**Effect of Reproductive Strategy on Demography Inference in Philippine Fishes**

Allison Fink

Rutgers University- New Brunswick, Pinsky Lab

Methods Co-Witten by Abby Etheridge, Erycka Molino

Advisors: Dr. Malin Pinsky, Dr. Brendan Ried, Kyra Fitz

**ABSTRACT**

The Philippines is considered one of the greatest biodiversity hotspots on the planet. However, this biodiversity is at risk. The life history of Philippine fishes can be divided into three categories, with each group having differing potential to transport offspring to new or established isolated habitats. Species with eggs that do not travel far from the parent habitat may be more at risk for a loss in genetic diversity within a species. This study uses PSMC to estimate the effective population size (Ne) of seventeen different species. These values were then used to compare species according to reproductive strategy. It was found that, on average, pelagic egg producers have the largest effective populations, brooding species have the smallest effective population, and demersal egg producers fall in-between. However, on a species-by-species basis this rule is not followed and there are additional factors to consider. Overall, reproductive strategy may be used as a metric for consideration in conservation decisions, but not as an only rule.

**INTRODUCTION**

The Indo-Malay-Philippines Archipelago is widely regarded as the most species-rich area of the world. Within this range, central Philippines contains the most biodiversity (Carpenter & Springer 2004). Due to economic, subsistence, and cultural significances, the Philippines is among the highest in fish catch by volume. Filipinos consume a substantial portion of protein via fish, and an estimated 1.6 million Filipinos rely on fish-related income. The demand for fish is expected to rise as the population of the country increases (Anticamara and Go 2016). A combination of the aforementioned high demand, along with relaxed fishing regulations, has led to overharvesting. As of 2009, over half of worldwide marine fish populations were under overharvesting pressure, and the Philippines is no exception. Despite the high fishing activity, the catch volume of Philippine fisheries has not increased over the past 30 years, indicating that fish populations are declining (Anticamara and Go 2016).

Many fishes are most vulnerable during the embryonic development stage, and as such a number of reproductive strategies have developed to optimize the survival of offspring (Duarte and Alcaraz 2004). This study separates Philippine fishes into three categories: pelagic egg producers, demersal egg producers, and brooders. Pelagic eggs are small, buoyant, and produced en masse (Duarte and Alcaraz 2004). After being laid directly into the water column, these eggs are at the mercy of the current (Ochs et al 2015). This results in high mortality, but eggs have the potential to be carried to ideal habitat (**Needs Source**). Demersal eggs, conversely, are large, non-buoyant, adhesive eggs that are produced in smaller quantities. Instead of traveling, these eggs hatch inside the already ideal parent habitat. (Ochs et al 2015, Duarte and Alcaraz 2004). It is not uncommon for demersal eggs to be guarded externally by the parents (**Needs Source**). Brooding is an extreme, specialized form of guarding where eggs are kept on or in the parent’s body. The fecundity and egg size of these species is limited to the mass the parent is capable of carrying, but these species are known for creating a “safe harbor” for eggs that bolsters survival rate (Rüber et al 2004).

Previous research suggests that reproductive strategy influences population structure. The Mandarinfish *Synchiropus splendidus* is a pelagic egg producer native to the Philippines. While abundant, the population tends to be localized. Leung et al (2020) studied population structure at six locations. Population genomic analyses revealed strong connectivity among all localities.

The same can be said for species with life histories that do not disperse offspring across long distances. The spiny chromis *Acanthochromis polyacanthus* is unique among damselfishes in that it produces demersal eggs. Previous observations noted this species had distinct phenotypes between isolated reefs, such as bold markings. The analysis of genetic markers of different populations within the species- regardless of the level of isolation among populations- all were deemed monophyletic (Planes, Doherty, and Bernard 2001). The Banggai Cardinalfish *Pterapogon kauderni* is unique among damselfishes in that it mouthbroods young late into the juvenile phase. This leads to a very constricted range (Bernardi and Vagelli 2003). Bernardi and Vagelli, after sequencing mitochondrial DNA of the species, found evidence that this species has low connectivity and frequent bottlenecking.

