Species Identification of a Commonly Believed Sinarapan Fish *Mystichtys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region

Jacob Anderson Cajuigan Sanchez1, 5, Venus Banguilan Buraga1, Carmelita Pingol Mapanao1, 4, Mary Beth Bacano Maningas1, 2 and Mudjekeewis Dalisay Santos3

1. Graduate School of University of Santo Tomas (UST), España, Manila 1008, Philippines
2. Molecular Biology and Biotechnology Laboratory, Research Center for Natural and Applied Sciences (RCNAS), University of Santo Tomas, Manila 1008, Philippines
3. Genetic Fingerprinting Laboratory, Bureau of Fisheries and Aquatic Resources-National Fisheries Research and Development Institute (BFAR-NFRDI), Quezon City 1011, Philippines
4. College of Arts and Sciences, Polytechnic University of the Philippines, Manila 1016, Philippines
5. College of Arts and Sciences, Pampanga State Agricultural University, Pampanga 2011, Philippines

**Abstract:** The paper aimed to determine the true identity of a presumed or commonly believed Sinarapan fish and compare their phylogenetic relationships by using DNA barcoding. The fish samples were collected by researchers from a fisheries research and development institution in April 2015 from four different lakes of Bicol Region, i.e., Lakes Buhi, Danao, Manapao and Bato. DNA was extracted using cetyl trimethyl ammonium bromide (CTAB) extraction buffer. The polymerase chain reaction (PCR) profile for the reaction was 94 °C for 10 min, followed by 35 cycles of 1 min at 94 °C, 1 min at 48 °C and 1.5 min at 72 °C, and a final extension of 10 min at 72 °C. The CO1 region with approximately 650 bp was amplified because of its capability to differentiate taxa. Sequencing was done by Macrogen while phylogenetic analysis was produced using a Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0. The findings showed that CO1 can be used as a DNA marker in molecular identification of the fish samples. Samples from two of the four lakes were successfully sequenced. From basic local alignment search tool (BLAST) results, the maximum parsimony (MP) and neighbor-joining (NJ) trees revealed that samples from Lakes Buhi and Bato are not species of Sinarapan but rather species of *Leiopotherapon plumbeus* and *Rhinogobius giurinus*, respectively. Furthermore, DNA barcoding is very useful in proving the true identity of unknown samples.

**Key words:** Goby fish, Sinarapan, DNA barcoding, CO1 gene, Bicol region.

1. **Introduction**

Sinarapan, *Mystichtys luzonensis* (Smith, 1902) is listed in the Guinness book of world records 2001 as the world’s smallest commercially-harvested fish [1] in freshwater lakes of Camarines Sur and other parts of Bicol region, Philippines [2]. The magnitude of this recognition has increased following various claims that the fish is endemic in Bicol region [3].

Known to be very small, fragile and delicious, this fish is facing possible extinction. To address this problem, the House Bill No. 1011 [4] and Senate Bill No. 1046 [5] during the Sixteenth Congress of the Republic of the Philippines declared the Sinarapan sanctuary to prevent over exploitation. The Sinarapan sanctuary covers Lake Manapao and Lake Makuao, and both are located in Barangay San Ramon. Lake Katugday is in between Barangay San Ramon and Barangay De Lafe in municipality of Buhi, Camarines Sur, and Lake Danao is located in the boundary of Barangay Danao, municipality of Polangui, Albay and...
the municipality of Buhi, near Barangay Makaangay. Lake Buhi, Buhi, Camarines Sur, Lake Danao, Danao, Masbate, Lake Manapao, Barangay San Ramon, Buhi Town, Camarines Sur, and Lake Bato, Bato, Camarines Sur are said to be the homes of Sinarapan and thus serve as main sources for fishing.

The Sinarapan is from the genus *Mystichtys* and Gobiidae family. They are morphologically transparent, except for black eyes, with an average length of 12.5 mm. Males are smaller than females. Previous reports showed that the biggest specimen only grew up to 15 mm with an average length of only 12.5 mm [2]. However, the fish started disappearing in the 1980’s with the advent of motorized fishing to capture them for commercial purposes [6].

