

## CHAPTER 3: METHODOLOGY

### 3.1: SAMPLING LOCATION

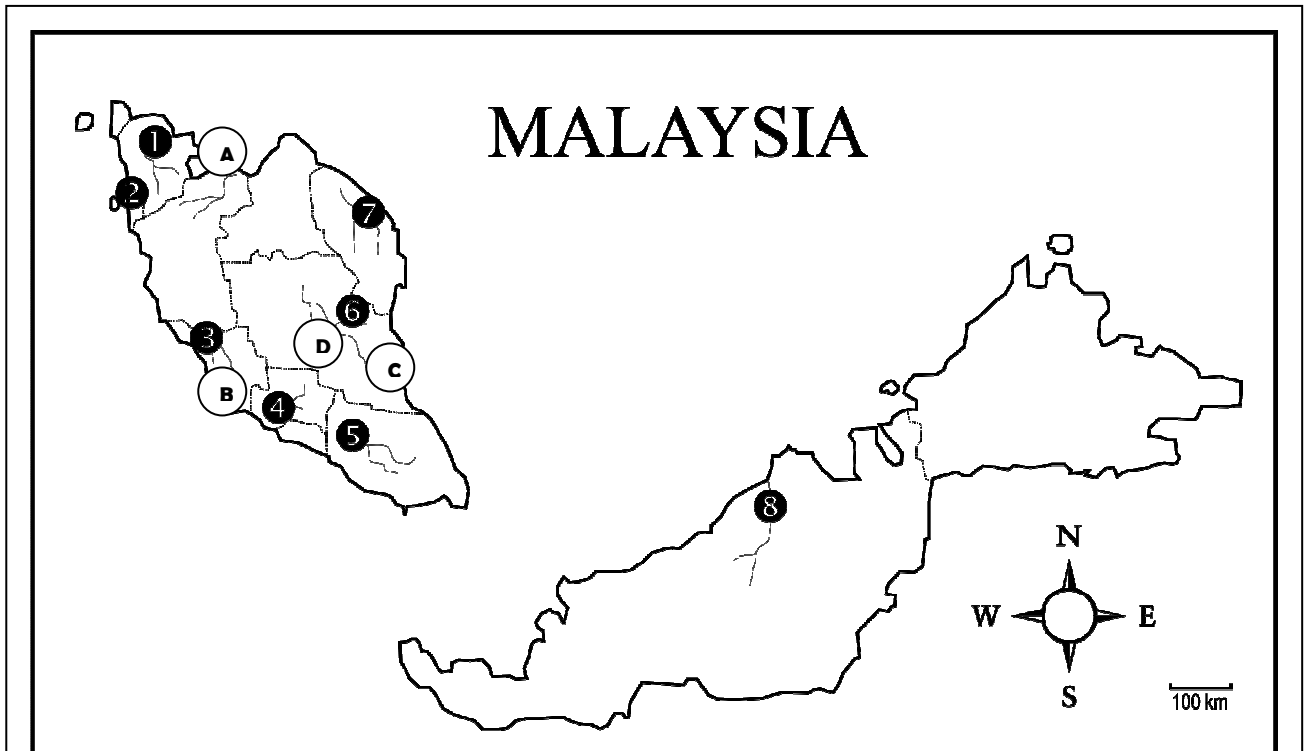
A total of 220 specimens of *C. striata* representing eight locations of Kedah, Pulau Pinang, Selangor, Negeri Sembilan, Johor, Pahang, Terengganu and Sarawak, and additional 83 specimens of 26 freshwater fish species from three locations of Pahang, Selangor and Perak were collected in this study. Refer **Figure 3.1** for sampling distribution map. All the samples were caught through electrofishing and simple fish net with the exception of *C. striata* from Sarawak which were purchased from local market. All the fish samples were positively identified to species level on the time of collection.

