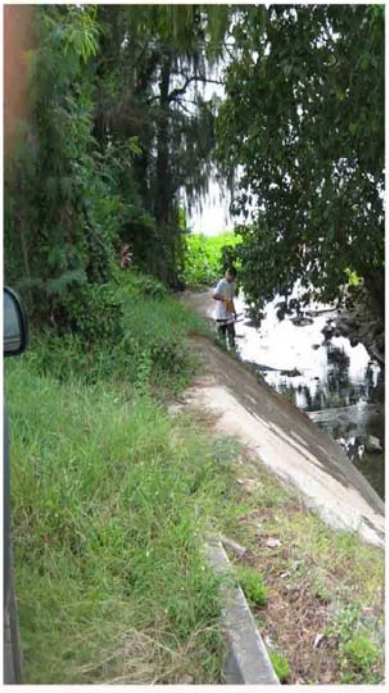

Molecular identification of freshwater fish species from the Commonwealth of the Northern Mariana Islands

Pacific Biological Survey

Final Report **January 2010**

FINAL REPORT

Confirming the identification of freshwater native and invasive fish species from the Commonwealth of the Northern Mariana Islands using molecular analysis

Prepared for:
Division of Fish and Wildlife
Saipan
Commonwealth of the Northern Mariana Islands

Prepared by:
Shelley A. James[1,2], Holly Bolick[1,2] & Arnold Suzumoto[1,3]
[1]Pacific Biological Survey
[2]Pacific Center for Molecular Biodiversity
[3]Ichthyology
Bishop Museum
Honolulu HI 96817

January 2010
Contribution No. 2010-001 to the Pacific Biological Survey

Confirming the identification of freshwater native and invasive fish species from the Commonwealth of the Northern Mariana Islands using molecular analysis

Introduction

It has long been recognized that DNA sequence diversity can be used to discriminate between species. Recent efforts have focused on developing a database of a single gene sequence, the mitochondrial gene cytochrome *c* oxidase subunit 1 (CO1), which is believed sufficient to differentiate the vast majority of animal species (e.g., Hebert *et al*., 2003; Ward, *et al.*, 2005). This process is commonly referred to as barcoding. The mitochondrial cytochrome *b* (cyt*b*) gene has frequently been used in molecular phylogeny studies, and can also be used for species identification (e.g., Parson *et al*., 2000; Hsieh *et al*., 2001; Tobe *et al*., 2009). GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) is the NIH genetic sequence database, consisting of more than 100 billion publicly available DNA sequences. This resource can be used to search for DNA sequences similar to that of an unknown specimen using *blast* algorithms.

Very little is known of the species that inhabit the freshwater habitats of the Commonwealth of the Northern Mariana Islands (CNMI). An extensive survey of freshwater interstitial streams on Saipan (McCagan *et al*., 2008) found seven native species and eight species of introduced fish, but were unable to confirm the identification of many of the species due to their size and morphological similarity between species. Here we use molecular barcoding techniques to help with species identification.

Materials and Methods

Sample collection

Samples were collected from freshwater systems on Saipan, CNMI (McKagan *et al*., 2008), frozen, and transported to the Pacific Center for Molecular Biodiversity, Bishop Museum for processing (Table 1). Upon thawing, muscle tissue was sampled and preserved in a salt-saturated solution of 20% dimethyl sulfoxide-EDTA (DMSO). Voucher specimens have been temporarily stored at -20C. Voucher DNA and tissues are housed at -80C in the Pacific Center for Molecular Biodiversity, Bishop Museum (PCMB 4840-4877).

DNA Extraction

Genomic DNA was extracted from approximately 25 mg DMSO-preserved muscle tissues using a DNeasy Blood and Tissue Kit (Qiagen Inc.) following the recommended protocol. The complete mitochondrial cytochrome *b* (cyt*b*) gene (approx. 1,200 bp) of each sample was amplified via polymerase chain reaction (PCR) using the primers GLU-L 5'TGA TAT GAA AAA CCA TCG TTG-3' (Palumbi *et al*., 1991) and H15915 5'-ACC TCC GAT CTY CGG ATT ACA AGA-3' (Aoyama *et al*., 2001). Approximately 655bp were amplified from the 5' region of the cytochrome *c* oxidase subunit 1 (CO1) gene were amplified via PCR using the primers FishF1 5'-TCA ACC AAC CAC AAA GAC ATT GGC AC-3' and FishR1 5'-TAG ACT TCT GGG TGG CCA AAG AAT CA-3' or FishR2 5'-ACT TCA GGG TGA CCG AAg AAT CAG AA-3' for samples PCMB 4848 and 4874 (Ward *et al*., 2005). Reactions were performed in a 50 μL solution containing 10 ng genomic DNA, 400 μM of each dNTP, 1.5 units Taq DNA polymerase (D4545, Sigma Chemical), 2 mM $MgCl_2$, each primer at 1 μM, and 1x PCR buffer. PCR cycling parameters were 94 C (1 min), 50 C (1 min), and 72 C (1 min) repeated for 35 cycles with a final extension of 72 C (5 min). PCR product was gel extracted