While it usually agreed upon within the marine biology community that there is a positive correlation between dispersal ability and connectivity, this is not always the case (Mattos, Seixas and Paiva 2018). Magsino et al (2019) compared the population structure of two rabbitfishes and found that the demersal egg producer had more heterozygosity than the pelagic egg producer. McCusker and Bentzen (2011) found that the population structure of two demersal egg producing wolffish species were comparable to that of unrelated pelagic egg producing marine fish. Despite its life history as a tail brooder and minimal habitat range, the messmate pipefish *Corythoichthys haematopterus* has high connectivity among lineages (Sogabe and Takagi 2013). While the findings of Bernardi and Vagelli (2003) were consistent with the hypothesis that demersal mouthbrooders would have low connectivity and high bottlenecking, the sample size was too small to confidently confirm this.

We used Pairwise Sequentially Markovian Coalescent (Li and Durbin 2011) to estimate ancient effective population sizes (Ne) of seventeen Philippine fish species. This method has been used to infer the historical demography of humans (Li and Durbin 2011), and of fish such as the corkwing wrasse (Mattings Dal et al 2019), and the Big-Eye Mandarin Fish (Lu, Zhao, and Li 2020).

We are assessing the relationships between reproductive strategy and ancient Ne of four pelagic egg producers, eight demersal egg producers, and five brooders. It is hypothesized that species producing pelagic eggs will have the largest historical population sizes, followed by demersal egg producers and then by brooders.

**METHODS**

***Species selection***

A total of seventeen species were selected from a pool of 26 species that had been pre-processed by the PIRE program. Initially, all 26 species were to be included, but nine species were excluded due to a lack of information or data on them. Most species had their reproductive strategy determined using the database [fishbase.se](https://www.fishbase.se/search.php), while others were determined by referencing literature on a species or genus level.

| **Scientific Name** | **Reproductive Strategy** |
| --- | --- |
| *Ambassis kopsii\** | Demersal Egg |
| *Ambassis urotaenia\** | Demersal Egg |
| *Atherinomorus duodecimalis* | Demersal Egg |
| *Corythoichthys haematopterus* | Brooder |
| *Gerres oyena\** | Pelagic Egg |
| *Halichoeres miniatus\** | Pelagic Egg |
| *Hypoatherina temminckii* | Demersal Egg |
| *Lethrinus variegatus* | Pelagic Egg |
| *Ostorhinchus chrysopomus* | Brooder |
| *Periophthalmus argentilineatus* | Demersal Egg |
| *Plotosus lineatus* | Demersal Egg |
| *Siganus spinus* | Demersal Egg |
| *Stethojulis interrupta* | Pelagic Egg |
| *Taeniamia biguttata* | Brooder |
| *Taeniamia kagoshimanus* | Brooder |
| *Taeniamia zosterophora* | Brooder |
| *Taeniamia zosterophora* | Brooder |

Fig 1. List of species included in the study and their corresponding reproductive strategies. \*Reproductive strategy was determined on a genus level.

***Subsampling and DNA extraction***

One individual per species was selected and three libraries were 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 in separate vials. Elution 1 was saved into a different vial than elution 2, and so on. Each sample was subjected to gel electrophoresis and fluorescent quantification using Biodiem AccuClear kit Spectramax M3 plate reader.

The best elution was identified and working plates were created using an Eppendorf EP motion fluidics robot. A quantity of 16.67 micrograms of DNA was transferred for each library/specimen. The extracts went through two rounds of paramagnetic bead cleanup (Omega Bio-Tek) and then were processed through the KAPA HyperPlus kit, which ran at 1/4x for each reaction.

Personnel followed the manual to optimize the enzymatic digestion of the DNA targeting approximately 300-500 base pair fragments. The KAPA HyperPlus kit includes PCR and DNA ligation using iTru stubby adapters (Travis Glenn, University of Georgia). PCR with iTru primers is used to uniquely index each library. Then using fluorescent quantification again and advanced analytic fragment length analyzer, the samples were combined in equimolar amounts and subjected to a size selection (200-600 bp) step on a BluePipen pulse field electrophoresis machine. The library was quantified using a

KAPA qPCR DNA quantification kit and was sent for sequencing at an Illumina NovaSeq 6000 S4 Flow Cell by NovoGene. 50 million read pairs per library were targeted.

***PIRE Shotgun Sequencing Library***

Shotgun sequencing was used to randomly shear DNA into tiny fragments, clone them, and then sequence them creating “reads” that are output to a fq.gz file for downstream analysis. Before assembling the genome, quality control was necessary to ensure that the reads were ready for PSMC.

