on resampling statistical tests).

In addition, we will use the SFS to estimate N_e change by maximum likelihood fitting of a population dynamics model. We will use fastsimcoal2 to fit models with pre-1900s change in population size, post-1900s change, or both. This approach successfully estimated recent N_e changes over 19 generations in two species of salamander (Nunziata *et al*. 2017).

Q1.4. Are there signatures of adaptation to a century of overfishing and habitat destruction?

Our ***hypothesis*** is that drastic changes in selective pressures driven by human activities will affect the genetic composition of the investigated species. While both population declines and intense selection can reduce N_e and have effects like a genetic bottleneck, intense directional selection also favors particular alleles at particular loci and will tend to drive them to higher frequency. Detecting these signatures of population-level adaptation often involves complex tests to infer historical dynamics from the frequency of variants, linkage disequilibrium across nearby loci, or differentiation among populations in modern samples (Vitti *et al*. 2013). In contrast, sampling from the same population through time provides a direct method for observing evolution in action. Changes in allele frequency at a limited set of loci that exceed the rates expected due to drift provide a clear genetic signature of selection, an approach that has helped uncover adaptation to contemporary anthropogenic activities in Atlantic cod and *Daphnia*, for example (Orsini *et al*. 2012, Therkildsen *et al*. 2013b).

Our RAD sequencing approach will allow us to interrogate 5000 loci for evidence of adaptation to selective pressures over the past century. To identify outlier loci potentially affected by changes in selection, we will use an Approximate Bayesian Computation (ABC) approach to infer the selection coefficient (*s*) consistent with the observed frequency change at each locus, as implemented in WFABC (Foll *et al*. 2015). In this case, *s* refers to the relative fitness difference among homozygotes (i.e., relative fitnesses 1 for AA, 1+s/2 for AB, and 1+s for BB). The method works by generating a large set of forward Wright-Fisher simulations that match the N_e and number of generations elapsed for the focal species, but where *s* is allowed to vary. The simulations matching the observations (degree of allele frequency change) most closely are retained and analyzed in the ABC framework to produce a posterior estimate of *s* (Bertorelle *et al*. 2010). The test can also be structured in an hypothesis testing framework by comparing the observations to the distribution of allele frequency change expected when there is no selection ($s = 0$).

Genome scan approaches with a few hundred or thousand loci (as in our data) have been criticized for scanning over relatively small fractions of the genome, particularly in species with relatively little linkage disequilibrium (Hoban *et al*. 2016, Lowry *et al*. 2016). On the other hand, RAD and other methods have also proved powerful for detecting selection across a wide range of species, including eels, salmon, and dolphins (Laporte *et al*. 2016, Cammen *et al*. 2015, Brieuc *et al*. 2015). We therefore will apply genome scans in our species with the realization that we are likely to miss some loci under selection, but that this does not negate any results if we do find genetic signatures of adaptive changes over the last century.

Q1.5. Are genetic changes congruent across species?

One of the most powerful aspects of our study design is the replication across 20 species from the same region. Even if signals of genetic diversity loss are weak, consistent signals of diversity decline across most species will provide convincing evidence that a decline has occurred. The same logic applies to signals of reduced N_e or of a genetic bottleneck. For example, the consistency of genetic diversity decline across 12 species groups in the meta-analysis of Pinsky & Palumbi (2014) provided more confidence that diversity loss was a common consequence of overfishing. Our ***hypothesis*** is that diversity loss, bottlenecks, and reduced N_e will be common across the species we will sample.

The Fisher method for combining *p*-values provides the simplest approach to integrating evidence across all species. Taking each of *k* species as an independent test of the hypothesis, a combined *p*-value can be computed from $\chi^2_{2k} = -2\sum_{i=1}^{k}\ln(p_i)$, where χ^2_{2k}follows a chi-squared distribution with 2*k* degrees of freedom. In addition, we will also use more sophisticated tests that allow for variance in demographic history and parameters across species. We will use hierarchical ABC methods (Ilves *et al*. 2010, Prates *et al*. 2016) to test a null model of no population decline against an alternative model of population decline. Hierarchical ABC (hABC) fits these models simultaneously for all species by drawing parameters for each species from a hyper-distribution. We will test allelic richness, Tajima's D, heterozygosity, and the SFS as summary parameters for the hABC methods. The posterior distribution of most interest from hABC will be the relative support for each of the hypotheses (decline or not).