using QIAQuick Gel Extraction Kit (Qiagen Inc.). Cycle sequencing of 100 fmol double-stranded PCR product was carried out with the forward primer (GLU-L) for cyt*b*, and both the forward and reverse primers for CO1 using a CEQ DTCS Quick Start Kit (Beckman-Coulter) following the recommended protocol. The products sequenced on a CEQ8000 genetic analysis system. Sequence data were subjected to nucleotide megablast (high similarity) search (http://www.ncbi.nlm.nih.gov/BLAST/) in order to identify the best taxon match for the resultant sequence for each specimen.

Results & Discussion:

Five of the specimens had 99-100% match for both the cyt*b* and CO1 sequences within GenBank, namely PCMB 4840, 4846 and 4847 (*Chanos chanos*), PCMB 4842 (*Eleotris acanthopoma*), PCMB 4845 (*Megalops cyprinoides*), and PCMB 4848 (*Gambusia affinis*) (Table 1). PCMB 4841, preliminarily identified only as a Goby, had a 98% (cyt*b*) and 99% identity with GenBank sequences for *Stiphodon elegans* (Appendix 1). *Stiphodon elegans* and *Megalops cyprinoides* have been previously reported as native to the Commonwealth of the Northern Mariana Islands (CNMI) (McCagan *et al*., 2008). *Chanos chanos* and *Gambusia affinis* are introductions to the CNMI, and have been previously recorded from the islands. *Eleotris acanthopoma* is recorded as native to Guam, Palau, and the Western Pacific (FishBase, 2009), and would be expected to also be native in Saipan's freshwater ecosystems.

Specimens PCMB 4846 and 4847 were originally identified as a mullet species (Table 1). Molecular analysis, however, determined the two specimens to be *Chanos chanos*, and a secondary inspection of morphological characters of the voucher specimens indicated that while the specimens superficially resembled mullets, they lacked the diagnostic characteristics of the Mugilidae (A. Suzumoto, pers. comm., 2009).

Specimens PCMB 4859 and 4866 were preliminarily identified as *Poecilia reticulata* (Table 1). Both specimens had identical cyt*b* and CO1 sequence data, and this confirmed the samples to be species in the genus *Poecilia*. However, no match to species level could be confirmed, most likely due to an absence of sequence data for the unknown species having been submitted to GenBank. CO1 and cyt*b* sequence data were available for *Poecilia reticulata*, but the Saipan specimens had significantly different sequence data (see Appendix 2).

The three tilapia species identified in this study were *Oreochromis* species (Table 1, Appendix 3). Two of the three species had 100% match for the CO1 gene as *Oreochromis mossambicus* (PCMB 4874 and 4877). PCMB 4872, however, had 99% identity match with *Sarotherodon galilaeus* for the cyt*b* gene, and 99% identity match for *Oreochromis niloticus* for the CO1 gene. This discrepancy between the two gene regions is unclear, but may be due to misidentification within the GenBank database: the CO1 gene submissions may be more reliable due to the strict identification protocols now required for barcoding projects in order to generate a standardized reference library (see Fish-Bol, http://www.fishbol.org/). Historically, the CNMI had aquaculture farms featuring tilapia (*O. niloticus* and *O. mossambicus*) (McCagan *et al*., 2008). Hybrids of *O. mossambicus x O. niloticus* are possible, but hybridization among species can not be recognized with the molecular techniques used in this study. Mitochondrial DNA is maternally inherited, and any hybrid or subsequent generation would have the maternal species DNA only.

References:

Aoyama, J., M. Nishida, and K. Tsukamoto. 2001. Molecular phylogeny and evolution of the freshwater eel, genus *Anguilla*. Molecular Phylogenetics and Evolution 20: 450-459.

Hebert, P.D.N., A. Cywinska, S.L. Ball, and J.R. deWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London, Series B 270: 313-321.