*Pre-Processing*

Step 1. Quality assessment (FastQC/MultiQC)

I needed to get a baseline assessment of the quality and quantity of our data by using FastQC and MultiQC. They are designed to give a brief assessment of the data and identify any problem areas. FastQC generates reports for each fq.gz file and MultiQC aggregates the information into one summary

file. In order to check the quality of data, we ran an sbatch job by using the code:

| “sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/Multi\_FASTQC.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta/shotgun\_raw\_fq" "fq.gz”. |
| --- |

Step 2. First Trim with fastp

The first trimming removed low quality bases and “bad” reads; those can include reads that are too short, reads that have too many low-quality bases, 3’ ends of reads, and adaptor sequences. I ran an sbatch job to perform the trimming using the code:

| “sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runFASTP\_1st\_trim.sbatch shotgun\_raw\_fq fq\_fp1”. |
| --- |

Step 3. Clumpify (De-duplicating)

At this point, I needed to reduce data size to make the assembly portion more efficient and to minimize effects of sequencing errors. Leftover reads from PCR amplification during the library prep needed to be removed in order to prevent said sequencing errors. Clumpify sorted reads in the fq.gz files into “clumps” by ordering them by similarity and then removed the duplicates. I then ran the sbatch job by running this code:

“bash /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runCLUMPIFY\_r1r2\_array.bash fq\_fp1 fq\_fp1\_clmp /scratch/aethr001 3”

Step 4. Second Trim with fastp2

For the second trim I had to remove the 5’ end of the reads if they have low quality bases. I ran the sbatch script to perform the second trim by using the code:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runFASTP\_2\_ssl.sbatch fq\_fp1\_clmp fq\_fp1\_clmp\_fp2”.

Step 5. Removing any contamination (FastQ Screen)

Contamination, non-target DNA, is a possibility during sample storage, extraction, and library prep. I ensured that contamination is removed as it can interfere with the assembly downstream. FastQ Screen searched our DNA against a database of other genomes to identify and remove extraneous DNA by running this code:

“bash /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runFQSCRN\_6.bash fq\_fp1\_clmp\_fp2 fq\_fp1\_clmp\_fp2\_fqscrn 6”

After checking that there are no errors, I then proceeded to run Multiqc bash script separately using the code:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runMULTIQC.sbatch fq\_fp1\_clmp\_fp2\_fqscrn fastqc\_screen\_report”

Step 6. Repair fastq\_screen paired end files

In this step, it is necessary to repair any read pairs that were separated during the previous steps and is to ensure the R1 and R2 fq.gz files are in the same order making assembly easier.  To run the script, this code was used:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/runREPAIR.sbatch fq\_fp1\_clmp\_fp2\_fqscrn fq\_fp1\_clmp\_fp2\_fqscrn\_repaired 40”

Alongside the repair script, Fastqc-Multiqc must be ran separately using the code:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_fq\_gz\_processing/Multi\_FASTQC.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta/fq\_fp1\_clmp\_fp2\_fqscrn\_repaired" "fq.gz"”

*Genome Assembly*

Step 1. Genome Properties

The sequences at this point were ready to be assembled, however, generating data on the sequences is imperative before assembling the genome. Jellyfish is a script that can create a histogram file that can then be uploaded to GenomeScope v1.0 and GenomeScope v2.0 using this code:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runJellyfish.sbatch "Sin" "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta/fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"”

The values of interest were heterozygosity, genome haploid length, and model fit and should be used to choose either the values derived from GenomeScope v1.0 or GenomeScope v2.0. Heterozygosity should be around 1% and model fit should be greater than 90%; the values for GenomeScope v2.0 aligned better with the preferred 1% and 90% so those values will be used in future steps.

Step 2. Assemble the Genome Using SPAdes

The estimated genome haploid length from GenomeScope v2.0 was then used to assemble the genome using SPAdes for each library and all the libraries combined using these codes:

“#1st library

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runSPADEShimem\_R1R2\_noisolate.sbatch "aethr001" "Sin" "1" "decontam" "675000000" "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"

#2nd library

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runSPADEShimem\_R1R2\_noisolate.sbatch "aethr001" "Sin" "2" "decontam" "675000000" "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"

#3rd library

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runSPADEShimem\_R1R2\_noisolate.sbatch "aethr001" "Sin" "3" "decontam" "675000000" "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"

#all libraries combined

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runSPADEShimem\_R1R2\_noisolate.sbatch "aethr001" "Sin" "all\_3libs" "decontam" "675000000" "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"”

SPAdes automatically runs QUAST which gives statistics on the genome noting number of contigs, size of largest contig, total length of assembly, N50, and L50.

