



Faculty of Resource Science and Technology

**Sequencing and Characterization of Complete Mitogenome DNA for
Rasbora spp. with Phylogenetic Consideration**

Cynthia Keranti Anak Kamar

**Master of Science
2019**

Sequencing and Characterization of Complete Mitogenome DNA for *Rasbora*
spp. with Phylogenetic Consideration

Cynthia Keranti Anak Kamar

A thesis submitted

In fulfilment of the requirements for the degree of Master of Science

(Molecular Biology)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2019

DECLARATION

I, Cynthia Keranti, declare that I am the sole author of this dissertation. I have fully acknowledged the ideas, methods, sources or works of other people either it is published or unpublished that related to my study in accordance with the standard referencing practices at Universiti Malaysia Sarawak. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Name: Cynthia Keranti Anak Kamar

Student No.: 15020870

Faculty of Resource Science and Technology

UNIVERSITI MALAYSIA SARAWAK

July 2019

ACKNOWLEDGEMENT

First and foremost, I praise and thank God, the almighty for His blessing that I was able to complete my thesis and conduct my research successfully. I thank Him for given me strength, wisdom, and good health throughout my research.

I want to express my deepest gratitude to my supervisor Dr Chung Hung Hui for this opportunity to do research and providing me all the facility that I need for my research work. I thank him for his invaluable guidance and encouragement during the whole period of research work that I was able to stay focused and motivated. I am truly grateful for what he has taught me. It was a great privileged working under his guidance.

A big thanks to Dr Lee Kui Soon for his kindness to allow me use his laboratory facility throughout my research. A giant thanks to the University of Hong Kong especially Dr Tommy Lam, Dr Chong Yee Ling, and Phd student, Yunshi Lao for assisting me to run the sequencing with their high-throughput sequencing technology. Next, I would like to express my deepest thanks to my lab colleagues, Jill, Aimi, Leonard, and Shamil for their support and helping one another throughout this research work. Special thanks to my close friends, Danny, Jill, and Diana for their encouragement, companionship, and assistances during sampling and finishing my thesis. They help me to establish and nurture a positive thinking in me.

Last but not least, I would like to express my greatest thanks to my beloved family for their love, prayer, care, and support so that I was able to stay firm and do my best to complete my study. I am blessed to have them in my life.

ABSTRACT

Rasbora are freshwater fishes that belongs to Cyprinidae family. *Rasbora* species share similar morphology and are closely related which makes them difficult to be differentiated between species. In this study, the morphometry analysis and mitochondrial DNA were used to distinguish the species and able to resolve the phylogenetic relationship between these species. A total of 58 *rasbora* specimens from four species (*Rasbora sumatrana* (20), *Rasbora hobelmani* (11), *Rasbora tornieri* (20), and *Rasbora borapetensis* (7)) were individually identified using morphometric analysis. The mitochondrial DNA were amplified by using the long polymerase chain reaction (PCR). Four sets of primer were designed for partial fragment of mitochondrial genome. The newly paired primer were made from these 4 sets of primer for complete fragments of mitochondrial genome. The complete mitochondrial DNA of 4 *Rasbora* species consists of 22 tRNA genes, 13 protein coding genes, and 2 rRNA genes with sequence full length of 16,539 bp (*Rasbora sumatrana*), 16,541 bp (*Rasbora hobelmani*), 16,573 bp (*Rasbora tornieri*), and 16,520 bp (*Rasbora borapetensis*) respectively. The phylogenetic position of these 4 *Rasbora* species was evaluated by using concatenated 13 protein coding genes of complete mitochondrial sequence. The morphometry and mitochondrial DNA study has successfully correlated these 4 *Rasbora* species by using discriminant function analysis and sequencing respectively. In conclusions, the molecular work produce more reliable results than morphometric study in identifying the species as the resolution of the species phylogeny can be observed.

Keywords: *Rasbora*, Mitochondrial DNA, morphometric

Urutan dan Ciri-Ciri Genom Mitokondria yang Lengkap untuk *Rasbora* spp. dengan Pertimbangan Filogenetik

ABSTRAK

*Spesies Rasbora adalah ikan air tawar yang tergolong dalam keluarga Cyprinidae. Spesies Rasbora berkongsi morfologi yang serupa dan berkait rapat. Ini menjadikan mereka susah untuk dibezakan antara spesies. Analisis morfometri dan DNA mitokondria digunakan untuk membezakan spesies dan dapat menyelesaikan hubungan filogenetik antara spesies ini. Sebanyak 58 ikan rasbora (*Rasbora sumatrana* (20), *Rasbora hobelmani* (11), *Rasbora tornieri* (20), dan *Rasbora borapetensis* (7)) telah digunakan dalam analisis morfometrik. DNA mitokondria pula diaplikasikan dengan menggunakan tindak balas rantai polimer (PCR) yang panjang. Empat set primer telah direka untuk sebahagian fragmen DNA mitokondria. Pasangan primer yang baru diambil daripada 4 set primer tersebut untuk fragmen DNA mitokondria yang lengkap. DNA mitokondria yang lengkap daripada 4 spesies *Rasbora* ini terdiri daripada 22 gen tRNA, 13 gen kod protein, dan 2 gen rRNA dengan urutan penuh masing-masing 16,539 bp (*Rasbora sumatrana*), 16,541 bp (*Rasbora hobelmani*), 16,573 bp (*Rasbora tornieri*), dan 16,520 bp (*Rasbora borapetensis*). Kedudukan 4 spesies *Rasbora* ini di filogenetik dinilai dengan menggunakan 13 gen kod protein yang bersambung dari urutan mitokondria yang lengkap. Kajian morfometri dan DNA mitokondria telah berjaya mengenalpasti 4 spesies *Rasbora* ini dengan menggunakan analisis diskriminasi fungsi dan penjujukan masing-masing. Secara kesimpulannya, teknik molekul menghasilkan lebih banyak hasil yang boleh dipercayai daripada kajian morfometrik dalam mengenal pasti spesies di mana resolusi spesies dapat diperhatikan.*

Kata kunci: *Rasbora, DNA mitokondria, morfometri*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Background Research	1
1.2 Problem Statement	3
1.3 Significance of Study	3
1.4 Objective of Study	4
CHAPTER 2: LITERATURE REVIEW	5
2.1 Family Cyprinidae	5
2.2 Biology of Genus <i>Rasbora</i>	5
2.2.1 <i>Rasbora sumatrana</i>	9
2.2.2 <i>Rasbora hobelmani</i>	10
2.2.3 <i>Rasbora tornieri</i>	11
2.2.4 <i>Rasbora borapetensis</i>	12
2.3 Taxonomy Classification of <i>Rasbora</i>	13
2.4 Morphometry Analysis	15

2.5 Mitochondrial DNA (mtDNA) Sequence	16
2.6 Phylogenetic Analysis	19
CHAPTER 3: METHODOLOGY	22
3.1 Site Descriptions and Sample Collection	22
3.2 Sample Assortment and Maintenance	25
3.3 Morphometry Procedure	25
3.3.1 Morphometric Measurements	25
3.3.2 Meristic Counts	29
3.3.3 Morphometry Analysis	29
3.4 Molecular Work	32
3.4.1 Mitochondrial DNA (mtDNA) Extraction	32
3.4.2 Primer Design and Synthesis	33
3.4.3 Long Polymerase Chain Reaction	34
3.4.4 Purification of PCR Amplified Product	35
3.4.5 DNA sequencing of PCR Amplified Product	36
3.4.6 Sequence Analysis	36
3.4.7 Phylogenetic Analysis	37
CHAPTER 4: RESULTS	38
4.1 Sample colouration	38
4.2 Morphometric Analysis	41
4.3 Meristic Analysis	50
4.4 Complete Fragment of Mitochondrial Genome Analysis	59
4.4.1 Annealing Temperature Optimization of Pair Primers	59
4.4.2 Sequencing Analysis	62

4.4.3 Phylogenetic Analysis	85
CHAPTER 5: DISCUSSION	87
5.1 Morphometric Analysis	87
5.2 Complete Mitochondrial Genome Analysis	89
5.3 Phylogenetic Analysis	92
CHAPTER 6: CONCLUSION AND RECOMMENDATIONS	94
6.1 Conclusion	94
6.2 Recommendations	95
REFERENCES	96
APPENDICES	104

LIST OF TABLES

	Page
Table 2.1 Classification level of <i>Rasbora</i>	14
Table 3.1 Sampling site descriptions	22
Table 3.2 Morphometric characters used for measurement	26
Table 3.3 Meristic characters used for counting	29
Table 3.4 Components of PCR	34
Table 4.1 Descriptive statistics for morphometric characters of <i>Rasbora</i> species	42
Table 4.2 ANOVA for size adjusted morphometric measurements among 4 <i>Rasbora</i> species	44
Table 4.3 Wilk's lambda values of 24 morphometric characters	45
Table 4.4 Eigenvalues, percentage of variance, cumulative percentage, canonical correlation, and standardized coefficients of morphometric characters	46
Table 4.5 Structure matrix of ordered correlation morphometric characters by its absolute size within function	47
Table 4.6 The classification of <i>Rasbora</i> species by DFA on morphometric measurements with their predicted group membership	49
Table 4.7 Descriptive statistics for meristic traits of <i>Rasbora</i> species	51
Table 4.8 Kruskal-Wallis test (non-parametric test) for comparison of meristic counts among the <i>Rasbora</i> species	52
Table 4.9 Values of Wilk's lambda and its significance for 11 meristic data	53
Table 4.10 The observed pooled within group correlation matrix among 11 traits of meristic data	54

Table 4.11	Eigenvalues, percentage of variance, cumulative percentage, canonical correlation, and standardized coefficients of meristic data	55
Table 4.12	Structure matrix of ordered correlation variables by its absolute size within function	56
Table 4.13	The classification of <i>Rasbora</i> species by DFA on meristic counting with their predicted group membership	58
Table 4.14	Primer sets for partial and complete fragments	59
Table 4.15	Forward and reverse primer characteristics for fragments of each <i>Rasbora</i> species	63
Table 4.16	Nucleotide BLAST descriptions of <i>Rasbora</i> species complete fragments	64
Table 4.17	Characteristics of complete mitochondrial genome of <i>Rasbora sumatrana</i>	67
Table 4.18	Characteristics of complete mitochondrial genome of <i>Rasbora hobelmani</i>	70
Table 4.19	Characteristics of complete mitochondrial genome of <i>Rasbora tornieri</i>	73
Table 4.20	Characteristics of complete mitochondrial genome of <i>Rasbora borapetensis</i>	76
Table 4.21	Whole nucleotide composition (%) of 4 <i>Rasbora</i> species	82
Table 4.22	Nucleotide composition of gene arrangements in percentage by MEGA	83

LIST OF FIGURES

	Page
Figure 2.1 ‘Seluang’ fish that can be found in local market Sibu, Sarawak with price of RM5 per plate	6
Figure 2.2 <i>Rasbora sumatrana</i> (left: presence of black tips, right: absence of black tips on the caudal lobes)	9
Figure 2.3 <i>Rasbora hobelmani</i>	10
Figure 2.4 <i>Rasbora tornieri</i>	11
Figure 2.5 <i>Rasbora borapetensis</i>	12
Figure 2.6 The circular gene map of <i>Araipaima gigas</i> mtDNA	19
Figure 3.1 Sampling site at Kanowit river in Sibu	23
Figure 3.2 Sampling site at drainage along Kuching Outer Ring Road at Kota Samarahan	24
Figure 3.3 Sampling site at Bajo Bergantong river, Lundu	24
Figure 3.4 Schematic illustration of morphometric characters used	28
Figure 3.5 Schematic illustration of meristic characters used	30
Figure 4.1 Samples before (left) and after (right) preserved in 95% of ethanol. a) <i>Rasbora sumatrana</i> ; b) <i>Rasbora hobelmani</i> ; c) <i>Rasbora tornieri</i> ; d) <i>Rasbora borapetensis</i>	40
Figure 4.2 Canonical discriminant function that classify <i>Rasbora</i> samples into its respective species based on morphometric analysis	48
Figure 4.3 Canonical discriminant function graph that separate the specimen by species based on meristic analysis	57

Figure 4.4	Gel electrophoresis results from gradient PCR of small fragment from both complete fragments primer. Lane L represents 1 kb ladder, Lane 1, 2, and 3 represents PCR amplicon from primer 1, Lane 4, 5, and 6 represents PCR amplicon from primer set 2 at the same temperature of 48.6°C, and Lane 7 and 8 represents negative control from both primer sets.	60
Figure 4.5	Gel electrophoresis results from gradient PCR of small fragment from set primer 1. Lane L represents ladder of 1 kb. Lane 3 represents negative control. Lane 1 and lane 2 represents PCR amplicon from <i>Rasbora borapetensis</i> sample.	61
Figure 4.6	Gel electrophoresis results from gradient PCR for large fragment. Lane L represents 1kb ladder. Negative control was represents by lane 3, 4, 9, and 10. a) Lane 1 and lane 2 represents <i>Rasbora sumatrana</i> at temperature of 47.6 °C. b) Lane 5 and lane 6 represents <i>Rasbora hobelmani</i> at temperature of 47.5 °C. c) Lane 7 and 8 represents <i>Rasbora tornieri</i> at temperature of 50.7 °C. d) Lane 11 represents <i>Rasbora borapetensis</i> at temperature of 50.0 °C.	62
Figure 4.7	An atypical anticodon stem-loop of secondary structure in <i>Rasbora sumatrana</i> tRNA genes by tRNA scan S.E server. a) tRNA ^{leu} (TAA); b) tRNA ^{ser} (TGA); c) tRNA ^{ser} (GCT); and d) tRNA ^{leu} (TAG)	66
Figure 4.8	Circular mtDNA genome of <i>Rasbora sumatrana</i>	78
Figure 4.9	Circular mtDNA genome of <i>Rasbora hobelmani</i>	79
Figure 4.10	Circular mtDNA genome of <i>Rasbora tornieri</i>	80
Figure 4.11	Circular mtDNA genome of <i>Rasbora borapetensis</i>	81

Figure 4.12 Phylogenetic tree of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora* 86
tornieri, and *Rasbora borapetensis* based on 13 protein coding genes using
Maximum Likelihood (ML) analysis. The numbers on the branches are
bootstrap values of ML.

LIST OF ABBREVIATIONS

A+T contents	Adenine + Thymine contents
ANOVA	Analysis of variance
ATP	Adenosine Triphosphate
BB	Binding buffer
BLAST	Basic local alignment search tool
blastn	Nucleotide BLAST
bp	Base pair(s)
COI	Cytochrome c oxidase subunit I
D-loop	Displacement loop
DNA	Deoxyribonucleic acid
DNTPs	Deoxyribonucleotide triphosphates
EB	Elution buffer
GC contents	Guanine + Cytosine contents
GPS	Global positioning system
MEGA	Molecular evolutionary genetics analysis
ML	Maximum Likelihood
mm	Millimetre
mtDNA	Mitochondrial DNA
NCBI	National center for biotechnology information
NGS	Next generation sequencing
PCA	Principal component analysis
PCR	Polymerase chain reaction
rRNA	Ribosomal RNA

SDS	Sodium dodecyl sulfate
spp	Species-species
SPSS	Statistical package for the social science
STD	Standard deviation
TAE	Tris-Acetate EDTA
<i>Taq</i>	<i>Thermus aquaticus</i>
tRNA	Transfer RNA
UV	Ultraviolet
WB	Wash buffer

CHAPTER 1

INTRODUCTION

1.1 Background research

Malaysia is one of the countries of Sundaland that have high diversity of freshwater fishes especially in Sabah (north Borneo) and Sarawak (northwest to west Borneo). This including described and undescribed fish species of order Cypriniformes. There are many areas in Sabah and Sarawak that has not been explored and unable to access. Sarawak is covered by a rich diversity of tropical rainforests where the fish diversity in Sarawak has been recorded as 254 species (Sulaiman and Mayden, 2012).

Cyprinidae family is the most species rich of freshwater fishes and belong to Cypriniformes order. This family consist of 11 subfamilies which include Acheilognathinae, Barbinae, Cultrinae, Cyprininae, Danioninae, Gobioninae, Labeoninae, Leuciscinae, Squaliobarbinae, Tincinae, and Xenocypridinae. *Rasbora* is a monophyly of rasborins group that belong to Danioninae subfamily (Liao *et al.*, 2010). The characteristics of Cyprinidae family are weak jaws, absence of adipose fin, some have barbells around the mouth, and grind their food with pharyngeal teeth. The breeding method of this family are mostly egg-scatters (Sim, 2002).

This small to moderate size of *Rasbora* can be found within Asia which include the Indian subcontinent, southern China, and Southeast Asia. Genus *Rasbora* contributes 77 valid species in the Cyprinidae family. This genus was first revised by Brittan in 1954. From his revision, three subgenera has been recognized which are *Rasbora*, *Rasboroides*, and *Megarasbora*. New genera of *Rasbora* has been generated by other authors beside Brittan which include the *Boraras*, *Brevibora*, *Horadandia*, *Kottelatia*, *Rasboroides*, *Trigonopoma*, and *Trigonostigma*. Then, *Rasbora* has been classified by Brittan into 8 species groups (initially

called as species complexes) which are *Lateristriata*, *Sumatrana-elegans*, *Caudimaculata*, *Trifasciata*, *Argyrotaenia*, *Daniconius*, *Einhovenii*, and *Pauciperforata* groups. However, many researcher are still referring to Brittan (1954) concept of this genus (Lumbantobing, 2014).

There are several studies conducted on *Rasbora* fish including morphometry analysis and mitochondrial DNA. Maurice Kottelat in collaboration with other researcher like Heok Hee Ng (2013), Heok Hui Tan (2011 and 2012), and Renny Kurnia Hadiaty (2009) has studied the identification of *Rasbora* by conducting morphometry analysis. For instance, *Rasbora dusonensis*, *Rasbora tornieri*, *Rasbora cryptica*, *Rasbora beauforti*, *Rasbora atranus*, and *Rasbora lacrimula*. Kottelat also has discovered a new species of *Rasbora* from Borneo like *Rasbora rheophila* (Kottelat, 2012) and *Rasbora dies* (Kottelat, 2007). A new *Rasbora* species from Peninsular Malaysia also has been discovered by Kottelat (2005) which is *Rasbora notura*. Daniel Lumbantobing also has described new *Rasbora* species from *Rasbora trifasciata* group (Lumbantobing, 2010) and *Rasbora sumatrana* group (Lumbantobing, 2014) by analysing morphological variation.

The study of complete mitochondrial genome on this small cyprinid has been conducted by several researchers like in *Rasbora trilineata* (Ho *et al.*, 2014), *Rasbora borapetensis* (Zhang *et al.*, 2014), *Rasbora aprotaenia* (Kusuma and Kumazawa, 2015), *Rasbora lateristriata* (Kusuma and Kumazawa, 2015), and *Rasbora argyrotaenia* (Kusuma *et al.*, 2017). The overall nucleotide composition and an overview of circle genome with genes arrangement were obtained from this molecular studies. These *Rasbora* morphological and molecular studies may provide an additional reference to construct the phylogenetic tree and resolve the taxonomy uncertainty of this genus.

1.2 Problem statement

Even though many taxonomist refer to Brittan revision of morphological characteristics in order to distinguish species among *Rasbora* genus, this information is not precise enough as cryptic species diversity issues often presence. The cryptic diversity are defined as two or more distinct species that were categorized as one due to their similar morphology (Trontelj and Fišer, 2009). In this case, many different species of *Rasbora* share similar morphological traits such as colour of caudal fin and presence of black stripe. Besides, *Rasbora* species are difficult to distinguish from each other due to its small body size and lack of obvious morphology. This may lead to misidentification of this genus.

Some species were wrongly listed where different names were used on similar species while error from previous studies were also repeated in later studies (Sim, 2002). Besides, inaccessibility of certain species also may lead to misidentification. This claims by Siebert (1997) when *Rasbora paucisqualis* has been confused with *Rasbora bankanensis* due to similar scales count and colour pattern. Furthermore, usage of single gene from mitochondrial DNA in species identification including Cytochrome *b*, Cytochrome oxidase and D-loop were found to provide little clue in resolving cryptic issue. Recent advancement in sequencing tools have provided the platform for a thorough sequencing of overall mitochondrial genome and it is hence hypothesized to provide more reliable molecular approach in phylogenetic analysis.

1.3 Significance of study

Species identification and molecular classification are important to this genus as there are probably many new species of *Rasbora* remain to be discovered and described especially in Sarawak. Malaysia has been identified as one of the biodiversity hotspots worldwide. Biodiversity hotspots are described as uneven biodiversity distribution and facing habitat loss in some areas (Hubert *et al.*, 2015).

The discovery of new species is very important so that the new generation are able to study, identify, and distinguish the *Rasbora* species. This can be an initiative to draw more attention and protection to this species as some may be extinct without being able to be recognized (Kottelat, 1984). Currently there about 120 species of *Rasbora* that has been discovered and described (Kottelat and Tan, 2011).

Due to cryptic diversity in *Rasbora* species, the phylogenetic relationships of *Rasbora* species still remain incomplete and scatter in the scientific literature. In order to resolve the ambiguity in *Rasbora* taxonomy knowledge, this research aim to analyse the *Rasbora* morphometry data, characterize the mitochondrial DNA fragment, and understand the phylogenetic relationship of four species of *Rasbora* from Sarawak, Malaysia which are *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis*.

1.4 Objective of study

Due to the limited access to most protected rehabilitation area in the state (limitation in permit - NCCD.94047(Jld13)-178), hence, sampling efforts were focussed mainly in the Southwest region of Sarawak, ranging from Machan, Kanowit to Lundu. Four selected species of *Rasbora* will be focussed in this study including *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* with objectives stated as below.

- i. To conduct the morphometry analysis of genus *Rasbora* from Sarawak.
- ii. To analyse the overall mitochondrial genome of selected *Rasbora* spp.
- iii. To establish the phylogeny of the species based on the mitochondrial DNA.

CHAPTER 2

LITERATURE REVIEW

2.1 Family Cyprinidae

The order Cypriniformes is the most diverse group of freshwater fishes worldwide and has its greatest diversity in Southeast Asia. The Cypriniformes consists of 6 families, 321 genera, and 3,268 species. Cyprinidae family are one of the 6 families that belong to Cypriniformes order. At least 220 genera and estimated 2,420 species are within this family (Sulaiman and Mayden, 2012). The main characteristics of Cyprinidae family include; one or two pairs of barbells are either presence or absence, no adipose dorsal fin, lateral line is either complete or incomplete, body with scales, one to three rows of pharyngeal teeth, no subocular spines, and well develop of opecular bones. The subfamilies within Cyprinidae family are namely Leuciscinae, Oreininae, Garrinae, Danioninae, and Cyprininae.

The Danioninae subfamily of Cyprinidae is a large group of minnows in South and Southeast Asian. Several genera from this subfamily are very popular in the ornamental fish trade. For example, *Boraras*, *Danio*, *Rasbora*, *Trigonostigma*, and *Sundadanio*. Blackwaters associated with peat swamp forests in Sundaland, are home to many species that are involved in the international ornamental fish trade. Aside from *Boraras* spp., *Betta* spp., and *Parosphronemus* spp., *Rasbora* spp. are no exception (Hubert *et al.*, 2015).

2.2 Biology of genus *Rasbora*

In Malaysia, *Rasbora* fish are commonly known as ‘Seluang’ fish. The terms are different depending on the locality. For example, ‘Buntong’ in Keningau Dusun, ‘Londoi’ in Dusun, and ‘Buntod’ in Orang Sungei. However, the local people apply the local name for this genus without distinguish the species (Inger and Chin, 1990). This fish are famous among local

people because of its affordability and tasteful flesh (refer to Figure 2.1). Usually this fish are sold in local market as food source and fish bait.



Figure 2.1: ‘Seluang’ fish can be found in local market Sibul, Sarawak with price of RM5 per plate.

Genus *Rasbora* are relatively small to moderate size of fishes with range of 5 cm to 20 cm in standard length (Kusuma *et al.*, 2016). Their habitats may vary from roadside drainage, small lakes, paddy fields, swamps, lower reaches of mountain streams, to black water. Generally, larger size of species are found in larger volumes of water like river while the smaller size of species in small volumes of water like pond and ditches. The *Rasbora* are schooling fishes and usually observed swimming in the upper layers of flowing water (Suvama and Rita, 2014). By gathering in groups, it will be harder for the predators to pick them off. They use the bodies of others for cover. This genus feeds on worms, crustaceans, and insects.

Several *Rasbora* are valued as an aquarium fishes because of their unique colour, small in size, and extremely active such as the *Rasbora trilineata* and *Rasbora heteromorpha* which also known as Harlequin Rasbora. *Rasbora heteromorpha* are found abundantly in Tasek Bera,

Pahang, Malaysia and at least RM 180,000 per annum generated from this species trading (Sim, 2002). Their size, colour pattern and shape among *Rasbora* species varies due to water condition, ecology, and food availability in their respective habitats.

Most freshwater fishes are seasonal breeders. However, according to Roberts (1989), small group of species from Borneo that belong to a diverse families such as Cyprinidae, Chacidae, Hemiramphidae, Syngnathidae, Mastacembelidae, Channidae, Belontiidae, Luciocephalidae, and Tetraodontidae have non seasonal or continuous reproduction. Several species from these families has been observed to reproduce actively regardless the time of the year. Whenever the species of small fishes like *Rasbora axelrodi*, *Rasbora bankanensis*, and *Betta dimidiata* was moderately abundant and extensive sampling were made, the very small juveniles were also found together.

The spawning method of this fish usually takes place under plant leaves as they attach their eggs to the underside of broad leaves or other suitable plants. Then, the fertilized eggs are left to develop and hatch on their own. *Rasbora* larvae are almost transparent making them hard to see and reduce the risk of being eaten. In some *rasbora* such as *Rasbora sumatrana*, the mature male are recognize by presence of whitish tubercles on the head meanwhile the mature female with enlarged ova (Inger and Chin, 1990). Lumbantobing (2014) also stated that male *Rasbora sumatrana* has 1 to 2 rows of antrorse tubercles on dorsoproximal side of pectoral fin rays and female body is deeper than male.

This genus are widely distributed around India subcontinents, Southern Eurasia including Indonesia, Malaysia, Brunei Darussalam, Philippines, Cambodia, and Thailand. At least 16 species of genus *Rasbora* in Borneo are hypothesized to be endemic to Borneo. This includes 14 *Rasbora* species that inhabit the Rajang River, Sarawak and Danau Sentarum National Park, Western Kalimantan. The *Rasbora* species that are endemic to Borneo are *Rasbora axelrodi*, *Rasbora borneensis*, *Rasbora ennealepis*, *Rasbora hubbsi*, *Rasbora*

johanae, *Rasbora kalbarensis*, *Rasbora kottelati*, *Rasbora laticlavia*, *Rasbora rutteni*, *Rasbora sarawakensis*, *Rasbora semilineata*, *Rasbora trifasciata*, *Rasbora tuberculata*, and *Rasbora volzii* (Sulaiman and Mayden, 2012).

Brittan (1954) has described the characteristics of *rasbora* including absence of barbells, elongate and slender body, large scale, upper lip is notched medially, and small or moderate fins. The belly are lighter in colour. Dark midlateral stripe extends from snout or opercle to the caudal fin base. There are light metallic stripe along the dark midlateral stripe on its dorsal margin. A dark streak is also present on the anal base and form a reticulate or chain mail pattern from the ventral view.

Rasbora has been categorized by Brittan (1954) into 8 species groups (species complexes) which are *lateristriata*, *sumatrana-elegans*, *caudimaculata*, *trifasciata*, *argyrotaenia*, *daniconius*, *einthovenii*, and *pauciperforata* (Lumbantobing, 2014). *Rasbora lateristriata* and *Rasbora baliensis* belong to *lateristriata* group but then this group was placed within *sumatrana* group based on the pigmentation of midlateral stripe, supra-anal, and basicaudal blotch (Kusuma *et al.*, 2016). The members of *sumatrana-elegans* group are *Rasbora sumatrana*, *Rasbora elegans*, *Rasbora atranus*, *Rasbora volzii*, *Rasbora hosii*, *Rasbora atridorsalis*, *rasbora dorsinotata*, *Rasbora cryptica*, *Rasbora paviana*, *Rasbora hobelmani*, *Rasbora notura*, *Rasbora vulgaris*, *Rasbora nematotaenia*, *Rasbora spilotaenia*, *Rasbora aprotaenia*, and *Rasbora bunguranensis* (Kottelat and Tan, 2011). Now, *sumatrana-elegans* group is refer as *sumatrana* group. *Rasbora arundinata*, *Rasbora haru*, *Rasbora maninjau*, and *Rasbora bindumatoga* are described as new members of *sumatrana* group by Lumbantobing (2014). The members of *caudimaculata* groups are *Rasbora caudimaculata*, *Rasbora trilineata*, and *Rasbora dorsiocellata* (Brittain, 1954). The members of *trifasciata* groups as claimed by Lumbantobing (2010) are *Rasbora amplistriga*, *Rasbora bankanensis*, *Rasbora beauforti*, *Rasbora dies*, *Rasbora ennealepis*, *Rasbora hubbsi*, *Rasbora johannae*,

Rasbora meinkenii, *Rasbora paucisqualis*, *Rasbora reticulata*, *Rasbora rutteni*, *Rasbora sarawakensis*, *Rasbora taytayensis*, *Rasbora tobana*, *Rasbora trifasciata*, *Rasbora tuberculata*, and *Rasbora vulcanus*. Based on Ng and Kottelat (2013), *argyrotaenia* group members are *Rasbora argyrotaenia*, *Rasbora dusonensis*, *Rasbora tornieri*, *Rasbora myersi*, *Rasbora aurotaenia*, *Rasbora borneensis*, *Rasbora everetti*, *Rasbora laticlavia*, *Rasbora leptosoma*, *Rasbora philippina*, *Rasbora rheophila*, *Rasbora septentrionalis*, and *Rasbora vaillantii*. *Daniconius* group members are including *Rasbora daniconius*, *Rasbora caverii*, *Rasbora kobonensis*, *Rasbora labiosa*, *Rasbora ornata*, and *Rasbora wilpita* (Liao et al., 2010). Members of *einthovenii* group are *Rasbora einthovenii*, *Rasbora cephalotaenia*, *Rasbora tubbi*, *Rasbora kottelati*, and *Rasbora jacobsoni* (Liao et al., 2010). Last but not least, *pauciperforata* group members are *Rasbora pauciperforata*, *Rasbora taeniata*, *Rasbora borapetensis*, *Rasbora kalochroma*, *Rasbora heteromorpha*, and *Rasbora maculata* (Brittain, 1954).

There are four species of *Rasbora* from Sarawak, Malaysia studied in this research which are *Rasbora sumatrana* (belong to *sumatrana-elegans* group), *Rasbora hobelmani* (belong to *sumatrana-elegans* group), *Rasbora tornieri* (belong to *argyrotaenia* group), and *Rasbora borapetensis* (belong to *pauciperforata* group).

2.2.1 *Rasbora sumatrana* (Bleeker, 1852)

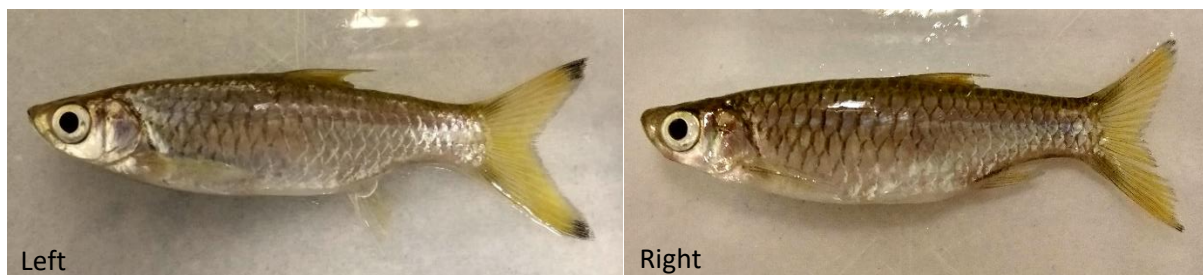


Figure 2.2: *Rasbora sumatrana* (left: presence of black tips, right: absence of black tips on the caudal lobes)

The *Rasbora sumatrana* is a medium to large size of *rasbora*. The mature male of *Rasbora sumatrana* are recognized by the small whitish tubercles on the head while the mature female with the enlarged ova. This *rasbora* usually can be found in a small clear stream with gravel at the bottom and a muddy river. They feed on terrestrial insects that fall into the streams (Inger and Chin, 1990).

The morphology of this *rasbora* including a dark midlateral stripe that extends from before dorsal to caudal peduncle with a roundish precaudal spot. The body colour is silvery yellowish. In living *rasbora*, a dull gold line over the dark lateral stripe are presence under reflection of light (Kottelat, 1984). However, this dull gold line are absence in the preserved sample. The reticulate pattern on the anal base are presence. The lobes of caudal fin of this *rasbora* are tipped with black but other fins are hyaline. There are variation of colour pattern that can be observed in this *rasbora* species. In some samples, the dark lateral stripe is presence but the precaudal spot is absence while in some this spot is presence. Sometimes, the stripe is absence or both stripe and spot are absence. Along with this stripe and spot, the black tips of caudal lobes (refer to Figure 2.2) may either be presence or absence in some samples (Brittan, 1954).

2.2.2 *Rasbora hobelmani* (Kottelat, 1984)



Figure 2.3: *Rasbora hobelmani*

Rasbora hobelmani are known as Kottelat *rasbora*. This species was first discovered by Maurice Kottelat in 1984 from Northern Thailand. The black midlateral stripe that runs from opercle to the vertical of dorsal fin are wider and become thinner towards the caudal peduncle. The precaudal spot are presence in this species. It has a deep head with a dark spot and thin caudal peduncle. The body colour is yellowish brown and greyish on the back and head. A black streak are presence on the anal base extends to the ventral edge of the caudal peduncle. All fins are hyaline. The lateral line is complete with scales range of 25 to 27 and additional of 1 scale on the caudal fin base (Kottelat, 1984).

This species are closely related to the *Rasbora sumatrana* as they share similar body shape and colour pattern. Both *Rasbora sumatrana* and *Rasbora hobelmani* are belong to *Rasbora sumatrana-elegans* group (Kottelat and Tan, 2011). However, the black tips on the caudal fins lobes (refer to Figure 2.3) is absence in *Rasbora hobelmani* (Rainboth and Kottelat, 1987).

2.2.3 *Rasbora tornieri* (Ahl, 1922)



Figure 2.4: *Rasbora tornieri*

Rasbora tornieri or yellow finned *rasbora* was first discovered by Ahl in 1922. Based on Ng and Kottelat (2013) article, *Rasbora tornieri* is identify by pre-dorsal scales of 12 to 14, bright coloured caudal fin, 14 circumpeduncular scales, and complete lateral line with 27 to 30 scales and additional of 2 scales on the caudal base. The caudal fin is deeply forked with pointed asymmetrical lobes where the lower lobe is longer than upper lobe (refer to Figure 2.4). This *Rasbora* species is known from Mekong River drainage in Cambodia, Perak River drainage in Malaysia, Siak River drainage in Sumatra, and Belait River drainage in Brunei.

According to Kottelat *et al.* (1993), the standard length of *Rasbora tornieri* is 120 mm, lateral line scale is range 33 to 35 with addition of 2 scale on caudal base, and 14 circumpeduncular scales. However, Parenti and Lim (2005) claims that *Rasbora tornieri* is the largest *Rasbora* genus in the Rajang River, Sarawak with standard length of 163 mm. This species has a slender body, silvery belly, gold stripe that extends from operculum to the caudal peduncle, and caudal fin with thin black rear edge. This fish is closely related to *Rasbora dusonensis* and belong to the *Rasbora argyrotaenia* group (Liao *et al.*, 2010).

2.2.4 *Rasbora borapetensis* (Smith, 1934)



Figure 2.5: *Rasbora borapetensis*

Rasbora borapetensis or also known as red-tailed *Rasbora* was first discovered by Smith in 1934. It is small in size and does not exceed 50 mm in total length. From the Figure 2.5, this *rasbora* has a black lateral stripe started from opercle to the base of caudal fin. On the upper check, there are fine melanophores observed. There is also a green-gold stripe along the black lateral stripe dorsally. The belly is silver in colour. The base of the caudal fin is reddish. All fins are hyaline. The closest relatives of this *rasbora* are *Rasbora taeniata* and *Rasbora pauciperforata* and belong to *Rasbora pauciperforata* group (Brittan, 1954). This species is usually distributed in a single lake, small streams, and roadside drainage.

2.3 Taxonomy classification of *Rasbora*

Taxonomy classification is a system of hierarchical with the purpose of classification organisms to the species level. This hierarchical system grouping the organisms into Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species was developed by Carolus Linnaeus in 1758. He believed that God created each species to be an unchanging entity. He developed the nomenclature system for species and hierarchical classification system for higher categories that are useful in organizing the rapidly expanding knowledge of plants and animals. Later, this system was modified to presume the evolutionary relationship among numerous taxa. Then, the evolutionary trees were drawn. This modern classification systems allows ichthyologist to study the evolutionary relationships among organisms (Moyle and Cech, 2004). The broadest classifications of organism are by domain and kingdom while the specific classifications are by genus and species (refer to Table 2.1).