Hsieh, H.-M., H.-L. Chiang, L.-C. Tsai, S.-Y. Lai, N.-E. Huang, A. Linacre and J.C. Lee**.** 2001. Cytochrome *b* gene for species identification of the conservation animals. *Forensic Science International* 122: 7-18.

McCagan, S., J. Herrod, P. Johnson, M. Tenorio, and M. Trianni. 2008. Freshwater Invasive Species Project. Division of Fish and Wildlife, CNMI.

Palumbi, A., S. Martin, S. Romano, W.O. McMillan, L. Stice, and G. Grabowski. 1991. The simple fool's guide to PCR, version 2. University of Hawaii at Manoa, Honolulu.

Parson W., K. Pegoraro, H. Niederstätter, M. Föger and M. Steinlechner. 2000. Species identification by means of the cytochrome *b* gene. International Journal of Legal Medicine 114: 23-28.

Tobe, S.S., A. Kitchener, A. Linacre. 2009. Cytochrome *b* or cytochrome *c* oxidase subunit 1 for mammalian species identification - an answer to the debate. Forensic Science International: Genetics Supplement Series 2: 306-307.

Ward, R.D., T.S. Zemlak, B.H. Innes, P.R. Last, and P.D.N. Hebert. 2005. Barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B 360, 1847–1857.

Table 1: Specimens analyzed in this project. Details for the collection sites can be found in McCagan *et al.* (2008). Species names in bold indicate a high probability in the maximum identity columns of genus and/or species match.

PCMB number	Collection site	Preliminary identification	cyt*b* maximum identity	CO1 maximum identity
4840	6B	*Chanos chanos*	100% ***Chanos chanos*** (700bp) 84% *Nemacheilus putaoensis*	100% ***Chanos chanos*** (655 bp) 84% *Barbonymus gonionotos*
4841	5A	Goby	98% ***Stiphodon*** *elegans* (400 bp) 85% *Gillichthys* sp. 85% *Rhinogobius* sp.	99% ***Stiphodon*** *elegans* (626 bp) 87% *Sicyopterus lagocephalus* 86% *Gnatholepis* sp.
4842	6B	*Eleotris* cf. *acanthopoma*	100-99% ***Eleotris acanthopoma*** (600bp) 98% *Eleotris sandwicensis* 84% *Eleotris melanosoma* 84% *Eleotris fusca*	99% ***Eleotris acanthopoma*** (644 bp) 99% *Eleotris sandwicensis* 86% *Eleotris picta*
4845	6B	*Megalops cyprinoides*	99% ***Megalops cyprinoides*** (414 bp) 89% *Megalops atlanticus*	99% ***Megalops cyprinoides*** (470 bp) 82% *Pervagor melanocephalus*
4846	6A	Mullet c.f. *Liza vaigiensis*	100% ***Chanos chanos*** (500 bp) 84% *Alosa pseudoharengus*	100% ***Chanos chanos*** (642 bp) 83% *Barbonymus gonionotus*
4847	6A	Mullet	No data.	99% ***Chanos chanos*** (458 bp)
4848	9A	*Gambusia* sp.	100-99% ***Gambusia affinis*** (700bp) 92% *Gambusia vittata* 91% *Gambusia atrora*	No data.
4859	6A	*Poecilia reticulata*	97% ***Poecilia*** *latipinna* (500 bp) 94% *Poecilia latipunctata* 92% *Poecilia gillii* 85% *Poecilia reticulata*	96% ***Poecilia*** *petenensis* (649 bp) 95% *Poecilia mexicana*
4866	6B	*Poecilia reticulata*	97% ***Poecilia*** *latipinna* (500 bp) 94% *Poecilia latipunctata* 92% *Poecilia gillii* 85% *Poecilia reticulata*	94% ***Poecilia*** *petenensis* (553 bp) 93% *Poecilia mexicana*
4872	9A	Tilapia	99% ***Sarotherodon*** *galilaeus multifasciatus* (600bp) 98% *Sarotherodon linnelii*	99% ***Oreochromis*** *niloticus* (600 bp) 99% *Oreochromis aureus* 98% *Sarotherodon galilaeus*
4874	11A	Tilapia	No data	100% ***Oreochromis mossambicus*** (400 bp)
4877	2A	Tilapia	95% ***Oreochromis*** *karongae* (370 bp) 95% *Oreochromis malagarasi* 95% *Oreochromis tanganicae* 94% *Oreochromis mossambicus*	100-94% ***Oreochromis mossambicus*** (636 bp) 95% *Oreochromis urolepis* 94% *Oreochromis niloticus*