The barcoded samples in this study consist of a total of 126 specimens which represent members of 27 previously described freshwater fish species from 23 genera, 7 families, and 3 orders. A total of 116 freshwater fish barcode sequences of 27 species from Peninsular Malaysia have been submitted for publication, excluding Sarawak *C. striata* samples which had not been obtained at the time of submission (Song, *et al*, 2011; manuscript submitted to *GENE*). These have been submitted to the Genbank database with accession numbers HM156340-HM156344, HM156350-HM156372, HM156377-HM156390, HM156392 and JF781170-JF781242, and had submitted to Barcode of Life Database (BOLD) under project code – FWFMM. Multiple individuals for each species were collected whenever possible to provide an indication of variation within species. In the case of estimation of the level of intra-species variation with the influence of geographically induced divergence, 43 *C. striata* individuals from eight sites across a wide

geographical range in Malaysia were used in this study to provide an estimate of the likely extent of intra-specific divergence present among Malaysia freshwater taxa.

A total of 43 *C. striata* specimens with sample size range from four to ten individuals as representatives from each location examined for DNA barcoding were used in investigating evolutionary history and population differentiation of *C. striata* in Malaysia. Briefly, all the mtDNA COI barcode sequence obtained for *C. striata* in this study was subsequently used for population genetic analysis of *C. striata*. All the 220 specimens of *C. striata* sampled from the eight surveyed locations in this study were used for morphology analysis at population level.

Muscle tissue of each sample was stored individually in microcentrifuge tube preserved with absolute ethanol and kept in -80°C freezer prior to genetic analysis. Most of the voucher specimens were preserved in absolute ethanol and lodged at the Repository Store located at Freshwater Fisheries Research Division of Jelebu (Department of Fisheries Malaysia, Freshwater Fisheries Research Division, FRI Glami Lemi, Jelebu, 71650 Titi, Negeri Sembilan, Malaysia; [www.pppat.gov.my](http://www.pppat.gov.my); [ffrc@pppat.gov.my](mailto:ffrc@pppat.gov.my)). Some voucher specimens such as *C. striata* from Sarawak were stored at the -20°C freezer in the laboratory located at University of Malaya. Refer **Appendix A** for specimen's details, **Table 3.1** for specimens used in DNA barcoding, and **Table 3.2** for *C. striata* specimens used in both molecular and morphometric analysis.



A-D: Sampling site of additional selected 26 freshwater fish species  
 1-8: Sampling site of *Channa striata*

	State	Site	Latitude	Longitude
A	Perak	Temenggor	5.51	101.36
B	Selangor	Kepong	3.230603	101.634806
C	Pahang	Lubuk Paku	3.5101406	102.7738548
D	Pahang	Kuala Mai	3.7736179	102.3664341
1	Kedah	Kubur Panjang	6.11427	100.54928
2	Pulau Pinang	Permatang Kuang	5.40665	100.36783
3	Selangor	Tanjung Karang	3.44285	101.1834
4	Negeri Sembilan	Sikamat	2.75925	101.94901
5	Johor	Batu Pahat	1.8586	102.93777
6	Pahang	Pahang River	3.69172	102.89516
7	Terengganu	Terengganu River	5.3795	103.08884
8	Sarawak	Miri River	4.6134368	114.3129719

Figure 3.1: Sampling locations of 220 specimens of *Channa striata* and additional 83 freshwater fishes constituting selected 26 freshwater fish species used in this study. Exact latitude and longitude for each sampling sites are being recorded. Number 1-8 and alphabet A-D correspond to locations listed in **Table 3.1**.

Table 3.1: List of 27 selected freshwater fishes species used in DNA barcoding study. Numbers and letters under location refer to **Figure 3.1**.