The taxonomy study provides useful information in order to develop an understanding of evolutionary relationships among fishes. There are six categories used in taxonomic studies of fish which are morphometric measurements, meristic traits, anatomical characteristics, colour patterns, karyotypes, and biochemical characteristics. Morphometric measurements are

including any standard measurements that can be made on fish body such as standard length and head length. These measurement are usually expressed as ratios to standard length and being standardized to make them more comparable among the fish species. Meristic traits are including anything on the fish that can be counted such as fin rays and scale rows. Anatomical characteristics are including shape, completeness, and position of features like lateral lines, internal organs or secondary sexual characteristics. Colour patterns in fish including pigmentation patterns such as dark spots and stripes. The colour may change with age, time, and habitats where the fish is found. Karyotypes are including number and morphology of chromosomes per cell. Biochemical characteristics are patterns in proteins and genetic material that used to quantify genetic differences among populations. Electrophoresis technique and various methods such as mitochondrial DNA and DNA fingerprinting used to analyse DNA composition and structure (Moyle and Cech, 2004).

Table 2.1: Classification level of *Rasbora*.

Classification Level	Name
Domain	Eukaryota
Kingdom	Animalia
Phylum	Chordata
Class	Actinopterygii
Order	Cypriniformes
Family	Cyprinidae
Genus	Rasbora

Fish can be divided into four vertebrate classes which are Actinopterygii (ray-finned fish), Chondrichthyes (cartilaginous fish), Cephalaspidomorphi (lampreys and lamperns), and Pteraspidomorphi (hagfish). Table 2.1 illustrates the classification level of *Rasbora* from

domain to genus level. Genus *Rasbora* belongs to vertebrate class of Actinopterygii. The vast majority of freshwater fish belong to this class. The main characteristic of fish from this vertebrate class is their fins are all supported by rays. These ray-finned characteristics enable them to be opened or folded at will and improve manoeuvrability in the water (Gilpin, 2011). The Cypriniformes order are belong to Actinopterygii class. The 5 important families in this order are Cyprinidae, Parapsilorhynchidae, Balitoridae, Cobitidae, and Psilorhynchidae. The characteristics of this order are including tooth less, head without scale, barbells around the mouth either presence or absence, single dorsal fin, absence of adipose dorsal fin except for some nemachelins and cobitins, and Weberian apparatus is mostly modified.

Genus *Rasbora* belongs to Cyprinidae family. There are many new *rasbora* species that has been identified. For example, *Rasbora maninjau* or knowns as Bada fish by Lumbantobing (2014) is described from northern Sumatra. This new species initially identified as *Rasbora sumatrana* because they share high morphological similarity such as a unique black pigmentation pattern of midlateral stripe, supra-anal, and blotch spot on caudal peduncle. This proved that cryptic species diversity are highly presence in *rasbora* species. According to Sulaiman and Mayden (2012), there are 14 species of this genus found in the Rajang River, Sarawak and Danau Sentarum National Park in western Kalimantan. It has been claimed that many taxonomic groups awaits for the monophyly test and accurate assessments of *Rasbora* diversity. Thus, morphometry (morphometric measurements and meristic traits) and molecular study such as mitochondrial DNA analysis are used in this research.

2.4 Morphometry analysis

The morphometry analysis is performed to provide the biology understanding of the fishes studied. This analysis have been commonly used to delineate stocks of fish (Turan *et al.*,

2006). The morphometric and meristic methods are the quantitative morphology that can be studied in the identification of selected species.

The morphometric measurement is a method used to describe the body form that encompasses size and shape of the fishes. Generally, this method is practice by the ichthyologist to looks at the measurable part of the fish anatomy such as size of body components and body length (Muchlisin, 2013).

Meanwhile the meristic method is a counting of the quantitative features in the fish, for example, the number of scales, fin rays, and gill rakers. Different species have different number of these features which are useful for their identification. The meristic traits are discrete variables which mean no decimal places. For example, number of pelvic fin rays is either 8 or 9, not 8.5. The meristic counts is usually done on dead fish sample that has been preserved in formalin or ethanol. This because the meristic features are clearly visible in the preserved solutions compared to live specimens especially in extreme small size specimen that require microscopic observation.

However, using morphological data alone are unable to resolve the taxonomic status on cryptic species. Molecular studies have been found to resolve many questions addressing fish systematics. Therefore, combining morphology and molecular data has enabled a much more feasible approach to resolve the taxonomy for most species.

2.5 Mitochondrial DNA (mtDNA) sequence

Mitochondria are organelles that can be found in nearly all eukaryotic cells that convert cellular metabolism into Adenosine Triphosphate (ATP) through oxidative phosphorylation process. Mitochondria is the only organelle (exclude nucleus) that contains its own genome, the mitochondrial DNA (mtDNA) and able to reproduce independently in the cells. The

mtDNA is located in the mitochondrial matrix. The structure and gene organisation of mtDNA is highly conserved in mammals (Taanman, 1999).

The mtDNA in mammals is a closed-circular, double-stranded DNA molecule of about 16.6 kb. The double strands of DNA molecule compose of GT base which results in different buoyant densities of each strand in denaturing caesium chloride (CsCl) gradients. There are two types of strand which are heavy (H) strand and light (L) strand. The heavy (H) strand coded for 2 ribosomal RNAs (rRNA), 14 transfer RNAs (tRNA), and 12 polypeptides. The light (L) strand coded for 8 transfer RNA (tRNA) and a polypeptide. The 13 proteins products are constituents of the enzyme complexes of the oxidative phosphorylation system (Taanman, 1999). The control region of mtDNA involved in controlling the replication of mtDNA and transcription of RNA. This control region also known as displacement loop (D-loop). During replication, one of the two helix strands is displaced by the synthesis of new strand. The D-loop in fishes is usually long and contains tandemly repeated segments (Kocher and Stepien, 1997).

Mitochondrial DNA is particularly useful in the study of population within species because this DNA lacks of recombination. This DNA has a simple inheritance pattern which is from the mother line and also evolve faster than nuclear DNA (Moyle and Cech, 2004). Taanman (1999) also recognized that mtDNA are only transmitted through maternal germ line. In mammalian the copy number of mtDNA in oocytes is higher than in sperm cells. This has been detected in the interspecific hybrids of mice species which are *Mus musculus* and *Mus spretus* from pronucleus stage to neonates stage of development using the polymerase chain reaction (PCR) method.

Mitochondrial DNA have been well studied in fishes. Restriction fragment length polymorphisms (RFLPs) in fish has been analysed where the whole mtDNA are digested with specific endonucleases. Then, the gel electrophoresis separated the product by size. The

knowledge of universal primer sequences are also accessible. As in the other vertebrates, the structure of mtDNA, gene order, and secondary structure in fishes are largely conserved as mtDNA is inherited as one unit and has been characterized as a single gene sampling (Kocher and Stepien, 1997).

According to Chen *et al.* (2016), the range of complete mtDNA in fish are 14 kbp to 18 kbp. Complete mtDNA genome provide information such as gene arrangement, nucleotide composition, gene content, and replication and transcription modes. These information are useful in species identification, population genetics, and phylogenetic relationship. The largest fish in Amazon basin which is *Arapaima gigas* has been successfully sequence for complete mitochondrial genome with total length of 16,433 bp (refer to Figure 2.6) by using PCR method (Hrbek and Farias, 2008). In Chen *et al.* (2016) study, the whole mitochondrial genome of *Brama japonica* was successfully characterized by PCR and primer-walking sequencing method. This method also effectively sequence the complete mitochondrial genome of *Rasbora aprotaenia* and *Rasbora lateristriata*. Mitochondrial genome in both *rasbora* species encodes 37 genes for 13 proteins, 22 tRNAs, and 2 rRNAs. All protein genes except for COI gene start with an ATG codon while COI gene start with GTG codon. A total of 7 protein genes contains stop codon and the remaining protein genes contains stop codon that are post-transcriptionally produced by polyadenylation (Kusuma and Kumazawa, 2015).

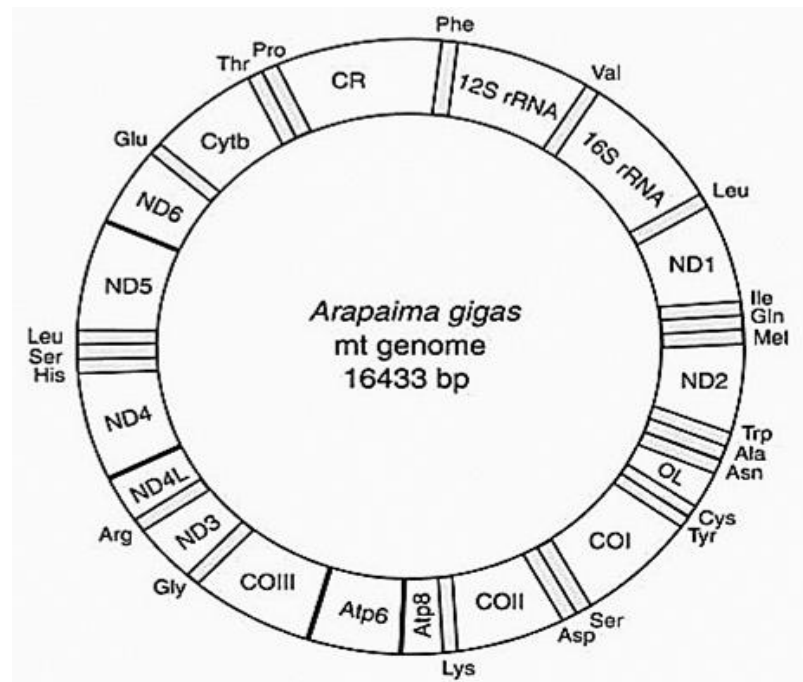


Figure 2.6: The circular gene map of *Arapaima gigas* mtDNA (Hrbek and Farias, 2008)

2.6 Phylogenetic analysis

A phylogenetic tree is a diagram that represents an evolutionary relationship between the organisms (also known as taxa). The taxa can be species, population, sequences or individual. Their evolutionary relationships are generated based on their genetic closeness (Mahapatro *et al.*, 2012). The tree structure is composed of nodes and branches. A single branch links any two adjacent nodes. The length of branch in horizontal line between the nodes represents the evolutionary lineages changing over time. The longer the horizontal lines of branch, the higher the amount of changes. The branch length units are the nucleotide substitutions per site. The vertical line of branch has no meaning. The nodes represent the taxonomic units. For example, sequences. Each node represents a speciation event in an evolution. There are 2 types of nodes which are external nodes and internal nodes. The external nodes or leaf represent the data under comparison which known as operational taxonomic units

(OTUs) whereas the internal nodes represent inferred ancestral units or known as hypothetical taxonomic units (HTUs).

Tree topology is the branching patterns of a tree. The difference branching patterns of trees can be used to illustrate the relatedness between the taxa. The patterns represents all the evolutionary relationships except for time and genetic distance. The tree root is the node that represents the common ancestor of all taxa (Talib *et al.*, 2016). In the rooted tree, time is represented by a single direction. Although the direction may be right, left, up or down, the root stay at the opposite end of the graph away from the leaf. A clade is a group of leaves that form a monophyletic group. Unrooted tree is portrayed without reference to the time direction. All the rooted tree are present in unrooted tree, where the root would be one of the tree edges. A clade in the unrooted tree would be any group of taxa that can be separated from the rest by removing a single edge.

There are many methods available that can be used to compute the phylogenetic tree. The main computational methods are cladistics (based on characteristics), and phenetics (distance-based). Cladistics method is composed of characteristics data for a number of analysed organisms. This method is able to map the characters or traits changes on the tree in order to see the time of specific traits evolved and which taxa is involved. Usually, a trait is possess from a common ancestor. This method works best for morphological data. Cladistics consists of maximum parsimony (MP) method and maximum likelihood (ML) method. Maximum parsimony builds tree based on minimum total number of evolutionary steps involved in given data (Kannan and Wheeler, 2012). Maximum likelihood method is a probability of DNA sequences given the tree. Phenetics method is a mathematical index of similarity among taxa to construct a branching network. This method is working for both morphological and molecular data. However, the results for morphological data is not reliable as the cladistics method because the comparative metric data such as percentage similarity was

used. Molecular evolution can be observe from the speciation pattern. Each node represents lineage of varying individuals that had a specific gene sequence which was ancestral prior of each hereditary species formation (Staton, 2015). Phenetics composed of neighbour joining (NJ) method and unweighted pair group method with arithmetic mean (UPGMA). NJ tree contains the distance of pairwise evolutionary between elements of sequences set. UPGMA method is a clustering of hierarchical tree and estimate a rate of evolution (Bíró, 2015).

CHAPTER 3

METHODOLOGY

3.1 Site descriptions and sample collection

The *Rasbora* samples has been obtained through several sites in Sarawak including Kanowit (Figure 3.1), Kota Samarahan (Figure 3.2), and Lundu (Figure 3.3). These sites were randomly selected depending on its accessibility by car and boat. In order to mark the location of each sampling site, the geographical coordinates were recorded using global positioning system (GPS) as in Table 3.1.

Table 3.1: Sampling site descriptions

Samples	Sites	North (N)	East (E)	Number of sample
<i>Rasbora sumatrana</i>	Bajo Bergantong River, Lundu	1°44'18.6"	109°52'05.9"	20
<i>Rasbora hobelmani</i>	Latong River, Machan, Kanowit, Sibu	1°59'55.0"	112°00'55.3"	11
<i>Rasbora tornieri</i>	Kanowit River, Sibu	2°01'33.4"	112°01'26.9"	20
<i>Rasbora borapetensis</i>	Drainage along Kuching Outer Ring Road, Kota Samarahan	1°25'36.7"	110°24'41.3"	7
Total sample				58

Cast net, scoop net and the fish traps were used to catch the *Rasbora* fishes. All samples were held in the portable fish bucket with oxygen supplied by the portable air pump. The samples were identified in the field whenever it is possible and photographed. The standard

length and total length of each specimen was measured at the sampling site in millimetre (mm) unit. The samples were then preserved in 95% of ethanol after anesthetized with tricaine solution and brought to the laboratory for proper identification.

The morphometry analysis and molecular work were conducted for species identification. A total of 58 fish samples were used for morphometry analysis where 20 from Bajo Bergantong river in Lundu, 11 from Latong river in Sibul, 20 from Kanowit river in Sibul, and 7 from the drainage along the road of Kota Samarahan.



Figure 3.1: Sampling site at Kanowit river in Sibul

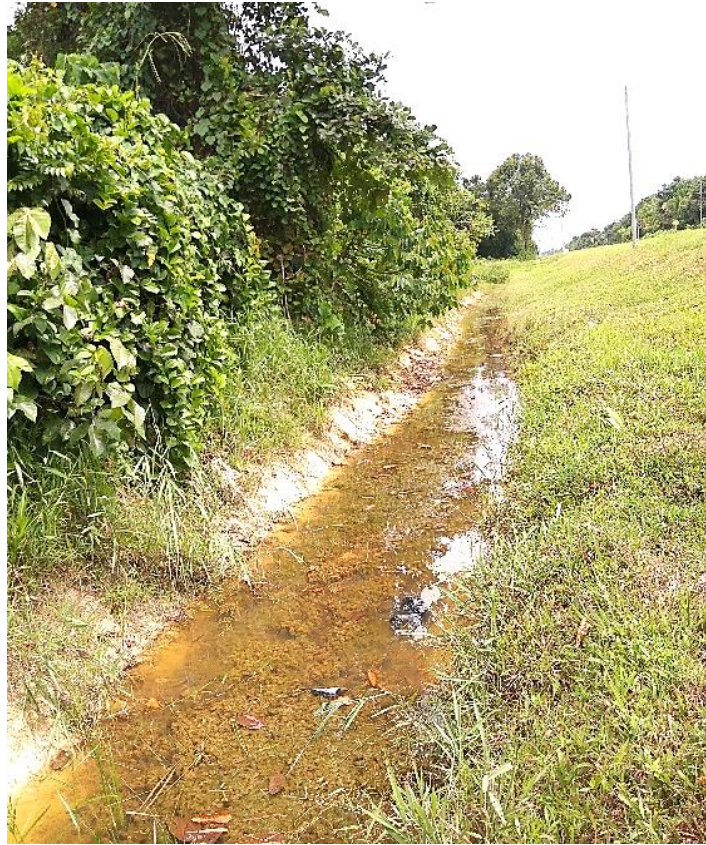


Figure 3.2: Sampling site at drainage along Kuching Outer Ring Road at Kota Samarahan



Figure 3.3: Sampling site at Bajo Bergantong river, Lundu

3.2 Sample Assortment and Maintenance

All samples were preserved in 95% ethanol before proceed with the molecular work in the laboratory. The purposes of the ethanol preservation used in this study were to reduce the DNA degrading of the specimens and to maintain the specimen tissue structure for a long-term storage (Berbel-Filho *et al.*, 2013). All the preserved samples were stored individually in 15 mL and 50 mL of centrifuge tube that fitting with their length. Bigger samples like *Rasbora hobelmani* were stored individually in a white clear plastic container. Smaller samples that less than 20 mm of total length like *Rasbora borapetensis* were stored individually in 2 mL of micro centrifuge tube. All fish samples were ensured to fully immerse in the preservative solution. The colour changes of the samples before and preserved in the 95% of ethanol were photographed. The replacement of the preservation solution was made when the solution changed colour.

3.3 Morphometry procedure

Every collected fish samples were measured on the left side of the body by using Vernier callipers with the accuracy of 0.05 mm. The samples were laid on a clean flat surface area to avoid an error while measuring. The meristic counts and morphometric measurements were taken according to Kottelat and Tan (2012) and Lumbantobing (2014) with minor modification.

3.3.1 Morphometric Measurements

A total of 24 characters were measured for the morphometric data of every samples by using the Vernier calliper. The measurements data were assorted by species. The following 24 morphometric characters were used in this research (Table 3.2) and illustrated in Figure 3.4.

Table 3.2: Morphometric characters used for measurement

No	Measurement Characters	Descriptions
1	Standard Length	Measurement from snout tip to the base of caudal fin.
2	Total Length	Measurement from the snout tip to the posterior end of caudal fin.
3	Head Length	Measurement of anterior end of snout to the posterior edge of operculum.
4	Pre-Dorsal Length	Measurement from snout tip to the base of first dorsal spine.
5	Pre-Anal Length	Measurement from snout tip to the first anal fin base.
6	Pre-Pelvic Length	Measurement from snout tip to the origin of pelvic fin.
7	Dorsal Depth	Vertical measurement from the first dorsal fin base to the first dorsal fin tip.
8	Body Depth	Vertical measurement between dorsal fin base and pelvic fin base.
9	Caudal Peduncle Depth	Depth measurement from dorsal caudal peduncle to ventral caudal peduncle.
10	Caudal Peduncle Length	Measurement from last anal fin base to base of caudal fin.

Table 3.2 continued

11	Dorsal Fin Base Length	Measurement from first dorsal fin base to last dorsal fin base in straight line.
12	Anal Fin Base Length	Measurement from first anal fin base to last anal fin base in straight line.
13	Pelvic Fin Length	Measurement from the farthest tip of pelvic fin to base of pelvic fin.
14	Pectoral Fin Length	Measurement from farthest tip of pectoral fin to base of pectoral fin.
15	Upper Caudal Lobe Length	Measurement from upper caudal fin tip to upper caudal fin base.
16	Median Caudal Lobe Length	Measurement from median caudal fin tip to median caudal fin base.
17	Lower Caudal Lobe Length	Measurement from lower caudal fin tip to lower caudal fin base.
18	Dorso-Hypural Distance	Measurement from last dorsal fin base to caudal fin base.
19	Eye Diameter	Diameter measurement from anterior to posterior end of eye orbit.

Table 3.2 continued

20	Snout Length	Measurement from snout tip to the front margin of eye orbit.
21	Head Width	Measurement from dorsal view between eye margins in straight line.
22	Head Depth	Vertical measurement between head top to lower jaw base, in a straight line to the middle eye.
23	Inter-Orbital Width	Measurement from dorsal view between snout margins.
24	Anal Depth	Measurement from first anal fin tip to first anal fin base.

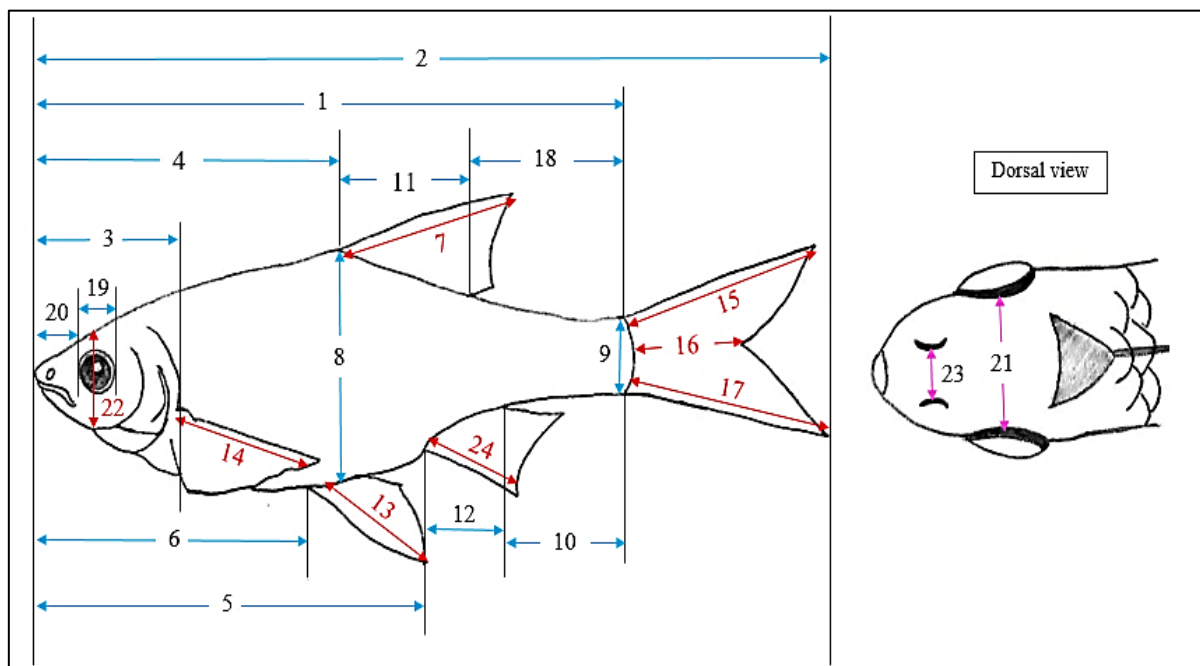


Figure 3.4: Schematic illustration of morphometric characters used

3.3.2 Meristic Counts

A total of 11 countable characters were taken into consideration for this research. Microscopic observation was needed for the small size samples like *Rasbora borapetensis*. The counting data were tabled by species. The following 11 countable characters were used in this research (Table 3.3) and illustrated in Figure 3.5.

Table 3.3: Meristic Characters used for counting

No	Meristic Characters	Descriptions
1	Dorsal Fin	Counted all rays in dorsal fin.
2	Pectoral Fin	Counted all rays in pectoral fin.
3	Pelvic Fin	Counted all rays in pelvic fin.
4	Anal Fin	Counted all rays in anal fin.
5	Caudal Fin	Counted all rays in caudal fin.
6	Dorsal Pro-Current	Counted number of rays at dorsal area of caudal fin base.
7	Ventral Pro-Current	Counted the number of rays at ventral area of caudal fin base.
8	Scales of Lateral Line	Counted the scales on the lateral line.
9	Pre-Dorsal Scale	Counted the scales from top of head to first dorsal fin base.
10	Transverse Scale	Counted the scales from the first dorsal fin base to lateral line and from the lateral line to the first anal fin base.
11	Circumpeduncular Scale	Counted the circumference scales around the caudal peduncle base.

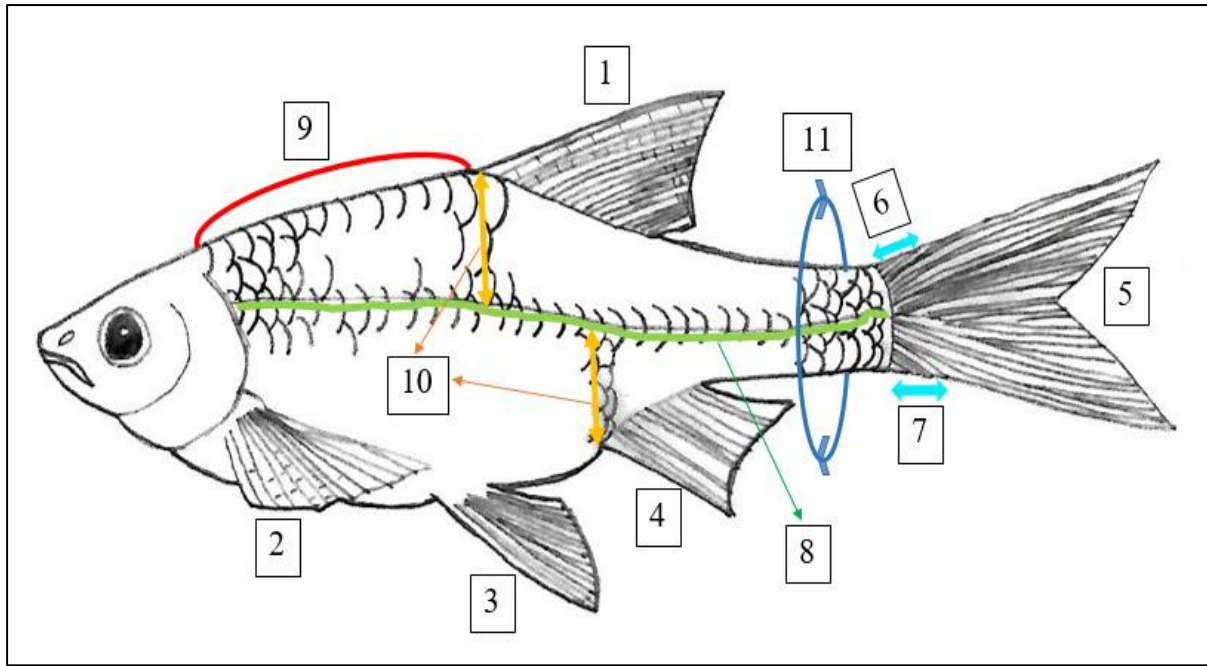


Figure 3.5: Schematic illustration of meristic characters used.

3.3.3 Morphometry Analysis

The morphometry analysis were composed of two separate data sets which are morphometric measurements (SI unit: millimetres) and meristic counts. All continuous and discrete data were tabled using Microsoft excel 2016. Then, these data were computed by using the statistical software which is Statistical Package for Social Science (SPSS) version 22.

The original data for morphometric measurements (SI unit: millimetres) for all *Rasbora* species were converted to percentage of standard length and percentage of head length. Mean, range, and standard deviation (std.) can be obtained from the converted data of morphometric measurements. The degree of differentiation of body size such as eye diameter and caudal peduncle length across the *Rasbora* species were able to be compared in these data. According to Turan *et al.* (2004), differences in body shape should be contributed to variation and not the relative size of the fish. Hence, for further analysis, the size adjustment was needed to remove size-dependent variation in morphometric data.

The original measurement of morphometric (SI unit: millimetres) data was adjusted using Elliott *et al.* (1995) formula to standardized the samples size between all samples.

$$M_{adj} = M \left(\frac{L_s}{L_o} \right)^b$$

Where M_{adj} : Adjusted measurement

M: Original measurement

L_s : Overall mean standard length for all fish samples

L_o : Standard length

b: Slope of regression of Log M on Log L_o using all fish samples.

Standard length character was excluded from size adjustment analysis. In order to do linear regression, all measurement data (SI unit: millimetre) were transformed to common logarithms. According to Hair *et al.* (1998), linearity and normality are usually more approximated by logarithms compared to original variables. The linear regression analysis of the morphometric character (logarithms transformed data) were computed by taking standard length as an independent variable and other character as dependent variable. The regression coefficient is the parameter b in the size adjustment formula.

The analysis of variance (ANOVA) was carried out on the size adjusted data to find the significant different of each morphometric character among the four *Rasbora* species with the significance level of 0.05. For further analysis, discriminant function analysis (DFA) was only performed on the significance morphometric character based on ANOVA using the size adjusted data. These analyses were used to identify and differentiate the *Rasbora* samples based on species.

The meristic data does not require any adjustment because it is discrete data. Besides, according to Barlow (1961), meristic characters are fixed in early development and remain unchanged throughout life, thus no size correction was applied. From the collected data of

meristic counting in Microsoft excel, range (Minimum-Maximum) and mode can be obtained. Non-parametric Kruskal-wallis test was applied to compare the meristic characters across the species (Mollah *et al.*, 2012). This test also used by Syaifullah *et al* (2015) study in morphometric variations of *Zenarchopterus buffonis*. The significance different ($p \leq 0.05$) of character among the fish can be determined by this test. Then, discriminant function analysis (DFA) was employed.

3.4 Molecular procedure

The molecular work was conducted after morphometric measurement and meristic counting were completed. Polymerase chain reaction is a technique used to amplify the mtDNA of *rasbora* species by using designated primers.

3.4.1 Mitochondrial DNA (mtDNA) extraction

The sample that has been preserved in the 95% (w/v) ethanol was used for DNA extraction. The DNA extraction protocol was followed according to Kumar *et al.* (2007) with minor modifications.

A volume of 910 μ L of lysis buffer was pipetted into 2 mL micro centrifuge tube. The contents of the lysis buffer were 200 mM Tris-HCL (pH 8), 100 mM EDTA (pH 8), and 250 mM NaCl. A volume of 60 μ L of 10% (w/v) SDS were added into the same tube. Then, tissue sample was cut for about 2 cm by using sterile blade on the clean board and was transfer into the same tube. A volume of 30 μ L of Proteinase K (10 mg/mL) was added.

The tube was then incubated at 48°C for 50 minutes in the water bath. The tube was centrifuged for 10 minutes at 15,000 x *g*. The supernatant was transferred into a new 2 mL micro centrifuge tube. An equal volume of freshly prepared PCI at ratio of 25:24:1 was added

into the tube and inverted gently to mix them properly. The tube was leaves for 10 minutes in order to precipitate the proteins and other nucleic acids. After 10 minutes, the tube was centrifuged at 15,000 x *g* for 10 minutes. Three phases of layer (top aqueous, interphase, and lower phase) can be seen in the tube. Only top aqueous layer were transferred into a new 1.5 mL micro centrifuge tube.

An equal volume of isopropanol was added and gently inverted the tube for several time to precipitate the DNA before centrifuged for 10 minutes at 15,000 x *g*. At this step, the DNA pellet can be observed. The supernatant was gently removed to reduce loss of DNA pellet. The DNA pellet was washed with 1 mL chilled 70% (w/v) of ethanol. Then, centrifuges for 5 minutes at 15,000 x *g*. The supernatant was discarded. The pellet was leaves for 30 minutes to dry via air drying in the fume hood. Lastly, the DNA pellet was resuspended in 100 µL of sterile distilled water and stored at -20°C with complete labelled until further use (Kumar *et al.*, 2007).

3.4.2 Primer design protocol

The National Centre for Biotechnology Information (NCBI) website (<http://www.ncbi.nlm.nih.gov/genbank/>) was accessed to search for the complete mitochondrial genome of *Rasbora* species. The FASTA format of the complete mitochondrial sequence of selected species were retrieved. In this study, three species of *Rasbora* namely *Rasbora sumatrana*, *Rasbora steineri*, and *Rasbora trilineata* were selected. Then, the multiple sequence alignment of the *Rasbora* species was conducted by using Clustal Omega interface webpage. The sequence alignment was copy into word document to identify the conserved gene regions between the *Rasbora* species. Primer3Plus software was used to design the forward and reverse primer of the identified conserved gene regions. The primer designated

was verified by using Oligonucleotide Properties Calculator. The primer sequences were sent to the First BASE Laboratories for subsequent synthesis.

3.4.3 Long polymerase chain reaction

The optimum annealing temperature of designated primers that were paired for the complete mitochondrial genome consisted of small and large fragment for each species were determined by using T100™ Thermal Cycler (Bio-Rad, USA). Final volume of 20 µL for PCR reaction was prepared which composed of double distilled water, 10X LA PCR Buffer II, 2.5 mM dNTPs, 10 µM forward primer, 10 µM reverse primer, *TaKaRa LA Taq* (5U/µL), and DNA template (Table 3.4). The PCR was performed for 40 cycles. This PCR for both small and large fragment taken 5 to 6 hours to complete.

Table 3.4: Components of PCR

No	Components	Final concentration	1X Volume (µL)	Negative Control (µL)
1	ddH ₂ O	N/A	13.00	13.80
2	10X LA PCR buffer II	1X	2.00	2.00
3	2.5 mM dNTPs	0.40 mM	3.20	3.20
4	10 µM forward primer	0.20 µM	0.40	0.40
5	10 µM reverse primer	0.20 µM	0.40	0.40
6	<i>TaKaRa LA Taq</i> (5U/µL)	0.05 units	0.20	0.20
7	DNA template	As required	0.80	-
Final volume			20.0	20.0

After long PCR, the electrophoresis was conducted by using Enduro™ Power Supplies (Labnet International, North America). Agarose gel with 1% (w/v) concentration was prepared by diluting the 0.25 g of agarose powder in 25 mL of 1X TAE buffer. The mixture was heated in microwave for 60 seconds. Every 10 seconds break, the mixture was checked and swirled until agarose powder dissolves completely in a buffer. The solution was cooled to 60°C before gently poured into a gel tray with comb that has been inserted earlier. The gel was allowed to solidify for 40 minutes. The 4 µL of 1 KB ladder was loaded into gel well. The 1 µL of loading dye was added into 5 µL of DNA sample and loaded into gel well (Chung, 2013). The gel was run at 80 V for 45 minutes in 1X TAE buffer. After that, the gel was post-stained with ethidium bromide for 45 minutes and visualised under UV transilluminator (MaestroGen).

3.4.4 Purification of PCR amplified product

Polymerase chain reaction (PCR) products were purified by using EasyPure PCR Purification kit (TransGene Biotech, Beijing). In a 1.5 mL of microcentrifuge tube, 5 volume of binding buffer (BB) were added to 1 volume of PCR products and mixed them briefly by gently pipetting up and down. All the mixture were transferred directly to the centre of a spin column with a collection tube that has been provided in the kit. The tube was incubated for 1 minute at room temperature in order to increase the yield of purified DNA.

After 1 minute, the tube was centrifuged at 15,000 x *g* for 1 minute. The filtrate was removed. A volume of 650 µL of wash buffer (WB) was added into the column tube directly to the centre. The column tube was centrifuged at 15,000 x *g* for 1 minute. The filtrate was discarded. The column tube was centrifuged for second times to remove the WB residual.

The spin column was then placed in a new 1.5 mL of micro centrifuge tube. A volume of 30 µL of elution buffer (EB) was added directly to the centre of column tube. The tube was

incubated at room temperature for 1 minute. Lastly, centrifuged the tube at 15,000 x *g* for 1 minute to elute the DNA. The tube was labelled properly and stored at -20 °C.

3.4.5 DNA sequencing of PCR amplified product

A total of 20 µL of the purified PCR product between the size of 7,000 bp to 12,000 bp was transferred into 1.5 ml micro centrifuge tube and labelled properly. Purified amplicons were sequenced pair-endedly with short reads via Illumina HiSeq 400 System. All sequencing reads were checked for quality and adapted-trimmed using cutadapt (Martin, 2011) before assembling via de novo assembler SPAdes (Bankeyich *et al.*, 2012).

3.4.6 Sequence analysis

The output results obtained from the sequencing were compared and blasted by using nucleotide BLAST (blastn) from NCBI website to obtain the score bits, blast identity, E-value, GC contents, gaps between the query and subject, and similar species. The sequence from small and large fragment were aligned to generate the complete mitochondrial genome for each *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis*.

The gene identification was according to Chen *et al.* (2016) study of *Brama japonica* mitochondrial DNA genome. The overall nucleotide composition of these 4 *Rasbora* species sequence were obtained by using molecular evolutionary genetic analysis (MEGA) version 7.0 software. This software was downloaded from MEGA software website (<https://www.megasoftware.net/>). The transfer RNA (tRNA) genes and anti-codons in genomic sequence of each species were identified by using tRNA scanner which is tRNA scan SE version 2.0 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) in a default search mode and vertebrate

mitochondrial as sequence source for structure prediction (Lowe and Chan, 2016). The circular mitochondrial genome map of each *Rasbora* species was generated by using the MitoAnnotator server from the website of MitoFish Mitochondrial Genome Database of Fish (<http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html>). Ribosomal RNA genes and protein coding genes were also determined by MitoFish webserver. Then, the coding sequence was translated into amino acid sequence by using MEGA 7.0 software to check whether the functional amplified domains are without frame-shifting or without premature stop codons.

3.4.7 Phylogenetic analysis

The phylogenetic relationship of these 4 *Rasbora* species were determined by constructed a phylogenetic tree using molecular evolutionary genetic analysis (MEGA) version 7.0 software. The concatenated 13 protein coding genes of 19 sequences were aligned by using ClustalW in MEGA 7.0 software. The best DNA model was identified by running the model test on these multi alignments. Then, the phylogeny test was conducted by using Tamura-Nei (TN93) model, 1000 bootstraps, and maximum likelihood method. The complete mitochondrial genome of *Rasbora lateristriata* LC021505.1, *Rasbora vaterifloris* AP011432.1, *Rasbora daniconius* AP011285.1, *Rasbora aprotaenia* LC021504.1, *Rasbora cephalotaenia* AP011430.1, *Rasbora argyrotaenia* LC269105.1, *Rasbora steineri* NC_020005.1, *Rasbora trilineata* NC_025336.1, *Rasbora borapetensis* AB924546.1, *Rasbora sumatrana* AP011221.1, *Trigonostigma espei* AP011449.1, *Trigonostigma heteromorpha* AP011421.1, *Danio rerio* AC024175.3, *Boraras maculatus* AP011420.1, and *Cheilinus undulates* KM461717.1 were downloaded from GenBank database to compare with complete mitochondrial genome of four studied *Rasbora* species. *Cheilinus undulates* KM461717.1 was added as outgroup in a phylogenetic tree.