Step 3. Running BUSCO

BUSCO tells how many expected genes were recovered by the assembly and code needs to ran for each library and combined libraries for both contigs and scaffolds using this code:

“#1st library - contigs

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-A\_decontam\_R1R2\_noIsolate" "contigs"

#2nd library -contigs

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-B\_decontam\_R1R2\_noIsolate" "contigs"

#3rd library - contigs

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-C\_decontam\_R1R2\_noIsolate" "contigs"

#all libraries - contigs

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_allLibs\_decontam\_R1R2\_noIsolate" "contigs"

#1st library -scaffolds

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-A\_decontam\_R1R2\_noIsolate" "scaffolds"

#2nd library - scaffolds

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-B\_decontam\_R1R2\_noIsolate" "scaffolds"

#3rd library - scaffolds

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_Sin-CPnd-C\_decontam\_R1R2\_noIsolate" "scaffolds"

#all libraries - scaffolds

sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runBUSCO.sh "/home/e1garcia/shotgun\_PIRE/REUs/2022\_REU/Abby\_Ethridge/stethojulis\_interrupta" "SPAdes\_allLibs\_decontam\_R1R2\_noIsolate" "scaffolds"”

Step 5. Identify Best Assembly

Selecting the best assembly for downstream analysis uses the results from QUAST and BUSCO. The metrics used in the determination is number of contigs, size of largest contig, total length of assembly, N50, L50, and BUSCO score.

Step 6. Assemble Contaminated Data From the Best Library

Finally after selecting the best assembly, assembling the contaminated data from the selected library was done before putting the data through PSMC. It can be assembled with this code:

“sbatch /home/e1garcia/shotgun\_PIRE/pire\_ssl\_data\_processing/scripts/runSPADEShimem\_R1R2\_noisolate.sbatch "e1garcia" "Sfa" "2" "contam" "635000000" "/home/e1garcia/shotgun\_PIRE/2022\_PIRE\_omics\_workshop/salarias\_fasciatus" "fq\_fp1\_clmp\_fp2\_fqscrn\_repaired"”

***PSMC***

After processing through the PIRE SSL pipeline was completed, the samples were processed using PSMC to infer Ne.

PSMC is built upon the concept of sequentially Markovian coalescence and the idea that a single organism’s genome carries a record of its lineage. Coalescence events are defined when two separate lineages within said genome fuse into a single ancestral lineage. Spotting these events requires a “backwards” view where base pairs are traced back to their common ancestors. A large count and frequency of coalescence events present in a genome is indicative of a small population, as smaller populations are more likely to share more recent ancestors, and vice versa (Mather et al 2019).

The Markov models in PSMC account for the amount of heterozygosity present in a genome. Multiple randomly generated strings of 150 base pairs each were processed and assigned a binary marker indicating if a heterozygous locus is present within those 150 base pairs. Higher occurrences of “heterozygous-positive” samples are associated with larger populations, as individuals in larger populations tend to have a larger effective population and gene pool. This leads to a more diverse genome (Mather et al 2019).

PSMC is not without limitations. The process assumes neutral evolution, a set mutation rate, and nonrandom mating, which do not always occur in nature. Contemporary population sizes cannot be estimated as results from more recently than 10,000 years ago are unreliable. Additionally, a single run of PSMC only displays one possibility of demography based on the probability of that result’s coalescence timing and frequency. [**Insert information on bootstrapping**].

This project uses methods outlined by Li and Durbin (2011). The following parameters were used by PSMC to estimate Ne: a manually set scaled mutation rate of 2.25x10^-8, manually determined species-specific generation times, an automatically determined recombination rate, and an automatically determined piecewise constant population size (Li and Durbin 2011). A PERL -language script was used to filter out scaffolds shorter than 20,000 base pairs. A modified version of the dDocent pipeline was used to map remaining shotgun reads to reference genomes. A samtools script calculated the mean depth of the mapped reads. A pipeline script modified from Harvard FAS Informatics tutorial, Applying PSMC to Neandertal data, & the PSMC documentation was used to call a consensus sequence from the mapped bam files. Only scaffolds ranging between one-third and two times the mean depth were used for this process (Li and Durbin 2011). The resulting scaffolds were converted to a PSMC format through an array script. The PSMC script was run applying both the re-formatted files and the species generation time as arguments. The PSMC was then run 100 more times via bootstrapping and combined into one graph by an additional script.

