Effect of Species Distribution on Demography Inference in Philippine Fishes

Eryka Molino, BS

Dr. Eric Garcia, PhD

Nichole Leach, MS

June 25, 2022

Abstract

The Philippines is the center of marine biodiversity of the world and it is also the center of adversity for marine systems. Using shotgun sequencing and a Pairwise Sequential Markovian Coalescence (PSMC) program, this project investigates the relationship between species range distribution and population history of marine reef fishes in the Philippines. First, shotgun sequencing was used to generate data for assembling the genome of the Spotted Gill Cardinalfish (*Ostorhinchus chrysopomus*) which is distributed from the Indo-Malay region to New Guinea and the Solomon Islands. This is considered to be shortly distributed. Second, the PSMC program was used to infer population size in the past for a total of 19 other fish species with existing shotgun data: 14 with long distributions (Africa to the West Pacific region), and 5 with short distributions (West Pacific region). It is hypothesized that fish species with larger distributions will have an average larger effective population size across time. This research will consolidate the relationship between current distribution and historical population changes, in order to make predictions on whether species with short or long distributions might be more vulnerable to future anthropogenic and environmental stressors.

Introduction

The Philippines is the epicenter of marine biodiversity, (Carpenter and Springer 2005) but it is not the epicenter of marine biology research (Panga et al.). The Philippine Islands are in the Coral Triangle (Sanciangco et al.). It has distribution data for 2983 fish species (Carpenter and Springer 2005) including endemic, regional and circumglobal species. As an island nation, the Philippines relies on fish for food, which has a high nutritional value due to the high concentration of protein, amino acids, fatty acids, vitamins, and minerals (Wahyuningtyas et al. 2017). Fish is also important due to its use for means of income, with 1.6 million Filipinos employed in fisheries-related occupations (Anticamara and Go 2016; BFAR 2011).

However, overfishing has caused a rapid decline in marine biodiversity in the Philippines(Silvestre et al. 2003); its conservation is critical to the interests of all humans, nations and governments (Dudgeon et al. 2005). Biodiversity is important to resisting environmental changes, and the goal of the PIRE Project is to assess how human interference has affected biodiversity (PIRE 2022). One example of this kind of study is the *Albatross* Philippine Expedition from 1907 to 1909, which assessed the aquatic resources of the Philippines by exploring fish, invertebrate specimens, hydrographic and fish data (Smith and Williams 1999). The current study focuses on distribution and population history and will use contemporary data that will be collected as temporal duplicates.

Distribution is the geographical range a species can survive in. This project will look specifically at latitudinal and longitudinal historical distribution patterns of fish species. There are 2 groups: short distribution and long distribution. There are 5 species

that are labeled as short distribution, which are localized to the West Pacific: Indo-Malay to New Guinea and the Solomon islands. There are 14 species that are labeled as long distribution, which range from Africa to the West Pacific. For the purpose of this study, population history will be viewed through the effective population size (N_E) which is defined as the portion of the overall population that can breed and produce viable offspring.

The goal of this project is to assess how distribution history affects demography inference, such as population history. This will be completed by using the Philippine International Research Experience (PIRE) Shotgun Sequencing Library (SSL) pipeline, and Pairwise Sequential Markovian Coalescence (PSMC) program. In order to do this we will be analyzing shotgun sequences from *Ostorhinchus chrysopomus* to generate a *de novo* genome assembly. This species is used for exploration of the pipeline because it will generate contemporary data only, since there is no current genome sequence data. The PSMC program will be used to infer population history of this species and 19 others up to 10,000 years ago. If the association between the two parameters can be identified, then this can be used to make predictions about future populations of fish species.

Methods: Lab Extraction

One individual per species is selected, and 3 libraries are started from each individual. The DNA was extracted in the lab at Old Dominion University using the Qiagen DNeasy blood tissue kit with the following modification: in the final step, elutions were saved into separate vials. Elution 1 was saved in a different tube from elution 2.

Each elution for each specimen was subjected to gel electrophoresis and fluorescent quantification using the biotium accu clear kit, a spectramax M3 plate reader. The best elution was identified and working plates were created using an eppendorf EP motion fluidics robot. We transferred 16.67micrograms of DNA for each specimen. The extracts were then subjected to two rounds of paramagnetic bead clean up (omega biotek). This went into the Kappa hyper prep plus kit, which ran at ¼ x for each reaction. Researchers followed the manial to optimize the initial enzymatic digestion of the DNA, targeting approximately 300-500 base pair fragments. The Kappa hyper prep plus kit includes DNA ligation, PCR, using iTrue stubby adaptors (Travis Glen at Georgia). PCR with iTrue primers to uniquely index each library. The indexing uniquely identifies the samples on the sequencer with specific artificial DNA sequences. The PCR adds the rest of the artificial sequence which includes 8 base pairs on each side of DNA, which is called an index.