Order	Family	Genus	Species	No of Individual	Local Name	Location	
Cypriniformes	Cyprinidae	Amblyrhynchichthys	<i>Amblyrhynchichthys truncatus</i>	5	Tempuling	(C)	
		Barbichthys	<i>Barbichthys laevis</i>	4	Bentulu	(C)	
		Barbonymus	<i>Barbonymus schwanefeldii</i>	5	Lampam Sungai	(C)	
		Chirrinus	<i>Cirrhinus caudimaculatus</i>	2	Selimang batu	(D)	
		Cyclocheilichthys	<i>Cyclocheilichthys apogon</i>	3	Temperas	(D)	
		Devario	<i>Devario regina</i>	3	Danio	(A)	
		Hampala	<i>Hampala macrolepidota</i>	3	Sebarau	(C, A)	
		Hypsibarbus	<i>Hypsibarbus wetmorei</i>	5	Kerai Kunyit	(C)	
		Luciosoma	<i>Luciosoma setigerum</i>	2	Nyenyuar	(D)	
		Osteochilus	<i>Osteochilus vittatus</i>	4	Terbul	(C)	
		Poropuntius	<i>Poropuntius smedleyi</i>	5	Tengas daun	(B)	
		Probarbus	<i>Probarbus jullieni</i>	1	Temoleh	(C)	
		Puntioplites	<i>Puntioplites bulu</i>	5	Tengalan	(D)	
			<i>Puntioplites proctozystron</i>	4	Kerengkek	(C)	
		Puntius	<i>Puntius banksi</i>	3	Tengas	(B)	
		Rasbora	<i>Rasbora sumatrana</i>	3	Seluang	(C, D)	
			<i>Rasbora paviana</i>	2	Seluang	(A, B)	
			Thynnichthys	<i>Thynnichthys thynnoides</i>	2	Lomah	(D)
		Perciformes	Anabantidae	Anabas	<i>Anabas testudineus</i>	1	Puyu
Channidae	Channa		<i>Channa striata</i>	43	Haruan	(1, 2, 3, 4, 5, 6, 7, 8)	
		<i>Channa lucius</i>	1	Bujuk	(B)		
Siluriformes	Bagridae	Hemibagrus	<i>Hemibagrus wyckii</i>	2	Baung kelulang	(D)	
			<i>Hemibagrus nemurus</i>	3	Baung	(A)	
	Pangasidae	Pangasius	<i>Pangasius nasutus</i>	4	Patin buah	(D)	
		Pseudolais	<i>Pseudolais micronemus</i>	5	Patin juara	(D)	
	Siluridae	Belodontichthys	<i>Belodontichthys dinema</i>	2	Gerahak	(D)	
	Sisoridae	Bagarius	<i>Bagarius yarrelli</i>	4	Kenderap	(D)	
<b>3 Order</b>	<b>7 Families</b>	<b>23 Genus</b>	<b>27 Species</b>	<b>126 fishes</b>	<b>27 species</b>	<b>12 Sites</b>	

Table 3.2: List of *Channa striata* specimens used in both molecular and morphometric analysis.

No	Location	Site	Latitude	Longitude	Number of individual collected	No. of individual used in molecular analysis	No. of individual used in morphometric analysis
1	Kedah	Kubur Panjang, Paddy Fields	6.11427	100.54928	30	4	30
2	Pulau Pinang	Permatang Kuang, Paddy Fields	5.40665	100.36783	30	5	30
3	Selangor	Tanjung Karang, Small Stream	3.44285	101.1834	30	5	30
4	Negeri Sembilan	Sikamat, Paddy Fields	2.75925	101.94901	30	5	30
5	Johor	Batu Pahat Dam	1.8586	102.93777	30	5	30
6	Pahang	Pahang River	3.69172	102.89516	30	5	30
7	Terengganu	Terengganu River	5.3795	103.08884	30	4	30
8	Sarawak	Miri River	4.6134368	114.3129719	10	10	10
<b>Total</b>	<b>8 locations</b>	<b>8 sites</b>			<b>220 specimens</b>	<b>43 specimens</b>	<b>220 specimens</b>

## 3.2: GENOMIC ANALYSIS

### 3.2.1: DNA EXTRACTION

DNA was extracted using AxyPrep<sup>TM</sup> Genomic DNA Extraction Kit according to the animal tissues spin protocol provided by the company (AxyPrep<sup>TM</sup> Genomic DNA Extraction Kit website: <http://www.axxygenbio.com/collections/vendors?page=2&q=Axygen>). About 0.02g of animal from each sample was transferred into the mortar. The tissue sample was ground in liquid nitrogen using a sterilized mortal and pestle until it formed a homogenate.