A report from the Department of Agriculture [7] revealed that *Mystichtys luzonensis* is threatened with extinction due to overfishing and introduction of exotic species, like Nile tilapia (*Oreochromis niloticus*) [8]. While recent efforts have been conducted for habitat enhancement [9] and water quality monitoring [10] for conservation, there are few studies that took account of the Sinarapan’s true identity.

In developing proper management strategies, the taxonomy will be fundamental since accurate identification of a species is a basic pre-requisite in fisheries management [11]. During the past decades, DNA barcoding and marker-assisted selection (MAS) have been used to identify fishery species [12]. A short fragment, which usually ranges around 648-705 base pairs [13] of cytochrome c oxidase 1 (CO1) marker—a mitochondrial DNA, can be used to identify fishes [14]. Previous researches proved that this identification system was able to differentiate the taxa of different fishes in Pantabangan dam [15] and has provided a quick solution to the challenges of using conventional taxonomy in fisheries based on morphological characteristics, especially when working on immature or damaged samples. The gene (CO1) has two important advantages: (1) the universal primers for this gene are very robust, enabling recovery of its 5’ end from representatives of most, if not all, animal phyla [16] and (2) CO1 has a greater range of phylogenetic signal than any other mitochondrial gene. Similar with other protein coding genes, its third-position nucleotides show a high incidence of base substitutions, leading to a rate of molecular evolution that is about three times greater than that of 12S or 16S rDNA [17].

Here, this paper aimed to use DNA barcoding using CO1 DNA marker on Sinarapan sample collections to confirm its true identity and determine the phylogenetic relationship of the collected samples.

### 2. Materials and Methods

#### 2.1 Samples

The Sinarapan samples were collected by the of researchers from the Bureau of Fisheries and Aquatic Resources-National Fisheries Research and Development Institute (BFAR-NFRDI) from four different lakes in Bicol Region, namely, Lakes Buhi (L1), Danao (L2), Manapao (L3) and Bato (L4). Muscle tissues of about 100 mg were obtained from randomly selected samples, fixed in 95% ethanol and stored at -20 °C until DNA extraction.

#### 2.2 DNA Extraction

Three samples were obtained from each lake for replication purposes. DNA was extracted using cetethyl trimethyl ammonium bromide (CTAB) extraction buffer following the method of Doyle [18] with modifications by Santos et al. [19]. The concentrations of the DNA extracts were measured using a Nano spectrophotometer with concentrations ranging from 144 ng/uL to 148 ng/uL.

#### 2.3 CO1 Amplification

A 24 µL reaction mixture was prepared containing water, 10× PCR buffer, 2 mM dNTP’s, 10 uM each of forward primer LCO1490: 5’GGTCACAAATCATAAAGATATTGG3’ and reverse primer HCO2198:
Species Identification of a Commonly Believed Sinarapan Fish *Mystichtys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region

5’TAAACTTCAGGTTGACCAAAAAATCA3’[16], 25 mM MgCl₂, 5 units/μL Taq polymerase and 2 μL of DNA template. The polymerase chain reaction (PCR) profile for the reaction was 94 °C for 10 min, followed by 35 cycles of 1 min at 94 °C, 1 min at 48 °C, 1.5 min at 72 °C, and a final extension of 10 min at 72 °C. CO1 amplicons were electrophoresed through a 1% agarose gel stained with GelRed and submerged in 1× TAE buffer. An approximately 650 bp fragment of the CO1 was amplified.

2.4 Purification and Sequencing

The PCR amplicons were purified in the Molecular Biology and Biotechnology Laboratory of University of Santo Tomas (UST) and were sent to Macrogen Korea for sequencing. The resequencing was done as needed.

2.5 Phylogenetic Analysis

The maximum parsimony (MP) and neighbor-joining (NJ) cladograms were created with fish species that were obtained from available online data. There is no reference sequence for *Mystichtys luzonensis* found in the NCBI reference sequence (http://www.ncbi.nlm.nih.gov/refseq). The sequences were aligned using Clustal X software and were determined to be suitable for the said phylogenetic trees [20]. All genetic distance analyses were established using Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0 [21]. Data for guanine and cytosine (GC) percentage were collected using the sequence analysis provided by Macrogen.