Appendix 1: Mitochondrial cytochrome *b* sequence data for *Stiphodon elegans* specimen included in this study (PCMB 4841), and data from GenBank. Dashes indicate an identical nucleotide match with *S. elegans*.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
S. elegans      AAATTGCAAA TCACGCACTA GTTGACCTCC CTGCCCCTTC AAATATTTCT
4841            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
S. elegans      GCATGATGAA ACTTTGGCTC CCTTCTAGGA CTCTGCCTTG GTGCCCAAAT
4841            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
S. elegans      TGTCACTGGG CTATTTCTTG CAATACACTA TACCTCTGAC ATTGCCACAG
4841            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
S. elegans      CTTTCTCATC TGTTGCACAC ATTTGCCGTG ATGTTAACTT TGGCTGACTA
4841            ---------- ---------- ---------- ----C----- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
S. elegans      ATCCGGAACA TACATGCTAA TGGTGCCTCC TTCTTCTTTA TTTGCATTTA
4841            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
S. elegans      CCTGCACTTA GGGCGAGGCC TTTATTATGG CTCATACCTT TACAAAGAAA
4841            -------G-- ---------- -----A---- --------C- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
S. elegans      CATGAAACAT TGGTGTAGTT CTACTACTAC TTGTAATAAT AACTGCCTTT
4841            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
S. elegans      GTAGGCTATG TTCTCCCCTG AGGACAAATA TCCTTCTGAG GGGCCACAGT
4841            ---------- ---------- ---------- ---------- ----------
```

Appendix 2: Mitochondrial cytochrome *b* sequence data for *Poecilia* species included in this study (PCMB 4859 & 4866), and data from GenBank for *Poecilia reticulata* (the preliminary species identification), and *P. latipinna* species (most closely matching species). Dashes indicate an identical nucleotide match with *P. latipinna.*

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
P. reticulata   --------G- -------C-- ----G-C--- ---------- -A--AA-A--
P. latipinna    CACCCCCTCT TAAAAATTGC AAACAATGCA CTAGTAGACC TCCCCGCCCC
4866            ---------- ---------- ----G----- ---------- ----------
4859            ---------- ---------- ----G----- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
P. reticulata   -C-A--T--T --A------- ----T----- A--T--C--- ---T------
P. latipinna    CGTCAACATC TCCGCCTGAT GAAACTTTGG CTCCCTACTA GGACTATGCC
4866            ---------- ---------- ---------- ---------- ----------
4859            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
P. reticulata   --G-T----- ----T----- --T------- ---------- ---T------
P. latipinna    TCATCGCCCA AATCCTAACC GGCTTATTTT TAGCAATACA CTACACCTCT
4866            -A-------- ---------- ---------- ---------- ----------
4859            -A-------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
P. reticulata   --T--T--T- ---------- ---C---G-- ---------- -------A--
P. latipinna    GACATCTCCA TAGCATTCTC ATCTGTAACC CACATTTGTC GAGATGTTAA
4866            -----T---- ---------- -------G-- --T------- ----------
4859            -----T---- ---------- -------G-- --T------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
P. reticulata   T--T------ --T--T---- -C--A--C-- T--T--A--- -----G--T-
P. latipinna    CTACGGATGA CTCATCCGCA ATATGCATGC CAACGGGGCC TCCCTATTCT
4866            ---------- ---------- ---------- ---T--A--- ----------
4859            ---------- ---------- ---------- ---T--A--- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
P. reticulata   -C--T---C- ----C-G--- ---------- -AT------- T--G------
P. latipinna    TTATCTGCAT CTACTTACAT ATTGGACGAG GCCTATACTA CGGTTCCTAC
4866            ---------- ---------- ---------- ---------- ----------
4859            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
P. reticulata   C-G-T----- -G-------- ------T--A --C--AT-T- –T--A-----
P. latipinna    TTATACAAAG AAACATGAAA CATTGGCGTT ATTCTCCTAC TACTTGTTAT
4866            ---------- ---------- ---------- ---------- ----------
4859            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
P. reticulata   ---------- --C------- –T--A----- ---------- ----------
```

```
P. latipinna  AATAACTGCT TTTGTAGGAT ACGTTCTCCC ATGAGGACAA ATATCCTTCT
4866          ---------- ---------- -T-------- ---------- ----------
4859          ---------- ---------- -T-------- ---------- ----------

              ....|....| ....|....| ....|....| ....|....| ....|....|
                      410        420        430        440        450
P. reticulata ----T--A-- ------T--- --T-----A- ---------- C---A-----
P. latipinna  GAGGCGCCAC TGTAATCACC AACCTCCTCT CTGCTGTACC TTACGTAGGC
4866          ---------- ------T--- --T------- ---------- ----------
4859          ---------- ------T--- --T------- ---------- ----------

              ....|....| ....|....| ....|....| ....|....| ....|....|
                      460        470        480        490        500
P. reticulata --TG-C---- ---------- ---------- --------A- -------A--
P. latipinna  GACACTCTCG TCCAATGAAT CTGAGGGGGA TTTTCAGTTG ACAACGCTAC
4866          --------T- ---------- ---------- ---------- ----------
4859          --------T- ---------- ---------- ---------- ----------
```

