Monterey Bay Aquarium Research Institute

Review and Expansion: Observing *Molidae* through eDNA

Kristina Samborski, MARC U*Star Scholar, University of West Florida Biology Department

Mentors: Francisco Chavez, Nathan Truelove

Summer Intern 2020

Introduction

Environmental DNA (eDNA) analysis has become a widely used method for diversity and migration studies. As potential uses for eDNA continues to grow rapidly, lack of data on hard-to-study animals, (including many marine animals such as *Mola tecta*) can be addressed.

The first record of *Mola mola* was made in 1758 by Carl Linnaeus (*WoRMS - World Register of Marine Species - Mola Mola (Linnaeus, 1758)*, n.d.) *(Linnaeus, 1758)*. Two centuries later, modern scientists have discovered a close cousin, The Hoodwinker Sunfish, *Mola tecta* (Nyegaard et al., 2017)(Yoshita et al., 2009). Due to the very recent species-distinction and the elusiveness of the animal, there is little available genetic information on *Mola tecta*. The first section of this paper attempts to address the only current genetic information on *Mola tecta* and expand upon it using alignment and analysis. D-loop and COI genes taken from the NCBI database were studied and aligned for potential use. The second section of this paper speculates on *Molidae* tracking through environmental DNA (eDNA) and amplicon sequence variant (ASV) analysis. The eDNA data set was taken from cruises (anywhere from 2008-2019) off the coast of Monterey Bay, California.

Methods 1

Twelve COI nucleotide sequences were available for *Mola tecta* on NCBI's Nucleotide Blast database (*Nucleotide BLAST: Search Nucleotide Databases Using a Nucleotide Query*, n.d.). The twelve nucleotides (649 base pairs (bp) each) were uploaded into Geneious Prime and aligned in order to study sequences side-by-side. From this multiple alignment, a consensus

identity sequence was made for the *Mola tecta* COI gene. The consensus is based on the provided sequences (see Figure 1 for visual) and includes variant genes. The variant genes of given COI alignment are noted. Unfortunately, for the purpose of identification and population dynamics, no variants or SNPs were detected. Although there were multiple genetic differences in the sequence ID: OMNZ X2017.18, at a minimum variant frequency of 0.25, the possible variants were not confirmed for this dataset.

```
Consensus      TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
OMNZ X2017.19 TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
OMNZ X2017.18 TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAACTA   60
NZ19          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ18          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ17          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ14          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ12          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ09          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ08          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NZ07          TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
NMNZ P.57679  TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
G06           TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA   60
```

Figure 1. Alignment Visual

Once the alignment was complete and the consensus was made, primers were designed using the Primer3 program. Though COI variants were not confirmed, with more data, SNP identification is possible. Primers were made to contribute to available literature. Primers selected were promising and showed to be without hairpin or dimer possibilities in Geneious Prime. The validity of the primers was checked in the NCBI Primer Designing Tool and the results are shown in Results section, Table 2a with a product of 351 base pairs (bp's).

The same method was used for available D-loop sequences for *Mola tecta*. There was a total of 23 D-loop sequences to align, ranging from 827 to 859 base pairs long. The validity of the primers was checked in the NCBI Primer Designing Tool and the results are shown in Results section, Table 2b with a product of 286 bp's. Variants were confirmed in the D-loop alignment (see Results table 3). The variants identified as SNPs are potential population identifiers and could be helpful in population dynamics.

Results 1

In Figure 2a, the COI consensus is shown. The COI consensus was made from twelve 649 bp long sequences. The consensus sequence also contains 649 bp. COI consensus has no variants with variant frequency over 0.25 or 25%.