3. Results and Discussion

3.1 CO1 Gene Amplified in All Samples

A total of 12 samples (Fig. 1), composed of the four different lakes with three replicates each, were used in this test. Except for the negative control containing no template, all samples from the four different lakes showed positive amplification of the CO1 gene with approximately 650 bp. This confirmed that the samples being tested belong to the animal phyla. It also affirmed the consistency of the optimized extraction and amplification protocols done in this experiment. The usefulness of CO1 gene in barcoding of unknown fish samples was achieved.

In addition, the CO1 gene can resolve potential mislabeling due to incorrect morphological identification, since Sinarapan shares similar phenotypic characteristics with a variety of juvenile fishes. It also was affirmed in the previous report that CO1 gene can be used to identify the incidence of mislabeling. The said study found that frozen “Tawilis” which is a Sardinella species, and “bluefin” tuna fillets are sold at a high price, yet consumers will later on realize that they bought a different kind of fish.

Hence, molecular identification using CO1 gene can be used to address the complexity and limitations of morphological characters in fishes to effectively discriminate different species. Furthermore, CO1 gene has been proved to be an effective tool for a reliable and fast identification of commercially-important fishes [15].

3.2 Purification of PCR Products

Two PCR products from each lake were chosen for DNA purification based on the intensity of bands. These are L1S1, L1S3, L2S1, L2S3, L3S1, L3S3, L4S1 and L4S3 (Fig. 2). However, results showed that L2S3, L3S1 and L3S3 have degraded DNA samples. Hence, only one sample from lake 2 can be used, while lake 3 is no longer represented.

The primer dimers found in each well called for another round of purification. Re-purification of the good samples in Fig. 2 resulted to a very low intensity of primer dimmers, indicating that the gene of interest, CO1, was already purified (Fig. 3). These samples were sent to Macrogen for sequencing.

The degradation of the DNA in samples of L2S3, L3S1 and L3S3 may be attributed to different factors, such as abrupt change in temperature, storage, frequent thawing, low DNA concentration and chemical
Species Identification of a Commonly Believed Sinarapan Fish *Mystichthys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region

Fig. 1 Amplification of CO1 gene from all samples.
PCR products showing positive amplification at ~650 bp across all lanes except for negative control.

Fig. 2 Five samples showing positive amplification.
Purified PCR products in all lanes except L2S3, L3S1, L3S3 and negative control.

Fig. 3 Purified PCR products showed positive amplification at 650 base pairs in all lanes except negative control.
Species Identification of a Commonly Believed Sinarapan Fish *Mystichthys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region

intervention. On the other hand, the purification of DNA resulted to the removal of possible contaminants, such as protein including other exogenous sources.

3.3 L1S1 (Sample from Lake Buhi) and L4S1 (Sample from Lake Bato)

From the five samples sent to Macrogen, there were three sets of re-sequencing and only two samples were successfully sequenced. These are L1S1 and L4S1. Sequence read lengths ranged from 542 to 693 base pair with an average length of 618 base pair long. No insertions, deletions or stop codons were observed, indicating the absence of nuclear mitochondrial pseudogenes (NUMTs) among sequences [15]. A simple evolutionary tree, representing the history of collected species, was inferred using MP and NJ methods (Figs. 4 and 5). After the MP and NJ trees generation, DNA barcoding reliably grouped the species according to their families. Bootstrap analysis was accomplished using 1,000 pseudo replications to increase the validity of the results.

The L1S1 formed a monophyletic group with *Leiopotherapon plumbeus* (bootstrap value of 99%) and *L. unicolor* (bootstrap value of 94%). Meanwhile, L4S1 formed a monophyletic group with *R. giurinus* (bootstrap value of 99%). A separate monophyletic group of *Rhinogobius* sp., and *R. brunneus* is indicative that the aforementioned species can be grouped with L4S1 vis-à-vis optimization in the parameters set in the MP tree.

