



UPPSALA
UNIVERSITET

UPTEC X 15 007

Examensarbete 30 hp
Juni 2015

DNA barcoding of freshwater fishes in Matang, Malaysia

Anna Fogelström



UPPSALA
UNIVERSITET

Bioinformatics Engineering Program

Uppsala University School of Engineering

UPTEC X 15 007	Date of issue 2015-06	
Author	Anna Fogelström	
Title (English)	DNA barcoding of freshwater fishes in Matang, Malaysia	
Title (Swedish)		
Abstract	DNA barcoding is a fairly recently developed method for species identification at the molecular level, often using a short DNA sequence from the mitochondrial genome. In this study DNA barcoding was applied to establish a framework at University of Malaya for rapid and efficient identification of freshwater fish of Malaysia instead of relying on morphological techniques that require high level of taxonomical knowledge and field experience. The results, when applying the established framework, indicate that the interspecific divergences calculated for our target group are clearly greater than intraspecific variations within the group, i.e. a barcode gap seems to exist. Indications are hence that <i>cytochrome c oxidase subunit 1</i> sequences as barcodes could be effective tools for rapid identification of freshwater fishes and results could provide a good start for researchers at University of Malaya to start build a reference library of barcode sequences.	
Keywords	DNA barcoding, freshwater fish, COI, coxI, BOLD, mitochondrial DNA,	
Supervisors	Amir F Merican University of Malaya	
Scientific reviewer	Mikael Thollesson Uppsala University	
Project name	Sponsors	
Language	Security	
ISSN 1401-2138	Classification	
Supplementary bibliographical information	Pages 53	
Biology Education Centre Box 592 S-75124 Uppsala	Biomedical Center Tel +46 (0)18 4710000	Husargatan 3 Uppsala Fax +46 (0)18 471 4687

DNA barcoding of freshwater fishes in Matang, Malaysia

Anna Fogelström

Populärvetenskaplig sammanfattning

DNA streckkodning (DNA barcoding) har under det senaste decenniet skördat framgång som en effektiv metod för att snabbt artbestämma biologiskt material med hjälp av DNA. Man har sedan länge vetat att varje organism bär sin unika uppsättning av DNA-sekvenser och att DNA-sekvenser därför borde vara lämpliga verktyg för artbestämning. Istället för att förlita sig på taxonomiska metoder eller komplexa analyser av hela genom kan man idag använda sig av en ungefär 650 baspar lång region av den mitokondriella genen *cytokrom c oxidas I (coxl)*. *CoxI* har visat sig extra lämplig för evolutionära analyser eftersom *mtDNA* har hög substitutionstakt, sequensen förändras alltså snabb från generation till generation, vilket leder till att variationen i basuppsättning inom *coxl* mellan näresläktade arter generellt är högre jämfört med variationen mellan individer inom samma art. För att DNA streckkodning ska fungera som metod för artbestämning måste alltså sekvensvariationen vara låg mellan individer inom samma art men hög nog mellan arter för att på så sätt kunna särskilja arter från varandra.

På University of Malaya i Kuala Lumpur, Malaysia, har man länge förlitat sig på taxonomiska metoder för artbestämning av sötvattenfiskar. 2009 beslutade man att påbörja arbetet med att fastställa processer och verktyg för att istället kunna använda sig av DNA streckkodning för artbestämning. Arbetet under detta examensarbete har därför gått ut på att studera och testa etablerade metoder för DNA streckkodning och därmed bidra till deras arbete att mer effektivt artbestämma bland annat sötvattenfiskar.

Resultaten av detta arbete kan sammanfattas i att DNA streckkodning baserat på genen *coxl* förmodligen kan vara en effektiv metod för artbestämning av sötvattenfiskar. Detta då ett så kallat ”barcode gap” kunnat påvisas mellan de sekvenser som analyserats, dvs. signifikant högre sekvensvariation mellan sekvenser från olika arter än mellan individuella sekvenser inom en art. Vidare så har ett arbetsflöde för artbestämning baserat på DNA streckkodning etablerats åt University of Malaya. Brister i antalet sekvenser som analyserats leder dock till att resultaten inte är statistiskt säkerställda. Resultaten kan därför endast ses som en positiv indikation på att det etablerade arbetsflödet förmodligen kan ge framgångrika resultat om det appliceras på större datamängder.

Examensarbete 30 hp

Civilingenjörsprogrammet Bioinformatik

Uppsala universitet, juni 2015

Table of content

List of figures and tables	i
Abbreviations.....	i
1 Introduction	1
1.1 Species identification	1
1.2 DNA barcoding	1
1.3 Purpose.....	3
1.4 Disposition.....	4
2 Materials and Methods.....	5
2.1 Target group for this study	5
2.2 Sampling.....	6
2.3 DNA Extraction, Amplification and Sequencing.....	7
2.4 Sequence analysis.....	11
2.5 Data management.....	14
3 Results	15
3.1 Sample collection and evaluation.....	15
3.2 DNA amplification.....	16
3.3 Sequence analysis.....	17
3.4 Publication - DNA-barcoding database	22
4 Discussion	23
4.1 DNA Barcoding workflow (<i>COI</i> barcode sequences)	23
4.2 Applying framework for species identification of freshwater fish	24
5 Acknowledgements.....	28
6 References.....	29
7 Appendices	32
Appendix A – Sequences per sample based on four bidirectional primers	32
Appendix B – Consensus Sequence Summary.....	39
Appendix C – Matang Specimen Info	45
Appendix D – Matang Specimen Taxonomy	51

List of figures and tables

Figure 1 Mangrove forests in the Sundaland Hotspot.....	5
Figure 2 Vast amount of fish collected during trawling in Matang	5
Figure 3 Fish F1/R1 and F2/R2 primer positions in relation to each other	10
Figure 4 Transitions (α) and transversions (β) definitions.....	13
Figure 5 Example of agarose gel of DNA extracted from fish specimen samples.	16
Figure 6 Example of agarose gel of PCR amplicons from DNA of samples.	16
Figure 7 Intra- and Interspecific Kimura-2-parameter distances for all samples.....	20
Figure 8 Intra- and Interspecific Kimura-2-parameter distances amongst unique sequences.....	21
Figure 9 NJ evolutionary relationships between the 9 unique <i>COI</i> sequences.	22
Figure 10 Workflow for DNA barcoding at UM	24
Table 1 Specimen samples generating successful <i>COI</i> barcodes and related taxonomy.....	17
Table 2 E-values and estimated sequence identity scores for bi-directional sequences.	18
Table 3 Four (4) highest BLAST sequence identity scores for consensus sequence CV15.....	20

Abbreviations

BOLD	Barcode of Life Database
bp	Base pair
CBOL	Consortium for the Barcode of Life
<i>COI</i>	Cytochrome c oxidase subunit 1
DNA	Deoxyribonucleic acid
iBOL	The International Barcode of Life project
<i>mtDNA</i>	Mitochondrial DNA
PCR	Polymerase chain reaction
K2P	Kimura two-parameter
UM	University of Malaya

1 Introduction

1.1 Species identification

Accurate species identification has long been dependent on morphological analyses performed by taxonomists. It is however known that morphological approaches to species identification have limitations, mainly since morphological similarities between closely related organisms create challenges to discriminate them from each other. In addition to this, taxonomists believe that only about 1.7 million species of the estimated 4-11 million that live on this planet have been morphologically identified and catalogued [1]. Other challenges relates to determination if similar specimens from different habitats are different species or simply look different due the adaptability of an organism to changes in its environment [2]. Morphological identification methods are also often dependent of gender and life stages of the species [2], this may lead to difficulties in recognition of for example juvenile specimens. Other studies show problems of regulatory matter. Rasmussen *et al.* [3] highlights the problem with identifying commercially important salmons at fish markets in North America due to the similar appearance of fillets from different salmon species. How can it be proved that a salmon fillet is sold to the right price when it is hard to perform a correct identification of the fish? Tools and processes for speeding up and improving methods for species identification are hence needed.

Opportunities to use DNA sequence data and refine molecular methods to support species identification have been investigated for long since it is known that each organism carries its own unique genetic setup of DNA. Already 30 years ago ribosomal DNA was used to establish relationships among species [4] and today there are well-established methods for aligning, comparing and establishing relationships among species using DNA sequences. Databases containing millions of DNA sequences from species all over the world are common and provide valuable tools for comparing and analysing relationships between DNA sequences.

Sequencing the whole genome of an organism is however still both time consuming and expensive and not one standardised way of identifying animals based on shorter regions in the genome exists. Researchers such as Herbert and others [2] are aiming to find new shortcuts to species identification and suggest that DNA barcoding could be one approach to address these concerns.

1.2 DNA barcoding

In 2003 Paul Hebert and his group presented DNA barcoding as a ground-breaking and rapid tool for identifying species in the article “Biological identifications through DNA barcodes” where DNA barcoding is described as method using a short DNA sequences from the mitochondrial genome for species identification at molecular level [2]. Their theory is easy to understand. First, find a DNA region that is short and suitable for rapid species identification, to serve as a species “barcode”. Then create a standardised

framework for capturing these short sequences from DNA samples and create a library containing such barcodes for every species in the world. Unknown DNA sequences can then be compared to already known barcodes with the help of this library [5].

Herbert and others hope that DNA barcoding can provide a new tool for species identification supplementing existing knowledge and also help non-experts to make a quick identification of a sample instead of always having to rely on morphological techniques that require high level of knowledge and field experience [6].

Today there are barcoding initiatives ongoing all over the world and most of them are in contact with The Consortium for the Barcode of Life (CBOL). CBOL is devoted to developing DNA barcoding as a standard process for species identification [7]. Their main tool in this process is the work with The International Barcode of Life project (iBOL). The major goal of iBOL is to create a DNA barcode reference library for all species; this implies it being the largest biodiversity genomics initiative in the world, collaborators in 25 different countries are currently involved in reaching the goal to have barcodes 500,000 different species in 2015 [8].

There are also many subgroups emerged from CBOL. The Fish Barcode of Life campaign (FISH-BOL) is one such group with the goal to gather fish barcodes in world in one huge database that will serve as a standardised reference library for fish species [9]. Bee-Bol, i.e. barcoding of bees, is another [10] example and the Sponge Barcoding Project a third example [11].

1.2.1 Required characteristics

DNA barcoding is based on using a short DNA sequence consisting of a mix of the four different nucleotides for comparison between species. Nucleotides within the sequence can be combined in a huge amount of ways, for example just a 15 bp long sequence allows for over a 1 billion different possible combinations of the nucleotides that can be used as distinct barcodes [2] to distinguish one species from others. In animals, it is suggested that a suitable barcode consists of approximately 600-800bp [12]. A shorter sequence might not provide enough data to allow for sufficient variation among nucleotides.

Establishing suitable barcodes is however more complicated than just finding a region in the genome having sufficient length to allow for enough variation among the nucleotides. As mentioned above, the short DNA sequence suggested by Herbet *et al.* [2] is part of mitochondrial DNA (*mtDNA*) and not nuclear DNA, the reasons are:

- Sequence evolution is fast in *mtDNA* (high substitution rate) [13].
This results in an increase of differences between closely related species and less sequence variation within species [14]. Compared with nuclear DNA, *mtDNA* will require a much shorter nucleotide sequence in order to differentiate species due to this. In order for DNA barcoding to work sequence variation must be high enough between species so that they can be distinguished from each other but low within the species [15].

- There are a large number of mitochondrial gene copies in every cell [16]. Amplification of the genes will generate a greater deal of DNA compared to using nuclear genes that only exists in one copy [17]. This is especially helpful when dealing with small or damaged samples.
- The genes are strongly conserved among animals [18]. This makes it possible to find a barcode gene that is universal in all animals.
- Limited recombination since *mtDNA* is maternally inherited in most animals [18]. Little recombination promotes the loss or fixation of *mtDNA* haplotypes [17]. This reduces diversity within species and makes species identification with molecular markers more successful.
- Fairly conserved flanking regions [19]. In order for universal primers pairs (used for large scale taxonomical applications) to be designed and developed the barcode sequence must be flanked by regions that are well conserved among species [19].

To summarize it, an ideal DNA barcode should i.) provide maximal discrimination among species and ii.) be easily generated with a single primer pair in bidirectional sequencing.

A short 648 base-pair region in the *mtDNA* cytochrome c oxidase 1 gene (*coxI*) and its resulting polypeptide (*COI*) qualifies to serve as a practical and standardized DNA barcode for all animals [20] in accordance with the above characteristics. The gene has successfully served as a barcode in many different animals such as birds, fish and insects [4, 14 & 21] mainly because it is well conserved, showing low levels of variance, within a species but it is at the same time showing enough divergence between species to allow for differentiation among many different species [2].

1.2.2 The “barcode gap”

As understood from previous sections in order for DNA barcoding to be a suitable tool for species identification there must be enough variation in sequences between different species (interspecific variation) so that they can be discriminated from each other. At the same time there must be little variation in sequences from individual samples within a species (intraspecific) to enable assignment of these sequences to the same species.

A threshold called a “barcode gap” has been defined and applied to examine whether or not sequence divergence is much larger among species than within species. A barcode gap exists if there are interspecific divergences that are clearly greater than intraspecific variation [2]. DNA barcoding becomes less effective as a tool the more overlap there is between the intra- and interspecific divergences, i.e. none or a very small “barcode gap” is identified [22].

1.3 Purpose

Barcode projects have been initiated all over the world and the field is constantly growing [55, 20, 5]. In Malaysia however, the field have not been deeply investigated

and not many barcoding projects have been undertaken. The purpose of this degree project was to support the Malaysian progress by initiating a DNA barcode project. This included i.) setting up a framework at UM for the workflow of barcode projects and ii.) test this DNA barcoding framework for species identification of Malaysian freshwater fishes.

1.4 Disposition

First, an overview of the process of DNA barcoding shall be outlined followed by the detailed methods applied on the samples used for this study. Finally the results by applying the process are presented. The intended readers of this report are future barcode students at UM and those who intend to continue the work that has been initiated. The methods applied have been described with the intention to get an initial understand of well-known DNA sequencing and bioinformatics tools but not to give the reader a deeper knowledge. The reader should read this report as a handbook for DNA barcoding of fishes.

This thesis is divided into the following two sub-projects.

1. **Field- and labwork;** covering activities related to specimen collection and DNA Sequencing to obtain DNA barcode sequences from collected specimens.
2. **Data management and analysis;** covering activities related to species identification and establishing relationships within and between species based on retrieved sequences along with database population of identified sequences as species identifiers.

2 Materials and Methods

2.1 Target group for this study

Malaysia is a country with high biodiversity. The country itself and the oceans surrounding it is a part of the Sundaland Hotspot which is defined as one of the world's 34 biodiversity hotspots [23]. The biodiversity is high in this region because of its wide range of different habitats, ranging from large rainforests to coral reefs and from muddy shores surrounded by mangrove forests and swamps to mountain forests. The region is particular rich in different species of freshwater fish. In 2007 there were about 1,000 known species of freshwater fish in the Sundaland Hotspot and about 200 new species have been discovered since 1997 [23].

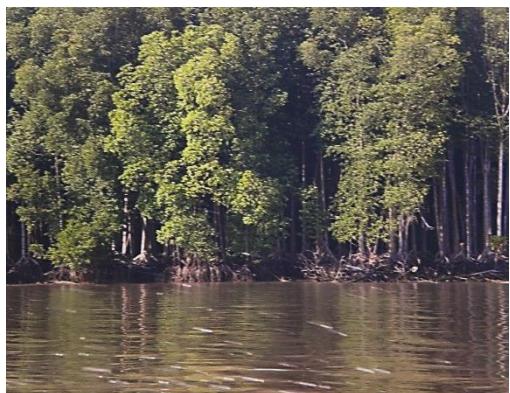


Figure 1 Mangrove forests in the Sundaland Hotspot



Figure 2 Vast amount of fish collected during trawling in Matang

Today nearly 10% of the DNA sequences available in the BOLD reference library comprise marine and freshwater fish species [9]. One of the first comprehensive and successful species identification studies applying DNA barcoding to freshwater fishes was published in Canada in 2008 [24] and after this several others also have followed in for example Brazil [26 & 27], Antarctica [50], Portugal [51], US [53] and Mexico and Guatemala [25], all showing successful results of species identification using the *coxI* gene for fishes.

This vast biodiversity initially made the amount of possible target groups for this project large. UM has however invested in several research projects related to the Sundaland Hotspot region and in particular in species biodiversity and habitants of one mangrove forest in Matang, a region situated in the Malaysian state Perak. As a result of deep knowledge at UM regarding species taxonomy and different ecosystems of freshwater fish in

this region it was decided that Freshwater fish of the mangrove forest in Matang shall be the target group for this thesis.

This hopefully implies easy identification of collected fish and also allows for simplified selection of relevant fishes for the study since for example cultured fishes and other fishes that are not natural habitants in this environment easily can be identified and removed from the study.

2.1.1 *The region*

The mangrove forest in Matang is situated in the Malaysian state Perak. Perak is bordered by Thailand in the north and the Strait of Malacca (a stretch of ocean between Peninsular Malaysia and the Indonesian island Sumatra) in the west. Mangrove forests are described by the World Wide Fund for Nature (WWF) as a unique type of ecosystem often found along sheltered coastlines. Here they grow in saline soil and brackish water and the trees in mangrove forests are familiar to many because of their characteristic root and leaf structures [28].

The region also is of great interest because the presence of mangrove areas is decreasing in Malaysia and today Mangrove forests are one of the world's most threatened tropical ecosystems. More than 35% of the world's mangroves are already gone [29]. According to WWF Mangrove forests are very important for retaining a stable ecosystem. In Malaysia they protect coastlines, prevent saltwater from intruding into rivers and are an important breeding ground for a great amount of species [28].

Many also believe that developing countries such as Malaysia benefit by sustaining mangrove areas rather than clearing them to make room for infrastructural projects or shrimp farms. This since they are important for the fishing industry and serve as a valuable tourist attraction. On the west coast of Peninsular Malaysia, about 50 % of all fish landings are associated with mangrove areas [28]. Highlighting work in mangrove areas could support sustainability since it increases knowledge about this particular ecosystem. It hopefully helps people understanding that destroying them effects the whole ecosystems of the country and such activities could also have negative economic effects long term.

2.2 Sampling

2.2.1 *Selection rules*

There are 1823 different fish species identified in Malaysia out of which 612 are freshwater fishes [30] and several of these freshwater fishes are found the in mangrove area of Perak. The following selection rules for fish sample collection were identified to ensure that only relevant samples were collected for this study:

1. For validation purposes it is crucial that it is possible for others to replicate results and collection should hence avoid rare species in the region. This could complicate ability to replicate this study. The first selection rule was set to:

Target collection of fish samples common to the area

2. Due to sensitivity of DNA extraction and PCR it is difficult to predict if methods applied for DNA management would generate successful result for all types of fish species. Fish sample collection should hence focus on different types of species and not be limited to a small set of species. The second selection rule was set to:

Target collection of samples from at least 15 different species

3. Further, sequences need to be generated from more than one sample of the same species in order to ensure that it can be reviewed if DNA barcoding is a suitable mechanism for species identification: The last selection rules was hence set to:

Target collection five (5) samples from each of the 15 species

In addition to the above rules, fishes need to be killed and managed in a DNA friendly matter to avoid negative impact on the quality of the DNA. All samples shall be kept on ice during the time between collection of the specimen and tissue (muscle or fin) sampling to avoid degradation of the DNA and the DNA sampling itself needs to be performed in ways to avoid DNA contamination. Tissue samples shall also be kept frozen and preserved in ethanol until further analysis to avoid degradation of the DNA.

Required information regarding the collection of samples shall be gathered on the collection sites (such as country, region, GPS-coordinates and photographs) for validation purposes and to start manage the sample in accordance with standards agreed by BOLD.

Expert taxonomists should be involved in validation of identity the voucher specimens and to provide additional information about the species to further ensure that the specimens collected are identified correctly at the sampling site and in accordance with the above rules.

2.3 DNA Extraction, Amplification and Sequencing

Generating DNA sequences that are intended for use as barcodes involve well established and standardised methods used in almost all types of DNA sequencing projects today. There are also several different ways of extracting DNA and different PCR cycles to use in generation of DNA sequences from animal specimens.

This thesis project relies on the knowledge and methods already used amongst the molecular scientists at UM. Their lab protocols have proven successful in previous projects with focus on sequencing fish DNA and are therefore followed to a great extent when working with these samples which are of similar character

2.3.1 DNA extraction

The *coxI* gene has been accepted as a standard barcode and there are now several standardized extraction protocols for DNA barcoding [31]. Ivanova and others [31] suggest using silica-based method for DNA extraction in fishes and such extraction methods have now been used in several fish barcoding projects [3, 4 & 24].

DNA shall be extracted from the tissue samples by using extraction protocols that are known to be appropriate for both muscle and fin organelles. Researchers at UM have successfully extracted highly purified DNA from fishes collected in Matang and the using and the following method shall hence be applied in this study as well:

Digestion - Step 1 of the extraction procedure

Prepare Eppendorf tubes labelled according to the samples used and take approximately 0.1 g of fresh or ethanol-preserved tissue and insert it into the corresponding Eppendorf tube. Insert 300 µl extraction buffer + 2 % SDS into each tube and add 30 µl proteinase K (protK) into each tube. Place tubes in water bath at 37°C overnight.

Extraction - Step 2 of the extraction procedure performed in a fume chamber

Add 330 µl of Phenol into each tube and shake the samples a few seconds by hand and then rotate them for 10 minutes. Spin the samples in microfuge for 30 seconds at 12000 rpm and add 330 µl of Chloroform into each tube post spinning. Shake the samples again for a few seconds by hand and then rotate them for 10 minutes in a microfuge at 12000 rpm.

The samples should now have two layers where the upper layer contains DNA. Remove the upper layer of each sample with a cut tip to avoid mixing the DNA and place it in a new labelled Eppendorf. Add 990 µl of Absolute Ethanol (ice cold) into each new tube to wash the DNA and shake the samples by hand (should see DNA). Place the samples to incubate in a freezer (-20°C) over night or at -80°C for at least 30 minutes.

Spin the samples the day after incubation in a microfuge for 10 minutes at 12000 rpm and then pipette off as much of the alcohol as possible. Centrifuge the tubes for 3 minutes at 12000 rpm and pipette off as much of the alcohol as possible without disturbing the DNA at the bottom. Cover tubes with foil and place in vacuum airer or oven at low temperature for approximately 15 minutes but be careful not to over dry the pellet.

Add 100 µl of filtered TE buffer with PH 7.2 into each tube and then flick them to liberate the pellet from the bottom of the tube and leave the tubes in fridge to dissolve overnight.

Confirm presence of DNA - Step 3 of the extraction procedure

Prepare an agarose gel with enough wells for all samples plus loading buffer and HIND III and add approximately 0.5µg of solution from each of the above samples and place it in one of these new Eppendorf tubes. Stain the 0.5µg samples with ethidium bromide (binds to DNA) and add loading buffer to each sample. Load the dyed samples on the agarose gel along with one well loaded with loading buffer and one with HIND III (known molecular weight ladder that will visualize 8 fragments from 125 to 23 130 bp).