Q1.6. Are there identifiable characteristics of fish populations more or less prone to genetic changes?

We expect common signals of genetic change across species. We also expect the magnitude of change to vary among species. Therefore, our final analysis will be to test whether the degree of genetic change can be explained by characteristics of the species. Specifically, we will test whether diversity loss differs between species affected by overfishing versus those affected by habitat loss. We also hypothesize that diversity loss will be greater in species with larger asymptotic body sizes and shorter generation times. Large species tend to have lower carrying capacities (i.e., pre-fishing population sizes) and therefore more susceptible to diversity loss upon further declines (McCusker & Bentzen 2010). Species with shorter generation times will have a larger number of elapsed generations and therefore more time for drift to act.

Gene Flow

Q2. Do overfishing and habitat loss reduce gene flow between marine populations?

Central Question 2 examines the consequences of the expected decreases in population size and genetic diversity on connectivity over time between different locations in the Philippines. Our ***hypothesis*** is that contemporary populations which have been subjected to a century of overexploitation and habitat degradation will have consistently lower connectivity estimates, and that decreased connectivity will be evident across all species. Overexploitation and habitat degradation are expected to reduce population sizes, which in turn decrease larval output, increase population fragmentation and genetic drift, and consequently produce stronger genetic structure between populations. Assessing the decrease in connectivity in both a spatial and temporal context in this study will provide insight into the evolutionary consequences of these factors. Centennial comparisons are possible using the *Albatross* and contemporary samples since at least ~20 specimens are needed to compare meaningful changes in gene flow between localities (e.g. Ackiss *et al*. 2013, Raynal *et al*. 2013), and at least five candidate species in the *Albatross* collection have adequate sample numbers from 2 to 7 disparate locations from which we can also collect modern samples during the course of this project. We will also try to increase beyond five species, as samples and genotyping allow. Population structure and demographic connectivity from disparate locations have already been demonstrated in modern marine populations in the Philippines (Carpenter *et al.* 2011, DeBoer *et al.* 2014a, 2014b; Raynal *et al.* 2014, Hanson 2015, Stockwell *et al*. 2016). Pairing historical with modern collections provides a unique opportunity to explicitly investigate genetic change and the evolutionary consequences of anthropogenic pressures (Nielsen & Hansen 2008, Therkildsen *et al*. 2013b, Habel *et al.* 2014).

The ability to subsample the genome using RAD sequencing increases the statistical power for distinguishing population units and gene flow related to connectivity (Funk *et al*. 2012). Summary statistics of within- and among-population diversity will be calculated for each population including F_{IS} and F_{ST} statistics (pairwise and overall). We will use the Bayesian clustering program Structure (Pritchard *et al*. 2000), as well as the individual-based principal component analysis with the R package ADEGENET (Jombart 2008) to identify unique genetic groups within each species. Tests for overall differentiation will be conducted using a hierarchical Analysis of MOlecular VAriance (Excoffier & Lischer 2010) with time points nested within populations. Coalescent analyses in Migrate-N will be used to test for directionality of gene flow (Beerli & Palczewski 2010) to determine if there have been changes over time as a result of human factors. We will also extend within-species examinations of population bottleneck signatures by fitting multi-population demographic models using fastsimcoal2 (Excoffier *et al.* 2013); and use BEAST to make extended Bayesian skyline plots to compare population structure that may have been in place long before the time of the *Albatross* samples (Bouckaert *et al*. 2014). Once data have been collected across multiple species, we will run meta-analyses of genetic signatures and connectivity to determine if there are consistent patterns across space and time.

Changes in Species Diversity Over Time

Q3. To what extent do genetic diversity and species diversity reveal similar changes over a century of human impacts?

In the Philippines, the consequences of both a peak in species diversity (Carpenter & Springer 2005, Allen 2008, Sanciangco *et al.* 2013) and a peak in threat to this diversity (Bryant *et al.* 1998, Roberts *et al*. 2002), particularly in the central Visayan region (Nañola *et al*. 2011, Lavides *et al.* 2016), can be directly measured because of another set of Smithsonian natural history collections that are particularly suited to measuring biodiversity changes over time. Collections completed in 1978 and 1979 in the central Visayan region of the Philippines resulted in 4,541 catalogued museum lots containing 37,392 fish specimens and included over 40 ichthyocide Philippine stations (Figure 5). PI Carpenter and Collaborator Alcala participated in these collections and are intimately familiar with their protocols and methodologies.