The MP cladogram of L1S1 and L4S1 (Fig. 4) had a consistency index of 0.273543, the retention index is 0.608696, and the composite index is 0.168478 for all sites and parsimony-informative sites. The percentage of replicate trees, in which the associated taxa clustered together in the bootstrap test, is shown next to the branches. The evolutionary distances were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site. The analysis involved 43 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 452 positions in the final dataset. Evolutionary analysis was conducted in MEGA version 6 [21].

The nucleotide composition of L1S1 is 2,777 (A), 3,811 (T), 1,559 (G) and 4,269 (C) with an average GC content of 47%. BLAST results revealed that L1S1 is 100% (five BLAST accessions) and 99% (five BLAST accessions) homologous to *L. plumbeus* (Table 1).

The nucleotide composition of L4S1 is 2,207 (A), 3,125 (T), 1,136 (G) and 3,107 (C) with an average GC content of 46%. BLAST results revealed that L4S1 is 98% (five BLAST accessions) and 99% (three BLAST accessions) homologous to *R. giurinus* (Table 2).

On the other hand, the NJ method for L1S1 and L4S1 (Fig. 5) produced an optimal tree with the sum of branch length = 1.71780854 is shown. The percentage of replicate trees, in which the associated taxa clustered together in the bootstrap test, is shown next to the branches. The evolutionary distances were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site. The analysis involved 43 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 452 positions in the final dataset. Evolutionary analyses were conducted in Molecular Evolutionary Genetics Analysis (MEGA 6).

Both cladograms (MP and NJ) showed that L1S1 and L4S1 belong to another type of species, revealing that the previous on-site collections are probably composed of a variety of fishes other than Sinarapan. Meanwhile, MP and NJ cladograms of L1S1 and L4S1 tell us that they are closely related to *L. plumbeus* and *R. giurinus*, respectively, and not directly associated with the other species.

It can be deduced that the reasons for obtaining different kinds of fish instead of Sinarapan are the following: (1) government initiative to restock different kinds of fish, (2) the introduction of invasive species, i.e., fish cages of tilapia, (3) the site of collection
Species Identification of a Commonly Believed Sinarapan Fish *Mystichthys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region

Fig. 4  The MP tree of L1S1 and L1S4.
Fig. 5  The NJ tree of L1S1 and L1S4.
has few population of Sinarapan due to overfishing. (4) Sinarapan can be confused with juvenile fishes because of similar morphological traits and (5) the collection was heterogenous, therefore, it is possible that the extracted DNA may have come from a different kind of fish. The approximately 100% homology of LISI and L4S1 with *L. plumbeus* and *R. giurinus* confirmed that the collections are not Sinarapan. This is indicative of diminishing population of Sinarapan which is associated to overfishing as it commands a very high market price.

Likewise, previous literatures mentioned that initiatives to restock Sinarapan was through the dispersal of “goby fishes” but did not cite if they are indeed Sinarapan or may be other members of the Goby family. In coming up with the relationships of

---

### Table 1  BLAST Result of LISI (Lake Buhi).

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. plumbeus</em> Voucher DB17.2 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,068</td>
<td>1,068</td>
<td>77%</td>
<td>0.0</td>
<td>99%</td>
<td>KJ013049.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher DB17.3 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,068</td>
<td>1,068</td>
<td>77%</td>
<td>0.0</td>
<td>99%</td>
<td>KJ013048.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher DB17.1 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,064</td>
<td>1,064</td>
<td>76%</td>
<td>0.0</td>
<td>100%</td>
<td>KJ013047.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher PNT-08 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,064</td>
<td>1,064</td>
<td>77%</td>
<td>0.0</td>
<td>99%</td>
<td>KC789541.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu1 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,064</td>
<td>1,064</td>
<td>77%</td>
<td>0.0</td>
<td>99%</td>
<td>HQ654716.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu1-LdB cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,055</td>
<td>1,055</td>
<td>75%</td>
<td>0.0</td>
<td>100%</td>
<td>HQ682700.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu2 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,055</td>
<td>1,055</td>
<td>75%</td>
<td>0.0</td>
<td>100%</td>
<td>HQ654715.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu3 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,055</td>
<td>1,055</td>
<td>75%</td>
<td>0.0</td>
<td>100%</td>
<td>HQ654714.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu4 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,055</td>
<td>1,055</td>
<td>75%</td>
<td>0.0</td>
<td>100%</td>
<td>HQ654713.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lplu5 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>1,044</td>
<td>1,044</td>
<td>75%</td>
<td>0.0</td>
<td>99%</td>
<td>HQ654712.1</td>
</tr>
<tr>
<td><em>L. plumbeus</em> Voucher Lu2 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>863</td>
<td>863</td>
<td>76%</td>
<td>0.0</td>
<td>94%</td>
<td>KF227959.1</td>
</tr>
</tbody>
</table>