Fix the geltray on voltage 60 mA to start migration and separation of DNA fragments in the solution, the shortest DNA fragments will move faster than the longer fragments which imply that fragments of DNA in each sample will be separated depending on their length. Determine presence of DNA by measuring the intensity of light absorbance of each sample of loaded DNA solution with a spectrophotometer. DNA absorbs UV light at 260 and 280 nanometres. Visualize the gel using a UV transilluminator and capture the results on photo.

2.3.2 Amplification - Polymerase chain reaction (PCR)

PCR is one of the most well-known methods for purifying DNA regions. One of the major limitations of PCR is however that at least one part of the sequence to be amplified must be known [32] in order to start amplification and the most crucial parts of DNA barcoding is often to design appropriate primer because small changes might have a large effect on barcode recovery so the first phase of DNA barcoding should involve identification of suitable primers [31]. The *coxI* region of *mtDNA* is highly conserved which makes design of universal primers and hence amplification possible.

There are several different amplification protocols to use for PCR. This study shall follow the protocols for *COI* amplification set up by Natalia Ivanova and Chris Grainger [33] for Canadian Centre for DNA barcoding (CCDB) as outlined below. CCDB is in collaboration with CBOL and their protocols are widely used in attempts to amplify barcode regions of approximately 655bp from the 5' end of the *coxI* gene using a mammal primer cocktail [34] and different combination of primers. The procedure for *coxI* amplification shall be based on the following basic recipe for PCR [34], please note that the 10% trehalose in the original recipe is replaced by ddH₂O:

Component	Amount
ddH₂O (autocalved water)	8.25 µl
10x Buffer	1.25 µl
50 mM MgCl₂	0.625 µl
Forward primer (10 µM)	12.5 µl
Reverse primer (10 µM)	12.5 µl
dNTPs (10 mM)	1 µl
Taq polymerase (5 U/µl)	0.25 µl
TOTAL	10.5 µl
Template DNA	2 µl per well

When mixing the PCR mixture the components presented above shall be mixed in the following order: ddH₂O – Buffer - MgCl₂ – dNTPs - Primers – Taq. The preparation of the mix shall be performed on ice with all components were kept cool (in order to prevent premature start of the reactions).

Bidirectional sequencing, i.e. using a combination of forward and reverse primers as reads, normally provides higher consensus sequences compared to unidirectional

sequencing and allows for independent confirmation of sequence information in two directions, not only one. The sequencing signal often becomes unclear near the end of a reading frame and bidirectional sequencing can help ensure that a clear signal from both ends is made available which potentially improves the generation of full-length sequences [31].

During the last years the following four bidirectional primers, or a combination of these, have successfully been applied in studies involving DNA barcoding of fish and will hence be used in this study as well [3, 4, 24 & 34]:

FishF1: 5' TCAACCAACCACAAAGACATTGGCAC 3'

FishR1: 5' TAGACTTCTGGTGGCAAAGAATCA 3'

FishF2: 5' TCGACTAACATAAAGATATCGGCAC 3'

FishR2: 5' ACTTCAGGGTGACCGAAGAATCAGAA 3'

The above four primes will all anneal to the same positions of the target sequence the fish gene which implies that the sequences generated will be interpretable and comparable when analyzing the sequences retrieved [34] in accordance with the below picture.

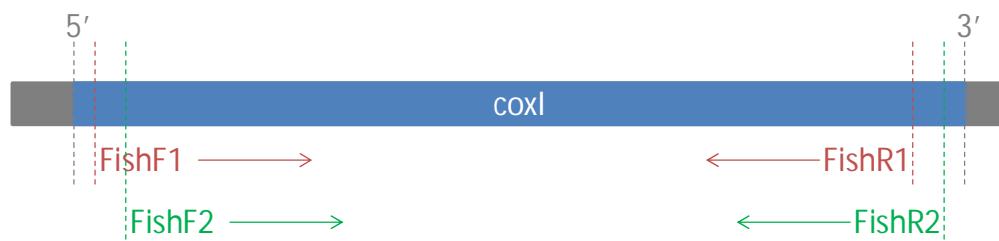


Figure 3 Fish F1/R1 and F2/R2 primer positions in relation to each other

The following PCR cycle [4] shall be applied in this study:

Number of cycles	Temperature	Time
1	95°C	2 min
35	94°C	30 sec
	54°C	30 sec
	72°C	1 min
1	72°C	10 min
HOLD AT:	4°C	

Upon completion of PCR it is important to make sure that the quantity of PCR product is enough for DNA sequencing [31]. Once the sequences are amplified and purified the resulting products shall therefore be analysed using agarose gel and electrophoresis. Presence of successful PCR products was determined by measuring the intensity of light absorbance of each sample of loaded DNA solution with a spectrophotometer using the same method as described above under DNA Extraction. Only the selected region will generate one single band during the electrophoresis if the PCR has amplified successfully [32].

2.3.3 DNA Sequencing

Successful amplification of DNA relies to a great extent on using purified high-quality DNA in the experiments. To prepare samples for sequencing the next step of the process is hence to purify, i.e. cleaning the DNA from nucleotides that are not part of the sequence and residual primers, the PCR products. The UM researchers have been using the QIAquick PCR Purification Kit for purification of up to 10 µg PCR products in other successful DNA extraction projects involving fish. It was therefore decided that this shall also be applied in this study. The kit includes all columns, buffers and collection tubes required and a purification process protocol [56].

There are various methods and instruments to be used for carrying out DNA sequencing post purification and choosing technique normally depends on the scale of the project and on the volume of sequencing reactions [31]. In this case, samples that successfully generates PCR products (indicate a single visible band) shall be sent together with the forward and reverse primers used in the PCR to 1stBase in Kuala Lumpur [35] for bidirectional sequencing with the forward and reverse primers described above. This company generates sequences for samples provided using BigDye® Terminator v3.1 cycle sequencing kit from Applied Biosystems Inc., ABI. The resulting fluorescent signals will be recorded as an electropherogram and provided in .ab1 format together with corresponding sequences in .seq fasta format.

2.4 Sequence analysis

The methods for analysing DNA data described in this section shall be applied in this study. The focus of these methods shall be to describe the evolutionary relationship between species by performing phylogenetic analyses and measuring the “barcode gap”. This shall provide an estimate of the ability of the *coxI* gene sequence to serve as a barcode for freshwater fish.

2.4.1 Individual sample management

All samples successfully sequenced will generate four sequences, i.e. one per primer, and shall be analysed and edited both automatically and manually.

First, sequence identity shall be reviewed by searching GenBank using algorithms built into the Basic Local Alignment Search Tool (BLAST). BLAST helps estimate similarity between the untrimmed barcode sequence records retrieved from sequencing

and sequence records already existing in the comprehensive Genbank database. Similarity between sequences is calculated using built in algorithms for local sequence alignments. These algorithms are based on statistical significance of matches expressed as Expect value (E-value) and Identity Scores (Ident). The E-value provides an estimate of the number of hits one can "expect" to see by chance when searching the database, the lower the E-value the more significant the match is of a particular size, and it decreases exponentially as the score of the match increase and the Ident value gives an estimate of the degree of correspondence between two sub-sequences (no gaps between the sequences) where a high identity score implies high similarity of structure or function [36]

All four sequences per sample will also be manually trimmed into the same length and aligned, reverse sequences complemented into forward, using the ClustalW method and the software MEGA 6 [37]. Further, to obtain one *coxI* sequence per sample a consensus sequence needs to be created from the four sequence records using MEGA 6. The consensus sequence will be established based on algorithms creating nucleotide patterns where each position in the sequence pattern represents the nucleotide that is most likely to occur, i.e. occurs most frequently at this position in the sequence [38].

2.4.2 Between species analyses

Once you have several sequences to analyse the first step of performing multiple alignments will be to align the sequences that shall be compared based on similarities. Challenges in this process normally relates to difficulties to determine the most likely alignment among dissimilarities and several combinations of gaps, matches and mismatches caused by indels and point mutations and several known software using predefined algorithms are often used to produce multiple alignment of sequence records. Multiple alignments in this study will be carried using the MEGA 6 software with Bootstrapping of 1000 replicates which is based on the Needleman-Wunsch dynamic programming algorithm [39] to maximize the number of matched nucleotides and select the alignments that generates that highest match score based on sequence similarity.

The evolutionary distance between the sequences (sequence divergence) will also be estimated once the sequences are aligned using substitution models that apply algorithms based on differences between sequences. The Kimura 2-parameter distance model (K2P) is up until now assessed as the most effective substitution model to use when calculating small evolutionary distances between short sequences with few substitutions such as the *coxI* sequence [2]. The K2P method is particular efficient because it aims at increasing accuracy of the sequence divergence calculation by taking the non-randomness of substantiations into account without complicating the calculations [40]. K2P assumes equal composition of base frequencies ($\pi = \pi^A = \pi^T = \pi^G = \pi^C = 1/4$) in sequences which is a simplification of the reality, but takes mutation frequencies, distinguishes between transitions (α) and transversions (β), into account [40].

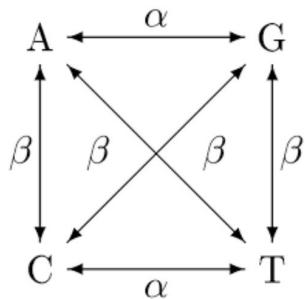


Figure 4 Transitions (α) and transversions (β) definitions.

To calculate sequence divergence, d , using K2P the following parameters are defined:

- P: observed number of transitional differences between the sequences
- Q: observed number of transversion differences between the sequences
- $w_1 = 1 - 2P - Q$ (1)
- $w_2 = 1 - 2Q$ (2)
- $d = -(1/2)x\ln(w_1) - (1/4)x\ln(w_2)$ (3)

Intra- and inter specific variation should be compared and expressed as the ratio of the two values sequence divergence among sequences from the same species, d_{intra} , and sequence divergence among sequences from different species, d_{inter} , when using the model to calculate the barcode gap [57].

- $K = d_{\text{intra}}/d_{\text{inter}}$ (4)

K indicates how different the sequences are. High genetic variation between the sequences would result in a high K -value and implies that there are many differences between the sequences (lineages are distinct). A low K -value means that it is hard to observe differences between the sequences. i.e. two distinct lineages cannot be found.

Further, a phylogenetic tree of sequence distances is normally constructed to help demonstrate a cluster of closely related sequences and is particularly useful in case of absence of a barcode gap (low K -value). There are several different types of methods of construction phylogenetic trees, Neighbor-Joining (NJ) phylogenetic tree [41] being one of the most widely used method since it is fast and easy to apply on large data sets. The NJ method also allows for bootstrapping, a tool for assigning measures of precision to sample estimates such as phylogenetic trees, in an easy way. NJ is a bottom-up clustering method that uses a distance matrix approach to display evolutionary distance between aligned sequences. Distances are often expressed as “the number of nucleotides that differ between aligned sequences using an iterative process of finding and joining pairs of aligned sequences that that gives the smallest sum of branch lengths until the tree is completely resolved and all branch lengths are known” [41].

The MEGA6 software shall be applied for calculation sequence divergence both within and between species in accordance with the above described methods and also to create rooted NJ trees of these estimated K2P distances that provide graphic representations of relationships between the *coxI* sequences.

2.5 Data management

Once DNA barcodes are generated at a lab they are most often used to be matched with other existing barcodes or to genetic material from voucher specimens by uploading them to for example GenBank or other databases on the Internet. Before this is possible, researchers at UM believe it is necessary to develop a workflow and tools for work with DNA barcoding. One of the main parts of a DNA barcoding project is data management and ensuring that information needed for barcode validation is provided in an appropriate manner. To facilitate these steps of the DNA barcoding workflow a database containing information gathered during barcoding work is therefore needed at UM.

CBOL provides a tool called Barcode of Life Database (BOLD) where users can upload barcode projects. BOLD have three main components. One part is based on a standardized framework that helps the user to ensure that all necessary data is provided in a barcoding project [42]. It also helps users to manage and analyse their barcode data. The main goal with BOLD is to create a reference library of DNA barcodes based on an identification engine. This part can be used to assign unknown specimens to already known ones based on the DNA barcodes [42]. Some other minor barcoding projects use other databases as well to help facilitate their work. FISH-BOL, for example, uses a database called FISH-BASE to help associate fish and corresponding taxonomic and environmental data with the aim to barcode all fishes over time [54]. There are databases developed apart from CBOL. One example is BioBarcode, a Korean initiative to serve as a platform for Asian biodiversity. They work in collaboration with CBOL and have agreed upon the same standards as CBOL but have decided to build their own database and tools to facilitate barcoding work in the Asian region [43].

At UM the goal is to join BOLD as soon as possible and the database developed in this project will serve as an internal tool to organize barcode data according to the standards set up by CBOL. The last step of a barcoding project is for the barcode to be approved and made public. This is most often done by adding the sequence to BOLD or to GenBank which is established by the National Center for Biotechnology Information, NCBI.

3 Results

During these six months in Malaysia a workflow for DNA barcoding was established along with the DNA barcode research performed. This workflow will be presented below along with the results from the actual DNA barcoding activities performed.

3.1 Sample collection and evaluation

Two field trips to Matang were accomplished during the course of the project. Trip A lasted for 3 days and two nights and trip B lasted for two days and one night. Fish was caught with support from local fishermen and their supporting equipment during both trips using trawls at five different locations in the mangrove area. The five locations were selected in order to collect as many specimens as possible amongst the most common species in the area.

A vast amount of fish was caught during trawls and, with initial support from the local fishermen, we managed to identify and collect fishes of interest based on our three selection rules. During trip A we collected 79 specimens from 23 different species and during trip B the collections resulted in 35 specimens from 14 different species. In total 114 specimens from 30 different species were collected (please refer to Appendix C Matang Specimen Info and Appendix D Matang Specimen Taxonomy for details).

All 114 specimens were morphologically identified by morphological specialists at UM and then labelled and archived in a museum collection at UM. The 30 species are distributed across 1 class, 6 orders, 19 families and 25 genera (please refer to Appendix D – Matang Specimen Taxonomy).

Fin and tissue samples were collected from the fishes immediately after each trawl to avoid contamination and DNA degradation to an as great extent as possible. Each fin and tissue sample was labelled as well as the corresponding fish specimen. Samples were then kept in 99.7 % ethanol and on ice until initiation of DNA extraction.

3.2 DNA amplification

DNA extraction was performed on 65 out of the 114 specimens out of which 61 successfully generated products containing DNA (53.0% of initial reference samples).

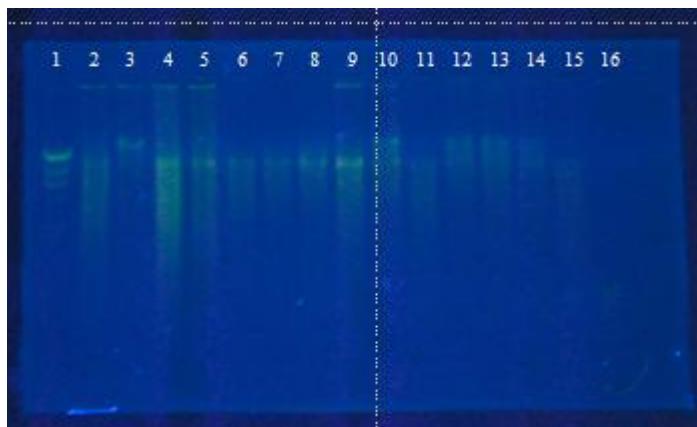


Figure 5 Example of agarose gel of DNA extracted from fish specimen samples.

Lane 1: Hind III DNA ladder and lane 2-15: samples of extracted DNA. This example illustrates that DNA fragments are to be found in each one of the 14 samples visualized on this gel.

The barcode region of the *coxI* gene (*COI-5P*) was PCR amplified using all combinations of the four universal fish primers in 45 out of these 61 samples. Visualization of the samples on agarose gel determined that 31 of the samples (70% of samples amplified) had successfully generated PCR products of an approximate length of 650-bp fragments.

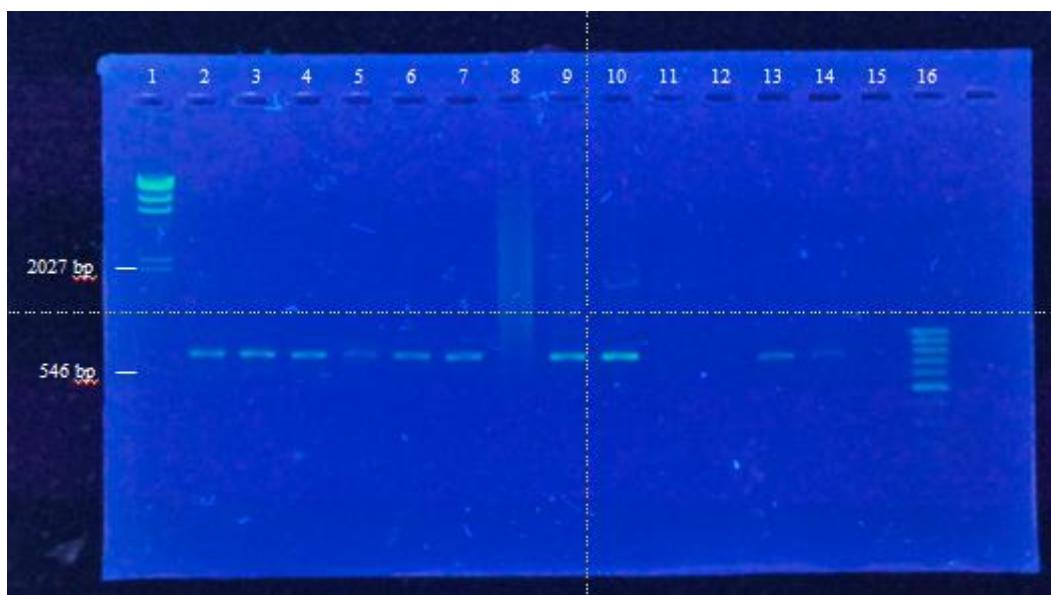


Figure 6 Example of agarose gel of PCR amplicons from DNA of samples.

Above illustrated samples are derived from primer pair F1, R1, F2 and R2. Lane 1: Hind III DNA ladder (size marker), lane 2-15: samples of extracted DNA and lane 16: low range DNA Ladder with 50-1000 bp. This example illustrates that DNA fragments of an approximate length of 650-bp are to be found in of the 10 samples visualized on this gel. Lanes 8, 11, 12 and 15 lack bands, indicating unsuccessful PCR amplification.

Due to economic constraints at UM, only the following 12 samples out of the above successfully extracted and amplified samples could be sent for sequencing post purification of the samples using the QIAquick PCR Purification Kit. All of the 12 samples amplified successfully during sequencing (please refer to Appendix A – Sequences per sample based on four bidirectional primers for detailed sequencing results). This implies that we successfully have obtained mitochondrial *COI* barcodes for 12 specimens belonging to 1 class, 5 families, 4 orders, and 5 species in 5 genera (Table 1 below).

Table 1 Specimen samples generating successful *COI* barcodes and related taxonomy.

Voucher ID	Species	Order	Family
DV 9	<i>Batrachomoeus trispinosus</i>	Batrachoidiformes	Batrachoididae
DV 7	<i>Batrachomoeus trispinosus</i>	Batrachoidiformes	Batrachoididae
DV 8	<i>Batrachomoeus trispinosus</i>	Batrachoidiformes	Batrachoididae
DV 10	<i>Scatophagus argus</i>	Perciformes	Scatophagidae
CV 5	<i>Arius maculatus</i>	Siluriformes	Ariidae
CV 6	<i>Arius maculatus</i>	Siluriformes	Ariidae
DV 13	<i>Pomadasys argenteus</i>	Perciformes	Haemulidae
CV 15	<i>Pomadasys argenteus</i>	Perciformes	Haemulidae
DV 15	<i>Pomadasys argenteus</i>	Perciformes	Haemulidae
DV 14	<i>Pomadasys argenteus</i>	Perciformes	Haemulidae
FV 3	<i>Trixiphichthys weberi</i>	Tetraodontiformes	Triacanthidae
FV 7	<i>Trixiphichthys weberi</i>	Tetraodontiformes	Triacanthidae

3.3 Sequence analysis

3.3.1 Sequence diversity

GenBank analyses indicate that all sequenced samples have been correctly identified morphologically based on low e-values and high sequence similarity scores to sequences from the same species already submitted to GenBank as indicated in Table 2 below. Sample CV15 was morphologically identified as *Pomadasys argenteus* but GenBank results for the all bi-directional primer sequences indicates, E-value 0,00 and average similarity score of 90%, high likelihood that this fish specimen was rather more likely to belong to the species *Pomadasys kaakan*.

Table 2 E-values and estimated sequence identity scores for bi-directional sequences.

Sequence (Primer name)	BLAST query result	E-value	Ident
1st_BASE_384227_P_D9_FISH_F1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	95%
1st_BASE_384228_P_D9_FISH_R1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	94%
1st_BASE_384229_P_D9_FISH_F2	No significant similarity found.	n/a	n/a
1st_BASE_384230_P_D9_FISH_R2	<i>Cyanocorax chrysops/cyanopogon</i> voucher MACN-Or-ct 1137 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	3e-17	86%
1st_BASE_384259_P_D7_FISH_F1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	95%
1st_BASE_384260_P_D7_FISH_R1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	93%
1st_BASE_384261_P_D7_FISH_F2	No significant similarity found.	n/a	n/a
1st_BASE_384262_P_D7_FISH_R2	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	93%
1st_BASE_384235_P_D8_FISH_F1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	95%
1st_BASE_384236_P_D8_FISH_R1	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	94%
1st_BASE_384237_P_D8_FISH_F2	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	2e-93	81%
1st_BASE_384238_P_D8_FISH_R2	<i>Batrachomoeus trispinosus</i> mitochondrial DNA, complete genome except for D-loop	0,0	94%
1st_BASE_384267_P_D10_FISH_F1	<i>Scatophagus argus</i> voucher KUT 1829 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384268_P_D10_FISH_R1	<i>Scatophagus argus</i> voucher KUT 1829 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384269_P_D10_FISH_F2	<i>Scatophagus argus</i> voucher KUT 1829 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384270_P_D10_FISH_R2	<i>Scatophagus argus</i> voucher KUT 1829 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384239_P_C5_FISH_F1	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384240_P_C5_FISH_R1	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384241_P_C5_FISH_F2	No significant similarity found.	n/a	n/a
1st_BASE_384242_P_C5_FISH_R2	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	100%
1st_BASE_384243_P_C6_FISH_F1	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384244_P_C6_FISH_R1	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384245_P_C6_FISH_F2	No significant similarity found.	n/a	n/a
1st_BASE_384246_P_C6_FISH_R2	<i>Arius maculatus</i> isolate MAM2 cytochrome oxidase subunit 1 (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384271_P_D13_FISH_F1	<i>Stolephorus indicus</i> isolate IOBML48 cytochrome oxidase subunit I-like (<i>COI</i>) gene, partial sequence; mitochondrial	2e-04	100%
1st_BASE_384272_P_D13_FISH_R1	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	96%
1st_BASE_384273_P_D13_FISH_F2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384274_P_D13_FISH_R2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	98%
1st_BASE_384247_P_C15_FISH_F1	<i>Pomadasys kaakan</i> cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	89%

Sequence (Primer name)	BLAST query result	E-value	Ident
1st_BASE_384248_P_C15_FISH_R1	<i>Pomadasys kaakan</i> cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	90%
1st_BASE_384249_P_C15_FISH_F2	<i>Pomadasys kaakan</i> cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	90%
1st_BASE_384250_P_C15_FISH_R2	<i>Pomadasys kaakan</i> cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	90%
1st_BASE_384251_P_D15_FISH_F1	<i>Thamnaconus modestus</i> voucher NSMK-PI-000104 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0002	97%
1st_BASE_384252_P_D15_FISH_R1	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384253_P_D15_FISH_F2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384254_P_D15_FISH_R2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	100%
1st_BASE_384275_P_D14_FISH_F1	No significant similarity found.	n/a	n/a
1st_BASE_384276_P_D14_FISH_R1	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	3e-120	90%
1st_BASE_384277_P_D14_FISH_F2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	99%
1st_BASE_384278_P_D14_FISH_R2	<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (<i>COI</i>) gene, partial cds; mitochondrial	0,0	100%
1st_BASE_384255_P_F3_FISH_F1	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384256_P_F3_FISH_R1	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	88%
1st_BASE_384257_P_F3_FISH_F2	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384258_P_F3_FISH_R2	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384231_P_F7_FISH_F1	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384232_P_F7_FISH_R1	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384233_P_F7_FISH_F2	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%
1st_BASE_384234_P_F7_FISH_R2	<i>Trixiphichthys weberi</i> cytochrome c oxidase subunit I (COXI) gene, partial cds; mitochondrial	0,0	87%

Comparing the four retrieved bidirectional *COI* sequence records per sample (see Appendix A – Sequences per sample based on four bidirectional primers) establishes that all *COI* sequences were longer than 600bp. One *COI* sequence per sample, a consensus sequence, was established and trimmed from each of the four sequences using MEGA6. All consensus *COI* sequences had a length of approximately 660bp (see Appendix B – Consensus Sequence Summary for details) and analysis also confirms that none of these sequences contain indels or stop codons.