CHAPTER 4

RESULTS

4.1 Sample colouration

The colour pattern of the specimens were faded after preserved in 95% ethanol (Figure 4.1). The most noticeable colour changes in all specimens after preserved in 95% ethanol was the eyes. Black eyes turned to fully white colour after stored in tube or container containing 95% ethanol for 30 minutes. However, some part of the specimen like caudal fin are able to retain their original colour. In some species, the lateral stripe, lateral line, and reticulate pattern become more visible in the preservative solution.

From the Figure 4.1, the colouration of living *Rasbora sumatrana* are shiny silvery - yellowish body, yellowish fins, and have black tips at the caudal lobes with thin black posterior margin. In some *Rasbora sumatrana*, the black tip at the caudal lobes are absence. This feature was stated by Lumbantobing (2014) where the triangular dark patch on each caudal lobe is cause by pigmentation. The dark margin presence on the fin rays are caused by small melanophores. The last pro-current ray and five longest principal rays of each caudal fin lobe are pigmented with dense melanophores. The reticulate pattern on the anal base extends to the caudal peduncle. In preservative solution, the body turn to brown, darker head, and the caudal lobes remain black. The dusky pigmentation is present on subpeduncular. The reticulate pigmentation in alcohol is thicker on the anal fin base and become thinner towards the caudal peduncle.

The colouration of living *Rasbora hobelmani* is yellowish brown and greyish on the back and the head. The pectoral fin, dorsal fin, and caudal fin are yellow with black posterior margin. The pelvic fin and anal fin are hyaline. The reticulate pigments is thicker on the anal base and become thinner towards the caudal peduncle. In preservative solution, the body colour

pattern of *Rasbora hobelmani* turn to brown, darker head, and yellow colour of caudal fin become less visible.

The colouration of living *Rasbora tornieri* are silvery body with hyaline fins except for caudal fin. The colour pattern of caudal fin is pale yellow with very thin grey or black posterior margin. In preservative solution, the colour pattern of *Rasbora tornieri* body turn to yellowish-brown. The lateral line and lateral stripe become visible. According to Ng and Kottelat (2013) article, the head and body of dorsal surfaces are brown after immersed in 70% ethanol. The ventral surfaces are unpigmented. On each caudal lobe, the yellow brownish pigments are presence on the proximal part of 4 to 5 longest principal rays.

The colouration of living *Rasbora borapetensis* are silvery at the bottom especially the belly, red colour of caudal fin base, visible black lateral stripe along with green-gold stripe and presence of reticulate pattern on the anal base extends to caudal peduncle. In preservative solution, this *rasbora* retained its red colour of caudal fin base, the upper body turn to brown while bottom body turn to silvery white, and other fins are hyaline. The black lateral stripe and reticulate pattern on the anal base become very visible in 95% ethanol.



Figure 4.1: Samples before (left) and after (right) preserved in 95% of ethanol. a) *Rasbora sumatrana*; b) *Rasbora hobelmani*; c) *Rasbora tornieri*; d) *Rasbora borapetensis*

4.2 Morphometric Analysis

Descriptive data for the 24 morphometric characters in Table 4.1 shows the range (Minimum-Maximum), mean, and standard deviation (std.) of 4 *Rasbora* species. All linear measurements were taken on the left side of each samples in millimetre (mm) by using Vernier calliper with accuracy of 0.05 mm. All measurements excluding standard length character were calculated to the means ratio of standard length and head length to compare the degree of differentiation of each character for each *Rasbora* species.

As stated by Kusuma *et al* (2016), *rasbora* species is small to moderate size. Table 4.1 show the standard length of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis*. *Rasbora borapetensis* is the smallest size compare to the rest of studied *rasbora* with mean of 21.95 and standard deviation of 3.68. This increasing size followed by *Rasbora sumatrana*, *Rasbora tornieri*, and *Rasbora hobelmani*. *Rasbora hobelmani* (23.8-25.9% SL) and *Rasbora tonieri* (23.5-25.8% SL) have a smaller head length compare to *Rasbora sumatrana* (23.8-28.4% SL) and *Rasbora borapetensis* (24.9-27.6% SL). *Rasbora borapetensis* (8.0-11.2% SL) has a thinner caudal peduncle followed by *Rasbora tornieri* (10.3-12.8% SL), *Rasbora sumatrana* (12.1-14.4% SL), and *Rasbora hobelmani* (13.1-14.6% SL). Upper caudal lobe (23.2-35.4% SL and 20.5-29.9% SL) of *Rasbora sumatana* and *Rasbora borapetensis* is longer than lower caudal lobe with 21.6-33.2% SL and 20.7-30.7% SL respectively. Meanwhile upper caudal lobe (29.2-34.1% SL and 23.6-28.7% SL) of *Rasbora hobelmani* and *Rasbora tornieri* is shorter than lower caudal lobe (29.2-37.7% SL and 20.7-30.7% SL). *Rasbora borapetensis* (31.9-45.0% HL) has the largest eye diameter while *Rasbora tornieri* (27.2-35.3% HL) has a smallest eye diameter.

Table 4.1: Descriptive statistics for morphometric characters of *Rasbora* species

Descriptions	<i>Rasbora sumatrana</i> N=20		<i>Rasbora hobelmani</i> N=11		<i>Rasbora tornieri</i> N=20		<i>Rasbora borapetensis</i> N=7	
	Range	Mean $\pm$ Std	Range	Mean $\pm$ Std	Range	Mean $\pm$ Std	Range	Mean $\pm$ Std
Standard length, SL (mm)	35.25-71.28	51.19 $\pm$ 11.17	66.20-98.35	78.12 $\pm$ 9.57	54.15-78.28	68.03 $\pm$ 7.41	18.13-29.33	21.95 $\pm$ 3.68
Percentage of standard length, SL								
Total length	116.17-134.38	128.68 $\pm$ 4.15	128.55-135.09	132.07 $\pm$ 1.92	124.71-132.53	127.47 $\pm$ 1.88	120.52-131.57	126.43 $\pm$ 4.17
Head length	23.84-28.42	26.26 $\pm$ 1.33	23.84-25.87	24.74 $\pm$ 0.57	23.47-25.79	24.65 $\pm$ 0.63	24.88-27.59	26.23 $\pm$ 1.14
Pre-dorsal length	52.41-55.57	53.78 $\pm$ 0.80	50.30-53.48	52.10 $\pm$ 1.06	51.57-56.98	54.89 $\pm$ 1.38	52.80-57.36	54.95 $\pm$ 1.64
Pre-anal length	67.20-71.69	70.17 $\pm$ 1.22	68.43-73.09	70.54 $\pm$ 1.47	68.03-73.50	70.86 $\pm$ 1.41	63.15-69.90	67.99 $\pm$ 2.28
Pre-pelvic length	48.54-52.70	50.82 $\pm$ 1.12	48.83-52.12	50.47 $\pm$ 1.01	47.56-50.36	49.21 $\pm$ 0.97	46.64-51.03	48.69 $\pm$ 1.58
Dorsal depth	12.00-26.67	21.92 $\pm$ 3.52	22.37-25.82	23.90 $\pm$ 1.09	17.06-22.18	20.78 $\pm$ 1.16	21.64-27.98	24.55 $\pm$ 2.08
Body depth	17.55-50.21	24.47 $\pm$ 6.46	22.94-26.94	24.61 $\pm$ 1.14	20.23-22.72	21.41 $\pm$ 0.68	18.21-25.49	21.99 $\pm$ 2.45
Caudal peduncle depth	12.07-14.46	13.04 $\pm$ 0.76	13.08-14.57	13.78 $\pm$ 0.55	10.33-12.84	11.36 $\pm$ 0.62	8.00-11.24	10.05 $\pm$ 1.06
Caudal peduncle length	16.49-23.52	19.73 $\pm$ 1.91	17.75-19.91	18.92 $\pm$ 0.79	18.65-22.99	20.43 $\pm$ 1.34	15.80-21.06	19.03 $\pm$ 2.14
Dorsal fin base length	5.55-11.60	9.83 $\pm$ 1.20	10.10-12.20	11.30 $\pm$ 0.71	8.12-10.61	9.42 $\pm$ 0.69	7.11-12.97	10.76 $\pm$ 1.83
Anal fin base length	8.39-11.30	9.74 $\pm$ 0.96	9.60-12.20	10.97 $\pm$ 0.78	7.86-10.40	9.03 $\pm$ 0.71	7.72-11.01	9.75 $\pm$ 1.30

Table 4.1 continued

Pelvic length	14.29-20.22	17.02 ± 1.64	18.13-20.59	19.18 ± 0.71	15.79-18.17	17.06 ± 0.61	16.17-21.13	17.92 ± 1.58
Pectoral length	16.42-26.14	21.20 ± 2.49	20.95-23.76	22.16 ± 0.84	21.21-24.38	22.83 ± 0.81	18.16-22.76	20.26 ± 1.85
Upper caudal lobe length	23.19-35.39	29.20 ± 3.23	29.23-34.06	31.40 ± 1.75	23.57-28.74	26.15 ± 1.48	20.52-29.93	25.12 ± 3.78
Median caudal lobe length	9.63-15.67	12.96 ± 1.76	12.16-16.83	13.50 ± 1.40	8.30-11.44	10.05 ± 0.95	9.86-16.15	13.98 ± 2.06
Lower caudal lobe length	21.56-33.24	28.32 ± 2.89	29.16-37.65	32.58 ± 2.17	22.29-31.83	26.95 ± 2.22	20.66-30.69	24.18 ± 3.91
Dorso-hypural Distance	33.85-39.53	36.48 ± 1.84	35.37-38.41	36.77 ± 1.04	33.29-38.29	36.41 ± 1.36	28.83-36.21	33.03 ± 2.92
Anal depth	11.36-20.44	17.42 ± 1.93	17.66-20.85	19.09 ± 0.99	14.04-18.79	16.13 ± 1.00	15.85-18.54	17.36 ± 1.13
Percentage of Head Length, HL								
Eye diameter	30.30-38.41	34.23 ± 1.91	28.31-35.29	31.13 ± 2.39	27.15-35.34	29.74 ± 1.82	31.88-45.00	37.76 ± 4.48
Snout length	21.18-28.74	24.89 ± 2.10	24.50-28.87	26.70 ± 1.36	19.94-26.18	23.88 ± 1.59	18.18-21.29	20.15 ± 1.21
Head width	33.22-41.70	39.14 ± 2.48	29.97-42.86	39.45 ± 4.47	29.17-39.09	34.20 ± 3.01	23.77-41.32	27.97 ± 6.07
Head depth	45.81-58.11	51.67 ± 3.09	48.69-100.78	56.09 ± 14.94	45.47-52.17	48.92 ± 1.83	54.96-65.98	60.54 ± 3.54
Inter-orbital Width	17.48-26.20	20.90 ± 2.35	19.56-24.74	21.65 ± 1.60	16.43-20.94	18.57 ± 1.11	6.00-17.77	9.76 ± 4.30

The letter 'N' represent the number of samples collected and used in the analysis.

Size adjustment was applied to the morphometric measurement data (SI unit: Millimetre) excluding standard length character by following the Elliot et al (1995) formula. Analysis of variance (ANOVA) was then carried out on size adjusted morphometric character among the four *Rasbora* species. Table 4.2 shows that all characters measurement are significantly different at level of 0.05.

Table 4.2: ANOVA for size adjusted morphometric measurements among 4 *Rasbora* species

Characters	F value	P value
Total length	132.545	0.000*
Head length	101.089	0.000*
Pre-Dorsal length	131.773	0.000*
Pre-Anal length	130.280	0.000*
Pre-Pelvic length	121.621	0.000*
Dorsal depth	87.512	0.000*
Body depth	51.989	0.000*
Caudal peduncle depth	27.748	0.000*
Caudal peduncle length	73.795	0.000*
Dorsal fin base length	12.627	0.000*
Anal fin base length	13.838	0.000*
Pelvic fin length	94.848	0.000*
Pectoral fin length	69.589	0.000*
Upper caudal lobe length	68.202	0.000*
Median caudal lobe length	27.334	0.000*
Lower caudal lobe length	81.989	0.000*
Dorso-hypural distance	128.055	0.000*
Eye diameter	21.879	0.000*
Snout length	128.436	0.000*
Head width	15.980	0.000*
Head depth	57.256	0.000*
Inter-orbital width	50.390	0.000*
Anal depth	64.848	0.000*

*indicate significantly difference at significance level of 0.05

Discriminant function analysis (DFA) was performed on the significance character from ANOVA. The Wilk's lambda value, eigenvalues, cumulative percentage, total variance percentage, canonical correlation, and species classification were generated from this analysis. Table 4.3 shows the Wilk's lambda values of 24 morphometric characters. A Wilk's lambda is used to test the significant differences between species. A highly significant inter-sample variation ($p < 0.05$) between all samples was discovered. The smallest values of wilk's lambda was showed by total length and pre-dorsal length characters with value of 0.120. These 2 characters were mostly contribute to discriminate the species.

Table 4.3: Wilk's lambda values of 24 morphometric characters

Characters	Wilk's Lambda	F value	Significance value
Total length	0.120	132.545	0.000
Head length	0.151	101.089	0.000
Pre-dorsal length	0.120	131.773	0.000
Pre-anal length	0.121	130.280	0.000
Pre-pelvic length	0.129	121.621	0.000
Dorsal depth	0.171	87.512	0.000
Body depth	0.257	51.989	0.000
Caudal peduncle depth	0.393	27.748	0.000
Caudal peduncle length	0.196	73.795	0.000
Dorsal fin base length	0.588	12.627	0.000
Anal fin base length	0.565	13.838	0.000
Pelvic fin length	0.160	94.848	0.000
Pectoral fin length	0.206	69.589	0.000
Upper caudal lobe length	0.209	68.202	0.000
Median caudal lobe length	0.397	27.334	0.000
Lower caudal lobe length	0.180	81.989	0.000
Dorso-hypural distance	0.123	128.055	0.000
Eye diameter	0.451	21.879	0.000
Snout length	0.123	128.436	0.000
Head width	0.530	15.980	0.000
Head depth	0.239	57.256	0.000
Inter-orbital width	0.263	50.390	0.000
Anal depth	0.217	64.848	0.000

In table 4.4, three functions were generated from discriminant function analysis with eigenvalues of 69.549, 6.318, and 2.977 respectively. The first function contribute most of the differentiating dimension. Percentage of variance is the proportion of discriminating ability of 4 *Rasbora* species which acts as variables found in a given function. In this analysis, function 1 accounts for 88.2% of discriminating ability of discriminating variables and followed by function 2 (8.0%) and function 3 (3.8%). The association of discriminant function and dependent variable was measured by canonical correlation. In function 1 and function 2 respectively pre-dorsal length (7.568) and total length (10.188) were the character that major contribute to discriminant. In function 3, the 3.8% discriminant ability was contributed by total length (13.780) and pre-dorsal length (13.032) characters.

Table 4.4: Eigenvalues, percentage of variance, cumulative percentage, canonical correlation, and standardized coefficients of morphometric characters

Function	1	2	3
Eigenvalues	69.549	6.318	2.977
Variance (%)	88.2	8.0	3.8
Cumulative (%)	88.2	96.2	100.0
Canonical correlation	0.993	0.929	0.865
Standardized coefficients;			
Total length	0.460	10.188	13.780
Head length	-1.449	1.893	-3.068
Pre-dorsal length	-7.568	-3.814	-13.032
Pre-anal length	5.040	-8.533	3.223
Pre-pelvic length	0.686	6.937	1.648
Dorsal depth	-1.623	-0.967	0.162
Body depth	-0.725	-1.716	0.154
Caudal peduncle depth	1.083	2.726	0.624
Caudal peduncle length	-0.458	-1.144	0.694
Dorsal fin base length	-0.166	0.596	0.138
Anal fin base length	0.820	-0.298	0.132
Pelvic fin length	-0.499	0.306	0.349
Pectoral fin length	1.959	-3.294	-0.791
Upper caudal lobe length	3.755	0.428	-1.394
Median caudal lobe length	-0.247	0.871	0.769
Lower caudal lobe length	-2.757	-2.602	1.026
Dorso-hypural distance	0.683	-0.807	-4.076
Eye diameter	1.119	0.517	0.178
Snout length	1.033	0.055	-0.059
Head depth	-0.104	-0.022	0.038

Table 4.5 display the structure matrix of ordered correlation morphometric characters by its absolute size within function. Snout length (0.312) and inter-orbital width (0.197) were the variables that mostly contribute in function 1. Anal depth (0.486), dorso-hypural distance (0.435), head length (0.426), pre-pelvic length (0.418), pre-dorsal length (0.407), pre-anal length (0.407), total length (0.405), eye diameter (0.403), caudal peduncle depth (0.402), lower caudal lobe length (0.396), pelvic fin length (0.392), head depth (0.386), upper caudal lobe length (0.382), median caudal lobe length (0.376), dorsal depth (0.365), body depth (0.357), caudal peduncle length (0.346), head width (0.315), pectoral fin length (0.309), anal fin base length (0.270), and dorsal fin base length (0.189) were variables that mostly contribute to function 2. There are no obvious characters that mostly contribute to function 3.

Table 4.5: Structure matrix of ordered correlation morphometric characters by its absolute size within function

Function	1	2	3
Snout length	0.312*	0.183	-0.234
Inter-orbital width	0.197*	-0.022	-0.052
Anal depth	-0.183	0.486*	-0.199
Dorso hypural distance	-0.284	0.435*	-0.323
Head length	-0.244	0.426*	-0.326
Pre-pelvic length	-0.279	0.418*	-0.289
Pre-dorsal length	-0.293	0.407*	-0.310
Pre-anal length	-0.293	0.407*	-0.285
Total length	-0.297	0.405*	-0.263
Eye diameter	-0.009	0.403*	-0.248
Caudal peduncle depth	-0.069	0.402*	-0.252
Lower caudal lobe length	-0.221	0.396*	-0.231
Pelvic fin length	-0.244	0.392*	-0.231
Head depth	-0.176	0.386*	-0.163
Upper caudal lobe length	-0.198	0.382*	-0.219
Median caudal lobe length	-0.095	0.376*	-0.030
Dorsal depth	-0.237	0.365*	-0.187
Body depth	-0.167	0.357*	-0.215
Caudal peduncle length	-0.208	0.346*	-0.341
Head width	0.016	0.315*	-0.179
Pectoral fin length	-0.209	0.309*	-0.276
Anal fin base length	-0.063	0.270*	-0.105
Dorsal fin base length	-0.082	0.189*	-0.061

*indicates the largest absolute correlation between each variable and its respective discriminant function.

All the samples were successfully separated by discriminant function analysis (DFA) into their respective species. As in Figure 4.2, *Rasbora borapetensis* samples were distantly separated from other three *rasbora* species. Meanwhile *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora tornieri* were clustered separately with each other. Two species which are *Rasbora sumatrana* and *Rasbora hobelmani* were closely clustered with each other. The prediction group membership of *Rasbora species* classification (Table 4.6) also shows that 2 samples of *Rasbora sumatrana* were counted as *Rasbora hobelmani* by discriminant function analysis. This shows that these 2 *Rasbora* species were closely related.

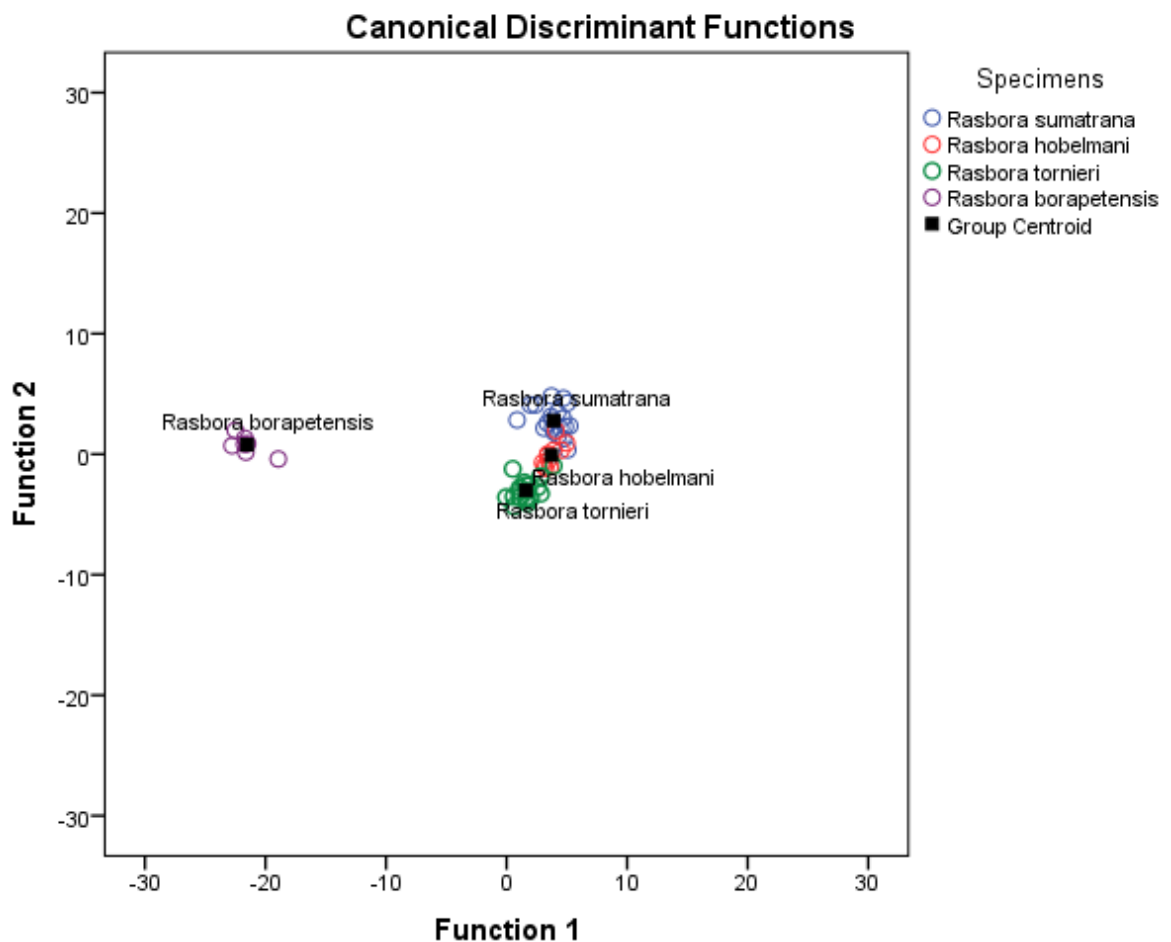


Figure 4.2: Canonical discriminant function that classify *Rasbora* samples into its respective species based on morphometric analysis.

Table 4.6 displays the classification of *Rasbora* species with their predicted group membership by discriminant function analysis (DFA). Samples of *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* were 100% classified into their respective *Rasbora* species meanwhile samples of *Rasbora sumatrana* was 90% classified as *Rasbora sumatrana*. About 10% of *Rasbora sumatrana* samples were classified as *Rasbora hobelmani*. Thus, 96.6% of original grouped cases were correctly classified in discriminant function analysis by SPSS.

Table 4.6: The classification of *Rasbora* species by DFA on morphometric measurements with their predicted group membership

Group cases		<i>Rasbora sumatrana</i>	<i>Rasbora hobelmani</i>	<i>Rasbora tornieri</i>	<i>Rasbora borapetensis</i>	Total	
Original	Count	<i>Rasbora sumatrana</i>	18	2	0	0	20
		<i>Rasbora hobelmani</i>	0	11	0	0	11
		<i>Rasbora tornieri</i>	0	0	20	0	20
		<i>Rasbora borapetensis</i>	0	0	0	7	7
	%	<i>Rasbora sumatrana</i>	90.0	10.0	.0	.0	100.0
		<i>Rasbora hobelmani</i>	.0	100.0	.0	.0	100.0
		<i>Rasbora tornieri</i>	.0	.0	100.0	.0	100.0
		<i>Rasbora borapetensis</i>	0	.0	.0	100.0	100.0

96.6% of original grouped cases correctly classified.

4.3 Meristic analysis

A total of 58 individuals were counted for 11 traits of meristic. Descriptive data in Table 4.7 shows the range (Minimum and Maximum) and mode of meristic counts for 4 *rasbora* species. Unlike morphometric analysis, the meristic does not require size correction.

From Table 4.7, all 4 *rasbora* species mostly have a similar dorsal fin count with mode of 8 rays. The pectoral fin of *Rasbora sumatrana* is frequently 15 rays (12-15 rays) and reaching the first ray of pelvic fin. *Rasbora hobelmani* (13-14 rays) and *Rasbora tornieri* (13-16 rays) are frequently 14 rays of pectoral fin and both pectoral fin are reaching the pelvic fin. *Rasbora borapetensis* is mostly 12 rays (12-15 rays) of pectoral fin and does not reaching the pelvic fin. The pelvic fin of 4 *rasbora* species are frequently 9 rays. Tip of appressed pelvic fin of *Rasbora sumatrana* and *Rasbora hobelmani* is about reaching the anus. Meanwhile the pelvic fin of *Rasbora tornieri* and *Rasbora borapetensis* are not reaching the anal opening.

The lateral line scale of *Rasbora hobelmani* is frequently 26 (25-29 scales) which is lesser compared to *Rasbora sumatrana*; 28 (27-32 scales), *Rasbora tornieri*; 33 and 34 (32-35 scales), and *Rasbora borapetensis*; 27, 29 and 31 (27-31 scales). The pre-dorsal scale of *Rasbora sumatrana* is lesser with mode of 12 (11-13 scales) followed by *Rasbora tornieri*; 13 (12-15 scales), *Rasbora borapetensis*; 13 scales, and *Rasbora hobelmani*; 14 (10-14 scales). *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora tornieri* have a similar mode of circumpeduncular scale which is 12 scales while *Rasbora borapetensis* is 10 scales.

Table 4.7: Descriptive statistics for meristic traits of *Rasbora* species

Meristic traits	<i>Rasbora sumatrana</i> N=20		<i>Rasbora hobelmani</i> N=11		<i>Rasbora torneri</i> N=20		<i>Rasbora borapetensis</i> N=7	
	Range	Mode (n)	Range	Mode (n)	Range	Mode (n)	Range	Mode (n)
Dorsal Fin	7-8	8 (19)	8	8 (11)	7-8	8 (18)	8-10	8 (6)
Pectoral Fin	12-15	15 (11)	13-14	14 (8)	13-16	14 (7)	12-15	12 (4)
Pelvic Fin	8-9	9 (16)	8-9	9 (9)	8-9	9 (19)	7-9	9 (3)
Anal Fin	6	6 (20)	6	6 (11)	6-7	6(17)	6-7	6 (4)
Caudal Fin	19	19 (20)	19	19 (11)	19	19 (20)	18-19	19 (5)
Dorsal Pro-current	3-6	4 (12)	4-7	5 (5)	3-6	5 (11)	5-6	6 (6)
Ventral Pro-current	4-6	4 (9)	4-7	5 (5)	4-6	6 (11)	5-7	5 (3), 6 (3)
Scales of Lateral Line	27-32	28 (8)	25-29	26 (5)	32-35	33 (9), 34 (9)	27-31	27 (2), 29 (2) ,31 (2)
Pre-dorsal Scale	11-13	12 (12)	10-14	14 (4)	12-15	13 (13)	13	13 (7)
Transverse Scale	6-8	7 (11)	6-7	6 (6)	7-8	7 (13)	7-8	8 (4)
Circumpeduncular Scale	10-13	12 (11)	10-12	12 (7)	11-14	12 (12)	9-11	10 (3)

Values in this table represent range in counts and mode (total samples with similar mode). N represent number of individuals sampled.

The non-parametric test which is Kruskal-Wallis test was performed on the meristic traits data using SPSS version 22 software. Among the 11 meristic traits, only one traits is not significance which is dorsal fins with its significance level of $p = 0.203$. The distribution of dorsal fins across 4 *rasbora* species is the same. The remaining 10 traits; pectoral fins, pelvic fins, anal fins, caudal fins, dorso pro-current, ventral pro-current, lateral line scales, pre-dorsal scales, transverse scales, and circumpeduncular scales are significantly different (refer to Table 4.8). This indicated that differences are mostly occurred in the fins and scales. Study of *Zenarchopterus buffonis* by Syaifullah *et al.* (2015) also found that most differences are occurred in fins organ.

Table 4.8: Kruskal-Wallis Test (non-parametric test) for comparison of meristic counts among the *Rasbora* species

No	Meristic traits	Chi-square	Degree of freedom	Significance ($p \leq 0.05$)
1	Dorsal fins	4.602	3	0.203
2	Pectoral fins	13.946	3	0.003*
3	Pelvic fins	10.260	3	0.016*
4	Anal fins	11.815	3	0.008*
5	Caudal fins	14.832	3	0.002*
6	Dorso pro-current	17.524	3	0.001*
7	Ventral pro-current	9.201	3	0.027*
8	Scales of lateral line	46.253	3	0.000*
9	Pre-dorsal scale	25.289	3	0.000*
10	Transverse scale	16.945	3	0.001*
11	Circumpeduncular scale	19.637	3	0.000*

*indicates the significance ($p \leq 0.05$)

Multivariate data analysis which is discriminant function analysis (DFA) was utilized in the meristic study. The eigenvalues, cumulative percentage, total variance percentage, and canonical correlation were generated from this analysis. Multivariate ANOVA (MANOVA) of meristic traits shows that 10 out of 11 traits are significantly different with level of 0.05 (Table 4.9). Only dorsal fin traits ($p=0.083$) is not significantly different. The smallest wilk's lambda value was showed by lateral line scales (0.151) that major contribute to species discriminant.

Table 4.9: Values of Wilks's lambda and its significance for 11 meristic traits

Meristic traits	Wilks' Lambda	F value	Significance value
Dorsal fin	0.885	2.343	0.083
Pectoral fins	0.716	7.155	0.000*
Pelvic fins	0.756	5.822	0.002*
Anal fins	0.793	4.707	0.005*
Caudal fin	0.740	6.331	0.001*
Dorso pro-current	0.700	7.709	0.000*
Ventral pro-current	0.839	3.465	0.022*
Scale of lateral line	0.151	101.066	0.000*
Pre-dorsal scale	0.614	11.303	0.000*
Transverse scale	0.697	7.808	0.000*
Circumpeduncular scale	0.584	12.814	0.000*

*the significantly different at $p<0.05$

The pooled within group correlation matrix of discriminant function analysis among the 11 meristic traits was shown in Table 4.10. The pairwise correlations among these traits were low which below than 0.5. Therefore, all the characters were retained for the rest of the analysis.

Table 4.10: The observed pooled within group correlation matrix among 11 traits of meristic data

	DF	PcF	PvF	AF	CF	DrPc	VrPc	SLL	PrDS	TrvS	CcpS
DF	1.000										
PcF	0.283	1.000									
PvF	-0.319	0.011	1.000								
AF	0.086	0.245	0.107	1.000							
CF	-0.481	-0.035	0.073	-0.058	1.000						
DrPc	-0.142	-0.025	-0.109	0.108	-0.045	1.000					
VrPc	-0.128	0.001	-0.163	-0.028	-0.081	0.497	1.000				
SLL	0.198	0.103	0.042	0.218	-0.219	0.187	-0.069	1.000			
PrDS	-0.104	0.082	-0.178	0.086	0.000	0.317	0.251	0.328	1.000		
TrvS	0.143	-0.184	0.108	-0.087	-0.169	-0.084	-0.094	0.209	-0.060	1.000	
CcpS	-0.025	0.085	0.130	-0.013	0.302	-0.002	-0.006	0.184	0.064	0.063	1.000

Traits code: DF; dorsal fin, PcF; pectoral fins, PvF; pelvic fins, AF; anal fin, CF; caudal fin, DrPc; dorso pro-current, VrPc; ventral pro-current, SLL; scale of lateral line, PrDS; pre-dorsal scale, TrvS; transverse scale, CcpS; circumpeduncular scale

From Table 4.11, three discriminant functions were generated. The highest eigenvalues was in function 1 which is 7.845 and followed by function 2 (1.757) and function 3 (0.445). The variance of each function was 78.1% (Function 1), 17.5% (Function 2), and 4.4% (Function 3) respectively. In function 1, function 2, and function 3, the characters respectively lateral line scale (1.132), anal fin (0.435), and pre-dorsal scales (0.865) were contributes the most in discriminating the species.

Table 4.11: Eigenvalues, percentage of variance, cumulative percentage, canonical correlation, and standardized coefficients of meristic traits

Function	1	2	3
Eigenvalues	7.845	1.757	0.445
Variance (%)	78.1	17.5	4.4
Cumulative (%)	78.1	95.6	100.0
Canonical correlation	0.942	0.798	0.555
Standardized Coefficients;			
Dorsal fin	-0.299	0.150	0.506
Pectoral fins	0.213	-0.351	-0.254
Pelvic fins	-0.051	-0.243	0.587
Anal fin	-0.159	0.435	-0.058
Caudal fin	0.216	-0.073	0.182
Dorso pro-current	-0.392	0.267	0.115
Ventral pro-current	0.336	0.121	0.192
Scales of lateral line	1.132	0.123	-0.152
Pre-dorsal scale	-0.211	0.183	0.865
Transverse scale	0.017	0.374	-0.293
Circumpeduncular scale	-0.093	-0.431	0.239

In Table 4.12, the structure matrix of ordered correlation variables was shown by its absolute size within the function. Lateral line scales variable with its absolute value of 0.832 was mostly contribute in function 1. Circumpeduncular scale (0.468), dorso pro-current (0.442), caudal fin (0.418), anal fin (0.381), transverse scale (0.362), pelvic fins (0.345), pectoral fins (0.287), and dorsal fin (0.217) variables were mostly contribute in function 2. Pro-dorsal scales (0.750) and ventral pro-current (0.329) variables were mostly contribute to function 3.

Table 4.12: Structure matrix of ordered correlation variables by its absolute size within function

Function	1	2	3
Scale of lateral line	0.832*	0.318	0.168
Circumpeduncular scale	0.187	-0.468*	0.344
Dorso pro-current	-0.066	0.442*	0.336
Caudal fin	0.068	-0.418*	0.128
Anal fin	0.026	0.381*	0.047
Transverse scale	0.151	0.362*	-0.233
Pelvic fins	0.108	-0.345*	0.225
Pectoral fins	0.180	-0.287*	0.001
Dorsal fin	-0.076	0.217*	-0.054
Pre-dorsal scale	0.157	0.325	0.750*
Ventral pro-current	0.043	0.272	0.329*

*indicates the largest absolute correlation between each variable and its respective discriminant function.

The discriminant function analysis (DFA) on meristic data has successfully separated the *Rasbora* samples into their respective species (Figure 4.3). The clustering of each species was clear and not too close to each other. The group centroid (black square symbol) of these 4 species were distant.