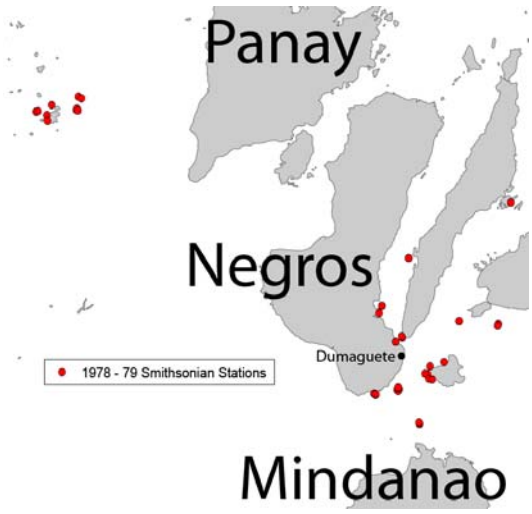


Figure 5. The central region of the Philippines known as the Visayas, shown with 1970s era Smithsonian collection sites.

Our goal will be to replicate more than half of these specific collection stations to address the question: what species diversity changes have occurred after four decades of habitat degradation and exploitation on coral reefs of the Philippines and how does this compare to changes in genetic diversity from comparisons of the *Albatross* material? Our expected outcome will be a direct measurement of changes in species composition and community structure with time, which will provide a corollary to the genetic investigations above. Unfortunately, the '78-'79 collections were preserved in formalin, prohibiting effective genetic analysis of these specimens. However, these two components of the project will allow a direct comparison of genetic and biodiversity changes over time.

Decadal comparisons of species changes will be carried out by resampling over 20 sites using the same methods that were used in the 1970's. The wind and current conditions and concentration of ichthyocide that prevailed at the time of the early samples will not exactly match resampling conditions and areas sampled during the course of this study. However, we are assuming that weather conditions will be more-or-less random by choosing similar times of the year for resampling and that the large number of resampling events at the same depths and types of habitats will be sufficient to overcome this random factor. We are also assuming that differences in collection methods will be minimal as PI Carpenter and collaborator Alcala took part in the 1978-79 collections, are confident that approximate depths and habitat types can be repeated, and PI Carpenter has recent experience in this methodology, having extensively replicated these ichthyocide methods in the Philippines as part of the on-going NSF grant DEB-1257632.

Nevertheless, the data will be treated as non-random, and nonparametric methods will be applied using a biodiversity assessment approach (e.g. Williams *et al*. 2010, Zuur 2007) comparing previous versus present day ichthyocide collections. All samples collected during this part of the study will be transported to the Smithsonian Natural History Museum and curated in compliance with Philippine laws (Williams & Carpenter 2015, Carpenter *et al*. 2017). Rarefaction and extrapolation analytical methods will be employed that have proven robust in comparisons of biodiversity using large sample sizes, while assuming data is non-standardized in terms of collection methods and area sampled (Guralnick & Van Cleve 2005, Beck & Kitching 2007, Biesmeijer *et al*. 2006, Colwell *et al*. 2012, Carvalheiro *et al*. 2013). Statistical treatments will be equivalent to comparisons from transect data that are currently used extensively (e.g. Nañola *et al*. 2011, Abesamis *et al.* 2016). Multivariate analyses on the changes in fish community structure will be carried out using PRIMER (Clarke & Gorley 2015). Within this package, similarity matrices will be generated to test for statistically significant clustering using SIMPROF (similarity profiles), and Multi-Dimensional Scaling (MDS) will be employed to highlight geographical patterns.

Additionally, the SIMPER (similarity percentage) routine, which uses dissimilarity matrices, will be used to identify species that contribute the most to the differentiation between the clusters. These analysis will be conducted on both the 1978-79 data set and the planned data set to compare differences in both spatial and temporal patterns in species communities.