Next, 650µl of Buffer G-A was added followed by 0.9µl of RNase A. The mixture was grinded gently for 30 seconds to homogenously mix the Buffer G-A with the ground tissue. Subsequently, 650µl of homogenate was collected and transferred to a 2ml microfuge tube. The collected homogenate was then incubated in a 65°C water bath overnight.

After incubation, 400µl of Buffer G-B was added followed by 1 ml of Buffer DV (pre-chilled to 4°C). The mixture was mixed vigorously and centrifuged at 12,000xg for 2 minutes. After centrifugation, the upper phase was aspirated off as much as possible. The inter-phase precipitate and the lower-phase were kept in the tube.

Next, 1 ml of Buffer DV (pre-chilled to 4°C) was added. The mixture was mixed well and centrifuged at 12,000xg for 2 minutes. After centrifugation, the upper surface of the solution was discarded, whereas the lower phase was transferred to a spin-filter (placed in a 2ml microfuge tube) and was centrifuged at 12,000xg for 1 minute. The spin-filter was

discarded in the subsequent step. Then, 400 $\mu$ l of Buffer BV was added to the flow-through and the mixture was mixed well by brisk, repeated inversion.

A miniprep column was placed to a 2ml Microfuge tube and the binding mixture from the previous step was transferred into the miniprep column. The binding mixture was then centrifuged at 12,000xg for 1 minute. After centrifugation, the resulting filtrate was discarded and the miniprep column was placed back into the 2ml microfuge tube.

Next, 500 $\mu$ l of Buffer W1 was added to the miniprep column and was centrifuged at 12,000xg for 1 minute. After centrifugation, the filtrate was discarded and the miniprep column was placed back into the 2ml microfuge tube.

The subsequent step was followed by addition of 700 $\mu$ l of Buffer W2 and centrifugation at 12,000xg for 1 minute. After that, the filtrate was discarded and the miniprep column was placed back into the 2ml microfuge tube.

The wash done by the addition of Buffer W2 (ethanol added) from the previous step was repeated by the addition of second 700 $\mu$ l aliquot of Buffer W2 (ethanol added) and centrifugation at 12,000xg for 1 minute. After centrifugation, the filtrate was discarded and the miniprep column was placed back into the 2ml microfuge tube. Next, another centrifugation was carried out at 12,000xg for 1 minute.

The miniprep column was then transferred into a clean 1.5ml microfuge tube after the centrifugation. After that, 200µl of Eluent (preheated at 65°C) was added to the center of the membrane to elute the genomic DNA.

Finally, the binding mixture was incubated overnight at room temperature and a centrifugation at 12,000xg was done for 1 minute to centrifuge the DNA. The extracted DNA was then checked for its purity through the spectrophotometric measurement of UV absorbance using a spectrophotometer (Eppendorf, Germany). The resulting optical density ( $OD_{260/280nm}$ ) ratio was read and the range was recorded.



### 3.2.2: POLYMERASE CHAIN REACTION (PCR)

The primer pair used in this study to amplify approximately 700bp DNA fragment of 5' end of mitochondrial cytochrome *c* oxidase subunit 1 (COI) was:

Fish F1: 5'TCAACCAACCACAAAGACATTGGCAC3'

Fish R1: 5'TAGACTTCTGGGTGGCCAAAGAATCA3' (Ward *et al.*, 2005).

Polymerase chain reaction (PCR) was performed using this primer pair. PCR amplification was carried out with total volume of 20  $\mu$ l of PCR mixture which contain 6.8 $\mu$ l of distilled water, 6 $\mu$ l of buffer (Promega), 2 $\mu$ l of MgCl<sub>2</sub> (25mM), 0.5 $\mu$ l of each dATP, dGTP, dCTP and dTTP (10mM), 0.5 $\mu$ l of each primer (10 $\mu$ M), 1.6 $\mu$ l of DNA (approx. 80ng) from each sample and 0.6 $\mu$ l Promega GoTaq® Polymerase (5u/ $\mu$ l). The cycling parameters on a Multigene Thermal Cycle were as followed: an initial denaturation at 96°C for 5 min, 38 cycles of denaturation at 94°C for 45sec, annealing at 45°C for 45sec, and extension at 72°C for 30sec. This was followed by a final extension at 72°C for 10 min and finally the PCR product was holds at 4°C. In order to minimise mis-priming and maximise amplification success prior to sequencing, both PCR mixture and annealing temperature were optimized for different samples. To achieve optimum condition for PCR reaction, the amount of 25mM MgCl<sub>2</sub> was optimized in a range of 2.0 to 2.3 $\mu$ l per reaction volume and annealing temperature was adjusted between 45 to 55°C. See **Appendix B** for details on optimum PCR mixture content and **Appendix C** for optimum thermal cycle profile.