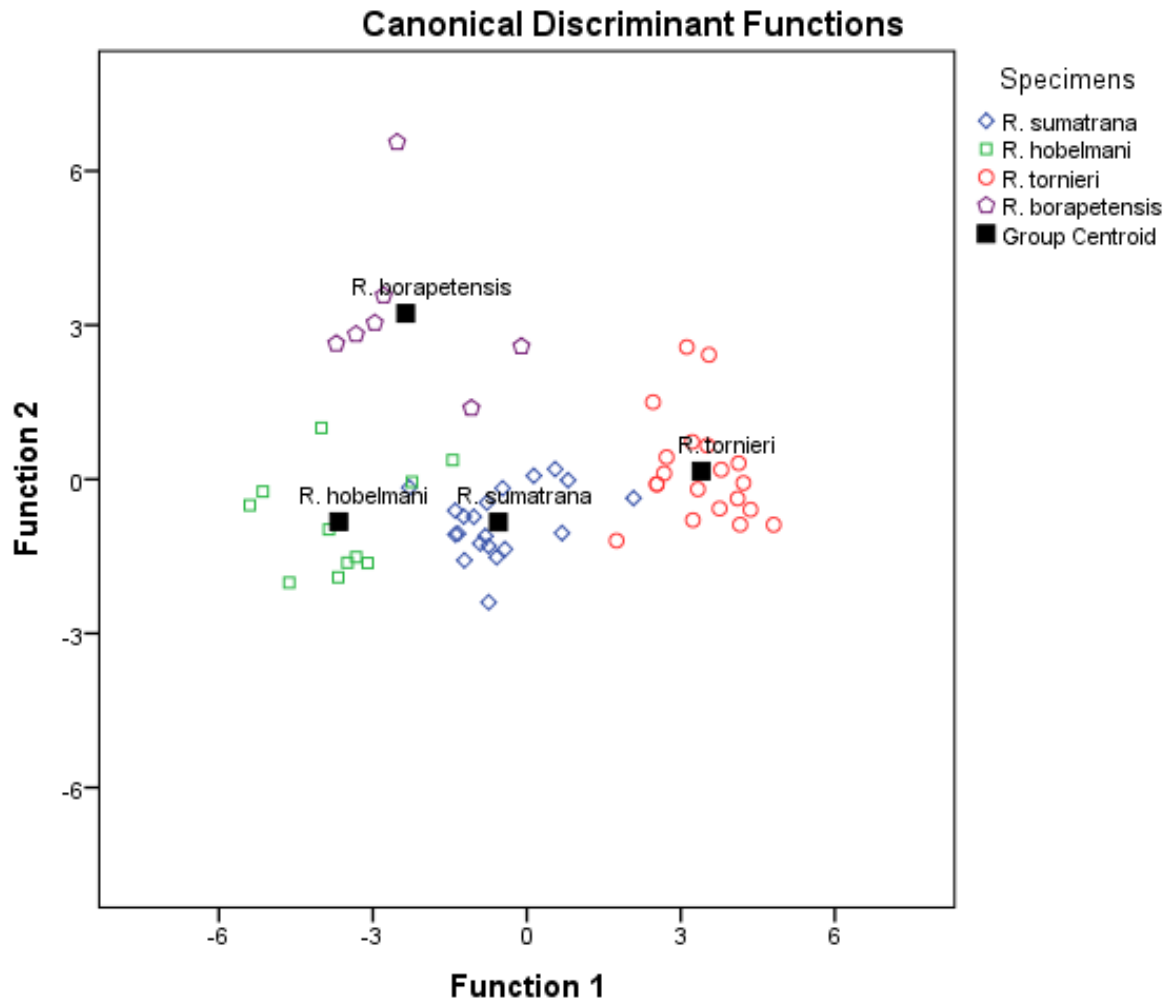


Figure 4.3: Canonical discriminant function graph that separate the specimen by species based on meristic analysis

Table 4.13 shows the classification of *Rasbora* species with their predicted group membership. One sample from *Rasbora borapetensis* were counted as *Rasbora sumatrana*. One sample from *Rasbora sumatrana* also counted as *Rasbora tornieri*. Samples of *Rasbora hobelmani* and *Rasbora tornieri* were 100% classified into their respective *Rasbora* species. A total of 95% *Rasbora sumatrana* were correctly classified as *Rasbora sumatrana* and the remaining of 5% were classified as *Rasbora tornieri*. *Rasbora borapetensis* was 85.7% was classified as *Rasbora borapetensis* and the remaining 14.3% of *Rasbora borapetensis* were classified as *Rasbora sumatrana*. Overall, 96.6% of original grouped cases were correctly classified by discriminant function analysis in SPSS.

Table 4.13: The classification of *Rasbora* species by DFA on meristic counting with their predicted group membership

Group cases		<i>Rasbora sumatrana</i>	<i>Rasbora hobelmani</i>	<i>Rasbora tornieri</i>	<i>Rasbora borapetensis</i>	Total	
Original	Count	<i>Rasbora sumatrana</i>	19	0	1	0	20
		<i>Rasbora hobelmani</i>	0	11	0	0	11
		<i>Rasbora tornieri</i>	0	0	20	0	20
		<i>Rasbora borapetensis</i>	1	0	0	6	7
	%	<i>Rasbora sumatrana</i>	95.0	.0	5.0	.0	100.0
		<i>Rasbora hobelmani</i>	.0	100.0	.0	.0	100.0
		<i>Rasbora tornieri</i>	.0	.0	100.0	.0	100.0
		<i>Rasbora borapetensis</i>	14.3	.0	.0	85.7	100.0

96.6% of original grouped cases correctly classified.

4.4 Complete fragment of mitochondrial genome analysis

4.4.1 Annealing temperature optimization of pair primers

Four sets of primer were designed and used in this study. These primers were tRNA his F1, ND5 R1, tRNA gln F2, tRNA asn R2, tRNA phe F3, 16s rRNA R3, CO1 F4, and tRNA lys R4. Each sets of these primer are for partial fragment of mitochondrial genome. So, a new paired of primer are made from these primers. The forward and reverse from each set of primer were paired with another set of primer. The new paired of primer were used to study the complete fragments (small and large fragments) of mitochondrial genome for every *Rasbora* species. As the complete fragments were studied, thus 2 sets of primer as shown in Table 4.14 were used on every species where each set was respectively used for small and large fragments.

Small fragment primer set 1 and small fragment primer set 2 as shown in Table 4.14 were applied at the same time to amplify the DNA extraction of single *rasbora* sample from every studied species. In optimization, the annealing temperature range was between 48.6 °C – 56.0 °C. Large fragment from both sets were also amplified at the same time with annealing temperature range between 47.6 °C – 54.0 °C.

Table 4.14: Primer sets for partial and complete fragments.

Set primer for partial fragment				
	1	2	3	4
Forward	tRNA his F1	tRNA gln F2	tRNA phe F3	CO1 F4
Reverse	ND5 R1	tRNA asn R2	16s rRNA R3	tRNA lys R4
New paired primer set for complete fragments				
	1		2	
Fragment	Small	Large	Small	Large
Forward	tRNA gln F2	tRNA his F1	tRNA phe F3	CO1 F4
Reverse	ND5 R1	tRNA asn R2	tRNA lys R4	16s rRNA R3
Expected size	8,923 bp	9,990 bp	7,866 bp	11,004 bp

From Figure 4.4, results from gel electrophoresis on lane 1 (*Rasbora sumatrana*), lane 2 (*Rasbora hobelmani*), and lane 3 (*Rasbora tornieri*) shows negative amplification results for small fragment from primer set 1. However, positive results which a single discrete band was shown on lane 4 (*Rasbora sumatrana*), lane 5 (*Rasbora hobelmani*), and lane 6 (*Rasbora tornieri*) for small fragment from primer set 2 at same temperature of 48.6 °C. The negative results were optimized with another temperatures. The positive results for small fragment of set primer 2 with expected size of 7,866 bp at temperature of 48.6 °C for *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora tornieri* were proceed to PCR purification for sequencing.

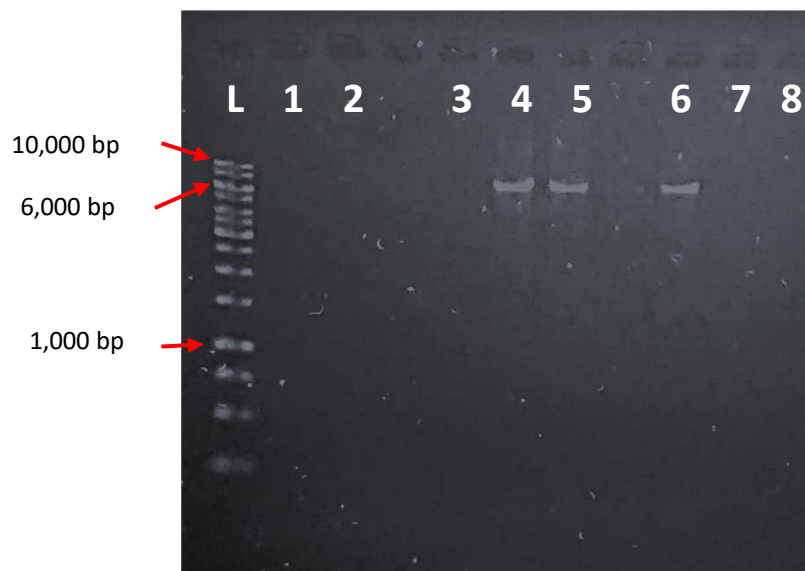


Figure 4.4: Gel electrophoresis results from gradient PCR of small fragment from both set complete fragments primer. Lane L represents 1 kb ladder, Lane 1, 2, and 3 represents PCR amplicon from primer set 1, Lane 4, 5, and 6 represents PCR amplicon from primer set 2 at the same temperature of 48.6 °C, and Lane 7 and 8 represents negative control from both primer sets.

Gel electrophoresis results as shown in Figure 4.5 shows the positive results of PCR amplicon with expected size of 8,923 bp from set primer 1 for small fragment of *Rasbora borapentensis* (lane 1 and lane 2) at temperature of 55.3 °C. The PCR product for small fragment of *Rasbora borapentensis* was proceed to PCR purification for sequencing.

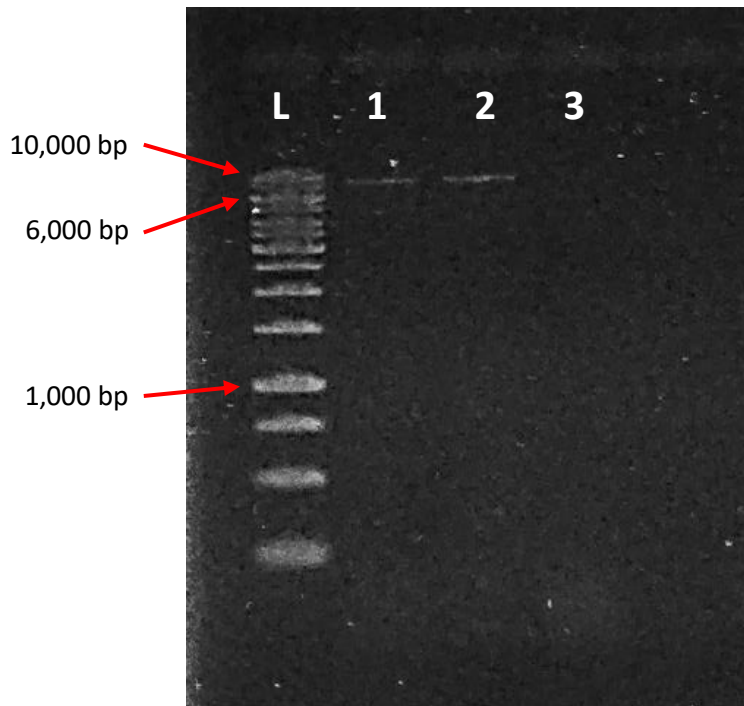


Figure 4.5: Gel electrophoresis results from gradient PCR of small fragment from set primer 1. Lane L represents ladder of 1 kb. Lane 3 represents negative control. Lane 1 and lane 2 represents PCR amplicon from *Rasbora borapetensis* sample.

Figure 4.6 shows the gel electrophoresis results from gradient PCR of same set of primer which is primer 2 for large fragment of *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora tornieri*. The expected amplicon size of this set primer 2 for large fragment is 11,004 bp. The PCR was successfully amplified the amplicon for *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora tornieri* at temperature of 47.6 °C, 47.5 °C, and 50.7 °C respectively. Meanwhile, set primer 1 successfully amplified large fragment of *Rasbora borapetensis* at temperature of 50.0 °C with expected amplicon size of 9,990 bp. Then, the PCR purification were proceed for sequencing.

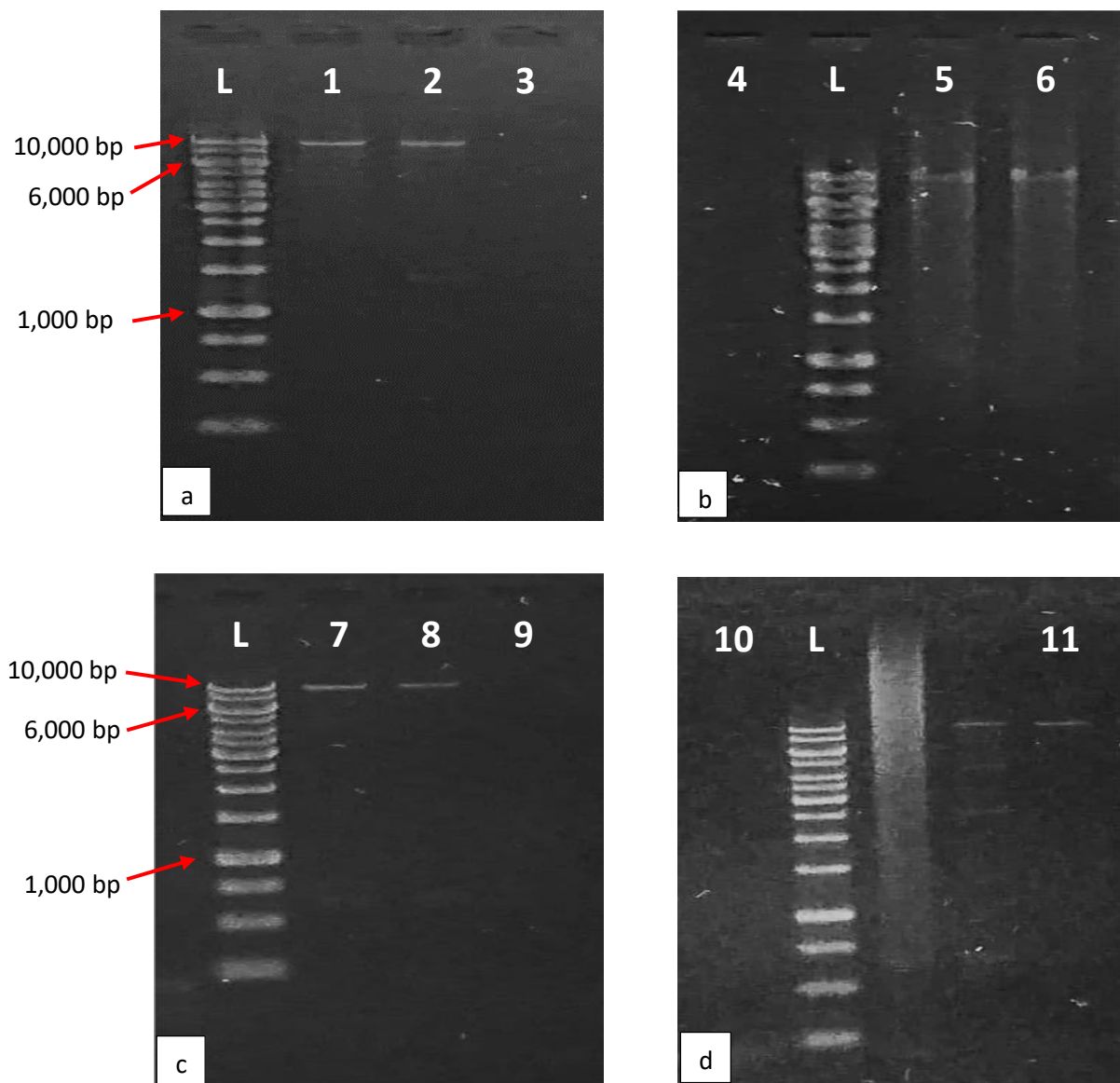


Figure 4.6: Gel electrophoresis results from gradient PCR for large fragment. Lane L represents 1kb ladder. Negative control was represents by lane 3, 4, 9, and 10. a) Lane 1 and lane 2 represents *Rasbora sumatrana* at temperature of 47.6 °C. b) Lane 5 and lane 6 represents *Rasbora hobelmani* at temperature of 47.5 °C. c) Lane 7 and 8 represents *Rasbora tornieri* at temperature of 50.7 °C. d) Lane 11 represents *Rasbora borapetensis* at temperature of 50.0 °C.

4.4.2 Sequencing analysis

The mitochondrial DNA (mtDNA) of four *Rasbora* samples were sent to University of Hong Kong for sequencing. All samples were consists of two fragments which are small and large fragments for sequencing. Table 4.15 shows the characteristics of forward and reverse primer for small and large fragments of each *Rasbora* species. Sample *Rasbora sumatrana*,

Rasbora hobelmani, and *Rasbora tornieri* were amplified using same set of primer (primer 2: tRNA phe F3, tRNA lys R4, CO1 F4, and 16s rRNA R3) while sample *Rasbora borapetensis* was amplified using primer 1 (tRNA gln F2, ND5 R1, tRNA his F1, and tRNA asn R2).

Table 4.15: Forward and reverse primer characteristics for fragments of each *Rasbora* species.

Sample	Fragment	Primer name	Sequence (5' - 3')	Size (bp)	Tm (°C)
<i>Rasbora sumatrana</i>	Small	tRNA phe F3	CAAAGGCATGGTCCCAAC	7866	48.6
		tRNA lys R4	AGGTGGTCGTTAATCAC		
	Large	CO1 F4	ATTCGCCATTATAGCAGG	11004	47.6
		16s rRNA R3	GTTTGGCTCAGAGTGATC		
<i>Rasbora hobelmani</i>	Small	tRNA phe F3	CAAAGGCATGGTCCCAAC	7866	48.6
		tRNA lys R4	AGGTGGTCGTTAATCAC		
	Large	CO1 F4	ATTCGCCATTATAGCAGG	11004	47.5
		16s rRNA R3	GTTTGGCTCAGAGTGATC		
<i>Rasbora tornieri</i>	Small	tRNA phe F3	CAAAGGCATGGTCCCAAC	7866	48.6
		tRNA lys R4	AGGTGGTCGTTAATCAC		
	Large	CO1 F4	ATTCGCCATTATAGCAGG	11004	50.7
		16s rRNA R3	GTTTGGCTCAGAGTGATC		
<i>Rasbora borapetensis</i>	Small	tRNA gln F2	GTGCTTCCTCTACACCAC	8923	55.3
		ND5 R1	TGATGTTGAGAAGGC TAC		
	Large	tRNA his F1	CCTATCTTACCGAGAAAG	9990	50.0
		tRNA asn R2	GAGGCCTTCCCATCTAGA		

Illumina Hiseq 400 System was used for DNA sequencing of small and large fragment. The small and large fragment were assembled to achieve full length of sequence. Then, each sequence of *Rasbora* sample was analysed individually and subjected to nucleotide BLAST (blastn) in NCBI website for species identification.

Table 4.16 shows the complete fragments description of each *Rasbora* species by nucleotide BLAST (blastn). These 4 species were identified by morphometry analysis before

proceed with mitochondrial genome study. Based on blastn, the *Rasbora sumatrana* (95%), *Rasbora hobelmani* (95%), and *Rasbora tornieri* (89%) were similar to *Rasbora sumatrana* with an accession number of AP011221.1. Meanwhile, the identified *Rasbora borapetensis* was 99% similar to *Rasbora borapetensis* in the Genbank with an accession number of AB924546.1. This makes the remaining 5% of *Rasbora sumtarana* and *Rasbora hobelmani*, 11% of *Rasbora tornieri*, and 1% of *Rasbora borapetensis* were their own identities.

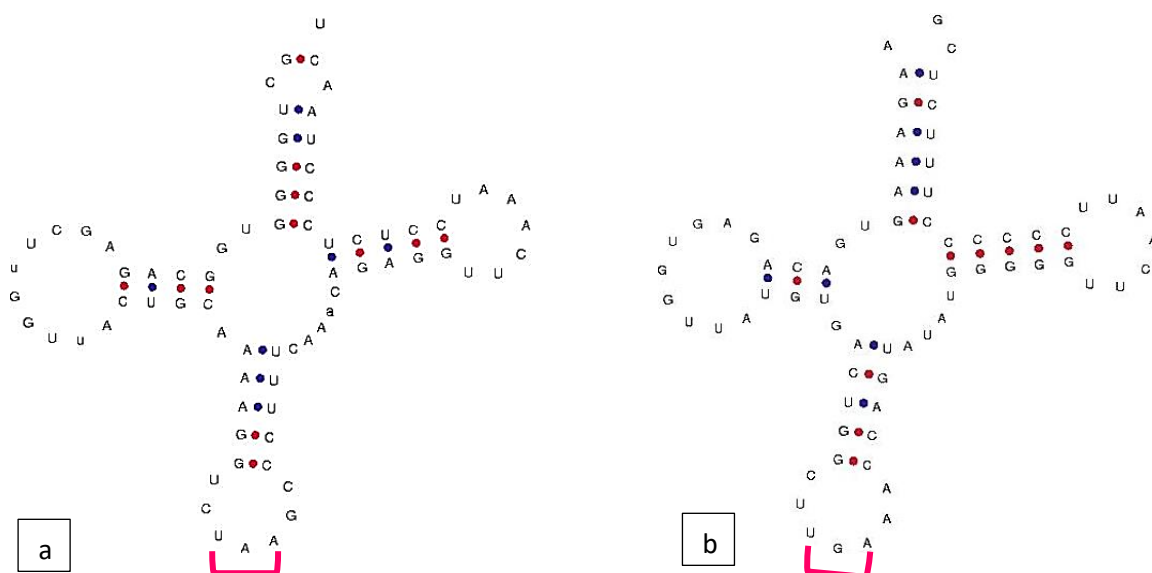
Table 4.16: Nucleotide BLAST description of *Rasbora* species complete fragments.

Description	<i>Rasbora sumatrana</i>	<i>Rasbora hobelmani</i>	<i>Rasbora tornieri</i>	<i>Rasbora borapetensis</i>
Length	16,539 bp	16,541 bp	16,573 bp	16,520 bp
Score	26,389 bits (29,266)	26,316 bits (29,184)	21,183 bits (23,492)	29,055 bits (32,222)
GC content	42 %	42 %	41 %	40 %
Query coverage	99 %	100 %	99 %	100 %
Expected value	0.0	0.0	0.0	0.0
Identities	95 % (15,787/16,543)	95 % (15,772/16,545)	89 % (14,690/16,596)	99 % (16,395/16,537)
Gaps	0 % (10/16,543)	0 % (10/16,545)	0 % (91/16,596)	0 % (44/16,537)
Similar species	<i>Rasbora sumatrana</i>	<i>Rasbora sumatrana</i>	<i>Rasbora sumatrana</i>	<i>Rasbora borapetensis</i>
(Accession number)	(AP011221.1)	(AP011221.1)	(AP011221.1)	(AB924546.1)

Gene classification for transfer RNA (tRNA) gene, protein coding gene, and ribosomal RNA (rRNA) gene were conducted by MitoFish server which is a server for mitochondrial genome database of fish. The arrangement of 22 transfer RNA (tRNA) genes and their respective anti codon were computed by using tRNA-Scan SE version 2.0 software. The

complete gene classification of mitochondrial DNA has been individually tabulated for *Rasbora sumatrana* (Table 4.17), *Rasbora hobelmani* (Table 4.18), *Rasbora tornieri* (Table 4.19), and *Rasbora borapetensis* (Table 4.20).

Based on Table 4.17, the complete mitochondrial DNA of *Rasbora sumatrana* consists of 22 tRNA genes, 13 protein coding genes, 2 rRNA genes, and D-loop region. A total of 14 tRNA genes which are phenylalanine (phe), valine (val), isoleucine (ile), methionine (met), tryptophan (trp), aspartic acid (asp), lysine (lys), glycine (gly), arginine (arg), histidine (his), serine (ser), threonine (thr), and two leucine (leu) encoded for heavy strands. Meanwhile, the other 8 tRNA genes which are glutamine (gln), alanine (ala), asparagine (asn), cysteine (cys), tyrosine (tyr), serine (ser), glutamic acid (glu), and proline (pro) encoded for light strands. A total of 4 tRNA genes which are leucine (leu), serine (ser), another serine (ser), and another leucine (leu) decoded for an atypical anticodon that are TAA, TGA, GCT, and TAG respectively and the stem-loop of secondary structures were illustrated in Figure 4.7.



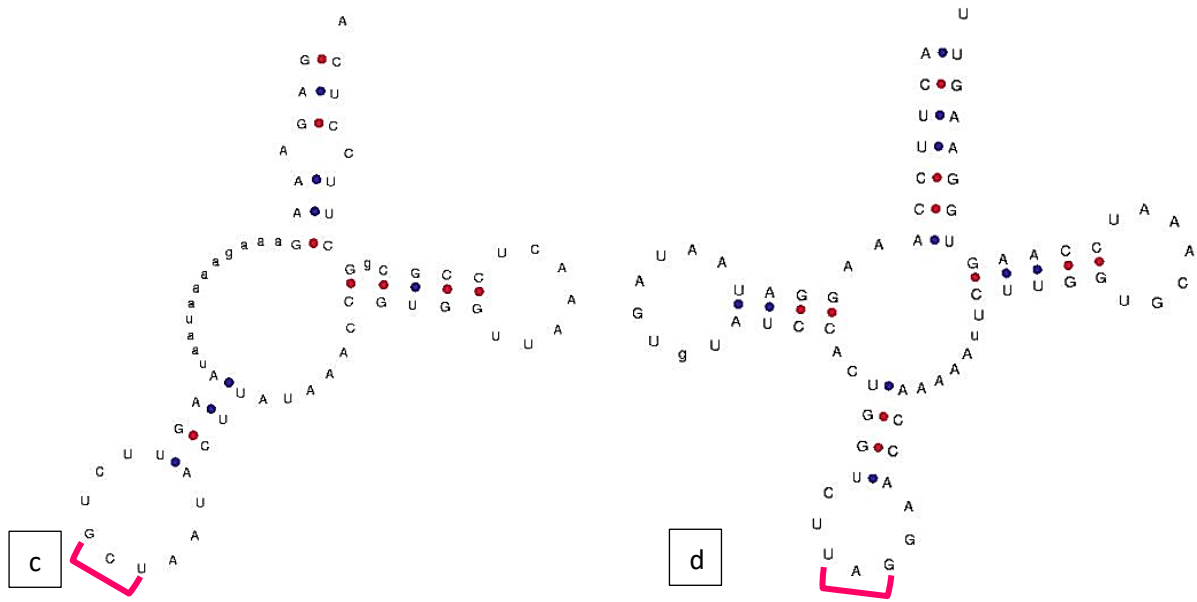


Figure 4.7: An atypical anticodon stem-loop of secondary structure in *Rasbora sumatrana* tRNA genes by tRNA scan S.E server. a) tRNA^{leu} (TAA); b) tRNA^{ser} (TGA); c) tRNA^{ser} (GCT); and d) tRNA^{leu} (TAG)

The big gap of nucleotides between asparagine (asn) and cysteine (cys) tRNA genes in Table 4.17 that are 34 bp nucleotides were the origin of light strand (O_L) replication. The 12S rRNA and 16S rRNA genes are encoded for heavy strands. A total of 12 protein coding genes which are ND1, ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND4L, ND4, ND5, and *Cytb* encoded for heavy strands. Only ND6 gene encoded for light strand. The start codon of protein coding genes in *Rasbora sumatrana* mitochondrial genome are ATG and GTG codons. Only COI gene started with GTG codon while the rest of protein coding genes initiated with ATG codon. A complete stop codon usage in *Rasbora sumatrana* protein coding genes are TAG (ND6) and TAA (ND1, COI, ATP8, ND4L, and ND5). An incomplete stop codons usage in *Rasbora sumatrana* protein coding genes are T (ND2, COII, ND3, and *Cytb*) and TA (ATP6, COIII, and ND4).

Table 4.17: Characteristics of complete mitochondrial genome of *Rasbora sumatrana*

Gene	Position (5' - 3')	Length (bp)	Start codon	Stop codon	Amino acid	Anti-codon	Intergenic nucleotides (bp)	Strand
tRNA ^{Phe}	1 - 70	70				GAA	0	H
12S rRNA	71 – 1,021	951					0	H
tRNA ^{Val}	1,022 – 1,092	71				TAC	0	H
16S rRNA	1,093 – 2,774	1,682					0	H
tRNA ^{Leu}	2,775 – 2,849	75				TAA	1	H
ND1	2,851 – 3,825	975	ATG	TAA	324		4	H
tRNA ^{Ile}	3,830 – 3,901	72				GAT	-2	H
tRNA ^{Gln}	3,970 – 3,900	71				TTG	1	L
tRNA ^{Met}	3,972 – 4,040	69				CAT	0	H
ND2	4,041 – 5,085	1,045	ATG	T- -	348		0	H
tRNA ^{Trp}	5,086 – 5,158	73				TCA	3	H
tRNA ^{Ala}	5,229 – 5,162	68				TGC	2	L
tRNA ^{Asn}	5,304 – 5,232	73				GTT	34	L
tRNA ^{Cys}	5,405 – 5,339	67				GCA	0	L
tRNA ^{Tyr}	5,476 – 5,406	71				GTA	1	L
COI	5,478 – 7,028	1,551	GTG	TAA	516		0	H
tRNA ^{Ser}	7,099 – 7,029	71				TGA	7	L
tRNA ^{Asp}	7,101 – 7,170	70				GTC	5	H

Table 4.17 continued

COII	7,176 – 7,866	691	ATG	T- -	230		0	H
tRNA ^{Lys}	7,867 – 7,941	75				TTT	2	H
ATP8	7,944 – 8,108	165	ATG	TAA	54		-7	H
ATP6	8,102 – 8,781	680	ATG	TA-	226		0	H
COIII	8,782 – 9,566	785	ATG	TA-	261		0	H
tRNA ^{Gly}	9,567 – 9,637	71				TCC	0	H
ND3	9,638 – 9,986	349	ATG	T- -	116		0	H
tRNA ^{Arg}	9,987 – 10,056	70				TCG	0	H
ND4L	10,057 – 10,353	297	ATG	TAA	98		-7	H
ND4	10,347 – 11,728	1,382	ATG	TA-	460		0	H
tRNA ^{His}	11,729 – 11,797	69				GTG	0	H
tRNA ^{Ser}	11,798 – 11,865	68				GCT	2	H
tRNA ^{Leu}	11,868 – 11,940	73				TAG	0	H
ND5	11,941 – 13,770	1,830	ATG	TAA	609		-4	H
ND6	14,288 – 13,767	522	ATG	TAG	173		0	L
tRNA ^{Glu}	14,357 – 14,289	69				TCC	9	L
Cytb	14,367 – 15,507	1,141	ATG	T- -	380		0	H
tRNA ^{Thr}	15,508 – 15,578	71				TGT	9	H
tRNA ^{Pro}	15,657 – 15,588	70				TGG	0	L
D-loop	15,658 – 16,539	882					-	-

The complete mitochondrial genome of *Rasbora hobelmani* consists of 22 tRNA genes, 13 protein coding genes, 2 rRNA genes, and D-loop region. A total of 14 tRNA genes were transcribed on heavy strand which includes phenylalanine (phe), valine (val), isoleucine (ile), methionine (met), tryptophan (trp), aspartic acid (asp), lysine (lys), glycine (gly), arginine (arg), histidine (his), serine (ser), threonine (thr), and two leucine (leu) genes. Meanwhile, the other 8 tRNA genes which are glutamine (gln), alanine (ala), asparagine (asn), cysteine (cys), tyrosine (tyr), serine (ser), glutamic acid (glu), and proline (pro) were transcribed on light strands. An atypical anticodon was found in 4 tRNA genes which are leucine (TAA), serine (TGA), another serine (GCT) which position at 11,798 - 11,865, and another leucine (TAG) gene at position of 11,868 – 11,940 as in Table 4.18.

The origin of light strand (O_L) replication were identified by 34 bp of intergenic nucleotides between an asparagine (asn) and cysteine (cys) tRNA genes (Table 4.18). The 12S rRNA and 16S rRNA genes are encoded for heavy strands. A total of 12 protein coding genes which are ND1, ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND4L, ND4, ND5, and Cytb encoded for heavy strands. Only ND6 gene encoded for light strand. ATG and GTG codons were identified as start codon in protein coding genes. Only COI gene was started with GTG codon while the rest of 12 protein coding genes initiated with ATG codon. A complete stop codon usage in *Rasbora hobelmani* protein coding genes are TAG (ND6) and TAA (ND1, COI, ATP8, ND4L, and ND5). An incomplete stop codons usage in *Rasbora hobelmani* protein coding genes are T (ND2, COII, ND3, and Cytb) and TA (ATP6, COIII, and ND4) as in Table 4.18. Protein coding genes that are ND4 and ATP6 both are overlaps 7 bp nucleotides with their adjacent genes which are ND4L and ATP8 respectively. Both distance tRNA^{Glu} and tRNA^{Thr} with their respectively adjacent genes; Cytb and Trna^{Pro} are 9 nucleotides in an intergenic spacer.

Table 4.18: Characteristics of complete mitochondrial genome of *Rasbora hobelmani*

Gene	Position (5'- 3')	Length (bp)	Start codon	Stop codon	Amino acid	Anti-codon	Intergenic nucleotides (bp)	Strand
tRNA ^{Phe}	1 – 69	69				GAA	0	H
12S rRNA	70 – 1,021	952					0	H
tRNA ^{Val}	1,022 – 1,092	71				TAC	0	H
16S rRNA	1,093 – 2,774	1,682					0	H
tRNA ^{Leu}	2,775 – 2,849	75				TAA	1	H
ND1	2,851 – 3,825	975	ATG	TAA	324		4	H
tRNA ^{Ile}	3,830 – 3,901	72				GAT	-2	H
tRNA ^{Gln}	3,970 – 3,900	71				TTG	1	L
tRNA ^{Met}	3,972 – 4040	69				CAT	0	H
ND2	4,041 – 5,085	1,045	ATG	T- -	348		0	H
tRNA ^{Trp}	5,086 – 5,158	73				TCA	3	H
tRNA ^{Ala}	5,229 – 5,162	68				TGC	2	L
tRNA ^{Asn}	5,304 – 5,232	73				GTT	34	L
tRNA ^{Cys}	5,405 – 5,339	67				GCA	0	L
tRNA ^{Tyr}	5,476 – 5,406	71				GTA	1	L
COI	5,478 – 7,028	1,551	GTG	TAA	516		0	H
tRNA ^{Ser}	7,099 – 7,029	71				TGA	1	L
tRNA ^{Asp}	7,101 – 7,170	70				GTG	5	H

Table 4.18 continued

COII	7,176 – 7,866	691	ATG	T- -	230		0	H
tRNA ^{Lys}	7,867 – 7,941	75				TTT	2	H
ATP8	7,944 – 8,108	165	ATG	TAA	54		-7	H
ATP6	8,102 – 8,781	680	ATG	TA-	226		0	H
COIII	8,782 – 9,566	785	ATG	TA-	261		0	H
tRNA ^{Gly}	9,567 – 9,637	71				TCC	0	H
ND3	9,638 – 9,986	349	ATG	T- -	116		0	H
tRNA ^{Arg}	9,987 – 10,056	70				TCG	0	H
ND4L	10,057 – 10,353	297	ATG	TAA	98		-7	H
ND4	10,347 – 11,728	1,382	ATG	TA-	460		0	H
tRNA ^{His}	11,729 – 11,797	69				GTG	0	H
tRNA ^{Ser}	11,798 – 11,865	68				GCT	2	H
tRNA ^{Leu}	11,868 – 11,940	73				TAG	0	H
ND5	11,941 – 13,770	1,830	ATG	TAA	609		-4	H
ND6	14,288 – 13,767	522	ATG	TAG	173		0	L
tRNA ^{Glu}	14,357 – 14,289	69				TTC	9	L
Cytb	14,367 – 15,507	1,141	ATG	T- -	380		0	H
tRNA ^{Thr}	15,508 – 15,578	71				TGT	9	H
tRNA ^{Pro}	15,657 – 15,588	70				TGG	0	L
D-loop	15,658 – 16,541	884					-	-

Table 4.19 show the characteristics of complete mitochondrial DNA of *Rasbora tornieri* which consists of 22 tRNA genes, 2 rRNA genes, 13 protein coding genes, and D-loop region. Phenylalanine (phe), valine (val), isoleucine (ile), methionine (met), tryptophan (trp), aspartic acid (asp), lysine (lys), glycine (gly), arginine (arg), histidine (his), serine (ser), threonine (thr), and two leucine (leu) of tRNA genes were transcribed on heavy strand. The remaining tRNA genes; Glutamine (gln), alanine (ala), asparagine (asn), cysteine (cys), tyrosine (tyr), serine (ser), glutamic acid (glu), and proline (pro) were transcribed on light strand. An atypical anticodon was found in 4 tRNA genes which are leucine (TAA), serine (TGA), another serine (GCT) which position at 11,809 - 11,876, and another leucine (TAG) gene at position of 11,879 – 11,951 as in Table 4.19. Isoleucine (ile) gene were overlap 2 nucleotides with glutamine (gln) gene. Threonine (thr) gene were 14 nucleotides away from proline (pro) gene. Asparagine (asn) and cysteine (cys) gene were distantly apart from each other by 37 nucleotides in the intergenic space. This big distance in intergenic space were the origin of light strand (O_L) replication.

The 12S rRNA and 16S rRNA genes are encoded for heavy strands. A total of 12 protein coding genes which are ND1, ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND4L, ND4, ND5, and Cytb encoded for heavy strands. Only ND6 gene encoded for light strand. ATG and GTG codons were identified as start codon in protein coding genes. COI and ND3 genes were started with GTG codon while the rest of 11 protein coding genes initiated with ATG codon. A complete stop codon usage in *Rasbora tornieri* protein coding genes are TAG (ND6) and TAA (ND1, COI, ATP8, ND4L, ND5, and Cytb). An incomplete stop codons usage in *Rasbora tornieri* protein coding genes are T (ND2, COII, and ND3) and TA (ATP6, COIII, and ND4) as in Table 4.19. Both ND4L and ATP8 genes were 7 nucleotides away from their respectively adjacent genes; ND4 and ATP6 genes (Table 4.19).