COI Consensus Sequence for *Mola tecta*

TTCGGTGCATGAGCCGGGATAGTGGGGACGGCCTTAAGCCTACTCATTCGAGCGGAGCTA 60

AGTCAACCTGGCGCTCTTCTTGGAGACGACCAAATTTACAATGTCATCGTCACAGCACAT 120

GCATTTGTAATAATTTTCTTTATAGTAATACCAATTATGATCGGGGGCTTCGGAAACTGA 180

CTGATCCCTCTTATGATTGGGGCCCCCGATATGGCCTTCCCCCGAATGAATAATATGAGC 240

TTTTGACTCTTGCCCCCCTCTTTTCTTCTCCTTCTTGCCTCCTCAGGCGTCGAAGCAGGT 300

GCCGGAACAGGATGAACTGTATACCCCCCTTTAGCCGGAAACTTAGCCCACGCAGGCGCC 360

TCTGTTGATTTAACAATCTTTTCCCTTCACCTGGCCGGTATCTCCTCAATTCTAGGGGCC 420

TCTGTTGATTTAACAATCTTTTCCCTTCACCTGGCCGGTATCTCCTCAATTCTAGGGGCC 480

CCCCTATTTGTATGGGCAGTCCTCATTACGGCAGTACTTCTTCTCCTCTCGCTTCCAGTT 540

CTTGCAGCCGGAATCACGATGCTTCTTACGGATCGAAACCTTAACACTACCTTCTTCGAC 600

CCGGCAGGCGGAGGAGACCCAATCCTGTATCAACATCTCTTCTGATTCT 649

Figure 2a. COI Consensus Sequence, *Mola tecta*

D-loop Consensus Sequence for *Mola tecta*

AAACTACTCTTTGATAGCGGCTATAT—ATATATATATGTATTATCACCATATATACATA 58

TATGTATTATCACCATATATATATATATACCATCAATCGATATTATTTCACGAGCAATAA 118

ATTATATGTGAAATAAAATGATTTCAAACATCACAAGAGGATGTAAAATTTGAATGTATA 178

AAAGATATAACATTAACTAGGACATCCTGACTAAAAGACCATCCTCTCAACATTCCGAAT 238

AACTTAAACAGATATACTTTGACTCAACATCTCTTTAAAAGAAATATTTAATGTAGTAAG 298

AACCGACCATCAGTTGATTTCTTAATGCATACTCTTATTGATGGTGAGGGACAATTATTT 358

GTGGGGGTCACACCCAGTGAATTATTCCTGGCATTTGGTTCCTATTTCAGGGCCATAAAC 418

TGAATAACTCCACACACTTTCATCGACGCTTGCATAAGTTAATGGTGGAACCCATATGGC 478

GCGATAACCCACCATGCCGGGCGTTCTTTCCAGCGGGTCACTGGTTCTCTTTTTTATTTT 538

CCTTTCATTTGACATCTCAGAGTGCACACGGTAATAGTGAAATAAGGTTGAACATTTCCT 598

TGGTTTCAATAAATACCGTTAATGGTGAAAGATATGGATTGAAAGGGCATACATAACTGA 658

TCTCAAGGACATAATACTGTAGTGTTCACTCGAAAGATCCCTATAAGTGCCCCGGGGTTT 718

CTACGCGTAAAACCCCCCTACCCCCCTAAACTACTAGGATGACTAACGTTCCTGAAAACC 778

CCCCGTAAACAGGAAGCCTCCGAGTAGTATATTTTTTACCTAAAAGTGTGTTTATTTACA 838

TTATTTAAATAATGCACTTGCCAGC 863

Figure 2b. D-loop Consensus Sequence, *Mola tecta*

In Figure 2b, the D-loop Consensus is shown. The D-loop consensus was made from 23 sequences ranging from 827 to 859 bp's. Variants were confirmed in the D-loop alignment. See Table 3 for variants.

Table 2a shows the COI forward and reverse primer sequences, and the corresponding Primer Blast analysis. Both primers are 20 bps each. The forward primer is located from 165→ 184 in the COI consensus sequence with an annealing temperature of 59.97°C. The reverse primer is located from 515→496 in the consensus sequence with an annealing temperature of 60.11°C. The difference between annealing temperatures is 0.14°C.