After PCR, the samples were cleaned and subjected to DNA fluorescent quantification. A fragment length analysis is used using an advanced analytical fragment length analyzer. The samples are then combined in equimolar amounts, and then are subjected to a size selection step on a pippin blue holst field electrophoresis machine. Size selected for 200-600 base pairs. Finally the library is quantified using the Kappa QPCR DNA quantification kit. The samples are set for sequencing using Illumina Novaseq6000 S4 Flow Cell by NovoGene. Researchers target 50 million read pairs per library.

Methods: Data Analyses

Shotgun sequencing is used to build the genome of *O. chrysopomus.* X amount of coverage was completed using a second-generation sequencing program, Illumina. This sequence was then pushed through the PIRE SSL Pipeline.

The Pairwise Sequential Markovian Coalescence program uses the genome from the contemporary species (all 20 studied) to infer population history from the last 10,000 years. This is done through using markers from loci positions and analyzing heterozygosity. A decrease in heterozygosity is associated with a coalescence event, which takes two species and traces it back to its most recent common ancestor (MRCA). This is likely to happen with a decline in population size.

RStudio was used to test the validity of the null hypothesis for this study using statistical analysis. First, the packages tidyverse, janitor, lubridate, ggpubr, and ggplotr were installed using the following lines of code:

```
library(tidyverse)
library(janitor)
library(lubridate)
#install.packages("ggpubr")
#install.packages("qqplotr")
library(ggpubr)
library(qqplotr)
```

The PSMC data was then read in using a tibble and piping (%>%) command. A plot was then created using a ggplot command, specifying the type of graph, x variable, and y variable. The following lines of code are examples to create a pointplot:

```
cars <- tibble(mtcars)
cars %>% ggplot(aes(x=mpg, y=wt)) +
geom_point() +
geom_smooth(method=lm, level=0.95) +
stat_cor(label.y = 2.5) +
stat_regline_equation(label.y=2)
```

In the case of distribution data, a linear regression model was used to fit the quantitative data using an lm command. ANOVA was then implemented to determine the differences between statistically significant groups using the same lm command. The statistical significance was determined by having a p value lower than 0.05, at which point is considered statistically significant, and the null hypothesis is rejected. The following is a command line of code to use linear regression and ANOVA:

```
caratprice<-lm(price~carat,data=dmd)
summary(caratprice)
```

We then test assumptions of normality, equal variance among groups, independence, and linearity for regression using a Q-Q plot:

```
cars %>% ggplot(aes(sample=mpg)) +
  stat_qq_band() +
  stat_qq_line() +
  stat_qq_point()

cars %>% ggplot(aes(sample=wt)) +
  stat_qq_band() +
  stat_qq_line() +
  stat_qq_point()
```

Results: Shotgun Sequencing of *Ostorhinchus chrysopomus*

Below is a diagram example with confidence bands of the SSL pipeline using Illumina:

The following is a portion of the contemporary genome of *O. chrysopomus* that has been generated:

Results: PSMC

- Current population of short distribution fish chart
- Population history of short distribution fish

- Current population of long distribution fish chart
- Population history of long distribution fish
- Include in diagrams confidence band/percent error

Discussion

Conclusion

- Tie to intro
- Future direction/implication of results findings/how can it be applied to other things/why was it important

References (use Zotero for formatting)

Anticamara, Jonathan A., and Kevin TB Go. "Spatio-temporal declines in Philippine fisheries and its implications to coastal municipal fishers' catch and income." Frontiers in Marine Science 3 (2016): 21.

BFAR (2011). Philippine Fisheries Profile 2011.

Chen, Ellson Y, et al. “Ordered Shotgun Sequencing, a Strategy for Integrated Mapping and Sequencing of YAC Clones.” Genomics, Academic Press, 25 May 2002, https://www.sciencedirect.com/science/article/abs/pii/S0888754383713856.