Research Significance

We currently lack even a basic understanding of the genetic consequences of intense human impacts in tropical marine ecosystems. Compared to temperate regions, cumulative human impacts are highest in the tropics and are also increasing at a greater rate (Halpern *et al.* 2015). Biodiversity is much higher in the tropics, with an epicenter in the Coral Triangle region and a peak within this epicenter in the Philippines (Sanciangco *et al.* 2011). Our current knowledge of historical genetic changes of warm-water fishes is limited to a single subtropical species (Longenecker *et al.* 2015) whose prehistoric-scale DNA data is limited to a 180 bp segment of a single mitochondrial gene. Over-exploitation is, by far, the greatest threat to marine fishes (McClenachan *et al*. 2011), even though marine fisheries did not become generally over-exploited until well into the second half of the 20th century (Caddy & Cochrane 2001). The past 100 years has been characterized by a dramatic increase in human population growth and an intense accumulation of human impacts in the oceans (Halpern *et al.* 2015). The *Albatross* material represents a veritable gold-mine of potential for unlocking insights into the ecological and evolutionary changes in tropical marine fishes over this most intense period of human impacts. The potential of this collection for uncovering population-level genetic changes can now be met given the advanced genomic methods available and our preliminary data demonstrating the high quality of the preserved DNA (similar to contemporary studies conducted by the PI and coPIs). The wealth of the *Albatross* and other Philippine collections housed at the Smithsonian National Museum of Natural History and the rich potential for research in present-day Philippines together provide an exciting opportunity for science.

1c. Plan for Educational Activities

This project will integrate research activities with field, laboratory, and classroom education. The research and education activities and goals are highly interdependent because of the need to complete research in the Philippines by students and early career professionals. This need stems from international agreements (i.e. Convention on Biological Diversity) that call into question transfer of raw genetic material between countries and represents permitting challenges in many countries. The PI, Co-PIs and Senior Personnel on this project have a proven track record of gaining the needed permits for collecting and for importing and exporting both raw and extracted genetic material between the U.S. and the Philippines. However, the optimal model to ensure no questions in this regard is for all the DNA extractions to be done within the country where the samples originate. This model was followed by a highly successful previous PIRE project (OISE-0730256, which had an entirely different research focus and set of collaborating institutions than the present proposed PIRE project) led by the PI Carpenter. It relies on the long-term residence of U.S. students and post-docs in the Philippines to complete initial genetic laboratory work alongside Filipino collaborators. This will fulfil one of the project's primary education goals: to provide an international research opportunity for U.S. students that promotes long-term research and education collaboration by working alongside Filipino students and scientists.

The summer semester of each year will be dedicated to field and educational activities in the Philippines. Early career professionals, graduate students, and undergraduates will spend one or more semesters per year in the Philippines completing all the field and initial laboratory work required. This will be coupled with the teaching of advanced genomic courses by post-docs and senior scientists in a classroom experience offered to both U.S. and Filipino students, and coordinated with the laboratory and field work of the research objectives. Based on our research plan and past experience with these types of projects, we anticipate that five years of field studies and analysis workshops will provide in-depth research experiences to approximately 20 U.S. REUs, 20 U.S graduate students, 6 U.S. post-docs and early career professionals, and 10 U.S. senior scientists. Also based on previous experience, we expect these overall numbers will be exceeded in terms of participation of Filipino graduate students and scientists.

The Philippines provides unique opportunities for U.S. students because of both the historical ties with the U.S. and because of its extreme concentration of marine biodiversity. English is widely spoken in the

Philippines and it is well known among diplomatic circles that Americans are genuinely liked among the vast majority of Filipinos! This provides a student-friendly introduction to collaboration in a very foreign setting and helps ensure success of the long-term residence of U.S. students in the Philippines.

Science in general and biology in particular is best learned through hands-on laboratory and field experiences. Students in residence in the Philippines will undertake the complete cycle of learning by participating in field experiences to collect specimens for genetic analysis in the extreme biodiversity settings unique to the Philippines, and then by following through with initial laboratory analysis and subsequent bioinformatic processing. The bioinformatics component will include courses taught every summer semester in the Philippines that use the advanced genomic data to teach the elements of high-powered computer access and use, assembly and SNP calling, SNP filtering, and population genetic analyses. The PIs and Senior Personnel on this project have held six successful advanced genomic workshops and trainings using remote access to high powered computing in Vietnam and the Philippines (OISE 1206614, PEER projects 177, 2-7, 3-100). These Philippine summer semester experiences will be coupled with relevant courses taken for college credit while completing theses and dissertation research requirements.