Appendix 3: Mitochondrial cytochrome *b* sequence data for an unknown Tilapia species included in this study (PCMB 4877), and data from GenBank for the most closely matching species of *Oreochromis*. Dashes indicate an identical nucleotide match with unknown specimen. CO1 data indicated a 100% match with *Oreochromis mossambicus*.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                    5         15         25         35         45
4877            CTAAAAATTG CAAACGACGC ACTAGTTGAC CTCCCAGCCC CCTCAAACAT
O. karongae     ---------- ---------- ---------- ---------- ----------
O. malagarasi   ---------- ---------- ---------- ---------- ----------
O. tanganicae   ---------- ---------- ---------- ---------- ----------
O. mossambicus  ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                   55         65         75         85         95
4877            TTCCGTTTGA TRAAACTTTG GATCTCTACT AGGCCTCTGT CTAGCCGCCC
O. karongae     ---------- -g-------- -g-------- ---------- ----------
O. malagarasi   c--------- -g-------- ---------- ---g------ ----------
O. tanganicae   ---------- -g-------- ---------- ---------- ----------
O. mossambicus  ---------- -g-------- ---------- ---------- ------a---

                ....|....| ....|....| ....|....| ....|....| ....|....|
                  105        115        125        135        145
4877            AAATCCTGAC AGGCCTTTTC CTAGCCATAC ACTATACTTC CGACATCGCC
O. karongae     ---------- ---------- ---------- ----c----- ----------
O. malagarasi   ---------- ---------- ---------- -------c-- ----------
O. tanganicae   ---------- ---------t ---------- ----c--c-- ----------
O. mossambicus  ---------- ---------t ---------- -------c-- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                  155        165        175        185        195
4877            ACAGCCTTCT CCTCCGTTGC TCATATTTGT CRAGATGTAA ACTACGGCTG
O. karongae     ---------- -------c-- c--c------ -g-------- ----------
O. malagarasi   ---------- -------c-- c--c------ -g---c---- ----------
O. tanganicae   ---------- -------c-- c--c------ -g-------- ----t-----
O. mossambicus  ---------- -------c-- c--c------ -g---c---- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                  205        215        225        235        245
4877            ACTCATTCRA AACATACATG CCAACGGCGC ATCTTTCTTC TTCATTTGTA
O. karongae     --------g- ---------- ---------- ---------- ----------
O. malagarasi   --------g- ---------- ---------- ---------- ----------
O. tanganicae   --------g- ---------- ---------- ---------- --------c-
O. mossambicus  --------g- ---------- ---------- ---------- --------c-

                ....|....| ....|....| ....|....| ....|....| ....|....|
                  255        265        275        285        295
4877            TTTATCTCCA CATCGGACGA GGCTTATATT ACGGCTCTTA CCTGTACAAA
O. karongae     ---------- t--t------ ---c------ ---------- ----------
O. malagarasi   ---------- ---t------ ---c------ ---------- ----------
O. tanganicae   ---------- ---t------ ---c------ ---------- ----------
O. mossambicus  ---------- ---t------ ---c------ ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                  305        315        325        335        345
4877            GAAACCTGAA ACATTGGAGT CATTCTCCTC CTCCTAACTA TAATAACAGC
O. karongae     ---------- ---------- t--------- ---------- ----------
```

```
O. malagarasi   ---------- ---------- t--------- --------c- ----------
O. tanganicae   ---------- ---------- t--c------ -----g---- ----------
O. mossambicus  ---------- ---------- catcctcctc ---------- ----------

                ....|....| ....|....|
                   355        365
4877            CTTCGTAGGC TACGTCCTCC
O. karongae     ---------t ----------
O. malagarasi   ---------- --------t-
O. tanganicae   ---------- ------tc-a
O. mossambicus  ---t-----t ----------
```