Additional analyses and GenBank results of the corresponding consensus sequence for the sample CV15 indicates, E-value 0,00 and average similarity score of 90%, high likelihood that this fish specimen was rather more likely to belong to the species *Pomadasys kaakan* or actually *Pomadasys hasta*, as indicated by the table below.

Table 3 Four (4) highest BLAST sequence identity scores for consensus sequence CV15

Description	Max score	Total score	Query cover	E value	Ident
<i>Pomadasys kaakan</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	841	841	98%	0.0	90%
<i>Pomadasys hasta</i> isolate FSCS052-06 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	833	833	98%	0.0	90%
<i>Pomadasys hasta</i> isolate FSCS051-06 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	830	830	98%	0.0	90%
<i>Pomadasys kaakan</i> voucher GF751 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	826	826	98%	0.0	90%

3.3.2 Sequence comparison

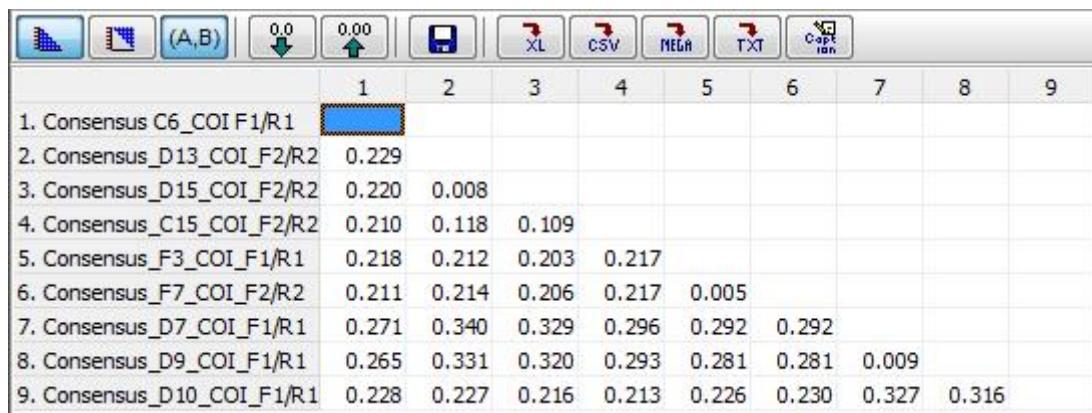
Pairwise multiple alignments of the consensus sequences using MEGA6 indicates that the sequence pairs C5 and C6, D14 and D15 as well as D8 and D9 constitute identical sequence pairs since the genetic divergence between them is being equal to 0 % in the below figure.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Consensus_C15_COI_F2/R2												
2. Consensus_C5_COI_F1/R1	0,209											
3. Consensus_C6_COI_F1/R1	0,209	0,000										
4. Consensus_D10_COI_F1/R1	0,213	0,228	0,228									
5. Consensus_D13_COI_F2/R2	0,118	0,228	0,228	0,227								
6. Consensus_D14_COI_F2/R2	0,109	0,219	0,219	0,216	0,008							
7. Consensus_D15_COI_F2/R2	0,109	0,219	0,219	0,216	0,008	0,000						
8. Consensus_D7_COI_F1/R1	0,296	0,271	0,271	0,327	0,340	0,329	0,329					
9. Consensus_D8_COI_F1/R1	0,293	0,265	0,265	0,315	0,330	0,320	0,320	0,009				
10. Consensus_D9_COI_F1/R1	0,293	0,265	0,265	0,315	0,330	0,320	0,320	0,009	0,000			
11. Consensus_F3_COI_F1/R1	0,217	0,217	0,217	0,226	0,212	0,203	0,203	0,292	0,280	0,280		
12. Consensus_F7_COI_F2/R2	0,217	0,211	0,211	0,230	0,214	0,205	0,205	0,292	0,281	0,281	0,005	

Figure 7 Intra- and Interspecific Kimura-2-parameter distances for all samples.

The table in the picture above contains estimated Intra- and Interspecific Kimura-2-parameter distances amongst all retrieved *COI* sequences and indicates that three sequence pairs constitute identical sequences given that the genetic divergence is equal to 0% within the pair.

Removal of one sequence per identical sequence pair, i.e. CV5, DV 14 and DV8, results in an high average interspecific Kimura-2-parameter distance at 24,73% between species and the remaining *COI* sequences shows low average intraspecific Kimura-2-parameter distances at 0.71% within species with the lowest intraspecific difference being 0.46%. This gives an indication that *COI* sequence differences between closely related species are higher than differences within species.



The screenshot shows a software interface for sequence analysis, specifically MEGA. At the top, there is a toolbar with various icons for file operations like Open, Save, and Print, as well as buttons for sequence selection (A,B), distance calculation (0.0, 0.00), and export formats (XL, CSV, MEGA, TXT, Capture). Below the toolbar is a menu bar with 'File', 'Edit', 'View', 'Analysis', 'Help', and 'About'. The main window displays a table of distance values. The columns are labeled 1 through 9, and the rows are labeled with sequence names. The first row, '1. Consensus_C6_COI_F1/R1', has a highlighted blue cell in the column for itself (labeled 1). The second row, '2. Consensus_D13_COI_F2/R2', has a value of 0.229 in the column for itself (labeled 1). The third row, '3. Consensus_D15_COI_F2/R2', has a value of 0.220 in the column for itself (labeled 1). The fourth row, '4. Consensus_C15_COI_F2/R2', has a value of 0.210 in the column for itself (labeled 1). The fifth row, '5. Consensus_F3_COI_F1/R1', has a value of 0.218 in the column for itself (labeled 1). The sixth row, '6. Consensus_F7_COI_F2/R2', has a value of 0.211 in the column for itself (labeled 1). The seventh row, '7. Consensus_D7_COI_F1/R1', has a value of 0.271 in the column for itself (labeled 1). The eighth row, '8. Consensus_D9_COI_F1/R1', has a value of 0.265 in the column for itself (labeled 1). The ninth row, '9. Consensus_D10_COI_F1/R1', has a value of 0.228 in the column for itself (labeled 1).

	1	2	3	4	5	6	7	8	9
1. Consensus_C6_COI_F1/R1									
2. Consensus_D13_COI_F2/R2	0.229								
3. Consensus_D15_COI_F2/R2	0.220	0.008							
4. Consensus_C15_COI_F2/R2	0.210	0.118	0.109						
5. Consensus_F3_COI_F1/R1	0.218	0.212	0.203	0.217					
6. Consensus_F7_COI_F2/R2	0.211	0.214	0.206	0.217	0.005				
7. Consensus_D7_COI_F1/R1	0.271	0.340	0.329	0.296	0.292	0.292			
8. Consensus_D9_COI_F1/R1	0.265	0.331	0.320	0.293	0.281	0.281	0.009		
9. Consensus_D10_COI_F1/R1	0.228	0.227	0.216	0.213	0.226	0.230	0.327	0.316	

Figure 8 Intra- and Interspecific Kimura-2-parameter distances amongst unique sequences.

The table shows that the within species (intraspecific) distances are i.) 0.46% between the unique *COI* sequences for samples FV3 and FV7 (*Trixiphichthys weberi*), ii.) 0.77% between samples DV13 and DV15 (*Pomadasys argenteus*) and iii.) 0.92% between samples DV7 and DV9 (*Batrachomoeus trispinosus*) resulting in an average Intraspecific distance at 0,71%. The average distance between species (Interspecific) is 24,73%. The average distance within an order, i.e. in this case the distance between sample D10 and D13, C15, D15 respectively, is 22,26%.

Phylogenetic trees were reconstructed based on the remaining 9 unique *COI* sequences. The NJ analysis showed that the *COI* sequences for each of the 5 species were clearly discriminated from each other, i.e. indicating deep intraspecific and interspecific divergences as outlined below. The mean distance in the constructed trees seem to be lower within orders as for example indicated by sample DV10 which belongs to the same order Perciformes as DV13, DV15 and C15.

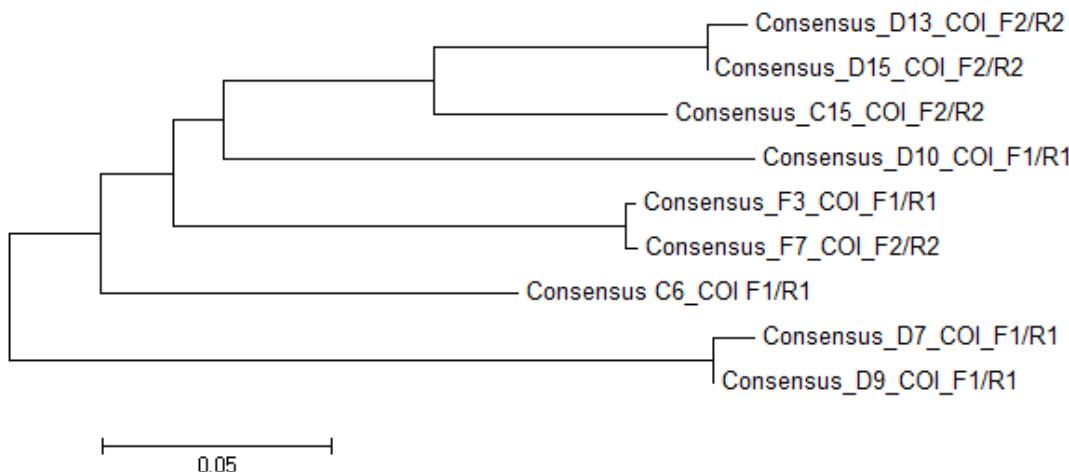


Figure 9 NJ evolutionary relationships between the 9 unique *COI* sequences.

Shown above is the optimal NJ tree illustrating relationships between the remaining 9 unique *COI* sequences calculated using the NJ method with the sum of branch length = 0.68079297 and there was a total of 657 positions in the final dataset.

3.4 Publication – DNA barcoding database

The sequences generated in this study have not been deposited in GenBank. All obtained sequences and related voucher information were stored in a local database at UM containing all required information in preparation for this final step of the DNA barcoding study. This database has remained in house at UM to date in preparation for the DNA barcoding workflow to be established.

4 Discussion

DNA barcoding as a tool for species identification is criticised from various aspects. Some are of more political character and some are directly from questioning scientists. When Herbert published his article in 2003 it gave spark to a debate regarding the importance of taxonomists and whether or not DNA barcoding was here to replace traditional morphological methods for species identification [1]. This position has changed and most researchers in the field of taxonomy now agree that DNA barcoding is a helpful tool in the process to identify and catalogue species. There are still researchers that doubt that a single gene can distinguish all species and refer to the fact that taxonomists that base their conclusions on morphological basis have a set of many different characters, not a single one, to help them recognize species [44]. For plants a combination of at least two barcodes are recommended to ensure identification and some argue that a combination of barcodes would be better for animal barcoding as well.

One of the goals of this study was to i.) set up a framework at UM for the workflow of barcode projects and ii.) test this DNA barcoding framework for species identification of Malaysian freshwater fishes. The results presented above will herein be presented below in light of these two goals.

4.1 DNA barcoding workflow (*COI* barcode sequences)

The steps and protocols applied in this study can be summarized into the following workflow:

1. Select target group for study
2. Collect required specimens/samples (via field work or other applicable repositories of biological material such as museums or herbaria)
3. Perform morphological analysis and identification of samples to ensure correct initial identification
4. If applicable, store corresponding voucher specimens for all samples at appropriate institution
5. Look into whether or not there are applicable barcoding protocols to follow to obtain DNA barcode sequences from the collected specimens
6. Extract DNA and perform amplification of the mitochondrial *coxI* gene of the samples (based on standardized protocols if applicable)
7. Sequence the extracted and amplified DNA samples (send to sequencing institution or perform sequencing at own lab facilities)
8. Perform sequence editing and analyses
 - Manual sequence editing (applicable mainly when bidirectional sequencing is used) to consolidate required information from obtained sequence records
 - Sequence alignment to arrange and prepare sequence records for similarity analyses

- Run BLAST queries for retrieved sequences against online databases (for example GenBank or CBOL) to verify morphological sequence identification based on finding the closest matching reference record to each sequence
 - Calculate interspecific and intraspecific divergences for sequences
9. Publication of sequence records and corresponding identification in applicable database

The following workflow from sample collection to publication of sequence record established during the first month of the project was hence established to support research at UM as a handbook for DNA barcoding of fishes.

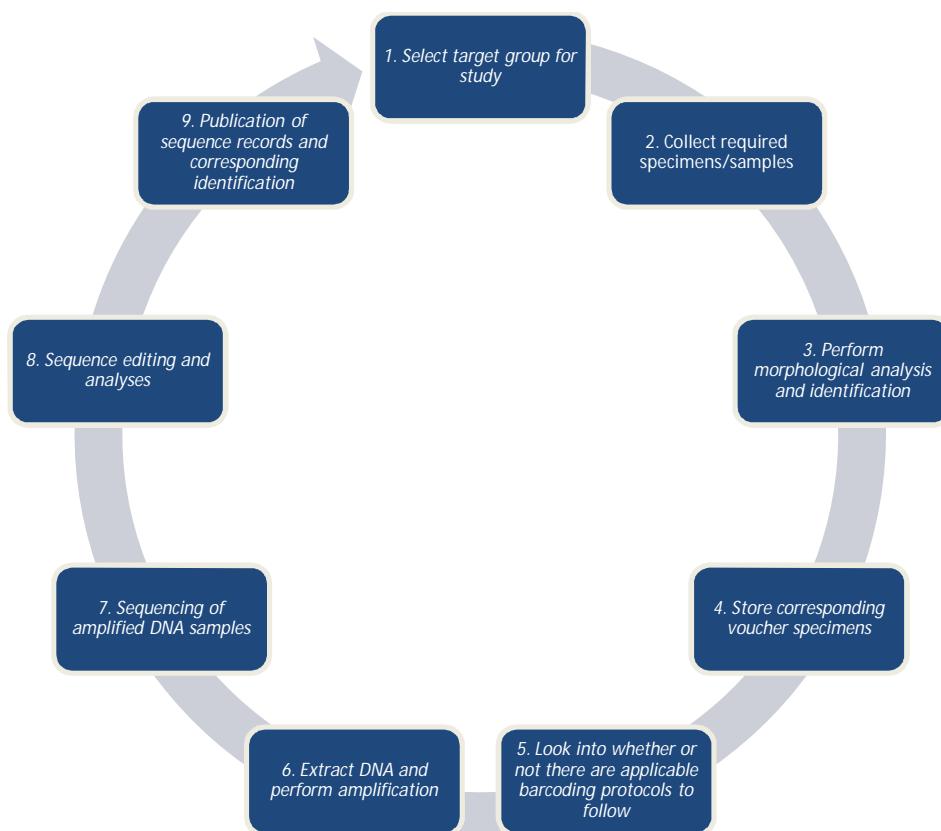


Figure 10 Workflow for DNA barcoding at UM

4.2 Applying framework for species identification of freshwater fish

4.2.1 Sample collection and amplification success

The selection rules established in the initiation of this study (target collection of five (5) fish samples common to the area from at least 15 different species) are general and were not established to support investigation of for example known taxonomic issues or to support other research initiatives ongoing within species from the region. The selection rules were also not adhered to fully during collection and samples were mainly selected based on availability. This was discussed frequently and the research team at UM in

Malaysia were aware of these aspects but established that the goal of this project was never to evaluate DNA barcoding as a tool for addressing known taxonomic issues in Malaysia, the goal was mainly to evaluate suitability of established protocols as tools for species identification and to set up a framework for further work with DNA barcoding at UM.

The application of universal fish primers for the *coxI* fish gene resulted in successful amplification in most cases and the results from DNA extraction shows that qualitative barcode sequences can be extracted from fish tissue using the suggested extraction protocols. The methods used could potentially be more time consuming, compared to for example silica-based methods for DNA extraction. Due to previous success at UM using this protocol and the availability of the equipment required at UM this was still assessed to be the most efficient method at the time. These results strengthens the argument that standardized methods to support quick and efficient methods for species identification helps provide time efficient tools allowing researchers to focus on the larger objectives of their studies rather than spending time on testing different methods for amplification and other steps in the process of retrieving DNA sequences for analysis.

PCR amplification failure (only 70% of samples amplified successfully) for some of the samples may be due to degradation of the DNA caused by inaccurate handling of the samples during collection. The environment on the boats during collection did not always allow for ideal conditions for managing DNA. The air temperature where samples were kept between collection of the specimen and tissue sampling, albeit on ice but not frozen, was always somewhere between 30°C and 33°C which are unideal thermal conditions and increases the risk of degradation of the DNA [45]. Other reasons for PCR amplification failure could be that samples were drying at too high temperatures during DNA extraction or that the storage of tissue samples in alcohol for too long time periods in warm temperatures has caused quicker molecular modifications, both resulting in degradation of the DNA. Testing additional PCR cycles could also be an option in order to increase PCR amplification success.

The actual tissue sampling from the specimens was also carried out on the boat under fairly inconvenient circumstances, i.e. many people working within the same physical area and time pressure to finalize the tissue sampling, which potentially has increased the risks of manipulation of specimens leading to errors such as mislabelling or cross-contamination of DNA. The identification error identified for sample CV15, was morphologically identified as *Pomadasys argenteus* but GenBank results indicates that this fish specimen was rather more likely to belong to the species *Pomadasys kaakan*, is likely caused by a mislabelling error during trawls but could also be caused by inaccurate morphological identification post sampling due to uncertainty of the specimen taxonomy. Relying too heavily on the result of GenBank analysis is also a risk. Sequences submitted to GenBank contain similar elements of inaccuracy in terms of errors related to species identification, mislabelling etc. as the sequences used for comparison in this study. GenBank results should hence not be considered to reflect the absolute truth. In this particular case for example the Genbank results showed high

similarity to both the species *Pomadasys kaakan* and *Pomadasys hasta*, which strengthens the argument that GenBank results should be taken into analyses as support but not as a tool providing 100% accurate match results. Keeping voucher specimen at an appropriate institution for validation purposes as highlighted by Seberg in 2003 [46] and his colleagues is now common practice in DNA barcoding. In this project all vouchers were labelled and are kept at the UM internal museum. If additional errors are identified during future work with these samples or on this set of data others can also go back and validate the accuracy or inaccuracy of morphological identifications made.

4.2.2 Individual and between sequence analyses

Key is to initially highlight that the number of sequences retrieved and later applied for sequences analyses are too few and cannot provide statistically significant results. The analysis results are only indicative and can mainly be used to create an initial understanding and support further more detailed analyses at UM of genetic variation in freshwater fish. The results of this study however indicate that the interspecific divergences calculated are clearly greater than intraspecific variations within the sample group of this study, i.e. a barcode gap seems to exist. These results hence support that *COI* sequences as barcodes could be effective for rapid identification of freshwater fishes as DNA barcoding only becomes effective as a tool for species identification if a clear barcode gap can be identified.

Mallet and Willmot [47] highlights that DNA barcoding is challenging to apply as method for species identification between groups of closely related organisms, this since there will be very small difference in terms of variation between sequences of closely related organisms. Discrimination accuracy when using distance-based methods for species differentiation such as those used with DNA barcoding will be unreliable when sequence differences are too small. Interspecific divergences calculated herein are clearly greater than intraspecific variations within the sample group of this study even between specimens from the same order (calculated as 22,26% between sample D10 and D13, C15, D15 respectively) which contradicts the challenges identified by Mallet and Willmot [47].

In 2003 Lipscomb *et al.* [48] also published an article focusing on the problems of aligning sequences of different length. There have been improvements in this area but manual editing of sequence records as applied in this study is still the most common way of preparing sequences for comparing unknown sequence records to reference sequences. Manual trimming increases the risk of inaccurate management of sequences leading to inaccurate result. The sequence comparison results outlined above should take errors caused by manual trimming into account in any future analysis. None of these sequences contain indels or stop codons post trimming and this supports the fact that the amplified sequences constitute functional genes and variation is hence most likely not a result of the presence of indels.

As mentioned earlier in this thesis, the *coxI* gene is widely used to serve as a DNA barcode for differentiation within in all animals and has successfully served as a

barcode in many different phylogenetic studies involving fish. This implies that the chances of finding high-scoring sequence matches in reference libraries, such as GenBank, are high. However, always taking into account that GenBank sequences contain same risk of including errors as the samples used for analysis. GenBank built in tools for sequence comparisons, in this case BLAST, are however not specifically implemented to be applied for DNA barcoding analyse and many of the barcode sequences submitted in GenBank are not of suitable length, i.e. approximately 648bp in length, to facilitate large-scale standardization of DNA barcoding [49]. This could complicate and reduce potential success of DNA barcoding analyses. In this study we found matches with low E-values 0,00 and average similarity score of 90% for 88% of all sequences retrieved which however strengthens the arguments for using *coxI* genes as barcodes in studies involving freshwater fish as it seems to rather facilitate rapid and easy identification of the retrieved sequences.

DNA barcoding as a method must also consider potential impacts to results as a result of differences in actual definitions of a “species”. There are many different definitions in place and so far none has been universally accepted. It is therefore important to point out that DNA barcoding does not seek to find a species definition it just uses sequence divergence as a tool for identification of groups that have similar and correlated sequences and can be identified as species. Others also highlight that it is sensationalistic to say that DNA barcoding can progress taxonomy, it is not supporting the taxonomists themselves it is only an additional tool for industries and governments to make species identification more efficient [52]. The result of this thesis could however provide a good start for researchers at UM to start build a reference library of barcode sequences from freshwater fishes.