The proven track record of all PIs in peer-reviewed publication success that includes students as authors will promote excellence in both science and education. This proven track record also includes close collaboration with international scientists from a previous PIRE project (OISE-0730256), including through the publication stage. This emphasis will provide an important learning experience for the U.S. students involved about the value of international collaboration. Their learning experience will take place alongside Filipino students and scientists who will also be trained in advanced genomic methodology and the use of data collected for their own research projects. The PI's previous NSF research experience in the Philippines has resulted in many Filipino student led research results (Casilagan *et al.* 2013, Maralit *et al.* 2013, Thomas *et al.* 2014, Willette & Padin 2014).

The education component of this proposed project will emphasize opportunities for groups underrepresented in the marine sciences. Special attention will be taken to encourage application of these groups in the recruitment and hiring of graduate students and post-docs at all institutions for work on this project. In addition, a REU summer semester will be offered to one undergraduate student (REU) from each of the U.S. institutions (ODU, ASU, RU, TAMUCC) for each of the active research years of the project, targeting recruitment specifically of minorities. ODU is a highly diverse campus, with 55% female and 32% underrepresented minority populations (Fall 2015). ASU-West Campus is especially focused towards serving first generation college students and minority students, through its well-established Trio-STEM Program. As of Fall of 2015, undergraduate students in STEM majors at ASU-West were 53% female and 49% from underrepresented minority groups. Additionally, 84% were Arizona residents, and 60% first generation college students. RU is more than 50% non-Caucasian and actively recruits students from underrepresented groups, including through the Rutgers Future Scholars program aimed at first-generation college students. TAMUCC is a Hispanic Serving Institution: 48% Hispanic, 61% non-Caucasian, 59% women, and 40% low income. There are several active programs at TAMUCC for under-represented groups in science, including LSAMP, AGEP, McNair Scholars, Trio-STEM, and S-STEM. Further, Co-PI Bird is the faculty mentor for the TAMUCC Society for the Advancement of Chicanos, Hispanics, and Native Americans in Science (SACNAS) chapter.

Each year each of the participating academic institutions will conduct a campus competition, with special encouragement of underrepresented groups in STEM, to recruit an REU to attend the summer research and education experience in the Philippines. Prior to departure, REUs, GRAs, and post-docs will receive basic Visayan language (spoken in the Silliman University province) and cultural training hosted online with facilitation from the ODU Filipino-American Cultural Center. The PIRE Project Administrator (see below) will ensure that all participants in the yearly research and education activities in the Philippines will have the proper insurance, medical clearances, medical preparation, and travel arrangements prior to departure.

1d. *Broader Impacts*

Philippine Shore Fishes Red List Assessments. In order to frame both the genetic and species level changes that we will directly measure over time with respect to changes in the overall marine biodiversity of the Philippines, we also propose to collaborate with current, on-going efforts in the Philippines to conduct a National Philippine Red List for marine species, based on assessment of the population and conservation status of all marine fishes present in Philippine waters. Essentially, data on genetic and species level changes from the proposed research will provide critical data to assess marine species population changes over time based on established International Union for the Conservation of Nature (IUCN) Red List assessment methodology (IUCN 2001).

An exhaustive evaluation of extinction risk theory is the basis for the categories and criteria of the IUCN Red List of Threatened Species (Mace *et al*. 2008, www.iucnredlist.org). Hundreds of biologists and conservation scientists have applied these criteria to infer extinction risk of species (e.g. Butchart *et al*. 2010, Hoffman *et al*. 2011, and citations therein). During the IUCN Red List assessment process, data produced from this proposal will be integrated into individual marine species assessments, which are reviewed and assessed for level of extinction risk in collaborative, expert workshops. In IUCN Red List methodology, prioritization of species for conservation can be based in part on the evolutionary or phylogenetically distinctness of species (Forest *et al*. 2015), in combination with population, or other surrogate measures of decline, correlated to or inferred from exposure to threats (IUCN 2001, 2014). Similarly, a number of phylogeny-based indices can be used to examine the historical dimensions of the evolutionary process that are responsible for the present-day patterns of biodiversity, particularly with regard to evolutionary patterns of extinction risk (Srivastava *et al*. 2012, Faith 2015). Thus, species determined to be at elevated risk of extinction during the Red List Assessment process can be plotted on available and enhanced phylogenies of all known Philippine marine fishes (Betancur *et al.* 2013, Near *et al*. 2014, Betancur *et al* 2015) to examine the evolutionary dimensions of extinction risk and provide another frame of reference for the research questions of this proposal.