Table 4.19: Characteristics of complete mitochondrial genome of *Rasbora tornieri*

Gene	Position (5'- 3')	Length (bp)	Start codon	Stop codon	Amino acid	Anti-codon	Intergenic nucleotides (bp)	Strand
tRNA ^{Phe}	1 – 69	69				GAA	0	H
12S rRNA	70 – 1,025	956					0	H
tRNA ^{Val}	1,026 – 1,096	71				TAC	0	H
16S rRNA	1,097 – 2,784	1,688					0	H
tRNA ^{Leu}	2,785 – 2,859	75				TAA	1	H
ND1	2,861 – 3,835	975	ATG	TAA	324		4	H
tRNA ^{Ile}	3,840 – 3,911	72				GAT	-2	H
tRNA ^{Gln}	3,980 – 3,910	71				TTG	1	L
tRNA ^{Met}	3,982 – 4,050	69				CAT	0	H
ND2	4,051 – 5,095	1,045	ATG	T- -	348		0	H
tRNA ^{Trp}	5,096 – 5,167	72				TCA	2	H
tRNA ^{Ala}	5,237 – 5,170	68				TGC	1	L
tRNA ^{Asn}	5,311 – 5,239	73				GTT	37	L
tRNA ^{Cys}	5,414 – 5,349	66				GCA	1	L
tRNA ^{Tyr}	5,485 – 5,416	70				GTA	1	L
COI	5,487 – 7,037	1,551	GTG	TAA	516		0	H
tRNA ^{Ser}	7,108 – 7,038	71				TGA	1	L
tRNA ^{Asp}	7,110 – 7,180	71				GTG	7	H

Table 4.19 continued

COII	7,188 – 7,872	685	ATG	T- -	228		0	H
tRNA ^{Lys}	7,873 – 7,947	75				TTT	4	H
ATP8	7,952 – 8,116	165	ATG	TAA	54		-7	H
ATP6	8,110 – 8,792	683	ATG	TA-	227		0	H
COIII	8,793 – 9,577	785	ATG	TA-	261		0	H
tRNA ^{Gly}	9,578 – 9,648	71				TCC	0	H
ND3	9,649 – 9,997	349	GTG	T- -	116		0	H
tRNA ^{Arg}	9,998 – 10,067	70				TCG	0	H
ND4L	10,068 – 10,364	297	ATG	TAA	98		-7	H
ND4	10,358 – 11,739	1,382	ATG	TA-	460		0	H
tRNA ^{His}	11,740 – 11,808	69				GTG	0	H
tRNA ^{Ser}	11,809 – 11,876	68				GCT	2	H
tRNA ^{Leu}	11,879 – 11,951	73				TAG	0	H
ND5	11,952 – 13,781	1,830	ATG	TAA	609		-4	H
ND6	14,299 – 13,778	522	ATG	TAG	173		0	L
tRNA ^{Glu}	14,368 – 14,300	69				TTC	6	L
Cytb	14,375 – 15,511	1,137	ATG	TAA	378		5	H
tRNA ^{Thr}	15,517 – 15,587	71				TGT	14	H
tRNA ^{Pro}	15,671 – 15,602	70				TGG	0	L
D-loop	15,672 – 16,573	902					-	-

Table 4.20 shows the 22 tRNA genes, 2 rRNA genes, 13 protein coding genes, and D-loop region of *Rasbora borapetensis* complete mitochondrial DNA. A total of 14 tRNA genes were transcribed on heavy strand (tRNA^{Phe}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Met}, tRNA^{Trp}, tRNA^{Asp}, tRNA^{Lys}, tRNA^{Gly}, tRNA^{Arg}, tRNA^{His}, tRNA^{Ser}, tRNA^{Thr}, and two tRNA^{Leu}). The remaining 8 tRNA genes were transcribed on light strand (tRNA^{Gln}, tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Glu}, and tRNA^{Pro} genes). An atypical anticodon was found in 4 tRNA genes which are leucine (TAA), serine (TGA), another serine (GCT) which position at 11,800 - 11,867, and another leucine (TAG) gene at position of 11,869 – 11,941 as in Table 4.20. The tRNA^{Ile} gene overlaps 2 nucleotides with tRNA^{Gln} gene. The distance of 35 nucleotides between tRNA^{Asn} gene and tRNA^{Cys} gene was the origin of light strand (O_L) replication. The tRNA^{Thr} gene was 14 nucleotides away from tRNA^{Pro} gene.

The 12S rRNA and 16S rRNA genes are encoded for heavy strands and does not overlap with their adjacent genes. A total of 12 protein coding genes which are ND1, ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND4L, ND4, ND5, and Cytb were encoded for heavy strands. Only ND6 gene encoded for light strand. ATG and GTG codons were identified as start codon in protein coding genes. Only COI was started with GTG codon while other 12 protein coding genes were initiated with ATG codon. A complete stop codon usage in *Rasbora borapetensis* protein coding genes are TAG (ND6) and TAA (ND1, COI, ATP8, ND4L, and ND5). An incomplete stop codons usage in *Rasbora borapetensis* protein coding genes are T (ND2, COII, ND3, and Cytb) and TA (ATP6, COIII, and ND4) as in Table 4.20. Both ATP8 and ND4L genes were 7 nucleotides away from their respectively adjacent genes; ATP6 and ND4 genes. ND4 gene was 4 nucleotides overlap with ND6 gene.

Table 4.20: Characteristics of complete mitochondrial genome of *Rasbora borapetensis*

Gene	Position (5'- 3')	Length (bp)	Start codon	Stop codon	Amino acid	Anti-codon	Intergenic nucleotides (bp)	Strand
tRNA ^{Phe}	1 - 69	69				GAA	0	H
12S rRNA	70 – 1,022	953					0	H
tRNA ^{Val}	1,023 – 1,093	71				TAC	0	H
16S rRNA	1,094 – 2,776	1,683					0	H
tRNA ^{Leu}	2,777 – 2,851	75				TAA	1	H
ND1	2,853 – 3,827	975	ATG	TAA	324		4	H
tRNA ^{Ile}	3,832 – 3,903	71				GAT	-2	H
tRNA ^{Gln}	3,972 – 3,902	71				TTG	1	L
tRNA ^{Met}	3,974 – 4,042	69				CAT	0	H
ND2	4,043 – 5,087	1,045	ATG	T- -	348		0	H
tRNA ^{Trp}	5,088 – 5,160	73				TCA	2	H
tRNA ^{Ala}	5,230 – 5,163	68				TGC	1	L
tRNA ^{Asn}	5,304 – 5,232	73				GTT	35	L
tRNA ^{Cys}	5,405 – 5,340	66				GCA	1	L
tRNA ^{Tyr}	5,476 – 5,407	70				GTA	1	L
COI	5,478 – 7,031	1,554	GTG	TAA	517		0	H
tRNA ^{Ser}	7,102 – 7,032	71				TGA	1	L
tRNA ^{Asp}	7,104 – 7,173	70				GTC	4	H

Table 4.20 continued

COII	7,178 – 7,868	691	ATG	T- -	230		0	H
tRNA ^{Lys}	7,869 – 7,943	75				TTT	2	H
ATP8	7,946 – 8,110	165	ATG	TAA	54		-7	H
ATP6	8,104 – 8,783	680	ATG	TA-	226		0	H
COIII	8,784 – 9,568	785	ATG	TA-	261		0	H
tRNA ^{Gly}	9,569 – 9,639	71				TCC	0	H
ND3	9,640 – 9,988	349	ATG	T- -	116		0	H
tRNA ^{Arg}	9,989 – 10,058	70				TCG	0	H
ND4L	10,059 – 10,355	297	ATG	TAA	98		-7	H
ND4	10,349 – 11,730	1,382	ATG	TA-	460		0	H
tRNA ^{His}	11,731 – 11,799	69				GTG	0	H
tRNA ^{Ser}	11,800 – 11,867	68				GCT	1	H
tRNA ^{Leu}	11,869 – 11,941	73				TAG	0	H
ND5	11,942 – 13,768	1,827	ATG	TAA	608		-4	H
ND6	14,286 – 13,765	522	ATG	TAG	173		0	L
tRNA ^{Glu}	14,355 – 14,287	69				TTC	6	L
Cytb	14,362 – 15,502	1,141	ATG	T- -	380		0	H
tRNA ^{Thr}	15,503 – 15,573	71				TGT	14	H
tRNA ^{Pro}	15,657 – 15,588	70				TGG	0	L
D-loop	15,658 – 16,520	863					-	-

The circular map of *Rasbora sumatrana* mitochondrial genome with its full length of 16,539 bp was illustrated in Figure 4.8. The light strand (tRNA^{Gln}, tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Glu}, tRNA^{Pro}, and ND6) was represented by inner circle that containing genes. The rest of genes that encoded on heavy strand were represented by outer circle. The genes in circular map were respectively represented by different colour; tRNA gene (red), rRNA gene (light brown), protein coding gene (black), and D-loop region (brown). The blue arrow represented the origin of light strand (O_L) replication.

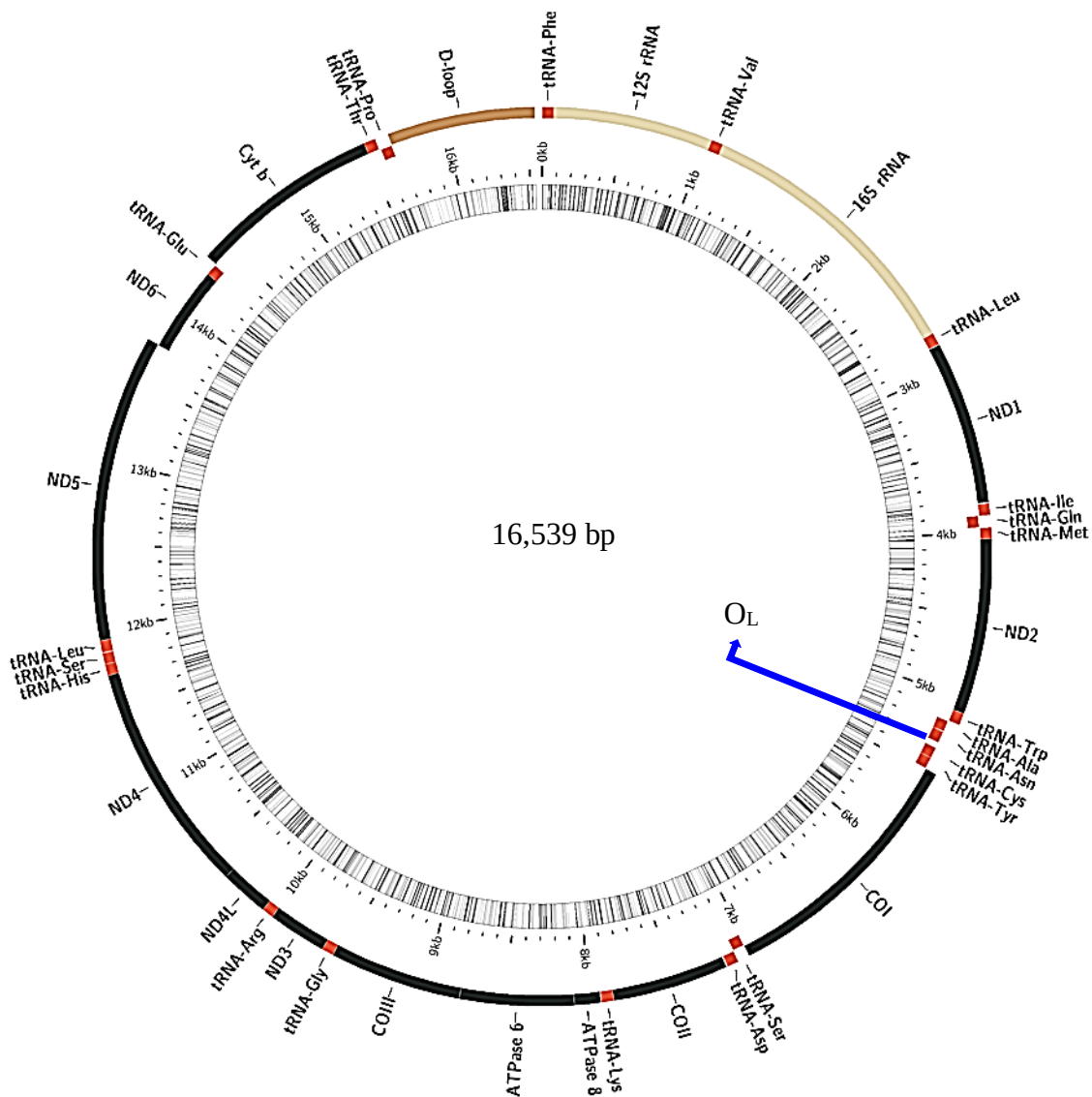


Figure 4.9 displays the circular map of *Rasbora hobelmani* mitochondrial genome with its full length of 16,541 bp. The light strand (tRNA^{Gln}, tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Glu}, tRNA^{Pro}, and ND6) was represented by inner circle that containing genes. The rest of genes that encoded on heavy strand were represented by outer circle. The genes in circular mitochondrial genome were respectively represented by different colours; tRNA gene (red), rRNA gene (light brown), protein coding gene (black), and D-loop region (brown). The blue arrow represented the origin of light strand (O_L) replication which located between the tRNA^{Asn} and tRNA^{Cys} genes.

Figure 4.9: Circular mitochondrial genome of *Rasbora hobelmani*

Referring to Figure 4.10, the complete length of *Rasbora tornieri* mitochondrial genome is 16,573 bp. Genes which encoded on the light strand (tRNA^{Gln}, tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Glu}, tRNA^{Pro}, and ND6) was represented by inner circle meanwhile genes that encoded on heavy strand were represented by outer circle. The features in circular map were represented by different colour; tRNA genes (red), rRNA genes (light brown), protein coding genes (black), and D-loop region (brown). The origin of light strand (O_L) replication was represented by long blue arrow.

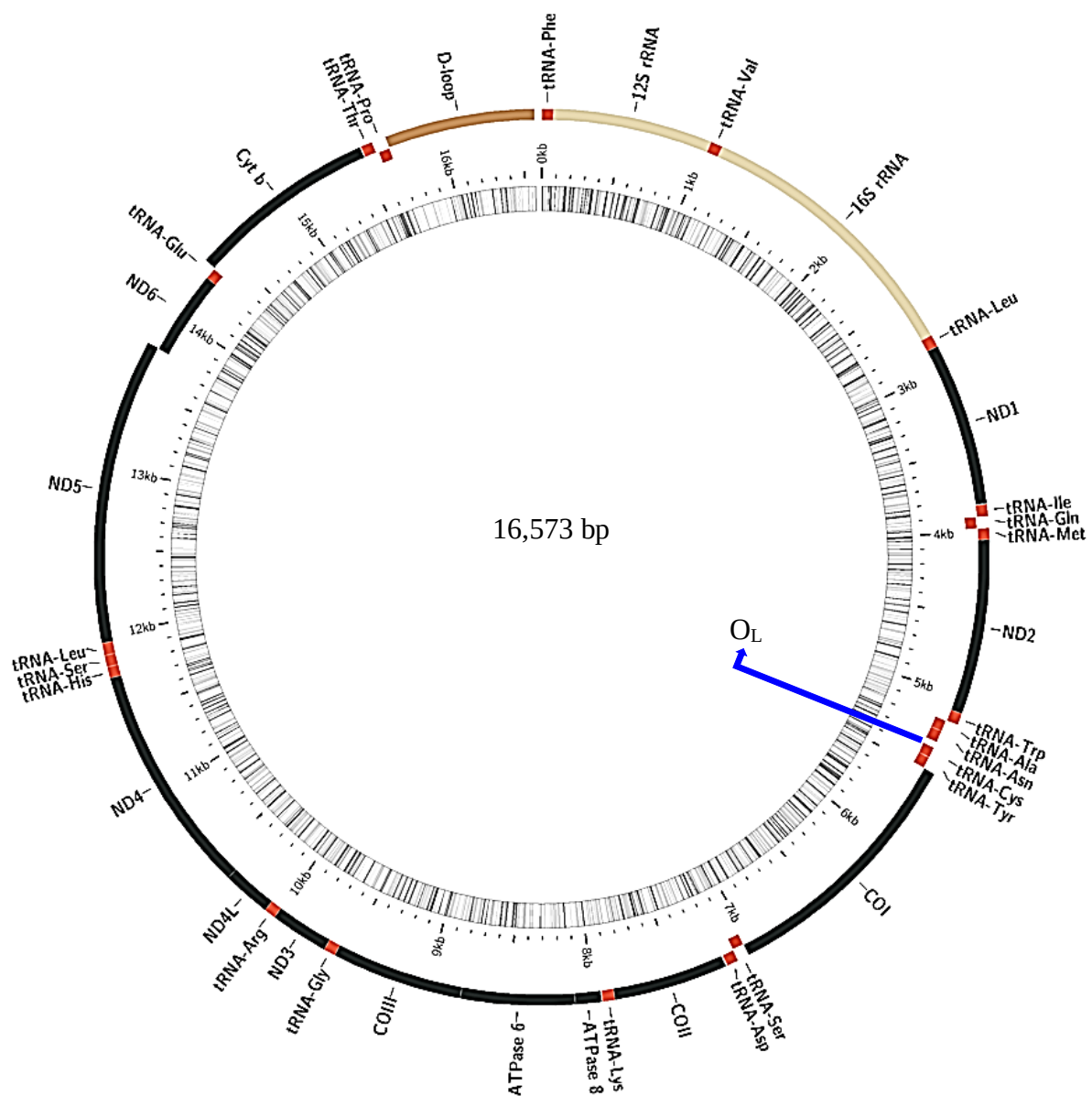


Figure 4.10: Circular mitochondrial genome of *Rasbora tornieri*

The complete length of *Rasbora borapetensis* mitochondrial genome is 16,520 bp. Genes that represented by inner circle were encoded on the light strand (tRNA^{Gln}, tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Glu}, tRNA^{Pro}, and ND6) meanwhile outer circle in the circular map represented the genes that encoded on the heavy strand. Features in circular map were represented by different colour; tRNA genes (red), rRNA genes (light brown), protein coding genes (black), and D-loop region (brown). The origin of light strand (O_L) replication was represented by long blue arrow as in Figure 4.11.

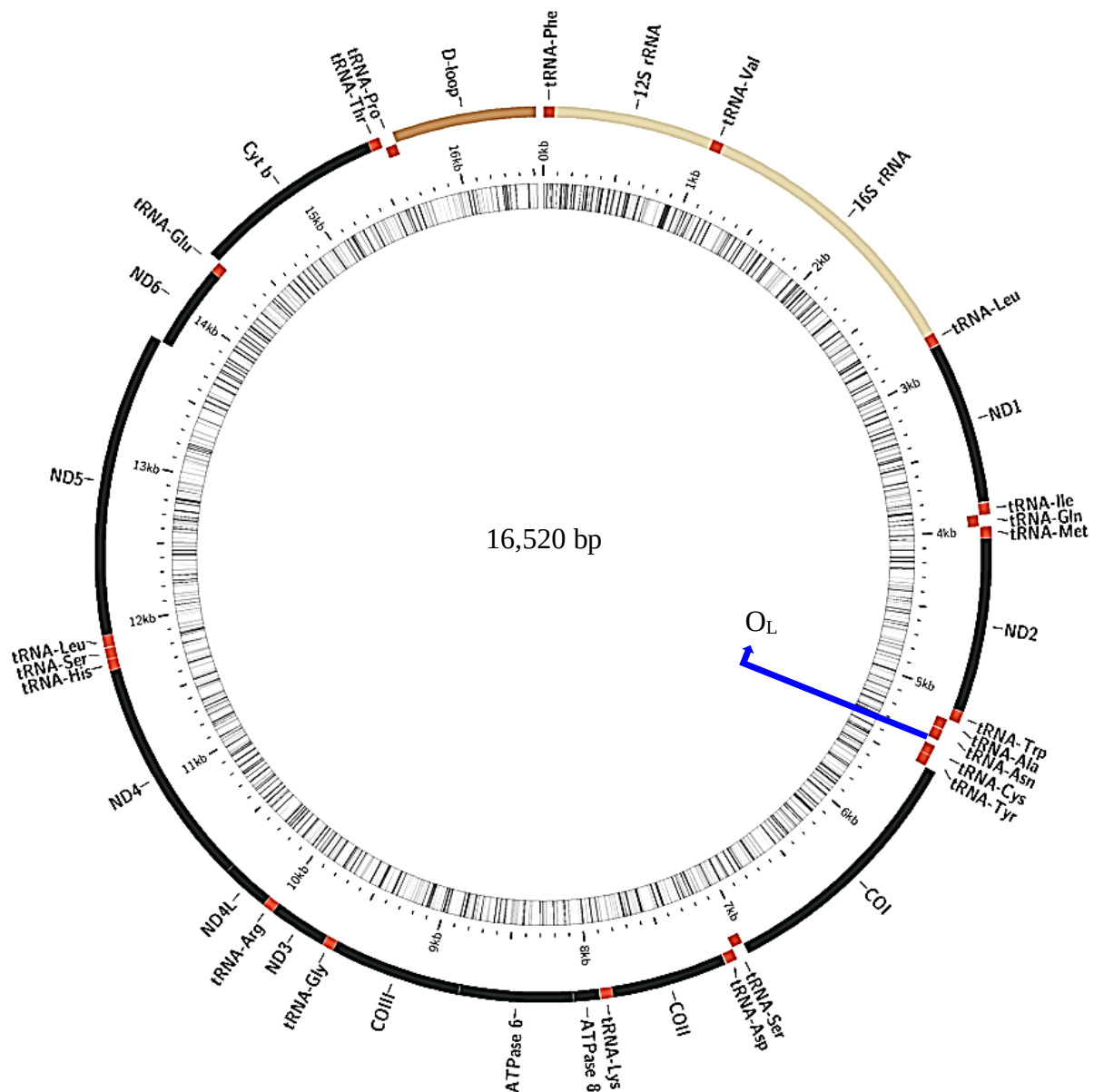


Figure 4.11: Circular mitochondrial genome of *Rasbora borapetensis*

Table 4.21 shows the whole nucleotide composition of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* in percentage by MEGA software. The highest AT contents was *Rasbora borapetensis* with 59.8% followed by *Rasbora tornieri* (59.2%), *Rasbora sumatrana* (57.9%), and *Rasbora hobelmani* (57.9%). The highest GC contents was *Rasbora sumatrana* with 42.1% followed by *Rasbora hobelmani* (42.0%), *Rasbora tornieri* (40.8%), and *Rasbora borapetensis* (40.2%). Concisely, the AT contents of 4 *Rasbora* species was higher than GC contents which indicate significant strand asymmetry. High contents of AT was frequently observed in fish (Shi *et al.*, 2016). In Table 4.22, the highest AT contents in *Rasbora sumatrana* and *Rasbora hobelmani* genome was tRNA^{Asp} with 68.5% and 68.6% respectively. In *Rasbora tornieri* genome, the highest AT contents was D-loop which is control region with 69.2%. The highest AT contents in *Rasbora borapetensis* genome was detected in the tRNA^{Gly} with 70.4% (Table 4.22).

Table 4.21: Whole nucleotide composition (%) of 4 *Rasbora* species

Nucleotide compostion	Size (bp)	T	C	A	G	AT	GC
<i>Rasbora sumatrana</i>	16,539	24.7	27.0	33.2	15.1	57.9	42.1
<i>Rasbora hobelmani</i>	16,541	24.8	26.9	33.1	15.1	57.9	42.0
<i>Rasbora tornieri</i>	16,573	25.6	26.3	33.6	14.5	59.2	40.8
<i>Rasbora borapetensis</i>	16,520	26.6	24.7	33.2	15.5	59.8	40.2

Table 4.22: Nucleotide composition of gene arrangements in percentage by MEGA

Genes		<i>Rasbora sumatrana</i>					<i>Rasbora hobelmani</i>					<i>Rasbora tornieri</i>					<i>Rasbora borapetensis</i>				
		T	C	A	G	AT	T	C	A	G	AT	T	C	A	G	AT	T	C	A	G	AT
rRNA	12S	18.7	26.2	34.5	20.6	53.2	18.9	26.2	34.5	20.5	53.4	19.1	25.9	34.8	20.1	53.9	19.4	25.6	32.8	22.1	52.2
	16S	19.3	24.4	37.8	18.6	57.1	19.4	24.3	37.8	18.5	57.2	20.2	23.6	37.9	18.4	58.1	21.0	22.7	37.4	18.9	58.4
	Phe	15.7	27.1	35.7	21.4	51.4	15.9	26.1	36.2	21.7	52.1	14.5	26.1	37.7	21.7	52.2	20.3	21.7	34.8	23.2	55.1
	Val	21.1	26.8	31.0	21.1	52.1	21.1	26.8	31.0	21.1	52.1	19.7	28.2	29.6	22.5	49.3	21.1	25.4	29.6	23.9	50.7
	Leu	25.3	25.3	25.3	24.0	50.6	25.3	25.3	26.7	22.7	52.0	22.7	25.3	26.7	25.3	49.4	25.3	24.0	25.3	25.3	50.6
	Ile	23.6	26.4	25.0	25.0	48.6	23.6	26.4	26.4	23.6	50.0	26.4	23.6	26.4	23.6	52.8	25.0	25.0	26.4	23.6	51.4
	Gln	25.4	25.4	35.2	14.1	60.6	25.4	25.4	35.2	14.1	60.6	23.9	26.8	35.2	14.1	59.1	25.4	25.4	33.8	15.5	59.2
	Met	21.7	29.0	31.9	17.4	53.6	21.7	29.0	31.9	17.4	53.6	20.3	31.9	30.4	17.4	50.7	23.2	29.0	29.0	18.8	52.2
	Trp	17.8	24.7	38.4	19.2	56.2	17.8	24.7	38.4	19.2	56.2	22.2	22.2	33.3	22.2	55.5	17.8	23.3	38.4	20.5	56.2
tRNA	Ala	27.9	25.0	35.3	11.8	63.2	27.9	25.0	36.8	10.3	64.7	29.4	25.0	35.3	10.3	64.7	29.4	25.0	35.3	10.3	64.7
	Asn	20.5	28.8	31.5	19.2	52.0	20.5	28.8	31.5	19.2	52.0	19.2	30.1	32.9	17.8	52.1	20.8	27.8	31.9	19.4	52.7
	Cys	23.9	28.4	29.9	17.9	53.8	23.9	28.4	29.9	17.9	53.8	24.2	28.8	31.8	15.2	56.0	24.2	25.8	33.3	16.7	57.5
	Tyr	19.7	32.4	33.8	14.1	53.5	19.7	32.4	33.8	14.1	53.5	21.4	31.4	30.0	17.1	51.4	20.0	31.4	28.6	20.0	48.6
	Ser	25.4	26.8	28.2	19.7	53.6	25.4	26.8	28.2	19.7	53.6	23.9	29.6	25.4	21.1	49.3	25.4	26.8	29.6	18.3	55.0
	Asp	31.4	17.1	37.1	14.3	68.5	30.0	18.6	38.6	12.9	68.6	28.2	21.1	35.2	15.5	63.4	32.9	17.1	37.1	12.9	70.0
	Lys	22.7	24.0	37.3	16.0	60.0	22.7	24.0	37.3	16.0	60.0	24.0	22.7	36.0	17.3	60.0	22.7	24.0	36.0	17.3	58.7
	Gly	28.2	23.9	31.0	16.9	59.2	26.8	25.4	31.0	16.9	57.8	29.6	21.1	33.8	15.5	63.4	32.4	15.5	38.0	14.1	70.4
	Arg	24.3	27.1	27.1	21.4	51.4	24.3	27.1	27.1	21.4	51.4	24.3	27.1	28.6	20.0	52.9	25.7	25.7	27.1	21.4	52.8
	His	29.0	23.2	34.8	13.0	63.8	30.4	21.7	34.8	13.0	65.2	30.4	21.7	34.8	13.0	65.2	30.4	21.7	34.8	13.0	65.2

Table 4.22 continued

	Ser	25.0	19.1	38.2	17.6	63.2	23.5	20.6	38.2	17.6	61.7	26.5	17.6	39.7	16.2	66.2	26.5	19.1	36.8	17.6	63.3
	Leu	27.4	17.8	34.2	20.5	61.6	27.4	17.8	34.2	20.5	61.6	27.4	17.8	34.2	20.5	61.6	27.4	19.2	34.2	19.2	61.6
	Glu	26.1	24.6	34.8	14.5	60.9	26.1	24.6	34.8	14.5	60.9	27.5	21.7	36.2	14.5	63.7	29.0	21.7	33.3	15.9	62.3
	Thr	25.4	26.8	23.9	23.9	49.3	26.8	25.4	23.9	23.9	50.7	25.4	26.8	25.4	22.5	50.8	28.2	22.5	26.8	22.5	55.0
	Pro	22.9	28.6	37.1	11.4	60.0	22.9	28.6	37.1	11.4	60.0	22.9	28.6	35.7	12.9	58.6	24.3	27.1	35.7	12.9	60.0
Protein coding	ND1	24.0	29.2	33.6	13.1	57.6	23.8	29.4	33.0	13.7	56.8	26.1	27.5	33.4	13.0	59.5	27.0	25.4	32.1	15.5	59.1
	ND2	21.8	31.7	35.5	11.0	57.3	22.0	31.5	36.0	10.5	58.0	21.7	30.4	37.6	10.2	59.3	24.7	27.3	36.1	12.0	60.8
	COI	28.9	25.5	28.1	17.5	57.0	28.9	25.3	28.0	17.7	56.9	29.5	25.0	28.4	17.2	57.9	30.5	23.2	28.3	18.0	58.8
	COII	28.2	23.6	31.3	16.9	59.5	28.2	23.6	31.4	16.8	59.6	28.6	24.8	30.8	15.8	59.4	27.9	23.4	33.0	15.6	60.9
	ATP8	30.9	26.1	32.7	10.3	63.6	30.9	26.1	32.7	10.3	63.6	29.1	27.9	35.8	7.3	64.9	33.3	23.0	35.2	8.5	68.5
	ATP6	30.9	26.1	32.7	10.3	63.6	30.3	27.5	30.6	11.6	60.9	31.2	26.6	31.5	10.7	62.7	30.3	25.9	32.2	11.6	62.5
	COIII	25.6	28.7	28.8	16.9	54.4	25.7	28.5	28.3	17.5	54.0	25.5	28.9	29.6	16.1	55.1	27.8	25.2	29.7	17.3	57.5
	ND3	27.3	29.6	29.6	13.5	56.9	27.2	29.8	29.5	13.5	56.7	29.2	27.8	29.5	13.5	58.7	29.5	26.6	30.1	13.8	59.6
	ND4L	26.3	30.3	29.0	14.5	55.3	26.3	30.3	29.0	14.5	55.3	28.3	27.9	30.0	13.8	58.3	28.3	28.3	30.6	12.8	58.9
	ND4	25.0	28.2	33.8	13.0	58.8	24.6	28.6	33.7	13.1	58.3	26.2	26.9	34.2	12.7	60.4	28.4	24.6	33.6	13.5	62.0
	ND5	24.8	28.2	34.6	12.5	59.4	25.1	27.9	34.6	12.4	59.7	26.2	26.8	35.5	11.5	61.7	27.2	25.4	34.9	12.5	62.1
	ND6	14.8	30.1	44.1	11.1	58.9	14.4	30.5	43.9	11.3	58.3	13.6	31.8	43.1	11.5	56.7	15.5	29.5	42.3	12.6	57.8
	Cytb	26.6	28.5	30.9	14.0	57.5	26.9	28.2	30.4	14.5	57.3	27.6	27.8	31.5	13.1	59.1	30.0	25.1	30.7	14.3	60.7
	D-loop	32.8	20.2	33.0	14.1	65.8	33.1	19.8	33.1	13.9	66.2	35.1	18.8	34.1	11.9	69.2	33.7	20.4	33.4	12.5	67.1

4.4.3 Phylogenetic analysis

The phylogenetic analysis was conducted by using MEGA software. A maximum likelihood tree was used to illustrate the phylogenetic position of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* among other *Rasbora* species (Figure 4.12). The maximum likelihood analysis was conducted using concatenated 13 protein coding genes of complete mitochondrial sequences under Tamura-Nei (TN93) + G + I substitution model with bootstrap probabilities of 1000 replications. A total of 14 sequences of Danioninae subfamily were obtained from the GenBank to compare with the 4 studied *Rasbora* (highlighted in Figure 4.12). An outgroup species which is *Cheilinus undulates* sequence was added into the phylogenetic analysis for tree rooting purposes. Overall, a total of 19 sequences were used in this phylogenetic analysis.

Among the 15 species in the phylogenetic tree, *Rasbora sumatrana* and *Rasbora hobelmani* shows a closer relationship with *Rasbora sumatrana* AP01121.1 with a bootstrap support of 100%. This means that the node is well supported. *Rasbora borapetensis* also shows a 100% of bootstrap support with *Rasbora borapetensis* AB924546.1. However, the bootstrap value of *Rasbora tornieri* was lower compared to other 3 studied *Rasbora* which is 64%.

In the phylogenetic tree, the pair of sister terminal taxa was shown by *Rasbora sumatrana* and *Rasbora hobelmani*. The studied *Rasbora tornieri* and *Rasbora borapetensis* has a distant relationship to studied *Rasbora sumatrana* and *Rasbora hobelmani*. This was illustrated by the deeper nodes connection between the *Rasbora tornieri* and *Rasbora borapetensis* with the *Rasbora sumatrana* and *Rasbora hobelmani*. The branches of *Rasbora tornieri* and *Rasbora borapetensis* were both connected with *Rasbora argyrotaenia* LC269105.1. This indicates that *Rasbora tornieri*, *Rasbora borapetensis*, and *Rasbora argyrotaenia* LC269105.1 has a common ancestor.

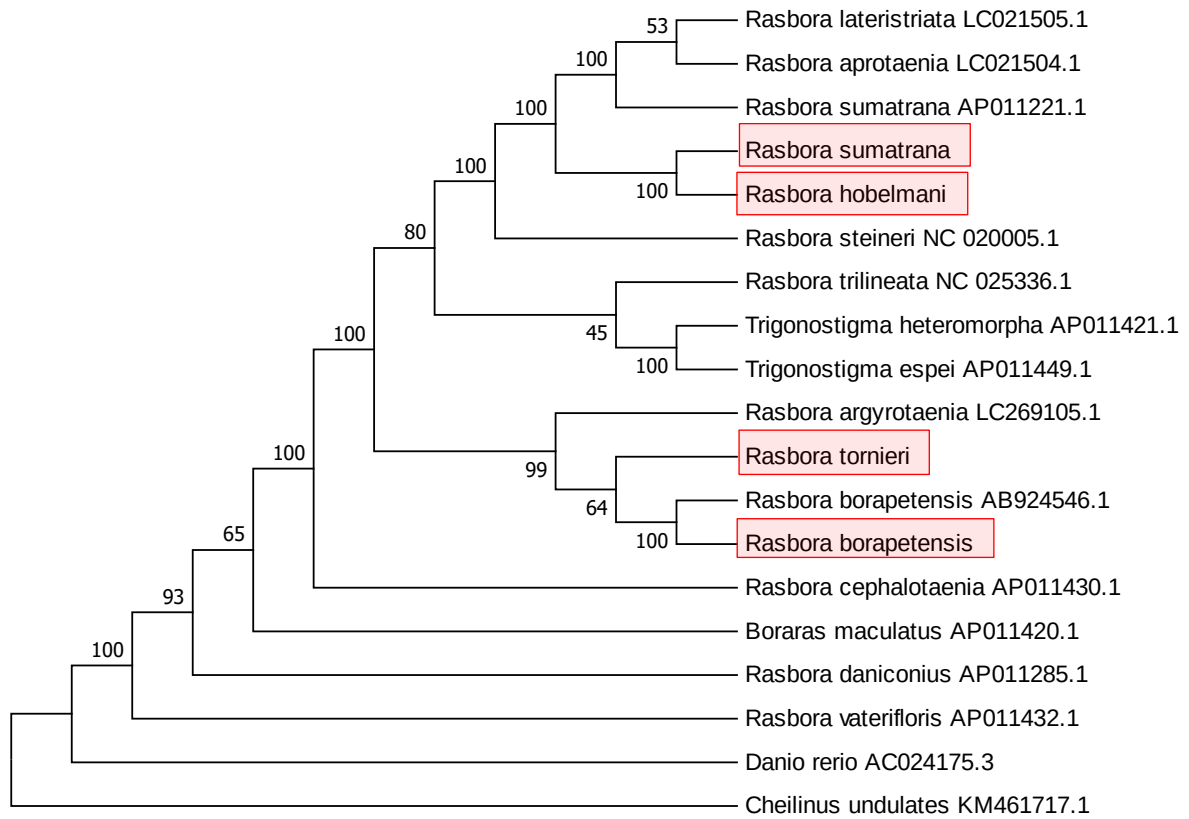


Figure 4.12: Phylogenetic tree of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* based on 13 protein coding genes using Maximum Likelihood (ML) analysis. The numbers on the branches are bootstrap values of ML.

CHAPTER 5

DISCUSSION

5.1 Morphometry analysis

The morphometry which consists of morphometric and meristic method is a useful tool in fish identification. Previously, the morphometric analysis has been reliable in distinguishing the species. For example, Syaifullah *et al.* (2015) used the morphological variations of *Zenarchopterus buffonis* from different area of West Sumatra. Muchlisin (2013) also used the morphological characters to separate the *Rasbora* species from Lake Laut Tawar, Indonesia into groups. Chandra and Prusty (2011) study on the fish fauna diversity of Keoladeo National Park wetland in India has recorded a new entry of fish species which is *Rasbora daniconius*. *Rasbora daniconius* fish has not been reported in this wetland until in 2005 where the identification was confirmed by morphometric method.