5 Acknowledgements

I want to first and foremost give my deepest thanks to the wonderful people at the Centre of Research for Computational Sciences and Informatics in Biology, Bioindustry, Environment, Agriculture and Healthcare (CRYSTAL) at UM and especially to my supervisor Associate Professor Dr. Amir F Merican for giving me the opportunity to perform this project and for providing excellent supervision and guidance.

Special thanks to my very supportive scientific reviewer at Uppsala University, Mikael Thollesson, for always taking the time to provide answers to my questions and directions when required, for me to get back on the right track. He has been particular helpful and flexible towards the end of this thesis completion and for this I will always be very thankful.

Additional thanks also needs to go out to my helpful scientific reviewer at UM, Dr. Rizman Idid. He made himself available on a daily basis to support and help drive the work forwards with extensive expertise and knowledge in the field. Prof. Dr. Chong Ving Ching and his team also require extra appreciation, without your extensive support in terms of performing morphological analyses of sample specimens this thesis would not be the same.

Appreciative thoughts also go out to Yusrizam bin Sharifuddin for excellent hospitality and support in my journey to navigate in a new cultural landscape and in getting to know the wonderful country Malaysia.

6 References

- 1: Pennisi, E (2003). Modernizing the Tree of Life. *Science*, vol. 300, June, pp. 1692-1697.
- 2: Hebert P.D.N., Cywinski A, Ball S.L., DeWaard J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 270, 313–322.
- 3: Rasmussen, R. S., Morrissey, T. M., and Hebert, P. D. N. (2009). DNA Barcoding of Commercially Important Salmon and Trout Species (*Oncorhynchus* and *Salmo*) from North America. *Journal of agriculture and food chemistry*, vol. 57, pp. 8379-8385.
- 4: Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. and Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical transactions of the royal society, Biological sciences*, vol. 360, pp. 1847-1857.
- 5: The International Barcode of Life project (iBOL). (February 2010). What is dna barcoding. <http://ibol.org/about-us/what-is-dna-barcoding/>.
- 6: Consortium for the Barcode of Life (CBOL). (December 2009). What Is DNA Barcoding?. <http://wwwbarcodeoflife.org/content/about/what-dna-barcoding>.
- 7: Barcode of Life. (December 2009). What Is cBol?. <http://wwwbarcodeoflife.org/what-is-cbol>.
- 8: Barcode of Life. (December 2009). What Is iBol?. <http://wwwbarcodeoflife.org/what-is-ibol>.
- 9: Fish Barcode of Life Campaign (FISH-BOL). (December 2009). Enabling Tools. http://www.fishbol.org/enabling_tools.php.
- 10: Bee Barcode of Life Initiative (Bee-BOL). (December 2009). About Bee BOL. <http://wwwbee-bol.org>.
- 11: Sponge Barcoding Project. (December 2009). Welcome to the Sponge Barcoding Project!. <http://www.spongebarcoding.org>.
- 12: Kress W.J., Wurdack K.J., Zimmer E.A., Weigt L.A., Janzen D.H. (2004). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences* vol. 102 no. 23, pp. 8369–8374.
- 13: Mindell DP, Sorenson MD, Huddleston CJ, Miranda HC Jr, Knight A, et al. (1997). Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In: DP Mindell, editor. *Avian molecular evolution and systematics*. New York: Academic Press. pp. 214–247.
- 14: Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2(10), e312.
- 15: Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough T.G., Savolainen V (2007). DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci U S A*, Feb 26;105(8):2923-8.
- 16: Galtier, N., Nabholz, B., Glémin, S., and Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, vol. 18(22), pp. 4541-4550.
- 17: Dawnay, N., Ogden, R., McEwing, R., Carvalho, G. R. & Thorpe, R. S. (2007). Validation of the barcoding gene CO1 for use in forensic genetic species identification. *Forensic Science International* 173, 1–6.
- 18: Gissi C, Iannelli F, Pesole G (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*. 2008;101:301–320.
- 19: Kress W.J., Erickson D.L. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *PNAS* 105 (8): 2761–2762.
- 20: Consortium for the Barcode of Life (CBOL). (Decmeber 2009). What Is DNA Barcoding?. <http://wwwbarcodeoflife.org/content/about/what-dna-barcoding>.

- 21: Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P.D.N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, 6, 959–964.
- 22: Meyer C.P., Paulay G (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol* 3:2229–2238.
- 23: Biodiversity hotspots of the World. (December 2009).The Sundaland Hotspot. <http://www.biodiversityhotspots.org/xp/hotspots/sundaland/Pages/biodiversity.aspx>.
- 24: Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., et al. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One*, 3(6), e2490.
- 25: Valdez-Moreno M, Ivanova NV, Elías -Gutierrez M, Contreras-Balderas S and Hebert PDN (2009). Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. *J Fish Biol* 74:377-402.
- 26: Carvalho D.C., Neto D.A.P., Brasil B.S.A.F., Oliveira D.A.A. (2011). DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. *Mitochondrial DNA* 22 (Supp. 1): 97–105.
- 27: Pereira H. M., Leadley P. W., Proenca V, Alkemade R, Scharlemann J. P. W., et al. (2010). Scenarios for global biodiversity in the 21st century. *Science* 330, 1496–1501.
- 28: World Wide Fund for Nature national conservation trust in Malaysia (WWF-Malaysia). (December 2009). Mangrove Forests. http://www.wwf.org.my/about_wwf/what_we_do/forests_main/the_malaysian_rainforest/types_of_forests/mangrove_forests.
- 29: The Sime Darby Group. (December 2009). Mangrove Replanting Project. http://www.simedarby.com/environment/MANGROVE/about_mangrove/threats_to_mangrove_forests.aspx.
- 30: The FishBase. (January 2010). <http://www.fishbase.org>.
- 31: Ivanova, N. V., DeWaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, 6, 998–1002.
- 32: Brown, T.A. 2002. Genomes. 2nd ed. 120-121 p.
- 33: The Canadian Centre for DNA Barcoding (CCDB). (December 2009). DNA Amplification – Animals. http://www.ccdb.ca/docs/CCDB_Amplification.pdf.
- 34: Ivanova, N., Zemlak, T. S., Hanner, R. H., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 6, 998–1002.
- 35: 1st BASE. DNA Sequencing Services. (February 2010). <http://www.base-asia.com/1st-base-dna-sequencing-services>.
- 36: The National Center for Biotechnology Information (NCBI). (February 2012). BLAST: Basic Local Alignment Search Tool. <http://www.ncbi.nlm.nih.gov/BLAST>.
- 37: Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30(12):2725–2729.
- 38: The National Center for Biotechnology Information (NCBI). (February 2010). Consensus Sequence – MeSH. <http://www.ncbi.nlm.nih.gov/mesh/68016384>.
- 39: Needleman, S. B. & Wunsch, C. D. (1970). General method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* 48, 443-453.
- 40: Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16, 111–120.

- 41: Saitou N & Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- 42: Boldsystems. (January 2010). About. <http://www.boldsystems.org>.
- 43: Lim J, Kim S-Y, Kim S, et al. (2009). BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. *BMC Genomics*. 2009;10(Suppl 3):S8. doi:10.1186/1471-2164-10-S3-S8.
- 44: Will K.W., Rubinoff D (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* (20) 47–55.
- 45: Moshe K, Zidon D, Polak P, Zalevsky Z, Shefi O. (2013). Thermal Degradation of DNA. *DNA and Cell Biology*. June 2013, 32(6): 298-301.
- 46: Seberg O, Humphries C. J., Knapp S, Stevenson D.W., Petersen G, Scharff N, Andersen N. M. (2003). Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18:63–65.
- 47: Mallet J and Willmott K (2003). Taxonomy: Renaissance or Tower of Babel?. *Trends Ecol. Evol.* 18:57–59.
- 48: Lipscomb D, Platnick N, Wheeler Q (2003). The intellectual content of taxonomy: A comment on DNA taxonomy. *Trends Ecol. Evol.* 18:65–66.
- 49: Waugh, J.(2007). DNA barcoding in animal species: Progress, potential and pitfalls. *BioEssays*, 29(2), 188–197.
- 50: Rock J, Costa F.O., Walker D.I., North A.W., Hutchinson W.F. and Carvalho G.R. (2008) DNA barcodes of fish of the Scotia Sea, Antarctica, indicate priority groups for taxonomic and systematics. *Antarct Sci* 20:253-262.
- 51: Costa F.O., Landi M, Martins R, Costa M.H., Costa M.E., Carneiro M, Alves M.J., Steinke D, Carvalho G.R. (2012). A Ranking System for Reference Libraries of DNA Barcodes: Application to Marine Fish Species from Portugal.
- 52: Ebach M.C., Carvalho M.R. (2010). Anti-intellectualism in the DNA Barcoding Enterprise. *Zoologica* 27 (2): 165-178.
- 53: April J, Mayden R.L., Hanner R.H., Bernatchez L (2011). Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the United States of America* 108: 10602-10607.
- 54: Ward RD, Hanner R, Hebert PDN (2009). The campaign to DNA barcode all fishes, FISHBOL. *J Fish Biol* 74:329–356.
- 55: The Canadian Centre for DNA Barcoding (CCDB). (January 2010). Home. <http://ccdb.ca/>.
- 56: QIAGEN. (January 2010). QIAquick PCR Purification Kit. <http://www.qiagen.com/products/dnacleanup/gelpcrsicleanupsystems/qiaquickpcrpurificationkit.aspx>.
- 57: Consortium for the Barcode of Life (CBOL). (January 2010). Non-COI Barcode Regions — Guidelines for CBOL Approval. <http://www.barcoding.si.edu/pdf/guidelines%20for%20non-co1%20selection%20final.pdf>.

7 Appendices

Appendix A – Sequences per sample based on four bidirectional primers

>1st_BASE_384239_P_C5_FISH_F1
NNNNNNNTGNTGNNGNTGCTGANNGAATAGTAGGAACCGCCCTAGCCTGTAATTGGCAGAATTAGCCAACC CGGCGCCCTC
TAGCCGATGATCAAATCTATAACGTATCGTACCGCCCACGCTTCGTAATAAATTCTTATAGTGATAACCAATCATAATTGGAGGCT
TTGGAATTGACTGTCCCCCTAATAATTGGAGCCCCGACATGGCTTCCCCGAATAAAATATGAGCTTCTGGCTCTCCCCCAT
CCTTCTACTCTCTTCTCATCAGGAGTGAAGCAGGGGGGAACAGGATGAACGTATATCCACCCCTGCTGGAAATCTCGCA
CACGCAGGAGCTTCTGAGACCTACTATTTCTCCCTCACCTAGCAGGAGTCTCATCAATTCTGGGGGCATCAACTCATCACAACT
ATCATTAATGAAACCTTCCAGCTATCTCACAATCAAACACTTATTTGTTGAGCCATTCTAATTACTGCGTACTTTACTCTCT
CCCTCCCAGTTCTGCTGCCGGCATCACTATACTATTAAACAGACCGAAATCTAACACCAACTTCTTGACCCCGCAGGAGGGGAGAC
CCAATCCTTACCAACATCTCTGATTCTTGGCCACCCAGAAAGTCTAA

>1st_BASE_384240_P_C5_FISH_R1
GNNNNNNNNNNNNNNNNNNNNNGNTCCCCCTCTGCGGGTCAAAGAAAGTGGTGTAAAGATTCGGTCTGTTAATAGTATAGTGATG
CCGGCAGCAAGAAGTGGAGGGAGAGAAGTAAAGTACGGCAGTAATTAGAATGGCTAACAAATAGAGGTGTTGATATTGTGAG
ATAGCTGGAGGTTTCATATTAGTATGGTGTGAAGTGTGAGATGGCCCCAGAATTGATGAGACTCTCTAGGTGGAGGGAGAAA
TAGTAAAGGCTACCGGAAGCTCTGGTGTGCGAGATTCCCAGAACGGGTGATATACAGTTCATCTGTCGCCCTGCTCAACT
CCTGATGAAGCAAGGAGAAGTAGGAAGGGTGGAGAACAGCTCATATTATTATCGGGGAAGGCATGTCGGGGCT
CCAATTATTAGGGGACAAGTCATAATTCCAAGCCTCAATTATGATTGTTACTATAAAGAAAATTACGAAAGCGTGGCG
TAAGATAACGTTATGATTGATCATCGCTAGAAGGGCCGGTTGGCTAATTCTGCCGAATTAGCAGGCTAAGGGCGGTTCCA
ACTATTCCGGCTCAGGCACCAAATACTAGGTAGAGGGTGCCTAATGTCCTTGTGGTTGGTGA

>1st_BASE_384241_P_C5_FISH_F2
NNNNNNNGGGGGGNNNCNNNANNCCCGANNAAGAGGGTGGTTAAGGTAAATATGGGACCTCATTGCTGAGCNTNGNTT
AAAAATAAAAGNNAAGAGAGGCCATGATAGCCNTGNCCAATGAGAAGTAAAANTAACGCAGGGTATGCCATGAATTGNTTG
ACNNGGACNAAAAGATGGGNGAATTTNNAAGCGNCCGATCCGTTATTACCTAGAATCAATAAGGGNCTGCCGACAAAAGACAA
ATTGATCCTCTGCTTGTCCCCTCCCCGGGTAGAAGGGGAGGCAGGAANGNTGAACCGGTATGCCAAGTCGNGGAAAGGGACG
ACGCAGGNGAGAAGTGGAAAATAAGAATCTACAAAAGAAGGGAAACCCCTTCTGGGCAATTTTACTTACCAACCGTCAATA
ACAAGAACCCCCAGGCCAATCCCATACCATATATAATTACACTGACCCGTGCACTACAACGAGAATAACATCATCGCTCGG
CCCCAATTAAAGCCGGGAGCAGCCCGATGCTCCCCACAAACCGAAATCTNNTCCN

>1st_BASE_38424_P_C5_FISH_R2
NNNNNNNGNNNNNNNNNNNNNCTCCNCNTGCGGGGCAAAGAAAGTGGTAAAGATTCGGTCTGTAAATAGTATAGT
ATGCCGCAGCAAGAACGTTGGAGGGAGAGAAGTAAAAGTACGGCAGTAATTAGAATGGCTCAAACAATAGAGGTGTTGATATTG
GAGATAGCTGGAGGGTTCATATTAGTATAGTTGATGAAGTTGATGCCCGAACATTGATGAGACTCTGTAGGTGGAGGGAGA
AAATAGTAAAGTCTACCGAACGCTCTCGTGCGAGATCCCCAGCAAGGGTGGATACAGITCATCTGTCTCCCCTGCTICA
ACTCTGTAGCAAGCAAGGAGAACAGTAGGAAGGGAGGGAGGACAGAACGCTCATATTATTCTGGGGAGGGCATCTGGGG
GCTCAATTATTAGGGGACAAGTCATTCTCAAAGCCTCAATTATGATTGTTATCATAAAAGAAAATTATTACGAAAGCTGGGC
GGTAACGATAACGTTATGATTGATCATGCCCTAGAGGGCCGGTTGGCTAATTCTGCCGAATTAGCAGGCTAAGGGCGTT
CCAATTCGGCTCAGGCACCAAAACTAGGTAGAGGGTCCCAATGCTTGTGGTTGGTGA

>1st_BASE_384243_P_C6_FISH_F1
NNNNNNNNNNNTNGGTCTNNNNNNNNNTNGNGNCNCCCATTAAAGCCTGNTAATTCGGGCAGAATTAGCCCAACCCGGCGCCCTTCTAGGCATGATCAAATCTATAACGTTATCGTTACCGCCCCACGCTTCGTAATAATTTCCTTATAGTGATACCAATCATAATTGGAGGCTTGGAAATTGACTTGTCCCCCTAATAATTGGAGCCCCGACATGGCCTTCCCCGAATAAATAATATGAGCTCTGGCTCTTCCCCCATCTTCTACTTCTGCTCATCAGGAGTTGAAGCAGGGCCGGAACAGGATGAACGTATATCCACCCCTGCTGGATATCTCGCACACGCCAGGGACCTTCCGTAAGCCTTACTATTTCCTCCCTCCACCTACAGGAGTCTCATCAATTCTGGGGGCATCAACTTCATCAAACACTTCAATATGAAACCTCCAGCTATCTCACAAATATCAAACACCTCTATTGTTGAGGCATTCTAATTACTGCCGTACTTTTACTCTCTCCCTCCAGCTTGTGGCAGGACATCAACTACTATTAAACAGGCCAACTTAAACACCACCTTCTTGACCCGCAGGAGGGGAGACCAATCCTTACCAACATCTTCTGATCTTGGCACNNNAAGCTAA

>1st_BASE_384244_P_C6_FISH_R1
NNNNNNNNNNNNNANTGNNNGNNCTANGNTCCATGCGCNGTCAAAGTAAAGTTGNNGTTAAGATTTCGGTCTGTTAATAGTAAAGTGTGAGATAGCTGGAGGTTTCAATTAAATGATAGTTGTGATGAAGTTGATGCCCCAGAATTGATGAGACTCCTGCTAGGTGGAGGGAGAAAATAGTAAGGTCTACCGAAGCTCCTGCGAGATCCCAGCAAGGGTGGATATACAGTTACCTCTGTCAGGCGCCCTGCTCAACTCCTGATGAAGCAAGGAGAAGTAGGAAGGATGGGGAGGAGGCCAGAACGTCATATTATTTCAGGCGCCATGTCGGGGCTCCAATTATTAGGGGACAAAGTCATAATTCCAAGGCCTCAATTATGATTGGTATCACAATAAGAAAATTATTACGAAAGCGGGCGGTAAACGATAACGTTATAGATTGATCATCGGCCAGAAGGGCGCCGGGTTGGCTAATTCTGGCCAAATTAGCAGGCTAAGGGCGTTTCAACTATTCCGGCTCAGGCCAAATACTAGGTAGAGGGTGCCTAGTCTTNTGTTGGTGA A

>1st_BASE_384245_P_C6_FISH_F2
NNNNNNNGNNNNGACATANNNNNNACACCCTNCCNNTGANNNTAGTAGTNNAGGATCGCTGGGTGNGAATTCTCGTT
ATNGACCTNNCCNTNCTTAGNTCTNNCNCNGNGNAATTAGNCTCAATCCTGACGACGATTCTAGGGTATAACATCGAG
NAGCTTATAAGCGNAGTAGTCGTTATACGAGGCANCAGCGCTCTCGNTAATAAAATNTGTCGTTTTTTTNTCCCCCCGCAA
GTGCTTAATTGCGAGGGTTTTGTTNAATGNAAGATTGANTGCCCATAGTAAGTAGNGAGAGNAACCAACCGANNNNG
ATATGGTGCCTGGTANCAGACTCCGATTAGTACATCAGAGACTNNAGTGANTGAGTAGATNNNTGTCNTGNCNTCGCTNTA
CCGACCAAACTACGTTNCGAATAAAACNNTTNCNTGCGGGTGTGGAGNNTAANTANCTAACGANGAAAATGGTGGGATA
GATGTCATGAGGACATAAGATCAACATGNGNGCCTGNCNGCCTGCGNCANANNANCNANATGAGTGTTCANNTNGAANATNAN
CAATGCGTGGTGACTIONGANTNNNNNNANNNCGGGCAGCENNNTGATTGANTCACGGNNAAAANNNNGNCGANAAAT
ANTTGCGNGTTNGANGCGCANNNNNNCTGTACANTNACCNATTNNNNNNNNACTNCCTTACCGCCANCGNNNN
NCGGANNNAANNCGAACGNTNNNATGGGAGANNCTNNTNNNNNTNNNTATTGACTNAACNNNNNNCCNNNGNCNN
NNNNNNNGGANCNGGNANNNNGNNNCGGNNNTGNNNNNGGGNNTATNNNNNNNNNGTTNNNNNNNNNNNNNN
NN

>1st_BASE_384246_P_C6_FISH_R2
GNNNNGGNANNNTGNGTNTNNNCCTCTGGGGTCAAGAAAGTGGTITAAGATTCGGCTGTTAATAGTATAGTGATGCCGGCAG
CAAGAACTGGGAGGGAGAGAAAGTAAAGTACGGCAGTAATTAGAATGGCTCAAACAAATAGAGGTGTTGATATTGTGAGATAGCTG
GAGGTTCTATATAATGATAGTGTGATGAAGTTGATGGCCCCCAGAATTGATGAGACTCCTGCTAGGTGAGGGAGAAAATAGTAAAG
GTCTACGGAAAGCTCTGGTGTGGAGATTCCCAGCAAGGGTGGATATACAGTTCATCCTGTTCCCGCCCTGCTCAACTCTGATG
AACGAAGGAGAAGTAGGAAGGATGGGGGAAGGGCAGAAGCTCATATTATTATTGCGGGGAAGGCCATGTCGGGGCTCCAATTAA
TTAGGGGCAAGTCATTCCAACAGCTCCAATTATGATTGGTACTCATATAAAAGAAAATTATACGAAAGCGTGGCGGTAAACGAT
AACGTTATAGATTGATCATCGCCTAGAAGGGCAGGGTGGCTAATTCTGCCGAATTAGCAGGCTAAGGGCGTTCAACTATTCC
GGCTCAGGCACCAATACTAGGTAGAGGGTGCCAATGTCTTGTGGTTGGTGA

>1st_BASE_384247_P_C15_FISH_F1
NNNNNCNGGGTCTGNAGCTGGAATANGTAGGCACGGCCCTAACGTCCTTATCCGTAGCAGAACTTAGTCACCCGGGCCCTCTCG
GGAGACGACCAAATTATAATGTTATCGTCACTGCACAGCCTTCGAATAATTCTTATAGTAATGCCATCTAACCGGGCTTC
GGAAACTGACTCGTCCCTAACGATGGAGCCCCGATATAGCATTCTCGAATAAAATAATAGCTCTGACTCTCCCCCTCT
TTTCCTTCT
ACGCAGGGCATCGTCGACCTAACAAATTCTCTCTCCACCTAGCAGGCGTCTCTCAATTCTTGAGCAATTAACTTCATCACAA
ATCATTAACATAAAACCCCGCTATCTCCAAATCAAACCCCTCTCTCGTCTGATCGCTCTAGTAACCGCTTCTCTTACTCT
CACTCCCAGTTCTGGCGCCGGCATTACAATCTTACAGATCGAAACTTAACACCACCTTCTCGACCCGGGAGGAGGTGAC
CCAATCTGTATCAACACCTTTCTGATTCGGNCCCNNGGAAGTANN