Completion of a Philippine National Red List, based in part on data produced from this proposal, will essentially serve as a Philippine equivalent of the US Endangered Species Act and as such, will transform conservation and fisheries management in the Philippines. The Philippine government has committed significant resources to a Philippine National Aquatic & Marine Red List, and a key researcher (Santos) on this proposal is in charge of coordinating this effort for marine fishes of the Philippines. Additionally, two other key researchers in this proposed research (Polidoro, Carpenter) have extensive experience in the application of IUCN Red List criteria (IUCN 2001, 2014), and together have produced global IUCN Red assessments for more than 12,000 marine species to date (Adams *et al*. 2013, Polidoro *et al*. 2012, Collette *et al*. 2011, Comeros-Raynal *et al*. 2012, Dulvy *et al*. 2014, Knapp *et al*. 2011, Sadovy de Micheson *et al*. 2012, among others) through the lead PI's (Carpenter's) management of the IUCN Marine Biodiversity Unit. **Generation length, a key element in Red List Assessments and useful in interpreting genetic results** (Fung & Waples 2017)**, will be compiled during the assessment process for each species used in genetic analyses for Research Questions 1 and 2.**

The long standing, close collaborations of the co-PIs on this proposal will ensure the use of the data on genetic and species level changes generated by the proposed research to directly inform the Philippine National Red List Assessments, which will ultimately generate both national and global recommendations for marine conservation and fisheries management. Final global Red List Assessments for each species, containing all Philippine information, will be published on the online IUCN Red List of Threatened Species, and regional Philippine assessments will be published online under the IUCN Red List regional initiatives.

Resource Management Impacts. The age of advanced genomics has resulted in unprecedented use of population genetics in fisheries management and conservation science (Allendorf 2017). The application of these methods to at least 20 species not yet analyzed in the Philippines will provide many insights for fisheries management. The participation of Senior Personnel Santos from NFRDI of the Philippine Bureau of Fisheries and Aquatic Resources will provide many opportunities to expand and apply the results of these studies to resource management. In addition to outreach activities outlined below, the final synthesis workshop of this project will target applications of the genetic material for resource

management with the participation of academic, government, and non-government institutions in the Philippines.

Outreach and Educational Impacts. Each summer, academic and research sessions will be accompanied by outreach that will ensure annual exposure of this study to popular audiences, local and federal governments, non-governmental organizations, and academic institutions. This will be accomplished by regular seminars for both academic and popular audiences, and resource and conservation institutions. Prior Informed Consent documents need to be obtained from each municipality where wild collections occur and an informational seminar and roundtable discussion will take place during each of the municipal visits required to obtain these permits. Regular press releases will accompany field activities and scientific publications, with particular attention to the potential for international media exposure that we expect to follow from discoveries stemming from centennial comparisons of *Albatross* era and contemporary populations.

Products from this research will be used for broader educational purposes and be adapted for use by wider audiences. For example, each postdoctoral researcher will be responsible for writing a popular science article (e.g., Jones 2013) focusing on the results from different research objectives. We will also work with the Data Nuggets project, an NSF Beacon Evolution in Action Initiative focused on developing K-12 worksheets and short analytical activities utilizing actual research data for students to learn skills in interpreting quantitative information and making claims based on evidence. As part of the outreach effort for this project, participating students will develop a Data Nugget based on the data from each of their respective projects. These will be submitted to the central Data Nugget repository as well as shared with local K-12 educators associated with the various host institutions.

Additional K-12 outreach will be conducted via development of multiple cross-coast/international Google Hangouts between a local Norfolk area high school and high-school classrooms in the Philippines. As very different coastal communities facing the global threats of climate change and marine resource overexploitation, these interactions will enable the exchange of ideas about coastal ecology and local threats and broaden the exposure of both communities to similar resource management issues in very different ecosystems.

Lastly, all data collected for IUCN Red List Assessments will be posted on global and national sites on the publicly available web-based IUCN Red List for use by numerous students and researchers globally.