Table 2b shows the D-loop forward and reverse primers with respective positioning to table 2a. Annealing temperature difference is 0.45°C.

In Table 3, seven pairs of variants are identified in the D-loop alignment with variant frequencies over 0.25 (25%). All variants are identified as SNP's.

Table 2a. Primer 1 Blast Report

Primer pair 1	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGGCTTCGGAAACTGACTGA	Plus	20	165	184	59.97	55	5	1
Reverse primer	ACTGCCGTAATGAGGACTGC	Minus	20	515	496	60.11	55	4	3
Product length	351								

Table 2b. Primer 2 Blast Report

Primer pair 2	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCACACACTTTCATCGACGC	Plus	20	428	447	59.84	55	4	2
Reverse primer	CCGGGGCACTTATAGGGATC	Minus	20	713	694	59.39	60	4	4
Product length									286

Table 3. SNP Identification in D-loop

Name	Type	Minimum	Maximum	Length	Direction	Coverage	Polymorpl	Variant Fre
C	Polymorpl	18	18	1	none	13	SNP	69.20%
T	Polymorpl	18	18	1	none	13	SNP	30.80%
C	Polymorpl	94	94	1	none	23	SNP	73.90%
T	Polymorpl	94	94	1	none	23	SNP	26.10%
C	Polymorpl	236	236	1	none	23	SNP	65.20%
T	Polymorpl	236	236	1	none	23	SNP	34.80%
A	Polymorpl	237	237	1	none	23	SNP	43.50%
G	Polymorpl	237	237	1	none	23	SNP	56.50%
C	Polymorpl	276	276	1	none	23	SNP	30.40%
T	Polymorpl	276	276	1	none	23	SNP	69.60%
G	Polymorpl	412	412	1	none	23	SNP	73.90%
T	Polymorpl	412	412	1	none	23	SNP	26.10%
A	Polymorpl	625	625	1	none	23	SNP	26.10%
G	Polymorpl	625	625	1	none	23	SNP	73.90%

Conclusions 1

In Tables 2a and 2b, Blast Reports analyze likelihood of primers working well. If the chosen primers work, there is also a possibility of self-annealing and making a primer dimer. Likelihood is analyzed through Primer Blast's self complementarity scores. A higher score is more likely to have primer dimers and self-annealing. The scores for our primer pairs range from 1-5, which is not ideal, but is less than a default score of 8. Because primer design is rather inconclusive until tested in a lab, these primers should be tested in lab for a conclusive analysis.

There were no variants to analyze in the COI data, but there were variants found in the D-loop. These variants were identified as SNP's, making the D-loop promising for *Mola tecta* population identification. Furthermore, if the D-loop primer design works, four of these SNP's will be within the amplification region, showing even more promise for population identification. Potentially, this research could provide a solid means to identify and track *Mola tecta* using D-loop amplification and SNP analysis.

Methods 2

eDNA samples were taken on cruises from 2013-2019. Samples were analyzed with Illumina NGS and sequenced using the MiFish primer below (Miya et al., 2015).

primer_sequence_F	primer_sequence_R
GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG

Primer sequence used→(Miya et al., 2015)

Molidae ASV's were found by referencing 12S master taxa and master metadata sheets from sample amplification data sheets. Of the Tetraodontiformes, the only ASV's detected are shown below, in Table 4.

Results 2

Table 4. ASV Detections and Reads

ASV	nDetections	Total Reads
ASV_90	37	44005
ASV_2467	2	595
ASV_3581	2	551
ASV_3601	2	295
ASV_3621	1	162

In Table 4, nDetections represents how many samples each ASV was present. Total Reads represents the total amount of hits in all samples. All ASV's were identified as Genus *Mola;* however, only ASV_2467 had a specificity identified. ASV_2467 was identified as *Mola mola*.