### Table 2  BLAST result of L4S1 (Lake Bato).

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>R. giurinus</em> Voucher 1F75 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>968</td>
<td>968</td>
<td>75%</td>
<td>0.0</td>
<td>99%</td>
<td>HQ536473.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> mitochondrion, complete genome</td>
<td>966</td>
<td>966</td>
<td>77%</td>
<td>0.0</td>
<td>98%</td>
<td>KP982753.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> mitochondrion, complete genome</td>
<td>966</td>
<td>966</td>
<td>77%</td>
<td>0.0</td>
<td>98%</td>
<td>KF371534.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 1F74 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>963</td>
<td>963</td>
<td>75%</td>
<td>0.0</td>
<td>98%</td>
<td>HQ536472.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 2F13 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>957</td>
<td>957</td>
<td>75%</td>
<td>0.0</td>
<td>98%</td>
<td>HQ536471.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 2F12 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>957</td>
<td>957</td>
<td>75%</td>
<td>0.0</td>
<td>98%</td>
<td>HQ536470.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher GB11159 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>841</td>
<td>841</td>
<td>64%</td>
<td>0.0</td>
<td>99%</td>
<td>JX679057.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher GB11160 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>830</td>
<td>830</td>
<td>64%</td>
<td>0.0</td>
<td>99%</td>
<td>JX679058.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 1F88 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>586</td>
<td>586</td>
<td>75%</td>
<td>2e-163</td>
<td>86%</td>
<td>HQ536462.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 1F84 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>586</td>
<td>586</td>
<td>75%</td>
<td>2e-163</td>
<td>86%</td>
<td>HQ536458.1</td>
</tr>
<tr>
<td><em>R. giurinus</em> Voucher 1F79 cytochrome oxidase subunit 1 (CO1) gene, partial cds: mitochondrial</td>
<td>580</td>
<td>580</td>
<td>75%</td>
<td>8e-162</td>
<td>86%</td>
<td>HQ536463.1</td>
</tr>
</tbody>
</table>
the different fish species, the MP and NJ cladograms using ClustalX and MEGA softwares are reported to be effective tools in species identification. Lastly, the need to study the genome of Sinarapan (*M. luzonensis*) is recommended, because there is no reference sequence in the online database for this kind of fish.

4. Conclusions

This study concluded that the samples collected from the different lakes are not species of Sinarapan, but rather species of *L. plumbeus* (L1S1) and *R. giurinus* (L4S1). Since these samples are members of different species, they did not form a single evolutionary group. Furthermore, DNA barcoding is very useful in proving the true identity of unknown samples.

Acknowledgments

The authors would like to express their gratitude to Dr. Drusila Esther Estenor Bayate of BFAR-NFRDI, Dr. Mudjekeewis Dalisay Santos and staff of Genetic Fingerprinting Laboratory of BFAR-NFRDI for their kindness, technical assistance, and for allowing the researchers to conduct this research in their laboratory. They are also very thankful to Dr. Mary Beth Bacano Maningas and staff of the Molecular Biology and Biotechnology Laboratory of the University of Santo Tomas—Research Center for Natural and Applied Sciences for their generosity and expertise. To God be the glory.

References


Species Identification of a Commonly Believed Sinarapan Fish *Mystichthys luzonensis* (Smith, 1902) in Lakes Buhi and Bato of Bicol Region