>1st_BASE_384248_P_C15_FISH_R1
NNNNNTNNCAGNNNTGGTCACCTCCCTCGCGGGGTCGAAGAAGGTGGTTAAAGTTTGATCTGTAAGAAGTATTGTAATGCCGGCGCGAAGAACTGGGAGTGAAGATAAGAGGAGAACGGCGTTACTAGGACGGATCAGACGAAGAGAGGGGTTGATATTGGGAGATAGCGGGGGTTTATGTTAATGATTGTTGTGATGAAGTTAATTGCTCCAAGAATTGAGGAGACGCTGCTAGGTGGAGAGAGAAAATTGTTAGGTCGACGGATGCCCTGCGTGGCTAGGGTTCGGCTAGAGGGGGTATACTGTTACCGGTACCGGCCCGCTCAACTCTGAGGAGGCAAGGAGGAGAAGGAAAGAGGGGGGGAGGAGTCAGAAGCTCATATTATTATTCGAGGGAATGCTATATCGGGGGCTCAATCATTAGGGCAGCTGAGTTCCGAAGGCCGATTAGGATGGCATTACTAAAGAAAATTATTACGAAGGCCTGTGCACTGACGATAACATTATAAATTGGTCGCTCCGAGAAAGGGCGCCGGGTTGACTAAAGTCTGCTCGATAAAGGAGGCTTAGGGCTGTGCCCTACCATTCAGCTCAGGCCAAATACTAAAGGGTGCCTAGTCTTNGGTT

>1st_BASE_384249_P_C15_FISH_F2
NNNGNNNCCTGAGCTGGATGGT GAGGCACAGCCCTAAGCCTCTTATCGAGCAGAACCTAGTCACCCGGCGCCCTCTCGGAGA
CGACCAAATTATAATGTTATCGTCACTGCACACGCCCTCGTAATAATTTCTTATAGTAATGCCCATCTAACCGGGCTTCGGAAA
CTGACTCGTCCCCCTAATGATTGGAGCCCCCGATATAGCATTCCCTCGAATAAAATAATGAGCTCTGACTCCTCCCCCTCTTCC
TCTCTCCTGCTCCCTCAGGAGTTGAAGCGGGGGCGGTACCGGATGAACAGTATACCCCCCTTAGCGGAAACCTAGCCCACGCAG
GGCATCGTCGACCTAACAAATTCTCTCCACCTAGCAGGGCTCTCTCAATTCTGGAGCAATTACCTCATCACACAATCATTA
ACATAAAACCCCCCGCTATCTCCAATACAAACCCCTCTCTCGTATCGGCTAGTAACC CGGTTCTCTTACTCTCACTCC
AGTCTTGGCCGGCGATTACAATACTCTTACAGATCGAAACTAACACCCACCTCTCGACCCGGGGAGGAGGTGACCCAATCC
TGATTAACACCTTTCTGATCTCTCGTCCCCCTCGAAGTAAAAAA

```
>1st_BASE_384250_P_C15_FISH_R2
NNNNNNNNNNNNATGGGTACCTCCTCCGGCGGGGTCGAAGAAGGTGGTTAAAGTTCGATCTGTAAGAAGTATTGTAATGCCGG
CGCAAGAACCTGGGAGTGAGACTAAGAGGAGAACGGCGTTACTAGGACGGATCAGACGAAGAGAGGGGTTGATATTGGGAGATA
GCGGGGGGTTTATGTTAATGATTGTTGTGATGAAGTTAATTGCTCCAAGAATTGAGGAGACGCCTGCTAGGTGGAGAGAGAAAATTG
TTAGGTCGACGGATGCCCTGCGTGGCTAGGTTCCGGCTAGAGGGGGTATACTGTTCATCCGGTACCGGCCCCGGCTCAACTCCT
GAGGAGGCAAGGAGGAGAAGGAAAGAGGGGGGAGGAGTCAGAAGCTCATATTATTGAGGAAATGCTATACGGGGGCTCCA
ATCATTAGGGCACGAGTCAGTTCCGAAGGCCCGATTAGGTGGGCTTAACTATAAAGAAAATTATTACGAAGGCCTGTCAGTGA
CGATAACATTATAAATTGGTCGCTCCGAGAAGGGCGCCGGGGTGAAGTCTGCTCGGATAAGGAGGCTTAGGGCTGTGCTACC
ATTCCACGTCAGGACCAAATACTAATAAAAGGTGCGCAATGCTCTTNTGN
```

```
>1st_BASE_384259_P_D7_FISH_F1
NNNNNNNNTGNANNCAATAGTGGCGCCGACTCAGCCTATTACTACGCACAGAACCTTCACAACCTGGCCCTTCTGGAAATG
ATCAAATCTATAACGTTGTGAAACAGCCCACGCATTGTAATATCTCTCATGGTCATACCCATTATAATCGCGGTTTCGGAACAT
GGCTCATCCTTAAATAATTGGTGCCCCAGACATAGCATTCCCACGAATAAAATATAAAGCTCTGACTCTCCACCACATTCTTC
TTCTCCTAGCTTCATCAACAGTAGAAGCAGGGGGGGAAACGGATGAACCATCTATCCTCACTAGCTAACAAACATTGCACATGCAGG
TGCCCTCGTAGACCTGACAATCTTCACTTCACCTAGCTGTGTCATCAATCCTGGTGTATTAACTTCATCACAAACAATTATAA
CATAAAACCAAAAGCCTCCACAAATACCAAAACCCCTTTTATCTGAGGACTAATAATCACAGCAGTTCTGCTTCTATCTTAC
AGTCTTGGCCGGAAATCACGATACTACTAACAGACCGTAACCTGAAACACAACATTCTTGACCCCTGGAGGGGGAGGGAGACCCATC
CTATTCGAAAGCTCTTCTGATTCTTCCGGNNNNNNAAACCTGTAAC
```

>1st_BASE_384260_P_D7_FISH_R1
NNNNNNNNNTGNNNNNNNGNNGNTCTCCTCCNTCCNGNNCAAAGAATGTTGTTCAGGTTACGGTCTGTTAGTAGTATCGTGA
TTCCGGCGGCAAGGACTGGTAAGATAGAAGCAGAAGAACTGCTGTGATTATTAGTCTCAGATAAAAAGAGGGGTTGGTATTGTG
GGAGGCTTTGGTTTATGTTATAATTGTTGATGAAGTTAATAGCACCAAGGATTGATGAGACACCAGCTAGGTGAAGTAAAAG
ATTGTCAGGTCTACGGAGGCACCTGCATGTCAATGTTGTTAGCTAGTGGAGGATAGATGGTCACTCCGTTCCGCCCTGCTCTAC
TGTGATGAAGTCAAGGAGAAGAAAGAATGATGGTGGGAGAAGTCAGAACGTTATATTATTATTGTTGGAATGCTATGCTGGGCA
CCAATTATTAAGGAATGAGCCAGTTCCGAAACGCCGATTATAATGGGTATGACCATGAAGAAGATTATTACAATGCGTGGGCT
TTACAACACGTTATAGATTGATCATTCAGAACGAAAGGGCCAGGTTGAAAGTCTGTGCGTAGTAATAGGCTGAGTGCGGCC
AACTATTGCGGTTCATGTACCAAAATACAAGTATAAGGTGCAATGCTTGGNNTTNGGTANAA

>1st_BASE_384261_P_D7_FISH_F2
NNNNANNACNAACCCCCCCCNNAGTNANCNTCCNNNCNAATTCTCCCTGACTACCGCTCTNNCTCATAAA
ATTGTTGCTTAATGCCNTTGAAACTTGGCACCTAGATGTCAGATAAGTCACTCCGTTTAAATGGTCTGGTGGTGT
CATCCAAATAATAGCGGGGTTGGGAAAGCGGCCATTCTTAAAGGCGCCAAAAAATGCTCCACCAAAATAT
TAAATCCTCCCTCTCCCCCACCACATCTTTCGGCTGGAAACTCACAAGAAGAAAAGGCCGAAACATAAGAACATCT
GTCCGCCACTACATAACAGTTGGAGAATGCCGGCTCCCTAAGGTTAAGAATCTTCACAGCTCGGGGGGGTCAACACCGA
ACTGGCAATATCACCAACCCCCCACCATAATGAAATACCATCACAGTAAAGGCGGCTTTTAACAACATCA
ATATTGGAACACTTCTGCTGGAGNACAGTGTGGCTTAGTTAAAGACGGATGAAGGAGAATAAGGCTGACCCAGCTAGAGCTGAAG
ACGCAATTAACCCAATAAAAGGTGCAAAAGTCTATTGTTTTGACTGAAAAATCAAATNTAAAAAAAT

>1st_BASE_384262_P_D7_FISH_R2
CNNANNTGANNAAANTTNNGNNNGTCGTGACGATTCACTCANTCTCACTNNNTCCGTCTGGTGGCGTTNCNGGGAGGNACCA
NNCTNGNTGGGGTTTCCGGNGCGCGATTAAATNTGGNTTTTGGCGAAAAAATAAAAAGTTGGTTGGTTGTGGAG
GCTTTGGTTTATGTTATAATTGTTGATGAAGTTAATAGCACCAAGGATTGAGACACCAGCTAGGTGAAGTAAAAGATTG
CAGGTCTACGGAGGCACCTGCATGTCAATGTTGTTAGCTAGTGGAGGATAGATGGTCACTCCGTTCCGCCCTGCTCTACTGTT
ATGAAGCTAGGAGAAGAAAGTATGGTGGGAGAAGTCAGAACGTTATATTATTATTCTGTGGAATGCTATGCTGGGACCAAT
TATTAAGGAATGAGCCAGTTCCGAAACGCCGATTATAATGGGTATGACCATGAAGAAGATTATTACAATGCGTGGGCTGTAC
ACAACGTTATAGATTGATCATTCAAANAAAGGGCCAGGTTGAAAGTCTGTGCGTAGTAATAGGCTGAGTGCGGCCAACTA
TGTGTTCATGTACCAAAATACAAGTATAAGGGCCAATGCTTGTGGTTGGTGAAGNN

>1st_BASE_384235_P_D8_FISH_F1
NNNNNNNNNNNNNNNNNGNNNTGNAGNCANNNTACNNTGGCGCCACTCAGCCTACTACTACGCACAGAACTTACAACCTGCC
CCTTCTGAAATGATCAAATCTATAACGTTGTAACAGCCACGCATTGTAATAATCTTCATGGTCATACCCATTATAATCG
CGGGTTGCGAAACTGACTCATTCTTAATATTGGTCCCCAGACATAGCATTCCACGAATAAAATAAAGCTCTGACTCTCC
CACCATCATTCTCTCTCAGCTCATCACAGTAAAGCAGGGCTGGAACCTGGATGAACCATCTATCTCCACTAGCTAAC
ATTGCACATGCGAGGCTCCGTAGACCTGACAATCTTCACTTCACCTAGCTGGTCTCATCAATCTTGTGCTTAACATC
ACAACAAATTATAACAAAACCAAAAGGCCACACAATACAAACCCCTTTTATCTGAGCAACTAAATACACAGCAGTTCT
GCTCTATCTTACAGTCTGCGCCGAATCACGATACTACTAACAGACCGTAACCTGAACACAAACATTCTTGACCCCTGGAGGG
GAGGAGACCCATCTATTCAACACCTCTTGATTCTTGCCACNNAAAGTCTAA

>1st_BASE_384236_P_D8_FISH_R1
ANNNNNNNNNAANNGANGGGTCTCCTCCCTCAGGGTCAAAGAATGTTGTTCAGGTTACGGTCTGTTAGTAGTATCGTATT
CGCGCGCAAGGACTGGTAAGATAGAAGCAGAAGAACTGCTGTGATTATTAGTGTCTCAGATAAAAAGAGGGGTTGGTATTGTG
GGCTTTGGTTTATGTTATAATTGTTGATGAAGTTAATAGCACCAAGGATTGAGACACCAGCTAGGTGAAGTAAAAGATTG
TCAGGTCTACGGAGGCACCTGCATGTCAATGTTGTTAGCTAGTGGAGGATAGATGGTCACTCCAGTTCCAGCCCTGCTCTACTGTT
GATGAAGCTAGGAGAAGAAAGTATGGTGGGAGAAGTCAGAACGTTATATTATTATTCTGTGGAATGCTATGCTGGGACCAA
TTATTAAGGAATGAGTCAGTTCCGAAACGCCGATTATAATGGGTATGACCATGAAGAAGATTATTACAATGCGTGGGCTGTAC
AACACGTTATAGATTGATCATTCAAAGAAAGGGCCAGGTTGAAAGTCTGTGCGTAGTAGTAGGCTGAGTGCGGCCAACT
ATTGCGGCTCATGTACCAAAATACAAGGTATAAGGTGCAATGTCCTTG

>1st_BASE_384237_P_D8_FISH_F2
NNNNNNNTGGTNCCTAAANAACCCCCCCCCATTGGAAGGAGGGGACGCTACTACGCGCGAACGTACGCTGNCTNNNTCN
GAAATGATCAAATGGTAACCTTGTAACACCCCCTCTTGTATAATCTTCATGGCCTACCCATTATAATGGGGTTTC
GAACCTGACTCATCTTATAATTGGAGCCCACATACCTCCATTCCACTGATAACCCATTATAAGCTCTGACTCTCC
TCTTCTCTCTCATCTTATAATTGGGGAACTGGATGAACCATCTATCTCCCTACTTAACACATTGCAAATC
CAGGAGACTAAAAACCTGACAATATTAGCACTCCCTAAGGGGACTCATCAATCTGGGCTATTAACTTCATCACAAAC
TATTAACTAAAACCAAACCCCCCACATTATAACAAACCCACCTTTATAAGAATAATAAAATCATGGTACTACTTCAAACAAAC
CTATCNTTATTGCTCTGCCAGCAAGGGACAGCTGGTTAAAGAACGGTGAAGAAAAGATTGAGTGAGGCGCCAAAAGGGGG
GGAAAACCTCATTATAAGGAACANTACGGATGATCTGGCATTAGTTNTAAAAAAATAANAAAN

>1st_BASE_384238_P_D8_FISH_R2
NNNNNTGANNNGNNNGCTNNNGCTGTCATCATGCTTGNCGATGTCATTCTCAAGTTTAGCATGTTGAAATTCGGGGTGAATG
CAATTGATGTGCACTTCAACGGGGCAGGGTTNTGTTTGTGTTGATGAGACACCAGCTAGGTGAAGTAAAAGATTG
GCTTTGGTTTATGTTATAATTGTTGATGAAGTTAATAGCACCAAGGATTGAGACACCAGCTAGGTGAAGTAAAAGATTG
CAGGTCTACGGAGGCACCTGCATGTCAATGTTGTTAGCTAGTGGAGGATAGATGGTCACTCCAGCCCTGCTCTACTGTT
ATGAAGCTAGGAGAAGAAAGTATGGTGGGAGAAGTCAGAACGTTATATTATTCTGTGGAATGCTATGCTGGGACCAAT
TATTAAGGAATGAGTCAGTTCCGAAACGCCGATTATAATGGGTATGACCATGAAGAAGATTATTACAATGCGTGGGCTGTAC
ACAACGTTATAGATTGATCATTCAGAACGAAAGGGCCAGGTTGAAAGTCTGTGCGTAGTAGTAGGCTGAGTGCGGCCAACT
TGTGCGCTCATGTACCAAAATACAAGGTATAAGGTGCAATGCTTGTGGTTGGTGAAG

>1st_BASE_384227_P_D9_FISH_F1

CNGNACTTGNNTTGGTNATGNNGCCGTAGTTGGCGCCGACTCAGCCTACTACTACGCACAGAACCTTCACAACCTGGCCCCTTCT
TGAAATGATAAATCTATAACGTTGTGAACAGCCCAGCATTTGAATAATCTCTCATGGTCATACCCATTATAATCGGGGTT
CGGAAACTGACTCATTCCTTAATAATTGGTGCCTCAGACATAGCATTCCACGAATAATAATAGCTGACTCTCCACATGCA
CATTCCTTCTCTCTAGCTCATCACAGTAGAAGCAGGGCTGGAACCTGGATGAACCATCTATCCTCCACTAGCTAACACATTGCA
CATGCAGGTGCCTCGTAGACCTGACAATTTCACTTCACTTAGCTGGTCTCATCACCTGGTCTATAACTCATCACACA
ATTATTAACATAAAACCAAAGCCTCCACACAATACCAACCCCTTTTATGAGCACTAATAACAGCAGCTAACCTGACACAA
TCTTACCACTGCTTGCCTGGGAATCACGACTACTAACAGACCGTAACCTGACACAAACATTCTTGACCTGGAGGGAGGAG
ACCCCATCCTATTCAACACCTTTGATTCTTGACNNN

>1st_BASE_384228_P_D9_FISH_R1

ANGNNNGAATNNAGGGGTCTCCCTCCAGGGTCAAAGAATGTTGTTAGGTTACGGTCTGTTAGTAGTATCGTGAATCCGGC
GGCAAGGACTGTAAGATAGAAGCAGAAGAACCTGCTGTGATTATTAGTGTCTAGATAAAAAGAGGGGTTTGTATTGTGAGGCT
TTGGTTTATGTAATAATTGTTGTGATGAGGTTAAGCACCAGGATTGATGAGACACCAGTAGGTGAAGTAAAAGATTGTCAG
GTCTACGGAGGCACCTGCATGTCAATGTTAGCTAGTGGAGGATAGATGGTCTCCAGGCTGCTTACTGTTGATG
AAGCTAGGAGAAGAAGAATGATGGTGGAGAAGTCAGAACGTTATATTATTATTGTTGAGGAAATGCTATGCTGGGACCAATT
TAAAGGAATGAGTCAGTTCGAAACCCCGATTATAATGGGTATGACCATGAAGAAGATTACAAATGCGTGGCTTACA
ACGTTATAGATTGATCATTCCAAGAAAGGGCCAGGTTGAAAGTCTGCGTAGTAGTAGGCTGAGTCGGCGCCA
GGCTCATGTACCAAATAAGGTATAAGGTGCAATGTTGGGTTGGTGA
AATAGA

>1st_BASE_384229_P_D9_FISH_F2

CNTGNTTGGGGGGNTCAAANCCCANATTGGNACAGCACATCTCTGATTCTCGGCACCCCTGAAGANTGCNAACNAAGGNGTGG
TTGAGTCTAGGAGGGTTTGGGANGNTNCAGAAANGTAAGAATTTTAAGGGCGGGACCGATAAAAAGGGGTTAACCTAA
ANNNAAAAAAACCTGTAGNTGGGNTCCCAAACCTACCCTCCAGAAACAACACTATAAGTGCGGACCTATCCAAACATCATT
ATTCTCCCTAGTTCAACCACCCCTAGAACGTTGGGGAGGAACGTGAGAACCATCTACCTCTACTTAAGAAGAGTACACATG
CAGGAGACAGCAGAAAGGTGAGAGTAGTTAGACGCCCTCCCTGTTCTCCAAATGCTGGGTTTTTAACTTCCCTACAATT
TATAGAACATAACAAAAGCCCCAACAAAGACCAACGCCCTTTTTGAGAACAAATAGGAATCAGGGAGCTTACCCA
GCTTCTATTTAACATTTCTGCGCAGGAAATCAGGATGAGAACATAACAAACGGGAAACCAANATTTTATTGGC
CCACACTCGAGGGAGAGAGAACGCTGTTGATTCCAAGGGCAACAGGATGCAAATGGACAGGGCCCTANGNN
NAGTATAAAAAA
AATAGA

>1st_BASE_384230_P_D9_FISH_R2

GNNNTNNNCNNNNNNNNGANCCNNNNNGTGNNGACNNAGNNNCGTTGAAAGAAAGCTGGTGCCTGATAATCTGTTATGATTA
GTCGGAGAGTCGTTGCCCTCCNGCGGGGGACTGGTATAGATAAGNNNGNNATACNNTNCNGAATATTNNNTGCTCTTT
TTTGGGGGGGGGGTGGTATTGATGNGGAGCTTNNNGTTNAAAAGAATTGTTGGTTTNTNATTGATAGCGCCGAGGGATT
GTATGGAACACCCANCTAGGGTGAAGTGGAAAGATTGGTCAAGGGTCTACCTGAGGCACCTCTTCTATGCCGCTTAGCTAG
GGGGAGGAGAGATGGTCTACCTGAGGGTCTACTGTTGAGGAAGCTAGGGAGAACAGGATGATGGTGTGAGAACGTC
AGACCCCCCTTAGTGTGTTGNAATGGTGTGTTGCTGCAATTAAATTAAATTGATTAATTTC
AAACCCGCAATTAAAGGGTCTGACNTTNTNTAGATTATTACAATTGAGGGCTGCTACACCACCTTAAATTAGATCATT
CTAAGAAAGGCAAACGTTGAAAAGGAGATGATGAATGGAAATTGAAAAAAAGGAGATGATGAAT
GGTTAGTTGNTGAAGAAA

>1st_BASE_384267_P_D10_FISH_F1

NNTGNNNTNCGGTCTGAGCAGGGAAAGTGGGACAGCCTTAAGCCTCTTATCCGTGCTGAACTAACGCCAACCAAGGGCTCTCTGG
GACGACAGATCTATAATGTGATGTAACGGCACATGCCCTCGTAATAATTCTTATAGTTATGCCAGTAATAATTGGAGGGTTGG
AAATTGACTGGTCTCCCTAATGATGGGGCACCGGATATAGCATTCCCCGGATAAAACATAAGCTCTGACTCTCCCCCTCT
CCTCTCTCTAGCTCTCTGGCTAGAGCCTGGGCTGAAACAGGATGAAACAGTCTACCCCCCTCGCTGTAATCTAGCAG
CGGGAGCCTCCGAGACTAACCATCTCACTTCAGGGATTCTCACTTCTGAGGCTTAACTTACCTGCTGAGCAGTCTAATT
TTAATATAAAACCCCTGCTCTCCAATATCAAACCTCTTATTCGCTGAGCAGTCTAATTACTGCTGCTTACTACTCTCT
ACCTGTTCTGCTGCTGGCATACAATACTCTTACAGATGAAACACCTTCTGATCCTGAGGAGGAGACCAA
TCCTTACCAACATCTATTGATTCTTGCCACCCAGAAGTCAA

>1st_BASE_384268_P_D10_FISH_R1

GNNNNTGTAAGAATTGGGCTCCCTCTGAGGATCAAAGAAAGAGGTGTCAGGTTCTGATCTGTAAGAAGTATTGATGCC
GCAGCAAGAACAGGTAGAGAGAGAACAGTAAAGACAGCAGTAATTAGGACTGCTCAGACGAATAGAGGAGTTGATATTGGAAGCA
GCAGGGATTTTATATAATAATAGGGTGTGAAAGTTAAGCCCCAAGGATTGAAAGAAATCCTGCAAGTGAAGTGAAGAAGATGG
TTAGGTCTACGGAGGCTCCCGATGTGCTAGATTACCGCAGAGGGGGTAGACTGTTCTGCTGAGCCGGCTTACGCC
GAGGAAGCTAGAAGGAGAAGGAAAGAGGGGGAGGAGTCAGAACGTTATGTTATTATCCGGGGAAATGCTATATCCGGTGGCC
ATCATTAGGGAAACAGTCATTCCAACCCCTCAATTACTGGCATAACTATAAGAAAATTACGAAGGCATGTGCGT
GATCACATTATAGATCTGGTGTCTCCAAGGAGAGCCCTGGTGTGAGTCAGCACGGATAAGGAGGCTAAGGCTGCT
CCTGCTCAGGCCACCGAATACTAGATAAAGGGTGCCTGATGTTGGNNT