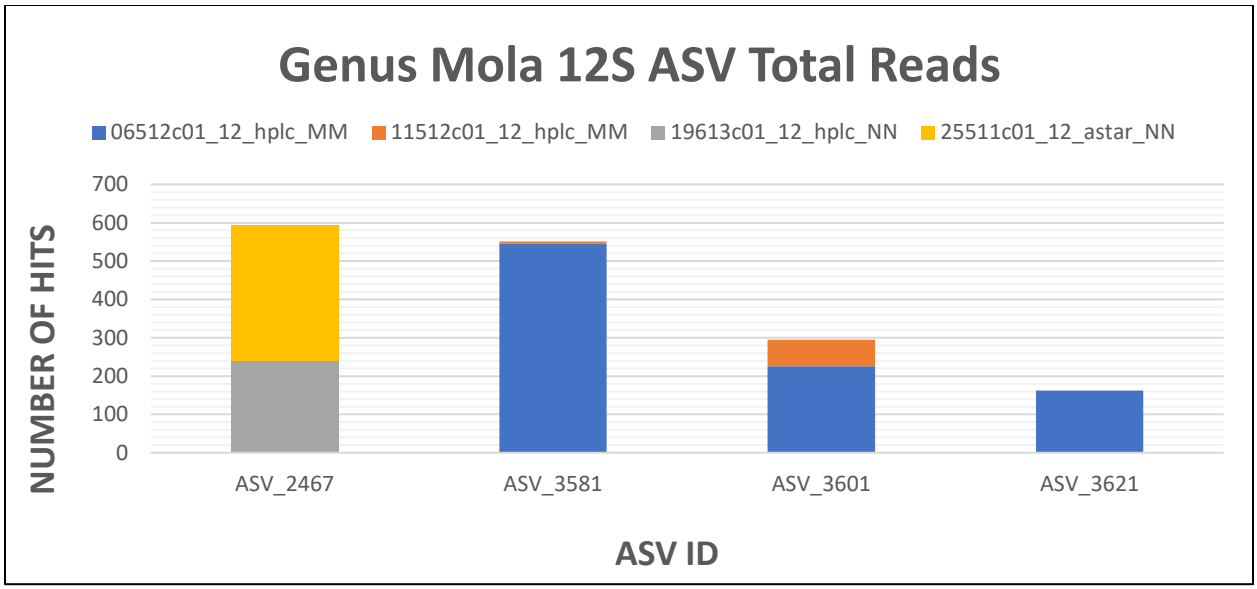

Figure 3. Genus *Mola* 12S ASV Total Reads vs Sample ID

Compares 12S ASV hits to sample ID. Quantifies presence of the different ASV's to site/sample. Applications to migration patterns based on location and time ASV was detected. Does not include ASV 90. ASV 90 found in *Mola mola* and is most abundant (see next slide for ASV 90).