**RESULTS**

***Corythoichthys haematopterus***

BUSCO and QUAST results for all three libraries and the combined library were produced for *Corythoichthys haematopterus*. When analyzed, these results showed that the library “allLibs” containing contaminated scaffolds produced the highest quality genome for C. *haematopterus*.

| **Library/Data Type/ SCAFIG** | **covcutoff** | **No. of contigs** | **Largest contig** | **Total length** |
| --- | --- | --- | --- | --- |
| **allLibs/contam/scaffolds** | off | 87399 | 552058 | 465127131 |
| allLibs/decontam/scaffolds | off | 76635 | 282571 | 378208088 |
| allLibs/decontam/contigs | off | 74819 | 282571 | 364837995 |
| Library#2/decontam/scaffolds | off | 14138 | 269079 | 65369174 |
| Library#2/decontam/contigs | off | 12295 | 176474 | 56786073 |
| Library#1/decontam/scaffolds | off | 9451 | 198527 | 198527 |
| Library#1/decontam/contigs | off | 8516 | 116066 | 45376391 |
| Library#3/decontam/scaffolds | off | 7408 | 182426 | 40835802 |
| Library#3/decontam/contigs | off | 6572 | 153476 | 36962939 |
| **Library/Data Type/ SCAFIG** | **N50** | **L50** | **Ns per 100 kbp** | **BUSCO single copy** |
| allLibs/contam/scaffolds | 5207 | 27175 | 48.94 | 838 |
| allLibs/decontam/scaffolds | 4860 | 26123 | 51.41 | 811 |
| allLibs/decontam/contigs | 4799 | 25780 | 0.00 | 798 |
| Library#2/decontam/scaffolds | 4268 | 4762 | 150.71 | 491 |
| Library#2/decontam/contigs | 4260 | 4145 | 0.00 | 479 |
| Library#1/decontam/scaffolds | 4773 | 2534 | 104.89 | 421 |
| Library#1/decontam/contigs | 4852 | 2278 | 0.00 | 406 |
| Library#3/decontam/scaffolds | 4851 | 1756 | 84.64 | 389 |
| Library#3/decontam/contigs | 5041 | 1498 | 0.00 | 369 |

Fig 1. Table of BUSCO and QUAST results for the genome assembly of *Corythoichthys haematopterus,* ranked by quality. The best assembly is highlighted.

PSMC of C. *haematopterus* was run using 658 scaffolds of the best library. The population had a potential maximum Ne of approximately 250,000 individuals approximately 100,000 years ago. The average line of the bootstrapping results (see Fig 2) reveals that this maximum Ne, while still 100,000 years ago, was more likely about 75,000.

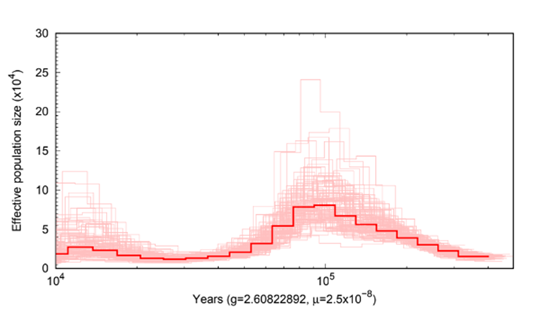


Fig 2. PSMC plot of *Corythoichthys haematopterus* including 100 rounds of bootstrapping. The dark red line shows an average trend.

***Species Trends***

Of all seventeen study species, [insert comments once results are totally completed].

| **Species** | **Best Library** | **# Scaffolds Used in PSMC** |
| --- | --- | --- |
| *Corythoichthys haematopterus* | allLibs/contam/scaffolds | 658 |
| *Ambassis kopsii* | ???/???/??? | 4732 |
| *Ambassis urotaenia* | Library#3/contam/scaffolds | 7427 |
| *Atherinomorus duodecimalis* | Library#3/contam/scaffolds | 6299 |
| *Hypoatherina temminckii* | Library#3/contam/scaffolds | 7427 |
| *Periophthalmus argentilineatus* | Library#1/contam/??? | 6299 |
| *Plotosus lineatus* | allLibs/contam/scaffolds | 3881 |
| *Salarias fasciatus* | allLibs/contam/scaffolds | ??? |
| *Siganus spinus* | ???/???/??? | ??? |
| *Gerres oyena* | Library#2/contam/scaffolds | 8340 |
| *Halichoeres miniatus* | Library#2/decontam/??? | 3777 |
| *Lethrinus* *variegatus* | allLibs/???/scaffolds | 14506 |
| *Stethojulis interrupta* | ???/???/??? | 5110 |
| *Ostorhinchus chrysopomus* | ???/???/??? | ??? |
| *Taeniamia biguttata* | Library#3/contam/scaffolds | ??? |
| *Taeniamia kagoshimanus* | ???/???/??? | ??? |
| *Taeniamia zosterophora* | Library#3/???/scaffolds | 11215 |