>1st_BASE_384269_P_D10_FISH_F2

TNNNNNTGNNNTNAGCAGGNNNNNTGNGGGACAGCCTTAAGCCTCTTATCCGTGCTGAACTAACGCCAACCAAGGGCTCTCTGGAG
ACGACAGATCTATAATGTGATGTAACGGCACATGCCCTCGTAATAATTCTTATAGTTATGCCAGTAATAATTGGAGGGTTGG
AAATTGACTGGTCTCCCTAATGATGGGGCACCGGATATAGCATTCCCCGGATAAAACATAAGCTCTGACTCTCCCCCTCT
CTCTCTCTAGCTCTCTGGCTAGAGCCTGGGCTGAAACAGGATGAAACAGTCTACCCCCCTCGCTGTAATCTAGCAG
GGGAGCCTCCGAGACCTAACCATCTCACTTCAGGGATTCTCACTTCTGAGGCTTAACTTACCTGCTGAGCAGTCTAATT
TAATATAAAACCCCTGCTCTCCAATATCAAACCTCTTATTCGCTGAGCAGTCTAATTACTGCTGCTTACTACTCTCT
CCTGTTCTGCTGCTGGCATACAATACTCTACAGATGAAACACCTTCTGATCCTGAGGAGGAGACCAA
CTTACCAACATCTATTGATTCTTGCCACCC

>1st_BASE_384270_P_D10_FISH_R2

GNNNNNNGGTAGNATTGGTCTCCTCTGCAGGATCAAAGAAAGAGGTGTCAGGTTGATCTGAAGAAGTATTGTGATGCC
AGCAGCAAGAACAGTAGAGAGAGAAAGTAGTAAGACAGCAGTATTAGGACTGCTCAGACGAATAGAGGGAGTTGATATTGGGAAGC
AGCAGGGGATTTATATTAAATAACTGGTGATGAAGTTAATAGCCCCAAGGGATTGAAGAAATCCCTGCCAAGTGAAGTGAGAAGATG
GTTAGGTCTACGGAGGCTCCCGCATCTGCTAGATTACCAGCGAGAGGGGGTAGACTGTTCATCTGTTCCAGGCCCGCTTACGCC
AGAGGAAGCTAGAAGGGAAAGGAAAAGGGGGAGGGAGTCAGAACGTTATGTTATTCGGGGGATGCTATATCCGGTGC
GATCATTAGGGGAACCAAGCTCAATTCCAAACCCCTCAATTATTACTGGCATAACTATAAGAAAATTATTACGAAGGCATGCGCTTA
CGATCACATTATAGATCTGGTCTCCAGGGAGAGCCCCTGGTGGCTTAGTTACAGCACGGATAAGGAGGCTTAAGGCTGTCCCAACT
ATCCCTGTCAGGCACCGAATACTAGATAAAGGGTGCCTAATGCTTTG

>1st_BASE_384271_P_D13_FISH_F1

NNNNNTGAGGANNTGNTCTTNNGCCACCCAGAAGTCAAGAATAGTCTCTCATAACGTATTGCTGTTTATGTGAGAACCC
TCTCTCGCAGGTTTGATTTCCTTAAATTGTTAGTTCAAGCTGAAGGGACTAACCTCTTACCAAGCTGATTCTTGGCCACCCAGAAGT
CTTAAAGTGGTGCAGANNTCCCGANGATGGCTGCACCAAGAAGAAAATAACAGGGCTGGGGGGGGAGAAGATGGTCCGTATCCT
GCGCTTCTCNGCTTGCAGATAGACGAGGGGGGGAGTTGGATTCACCTTCTCCNCCTCACCAACTACGCAGGNNAAGGG
AGAAGGAAAGATTGGAGANGAACGACCCCTCNCCTGGGGTTGGGGGGGGGGAGCCGGTGGGGACCCAATGTAACAGGGCCAAT
TACCCAAGCCCTCGAACATTGGCATATGTTTAAATAAAATAAAAAAAACAAAATGCTCGACAATAAACATTAAACATATGNATC
TTGTNACCACACCCCCCAGGGTGCCTAGCTCCACTTGANAAAGATTTTTGGTAGGCCGGCAGCGAGGTTCCGGCCCCCGA
ACACACAAACAGTAACGAGCGAAGGTCTTGGAGCTTAGGGAAAACCN

>1st_BASE_384272_P_D13_FISH_R1

NNNNNTGNNNNNNGGTNCGCCATGCCTTGTGGTGGGTGATGTTGAGGGTTGGATCTGCACCAAGGATTGTCATGCCGC
CGTGAGAACAGGTATGGCAGCAGGTATCAGAACCGCAGTTACTAGGACTGATCAAGAGAATAGGGGGCTGATATTGGAAATAG
TTGGGGTTCTGATAATGTTGTGATGAAGTTAATGCCCAAGAATTGAGGAACACCTGCCAGGTGGAGGGAGAAAGATGGT
TAGGTCACCGATGCCCTGGCTGGGCCAGATTCGGCTAGAGGGGGTAACCGGTTCCATCGACTACCAGCTCCGGCTCAACTCCT
GAGGAGGAAGGGAGAAGGGAGGAAGGGGGAAAGGAGTCAAAGCTATGTTGTCATTGAGGAAATGCCATGTCGGAGCTCA
ATCATTAGGGCACTAGTCAGTTCCGAAGGCCGATTAGGATTGGTATTACCATGAAGAAAATTATCACACCGATGTCAGTAA
CGATAACATTATAGATCTGATCGTCCCCAAGGAGAGCGCCGGTGGCTGAGTTCTGCTCGATAAGCAGGCTTAGGGCTGCTACT
ATTCCGGCCCAGGCACCAAATCAAAAGGGTGGCGATATTTTGAATT

>1st_BASE_384273_P_D13_FISH_F2

NNNNNGTGGTCTGNANTAGTATGCACGCCATAAGCCTGTTATCGAGCAGAACTCAGCAACCGGGCGCTCCCTGGGACG
ATCATGATCTATAATGTTATCGTACTGCACATGCGTTGATAATAATTCTTCATGGATAACCAATCTAATCGGCGCTTCGAGAACT
GAAGTAGTCCCCATAATGATTGGAGCTCGGACATGGCATTCCCTGAATGAACAACATGAGCTTTGACTCCTCCCCCTCCTCC
TTCTCCTTCCTCCAGAGGTGAAGCCGGAGCTGGTACTGGATGAACCGTTTACCCCCCTAGCCGGATACTGGGCCACCGAGGG
GCATCGGTTGACCTAACCATCTCTCCCTCCACCTGGCAGCTGTTCTCAATTCTGGGCAATTAACTTCATCACAAACATTAAC
ATGAAACCCCCCGCTATTCCCAATATCAGACCCCCCTATTGCTGTACGCCCTAGTTACCGCGTCTCCCTGCTCCCTACCCG
TTCTCGCCGCCGGCATTACAATGTTTACAGATCGAAACCTAACACCACCTCTCGACCCGCCGGAGGGAGGTGACCCAATCTG
TATCAGCACCTTTCTGATTCTCGNNCCN

>1st_BASE_384274_P_D13_FISH_R2

NNNNNGNNNCNNNGNNTGTGGACNCCTCCTCCGTGGGGTCAAGAAGGTGGTITTAGGTTCGATCTGAATATGCAATTGAAATGCC
CGAGGCTGACAACGGGTAGGGAGAGCAGGAGGAACCGCTTAACTAGGACGGATCAGACGAATAAGGGGTCTGATATTGGGAAA
TAGCGGGGGTTCATGTTAATGTTGTGATGAGGTTAATTCCCCAAGAATTGAGGAAACACCTGCCAGGTGGAGGGAGAAGAT
GGTTAGGTCACAGGATCCCCCTGCGTGGGCCAGATCCCGCTAGAGGGGGTAAACGGTTCATCCAGTACCAGCTCCGGCTCAACT
CTCGAGGAGGAAGGAGGAAGGGAGGGAGGACTCAAAGCTCATGTTCATTCGAGGGAATGCCATGTCGGAGCT
CCAATCATTAGGGGCACTAGTCAGTTCGCAAGGCCCGATTAGGATTTGATTACATGAAGAAAATTATTACAAACGCATGTGCAGT
AACGATAACATTATAGATCTGATGTCCTCCAAGGAGAGGCCGGTTGGCTGAGTTCTGCTGGATAAGCAGGCTTAGGGCTGTGCCT
ACTATTCCGGCCCAGGCACCAAATCTAAATAAGGGTGGCGATATCTTGGAAA

>1st_BASE_384275_P_D14_FISH_F1

NNNNNNNGGANNTGATTCTTGGCCACCCAGAAGTCTAAATGTCGATAACTCTGCCCTCTGAAGAAGTGTCTATGGAATTAAAC
CGGGGACGGTCAAGGGATGATTGGGGGACCGNTAACGTCAAACACCCAGAAGTCTAAATTGAAACCGAAGTCTAAAANGANTAA
NCTTTGTITANGTTNATGTTCCGCTGGTCGTCGCTAGGAAACGACAANACCGCATATGAAACTTATTGTAAGTCGCGGATGCCCT
TCGTATCCCAAACAGGCCAGGGGGGGGAAGGACAACGANTCGTCCCCCCTGCTCCANCTCCGAACTGATGAAGAAAAGAAAAAGA
GCATGCTGCTAGAACGCCCTGNCCCGGTATTCTATTGGGGTAGCTTGAAGCCACGGCTGCTGNGGGGGAGCCGGTCTCC
TGCGCTGCCACCTTCAACGCGGTCTTTTGTAAAGAACACNCATTAACGATGCTCTATAAAATGATCTGCTGTTGCTGTG
CNCCCCACCGCGCCGCCGGGTGGACCGAGAATNACCNGTTTGTCTTNGCAGCNCGTACGCTTCANGTCCCGTCAACCA
AGANAATCAAACGANTNNNNACATC

>1st_BASE_384276_P_D14_FISH_R1

NNNNNNNNNNNCAGGTGTACGTCCAATGTCCTTGGTTGGTGAGTTAATGTTGCATTCTTAAAGCATTGTTATGCCGGCGGTGATAATTGTTGAAAATAACAGGATGATGACGGAAGGAACCTAGGACTTATCGTGACGATAAGGGGGTCTGATATTGGGAAATAGCGGGGGGTTCTGTTAATGTTGTTGATGAAGTAAATTGCCCCAAGAATGAGGAAACCTCTGCCAGGTGGAGGGAGAAGATGTTAGGTCAACGGATGCCCTCGCTGGGCCAGATTCCGGTAGAGGGGGTAACGGTCTACCTCAGTACCACTCGCTTCACCTCTGAGGAGGCTAGGAGAAATGAGGAAAGATGTTGGGAAGCTCCCCACCNGCTTGTGTCCTCAATTGNTGTATTACTTCACCATNGATCATGAANGTCTCCTCCCCCTATNTCCCTGACATACANNTGAAGAACCGAACAA

>1st_BASE_384277_P_D14_FISH_F2
NNNNNGCTGGNCGGATAGTAGGCACAGCCCTAACGCTGCTTATCGAGCAGAACTCAGCCAACCGGGCGCTCTCCTTGGGAGAT
CAGATCTATAATGTTATCGTTACTGCACATGCCTTGTAATAATTTCATGGTAATACCAATCCTAATCGGGCGCTCGGAAACTGA
CTAGTGCCCTAATGATTGGAGCTCCGGACATGGCATCCCTCGAATGAACAACATGAGCTTGAECTCTCCCCCTCCTCCT
CTCCTGCCCTCAGGAGITGAAGCCGAGCTGTACTGGATGAACGTTACCCCTCTAGCCGGAACTGGCCACGCAGGGGC
ATCCGTTGACCTAACCATCTCCCTCACCTGGCAGGTGTTCTCAATTCTGGGCAATTAACTCATCACAACAATCATTAACAT
GAAACCCCCCGCTATTCCCAATATCAGACCCCTTATTGCTGATCGCTAGTACAGTAAACACCCTCTGACCCCGGGAGGAGGTGACCAATCCTGTA
TCAGCACCTTCTGATTCTCGNCCCCGGAAAAGTATAA
>1st_BASE_384278_P_D14_FISH_R2
NNNNNTNGGTACCTCCCGCGGGGCGAAGAAGGTGGTGTAGGTTGATCTGAAGAAGCATTGTAATGCCGGCGAG
AACGGTAGGGAGAGCAGGAGGAACGGCGGTAACTAGGACGGATCAGACGAAATAAGGGGCTGATATTGGAAATAGCGGGGG
GTTCATGTTATGATTGTTGATGAAGTTAATTGCCCAAGAATTGAGGAACACCTGCCAGGTGGAGGGAGAAGATGGTAGGTC
AACGGATGCCCTCGCTGGGCCAGATCCGGCTAGAGGGGGTAAACGGTTCATCAGTACAGCTCCGGCTCAACTCTGAGGAG
GCAAGGAGAAGGAGGAAGGAGGGGGAGGAGTCAAAGCTCATGTTGTTATCGAGGGAATGCCATGTCGGAGCTCAATCATT
AGGGGCACTAGTCAGTCCGAAGCCGGATTAGGATGGTATTACCATGAAGAAAATTATTACAAACGCATGTCAGTAACGATAA
CATTATAGATCTGATCGTCCCAGGAGAGCGCCCGTGGCTGAGTCTGTCGATAAGCAGGCTTAGGGCTGTGCCTACTATTCCG
GCCAGGCACCAAATACTAAATAAGGGTGCCTATCTNNNNNTTAGTCGAA
>1st_BASE_384251_P_D15_FISH_F1
NNCCNCNNNTGNNTTTATGAANCATGTGNTCNNCACACACACACTCACACCATATGCCCTCTCAAAACCCACCAGTGTCACTT
GCTGACTGCTCAGGGATGAATATTAAAGNCACCCACGAANTCAAATGTTCTTATTGCCACCCCTGGCTCAAAGATGTCT
AACAGATGGTGGCAGATCTTATGATTAGTCGAACCCCGGAAAAAAAAGAAGATTGTGAATCCTCCCCCTGACTTCTCT
TCTGCTCAGCTGGCGAAAAGGGGGGGGGAGATGATGAAAGTGAGCACCCCTTGTATTGGAAATTGGAAAGCGGGGA
GACTAAGGAGACTGGAAGGTTCTCTCCCATAGGTTGTTCTAAATGACTGGGGGGGATATAATTCCATCCAACCAAGTAAA
AAATTCCATTACCCGCCCATACTGATAACACACTCTCTGTTAGAGAGATAAAAAGAATGACTGGTTGCTAATCCTTCTATCTCT
AACTGTTCTCTCAAGGGCTANTACTACTACTCTTACTGCTCATCAGATATTAAAGACACTGGTTGCTCACCCCTGGAAAGGAGGAGA
AACTTCTTATTACAGGATAAAAAGANAAGTGTGCGAGCTAGGAAAGANTATAANNN
>1st_BASE_384252_P_D15_FISH_R1
GGTNNGTGTNNNTNNNNNNNTGTANGNGCCGTGCGCTNTCTGAGTTGGGTGATGTTAGGTTGATCTGAAGAAGCAT
TGTAATGCCGGCGCGAGAACGGTAGGGAGAGCAGGAGGAACGGCGTAACTAGGACGGATCAGACGAAATAAGGGGTCTGAT
ATTGGAAATAGCGGGGGGTTCTGTTAGGTTGATGAAGTTAATTGCCCAAGAATTGAGGAACACCTGCCAGGTGGAG
GGAGAAGATGGTCAACGGATGCCCTCGTGGCCAGATTCCGGCTAGAGGGGGTAAACGGTTCATCCAGTACCAAGCTCCG
GCTCAACTCTGAGGAGGCAAGGAGAAGGAGGAAGGAGGGGGAGGAGTCAAAGCTCATGTTGTTCTGAGGGAATGCCATG
TCCGGAGCTCCAATCATGGGCACTAGTCAGTTCCGAAGCCGGATTAGGATTGGTATTACCATGAAGAAAATTATTACAAACGC
ATGTCAGTAACGATAACATTATAGATCTGATCGTCCCAGGAGAGCGCCGGTGGCTGAGTCTGTCGATAAGCAGGCTT
GCTGTGCTACTATTCCGGCCAGGCACCAAATACTAAATAAGGGTGCCTATCTTGTATT
>1st_BASE_384253_P_D15_FISH_F2
NNNNNNNNNTGNCNNNNNGNCNNNNNTGAGGCACAGCCCTAACGCTGCTTATCGAGCAGAACTCAGCCAACCGGGCGCTCTCC
TTGGGAGCATGATCTATAATGTTATCGTTACTGCACATGCCTTGTAATAATTTCATGGTAATACCAATCCTAATCGGGCGCT
CGGAAACTGACTAGTGGCCCTAATGATTGGAGCTCCGGACATGGCATCCCTCGAATGAACAACATGAGCTTGTACTCTCCCCCT
CCTCCTCTCTCTCTGCTCAGGAGTTGAAGCCGGAGCTGGTACTGGATGAACCGTTACCCCTCTAGCCGGAACTCTG
ACCGAGGGCATCCGTGACTAACATTCTCTCCACCTGGAGGTGTTCTCAATTCTGGGCAATTAACTTCACTACAACAA
TCATTAACATGAAACCCCCCGTATTCCAATATCAGACCCCTTATTGCTGATCCGTCTAGTTACGGCGTTCTCTGCTCT
CCTACCGTCTCGCCGGCATTACAATGCTTACAGATGAAACCTAACACCCTCTGACCCCGGGAGGAGGTGACC
CAATCCTGTATCAGCACCTTCTGATTCTGGTCACCCCTGAAGTANN
>1st_BASE_384254_P_D15_FISH_R2
GNNNNGCANTNNTGNGNCCNNNTNNNNCGGGGGTCAGAGAAGGTGGTGTAGGTTGATCTGAAGAAGCATTGTAATGCCGG
CGCGCAGAACGGTAGGGAGAGCAGGAGGAACGGCGTAACTAGGACGGATCAGACGAAATAAGGGGTCTGATATTGGAAATA
GCGGGGGGTTCTGTTAGGTTGATGAAGTTAATTGCCCAAGAATTGAGGAACACCTGCCAGGTGGAGGGAGAAGATGG
TTAGGTCAACGGATGCCCTCGTGGCCAGATTCCGGCTAGAGGGGGTAAACGGTTCATCCAGTACCAAGCTCCGGCTCAACTCT
GAGGAGGCAAGGAGAAGGAGGAAGGAGGGGGAGGAGTCAAAGCTCATGTTGTTCTGAGGGAATGCCATGTCGGAGCTCA
ATCATTAGGGCACTAGTCAGTTCCGAAGCCGGATTAGGATTGGTATTACCATGAAGAAAATTATTACAAACGCATGTCAGTAA
CGATAACATTATAGATCTGATCGTCCCAGGAGAGCGCCGGTGGCTGAGTCTGTCGATAAGCAGGCTTAGGGCTGTGCCTACT
ATTCCGGCCAGGCACCAAATACTAAATAAGGGTGCCTATCTNNNNNTTAGTCGAA
>1st_BASE_384255_P_F3_FISH_F1
NNTTANTGGNNNTGGGGCTTGGGCGGAATAATAGGCACGCCCTAACGCTCTGATCCGAGCAGAACTCAGCCAACCTGGCGCTT
CTGGGGGACGACCAAATCTACATGTTATCGTTACAGCACATGCATTGTAATAATTTCATGGTCATGCCATCATAATTGGAGG
CTTGGAAATTGACTGGTTCTCTAATAATTGGTGCCTGATATGGCTTCCCTCGAATAAAATAATATAAGCTCTGATTACTCT
CTCTCTCTCTCTCTGCTCAGGTGAGAAGCAGGGCGGAACGGTTGAACAGTTATCCCCCTAGCGGGCAACCTGG
ACATGCAGGAGGACATGGAACCTAACCATCTCTCCACATTAGCAGGGGTATCTCAATTCTGGGCAATTAAATTATTACAAAC
CATTATTAATATGAAACCCCCGGCATCTCCAGTATCAGACACCTTATTGCTGAGGCCGTTCAATTACAGCAGTCTACTCT
TCTCTCCAGTTAGCTGGGAATCACTATGCTTCTACAGATGAACTTAATACAACCTCTGACCCCTGCTGGGGAGGGAC
CCTATTCTATACCAACACCTATTCTGATTCTCGNCCCCN

>1st_BASE_384256_P_F3_FISH_R1

NNGGNNNGNTNNNAATAGGGTCCCTCCGCCAGCAGGGTCGAAGAAGGTTGTATTAAAGATTCGATCTGTGAGAAGCATACTGATTCCG
GCAGCTAAAACGGAGAGAAAAGAAGAAGTAGAACTGCTGAATTAGAACGGCTCAGACAAATAAAGGTGCTGATACTGGGAGATG
GCCGGGGTTTCATATTAAATAATGGTTGAATAAAATTAAATGGCCCAAGAATTGAGGATACCCCTGCTAAATGTAGGGAGAAGATGG
TTAGGTCCACTGATGCTCCTCATGTGCCAGGGTGCCTCGTAAGGGGGATAAACTGTTCAACCAGTCCGGCCCTGCTTACACCT
GAGGAAGCAAGGAGAAGAAGGAAAGAGGGGGAGTAATCAGAAAGCTTATTAACTTGGAGGGCTGAGTTCTGCTGGATCAGGAGGTAAGGGCGGTGCCTACTA
ATTATTAGAGGAACCAAGTCATTCCAAAGCCTCAATTATGATAGGCATGACCATGAAGAAAATTAACTACAAATGCATGTGCTGAAC
GATAACATTGTAGATTGGTGTCCCCAGTAAAGGCCAGGGTGGCTGAGTTCTGCTGGATCAGGAGGTAAGGGCGGTGCCTACTA
TCGGCTCAAGCACCAAATACAAGATAAAGGGTGCCAATGTCTTGGTT

>1st_BASE_384257_P_F3_FISH_F2

CNNNNNNNNNTGGGTGCTTNGCCGNATAGTAGGCACCGCCCTAGCCTCTGATCCGAGCAGAACTCAGCCAACCTGGCGCTTT
ACTGGGGGACGACCAAATCTACAATGTTATCGTTACAGCACATGCATTGTAATAATTCTTCATGGTATGCCATCATAATTGGAG
GCTTGGAAATTGACTGGTCTCTAATAATTGGTGCCTGATATGGCTTCCCTGAATAAATAATATAAGCTCTGATTACTCCCC
CCTCTTCTCTCTCTGCTCTCAGGTGAGAAGCAGGGGCCGAACGGTTGAACAGTTATCCCCCTAGCGGAAACCTGG
CACATGCAGGAGCATGAGCTAACCATCTCCCTACATTAGCAGGGTATCCTCAATTCTGGGGCATTAAATTITATTACAA
CCATTATTAAATGAAACCCCCGGCATCTCCAGTATCAGACACCTTATTGTCTGAGGGCTCTAATTACAGCAGTTACTCTC
TTCTCTCCCAGTTAGCTGCCGAATCACTATGCTCTCACAGATCAGAAATCTTAATACAACCTCTCGACCCGCTGGGGAGGG
ACCCATTCTATACCAACACTTCTGATTCTCGNNNCNNCTGGAAAGTATA