The results of univariate analysis (ANOVA test) on morphometric showed that the 24 characters were significantly different between *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis*. According to Aprilian *et al.* (2016) study of *Rasbora maninjau*, the differences of water depth and intensity of light may affect the size of fish eyes. The fish with larger eyes usually living in the shallow waters. In this study, referring to Table 4.1, *Rasbora borapetensis* (31.9-45.0% HL) has the largest eye diameter and the decreasing diameter followed by *Rasbora sumatrana* (30.3-38.4% HL), *Rasbora hobelmani* (28.3-35.3% HL), and *Rasbora tornieri* (27.2-35.3% HL). The fish samples of *Rasbora borapetensis* were obtained from drainage along the road which is the shallowest waters among the sampling areas.

Based on multivariate analysis (discriminant function analysis) of morphometric, total length and pre-dorsal length characters were mostly contributing to species differentiation. As

stated by Lumbantobing (2014), in Cyprinidae family, genus *Rasbora* is a small to moderate size of fish. Total length character is a measurement of snout tip toward the posterior end of caudal fin. In this study, the smallest size of total length was *Rasbora borapetensis* (120.52-131.57% SL). The increasing total length was followed by *Rasbora tornieri* (124.71-132.53% SL), *Rasbora sumatrana* (116.17-134.38% SL), and *Rasbora hobelmani* (128.55-135.09% SL). The variation in fish sizes may due to environmental factors in their habitat such as food supply and hydrodynamic condition. In Syaifullah *et al.* (2015) study, the total length and pre-dorsal length characters were among the characters that contribute to variation in *Zenarchopterus buffonis* from different area.

Pre-dorsal length is a measurement of snout tip toward the base of first dorsal spine. In this study, the smallest size of pre-dorsal length was *Rasbora hobelmani* (50.30-53.48% SL) and increasing size followed by *Rasbora sumatrana* (52.41-55.57% SL), *Rasbora tornieri* (51.57-56.98% SL), and *Rasbora borapetensis* (52.80-57.36% SL). The study of *Rasbora* group variation in morphometric by Muchlisin (2013) has revealed that the head and caudal region were the major characters that lead to group distinguished. It is probably related to the nutrient and speed of water flow differences. Matthews (1998) also stated that the head, jaw, and propulsive musculature adaptation in obtaining the food in their habitat influenced the morphological variation of fish species.

Based on non-parametric test (Kruskal-Wallis) of meristic, only dorsal fin character was not significant. Then, the multivariate analysis (discriminant function analysis) was applied on meristic. The lateral line scales and pre-dorsal scales characters were found to be the major contribution to *Rasbora* species discriminant. Lateral line scale was counted along the lateral line. The lateral line scales of *Rasbora hobelmani* (25-29 scales) was lesser with mode of 28. Then followed by *Rasbora sumatrana*, mode of 28 (27-32 scales), *Rasbora borapetensis*, mode of 27, 29, and 31 (27-31 scales), and *Rasbora tornieri*, mode of 33 and 34

(32-35 scales). In Chandra and Prusty (2011) study, the lateral line scale was the main character that distinguished *Rasbora daniconius* from other *Rasbora* species in the Keoladeo National Park. The lateral line in fish was involved in hydrodynamic motion detection. Based on the Spiller *et al.* (2017) study, the specialisation of lateral line was displayed by *Melanotaenia australis* in different habitats. Water velocity, risk to predator, habitat structure, and availability of prey were the factors that leads to lateral line diversity. In this study, *Rasbora tornieri* has higher number of lateral line scales among studied *Rasbora*. *Rasbora tornieri* samples were obtained from large river which require great energy in finding food and hydrodynamic motion detection.

Pre-dorsal scale was counted from top of head towards the first dorsal fin base. The pre-dorsal scales of *Rasbora sumatrana* was lesser with mode of 12 (11-13 scales) followed by *Rasbora borapetensis* mode of 13 (13scales), *Rasbora tornieri*, mode of 13 (12-15 scales), and *Rasbora hobelmani*, mode of 14 (10-14 scales). In Syaifullah *et al.* (2015) study on *Zenarchopterus buffonis* morphology from different area, the circular scales on the caudal peduncle character was contributed to variation. In general, the morphometric and meristic analysis are able to separate the 4 presumed *Rasbora* into their respectively species.

5.2 Complete mitochondrial genome analysis

The four complete mitochondrial DNA genome isolated from this study for *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* respectively were identified by nucleotide BLAST search which result in zero expected value. In this case, these four mitochondrial genomes were significantly similar with those *Rasbora* genomes deposited in the public database. Nucleotide BLAST search indicates that a 95 % of similarity was established between studied *Rasbora sumatrana* and *Rasbora hobelmani* with *Rasbora*

sumatrana AP011221.1 in the GenBank database. Meanwhile the studied *Rasbora tornieri* and *Rasbora borapetensis* resulting in 89 % and 99 % respectively similar with *Rasbora sumatrana* AP011221.1 and *Rasbora borapetensis* AB924546.1 in the GenBank database. The frequent similarity to *Rasbora sumatrana* AP011221.1 in the public database indicates the absence of availability of mitochondrial genome for *Rasbora* genus. Therefore, more mitochondrial genome needs to be available in the public database in order to evaluate and distinct the *Rasbora* genus accurately.

The full length of the complete mitochondrial genome for four *Rasbora* species were 16,539 bp (*Rasbora sumatrana*), 16,541 bp (*Rasbora hobelmani*), 16,573 bp (*Rasbora tornieri*), and 16,520 bp (*Rasbora borapetensis*) respectively. The shortest mitochondrial genome was *Rasbora borapetensis* and the *Rasbora tornieri* has the longest mitochondrial genome. The characterization of complete mitochondrial genome sequences of the four *Rasbora* species are canonically identical and consistent with each other.

The range length of 12S ribosomal RNA gene between four *Rasbora* species was 951 bp to 956 bp where respectively *Rasbora sumatrana* (951 bp), *Rasbora hobelmani* (952 bp), *Rasbora tornieri* (956 bp), and *Rasbora borapetensis* (953 bp). Meanwhile the length of 16S ribosomal RNA gene was ranging from 1,682 to 1,688 bp where *Rasbora sumatrana* (1,682 bp), *Rasbora hobelmani* (1,682 bp), *Rasbora tornieri* (1,688 bp), and *Rasbora borapetensis* (1,683 bp) respectively. These 2 ribosomal RNA genes were encoded on heavy strand. As in other vertebrates, the 12S rRNA genes was located between tRNA^{Phe} and tRNA^{Val} genes, whereas the 16S rRNA was located between tRNA^{Val} and tRNA^{Leu} genes (Prosdocimi *et al.*, 2011).

In this study, the origin of light strand (O_L) replication of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* was detected in similar region which

is between the big distance of tRNA^{Asn} and tRNA^{Cys} genes. In *Rasbora sumatrana* and *Rasbora hobelmani*, the O_L replication was 34 intergenic nucleotides. Meanwhile in *Rasbora tornieri* and *Rasbora borapetensis*, the O_L replication was 37 and 35 intergenic nucleotides respectively. The origin of light strand in Moreira *et al.* (2016) study of mailed catfish *Corydoras nattereri* from south eastern Brazil was 35 nucleotides. This catfish is small in size and inhabit coastal river drainage which similar to *Rasbora borapetensis*.

The 22 transfer RNA genes of 4 *Rasbora* species were ranging from 66 to 75 bp. There are 2 forms of tRNA^{Leu} and tRNA^{Ser} were identified. The range of 22 tRNA genes (66 - 75 bp) were similar in study of Prosdocimi *et al.* (2011) on *Nannoperca obscura* and *Nannoperca australis* fish. These perciformes fishes are apparently found in similar habitat to *Rasbora* species such as lake and river basin of south eastern Australia.

The 13 protein coding genes that commonly found in other vertebrates mitochondrial genome were also presented in *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* with the same organization. In *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora borapetensis*, only COI (cytochrome oxidase c subunit 1) gene was initiated with GTG codon while the rest of protein coding genes were started with ATG codon. The COI gene in catfish *Cranoglanis boudierius* (Peng *et al.*, 2006) and *Rasbora steineri* (Chang *et al.*, 2013) also used GTG codon as a start codon. Meanwhile, in *Rasbora tornieri*, 2 protein coding genes which are COI and ND3 (NADH dehydrogenase subunit 3) genes were started with GTG codon. This alternative start codon was translated as methionine although it encodes valine (GTG) amino acid. A GTG codon was rarely used as a start codon in other genes.

As in other vertebrates, *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis* used TAA and TAG as stop codons. The TAA stop codon is the most prevalent codon. In *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora borapetensis*, TAA

codon was used in ND1, COI, ATP8, ND4L, and ND5 genes. In *Rasbora tornieri*, TAA stop codon was used in ND1, COI, ATP8, ND4L, ND5 and Cytb genes. A TAG codon was used in ND6 gene of all studied *Rasbora*. As compared to helmet catfish *Cranoglanis bauderius* (Peng *et al.*, 2006), TAG stop codon was the most prevalent where it applied to ND1, ND2, COI, ND3, and ND6 genes. Two of incomplete stop codon which are T- - and TA- were detected in all studied *Rasbora*. An incomplete T- - stop codon was used in ND2, COII, ND3, and Cytb genes of *Rasbora sumatrana*, *Rasbora hobelmani*, and *Rasbora borapetensis*. The similar stop codon in *Rasbora tornieri* was used in ND2, COII, and ND3 genes. This T- - stop codon in minnow *Aphyocypris moltrechti* was also found in COII and Cytb genes. This incomplete stop codon was presumably completed by polyadenylation of RNA messenger after cleavage (Jang-Liaw *et al.*, 2013). Another an incomplete TA- stop codon was used in ATP6, COIII, and ND4 genes of all studied *Rasbora*.

5.3 Phylogenetic analysis

Phylogenetic analysis was performed to elucidate the interspecific relationships of *Rasbora* species by using DNA sequences from complete mitochondrial DNA of single individual that represents each species. The concatenated 13 protein coding genes of each *Rasbora* species in this study were used to estimate the phylogenetic position by using Maximum Likelihood (ML) method in MEGA software. Tian *et al.* (2017) study also using concatenated 13 protein coding genes of *Gymnocypris scleracanthus* fish to study its taxonomic position. Jiang *et al.* (2014) study on Cyprinid fish *Aspiorhynchus laticeps* complete mitochondrial as well using the concatenated 13 protein coding genes in phylogenetic analysis.

In this study, the phylogenetic analysis of 18 cyprinid fishes and 1 outgroup saltwater fish indicated that *Rasbora sumatrana* had the closet relationship with *Rasbora hobelmani* with

a strong bootstrap support of 100 %. Kottelat (1984) has described that *Rasbora hobelmani* was closely related to *Rasbora sumatrana* as they had similar body shape and colour pattern. Meyer *et al.* (1993) stated that the closely related species have genes in common as they also shared ancestors in which some of the genes may absence and modified among distant relatives.

The position of *Rasbora tornieri* and *Rasbora borapetensis* in the phylogenetic tree displayed that both has a common ancestor with *Rasbora argyrotaenia* LC269105.1. Based on Kusuma *et al.* (2017) phylogenetic study, the relationship of *Rasbora argyrotaenia* is closer to *Rasbora borapetensis* compared to other *Rasbora* fish in Java Island, Indonesia. In this case, the relationship of *Rasbora borapetensis* with *Rasbora argyrotaenia* LC269105.1 showed a strong bootstrap support of 99 % (refer to Figure 4.12).

The phylogenetic analysis in this study also evaluated that the relationship between *Rasbora* species were non-monophyletic due to the position of *Trigonostigma heteromorpha* AP011421.1, *Trigonostigma espei* AP011449.1, and *Boraras maculatus* AP011420.1. This results was similar to Fang *et al.* (2008) study where the non-monophyletic of *Rasbora* was also due to the location of *Trigonostigma* and *Boraras* in the phylogenetic tree. In Mayden *et al.* (2007) study as well showed that *Microrasbora* and *Rasbora* genus were non-monophyletic assemblages, whereas *Boraras* genus was monophyletic.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

In conclusions, morphometry analysis had classified these 4 *Rasbora* into their respective *Rasbora* species. The complete mitochondrial genome has been successfully isolated from *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and *Rasbora borapetensis*. The 22 transfer RNA genes, 13 protein coding genes, 2 ribosomal RNA genes, and D-loop region of complete mitochondrial genome of 4 studied *Rasbora* characterized and mapped into circular. The phylogeny of these 4 studied *Rasbora* were established. The phylogenetic relationship of each studied *Rasbora* with other 10 *Rasbora* genus members from Genbank was discovered by using Maximum likelihood method. Thus, the molecular analysis produced more reliable data than morphometry analysis as these *Rasbora* mitochondrial genome data are essential for *Rasbora* evolution studied in the future.

6.2 Recommendations

The partial fragment of mitochondrial study of *Rasbora* was abundant in Genbank. This makes the blast search more accurate in matching the sequences. However, there are only 10 complete mitochondrial sequence of *Rasbora* species that has been submitted to the Genbank which including *Rasbora lateristriata*, *Rasbora vaterifloris*, *Rasbora daniconius*, *Rasbora aprotaenia*, *Rasbora cephalotaenia*, *Rasbora argyrotaenia*, *Rasbora sumatrana*, *Rasbora steineri*, *Rasbora borapetensis*, and *Rasbora trilineata*. There are great diversity of *Rasbora* in Sarawak that still not yet discovered and study. The extending to this study on complete mitochondrial genome of other *Rasbora* species from Sarawak, Malaysia will yield significant

insights about phylogeny of the *Rasbora* group. There will be more available data in the Genbank for the next generation to refer.

Second recommendation is to study the *Rasbora* breeding in the laboratory environment. Mainly, wild *Rasbora* which caught from the local river was used in a lot of research. However, there is limited study of *Rasbora* breeding. By studying *Rasbora* breeding in the laboratory environment, the breeding strategy and mating behaviour data of *Rasbora* are obtain. The laboratory *Rasbora* fish can be use in biomedical studies. Contamination in research also can be reduced by using laboratory *Rasbora* compared to wildly caught *Rasbora*. *Rasbora* may capable of becoming next fish model organism apart from *Danio rerio*, the zebrafish.

REFERENCES

- Aprilian, E., Roesma, D. I., & Tjong, D. H. (2016). Morphological variation of Bada fish (*Rasbora maninjau*, Lumbantobing) in Maninjau Lake, West Sumatra. *Journal of Entomology and Zoology Studies*, 4(2), 541-544.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455-477.
- Barlow, G. W. (1961). Causes and Significance of Morphological Variation in Fishes. *Systematic Zoology*, 10(3), 105-117.
- Berbel-Filho, W. M., Jacobina, U. P., & Martinez, P. A. (2013). Preservation effects in geometric morphometric approaches: Freezing and alcohol in a freshwater fish. *Ichthyological Research*, 60(3), 268-271.
- Bíró, R. K. (2015). *Constructing Phylogenetic Trees* (Unpublished bachelor's thesis). Eötvös Loránd University, Budapest, Hungary.
- Brittan, M. R. (1954). The cyprinid fish genus *Rasbora* in Malaya. *The Raffles Bulletin of Zoology*, 25, 129-156.
- Chandra, R., & Prusty, B. A. K. (2011). A new record of a fish species, *Rasbora daniconius*, in the wetlands of Keoladeo National Park, Bharatpur, India, and conservation concerns. *Environmental Research Journal*, 4(5-6), 499-514.
- Chang, C., Tsai, C., & Jang-Liaw, N. (2013). Complete mitochondrial genome of the Chinese rasbora *Rasbora steineri* (Teleostei, Cyprinidae). *Mitochondrial DNA*, 24(3), 183-185.

- Chen, F., Ma, H., Ma, C., Zhang, H., Zhao, M., Meng, Y., Wei, H. & Ma, L. (2016). Sequencing and characterization of mitochondrial DNA genome for *Brama japonica* (Perciformes: Bramidae) with phylogenetic consideration. *Biochemical Systematics and Ecology*, 68, 109-118.
- Chung, H. H. (2013). Ontogenic expression and functional characterization of FADS2, an enzyme for long-chained unsaturated fatty acid desaturation in zebrafish (Ph.D dissertation). Universiti Sains Malaysia, Pulau Pinang, Malaysia.
- Elliott, N. G., Haskard, K. & Koslow, J. A. (1995). Morphometric analysis of orange roughy (*Hoplostethus atlanticus*) off the continental slope of southern Australia. *Journal of Fish Biology*, 46(2), 202-220.
- Fang, F., Nori, N., M., Liao, T. Y., Källersjö, M., & Kullander, S. O. (2009). Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). *Zoologica Scripta*, 38(3), 237-256.
- Gilpin, D. (2011). *The complete illustrated guide to animals, birds & fish of the British Isles*. Anness Publishing, London, United Kingdom.
- Hadiaty, R. K. (2009). *Rasbora lacrimula*, a new species of cyprinid fish from eastern Borneo (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 20(2), 105-109.
- Hair, J. F., Tatham, R. L., Anderson, R. E. & Black, W. (1998). *Multivariate Data Analysis* (5th Ed.). Upper Saddle River, USA: Prentice Hall.
- Ho, C., Liu, M. & Chen, M. (2014). Complete mitochondrial genome of *Rasbora trilineata* (Cypriniformes, Cyprinidae). *Mitochondrial DNA*, 1-3.
- Hrbek, T., & Farias, I. P. (2008). The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). *Genetics and Molecular Biology*, 31(1), 293-302.

- Hubert, N., Wibowo, A., Busson, F., Caruso, D., Sulandari, S., Nafiqoh, N., Pouyaud, L., Ruber, L., Avarre, J., Herder, F., Hanner, R., Keith, P., Hadiaty, R. K. & Kadarusman. (2015). DNA Barcoding Indonesian freshwater fishes: challenges and prospects. *DNA Barcodes*, 3(1), 144-169.
- Inger, R. F. & Chin, P. K. (1990). *The freshwater fishes of North Borneo* (Vol. 45). Sabah, Malaysia, Kota Kinabalu: Sabah Zoological Society.
- Jang-Liaw, N., Tsai, C., Chang, C., & Watanabe, K. (2013). Complete mitochondrial genome of the Moltrechts minnow, *Aphyocypris moltrechti* (Teleostei, Cyprinidae), in comparison with *A. kikuchii*. *Mitochondrial DNA*, 24(2), 117-119.
- Jiang, M., Yang, C., & Wen, H. (2014). The complete mitochondrial genome of *Aspiorhynchus laticeps* and its phylogenetic analysis. *Meta Gene*, 2, 218-225.
- Kannan, L., & Wheeler, W. C. (2012). Maximum Parsimony on Phylogenetic networks. *Algorithms for Molecular Biology*, 7(9), 1-10.
- Kashyap, A., Awasthi, M. & Serajuddin, M. (2015). Intraspecific Diversity of Freshwater Murrel, *Channa punctatus*. *International Journal of Fisheries and Aquatic Studies*, 3(2), 166–173.
- Kocher, T. D. & Stepien, C. A. (1997). *Molecular systematics of fishes*. San Diego, CA: Academic Press.
- Kottelat, M. (1984). A new *Rasbora* s.l. (Pisces: Cyprinidae) from Northern Thailand. *Revue Suisse de Zoologie*, 91, 717–723.
- Kottelat, M., Whitten, A. J., Kartikasari, S. N. & Wirjoatmodjo, S. (1993). *Freshwater fishes of western Indonesia and Sulawesi*. Hong Kong: Periplus Editions.
- Kottelat, M. & Tan, H. H. (2011). Identity of *Rasbora beauforti*, a cyprinid fish from Borneo (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 22(4), 319-326.

- Kottelat, M. & Tan, H. H. (2011). *Rasbora atranus*, a new species of fish from central Borneo (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 22(3), 215-220.
- Kottelat, M. & Tan, H. H. (2012). *Rasbora cryptica*, a new species of fish from Sarawak, Borneo (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 23(1), 37-44.
- Kottelat, M. (2005). *Rasbora notura*, a new species of cyprinid fish from Malay Peninsula (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 16(3), 265-270.
- Kottelat, M. (2007). *Rasbora dies*, a new species of cyprinid fish from eastern Borneo (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, 18(4), 301-305.
- Kottelat, M. (2012). *Rasbora rheophila*, a new species of fish from northern Borneo (Teleostei: Cyprinidae). *Revue Suisse De Zoologie.*, 119(1), 77-87.
- Kumar, R., Singh, P. J., Nagpure, N. S., Kushwaha, B., Srivastava, S. K. & Lakra, W. S. (2007). A non-invasive technique for rapid extraction of DNA from fish scales. *Indian Journal of Experimental Biology*, 45, 992-997.
- Kusuma, W. E. & Kumazawa, Y. (2015). Complete mitochondrial genome sequences of two Indonesian rasboras (*Rasbora aprotaenia* and *Rasbora lateristriata*). *Mitochondrial DNA Part A*, 27(6), 4222-4223.
- Kusuma, W. E., Ratmuangkhwang, S. & Kumazawa, Y. (2016). Molecular phylogeny and historical biogeography of the Indonesian freshwater fish *Rasbora lateristriata* species complex (Actinopterygii: Cyprinidae): Cryptic species and west-to-east divergences. *Molecular Phylogenetics and Evolution*, 105, 212–223.
- Kusuma, W. E., Samuel, P. D., Wiadnya, D. G., Hariati, A. M. & Kumazawa, Y. (2017). Complete mitogenome sequence of *Rasbora argyrotaenia* (Actinopterygii: Cyprinidae). *Mitochondrial DNA Part B*, 2(2), 373-374.

- Liao, T. Y., Kullander, S. O. & Fang, F. (2010). Phylogenetic analysis of the genus *Rasbora* (Teleostei: Cyprinidae). *Zoologica Scripta*, 39, 155-176.
- Lowe, T.M. and Chan, P.P. (2016) tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. *Nucleic Acids Research*, 44, W54-57.
- Lumbantobing, D. N. (2010). Four New Species of the *Rasbora trifasciata*-Group (Teleostei: Cyprinidae) from Northwestern Sumatra, Indonesia. *Copeia*, 2010(4), 644-670.
- Lumbantobing, D. N. (2014). Four new species of *Rasbora* of the Sumatrana group (Teleostei: Cyprinidae) from northern Sumatra, Indonesia. *Zootaxa*, 3764(1), 1-25.
- Mahapatro, G., Mishra, D., Shaw, K., Mishra, S., & Jena, T. (2012). Phylogenetic Tree Construction for DNA Sequences using Clustering Methods. *Procedia Engineering*, 38, 1362-1366.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10-12.
- Matthews, W. J. (1998). *Patterns in Freshwater Fish Ecology*. New York, USA: Chapman & Hall.
- Meyer, A., Biermann, C. H., & Orti, G. (1993). The phylogenetic position of the zebrafish (*Danio rerio*), a model system in developmental biology: An invitation to the comparative method. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 252(1335), 231-236.
- Mollah, M., Yeasmine, S., Hossen, M. B. & Ahammad, A. K. (2012). Landmark-based morphometric and meristic variations of *Glossogobius giuris* in three stocks. *Journal of the Bangladesh Agricultural University*, 10(2), 375-384.
- Moreira, D. A., Buckup, P. A., Britto, M. R., Magalhães, M. G., Andrade, P. C., Furtado, C., & Parente, T. E. (2016). The complete mitochondrial genome of *Corydoras nattereri* (Callichthyidae: Corydoradinae). *Neotropical Ichthyology*, 14(1), e150167.

- Moyle, P. B., & Cech, J. J., Jr. (2004). *Fishes: An Introduction to Ichthyology* (5th ed.). USA: Prentice Hall.
- Muchlisin, Z. A. (2013). Morphometric Variations of Rasbora Group (Pisces: Cyprinidae) in Lake Laut Tawar, Aceh Province, Indonesia, Based on Truss Character Analysis. *HAYATI Journal of Biosciences*, 20(3), 138-143.
- Ng, H. H. & Kottelat, M. (2013). The identity of the cyprinid fishes *Rasbora dusonensis* and *R. tornieri* (Teleostei: Cyprinidae). *Zootaxa*, 3635(1), 62-70.
- Parenti, L. R. & Lim, K. K. (2005). Fishes of the Rajang Basin, Sarawak, Malaysia. *The Raffles Bulletin of Zoology*, (13), 175-208.
- Peng, Z., Wang, J., & He, S. (2006). The complete mitochondrial genome of the helmet catfish *Cranoglanis boudierius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene*, 376(2), 290-297.
- Prosdocimi, F., Carvalho, D. C., Almeida, R. N., & Beheregaray, L. B. (2011). The complete mitochondrial genome of two recently derived species of the fish genus *Nannoperca* (Perciformes, Percichthyidae). *Molecular Biology Reports*, 39(3), 2767-2772.
- Rainboth, W. J. & Kottelat, M. (1987). *Rasbora spilocerca*, a New Cyprinid from the Mekong River. *Copeia*, 1987(2), 417-423.
- Roberts, T. R. (1989). *The freshwater fishes of Western Borneo (Kalimantan Barat, Indonesia)*. San Francisco: California Academy of Sciences.

- Shi, X., Tian, P., Lin, R., Huang, D., & Wang, J. (2016). Characterization of the Complete Mitochondrial Genome Sequence of the Globose Head Whiptail *Cetonus globiceps* (Gadiformes: Macrouridae) and Its Phylogenetic Analysis. *Plos One*, 11(4), 1-17.
- Siebert, D. J. (1997). The identities of *Rasbora paucisqualis* Ahl in Schreitmuller, 1935, and *Rasbora bankanensis* (Bleeker, 1853), with the designation of a lectotype for *R. paucisqualis* (Teleostei: Cyprinidae). *The Raffles Bulletin of Zoology*, 45(1), 29-37.
- Sim, C. H. (2002). *A field guide to the fish of Tasik Bera Ramsar Site, Pahang, Malaysia* (1st ed.). Wetlands International-Malaysia Programme.
- Spiller, L., Grierson, P. F., Davies, P. M., Hemmi, J., Collin, S. P., & Kelley, J. L. (2017). Functional diversity of the lateral line system among populations of a native Australian freshwater fish. *The Journal of Experimental Biology*, 220(12), 2265-2276.
- Staton, J. L. (2015). Understanding phylogenies: Constructing and Interpreting Phylogenetic trees. *Journal of the South Carolina Academy of Science*, 13(1), 22-29.
- Sulaiman, Z. H. & Mayden, R. L. (2012). Cypriniformes of Borneo (Actinopterygii, Otophysi): An Extraordinary Fauna for Integrated Studies on Diversity, Systematics, Evolution, Ecology, and Conservation. *Zootaxa*, 3586, 359–376.
- Suvarna, D. S. & Rita, K. S. D. (2014). Morphometric and electrophoretic studies of *Rasbora dandia* (Valenciennes) (Pisces; Cyprinidae) from varied habitats in Kerala. *Journal of Aquatic Biology & Fisheries*, 2(1), 290-296.
- Syaifullah, S., Fajri, H., & Roesma, D. I. (2015). Morphometric variations of halfbeak fish (*Zenarchopterus buffonis*) from estuary of west Sumatra, Indonesia. *Aquaculture, Aquarium, Conservation & Legislation International Journal of the Bioflux Society*, 8(2), 168-176.

- Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2), 103–123.
- Talib, Y., Farooqui, A., Fatema, M., & Khan, W. (2016). Phylogenetic tree construction of biosurfactant producing organisms. *Journal of Global Bioscience*, 5(5), 4105-4108.
- Tian, F., Zhao, K., & Tang, Y. (2017). Characterization of the complete mitochondrial genome of *Gymnocypris scleracanthus* (Teleostei: Cyprinidae). *Conservation Genetics Resource*, 10(3), 467-469.
- Trontelj, P. & Fišer, C. (2009). Perspectives: Cryptic species diversity should not be trivialised. *Systematics and Biodiversity*, 7(1), 1-3.
- Turan, C., Erguden, D., Gurlek, M., Basusta, N. & Turan, F. (2004). Morphometric Structuring of the Anchovy (*Engraulis encrasicolus* L.) in the Black, Aegean and North eastern Mediterranean Seas. *Turkish Journal of Veterinary and Animal Sciences*, 28, 865-871.
- Turan, C., Oral, M., Öztürk, B. & Düzgüneş, E. (2006). Morphometric and meristic variation between stocks of Bluefish (*Pomatomus saltatrix*) in the Black, Marmara, Aegean and northeastern Mediterranean Seas. *Fisheries Research*, 79(1-2), 139-147.
- Zhang, S., Cui, J., Li, C., Mahboob, S., Al-Ghanim, K., Xu, P. & Sun, J. (2014). The complete mitochondrial genome of *Rasbora borapetensis* (Cypriniformes: Cyprinidae: Rasbora). *Mitochondrial DNA*, 27(2), 861-862.

APPENDICES

Appendix 1

Materials and apparatus

The materials and apparatus used in the molecular work were listed in Table 1.

Table 1: Materials and apparatus used

No	Materials and Apparatus	Brand
1	Proteinase K	Amresco
2	Phenol	R&M Chemicals
3	Chloroform : Isoamyl alcohol (Biotechnology grade)	Amresco
4	Agarose powder (Biotechnology grade)	1 st Base
5	1x TAE buffer (pH 8.0)	1 st Base
6	1 kb ladder	Fermentas Gene Ruler
7	6x Loading dye	Fermentas Gene Ruler
8	<i>TaKaRa LA</i> Taq (5 U/ μ L)	<i>TaKaRa</i>
9	2.5 mM dNTPs	<i>TaKaRa</i>
10	10x LA PCR buffer II	<i>TaKaRa</i>
11	Isopropanol	Amresco
12	Tricaine	Nacalai
13	Water bath incubator	Major Science
14	Wise Spin	Wisd Laboratory Instruments
15	T100 TM thermal cycler	Bio-Rad, USA
16	EasyPure PCR purification kit	TransGene Biotech, Beijing
17	Enduro TM Power Supplies for electrophoresis	Labnet International, North America
18	UV transilluminator	MaestroGen

Appendix 2

13 protein coding genes of *Rasbora sumatrana*, *Rasbora hobelmani*, *Rasbora tornieri*, and

Rasbora borapetensis:

>*Rasbora sumatrana*

ATGTTTAAACACCCTAATAAACCACCTAATCAACCCCTTAATCTACATCGTATTTGTTTTATTAGCAG
TAGTCTTTCTAACCCTAGTTGAACGAAAAGTACTAGGATATATACAACCTCCGAAAAGGACCAAAC
GTAGTAGGACCTTACGGCACAATTCAACCAATCGCCGACGGTATTAAACTATTTATTAAGAACCG
ATCCGCCCATCATCATCATCACCCATTTTATTCCTAATCACACCAATCCTTGCACTAACCCTGGCCA
TAATACTATGAGCACCAATACCCCTGCCACACGCAATAACAAACCTAAACCTAGGCATGCTCTTTA
TCCTAGCCCTATCAAGCCTGGCAGTATACTCAATCCTAGGATCAGGATGAGCCTCAAACCTCAAAT
ACGCGCTAGTCGGAGCCCTACGAGCTGTAGCCCAAACAATTTTCATATGAAGTAAGCCTAGGACTA
ATCGTACTATCTATTATAATCTTCTCAGGAGGCTACTCCCTACAAACCCTAAGCACTGCACAAGAA
AAAATCTGACTACTAGTCCCAGCCTGACCACTAGCCACAATATGGTACATCTCAACCCTAGCAGAA
ACCAACCGGGCACCATTTCGACCTAACAGAAGGAGAATCAGAACTAGTATCAGGATTTAATGTAGA
ATACGCAGGAGGACCATTTCGCACTATTCTTCCTAGCTGAATACGCCAACATTCTACTAATAAATAC
CCTATCGGCAGTTCTATTTCTAGGAACATCTTACCTACCAAACACCCCAGAACTATCAACAATCCT
CATCATAACCAAAACCGCCCTACTAGCTGCAATATTCTATGAGTACGAGCATCCTACCCACGATT
CCGATATGATCGACTTATACACCTAATTTGAAAAAACTTCCTCCCAATCACATTAGCCTTTGTCCTA
TGACACACATCCCTACCAATCGCACTAGCAAGCCTACCACCACAACTATAAATGAATCCATACGTA
CTAGCAATTCTACTATTACGCCTAGGACTAGGCACCACCATAACATTTTCCAGCTCACACTGATTA
CTTGCTGAATAGGACTAGAAATTAATACCCTAGCAATCACCCCGCTAATAGCCCAACAACACCAC
CCACGAGCAGTAGAAGCAACCACAAAATATTTCTCATCCAAGCCACGGCAGCAGCAATAATCCT
ATTTGCAAGCACTACAAATGCCTGACTCACAGGAGAATGAAATATCAATAACGTATCCAACCCAA
TCACTAACATAATAATTATTTCTGCACTAGCACTGAAAATTGGACTAGCACCAGCACACTTCTGGC
TGCCAGAAGTTCTACAAGGATTAAACCTACTAACAGGACTAATCCTGTCCACATGACAAAAAATG
GCTCCATTTCGCCCTAATTATTCAAGTAGCCCAAAACACCAACCCGACCCTTCTAACACTATTAGGA
ATCTCATCAACACTAGTGGGAGGATGAGGAGGCCTAAACCAAACCCAGCTGCGAAAAATCTTAGC
CTACTCATCAATCGCCACATAGGGTGAATAATAATTATTATTACATTACGCCCCACAACCTACCAC
CATCGCCCTAATCACTTACATCTTCATAACATCCGCAGCATTCCTCACCCCTAAAAACACTAAACAC
AACAAAAATTAATACACTTTCAACAGCTTGATCAAAAAACCCAACTCTAACAGCAATCACCCCTCT
CGTCTACTATCACTAGGCGGCCTCCCACCAATAACTGGTTTCACACCAAAATGACTCATCATCCA
AGAAGTAGCAAAACAAAACCTCCCCCTTACAGCCACAATCATAGTTTTAACCAGCCCTACTAAGCCT
ATACTTCTACCTACGACTATGCTATACAGCAACACTAACCATTAAACCCCAACACAATTAACCTAAC
CACCCCTGACGAATAAAACAACCACAAATCTCTTACCCTGGCACTAACAATCATTATTACCTT
AGGGCTTCTACCACTAACCCCAACCATCCTAGCACTAACCACCTGTGATAATCACGCGTTGACTCT
TCTCTACTAATCACAAAGACATTGGCACCCCTTTATCTCGTATTTGGTGCCTGAGCCGGAATAGTTG
GAACCGCCCTTAGTCTTCTCATTCTGCTGAACCTCAGCCAACCGGGATCACTTTTAGGGGACGATC
AAATCTATAATGTAATTGTAAGTGCACGCTTCGTAATAATTTTCTTTATAGTTATGCCAATATT
AATCGGGGGATTTGGAAACTGATTGGTCCCACTAATAATCGGAGCACCAGACATGGCATTCCAC
GAATAAACAAACATAAGCTTCTGACTTCTTCCACCATCATTTCTCCTCCTATTGGCCTCTTCTGGCGT
TGAAGCCGGAGCCGGGACAGGGTGAACAGTATACCCACCCTAGCAGGCAACCTTGCCACGCGAG
GAGCATCAGTAGACCTAACAATTTTTTCACTCCACTTGGCAGGTGTATCATCAATTCTAGGGGCTA
TTAATTTTATTACAACAATTATTAATATGAAACCCCCAGCCATCACTCAATACCAAACCCGCTATT
TGTATGAGCAGTGCTAGTTACGGCTGTTTTACTACTTCTATCACTACCTGTGCTAGCCGCTGGAATT
ACAATACTTCTTACAGACCGAAACCTTAATACCACATTCTTCGACCCAGCCGGCGGAGGGGACCCG
ATTCTATATCAACACCTATTTTGATTCTTTGGTCACCCAGAAGTTTATATTCTTATTTTACCTGGGT
CGGAATCATCTCCCACGTCGTAGCTTACTACGCAGGAAAAAAGAACCATTGGGTATATGGGCA
TAGTCTGAGCTATAATGGCCATTGGCCTTCTAGGGTTCATCGTATGAGCCCACCACATATTTACCGT
TGGAATAGACGTAGACACCCGAGCATACTTTACATCCGCAACAATAATTATCGCCATCCCAACAG
GTGTAAGTATTTCAGCTGACTAGCCACACTTCACGGAGGTTCTATTAAATGAGAAACACCAATAC