### 3.2.3: GEL ELECTROPHORESIS

Success of the amplicon was analyzed using 1% agarose gel. To prepare 1% agarose gel, an appropriate amount of agarose powder was added with 1X TBE buffer (refer **Appendix D** for components of 10X TBE buffer). Specifically, 0.25g of agarose powder was added to measuring cylinder and 25ml of 1X TBE buffer was measured and poured into the measuring cylinder. This proportion of mixture will give a final product of 1% agarose gel for further genetic analysis. The mixture was heated in the microwave oven (SHARP) for about 1 minute until the agarose powder was fully dissolved. In the mean time, a gel cast was assembled according to the number of samples needed to be loaded into the gel.

The dissolved agarose gel after heating was allowed to be cooled down and was poured gently into the gel cast assembled earlier. The gel was let to be solidified at room temperature for about 30 minutes. Next, the comb was removed and the solidified gel was transferred into the gel electrophoresis tank. The tank was then filled with 1X TBE buffer until the gel's surface was slightly covered by the buffer.

Before the gel electrophoresis started, 5 $\mu$ l of PCR product from each PCR tube was loaded into the separate wells of the gel. After loading all the samples, the tank was connected to the power supply (BioRad) at 80volt for about 35 minutes to allow the migration of DNA from negative pole to positive pole. Next, the gel was then transferred into the ethidium bromide (EtBr) container. The gel was soaked into the EtBr solution for about 30 minutes. Finally, the gel was viewed under the UV light to detect the presence of bands using Alpha Imager<sup>TM</sup> 2200 transilluminator.

### 3.2.4: GEL PURIFICATION

Gel purification was carried out using AxyPrep™ DNA Gel Extraction Kit according to the protocol provided by the company (AxyPrep™ DNA Gel Extraction Kit website : [http://cdn.shopify.com/s/files/1/0001/7488/files/05-AxyPrep\\_DNA\\_Gel\\_DYE\\_Ext\\_Kit\\_ver20090727.pdf?1262654303](http://cdn.shopify.com/s/files/1/0001/7488/files/05-AxyPrep_DNA_Gel_DYE_Ext_Kit_ver20090727.pdf?1262654303)). All the steps performed in gel purification are as follow.

The PCR product was run in the 1% agarose gel for 35 minutes at 80volt. After gel electrophoresis, the gel was soaked into the EtBr solution for about 30 minutes before transferred to the dark room. In the dark room, the gel was viewed under UV illumination and the selected bands with the correct size were excised from the agarose gel with a clean scalpel. The excised gels with the selected bands were sliced into small pieces with the scalpel and were transferred into 2ml microfuge tube.

Next, the 2ml microfuge tube with the sliced gel was then centrifuged at 12,000xg for about 30 seconds. After centrifugation, 300µl of Buffer DE-A was added to the gel. The microfuge tube containing the mixture of gel and Buffer DE-A was voltex for a few seconds. After voltexing, the tube was incubated in the water bath at 75°C for about 7 minutes with intermittent voltex every 2 minutes until the gel was completely dissolved.

After incubation, 150 $\mu$ l of Buffer DE-B was added into the tube. An Axyprep column was placed into a new 2ml microfuge tube after the addition of Buffer DE-B. The solubilized gel was then transferred into the Axyprep column and was centrifuged at 12,000xg for 1 minute. After centrifugation, the filtrate will be discarded from the microfuge tube and the Axyprep column was placed back into the microfuge tube.