>1st_BASE_384258_P_F3_FISH_R2

NNNNNNNGNNNATAGGGCNNTCCGCCAGCAGGGTCGAAGAAGGTTGTATTAAAGATTCGATCTGTGAGAAGCATACTGATTCCG
GCAGCTAAAACGGAGAGAAAAGAAGAAGTAGAACTGCTGAATTAGAACGGCTCAGACAAATAAAGGTGCTGATACTGGGAGATG
GCCGGGGTTTCATATTAAATAATGGTTGAATAAAATTAAATGGCCCAAGAATTGAGGATACCCCTGCTAAATGTAGGGAGAAGATGG
TTAGGTCCACTGATGCTCCTCATGTGCCAGGGTGCCTCGTAAGGGGGATAAACTGTTCAACCAGTCCGGCCCTGCTTACACCT
GAGGAAGCAAGGAGAAGAAGGAAAGAGGGGGAGTAATCAGAAAGCTTATTAACTTGGAGGGAAAGCCATATCAGGGGCCA
ATTATTAGAGGAACCAAGTCATTCCAAAGCCTCAATTATGATAGGCATGACCATGAAGAAAATTAACTACAAATGCATGTGCTGAAC
GATAACATTGTAGATTGGTGTCCCCAGTAAAGGCCAGGGTGGCTGAGTTCTGCTGGATCAGGAGGTAAGGGCGGTGCCTACTA
TCGGCTCAAGCACCAAATACAAGATAAAGGTGCCAATGTCTNTNGNNTTGGTTGAA

>1st_BASE_384231_P_F7_FISH_F1

NNNNNGGGNTGGTGTGAGCCNGATAGTAGGCACCGCCCTAGCCTCTGATCCGAGCAGAACTCAGCCAACCTGGCGCTTTACT
GGGGGACGACCAAATCTACAATGTTATCGTTACAGCACATGCATTGTAATAATTCTTCATGGTATGCCATCATAATTGGAGGCTT
TGGAAATTGACTGGTCTCTAATAATTGGTGCCTGATATGGCTTCCCTGAATAAATAATATAAGCTCTGATTACTCCCCCTC
TTCTCTCTCTCTGCTCTCAGGTGAGAAGCAGGGGCCGAACGGTTGAACAGTTATCCCCCTAGCGGGCAACCTGGCACA
TGCAGGAGCATGAGCTAACATCTCTCCCTACATTAGCAGGGTATCCTCAATTCTGGGGCATTAAATTITATTACACCAT
TATAATGAAACCCCCGGCATCTCCAGTATCAGACACCTTATTGTCTGAGGGCTCTAATTACAGCAGTACTACTTCTCT
CTCCAGTTAGCTGCCGAATCACTATGCTCTCACAGATCAGAAATCTTAATACAACCTCTCGACCCGCTGGGGAGGGACCC
TATCCTATACCAACACCTTCTGATTCTCGGTCCNNNGAAAGTATA

>1st_BASE_384232_P_F7_FISH_R1

NNNNNNNGGTNGGAAGGGTCCCTCCGCCAGCAGGGTCGAAGAAGGTTGTATTAAAGATTCGATCTGTGAGAAGCATACTGATTCCG
GCAGCTAAAACGGAGAGAAAAGAAGAAGTAGTACTGCTGAATTAGAACGGCTCAGACAAATAAAGGTGCTGATACTGGGAGATG
GCCGGGGTTTCATATTAAATAATGGTTGAATAAAATTAAATGGCCCAAGAATTGAGGATACCCCTGCTAAATGTAGGGAGAAGATAG
TTAGGTCCACTGATGCTCCTCATGTGCCAGGGTGCCTCGTAAGGGGGATAAACTGTTCAACCAGTCCGGCCCTGCTTACACCT
GAGGAAGCAAGGAGAAGAAGGAAAGAGGGGGAGTAATCAGAAAGCTTATTAACTTGGAGGGAAAGCCATATCAGGGGCCA
ATTATTAGAGGAACCAAGTCATTCCAAAGCCTCAATTATGATAGGCATGACCATGAAGAAAATTAACTACAAATGCATGTGCTGAAC
GATAACATTGTAGATTGGTGTCCCCAGTAAAGGCCAGGGTGGCTGAGTTCTGCTGGATCAGGAGGTAAGGGCGGTGCCTACTA
TCGGCTCAAGCACCAAATACAAGATAAAGGTGCCAATGTCTTNGGTTGGTT

>1st_BASE_384233_P_F7_FISH_F2

NNNNNNNNNGAGCCGGATAGTGAGGCACCGCCCTAGCCTCTGATCCGAGCAGAACTCAGCCAACCTGGCGCTTTACTGGGGACGA
CCAAATCTACAATGTTATCGTACAGCACATGCATTGTAATAATTCTTCATGGTATGCCATCATAATTGGAGGCTTGGAAATTG
ACTGGTCTCTCTAATAATTGGTGCCTGATATGGCTTCCCTGAATAAATAATATAAGCTCTGATTACTCCCCCTCTTCTCT
CTCTGCTCTCTCAGGTGAGAAGCAGGGGCCGAACGGTTGAACAGTTATCCCCCTAGCGGGCAACCTGGCACATCAGGGC
ATCAGTGGACCTAACATCTCTCCCTACATTAGCAGGGTATCCTCAATTCTGGGGCATTAAATTITATTACAACCATTTAAT
GAAACCCCCGGGCATCTCCAGTATCAGACACCTTATTGTCTGAGGGCTCTAATTACAGCAGTACTACTCTCTCTCCAGTT
TTAGCTGCCGAATCACTATGCTCTCACAGATCAGAAATCTTAATACAACCTCTCGACCCGCTGGGGAGGGACCCATCCTATA
CCAACACCTTCTGATTCTCGNNCCCTGGAAAGTATA

>1st_BASE_384234_P_F7_FISH_R2

GNNNNNNGTTGGAAAGGGTCCCTCCGCCAGCAGGGTCGAAGAAGGTTGTATTAAAGATTCGATCTGTGAGAAGCATACTGATTCCGG
AGCTAAAACGGAGAGAAAAGAAGAAGTAGTACTGCTGAATTAGAACGGCTCAGACAAATAAAGGTGCTGATACTGGGAGATGG
CGGGGGTTTCATATTAAATAATGGTTGAATAAAATTAAATGGCCCAAGAATTGAGGATACCCCTGCTAAATGTAGGGAGAAGATAGTT
AGTCCACTGATGCTCCTCATGTGCCAGGGTGCCTCGTAAGGGGGATAAACTGTTCAACCAGTCCGGCCCTGCTTACACCTGA
GGAAGCAAGGAGAAGAAGGAAAGAGGGGGAGTAATCAGAAAGCTTATTAACTTGGAGGGAAAGCCATATCAGGGGCCA
TATTAGAGGAACCAAGTCATTCCAAAGCCTCAATTATGATAGGCATGACCATGAAGAAAATTAACTACAAATGCATGTGCTGAAC
ATAACATTGTAGATTGGTGTCCCCAGTAAAGGCCAGGGTGGCTGAGTTCTGCTGGATCAGGAGGTAAGGGCGGTGCCTACTA
TCGGCTCAAGCACCAAATACAAGATAAAGGTGCCAATGTCTTGGTT

Appendix B – Consensus Sequence Summary

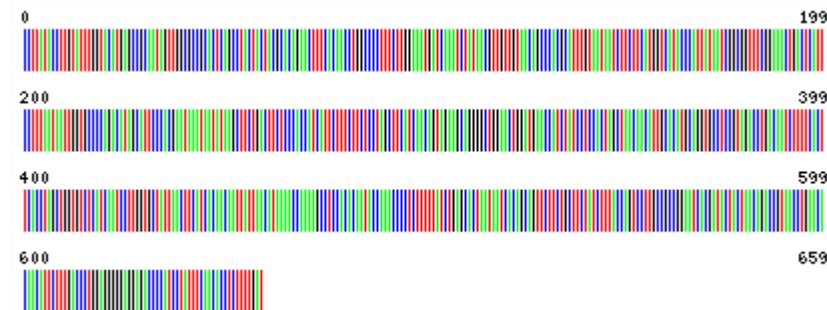
DV 9:

Sequence: 660 bp

Composition: A (184), G (98), C (188), T (190)

```
CCTTATACTTGATTTGGTACATGAGGCCAATAGTGGCGCCGACTCAGCCTACTACTACGCACAGAACATTACAACCTGGCCCC
TTCTTGGAAATGATCAAATCTATAACGTTGTAAACAGCCCACGCATTGTAAATAATCTTCTCATGGTCATACCCATTATAATCGGC
GGTTTGGAAACTGACTCATTCTTAATAATTGGTGCCCGACAGATAGCATTCCCACGAATAAAATAATATAAGCTCTGACTCTCCCA
CCATCATTCTTCTCTCTAGCTTCATCAACAGTAGAACGAGGGGCTGGAACGGATGAACCACATCTATCCTCACTAGCTAACACAT
TGCACATGCAGGTGCCCGTAGACCTGACAATCTTCACCTCACCTAGCTGGTGTCTCATCAATCCTGGTGTATTAACTTCATCAC
AACATTATAACATAAAACCAAAAGCCTCACACAATACCAACCCCTCTTTTATCTGAGCACTAATAATCACAGCAGTTCTCTGC
TTCTATCTTACCACTGCTTGGCGCCGAATCAGATACTAACAGACCGTAACCTGAACACACATTCTTGACCCCTGGAGGGGAG
GGAGACCCCATCTATTCAACACCTCTTTGAT
```

Illustrative barcode:



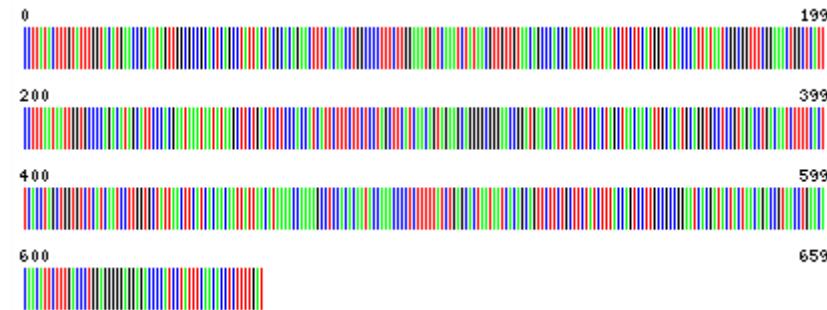
DV 7:

Sequence: 660 bp

Composition: A (184), G (99), C (187), T (190)

```
CCTTATACTTGATTTGGTACATGAACCGCAATAGTGGCGCCGACTCAGCCTATTACTACGCACAGAACATTACAACCTGGCCCC
TTCTTGGAAATGATCAAATCTATAACGTTGTAAACAGCCCACGCATTGTAAATAATCTTCTCATGGTCATACCCATTATAATCGGC
GTTTGGAAACTGGCTCATTCTTAATAATTGGTGCCCGACAGATAGCATTCCCACGAATAAAATAATATAAGCTCTGACTCTCCAC
CATCATTCTTCTCTCTAGCTTCATCAACAGTAGAACGAGGGGCGGAACGGATGAACCACATCTATCCTCACTAGCTAACACATT
GCACATGCAGGTGCCCGTAGACCTGACAATCTTCACCTAGCTGGTGTCTCATCAATCCTGGTGTATTAACTTCATCAC
ACAATTATAACATAAAACCAAAAGCCTCACACAATACCAACCCCTCTTTTATCTGAGCACTAATAATCACAGCAGTTCTCTGCT
TCTATCTTACCACTGCTTGGCGCCGAATCAGATACTAACAGACCGTAACCTGAACACACATTCTTGACCCCTGGAGGGGAG
GAGACCCCATCTATTCAACACCTCTTTGAT
```

Illustrative barcode:



DV 8:

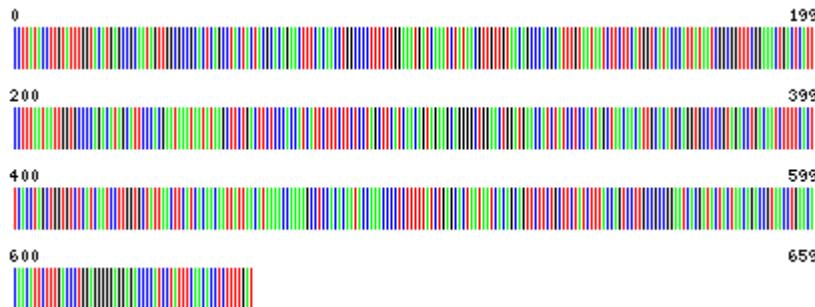
Sequence: 660 bp

Composition: A (184), G (98), C (188), T (190)

```
CCTTATACTTGATTTGGTACATGAGGCCAATAGTGGCGCCGACTCAGCCTACTACTACGCACAGAACATTACAACCTGGCCCC
TTCTTGGAAATGATCAAATCTATAACGTTGTAAACAGCCCACGCATTGTAAATAATCTTCTCATGGTCATACCCATTATAATCGGC
GGTTTGGAAACTGACTCATTCTTAATAATTGGTGCCCGACAGATAGCATTCCCACGAATAAAATAATATAAGCTCTGACTCTCCCA
```

CCATCATTCTTCTCTCATCACAGTAGAACAGCAGGGCTGGAACGGATGAACCACATCTATCCTCACTAGCTAACACAT
TGACATGCAGGTGCCCGTAGACCTGACAATCTTCACCTCACAGCTGGTCTCATCAATCCTGGTCTATTAACTTCATCAC
AACATTATAACATAAAAACCAAGCCTCACACAATACCAACCCCTTTTATCTGAGCAACTAATAACAGCAGTTCTGC
TTCTATCTTACCACTGCTGCCGCCGAATCACGATACTACTAACAGACCGTAACCTGAACACAAACATTGACCTGGAGGGGA
GGAGACCCATCCTATTCAACACCTCTTGAT

Illustrative barcode:



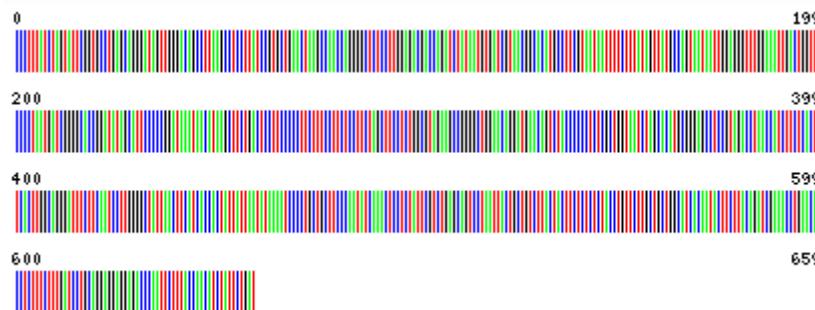
DV 10:

Sequence: 660 bp

Composition: A (154), G (122), C (183), T (201)

CCCTTATCTAGATTGGTGCTGAGCAGGGATAGTTGGACAGCCTTAAGCCTTATCCGTGCTGAACTAAGCCAACCAGGGCT
CTCCTGGAGACGACCAGATCTATAATGTGATCGTAACGGCACATGCCCTCGTAATAATTCTTATAGITATGCCAGTAATAATTGGA
GGGTTGGAAATTGACTGGTCCCTAATGATCGGGCACCGGATATAGCATTCCCCGATAAATAACATAAGCTCTGACTCCTCC
CCCTCTTCCTCTCCCTAGCTCCCTCTGGCGTAGAAGCCGGGCTGGAACAGGATGAACAGTCTACCCCCCTCGCTGGTAATCT
AGCACATGCGGAGGCCCGTAGACCTAACCATCTTCACTTCACTGGCAGGGATTCTCAATCCTGGCTTAACTTCATCAC
CACTATTATAATATAAAATCCCCTGCTGCTTCCAAATCAAACCTCTATTGCTGAGCAGCTCAATTACTGCTGTCTTACTACTT
CTCTCTCACCTGTTCTGCTGGCATCACAAACTTACAGATGAAACACCTTGAACACCTTGTGATCCTGCAGGAGGAGGA
GACCCAATTCTTACCAACATCTATTGAT

Illustrative barcode:



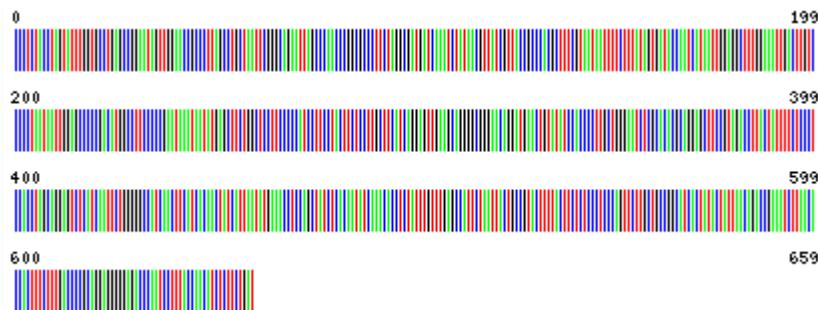
CV 5:

Sequence: 660 bp

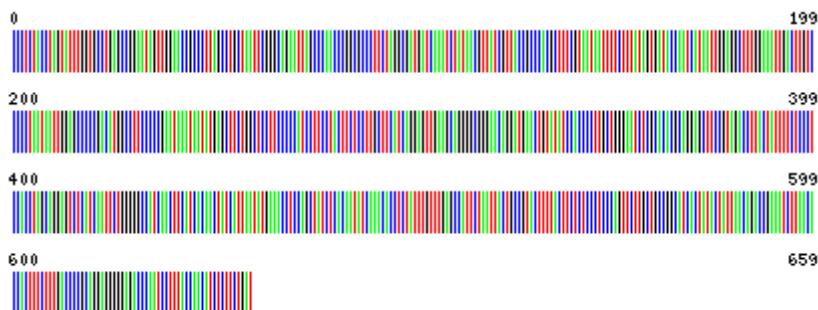
Composition: A (160), G (114), C (196), T (190)

CCCTCTACCTAGTATTGGTGCTGAGCCGAATAGTGGAACGCCCTAGCCTGTAATTGGCAGAATTAGCCAACCCGGGCC
CTCTAGCGATGATCAAATCTATAACGTTATCGTACCGCCACGCTTCGTAATAATTCTTATAGTGTACCAATCATAATTGGA
GGCTTGAAATTGACTTGCCCCCTAATAATTGGAGCCCGACATGGCCTTCCCAGAATAAAATATGAGCTCTGGCTCCTCC
CCATCCTCCTACTCTCTGCTCATCAGGAGTTGAAGCAGGGGGGGAAACAGGATGAACGTATATCCACCCCTGGTGGGAATCT
CGCACACGCAGGAGCTCCGTAGACCTTACTATTCTCCCTCACCTAGCAGGAGTCTCATCAATCTGGGGCCATCAACTCATCAC
AACTATCATTAAATGAAACCTCCAGCTATCTCACAAATATCAAACACCTTAACTTGTGAGCCATTCTAATTACTGCGTACTTTACT
TCTCTCCCTCCAGTTCTGCTGCCGCATCACTACTATTAAACAGACCGAAATCTAACACCACCTTGTGACCCCGCAGGAGGGGG
AGACCCAATCCTTACCAACATCTCTGAT

Illustrative barcode:

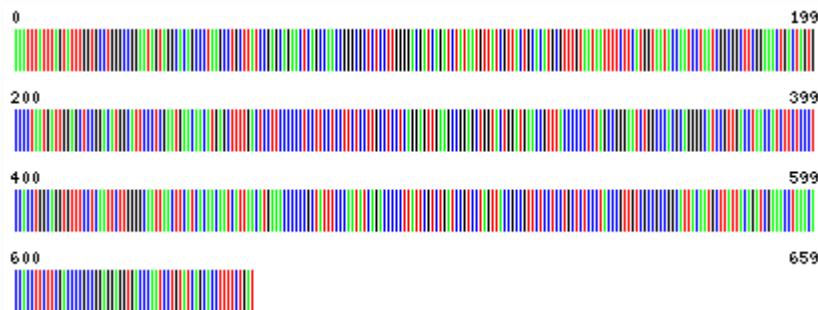
**CV 6:****Sequence:** 660 bp**Composition:** A (160), G (114), C (196), T (190)

CCCTCTACCTAGTATTTGGCCTGAGCGGAATAGTTGAACGCCCTAGCCTGTAATTGGCAGAATTAGCCAACCCGGCGCC
CTTCTAGGCATGATCAAATATAACGTTATCGTACCGCCACGCTTCGTAAATAATTCTTATAGTGATACCAATCATATTGGA
GGCTTGGAAATTGACTTGTCCCCTAATAATTGGAGCCCCGACATGGCCTTCCCCGAATAAAATAATATGAGCTCTGGCTCCTCCC
CCATCCTCTACTTCTCTGCTTCATCAGGAGTTGAAGCAGGGGGGGAAACAGGATGAACGTATCCACCCCTGCTGGGAATCT
CGCACACGAGGAGCTCCAGCTACATTTCTCCACCTAGCAGGAGTCTCATCAATTCTGGGGCCATCAACTTCATCAC
AACTATCATTAAATGAAACCTCCAGCTATCTCACAAATATCAAACACCTTATTTGTTGAGCCATTCTAATTACTGCCGTACTTTACT
TCTCTCCCTCCAGTTCTGCTGCCGGCATCACTACTATTAAACAGACCGAAATCTAACACCACTTCTTGACCCCGAGGAGGGGG
AGACCCAATCTTACCAACATCTCTTGAT

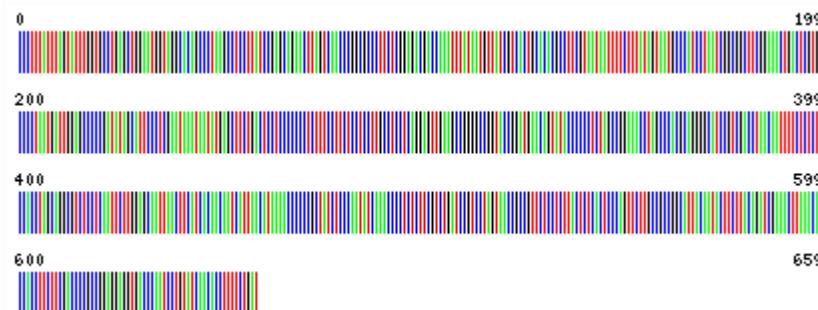
Illustrative barcode:**DV 13:****Sequence:** 660 bp**Composition:** A (144), G (127), C (209), T (180)