TATGAGCACTAGGATTTATTTTCCTATTCACAGTAGGTGGACTAACAGGAATTATTTTAGCTAACTC
CTCACTAGATATTGTACTTCATGACACCTATTATGTCGTCGCGCATTTCCACTACGTACTATCTATA
GGTGCCGTATTCGCCATTATAGCAGGCTTTGTCCACTGATTCCCCTACTTTACAGGCTACACTCTAA
GCAGCGCCTGAACAAAGATCCACTTTGGAGTAATGTTTATTGGTGTAATCTTACATTTTCCCCCA
ACACTTCCTCGGACTAGCAGGCATGCCACGACGATACTCTGACTACCCAGATGCCTACGCCCTATG
AAACACAGTATCATCAATTGGGTCACTGATTTCTCTAGTAGCAGTAATTATATTCTTATTTATTATC
TGAGAAGCCTTCGCCGCTAAACGAGAAGTACTATCTATTGAACTAACCTCAACAAATGCAGAATG
ACTTCATGGATGCCCTCCACCATAACCACACATTTGAAGAACCGGCATTTGTTATAACCCAGCCAAA
CTAAATGGCACACCCCACTAGGATTCCAAGATGCAGCATCACCTGCTATAGAAGAGCTTCT
CAGCTTTCATGACCATGCCCTAATAATTGTATTTTAATCAGCACCCCTAGTACTTTACATTATTGTT
GCTATAGTATCAACTAACTAACTAATAAATTTATTCTGGACTCCCAAGAAATCGAAATTTGTATGA
ACAGTACTACCAGCTATCATTCTAATTCTAATCGCCCTTCCATCTCTCCGAATTTTATATCTTATAG
ACGAAATTAATGATCCGCACGTAACAGTAAAAGCCATAGGACATCAGTGATACTGAAGTTATGAG
TATACAGACTATGAAAACCTGGGGTTTGACTCATATATAGTACCCACCCAAGACTTAACCCCGGGA
GGATTCCGACTATTAGAACTGATCACCGAATGGTAATCCCAAAAGAATCCCAATCCGCATTTTA
GTGTCTGCCGAAGATGTCCTCCACTCTTGAGCAGTCCCATCACTAGGTGTTAAAATAGACGCAGTG
CCGGGACGACTTAATCAGGCTGCCTTTATTGTTTCACGACCTGGTATATTTTACGGACAATGCTCTG
AAATCTGCGGGGCTAACCCACAGCTTTATACCAATTGTAGTTGAAGCAGTACCCCTAGAAAATTTTG
AAAAATGATCTTCATTAATGCTAGAGGACGCCTATGCCCAATTGAACCCACTTCCATGGTTTACA
ATCCTAGTCTTTTCATGAATCGTTCTTCTCACCTTTACCCCAACCAAAGTTCTAGACCACATTCAGC
CGAATGAATTTATTCTACTAGACGTTAAAAAATACCAAACCTTAATTGAATCTGACTATGATAAA
TGCCCAATTGAACCCACTTCCATGGTTTACAATCCTAGTCTTTTCATGAATCGTTCTTCTCACCTTT
ACCCCAACCAAAGTTCTAGACCACATTCAGCCGAATGAATTTATTCTACTAGACGTTAAAAAATAC
CAAAACCTTAATTGAATCTGACTATGATAAATGGCCCACCAAGCACACGCATACCACATAGTTGAC
CCAAGCCCATGACCACTAACAGGAGCTGTCGGCGCCCTACTAATAACATCTGGCCTAGCAATCTGA
TTCCACTTTCACTCAATAACACTCATTACCCTTGGACTAACCTTACTAATCCTCACCATAATTCAAT
GATGACGAGACATTATCCGAGAGGGGACATTTTCAAGGTCACCACACACCACCCGTACAAAAAGGA
CTCCGATATGAATAAATTCTCTTTATTACCTCCGAAGTATTCTTTTCTCTGGGCTTTTCTGAGCCTT
TTACCACTCAAGCCTAGCCCCACACCTGAGCTAGGAGGATGCTGACCACCGACAGGAATCATAA
CTTTAGACCCATTTGAAGTCCCCCTACTTAACACAGCTGTCTACTAGCATCTGGGGTAACAGTAA
CATGAGCCCACCACAGCATTATAGAGGGGCAACGAAAACAAGCTATCCAATCTCTCGCACTAACA
ATCCTTCTAGGATTATACTTTACAGCCCTCCAAGCAATAGAATACTATGAAGCACCATTACCATC
GCAGATGGTGTATACGGATCAACATTCTTTGTAGCTACGGGGTTCCACGGACTACACGTGATTATT
GGATCAACCTTCCTAGCTGTGTGCCTCCTACGTCAAATCCAATACCATTTTACATCAGAACATCACT
TCGGCTTTGAGGCCGCAGCATGATATTGACACTTTGTGCGACGTAGTCTGACTATTCCTATATGTATC
CATCTACTGATGAGGCTCATAATGAACCTTATCATAACTATTATTATAATTACACTAATCGTACCAT
CAATCCTAGCATTTCATCTCATTCTGACTTCCCCAGATAGACCCAGATACAGAAAACTATCACCT
ATGAGTGCGGATTTGACCCACTAGGTTCTGCCGGCTTCCATTCTCACTGCAATTTTCTCTAGTGGC
AATCCTATTCTCCTATTTGACCTGGAAATTGCCCTACTCCTCCCCCTCCCATGAGGAGACCAACTT
TATAACCCAACCGAAACACTCTTCTGAGCCACAACAATCCTAATTCTACTCACCTAGGACTAATC
TATGAATGAGCCCAAGGAGGTCTAGAATGAGCAGAAATGACACCAGTACATTTTAGCTTCACATC
GGCATTTACCTTAGGATTAATAGGACTAACATTTAACCGCATACACCTACTCTCAGCACTTTTATGT
TTAGAAGGAATAATACTATCCTTATTTATCGCGCTAGCCCTATGGGCACTCCAATTCGAATCAACA
GGTTTCTCCTCCACCCCTATGCTACTACTGGCCTTCTCAGCCTGTGAAGCAAGCACAGGCCTTGCAC
TACTAGTAGCCACTGCCCCGAACCCATGGAACAGACCACCTACAAAACCTCAACCTCCTACAATGCT
AAATGCTAAAAGTACTAATCCCAACAATTATACTATTTCCCACAATTTGACTTACCCCGCAAAAC
TTCTCTGAACAACACTACAACAATACACAGCTCCCTAATTGCTTTAATCAGCCTAACATGATTAAAAT
GAACATCAGAAACAGGATGAGCCACTTCCAACACATACATGGCCACAGACCCACTATCAACCCCC
TTCCTAACATTAACATGCTGGTTACTCCCACTAATAGTCCTAGCCAGCCAAAACCATATTAACCCA
GAACCAATCAACCGCCAACGCCTATATATCTCATTACTCGCATCCCTCCAACTTTTCTAATCATA
GCATTTCGGTGCCACAGAGATCATTATTTTACATTATATTTGAAGCCACATTAATCCCAACCTA
ATTATCATCACCCGATGAGGAAACCAGGCTGAGCGCCTAAATGCAGGAATTTACTTTTTATTCTAC
ACACTTGAGGTTCCCTACCCCTCTTAATCGCACTGCTACTCCTCCAACAAACCACCGGAACACTA
TCCATACTAATTCTTCAATATGCACAACCAATTGCATTAGACTCCTGAGGTCATAAAATTTGATGA
GCCGGCTGCCTAATCGCATTCTTAGTAAAAAATACCACTCTATGGGGTACATCTGTGATTACCAAAA
GCACATGTTGAAGCCCCAGTAGCAGGATCCATAGTACTAGCAGCAATCCTACTAAAACCTAGGAGG

CTATGGTATGATACGAATAATAATCGTATTAGACGCATTCAATAAAGAACTAACATACCCATTTAT
 CATCCTAGCCCTATGAGGAATCATTATAACAGGCTCAATCTGCCTACGACAAACAGACCTTAAATC
 TTTAATTGCCTACTCATCTGTAAGCCACATGGGCTGGTAGCAGGAGGAATTCTAATTCAAACTAC
 ATGAGGATTCTCAGGAGCAATCGCCCTTATAATTGCCACGGGTAAACATCCTCAATACTATTCTG
 CCTAGCTAACACAACATATGAACGAGTTCACAGCCGAACCATAATTCTAATTCGAGGGTTACAAAT
 AATCTTCCCACCTTGACGAATATGATGATTCAATTGCCAACCTAGCTAATCTAGCACTCCCACCCCTA
 CCAAATCTAGTAGGAGAATTAACAATTATCACACAATATTTAACTGATCCCCATGAACCATCGCA
 CTTACAGGAATAGGAACACTAATTACAGCAAGCTACTCCCTATACATATTCCTAATAACACAACGA
 GGTC AAGCACCAAGCCACATTACAAACCTACCACCCTACCACACCCGAGAACACCTATTAATAGC
 CCTCCACCTAATCCCCATCACCCCTACTAGTAGCAAAACCAGAACTCCTGCTGGGTTGATGTTATTA
 ATGACATATGTATTAACATCGCTCATACTCCTATCGATCACAACCTTAATCTACCCACTACTAAAA
 ACACTCAGCCCCAATCTAAAAAAACCAAACCAACTAGCCACAGATATCAAAACCGCCACAAGCCA
 CTCATTTTACATTAGCCTTACCGCACTTATAATCTTTATTAACCAAGGAGTGGAGAACATCTCTACC
 AGTTGATACTGAACAAGTATTTTTACACTTGACATTTTCGTAGGCTTTAACTTTGACCACTACTCCC
 TTATCTTCACACCAGTCGCATTATTCGTTGCCTGGTCGATCCTAGAATTTGCACTCTGGTACATACA
 CGCAGACCCAGACAAAGACCGATTCTTTAAATACCTACTCCTATTTCTTGTAGCCATAATTACCCT
 AGTCACAGCTAACAACATATTTCAACTATTCATTGGCTGAGAGGGGGTAGGAATTATATCATTCTCT
 CCTAATTGGATGATGATCTGGACGAGCAGATGCTAACACATCAGCCCTTCAAGCCATTATCTACAA
 CCGAGTAGGAGACATCGGATTTATTATGGCCATCGCATGATTTGCCACACAAATAAATACATGAA
 ACATTCAAGAAATCCTTACAGAATCAGAAATACACAACCTCAATAATCCCCTAGCAGGAATTATTT
 TGGCAGCCATAGGAAAATCAGCCCAATTTACTCTCCAGCACCATAGTTGTCTGCTGGAATCTTCTACTAGTTC
 CCAACACCCGTATCTGCCCTACTTCACTCCAGCACCATAGTTGTCTGCTGGAATCTTCTACTAGTTC
 GCCTTCACCCCTTAATACAAAACAATAACACAGCACTAATAGCCTGCCTATACCTCGGCCTAGCAA
 CAACACTATTCTCTGCTGCCTGCGCCCTGACTCAAAATGACATCAAAAAAATCGTAGCCTTCTCAA
 CATCAAGCCAACTAGGACTTATAATAATAGCAATCGGACTAAACCAACCTCAACTAGCATTTTTCC
 ACATCTGCACCCACGCCTTTTTTAAAGCCATACTATTCTTATGCTCAGGCATAATTATTCACAAGCT
 AAAAGACGAACAAGACATCCGAAAAATAGGAAATCTAATAACACTAATACCAACCACCACAACA
 TACCTATTAATTGGCAACATAGCACTGTGAGGAATCCCATTTCTAGCTGGCTTCTACTCAAAAGAC
 GCCATCCTTGAATCCCTAGTCACATCCAAGCTAAGCATCCTTACCATCATCCTTACACTTATCGCCG
 CGTCATTACCCGAGCCTACAGTTACCGACTAATATACCATGTAACCCTAGGACCAGGATTAATCA
 ACCCCAAAATCATACCAACCGAAGACGCAATAACTGTGCTCAAACCTATCAAACGACTTGCCTGA
 GGAAGCATCTTTATAGGATTTGTAATCTGACACAACATACTACCCGAGAAAACACCAACACTAAC
 AATACCTATGTACCTTAACTAACAGCTGTGCGAGTAATCATCATTGGCTTCTCACAGCAAAATG
 AATCTCCACATTAACGAAGAACAATAATAAACACACCAACAGCCATGCTATTTTATTTCATTCAA
 CATACTAGGGTACTGCCCTGCATTAGTTCACCGACTCTTCCCAAACTAAACTAATTTAGGCCA
 CACAATCGGCACAACACTAGACAAAGCATGACAAGTCTTATGAGTAGAAAAAATACTATGAACAT
 TCACTGAAACCGCCACATACCTAAACGACGCTCAACAGGCAAAAATTAACACATACCTAACTATC
 TTCTCCATTTCCATAACAACACTAATCTTAACATACTCCATCATCCTCTAACTAACTGACCGTAAG
 CTGCCACGATCTAAACCACGGGTAAATTCTAACACAACAAACAGCGTTAAGAGCAATATTCAAGC
 ACAAATTACCAATATACCACCGCCAAAAGAATAAACAAACAGCCACACCCTTACATCATTACGCA
 ATACAGAAAATCTTTTAAACCATCAACAACCACCCAATAACCATCATACCACCCACCTCAAAAA
 AACCCAGCCACCAAAACCCACCACAAACAAATACACTAAAACATACTCTGCAACAGAACGACCGCC
 CCAAGCCTCAGGAAAAGGTTCCGCAGCTAAAGCCGCTGAATAAGCAAACACTACAAGCATGCCTC
 CAAGATAAATTAAAAAAAGAATTAAAGACAAAAAAGAGCCCCCACAATAAACTAAAACCCCA
 CCCCACACCAGCTGTAACCTACCAAACCAACAGCCGCAAAATAAGGAGTCGGATTTGAAGCAACCG
 CAACAAGGCCCACAACCTAAGATTACTAATAATAAAGACACTGAATATATCATATGGCAAGCCTAC
 GCAAAACACATCCCCTAGCTAAAATTGTTAACGACGCACTAATTGACTTACCAACACCAGCCAATA
 TTTACAGCATGATGAACTTTGGATCACTACTAGGACTATGCTTAATTACACAAATCCTAACAGGAT
 TGTTCCTAGCTATACACTACACCTCAGACATTTCAACTGCCTTTTCATCAGTAGTCCACATCTGCCG
 AGATGTAAACTACGGATGACTGATCCGAAACATGCATGCCAACGGGGCATCATTCTTCTTTATCTG
 CCTTTACATTACATTGCCCCGAGGCCTATACTACGGCTCATACCTATATAAAGAACTTGAAACAT
 CGGAGTAGTCTGTTCCTACTAGTTATAATAACAGCATTCGTAGGCTATGTACTCCCATGAGGACA
 AATATCATTTTGAGGTGCCACAGTAATTACAAACCTGCTATCCGCAGTCCCATACGTGGGTGATAT
 ACTAGTACAATGAATCTGAGGGGGATTCTCAGTAGACAACGCAACACTCACACGATTCTTCGCATT
 CCACTTCCTCTACCATTCATCGTTGTGGCCATAACAGTCCTCCACCTACTTTTCTTACACGAAACA
 GGATCAAAACAACCCACTTGGCCTTAACCAAATAACAGATAAAATTTCTTTCCACCCATACTTTTCA

TACAAAGACCTCCTAGGGTTTATAATCATGTTACTAGCCCTAGCACTTCTAGCACTATTCTCCCCAA
 ACCTCCTAGGAGACCCAGAAAACCTTACCCCAGCCAACCCGCTAGTCACACCACCACACATTA
 CCAGAGTGATACTTCTATTTGCCTATGCAATCTACGTTCAATCCCAAACAACTAGGAGGAGTC
 CTTGCACTACTATTCTCAATTCTAATCCTACTAGTAGTACCCTCCTTCATACCTCAAAACAACGAG
 GAATAACATTCCGCCCCACTCACCCAATTCTCTTCTGAACCCTTGTAGCAGACATAATCATCCTTAC
 ATGAATCGGGGGTATACCAGTAGAACCCCATTCATCATTATTGGACAAATCGCATCCGTACTATA
 CTTCTCACTATTCTTAATCCTATTCCCCCTAGCAGGATGATTAGAAAACAAAGCACTAGATCGGGA
 CT

>Rasbora hobelmani

ATGTTTAAACACCCTAATAAACCACCTAATCAACCCCTTGATCTACATCGTATTTGTTTTACTAGCAG
 TAGCCTTTCTAACCCTAGTTGAACGAAAAGTACTAGGATATATACAACCTCCGAAAAGGCCCAAAC
 GTAGTAGGACCTTACGGTACAATTCAACCAATCGCCGACGGCATTAACTATTTATTAAGAACC
 ATCCGCCCATCATCATCATCACCCATTTTATTCTAATCACACCAATCCTTGCACTAACCTGGCTA
 TAATGCTATGAGCACCAATGCCCTGCCACACGCAATAACAAACCTAAACCTAGGCATGCTCTTTA
 TCCTAGCCCTATCGAGCCTAGCAGTATACTCAATCCTAGGATCGGGGTGAGCCTCAAACCTAAAAT
 ACGCGCTAGTCGGAGCCCTACGAGCTGTAGCCCAAACAATTTTCATATGAAGTAAGCCTGGGACTA
 ATCGTACTATCTATTATAATCTTCTCAGGAGGATACTCCCTACAAACCCTAAGCACTGCACAAGAA
 AAAATCTGATTACTAATCCCAGCCTGACCACTAGCCACAATATGATACATCTCAACCCTAGCAGAA
 ACCAACCGGGGACCATTCGACCTAACAGAAGGAGAATCAGAGCTAGTATCAGGATTTAATGTAGA
 ATACGCAGGAGGACCATTCGCACTATTCTTCTAGCTGAATACGCCAACATTCTACTAATAAATAC
 CCTATCGGCAGTTCTATTTCTAGGAACATCTTACCTACCAAACACCCCAGAACTATCAACAATCCT
 CATCATAACCAAAACCGCCCTACTAGCTGCAACATTCTATGAGTGCGAGCATCCTACCCACGATT
 CCGATATGATCGACTTATACACCTAATTTGAAAAAACTTCCTCCCAATCACATTAGCCTTCGTCCTA
 TGACACACATCCCTACCAATCGCACTAGCAAGCCTACCACCACAACCTATAAATGAATCCATATGTA
 CTAGCAATTCTACTATTCAGCCTAGGACTAGGCACCACCATAACATTTTCCAGCTCACACTGATTA
 CTTGCCTGAATAGGACTAGAAATTAATACCCTAGCAATCACCCCACTAATAGCCCAACAACACCAC
 CCACGAGCAGTAGAAGCGACCACAAAATATTTCTCATCCAAGCCACAGCAGCAGCAATAATCCT
 ATTTGCAAGCACTACAAATGCCTGACTTACAGGAGAATGAAATATCAATAACATATCCAACCCAA
 TCACTAACATAATAGTTATTTCTGCACTAGCACTAAAAATTGGACTAGCACCAGCACACTTCTGGC
 TGCCAGAAGTTCTACAAGGATTAAACCTACTAACAGGACTAATCCTGTCCACATGACAAAAACTG
 GCCCCATTGCCCCTAATTATTCAAGTAGCCCAAAACACCAACCCAACCCTCCTAACACTATTAGGA
 ATCTCATCAACGCTAGTGGGAGGATGAGGAGGCCTAAACCAAACCCAACCTACGAAAAATCCTAGC
 CTACTCATCAATCGCCACATAGGATGAATAATAATTATTATTACGCCCCACAACCTACCAC
 CATCGCCCTAATCACTTACATCTTCATAACATCCGCAGCATTCCTCACCTTAAAAACACTAAACAC
 AACAAAAATTAATACACTTTCAACAGCTTGATCAAAAAACCCAACCTCTAACAGCAATCACCCCCCT
 CGTCCTACTATCACTAGGCGGCCTCCCACCAATAACTGGTTTCGCACCAAAATGACTTATCATCCA
 AGAACTAGCAAAACAAAACCTCCCCCTTACAGCCACAATCATAGTTTTAACCGCCCTACTAAGCCT
 ATACTTCTACCTACGACTATGCTACACAGCAACACTAACCATTAACCCTAACACAATTAACCTAAC
 CACCCCTGACGAACTAAAACAACCACAAATCTCTTACCCTTAACACTAACAATTATTGTTACCTT
 AGGGCTTCTACCACTAACCCCAACCATCCTAGCACTAACCACCTGTGATAATCACGCGTTGACTCT
 TCTCTACTAATCACAAAGACATTGGCACCCCTTTATCTCGTATTTGGTGCCTGAGCCGGAATAGTTG
 GAACCGCCCTTAGTCTTCTCATTCTGTGCTGAACCTCAGCCAACCGGGATCACTTTTAGGGGACGATC
 AGATCTATAATGTAATTGTAACCTGCCACGCCTTCGTAATAATTTTCTTTATAGTTATGCCAATATT
 AATCGGGGGATTTGGAACTGATTGGTCCCACTAATAATCGGAGCACCAGACATGGCATTCCCAC
 GAATAAACAACATAAGCTTCTGACTTCTTCCGCCATCATTTCTCCTTCTATTGGCCTCTTCTGGCGT
 TGAAGCTGGAGCCGGGACAGGGTGAACAGTATACCCACCCTAGCAGGCAACCTTGCCACGCAG
 GAGCATCAGTAGACCTAACAATTTTTTCTACTCCACTTGGCAGGTGTATCATCAATTCTAGGGGCTA
 TTAATTTTATTACAACAATTATTAATATGAAACCCCCAGCCATCACTCAATACCAAACCCCGCTATT
 TGTATGAGCAGTGCTAGTTACGGCTGTTTTACTACTTCTATCACTACCAGTGCTAGCCGCTGGGATT
 ACAATACTTCTTACAGACCGAAACCTTAATACTACATTCTTCGACCCAGCCGGCGGAGGAGACCCA
 ATTCTATATCAACACCTATTTTGATTCTTTGGTCACCCAGAAGTTTATATTCTTATTTTACCTGGGT
 CGGAATCATCTCCACGTCGTAGCTTATTACGCAGGAAAAAAGAACCATTGGATATATGGGCAT
 AGTCTGAGCTATAATGGCCATTGGCCTTCTAGGGTTCATCGTATGAGCCACCACATATTTACCGTT
 GGAATAGACGTAGACACCCGAGCATACTTTACATCCGCAACAATAATTATTGCCATCCCAACAGGT

GTAAAAGTATTCAGCTGACTAGCCACGCTTCACGGAGGTTCTATTAAATGAGAAACGCCAATACTA
 TGAGCACTAGGATTTATTTTCCTATTACAGTAGGTGGACTAACAGGAATTATTTTAGCTAACTCCT
 CACTAGATATTGTACTTCATGACACCTATTATGTCGTCGCGCACTTCCACTACGTATTATCTATAGG
 TGCCGTATTCGCCATTATAGCAGGCTTTGTCCACTGATTCCCACTATTTACAGGCTACACCCTAAGC
 AGCGCCTGAACAAAGATCCACTTCGGAGTAATGTTTATTGGTGTAAATCTTACATTTTTCCCCCAA
 CACTTCCTCGGACTAGCAGGCATGCCACGACGATACTCTGACTACCCAGATGCCTACGCCCTATGA
 AACACAGTATCATCAATTGGATCACTGATTTCTCTAGTGGCAGTAATCATATTCTTATTATTATCT
 GAGAAGCCTTCGCCGCTAAACGAGAAGTACTATCTATTGAACTAACCTCAACAAATGCAGAATGA
 CTTTCATGGGTGCCCTCCACCATAACCACACATTTGAAGAACCAGCATTTGTTATAACCCAGCCGAAC
 TAAATGGCACACCCCACTAGGATTCCAAGATGCAGCATCACCTGCTATAGAAGAGCTTCTC
 AGCTTTCATGACCATGCCCTAATAATTGTATTTTTAATCAGCACCTTGGTACTTTACATTATTGTTG
 CTATAGTATCAACTAACTAACTAATAAGTTTATTCTGGACTCCCAAGAAATCGAAATTGTATGAA
 CAGTACTACCAGCTATCATTCTAATTCTAATCGCCCTTCCATCTCTCCGAATTTATATCTTATAGA
 CGAAATCAATGATCCACACGTAACAGTAAAAGCCATAGGACATCAGTGATACTGAAGTTATGAGT
 ATACAGACTATGAAAACCTGGGATTTGACTCATATATAGTACCTACCCAAGACTTAACCCCGGGAG
 GATTCCGACTATTAGAACTGACCACCGAATGGTAATCCCAAAAGAATCCCAATCCGCATTTTAG
 TGTCTGCCGAAGATGTCTCCACTCTTGAGCAGTCCCATCACTAGGTGTTAAAATAGACGCAGTAC
 CGGGACGACTTAATCAGGCTGCCTTTATTGTTTCACGACCTGGTATATTTTACGGACAATGCTCTGA
 AATCTGCGGGGCTAACCACAGCTTTATACCAATTGTAGTTGAAGCAGTACCCCTAGAAAATTTTGA
 AAAATGATCTTCATTAATGCTAGAGGACGCCTATGCCCCAATTGAACCCACTTCCATGGTTTACAA
 TCCTAGTCTTTTCATGAATCGTTCTTCTCACCTTCACCCCAACCAAGTTCTAGATCACATTCAGCC
 GAATGAATTTATTCTACTAGACGTTAAAAAATACCAAAACCTTAATTGAATCTGACTATGATAAAT
 GATAATCAGCCTCTTCGACCAATTTGCAAGCCCCCACTCATTGGTATCTCACTAGTCTTCATTGCA
 CTATTAGTACCACTACTCTTCCCAAACTCATTCCACGACTAAACAACAACCGATTCTTTACAACCT
 AGACATGATTAATTAACAAGTTTCGCCAGCCAACTAATAGCTCCACTAAATATTGGCGGACACAAA
 TGAGCCCTTTTATTCACCTCACTAATATTATATCTTATCTCAATAAACGTACTTGGACTACTCCCAT
 ACACTTTTACACCCACAACACAATTATCAATAAATATGGGATTTGCAGTACCACTATGACTTGCCA
 CAGTAATTGTTGGTATACTAAATCAACCAACATCTTCTTTAGGTCATCTTCTACCAGAGGGAACAC
 CAACCCTCCTAATCCCAGTTCTAGTAATTATTGAAACACTTAGCTTATTTATCCGACCACTAGCCCT
 GGGTGTTCGACTTACAGCCAATCTAACCAGCAGGCCACCTACTTATTCAACTTATCTCAATAGCCGT
 ACTTGTCTTACTACCTATAATGCCAGCAGTAGCACTTCTGACCTCTACAATTCTTGTTATACTTACA
 CTTCTAGAGGTTGCAGTAGCTATAATTCAAGCCTACGTTTTTATTCTGCTATTAAGCCTATATCTTC
 AAGAAAACACCTAATGGCCCAACCAAGCACACGCATACCACATAGTTGACCCAAGCCCATGACCAC
 TAACAGGAGCTGTGCGCGCCCTACTAATAACATCTGGCCTAGCAATCTGATTCCACTTTCACTCAA
 TAACACTCATTACCCTTGGAATAACCCTACTAATCCTCACCATAATTCAGTGATGACGAGACATTA
 TCCGAGAGGGGACATTTAGGGTCACCACACACCACCCGTACAAAAAGGACTCCGATATGGAATA
 ATTCTCTTTATTACCTCCGAAGTATTCTTTTTCTGGGTTTTTTCTGAGCCTTTTATCATTCAAGCCT
 AGCCCCCACACCTGAGCTAGGAGGATGCTGACCACCGACAGGAATCATAACCTTAGACCCATTTG
 AAGTCCCCTTACTTAACACAGCTGTCTACTAGCATCCGGGGTAACAGTAACATGAGCCCACCACA
 GCATTATAGAGGGGCAACGAAAACAAGCTATCCAATCTCTCGCACTAACAATCCTTCTAGGATTAT
 ACTTTACAGCCCTCCAAGCAATAGAATACTATGAAGCACCATTACCATCGCAGATGGCGTATACG
 GGTCGACATTCTTTGTGGCTACAGGGTTCCACGGACTACACGTGATTATTGGATCAACCTTCCTAG
 CTGTGTGCCTCCTGCGTCAAATCCAATACCATTTTACATCAGAACATCACTTCGGCTTTGAGGCCGC
 AGCATGATATTGACACTTTGTGACGCTAGTCTGACTATTCTATATGTATCCATCTACTGATGAGGC
 TCATAATGAACCTTATCATAACTATTATTATAATTACACTAATCGTACCATCAATCCTAGCATTTCAT
 CTCATTCTGACTTCCCCAGATAAACCCAGATACAGAAAACTATCACCTATGAATGCGGATTTGA
 CCCACTGGGTCTGCCCCGACTTCCATTCTCACTGCAATTTTTCTAGTGGCGATCCTATTCTCTCTGT
 TTGACCTGGAAATTGCCCTACTCCTCCCCCTCCCATGAGGAGACCAACTTTACAACCCAACCGAAA
 CACTCTTCTGAGCCACAACAATCCTAATTCTACTACCCTAGGACTAATCTATGAATGAGCCCAAG
 GAGGTCTAGAATGAGCAGAATATGACACCAGTACATTTTAGCTTCACATCGGCATTTACCTTAGGA
 TTAATAGGACTAACATTTAACCGCATACACCTACTCTCCGCATTTTATGTTTAGAGGGAATAATA
 CTATCATTATTTATCGCACTAGCCCTATGGGCACTCCAATTCGAATCAACAGGTTTCTCCTCCACCC
 CCATGCTACTACTGGCCTTCTCAGCCTGTGAAGCAAGCACAGGCCTTGCACTACTAGTAGCCACTG
 CCCGAACCTCATGGAACAGACCACCTACAAAACCTCAACCTCCTACAATGCTAAATGCTAAAAGTA
 CTAATCCCAACAATTATACTATTTCCCACAATTTGACTTACCCCCGAAAACCTTCTCTGAACAACCTA
 CAACAATACACAGCTCCCTAATTGCTTTAATTAGCCTAGCATGATTAAATGAACATCGGAAACAG

GGTGAGCCACTTCCAACACATACATGGCCACAGACCCACTATCAACCCCATTCTTAACACTAACAT
GCTGATTACTCCCCTAATAATCCTAGCCAGCCAAAACCATATTAGCCCAGAACCAATCAACCGCC
AACGCCTATATATCTCACTACTCGCATCCCTCCAACTTTCTAATCATAGCATTCCGGTGCCACAGA
GATCATTATATTTTACATTATATTTGAAGCCACATTAATCCCAACCCTAATTATCATCACCCGATGA
GGAAACCAGGCTGAGCGCTAAATGCAGGAATTTACTTTTTATTCTACACACTTGCAGGTTCCCTA
CCCCCTCTAATCGCACTGCTACTCCTCCAACAAACCACCGGAACACTATCCATATTAATTCTTCAAT
ATGCACAACCAATTGCATTAGACTCCTGAGGTCATAAAATTTGATGAACCGGCTGCCTAATCGCAT
TCCTAGTAAAAATACCACTCTACGGGGTACATCTGTGATTACCAAAAGCACATGTTGAAGCCCCAG
TAGCAGGATCCATAGTACTAGCAGCAATCCTACTAAAACCTAGGAGGCTATGGTATGATACGAATA
ATAATCGTATTAGACGCATTCAATAAAGAACTAACATACCCATTTATCATCCTAGCCCTATGAGGA
ATCATTATAACAGGCTCAATCTGCCTACGACAAACAGACCTTAAATCTTTAATTGCCTACTCATCT
GTAAGCCACATGGGCCTAGTGGCAGGAGGAATTCTAATTCAAACTACATGAGGATTCTCAGGAGC
AATCGCCCTTATAATTGCCACGGACTAACATCCTCAATACTATTCTGCCTGGCTAACACAACATA
TGAACGAGTTCACAGCCGAACCATAGTTCTAATTCGAGGGTTACAAATAATCTTCCCACTTGCAGC
AATATGATGATTCAATTGCCAACCTAGCTAATCTAGCACTCCACCCCTACCAAACCTAGTAGGAGA
ATTAACAATTATCACAACAATATTTAACTGATCCCCATGAACCATCGCACTCACAGGAATAGGAAC
ACTAATTACAGCAAGCTACTCCCTATACATATTCCTAATAACACAACGAGGTCAAGCACCAAGCCA
CATCACAACCTACCACCCTACCACACCCGAGAACACCTATTAATAGCCCTCCACCTAATCCCCAT
CATCCTACTAGTAGCAAAACCAGAACTCCTGCTGGGTTGATGTTATTAATGACATATGTATTAACA
TCGCTCATACTCCTATCAATCACAACCTTAATCTACCCACTACTAAAAACACTTAGCCCCAATCTA
AAAAAGCCAAGCCAAGCTAGCCACAGATATCAAAACCGCCACAAGCCACTCATTTTACATTAGCCT
TACCGCACTTATAATCTTTATTAACCAAGGGGTGGAGAACATCTCTACCAGTTGATACTGAACAAG
TATTTTTACACTTGACATTTTTCGTAGGCTTTAACTTTGACCATTACTCCCTTATCTTTACACCAGTCG
CATTATTTGTTGCCTGATCGATCCTAGAAATTTGCACTCTGATACATACACGCAGACCCAGACAAAG
ACCGATTCTTTAAATACCTACTCCTATTTCTTGTAGCCATAATTACTCTAGTCACAGCCAACAACAT
ATTTCAACTATTCAATTGGCTGAGAGGGGGTAGGAATTATATCATTCCCTCCTAATTGGATGATGATC
TGGACGAGCAGATGCTAACACATCAGCCCTTCAAGCCATTATCTACAACCGAGTAGGAGACATCG
GATTTATTATGGCCATCGCATGATTTGCCACACAAATAAATACATGAAACATTCAAGAAATCCTTG
CAGAATCAGAAATACACAACCTCAATAATCCCATTAGCAGGAATTATTTTGGCAGCCATAGGAAAA
TCAGCCCAATTTACTCTCCACCCGTGACTCCCAGCAGCCATGGAAGGTCCAACACCCGTATCTGCC
CTACTTCACTCCAGCACCATAGTTGTCGCTGGAATCTTCTACTAGTTGCGCTTCACCCCTTAATAC
AAAACAATAACACAGCACTAACAGCCTGCCTATACCTCGGTCTAGCAACAACACTATTCTCCGCTG
CCTGCGCCCTGACTCAAAATGACATCAAAAAAATCGTAGCCTTCTCAACATCAAGCCAAGTAGGA
CTTATAATAATAGCAATCGGACTAAACCAACCTCAACTAGCATTTTCCACATCTGCACCCACGCC
TTTTTTAAAGCCATACTATTCTTATGCTCAGGTATAATTATTCACAAGCTAAAAGACGAACAAGAC
ATCCGAAAAATAGGAAATCTAATAACACTTATACCAACCACCACAACATACCTATTAATTGGCAA
CATAGCACTGTCAGGAATCCCATTCTAGCCGGCTTCTACTCAAAAGACGCCATCCTTGAATCCCT
AGTCACATCCAAGCTAAGCATCCTTACCATCATCCTTACACTTATCGCTGCATCATTCACCGCAGCC
TACAGTTACCGACTAATATACCATGTAACCCTAGGACCAGGATTAATCAACCCCAAAATCATACCA
ACCGAAAACGCAATAACTGTACTCAAACCTATCAAACGACTTGCCTGAGGAAGCATCTTTATGGG
ATTTGTAATCTGACACAACATACTACCCGAAAAAACACCAACACTAACAATACCTATGTATCTTAA
ACTAACAGCTGTGCGAGTAATCATCATTGGTTTCCTCACAGCAAAATGAATCTCCACACTAAACGA
AGAACAATAATAAACACACCAACAGCCATGCTATTTTATTCATTCAACATACTAGGGTACTGCC
TGCATTAATTCACCGACTCTTCCCAAACTAAAACTAACCTTAGGCCACACAATCGGCACAACACT
AGACAAAGCATGACAGGTCCTATGAGTAGAAAAAATATTATGAACATTCACTGAAACCGCCACAT
ACCTAAACGACGCTCAACAGGCAAAAATCAAAACATACCTGACTATCTTCTCCATTTCCATAATAA
CACTAATCTTAACATACTCCATCATCCTCTAACTAACTGACCGTAAGCTGCCACGATCTAGACCA
CGGGTTAACTCTAACACAACAAACAACGTTAAAAGCAATATTCAAGCACAAATTACCAATATACC
ACCGCCAAAAGAATAAACAAACAGCTACACCCTTACATCATTACGCAACACAGAAAACCTCTTTTA
AACCATCAACAACCACCCAATAACCATCATACCACCCACCTCAAAAAAACCCAGCCACCAAAACCC
ACCACAAACAAATATACTAAAACATACTCCACAACAGAACGACCGCCCCAAGCCTCAGGAAAAGG
TTCCGCAGCTAAAGCCGCTGAATAAGCAAAACACTACAAGCATGCCTCCGAGATAAATTAATAAAAA
GAATTAAGACAAAAAAGAGCCCCACAATAAACTAAAACCCACACCCACACCCAGCTGTAAACC
ACCAAACCAACAGCCGCGAAATAAGGAGTCGGATTTGAAGCAACCGCAACAAGGCCCAACACTA
AGACTACTAGTAATAAAGACATTGAATATATCATATGGCAAGCCTACGCAAAACACACCCCTAG
CTAAAATTGTTAACGACGCACTAATTGACTTACCAACACCAGCCAATATTTTCAGCATGATGAACT

TTGGATCACTACTAGGACTATGCTTAATTACACAAATCCTAACAGGATTGTTCTAGCTATGCACT
ACACCTCAGACATTTCAACTGCCTTTTCATCAGTAGTCCACATCTGCCGAGATGTAAACTACGGGT
GACTGATCCGAAACATGCATGCCAACGGGGCATCATTCTTCTTTATCTGCCTTTACATTACATTGC
CCGAGGCCTATACTACGGCTCATACTATATAAAGAACTTGAAACATCGGAGTAGTCCTGTTCTT
ACTAGTTATAATAACAGCATTTCGTAGGCTATGTACTCCCATGAGGACAAATATCATTTTTGAGGTGC
CACAGTAATTACAAACCTGCTATCCGCAGTCCCATACGTGGGTGATATACTAGTACAATGAATTTG
AGGGGGGTTCTCAGTAGACAACGCAACACTCACACGATTTTTTCGCATTCCACTTCCTCCTACCATT
CATCGTTGTGGCCATAACAGTCTCCACCTACTTTTTCTTACACGAAACAGGATCAAACAACCCACT
TGGCCTTAACCAAATAACAGATAAAATTTCTTTCCACCCATACTTTTCATATAAAGACCTCCTGGG
GTTTGTAATCATATTATTAGCCCTAGCACTTCTAGCACTATTCTCCCCGAACCTCCTAGGAGACCCA
GAAACTTCACCCCAGCCAACCCGCTAGTCACACCACCACATATTAAACCAGAGTGATACTTCCTA
TTTGCCTACGCAATCCTACGTTCAATCCCAAACAACTAGGAGGAGTCCTTGCACTACTATTCTCA
ATTCTAATCCTACTTATAGTACCACTCCTTCATACCTCAAACAACGAGGAATAACATTCCGCCCA
TTCACCCAATTCTCTTCTGAACCCTTGTGGCAGACATAATCATCCTTACATGAATCGGGGGTATAC
CAGTAGAACACCCATTTCATCATTATTGGACAAATCGCATCCGTACTATACTTCTCACTATTCTTAAT
CCTACTCCCCCTAGCAGGATGATTAGAAAACAAGCACTAGATCGGGACT