Fig 3. Best libraries for all study species

On average, each species used [Number of scaffolds] scaffolds within their best library to run PSMC. [**Insert stuff about overall trends in population size for certain years- did a lot of species have a peak or dip at a certain point in time? LOOK AT ALL COMBINED EPS PLOTS**]

[**Insert samples of different eps plots that show a variation of peaks/dips over time**]

ANOVA for these PSMC results revealed the following p-values: 0.09993 10k years ago, 0.1896 25k years ago, 0.09552 100k years ago, 0.0004582 500k years ago, and 0.0008474 Theta0. The results for 500k years ago and Theta0 are significant.

Overall, pelagic egg producers maintained higher populations than other life histories, followed by demersal and brooder species in order of Ne. Pelagic egg producers were more populous than the other reproductive strategy types in more ancient years. However, there were demersal egg producing and brooding species that were more populous than certain pelagic egg producing species during this time. In ancient years, it appears that there were a number of brooding species that had a higher Ne than demersal egg producing species, and vice versa. In more recent, insignificant years, the trends between the three reproductive types become more blurred and a definite ranking of Ne by egg type is not clear.

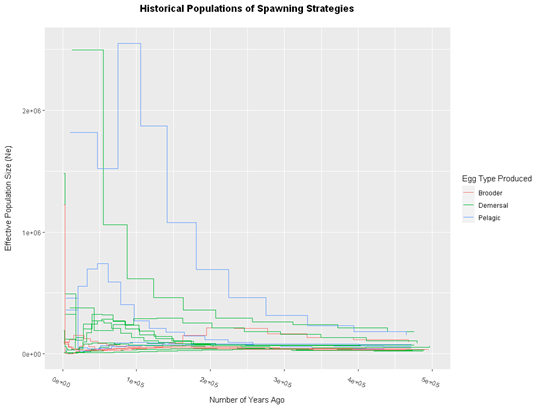


Fig 4. Visualization of PSMC of every study species over time, with grouping by reproductive strategy.

The most populous species within insignificant time periods were a demersal species at 10k and 25k years ago and a pelagic species at 100k years ago. The least populous species for insignificant years was a brooder species at 10k years ago and demersal species at 25k and 100k years ago. The species with a median Ne for insignificant years were a demersal species at 10k years ago, a brooder species at 25k years ago, and a demersal species at 100k years ago.

The most populous species at the significant time period, 500k years ago, was pelagic, followed by a pelagic species with the median Ne for this time, and a demersal species with the smallest Ne 500k years ago. A pelagic species had the highest overall (theta0) population, followed by a pelagic species with the median overall Ne and a demersal species with the overall lowest.

| **10,000 Years Ago** |  |  |
| --- | --- | --- |
| **Species** | **Egg Type** | **Ne** |
| *Hypoatherina temminckii* | Demersal | 2,493,790.428 |
| *Periophthalmus argentilineatus* | Demersal | 1,21,707.293 |
| *Corythoichthys haematopterus* | Brooder | 1,152.230 |
| **25,000 Years Ago** |  |  |
| **Species** | **Egg Type** | **Ne** |
| *Hypoatherina temminckii* | Demersal | 2,493,790.428 |
| *Corythoichthys haematopterus* | Brooder | 61,491.709 |
| *Plotosus lineatus* | Demersal | 9,113.726 |
| **100,000 Years Ago** |  |  |
| **Species** | **Egg Type** | **Ne** |
| *Stethojulis interrupta* | Pelagic | 2,549,283.50 |
| *Atherinomorus duodecimalis* | Demersal | 93,096.40 |
| *Plotosus lineatus* | Demersal | 16,652.62 |
| **500,000 Years Ago** |  |  |
| **Species** | **Egg Type** | **Ne** |
| *Hypoatherina temminckii* | Demersal | 182,579.00 |
| *Lethrinus variegatus* | Pelagic | 57,491.45 |
| *Atherinomorus duodecimalis* | Demersal | 8,569.762 |
| **Theta0** |  |  |
| **Species** | **Egg Type** | **Ne** |
| *Stethojulis interrupta* | Pelagic | 1.224384 |
| *Halichoeres miniatus* | Pelagic | 0.350770 |
| *Plotosus lineatus* | Demersal | 0.040264 |

Fig 4. Table listing the most, median, and least populous species for each time period and Theta0, including egg type and estimated Ne.