In the subsequent step, 500 $\mu$ l of Buffer W1 was added into the Axyprep column and the microfuge tube was centrifuged at 12,000xg for 30 seconds. After that, the filtrate will be discarded from the microfuge tube and the Axyprep column was placed back into the microfuge tube.

Then, the microfuge tube was then added with 700 $\mu$ l of Buffer W2 (ethanol added) and was centrifuged at 12,000xg for 30 seconds. After centrifugation, the filtrate will be discarded from the microfuge tube and the Axyprep column was placed back into the microfuge tube. Next, another 700 $\mu$ l of Buffer W2 (ethanol added) was added as the previous step and the tube was centrifuged at 12,000xg for 1 minute. After that, the filtrate will be discarded from the microfuge tube and the Axyprep column was placed back into the microfuge tube. The microfuge tube was then centrifuged again at 12,000xg for 1 minute.

Then, the Axyprep column was transferred into a new 1.5 microfuge tube. Finally, 30 $\mu$ l of Eluent (preheated at 65°C) was added into the center of the membrane. The 1.5 microfuge tube was then allowed to stand at room temperature overnight. In the next day, the microfuge tube was centrifuged at 12,000xg for 2 minutes.

### 3.2.5: DNA SEQUENCING

The success of gel purification was determined by 1% agarose gel electrophoresis, with product visualized by staining with Ethidium Bromide. The clean product was subjected to sequencing reactions in both directions using the Big Dye version 3.1 chemistry following the manufacturer's instructions (Applied Biosystem website: [http://www.ibt.lt/sc/files/BDTv3.1\\_Protocol\\_04337035.pdf](http://www.ibt.lt/sc/files/BDTv3.1_Protocol_04337035.pdf)). After sequencing reactions, sequencing products were cleaned using an ethanol precipitation and analyzed on a ABI 3730 capillary sequencer (Applied Biosystems). Finally, these sequenced products were ready for further data analysis.

### 3.2.6: MOLECULAR DATA ANALYSIS

Prior to data analysis, all sequence reads were checked against chromatograph data using Chromas Software (Technelysium Pty Ltd) to ensure the high quality of the sequence and to remove ambiguous bases. Next, nucleotide sequences were trimmed to a length of 582bps using Mega Software version 4.1 (Tamura *et al.*, 2007). The final alignment began at position 6552 of the complete mitochondrial genome of *Oncorhynchus mykiss* (Genbank accession number: NC\_001717.1).

The obtained DNA sequences were aligned using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1997) to check for homology to other vertebrate COI sequences in order to ensure that the desired COI region had been obtained. During DNA barcoding, there are some cases where non-target fragments such as nuclear mtDNA pseudogenes (numt) may be amplified (Shafer *et al.*, 1999; Bensasson *et al.*, 2001). To ensure that our current data set does not involve these non-target fragments, all sequences were screened for the presence of stop codons and insertion-deletion mutations using the software Mega Version 4.1 (Tamura *et al.*, 2007). After sequence alignment, the software DnaSP v5 (Librado & Rozas, 2009) was used to summarise haplotype information. Within group mean genetic distances were calculated as within species, within genera and, where sampling was appropriate, within one family (Cyprinidae) using the Kimura 2 Parameter (K2P) distance model (Kimura, 1980) in the software Mega Version 4.1.

Phylogenetic trees were constructed to illustrate the patterns of divergence among species and within species using the obtained haplotype sequences. Only one individual was included in the phylogenetic analyses if multiples individuals possessed identical haplotypes. In the case to confirm accuracy of the divergence pattern, two species, *Amia calva* (genbank accession number: EU524435.1) and *Atractosteus tristoechus* (genbank accession number: FN545592.1) were included as outgroup taxa. There are two models of phylogenetic trees constructed in this study. First, a Neighbour-joining (NJ) tree (Saitou & Nei, 1987) was constructed in Mega Version 4.1 using the K2P distance model with 1,000 bootstrap replicates. Secondly, a Bayesian tree was estimated using MrBayes software (Huelsenbeck & Ronquist 2001). For this analysis parameters not part of the default settings were a 4by4 nuclear model, an nst of 6, and an invariable sites with gamma model. This Bayesian analyses test was run for 200,000 generations of simulations under a Markov Chain Monte Carlo (MCMC) model. The trees were sampled every 100 generations and the first 25% of samples were discarded as burn-in.