Carpenter, K.E., Springer, V.G. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ Biol Fish 72, 467–480 (2005). https://doi.org/10.1007/s10641-004-3154-4

Dudgeon, David, et al. “Freshwater Biodiversity: Importance, Threats, Status and Conservation Challenges.” Biological Reviews of the Cambridge Philosophical Society, U.S. National Library of Medicine, 12 Dec. 2005, https://pubmed.ncbi.nlm.nih.gov/16336747/.

Fraser, T.H., 2012. A new species of deeper dwelling West Pacific cardinalfish (Percomorpha: Apogonidae) with a redescription of *Ostorhinchus atrogaster*. Zootaxa 3492:77-84. (Ref. 91763)

Kaiser, Olaf, et al. “Whole Genome Shotgun Sequencing Guided by Bioinformatics Pipelines-an Optimized Approach for an Established Technique.” Journal of Biotechnology, Elsevier, 19 Nov. 2003, https://www.sciencedirect.com/science/article/abs/pii/S0168165603002293.

Libretexts. “7.13F: DNA Sequencing Based on Sanger Dideoxynucleotides.” Biology LibreTexts, Libretexts, 3 Jan. 2021, https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Boundless)/7%3A_Microbial_Genetics/7.13%3A_Bioinformatics/7.13F%3A_DNA_Sequencing_Based_on_Sanger_Dideoxynucleotides#:~:text=Sanger%20sequencing%2C%20also%20known%20as,specifically%2C%20dideoxynucleotides%20(ddNTPs).

Liu, Shenglin, and Michael M Hansen. “PSMC (Pairwise Sequentially Markovian Coalescent) Analysis of RAD (Restriction Site Associated DNA) Sequencing Data.” Wiley Online Library, 8 Oct. 2016, https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12606.

Mather, Niklas, et al. “A Practical Introduction to Sequentially Markovian Coalescent Methods for Estimating Demographic History from Genomic Data.” Ecology and Evolution, John Wiley and Sons Inc., 7 Dec. 2019, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6972798/.

"Philippines Pire Project - Old Dominion University." The Philippines PIRE Project, 2022, https://sites.wp.odu.edu/PIRE/philippines/.

Panga, Fleurdeliz M, et al. "Through the Boundaries: Environmental Factors Affecting Reef Benthic Cover in Marine Protected Areas in the Philippines." Frontiers, Frontiers, 1 Jan. 2001, https://www.frontiersin.org/articles/10.3389/fmars.2021.702071/full.

Roach, Jared C., et al. "Pairwise End Sequencing: A Unified Approach to Genomic Mapping and Sequencing." Genomics, Academic Press, 21 Oct. 2002, https://www.sciencedirect.com/science/article/abs/pii/088875439580219C?via%3Dihub.

Reed, Sharon E., et al. "Shotgun Sequencing of the Negative-Sense RNA Genome of the Rhabdovirus Maize Mosaic Virus." Journal of Virological Methods, Elsevier, 6 July 2005, https://www.sciencedirect.com/science/article/abs/pii/S0166093405001667.

Sanciangco, Jonnell C., et al. "Habitat Availability and Heterogeneity and the Indo-Pacific Warm Pool as Predictors of Marine Species Richness in the Tropical Indo-Pacific." PLoS ONE, vol. 8, no. 2, 2013, https://doi.org/10.1371/journal.pone.0056245.

Siegel, Andrew F., et al. "Modeling the Feasibility of Whole Genome Shotgun Sequencing Using a Pairwise End Strategy." Genomics, Academic Press, 25 May 2002, https://www.sciencedirect.com/science/article/abs/pii/S0888754300963031.

Silvestre, G. T., Garces, L. R., Stobutzki, I., Ahmed, M., Santos, R. A. V., Luna, C. Z., et al. (eds.). (2003). "South and South-East Asian coastal fisheries: their status and directions for improved management: conference synopsis and recommendations," in Assessment, Management and Future Directions for Coastal Fisheries in Asian Countries (Penang: WorldFish Center Conference Proceedings), 1–40. Available online at: http://aquaticcommons.org/1282/

Smith, David G, and Jeffrey T Williams. "The Great Albatross Philippine Expedition and Its Fishes ." 1999.

Wahyuningtas, L.A., Nurilmala, M., Sondita, M. F. A., Taurusman, A. A., Sudrajat, A. O., 2017. Nutritional profile of Rabbitfish (Siganus spp.) from the Kepulauan Seribu (Thousand Islands), Jakarta, Indonesia. International Food Resource Journal.