For all time periods and Theta 0, the average Ne was the lowest for brooder species, the second lowest for demersal species, and the highest for pelagic species (see Fig 5).

| **10,000 Years Ago** |  |  | **Theta0** |  |
| --- | --- | --- | --- | --- |
| **Egg Type** | **Avg Ne** |  | **Egg Type** | **Avg Ne** |
| Pelagic | 6,66,249.90 |  | Pelagic | 0.7345700 |
| Demersal | 480,423.70 |  | Demersal | 0.4559230 |
| Brooder | 114,838.90 |  | Brooder | 0.2842964 |
| **25,000 Years Ago** |  |
| **Egg Type** | **Avg Ne** |
| Pelagic | 611,094.80 |
| Demersal | 400,016.10 |
| Brooder | 124362.10 |
| **100,000 Years Ago** |  |
| **Egg Type** | **Avg Ne** |
| Pelagic | 746,625.42 |
| Demersal | 195,056.77 |
| Brooder | 77,075.53 |
| **500,000 Years Ago** |  |
| **Egg Type** | **Avg Ne** |
| Pelagic | 80,092.19 |
| Demersal | 64,774.49 |
| Brooder | 63,179.59 |

Fig 5. Average Ne Per Egg Type, Per Year, Listed Highest to Lowest

**DISCUSSION**

***Corythoichthys haematopterus***

The Ne of *Corythoichthys haematopterus* had a noticeable peak approximately 100,000 years ago. [What happened during this time?]. [Talk about how Cha has a larval stage, but this may be better for a different portion of the discussion]. Ne of *Corythoichthys haematopterus* was estimated using 658 scaffolds- this makes it the smallest genome by far, [**percentage**] below the average. It was reported that the sample collected was of low quality.

**Significant Time Period (500,000 Years Ago)**

*Pelagic Egg Producers*

On average, demersal egg producers were listed having the highest Ne compared to pelagic egg producers and brooding species. This aligns with the hypothesis. However, the results are not as consistent on a species level.

500,000 years ago, pelagic egg producers ranked 2nd, 6th,9th, and 12th in Ne compared to other reproductive strategies. Should the hypothesis be followed exactly, the pelagic egg producers should have ranked 1st – 4th in Ne. Only one species, *Stethojulis interrupta*, follows this hypothesis. [**WHY? Why is only one pelagic species in the top 5, and what makes this species so different? What was happening 500,000 years ago that could have affected populations? Why are some pelagics so low- do they have small ranges, or maybe demersal larvae?]**

*Demersal Egg Producers*

On average, demersal egg producers were listed having a median Ne compared to pelagic egg producers and brooding species. This aligns with the hypothesis. However, the results are not as consistent on a species level.

500,000 years ago, demersal egg producers ranked 1st, 5th, 7th, 8th, 11th, 14th,16th, and 17th in Ne compared to other reproductive strategies. Should the hypothesis be followed exactly, the brooding species should rank 12th-5th in Ne. Half of the species in this category follow this assumption, while half do not. [**WHY? What was happening 500,000 years ago that could have affected populations? Why is there such an even split? Why is the population of a demersal egg layer higher than all of the pelagics?**]

*Brooders*

On average, brooders were listed as having the lowest Ne, which aligns with the hypothesis. However, the results were not as consistent on a species level.

500,000 years ago, brooding species ranked 3rd, 4th, 10th, 13th, and 15th in Ne. Should the hypothesis be followed exactly, the brooding species rank 17th – 13th in Ne. While *Taeniamia biguttatar* and *Taeniamia kagoshimanus* are consistent with this assumption, the majority of species are not. [**WHY? What was happening 500,000 years ago that could have affected populations? What quality of the *Taeniamia* genus keeps Ne so small? Why are the 3rd and 4th populations so high (look into the possibility of a pelagic larval stage?)]**

***Theta 0 (Overall population trends)***

While 500,000 years ago was the only time period with statistical significance, the population trend over all time also has statistical significance. Like 500,000 years ago, Theta 0 follows the same ranking pattern of pelagic egg producer, demersal egg producer, and then brooding species (largest Ne to smallest). However, the trends of Theta 0 are not as clear on a species scale.