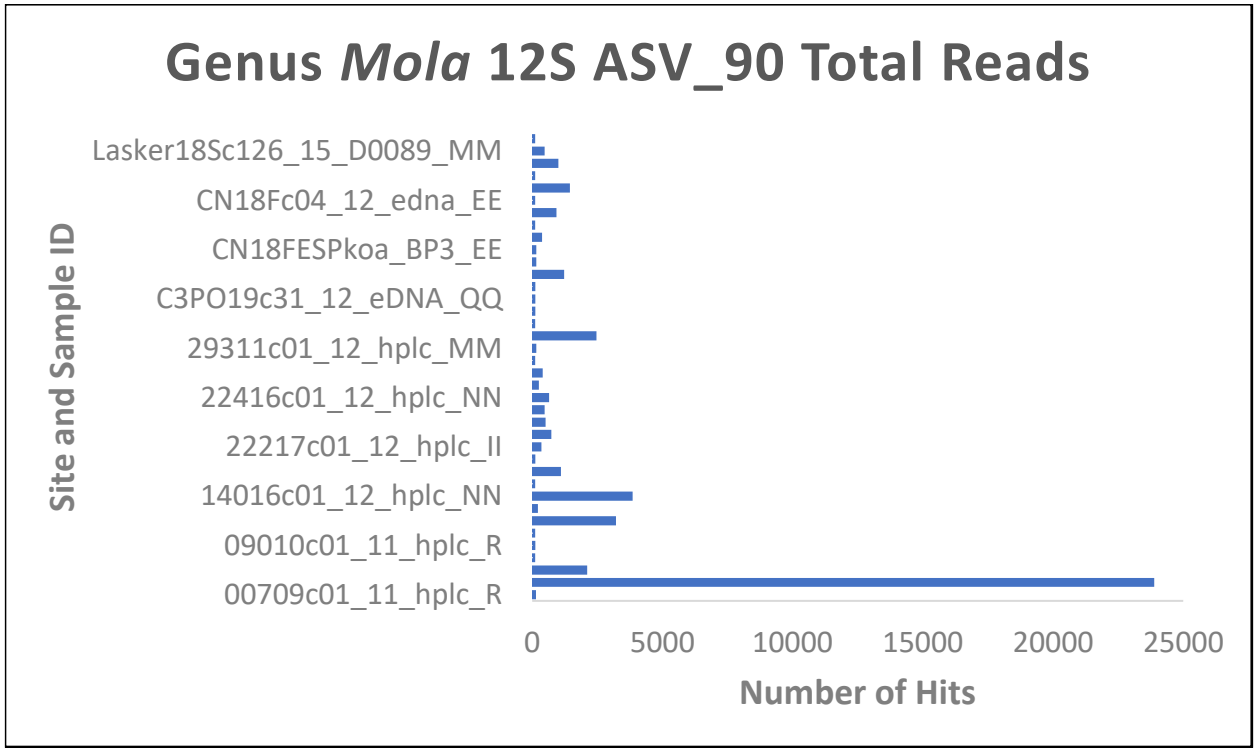

Figure 4. Site ID vs Total Reads of ASV_90 Hits

Description: Excludes SampleID 06512c01_12_hplc_MM with the highest number of hits, at 23,786.

Conclusions 2

Of the five ASV's found in the 12S *Molidae* sequences, ASV_90 is much higher than any other ASV present (73 times higher in total reads and 18 times higher in number of detections). ASV_90 is clearly the most abundant and most frequent. Once ASV_90's species is identified, it could be used as a marker for presence of a specific *Molidae*. ASV_2467 was identified as *Mola mola*, and can therefore be used as a tracker for *Mola mola*. Interestingly, presence of ASV is consistently found throughout various years at a similar latitude and longitude. More research is needed to draw conclusions, relationships, and patterns for the other ASV's. The potential for tracking and studying *Molidae* with eDNA remains.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7). https://doi.org/10.1098/rsos.150088

Nucleotide BLAST: Search nucleotide databases using a nucleotide query. (n.d.). Retrieved December 25, 2020, from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Nyegaard, M., Sawai, E., Gemmell, N., Gillum, J., Loneragan, N. R., Yamanoue, Y., & Stewart, A. L. (2017). Hiding in broad daylight: molecular and morphological data reveal a new ocean sunfish species (Tetraodontiformes: Molidae) that has eluded recognition. *Zoological Journal of the Linnean Society*, *182*(3), 631–658. https://doi.org/10.1093/zoolinnean/zlx040

WoRMS - World Register of Marine Species - Mola mola (Linnaeus, 1758). (n.d.). Retrieved December 24, 2020, from http://www.marinespecies.org/aphia.php?p=taxdetails&id=127405

Yoshita, Y., Yamanoue, Y., Sagara, K., Nishibori, M., Kuniyoshi, H., Umino, T., Sakai, Y., Hashimoto, H., & Gushima, K. (2009). Phylogenetic relationship of two Mola sunfishes (Tetraodontiformes: Molidae) occurring around the coast of Japan, with notes on their geographical distribution and morphological characteristics. *Ichthyological Research*, *56*(3), 232–244. https://doi.org/10.1007/s10228-008-0089-3