In the case of understanding the population evolutionary of *C. striata*, a haplotype network which connecting all the haplotypes by mutational step was performed using software TCS version 1.21 (Clement *et al.*, 2000). In order to measure the population distance, the Fst value which determines the population differentiation was then calculated using software Arlequin version 3.11 (Excoffier *et al.*, 2005). This population comparison was performed by computing pairwise differences with number of permutation of 100 and significant level at  $p < 0.05$ .

### 3.3: MORPHOMETRIC ANALYSIS

Samples were brought back from the study field to the laboratory in which 38 morphometric measurements were taken to the nearest centimeter according to the Truss Network Measurement (Bhassu & Rashid, 2009). See **Figure 3.2(a) & 3.2(b)** for all the linear measurements. A box-plot analysis was done using SPSS software version 16.0 (SPSS, Inc., Chicago IL) to exclude the extreme measurement which indicate outlier. A total of eight morphometric measurements were excluded due to the presence of outlier measurements. Prior to multivariate discriminant analysis, transformation of size-independent shape variables was carried out using SPSS software version 16.0 with the formula of:  $M_{adj} = M(L_s/L_o)^b$  (Elliot *et al.*, 1995), where  $M_{adj}$  is the transformed morphometric measurement to be used in multivariate analysis,  $M$  is the original morphometric measurement,  $L_s$  is the overall mean of standard length obtained from all samples for each variable,  $L_o$  is the standard length of the fish, and  $b$  is the slope of regression of  $\log M$  over  $\log L_o$ . Parameter  $b$  is calculated to check the effect of size on the measurement and transformed  $M_{adj}$  variables were used in the following multivariate discriminant analysis when the transformed variable did not show significant correlation with standard length.

All the 30 successfully transformed measurements were included in subsequent multivariate discriminant analysis using software STATISTICA version 8 (StataCorp. 2003. *Stata Statistical Software: Release 8*. College Station, TX: StataCorp LP). This analysis highlighted the two traits which had greatest contribution in discriminating populations and the pattern of variation will be illustrated in two-dimensional graph. Specifically, there are unobserved pattern of variation lying behind each observed measure



and multivariate discriminant analysis aim to group those underlying variation which are correlated to each other under the same root. The pattern of variation underlined by each root can be seen in two-dimensional graph in which those populations located at positive region of each root acquired unique extreme measurements of the two highly correlated traits which could discriminated them from other population located at negative region of the same root. These two discriminant traits will be referred to the “List of Standardized Coefficients of Canonical Variables” which will highlight the two highly correlated traits that had greatest contribution of the variation for each root. These discriminant traits are those traits with the two top coefficient values in the mentioned list ignorance to the “positive” and “negative” sign of the value. The comparative small or comparative large measurement of the trait in relation to others will then be referred to the “negative” and “positive” sign of the coefficient value respectively (DeCoster, 1998).

The first stage of multivariate discriminant analysis was obtained to illustrate the morphological population divergence pattern of *C. striata* in Peninsular Malaysia and so was included all the *C. striata* populations surveyed in Peninsular Malaysia. In order to achieve balance sample size in current data set for further analysis, locations in Peninsular Malaysia with overlapping centroid mean were considered as single group and the subset of each group from Peninsular Malaysia together with the group of Sarawak were included in second stage of multivariate discriminant analysis to classsify the morphological divergence pattern of *C. striata* throughout Malaysia.