AAATTATTTAGTATTTGGCCTGGCGGAATAGTAGGCACAGCCCTAACGCTGCTTATCCGAGCAGAACCTAGCCAACCGGGCGCT
CTCCTGGGGACGATCAGATCTATAATGTTATCGTACTGCACATGCGTTGTAATAATTCTTATGGAATACCAATCTAACCGC
GGCTCGGAAACTGACTAGTAGTGCCTCAATGATTGGAGCTCCGGACATGGCATTCCTCGAATGACAACATGAGCTTTGACTCCTCC
CCCTCCTCCCTCTCCCTGCCTCTCAGGAGTTGAAGCCGGAGCTGGTACTGGATGAACCGTTACCCCTCTAGCCGGAAATCT
GGCCCACGCAGGGCATCGTTGACCTAACCATCTCTCCCTCACCTGGCAGGTGTTCTCAATTCTGGGGCAATTAACTTCATCAC
AACAACTCATTAAACATGAAACCCCCCGCTATTCCAATATCAGACCCCTTATTGCTGATCCGCTAGTTACCGCCGTTCTCCTCCT
GCTCTCCCTACCGTTGTCGCCGGCATACAATGTTATTACAGATGAAACACCACTTCTGACCCCGCCGGAGGAG
GTGACCCAATCTGTATCAGCACCTTCTGAT

Illustrative barcode:

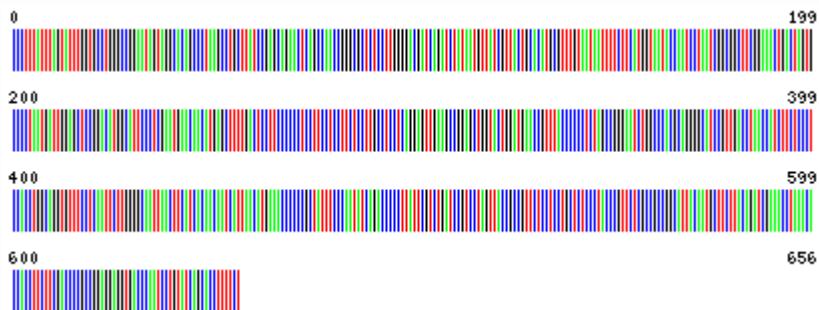
**CV 15:****Sequence:** 660 bp**Composition:** A (154), G (110), C (219), T (177)

CCCTTATTAGTATTGGTGCTGAGCTGGAATGGTAGGCACAGCCTAAGCCTCTTATCGAGCAGAACTTAGTCACCCGGGCC
CTTCTCGGAGACGACCAAATTATAATGTTATCGTCACTGCACACGCCCTCGTAATAATTCTTATAGTAATGCCCATCCTAATCGGC
GGCTTCGAAACTGACTCGTGCCCTAATGATTGGAGCCCCGATATAGCATTCCCTCGAATAAATAATGAGCTTGTACTCCTCCC
CCCCCTTCTCTCTCCCTCAGGAGTTGAAGCCGGGGCGTACCGATGAACAGTATACCCCCCTAGCCGGAAACCT
AGCCCACGCAGGGCATCCGTCACCTAACAAATTCTCTCCACCTAGCAGGGCTCCCTAATCTGGAGCAATTAACTTCATCAC
AACAACTATTAACATAAAACCCCCGCTATCTCCAATATCAAACCCCTCTTCGCTGATCCGCTCTAGTAACCGCCGTTCTCCTCTT
ACTCTCACTCCAGTTGCGCCGGCATACAATACTTCTACAGATCGAAACTAAACACCACCTTCTCGACCCCGCCGGAGGAG
GTGACCAATCCTGTATCAACACCTTTCTGAT

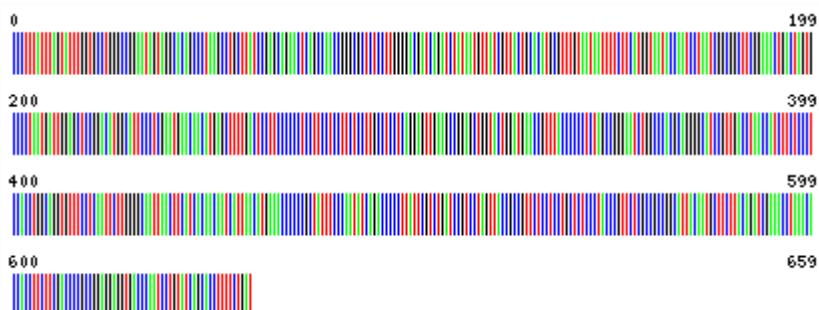
Illustrative barcode:**DV 15:****Sequence:** 657 bp**Composition:** A (139), G (125), C (214), T (179)

CCCTTATTAGTATTGGTGCTGGCCGAATAGTAGGCACAGCCTAAGCCTGTTATCGAGCAGAACTCAGCCAACCGGGCGCT
CTCCTGGGGACGATCAGATCTATAATGTTATCGTACTGCACATGCGTTGTAATAATTCTCATGGAATACCAATCCTAATCGGC
GGCTTCGAAACTGACTAGTGCCCTAATGATTGGAGCTCCGGACATGGCATTCCCTCGAATGAAACATGAGCTTTGACTCCTCC
CCCCCTTCTCTCTCCCTCAGGAGTTGAAGCCGGAGCTGGTACTGGATGAACCGTTACCCCCCTAGCCGGAAATCT
GGCCACGCAGGGCATCCGTTGACCTAACCATCTCTCCACCTGGCAGGTGTTCCCTCAATTCTGGGCAATTAACTTCATCAC
AACAACTATTAACATGAAACCCCCGCTATTCCAATATCAGACCCCTTATCGTCTGATCCGCTCTAGTTACCGCCGTTCTCCTCT
GCTCTCCCTACCCGTTCTCGCCGCCGGATTACAATGCTTACAGATCGAAACCTAAACACCACCTTCTCGACCCCGCCGGAGGAG
GTGACCAATCCTGTATCAGCACCTTTCTNNN

Illustrative barcode:

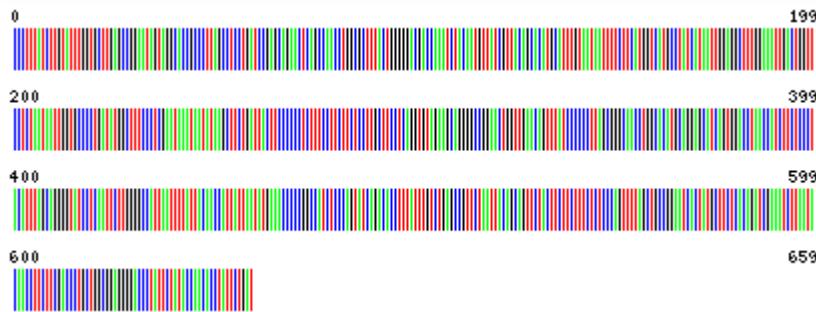
**DV 14:****Sequence:** 660 bp**Composition:** A (140), G (126), C (214), T (180)

CCCTTATTAGTATTGGTGCCTGGGCGGAATAGTAGGCACAGCCCTAACGCCGCTTATCCGAGCAGAACTCAGCCAACCGGGCGCT
CTCCTGGGGACGATCAGATATAATGTTATCGTACTGCACATGCCGTTGAATAATTTCATGGTAATACCAATCCTAATCGC
GGCTTCGAAACTGACTAGTGCCCCATGATTGGAGCTCCGGACATGGCATCCCTCGAATGACAACATGAGCTTTGACTCCTCC
CCCCTCCTCCCTCTCCCTCAGGAGTTGAAGCCGGAGCTGACTGGATGAACCGTTACCCCTCTAGCCGGAAATCT
GGCCCACGCAGGGCATCCGTTGACCTAACATCTTCCCTCCACCTGGCAGGTGTTCTCAATTCTGGGCAATTAACTTCATCAC
AACAACTATTAAACATGAAACCCCCCGCTATTCCCACATTCAGACCCCTTATTCGCTGATCCGCTCAGTTACGCCGTTCTCCCT
GCTCCCTACCGTTCTGCCGCCGGCATACATGCTTACAGATGAAACCTAACACCACCTCTCGACCCGCCGGAGGAG
GTGACCCAATCTGTATCAGCACCTTTCTGAT

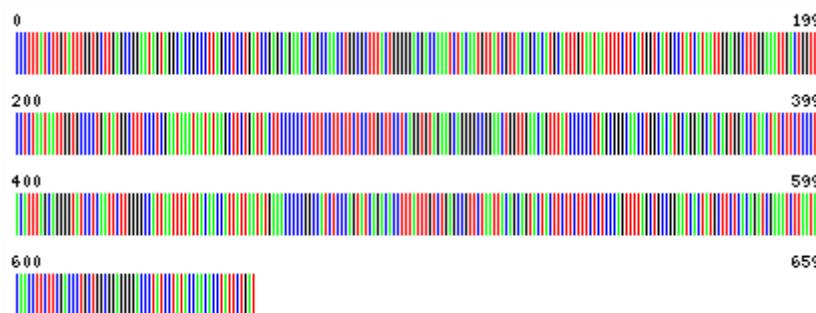
Illustrative barcode:**FV 3:****Sequence:** 660 bp**Composition:** A (153), G (121), C (180), T (206)

CCCTTATCTGTATTGGTGCCTGGAGCCGAATAGTAGGCACGCCCTAACGCCGCTTATCCGAGCAGAACTCAGCCAACCTGGCGCT
TTACTGGGGACGACAAATCTACAATGTTATCGTACAGCACATGCCGTTGAATAATTTCATGGCATGCCATCATAATTGGA
GGCTTGAAATTGACTGGTCTCTAATAATTGGTGCCTGATATGGCTTCCCTCGAATAAAATAATATAAGCTCTGATTACTTCCC
CCCTCTTCTCTCTCCTGCTCAGGTGAGAACGAGGGCCGAACGGTTGAACAGTTATCCCCCTAGCCGGAAACCTG
GCACATGCAGGAGCATCAGTGACCTAACATCTTCCCTACATTAGCAGGGGTATCCTCAATTCTGGGGCATTAATTATTACA
ACCATTATTAAATATGAAACCCCCGGCATCTCCAGTATCAGACACCTTATTGCTGAGCCGTTCAATTACAGCAGTTACTTCTT
CTTCTCTCCAGTTAGCTGCCGAATCACTATGCTTACAGATCGAAATCTAACACCTCTCGACCCGCTGGGGAGGG
GACCCATTCTATACCAACACCTATTCTGAT

Illustrative barcode:

**FV 7:****Sequence:** 660 bp**Composition:** A (154), G (121), C (180), T (205)

CCCTTTATCTTGATTGGTGCCTGACCGGAATAGTAGGCACCGCCCTAGCCTCTGATCCGAGCAGAACTCAGCCAACCTGGCGCT
 TTACTGGGGACGACCAAATCTACAATGTTACAGCACATGCATTGTAATATTTCATGGTCATGCCATCATATAATTGGA
 GGCTTGGAAATTGACTGGTCTCTAATAATTGGTGCCTGATATGGCTTCCCTGAATAAAATAAGCTTCTGATTACTTCCC
 CCCCTTTCTCTCTCTGCTCCCTCAAGGTGAGAACGCAGGGCCGAACTGGTGAACAGTTATCCCCCTAGCGGGCAACCTG
 GCACATGCAGGAGCATCAGTGGACCTAACTATCTCTCCCTACATTAGCAGGGGTATCCTCAATTCTGGGCCATTAAATTATTACA
 ACCATTAAATATGAAACCCCGGCCATCTCCAGTATCAGACACCTTATTGCTGAGCCGTTCAATTACAGCAGTACTACTTCTT
 CTTCTCTCCAGTTAGCTGCCGAATCACTATGCTCTCACAGATCGAAATCTAACACCTCTCGACCCGTGGCGGAGGG
 GACCCTATCCTATACCAACACCTATTCTGAT

Illustrative barcode:

Appendix C – Matang Specimen Info

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
BS 1	BV 1	T1, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Acentrogobius caninus</i>	Tropical sand goby	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°51'01.1, E: 100°31'13.5	n/a
DS 6	DV 6	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 7	DV 7	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 8	DV 8	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 9	DV 9	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 10	DV 10	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 11	DV 11	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Allenbatrachus grunniens</i>	Grunting toadfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
AS 4	AV 4	T1, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Ambassis gymnocephalus</i>	Bald glassy	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'06.7, E: 100°29'11.6	n/a
CS 23	CV 23	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Anodontostoma chacunda</i>	Chacunda gizzard shad	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 1	CV 1	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 3	CV 3	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 5	CV 5	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 6	CV 6	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 7	CV 7	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
FS 16	FV 16	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Arius maculatus</i>	Spotted catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
AS 17	AV 17	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Hexanematichthys sagor</i>	Sagor catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
CS 2	CV 2	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Hexanematichthys sagor</i>	Sagor catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
DS 1	DV 1	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Hexanematichthys sagor</i>	Sagor catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
HS 4	HV 4	T4, St. 4	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Hexanematichthys sagor</i>	Sagor catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'26.7, E: 100°29'05.2	n/a
CS 4	CV 4	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Cryptarius truncatus</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
DS 3	DV 3	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Cryptarius truncatus</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 4	DV 4	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Cryptarius truncatus</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
HS 1	HV 1	T4, St. 4	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Cryptarius truncatus</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'26.7, E: 100°29'05.2	n/a
HS 2	HV 2	T4, St. 4	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Cryptarius truncatus</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'26.7, E: 100°29'05.2	n/a
FS 12	FV 12	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Arius oetik</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
FS 13	FV 13	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Arius oetik</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
HS 3	HV 3	T4, St. 4	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Arius oetik</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'26.7, E: 100°29'05.2	n/a
DS 2	DV 2	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Arius venosus</i>	Veined catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 5	DV 5	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Arius venosus</i>	Veined catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
FS 17	FV 17	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Arius venosus</i>	Veined catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
GS 2	GV 2	T4, St. 3	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Arius venosus</i>	Veined catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
GS 4	GV 4	T4, St. 4	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Arius venosus</i>	Veined catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'26.7, E: 100°29'05.2	n/a
BS 6	BV 6	T3, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Butis koilimatomodon</i>	Mud sleeper	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°51'57.7, E: 100°31'20.0	n/a
CS 17	CV 17	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Butis koilimatomodon</i>	Mud sleeper	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
FS 11	FV 11	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Carangidae spp.</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
FS 14	FV 14	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Carangidae spp.</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
CS 18	CV 18	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Chelonodon patoca</i>	Milkspotted puffer	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 19	CV 19	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Chelonodon patoca</i>	Milkspotted puffer	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 20	CV 20	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Chelonodon patoca</i>	Milkspotted puffer	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
ES 5	EV 5	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Chelonodon patoca</i>	Milkspotted puffer	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
ES 6	EV 6	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Chelonodon patoca</i>	Milkspotted puffer	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
AS 9	AV 9	T2, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Cynoglossus lingua</i>	Long tongue sole	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'22.5, E: 100°29'06.2	n/a
AS 15	AV 15	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Cynoglossus lingua</i>	Long tongue sole	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 16	AV 16	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Cynoglossus lingua</i>	Long tongue sole	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 18	AV 18	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Cynoglossus lingua</i>	Long tongue sole	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 19	AV 19	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Cynoglossus lingua</i>	Long tongue sole	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 14	AV 14	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Terapon theraps</i>	Largescaled terapon	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
FS 9	FV 9	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Gerres erythrourus</i>	Deep-bodied mojara	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 15	FV 15	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Gerres filamentosus</i>	Whipfin silver-biddy	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
DS 12	DV 12	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Glosogobius spp.</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
AS 2	AV 2	T1, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'06.7, E: 100°29'11.6	n/a
AS 3	AV 3	T1, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'06.7, E: 100°29'11.6	n/a
AS 6	AV 6	T2, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'22.5, E: 100°29'06.2	n/a
AS 7	AV 7	T2, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'22.5, E: 100°29'06.2	n/a
AS 8	AV 8	T2, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'22.5, E: 100°29'06.2	n/a
DS 23	DV 23	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 24	DV 24	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 25	DV 25	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius belangerii</i>	Belanger's croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
AS 1	AV 1	T1, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius carouna</i>	Caroun croaker	Gatu UM	n/a	UM	n/a	Sexual	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'06.7, E: 100°29'11.6	n/a
AS 20	AV 20	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius carouna</i>	Caroun croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
BS 2	BV 2	T1, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius carouna</i>	Caroun croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°51'01.1, E: 100°31'13.5	n/a
BS 3	BV 3	T2, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Johnius carouna</i>	Caroun croaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'57.3, E: 100°31'17.1	n/a

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
DS 17	DV 17	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius carouna</i>	Caroun cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.7	n/a
CS 21	CV 21	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
DS 18	DV 18	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.8	n/a
DS 19	DV 19	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 20	DV 20	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 21	DV 21	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 22	DV 22	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
FS 5	FV 5	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 10	FV 10	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
FS 19	FV 19	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
FS 20	FV 20	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
FS 21	FV 21	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
GS 3	GV 3	T4, St. 3	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Johnius weberi</i>	Weber's cownaker	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
ES 4	EV 4	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Mystus galio</i>	Long whiskers catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
CS 16	CV 16	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Oreochromis niloticus</i>	Nile tilapia	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
FS 18	FV 18	T4, St. 3	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Osteogeneiosus militaris</i>	Soldier catfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
CS 15	CV 15	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Pomadasys kaupan</i>	Javelin grunter	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
DS 13	DV 13	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Pomadasys kaupan</i>	Javelin grunter	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 14	DV 14	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Pomadasys kaupan</i>	Javelin grunter	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
DS 15	DV 15	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Pomadasys kaupan</i>	Javelin grunter	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a
FS 1	FV 1	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus striatus</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
FS 2	FV 2	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 3	FV 3	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 4	FV 4	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 6	FV 6	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 7	FV 7	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
CS 8	CV 8	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 9	CV 9	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 10	CV 10	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 11	CV 11	T1, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'33.0, E: 100°35'47.6	n/a
CS 12	CV 12	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 13	CV 13	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 14	CV 14	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
DS 26	DV 26	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 27	DV 27	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
DS 28	DV 28	T2, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.6, E: 100°35'44.2	n/a
ES 1	EV 1	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
ES 2	EV 2	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
ES 3	EV 3	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
FS 8	FV 8	T1, St. 5	Anna Fogelström	2009-04-12	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'01.9, E: 100°29'10.1	n/a
GS 1	GV 1	T4, St. 3	Anna Fogelström	2009-04-13	UM	n/a	n/a	<i>Scatophagus argus</i>	Spotted scat	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'38.2, E: 100°29'12.8	n/a
DS 16	DV 16	T1, St. 1	Anna Fogelström	2009-03-13	UM	n/a	n/a	<i>Siganus canaliculatus</i>	White-spotted spinefoot	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'41.7, E: 100°35'37.6	n/a

Specimen ID	Voucher ID	Collection Code	Collectors	Collection Date	Storing Institution	Sample Donor	Donor E-mail	Taxonomy	Common name	Identifier	Identifier E-mail	Identifier Institution	Sex	Reproduction	Life Stage	Extra Info	Notes	Continent/Ocean	Country	State/Province	Region	Exact Site	GPS Coordinates (in)	Elevation Depth
CS 24	CV 24	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Stolephorus spp.</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 25	CV 25	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Stolephorus spp.</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
CS 22	CV 22	T2, St. 2	Anna Fogelström	2009-03-12	UM	n/a	n/a	<i>Thryssa kammalensis</i>	Kammal thyssa	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'35.8, E: 100°35'50.5	n/a
AS 10	AV 10	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 11	AV 11	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 12	AV 12	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
AS 13	AV 13	T3, St. 4	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'37.3, E: 100°29'10.8	n/a
BS 4	BV 4	T2, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'57.3, E: 100°31'17.1	n/a
BS 5	BV 5	T2, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°50'57.3, E: 100°31'17.1	n/a
BS 7	BV 7	T3, St. 3	Anna Fogelström	2009-03-11	UM	n/a	n/a	<i>Trypauchen vagina</i>	n/a	Gatu UM	n/a	UM	n/a	n/a	Adult	n/a	n/a	Andaman Sea (Malacca Strait)	Malaysia	Perak	Matang	Matang Mangrove forest	N: 4°51'57.7, E: 100°31'20.0	n/a

Appendix D – Matang Specimen Taxonomy

Species	Common name	Class	Order	Family
<i>Acentrogobius caninus</i>	Tropical sand goby	Actinopterygii	Perciformes	Gobiidae
<i>Allenbatrachus grunniens</i>	Grunting toadfish	Actinopterygii	Batrachoidiformes	Batrachoididae
<i>Ambassis gymnocephalus</i>	Bald glassy	Actinopterygii	Perciformes	Ambassidae
<i>Anodontostoma chacunda</i>	Chacunda gizzard shad	Actinopterygii	Clupeiformes	Clupeidae
<i>Arius maculatus</i>	Spotted catfish	Actinopterygii	Siluriformes	Ariidae
<i>Hexanematichthys sagor</i>	Sagor catfish	Actinopterygii	Siluriformes	Ariidae
<i>Cryptarius truncatus</i>	n/a	Actinopterygii	Siluriformes	Ariidae
<i>Arius oetik</i>	n/a	Actinopterygii	Siluriformes	Ariidae
<i>Arius venosus</i>	Veined catfish	Actinopterygii	Siluriformes	Ariidae
<i>Butis koiomatodon</i>	Mud sleeper	Actinopterygii	Perciformes	Eleotridae
<i>Carangidae spp.</i>	n/a	Actinopterygii	Perciformes	Carangidae
<i>Chelonodon patoca</i>	Milkspotted puffer	Actinopterygii	Tetraodontiformes	Tetraodontidae
<i>Cynoglossus lingua</i>	Long tongue sole	Actinopterygii	Pleuronectiformes	Cynoglossidae
<i>Terapon theraps</i>	Largescaled terapon	Actinopterygil	Perciformes	Terapontidae
<i>Gerres erythrourus</i>	Deep-bodied mojarra	Actinopterygil	Perciformes	Gerreidae
<i>Gerres filamentosus</i>	Whipfin silver-biddy	Actinopterygil	Perciformes	Gerreidae
<i>Glossogobius spp.</i>	n/a	Actinopterygii	Perciformes	Gobiidae
<i>Johnius belangerii</i>	Belanger's croaker	Actinopterygii	Perciformes	Sciaenidae
<i>Johnius carouna</i>	Caroun croaker	Actinopterygii	Perciformes	Sciaenidae
<i>Johnius weberi</i>	Weber's croaker	Actinopterygii	Perciformes	Sciaenidae
<i>Mystus gulio</i>	Long whiskers catfish	Actinopterygii	Siluriformes	Bagridae
<i>Oreochromis niloticus</i>	Nile tilapia	Actinopterygii	Perciformes	Cichlidae
<i>Osteogeneiosus militaris</i>	Soldier catfish	Actinopterygii	Siluriformes	Ariidae
<i>Pomadasys kaakan</i>	Javelin grunter	Actinopterygii	Perciformes	Haemulidae
<i>Pseudotriacanthus strigilifer</i>	Long-spined tripodfish	Actinopterygii	Tetraodontiformes	Triacanthidae
<i>Scatophagus argus</i>	Spotted scat	Actinopterygii	Perciformes	Scatophagidae
<i>Siganus canaliculatus</i>	White-spotted spinefoot	Actinopterygii	Perciformes	Siganidae
<i>Stolephorus spp.</i>	n/a	Actinopterygii	Clupeiformes	Engraulidae
<i>Thryssa kammalensis</i>	Kammal thryssa	Actinopterygii	Clupeiformes	Engraulidae
<i>Trypauchen vagina</i>	n/a	Actinopterygii	Perciformes	Gobiidae