[**Repeat Above: separate section explaining each egg type, how it met or didn’t meet expectations, and why this may have been. Instead of 500,000 years ago look at long-term issues.]**

**CONCLUSION**

Summary:

The Philippines is considered to contain the highest concentration of marine biodiversity in the world, but this biodiversity is expected to decline due contemporary issues such as overharvesting, climate change, and habitat destruction. The genomes of seventeen species were assembled and then had Ne estimated using PSMC. These species were compared to one another and within groups based on egg type.

Points to make:

Despite not all species following the expected ranking pattern (for example, a demersal having bigger population than a pelagic egg laying species) when averaged out, the pattern of pelagic-demersal-brooder is followed. This makes sense when it’s seen that the p-value for theta 0 (i.e. over all time instead of a specific year) is significant. The only significant year being 500,000 years ago is consistent with the fact that PSMC is most accurate when used to estimate ancient populations- so the other more recent time periods could be more consistent but a different method would have to be used to estimate the contemporary populations [**come up with some suggestions**].

Overall, brooder and demersal species were shown to have more delicate ancient populations than that of pelagic egg producers.

**ACKNOWLEDGEMENTS**

*Special thanks to:*

The Philippines PIRE Project

NSF Award #1743711

The following partners:

Old Dominion University, Rutgers University, Texas A&M University- Corpus Christi,

NOAA Fisheries, Republic of the Philippines Department of Agriculture, Arizona State University,

Siliman University, University of the Philippines- Mindanao, The Smithsonian,

Mindanao State University, and the Bureau of Fisheries and Aquatic Resources

Dr. Brendan Reid, Dr. Malin Pinsky, Dr. Rene Clark, and Kyra Fitz for advising me throughout this project.

Abby Ethridge and Erycka Molino for co-authoring the methods sections

Ivan Lopez, George Bonsall, and Eric Garcia for support while abroad

**WORKS CITED**

al, P. T. Y. L. e. (2020). Population Genetic Structure of A Marine Pelagic Egg Producer and Popular Marine Aquarium Species, the Mandarinfish Synchiropus splendidus [Open Access]. Zoological Studies, 59(68). https://doi.org/10.6620/ZS.2020.59-68

Anticamara, J. A. (2016). Spatio-Temporal Declines in Philippine Fisheries and its Implications to Coastal Municipal Fishers’ Catch and Income. Frontiers in Marine Science, 3(21). https://doi.org/doi: 10.3389/fmars.2016.00021

Bernardi, G. (2003). Population structure in Banggai cardinalfish, Pterapogon kauderni, a coral reef species lacking a pelagic larval phase [Research]. Marine Biology. https://doi.org/10.1007/s00227-004-1355-1

Carpenter, K. E. (2004). The center of the center of marine shore fish biodiversity: the Philippine Islands. Environmental Biology of Fishes, 72, 468-480. Indo-Pacific biogeography, vicariance, island integration, conservation

Magsino, R. M. (2007). The influence of contrasting life history traits and oceanic processes on genetic structuring of rabbitfish populations Siganus argenteus and Siganus fuscescens along the eastern Philippine coasts. Marine Biology, 154, 519-532. https://doi.org/10.1007/s00227-008-0946-7

Mather, N. (2019). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data [Review]. Ecology and Evolution, 10(1), 579-589. https://doi.org/DOI: 10.1002/ece3.5888

Mattos, G. (2019). Comparative phylogeography and genetic connectivity of two crustacean species with contrasting life histories on South Atlantic sandy beaches. Hydrobiologia, 826, 319 - 330. https://link.springer.com/article/10.1007/s10750-018-3744-3#citeas

McCusker, M. R. (2011). Limited population structure in Northern and Spotted Wolffishes (Anarhichas denticulatus and A. minor) despite low apparent dispersal potential. Marine Biology, 158(8). https://doi.org/DOI:10.1007/s00227-011-1698-3

Planes, S. (2001). Strong genetic divergence among populations of a marine fish with limited dispersal, Acanthochromis polyacanthus, within the Great Barrier Reef and the Coral Sea. Evolution, 55(11), 2263-2273. https://doi.org/10.1554/0014-3820(2001)055[2263:SGDAPO]2.0.CO;2