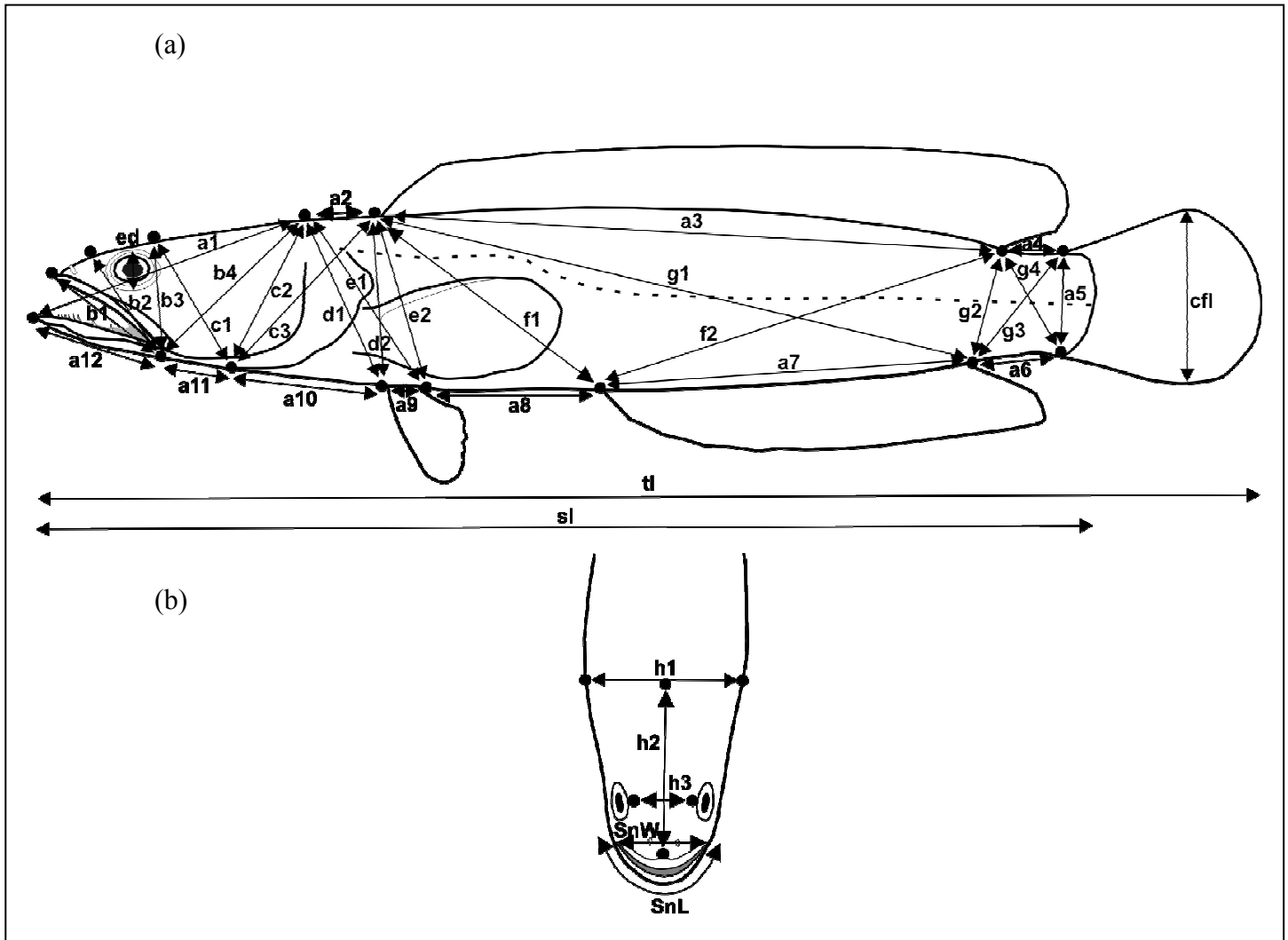


Figure 3.2 (a) & (b): Schematic diagram of 38 morphometric measurement of *Channa striata* (Total length (tl), standard length (sl), caudal fin length (cfl), eyes diameter (ed), snout width (SnW), snout length (SnL), measurement of each body part (a1-a12, b1-b4, c1-c3, d1-d2, e1-e2, f1-f2, g1-g4, h1-h3)) based on the Truss Network Measurement (Bhassu & Rashid, 2009).