>Rasbora tornieri

ATGCTTAACACCCTAATTAATCACCTAATCAACCCCTAATCTATATCGTATTTGTCTTACTAGCAG
TAGCCTTTCTAACCCTGATCGAACGAAAAGTACTAGGATATATACAGCTCCGAAAAGGCCCAAC
GTAGTAGGACCTTACGGCACAATTCAACCAATCGCCGATGGTGTTAAATTATTTATTAAGAACCA
GTCCGCCCATCATCATCATCACCCATTCTATTCTAACTACACCAATCCTTGCACTAACCCTAGCCA
TAATATTATGAGCGCCAATGCCCTACCACACGCAATAACAAACCTCAACCTAGGCATACTTTTTTA
TTTTAGCCCTATCAAGCCTAGCAGTATACTCAATTTTAGGATCAGGATGAGCATCAAATTCAAAAT
ACGCACTAGTCGGAGCTCTACGAGCCGTAGCCCAAACAATCTCATATGAAGTAAGCCTAGGACTA
ATTGTACTATCCATCATAATCTTCTCAGGGGGTTATTCCCTACAAACCTTAAGCACAGCACAAGAA
AAAATTTGACTACTCATCCAGCTTGACCACTAGCTACAATATGATACATCTCAACCCTAGCAGAA
ACTAACCGAGCACCATTGACCTAACAGAAGGTGAATCAGAAGTATCCGGGTTTAATGTAGA
ATACGCAGGAGGACCATTGTCATTATTCTTTCTAGCTGAATACGCCAACATTCTATTAATAAACAC
CCTATCAGCAATTTTGTTCCTAGGTACATCATATTTACCAGACACCCCTGAACCTATCTACAATCTAT
ATCATAACTAAACTGCCTACTGGCCACAATATTTCTATGAGTACGAGCATCCTACCCACGATTC
CGATATGATCGACTCATACACTTAATCTGAAAAAATTCCTACCAATCACACTAGCCTTCGTCTTA
TGACATACAGCCTTACCAATCGCATTAGCAAGCCTACCACCACAACCTATAAATGAATCCATACGTA
CTAGCAATTCTACTATACAGCTTAGGACTAGGCACCACAATAACATTCGCCAGCTCACACTGACTA
CTCGCCTGAATGGGACTGGAATCAACACCCTAGCAATCACTCCATTAATAGCCCAACAACATCAC
CCACGTGCAGCAGAAGCCACCACAAAATACTTCCTCATCCAAGCCACAGCAGCAGCAATAATCCT
GTTTCGAAGCACAACAAATGCCTGACTAACAGGAGAATGAAACATTAACAATACATCCAACCCAA
TAACTAACATAATAATCATCTCTGCCCTAGCACTAAAAATTGGACTAGCACCAGCACACTTCTGAC
TACCAGAAGTTCTACAGGGATTAAACCTACTAACAGGACTAATCCTATCTACCTGACAAAAACTAG
CCCCACTTATCCTAATAATTCAAGTAACCCAAAAACCAACCCCAACCCTCCTAACACTACTAGGAA
TCTTATCGACACTAGTAGGAGGATGAGGGGGTCTAAACCAAACCCAGTTACGAAAAATCCTAGCC
TACTCATCAATTGCCACATAGGATGAATAATAATTATTATCCACTACACCCCACTTACCATA
ATTGCCCTAATTACCTACATTTTTATAACATCTGCAGCATTCTTACCCTAAAAACATTAAACACAA
CAAAAATTAATACACTTTCAACAGCATGATCAAAAAACCAACCCTAACAGCAATTACCCCCCTTA
CTATATTATCACTAGGAGGCCTACCACCAATAACCGGATTTGTACCAAAATGACTAATCATTCAAG
AACTAGCAAAAACAAAACCTCCCCCTCACAGCTACAATCATAGTTTTAGCCGCCTTACTAAGCCTAT
ACTTTTACCTACGACTATGTTATACAGCAACACTAACCATCAACCCAAACACAATTAACGCAACAA
CCCCATGACGAATAAATCAACCCAAAACCTTTTACCCTTAACAGTCACAATCACCACCACACTAG
GATTCCTACCAATAACCCCAACTATCCTCATACTAACCGCTTGTGATAATCACGCGTTGATTCTTCT
CTACTAATCACAAAGACATTGGCACCCCTTATCTCGTATTCCGGTGCCTGAGCTGGAATAGTTGGAA
CCGCCCTTAGCCTTCTTATCCGTGCTGAGCTTAACCAACCGGGTCTCTTTTAGGAGATGACCAAA
TTTATAATGTAATTGTAAGTGGCCATGCTTTCGTAATAATTTTCTTTATAGTTATACCAATGTTAATT
GGGGGGTTTGAAATTGATTAGTTCCCTGATAATCGGGGCACCAGACATAGCATTCCCACGAATA
AACAACATAAGTTTCTGACTTCTTCCCCCATCATTTCTTTTATTGTTGGCCTCTTCTGGCGTCGAAG
CTGGGGCCGGAACAGGGTGAACAGTATACCCACCCTTGCAGGTAATCTTGCCACGCAGGAGCA

TCAGTAGACCTAACAATTTTTTCACTTCACCTAGCAGGTGTATCATCTATTTTAGGGGCTATCAATT
TTATTACAACCATCATTAACATAAAACCCCCAGCCATTACTCAGTACCAAACCTCCACTATTTGTGT
GAGCAGTACTAGTCACAGCTGTACTCTTACTTCTATCCCTACCAGTTTTAGCCGCTGGGATTACAAT
ACTGCTCACAGATCGAAATCTTAACACCACATTCTTCGACCCAGCCGGAGGAGGAGACCCAATCCT
ATATCAACACCTATTCTGATTCTTTGGTCACCCAGAGGTTTACATTCTTATTTTACCAGGATTCCGA
ATTATCTCTCACGTTGTAGCCTACTACGCAGGAAAAAAGAACCCTTTGGGTATATGGGCATGGTC
TGAGCTATAATAGCTATCGGCCTCTTAGGATTCATTGTATGAGCCCACCACATATTTACCGTCGGA
ATAGACGTAGACACCCGTGCATATTTACATCCGCAACAATAATTATTGCTATTCCACGGGCGTT
AAAGTATTTAGCTGACTAGCTACGCTCCACGGAGGCTCAATTAATGAGAAACACCCATACTATG
AGCACTAGGATTTATTTTCCTATTACAGTAGGTGGACTAACAGGAATCATCTTAGCCAACTCCTC
ATTAGACATTGTACTCCACGACACCTATTACGTTGTTGCACACTTCCATTATGTATTATCAATGGGT
GCCGTATTCGCCATCATAGCAGGCTTTGTCCACTGATTCCCACTATTTACAGGATATACCCTAAAC
AGCGCATGAACAAAAATCCACTTCGGAGTAATATTTATCGGTGTAAACCTCACATTCTTCCCACAA
CACTTCCTAGGACTAGCAGGAATACCACGACGATACTCTGACTACCCAGATGCCTACGCCCTATGA
AACACAGTTTCATCCATTGGGTCAATTAATTTCTTTAGTAGCAGTAATTATATTCCTATTTATTATTG
AGAAGCCTTTGCCGCTAAACGAGAAGTACTATCTATTGAAATAACCGCAACAAATGCAGAATGAC
TCCATGGTTGTCTCCACCATAACCACACATTTGAAGAACCAGCATTTCGTTATGACCCAACCAAAC
AAATGGCACACCCACCCAATTAGGATTCCAAGATGCAGCCTCACCTGTTATAGAGGAACTTTTAA
ACTTCCATGATCATGCCCTAATAATTGTACTCTTAATTAGCATCTTAGTACTTTACATTATTGCCGC
AATGGTATCAACTAACTGACCAATAAATTCATCCTAGATTCCCAAGAAATTGAAATTGTGTGGAC
TGCTTACCAGCTATTATCCTAATCTCAATTGCTCTTCCATCCCTACGAATCTTATACCTCATAGAC
GAAATTAACGACCCCAACGTAAACAGTAAAAGCCATAGGGCACCAATGATATTGAAGCTATGAGTA
TACAGATTATCATAATATAGGATTTCGACTCCTACATAATCCCAACTGAGCTCCTAGCCCCAGGAGA
ATTCCGACTCTTAGAGACTGACTACCGAGTAGTGCTCCCAAAAGAAGCCCCAGTCCGCGTCCTAGT
ATCTGCTGAAGATGTCTCCACTCATGAGCTGTCCCATCACTAGGCATTAATAATAGATGCAGTACC
CGGACGACTCAACCAAATTGCCTTTATCATTTTACGTCCTGGTGTATTTTATGGACAATGCTCTGAA
ATTTGCGGAGCTAACCATAGCTTCATACCAATTGTAATTGAAGCAGTACCACTAGAATATTTTGAA
TTTTGATCATTATTAGTATACGCCTATGCCCAATTGAATCCACTTCCTTGATTCACAATCCTAATT
TTTTCATGAGCCATCCTTCTCACTTTTACACCAACCAAAGTTCTAAACCACACCCAACCAAATGACT
TCATCTTACTAGACATTAAAAAATATCAAGACCTAACTGAATCTGACTATGATAAATGATAATTA
GCCTTTTCGATCAATTTGCAAGCCCTAAGTTTATAGGAATTTCACTAATCTTTATTGCACTTATCGT
ACCGCTAATCTTTTTTCCCAAACCTCACTTCCACGATGAACCAATAACCGACTCTTCACAACCTCAAAC
ACTGTTAATTAACCGATTGTAAATCAACTAATAACCCCACTAAACATTAACGGACATAAATGAGC
CCTCTTATTTACCTCACTAATCTTATATCTTATATCAATAAATTTACTTGGATTACTCCCATACACTT
TTACACCCACAACACAACCTATCAATAAATATAGGATTTGCAGTACCACTATGGCTTGCCACAGTAA
TTATTGGCATACTTAATCAACCCACATCCTCTCTAGGACATCTCCTCCCAGAAGGAACACCAACCC
TCCTCATTCAGTCCTAGTAATTATTGAAACAATCAGCCTATTTATTTCGACCATTGGCCCTAGGTGT
ACGACTCACAGCCAACCTAACCGCAGGTCACCTACTTATTCAACTTATCTCAACAGCAGTATTTGT
TTTAACATCACTATTACCGACCGTAGCATTCCTGACTTCTACAATTCTTGTTATACCTTACGCTCTTA
GAAGTTGCAGTAGCAATAATTCAGGCCTATGTCTTTATCCTCCTACTAAGCCTGTACCTACAAGAA
AACATTTAATGGCTCACCAAGCACACGCATACCACATAGTTGACCCAAGCCCATGGCCTTTAACAG
GAGCTATAGGTGCTCTACTAATAACATCTGGCCTAGCAATCTGATTCCACTTTAACTCAATAACAC
TCATTATCCTAGGACTAACCCCTACTTATCCTTACCATAATCCAATGATGACGAGACGTAATCCGAG
AAGGAACATTCCAAGGCCACCACACCCAACCCGTACAAAAAGGACTCCGATATGGAATAATTCTC
TTCATTACCTCCGAAGTATTCTTTTTCTAGGATTTTTCTGAGCCTTTTATCACTCAAGCCTAGCCCC
CACCCCGAACTAGGAGGATGCTGACCACCAACAGGAGTCATTACCCTAGACCCATTTGAAGTCC
CCTTACTCAACACAGCTGTCCTACTAGCATCCGGGGTAACAGTAACATGAGCTCACCACAGCATTA
TAGAAGGAGAACGAAAACAAGCCATCCAATCCCTCGCACTAACAATCCTCCTAGGACTTTACTTCA
CAGCCCTCCAAGCAATAGAGTACTACGAAGCACCATTTACCATCGCAGACGGAGTGACGGATCA
ACATTTTTTGTAGCCACAGGATTCCACGGACTACATGTTATTATTGGGTCAACTTTTCTAGCCGTAT
GCTTCCTACGTCAAATCCAATATCACTTCACATCAGAACATCACTTCGGCTTTGAAGCTGCAGCAT
GATACTGACATTTGTTGACGTAGTTTGACTTTTCTCTACGTATCCATCTACTGATGAGGCTCATA
GTGAATCTCATTATAACCATTATCATTATTACACTAATTGTACCATCAATCTTAGCATTTATCTCAT
TTTGACTCCCCCAAATAGACCCAGACACAGAAAACTATCGCCCTACGAATGCGGATTTGATCCAC
TAGGATCTGCCCGACTCCCATTTTCACTACAATTTTTTCTAGTGGCAATCTTATTCCTATTGTTTCGAC
CTAGAAATCGCCCTTCTCCTCCCCCTCCCATGAGGAGATCAACTTTATAACCCAACCGAAACACTC

TTCTGAGCTACAGCAATCCTAATTCTACTCACCTAGGATTAATCTATGAATGAACCCAAGGAGGC
 TTAGAGTGGGCAGAATATGACACCAGTACATTTTAGCTTCACATCAGCATTTACTTTAGGACTAAT
 AGGACTGACATTTAACCGCATACACCTACTCTCTGCACTTCTGTGCTTAGAAGGAATAATATTATC
 ATTATTTATTGCACTAGCCCTGTGAGCACTACAATTTGAATCAACAGGCTTCTCCTCTACCCCAATA
 CTACTACTAGCTTTCTCTGCCTGTGAAGCAAGCACAGGTCTTGCACTATTAGTAGCCACTGCTCGA
 ACCCATGGAACAGACCACCTACAAAACCTCAACCTCCTACAATGCTAAATGCTAAAAGTACTAATT
 CCAACAATCATATTATTCCCCACAATTTGACTCACCCCCGCAAAATTCTTATGAACAACCACAACA
 ATGCACAGCTCCCTAATTGCCCTAATTAGCCTGACATGACTAAAATGAACATCAGAAACAGGATG
 AGCCGCTCCAACATATTTATAGCCACAGACCCATTATCAATCCCCCTCCTAACACTAACGTGCTG
 ACTACTTCCACTAATAATTCTAGCTAGTCAAAACCACATCAATCCAGAACCAATTAACCGCCAACG
 TCTGTACATCTCATTACTTGCGTCCCTTCAAACCTTTTCTAATCATAGCATTTAGTGCCACAGAAATC
 ATTATGTTTTATATTATATTTGAAGCCACACTAATCCCAACCTTAATTATTATCACCCGATGAGGAA
 ACCAAGCCGAACGTCTAAATGCAGGAATTTACTTTTTATTTTACACACTCGCGGGTTCTCTACCACT
 CCTAATTGCACTTCTACTCCTTCAACAAACCACCGGAACACTATCCATGCTAATCCTCCAATACGC
 ACAACCAATCGCACTTGACTCTTGAAGCCATAAAATTTGATGAGCCGGATGCTTAATCGCATTTCT
 AGTAAAAATACCACTCTATGGGGTACACCTATGACTACCCAAAGCACATGTAGAAGCCCCAGTAG
 CAGGATCAATAGTACTAGCAGCAATTTTACTAAAAGTAGGAGGATATGGTATAATACGAATAATA
 ATTGTAAGTACGCACTTCAACAAAGAACTGGCATAACCCATTCAATATCTTGGCCTTATGAGGAATT
 ATCATAACAGGATCAATCTGCCTACGACAAACAGACCTAAAATCTTTAATTGCCTACTCATCTGTGA
 AGTCACATAGGTCTAGTAGCAGGAGGGATTTTAAATCCAAACCCCATGAGGATTTTCAGGAGCAATT
 GCCTTAATAATTGCCACGGACTAACATCCTCAATATTATTTTGCCTCGCCAACACAACATATGAA
 CGAATCCACAGCCGAACCATAATTCTCATTGAGGGCTACAAATAATCTTTCCATTAAACAGCAACA
 TGATGATTCAATGCTAACCTAGCTAACCTAGCGCTTCCACCCCTACCAAACCTAGTAGGAGAACTA
 ACGATCATTACAACAATATTCAACTGATCTCCATGAACCATCGCACTTACAGGAATAGGGACACTA
 ATTACAGCAAGCTACTCTCTATACATATTCTTAATAACACAACGAGGTCAAATACCAAACCACATT
 ATAAACCTCCCACCTATCACACCCGAGAACACCTATTAATAGCCCTTACCTAGTCCCCATTATTT
 TACTCGTAGCAAAACCAGAACTCCTACTAGGATGATGCTATTAATGACATACACATTAACATCACT
 CATACTCCTAGCAATCACTACCTTAATCTACCCACTACTAAAAACACTACACCCTAAACCAAAAAA
 ACCAAACCAACTAGCCATAGACATTAAAACCGCCGTAAGCCACTCATTCTATATCAGCTTTGCCGC
 ACTCATAATTTTCATTAAACCAAGGAGCGGAAAACATCTCCACTAGCTGATACTGAATAAGCATTTA
 TACACTTGATGTCTTCGTGGGCTTTAATTTGACCACTACTCCCTAATTTTTTACACCAGTTGCATTA
 TTCGTGCGCTGATCGATCCTAGAGTTTGCCTCTGATATATACACGCAGACCCGAATAAAGACCGA
 TTCTTCAAATACCTCCTTCTATTCTCGTAGCCATAATCATCCTAGTCACAGCCAACAACATATTTT
 AATTATTCATCGGCTGAGAGGGAGTAGGAATTATATCATTCTTCTAATTGGATGATGATCAGGAC
 GAGCAGATGCTAACACATCAGCCCTCCAAGCCATTATCTATAACCGAGTAGGAGACATTGGATTTA
 TTATAGCCATTGCATGATTTGCCACACAAATAAACACATGAAATATTCAAGAAATCCTAACAGAAT
 CAAAAACACATAACTCAATAATCCCACTAGCAGGCATTATTCTAGCAGCTATAGGAAAATCTGCC
 AATTTACACTCCACCCATGACTTCCAGCAGCCATAGAAGGCCCTACACCCGTATCTGCCCTACTCC
 ACTCCAGCACCATAGTTGTGCTGGGATCTTCTACTAATCCGCCTTACCCAATTATACAAAATA
 GTAACACAGCACTAACAGCTTGCTTATACCTTGTTTAAATAACAACACTATTTTCCGCTGCCTGCGC
 CCTAACCCAAAATGACATTAAAAAAATCGTAGCCTTCTCAACATCAAGTCAACTAGGACTAATAAT
 AATAGCCATCGGCCTAAACCAACCCCAACTAGCATTCTTCCATATCTGCACCCATGCATTTTTTAA
 AGCCATACTATTTTTATGCTCAGGTATAATTATCCACAACTAAAAGATGAACAAGACATCCGAAA
 AATAGGAAATCTATTAACCCTAATACCAACCACCACAACATACCTATTAATTGGTAACATAGCACT
 ATCAGGAATTCCCTTCTAGCCGGCTTTTATTCAAAGATGCCATTCTAGAATCCCTAGTCACATCT
 AAATAAGCATTCTTACAATTACCCTCACACTTATCGCCGCATCATTACCGCAGCTTATAGCTATC
 GACTAATATATCACGTAACCTTAGGACCAGGACTAATTAACCCAAAAATTATACCAACCGAAGAT
 GCAATAACAGTACTTAAGCCCATCAAACGACTTGATGAGGAAGCATTTTGATAGGATTTGTAATC
 TGACATAATATTCTCCCTGAAAAAACCAACCTTAACAATACCCCTACACCTTAACTAACAGCT
 ATTGCAGTGATTATTATTGGCTTCTCACAGCAAAATGAATTTCTATATTAGATCAAAAAACAAATC
 AAAAATACACCAATATCCATACTATTCTACTCATTCAAGTATGCTAGGGTACTGTCCAACACTAATT
 CACCGACTTTTTCCAAAACCTAAACTAACCTAGGCCACACAATCGGTACAGCACTAGACAAAAC
 GTGACAAGTTTTATGAGTAGAAAAAATACTATGAACACTAACTAAAACCACTACATACCTAAACG
 ACGCTCAAAAAGCAAAAATTAACACATACCTAACTATCTTCTCCATCTCCTTAACCATCCTAATCT
 CAACATACTCCATCCTCTCTAACTAACTGACCGTAAGCTGCCACGATCTAAACCACGAGTCAGC
 TCCAATACAACAAACAACGTTAAAAGCAACACCCAGGCACAAATCACCAACATACCCCCACCAAA

AGAATAAACACAGCCACACCACTCATATCGTTACGTAATACAGAAAACCTCTTTAAGACCATCAA
CAACCACCCAATAACCATCGTACCAACCCCCCAAAAAACCACGCCACCAAACCTACTCCAAAC
AAATATATTAGAACATACTCTGCAACTGAACGTCCACCTCAAGCTTCCGGAAGGATCAGCAGC
TAAAGCCGCTGAATAAGCAAATACTACAAGCATGCCCCAAGATAAATCAAAAAAGAACCAGTG
ACAAAAAAGAACCCCCACAACAACTAAAACTCCACACCCGACACCAGCCGTAACCACCAAGCCA
ACAGCCGCAAAATTTGGGGTAGGATTAGAAGCAACTGCAACAAGACCTACAACCTAAAACCACCA
CAACAAAGACATTGAATATATCATATGGCAAGCCTACGAAAAACACACCCCACTATTAAAAATCGC
TAACGACGCGCTAATTGATCTACCAACCCCAGCTAGCATCTCAGCATGATGAAACTTTGGATCCCT
ACTAGGACTATGCCTAATAGCCCAAATCCTAACAGGACTATTCTAGCCATACACTACACCTCAGA
TATCTCAACCGCTTTCTCATCAGTGGTCCATATCTGCCGAGATGTAAACTACGGATGAATAATCCG
TAATATTTCATGCCAACGGAGCATCATTCTTTTTTATCTGCCTATATATCCACATTGCCCGAGGCCTA
TACTACGGCTCATACCTATATAAAGAAACCTGAAACATTGGAGTAATCCTCTTCTTACTAGTCATG
ATGACAGCATTTCGTAGGTTATGTACTTCCATGAGGACAAATATCATTCTGAGGCGCTACAGTAATT
ACAAACCTATTATCCGCGAGTACCCTACGTGGGTGACACATTAGTCCAATGAATCTGAGGTGGATTC
TCCGTAGATAACGCAACACTCACACGATTCTTCGCATTTTCCTTCTTACCATTATTGTAGCAG
CTATAACAATCCTCCACCTACTCTTCTTACACGAAACAGGATCAAACAACCCAATAGGCCTCAACC
CAAACGCAGACAAAGTCTCTTTCCACCCATACTTCTCATATAAAGACCTTTTAGGATTTGTAATTAT
ATTATCAGCCCTATCATTCTAGCACTATTTTCCCCAAACATATTAGGAGACCCAGAAAACCTTAC
CCCAGCCAACCCCTCTAGTTACACCCCTCACATTAAACCAGAATGATATTTCTATTTGCCTACGCC
ATCCTACGCTCAATTCCAAACAACTAGGAGGAGTCTCGCACTACTATTCTCAATCCTAATCCTA
TTAGTAGTACCATTCTCCACACCTCAAACTACGAGGAATAACATTCCGTCCACTTACCCAATTC
CTCTTCTGAGTACTTGTAGCAGACATAATTATCCTCACATGAATTGGAGGAATACCAGTAGAACAA
CCATTCAATTATTATCGGGCAAATCGCATCAGTATTATACTTCTCACTCTTCTAATTCTATTCCCCCT
AGCAGGATGATTAGAAAATAAATCTATAAACTAA

>Rasbora borapetensis

ATGCTAAACACCCTATTAAATCACCTAATTAATCCCCTGATCTATATTGTATTTGTTCTACTAGCAG
TAGCTTTTTTAACCCTGGTTGAACGGAAAGTATTGGGGTATATACAACCTCGAAAGGGCCCAAACG
TAGTGGGACCTTATGGTGTAATCCAACCAATTGCCGATGGAGTGAAGCTATTTATTAAGAACCAA
TCCGCCCATCATCATCACCAATTTATTTTATTTAGTTGCACCCATCCTTGCCTAACCTTAGCCAT
AATACTATGAACCCCAATACCCCTCCCACACGCAATAATAAATCTAAATCTAGGGATACTATTTAT
TCTAGCCCTATCAAGCCTAGCAGTATATTCAATTCTAGGATCAGGGTGAGCATCAAACCTCAAATA
TGCCTAGTGGGAGCCTTACGAGCAGTAGCCCAACAATTTTCATACGAGGTAAGCTTGGGGTTGA
TCGTATTATCAATTATAATGTTCTCCGGGGGGTACTCCCTACAAACCCTATCCACAGCACAAGAAA
AAATCTGATTACTAATTCCAGCCTGACCACTAGCGACAATGTGATACATCTCAACCCTAGCGGAAA
CCAACCGAGCACCATTTCGACCTAACAGAAGGAGAATCAGAATTAGTATCAGGCTTTAATGTAGAG
TATGCAGGAGGACCCTTCGCACTATTTTTTCTAGCTGAATACGCCAACATTCTACTAATAAACACC
TTATCAGCAATTATATTCTAGGAACATCATACCTACCAAACGCCCCAGAGCTATCAACAATTTTC
ATTATAACCAAAGCCGCACTATTAGCTGCAGTATTCTTGTGAGTACGAGCATCTTACCCACGATTC
CGATATGACCAGCTTATACACTTAATTTGAAAGAGCTTCCTTCCAATTACACTAGCCTTTGTCTAT
GACACACAGCCCTACCAATTGCACTAGCAAGCCTGCCACCACAACCTATAAATGAACCCTTACGTAC
TAGCAATTTTACTATGCAGCCTAGGATTAGGAACCACTATAACTTTTGCCAGTTTCACTGACTAC
TTGCCTGAATAGGATTAGAAATTAACACCTTGGCAATCGCCCCATTAATAGCCCAACAACACCACC
CACGCGCAGTAGAAGCAACCACAAAATATTTCTTATTCAAGCCACAGCAGCAGCGATAATCCTG
TTTGCAAGCACAACAAATGCATGACTCATAGGAGAATGAAACATTAATAACGTATCAAACCCAAT
AACAAATATAATAATTATTTCTGCACTAGCACTAAAAATTGGATTAGCACCAGCACACTTCTGACT
GCCAGAAGTCTTACAGGGCCTAAACCTAACTACAGGACTAATCCTGTCTACCTGACAAAACTAG
CCCCATTTGCCCTAATTATTCAAGTGGCCCAAAACACAAACCCAACCTTCTAACACTACTAGGAG
TGTTATCAACACTCATTGGAGGGTGAGGAGGACTAAACCAAACCCAGCTACGGAAAATCTTAGCT
TACTCATCAATTGCTCACATAGGATGAATAATAATTATTACTACGCCCCACAACCTTACTATTA
TTGCCCTAATTACTTACATTTTCATAACAGCAGCAGCATTCTCACGTTAAAAACACTAAATACGA
CAAAGATTAATACACTATCTACATCCTGATCAAAAAACCAACCTTAACGGCAGTAACCCCACTAA
TCCTATTATCACTAGGGGGATTACCACCAATAACCGGATTCGCACCAAAATGGCTTATTATCCAAG
AACTAACAAAACAAACCTTCCCCTCACAGCCACAATTATAATTTTAACCGCCCTATTAAGCCTTT
ACTTTTATTTACGACTGTGCTATACAGCAACACTAATTAACCCCTAACACAATCAACGCAACTG

CCCCCTGACGAACTAAAACGACCCAAAACCTGTTGCCATTAACATTAACAATTGTTATTACTTTAG
 GACTTCTACCAATAACCCCAACTATTCTAACACTAACCACCTGTGACAATCACGCGTTGATTCTTCT
 CTACTAACCAACAAAGATATTGGCACCCCTTTATCTCGTATTTCGGTGCCTGAGCCGGAATAGTTGGAA
 CTGCCCTTAGTCTTCTCATCCGTGCTGAACTCAGCCAACCGGGATCACTCCTAGGGGATGACCAAA
 TTTATAACGTAATCGTAACTGCCCATGCTTTTCGTAATGATTTTTCTTTATAGTTATACCAATATTAAT
 CGGAGGATTGGAACCTGATTAGTCCCACTAATAATTGGGGCACCAGATATAGCATTCCCACGAAT
 AAATAATATAAGTTTCTGACTTCTTCCACCATCATTTCTTCTACTATTGGCTTCTTCCGGTGTTGAA
 GCCGGGGCCGGAACAGGGTGAACAGTATACCCACCACCTCGCAGGTAATCTTGCTCACGCAGGGGC
 ATCAGTAGATTTAACAATTTTTCTCTCCACTTAGCAGGTGTATCATCAATTTTAGGGGCTATTAAC
 TTCATCACAACAATTATTAACATGAAGCCCCCAGCCATCACCCAATATCAAACCCCAATTATTTGTTT
 GAGCAGTGTTAGTCACAGCTGTATTATTACTACTATCACTACCAGTTTTAGCCGCTGGAATCACAA
 TGCTTCTTACAGACCGAAACCTAAATACCACATTCTTCGACCCAGCCGGAGGAGGAGACCCAATTC
 TATATCAACACCTATTCTGATTCTTTGGTCACCCAGAAGTTTATATTCTTATTCTACCAGGATTG
 AATTATCTCCCACGTCGTAGCCTATTATGCAGGAAAAAAGAACCATTTGGGTATATGGGTATAGT
 ATGGGCTATGATGGCTATTGGCCTACTAGGATTTATTGTATGAGCTCACCACATGTTTACCGTTG
 AATAGACGTAGACACCCGCGCATATTTTACGTCCGCAACAATAATTATTGCCATCCCAACAGGGGT
 AAAAGTATTTAGCTGATTAGCCACACTGCATGGAGGTTCAATTAAATGAGAAACACCAATACTAT
 GGGCACTAGGGTTTATCTTCTTATTTACAGTAGGAGGATTAACAGGAATCATTCTAGCTAATTCAT
 CACTAGACATTGTTCTTCATGATACTTATTACGTAGTTGCACACTTCCATTACGTATTATCAATAGG
 TGCCGTATTTGCCATTATAGCAGGATTTGTCCACTGATTCCCCCTATTTACAGGATACACCCTAAGT
 AGCGCTTGAACAAAAGTTCACTTCGGGGTAATATTTATTGGGGTAAATCTCACATTCTTCCCTCAA
 CATTTCCCTCGGATTAGCAGGTATACCACGACGATACTCCGATTACCCCGATGCCTACGCCTTATGA
 AATACAGTGTCATCTATTGGGTCACTAATCTCTTTAGTAGCAGTAATTATATTCTTATTTATCATTT
 GAGAAGCCTTCGCGCTAAGCGAGAGGTGCTATCTGTTGAATTAACCGCAACAAATGTGGAGTGA
 CTTTCATGGCTGCCCTCCACCTTATCACACATTTGAAGAACCAGCATTTCGTTATAGCTCAGCCAAAC
 AATTAAATGGCACACCCCACTAGGATTTCAAGATGCAGCATCACCTGTAATAGAAGAGCT
 TCTCGGCTTCCATGATCACGCCCTAATAATTGTATTTTAATTAGCACCTGGTACTTTACATTATT
 GTCGCAATAGTATCAACCAAACTAACCAATAAATTTATTTTAGATTCCCAAGAAATTGAAATTGTA
 TGAACAGTCTTACCAGCAATTATCCTAATTTTAATTGCCCTCCCATCCCTCCGAATTTTATACCTTA
 TAGACGAACTAATGACCCACACGTAACAGTAAAAGCCATAGGACATCAATGATACTGAAGCTAT
 GAATATACAGACTATGAAAATTTAGCATTTGACTCATATATAGTACCAACCCAAGACTTAACTCCA
 GGGGGTTTCCGACTCTTAGAACTGATAACCGAATAGTAGTCCCCAAAGAATCCCCAATTCGTGTT
 CTAGTGTCTGTGAAGACGTCCTTCACTCCTGAGCAGTCCCATCACTAGGCGTCAAATAGACGCA
 GTACCAGGACGACTAAACCAAACCTGCCTTTATTGTTTCACGACCAGGAGTATTCTATGGACAATGC
 TCTGAAATCTGTGGAGCTAATCACAGCTTTATACCAATTGTAGTTGAAGCAGTACCACTAGAAAAT
 TTTGAAAAGTGATCTTCATTAGTACTAGAATACGCCTATGCCTCAATTGAATCCACTTCCATGATTC
 ACAATTTTAATCTTTTCATGAGCCGTTCTTCTCACTTTTACACCAACCAAAGTTCTAGATCATATTC
 AACCAAATGAATTTATTTTACTAGACGTTAAAAAATATCAAAACCCTAACTGAACCTGACTATGAT
 AAATGATAATTAGCCTATTCGACCAGTTTGCAAGCCCTAACTAATAGGAATTTCACTAATCTTTA
 TTGCACTCCTAGTACCACTACTTTTCCCAAACCTCACTCCACGATGAACTAATAACCGATTCTTCAC
 AACCCAATCATGACTAATTAACCGATTTCGTCAACCAACTAATAACACCCCTAAACATTGGAGGAC
 ACAATGAGCCCTTTTATTACCTCATTAACTTTACCTTATCTCAATAAATTTAATAGGATTACT
 CCCATACACTTTTACACCTACAACACAATTATCAATAAATATAGGATTGTCAGTGCCACTATGACT
 AGCCACAGTAATCATTGGCATACTAAATCAACCAACAGCCTCTTTAGGCCATCTCCTACCAGAAGG
 AACACCAACCCTCCTGATCCCAATCTTGGTAATTATTGAAACAATTAGCCTATTAATCCGACCACT
 GGCCTTAGGTGTACGACTTACGGCTAATTTAACTGCGGGCCACTTACTTATTCACTTATTTCAACA
 GCCGTATTTGTTTTATTATCAATAATACCAGTAGTAGCATTAGTGACCTCAACAGTCCTTGTATGC
 TTACACTTTTAGAAATTGCAGTAGCAATGATTCAGGCCTATGTTTTATTCTTCTACTAAGCCTTTA
 CCTCCAAGAAAACATCTAATGGCCCCACCAAGCACACGCATACCATATAGTTGACCCAAGCCCATG
 ACCACTAACAGGAGCTGTAGGCGCCCTACTAATAACATCTGGCCTAGCAATCTGGTTCCATTTTCA
 CTCAATAACACTAATTATTCTTGACTAATTCTATTAGTTCTCACAATAATTCAATGATGACGAGAT
 ATTGTCCGGGAAGGGACATTCCAAGGACACCACACCCCAACCCGTACAAAAAGGGCTCCGATATGG
 AATAATCCTTTTTATTACCTCAGAAGTATTCTTTTTCTAGGCTTTTTCTGAGCCTTTTACCACTCAA
 GTTTAGCCCCCACACCCGAACCTAGGAGGGTGCTGACCACCAACAGGGATTACAACATTAGACCCA
 TTCGAAGTTCCCTTACTCAACACAGCTGTTCTATTAGCATCAGGAGTGACAGTAACATGAGCTCAT
 CACAGCATTATAGAAGGTGAACGAAAACAAGCAATCCAATCACTCACACTAACAATTTTATTAGG

ACTGTATTTTACAGCTCTTCAAGCAATAGAGTATTACGAAGCACCATTACAAATCGCAGACGGAGT
 ATATGGGTCAACATTCTTTGTGGCCACGGGGTTCCACGGATTACATGTTATTATTGGGTCAACCTTT
 TTAGCTGTTTGCCTTCTACGACAAATCCAATACCACTTTACATCAGAACATCACTTCGGCTTTGAAG
 CCGCAGCATGGTACTGACACTTTGTTGACGTAGTTTGACTATTCTCTACGTATCAATCTACTGATG
 AGGCTCATAATGAGCCTTATTATAACAATTATTATAATCGCACTAGTTGTGCCATTAATTCTGGCAT
 TTATTTCAATTCTGATTACCTCAAATAGACCCAGACACAGAAAACTATCACCATACGAATGCGGAT
 TTGACCCACTAGGATCCGCCCCGACTTCCATTTTCACTTCAATTCTTCCTAGTAGCTATCTTATTCCTC
 CTATTTGATCTAGAAATTGCCCTTCTCCTTCCCCTTCCATGAGGAGACCAACTCTACAACCCAACCG
 AAACACTATTCTGAGCCACAGCAATCCTAATTCTACTCACTCTAGGATTAATATATGAGTGAGCCC
 AAGGAGGCCTAGAATGAGCAGAATATGACACCCGTACATTTTAGCTTCACATCAGCATTTACTTTA
 GGACTAATAGGATTAACATTTAATCGCACACACCTACTCTCCGCACTCCTATGCTTAGAAGGAATA
 ATATTATCCCTATTTATTGCATTAGCCCTATGAGCACTACAATTTGAATCAACAGGTTTTTCAACCA
 CCCCCATATTACTTCTTGCAATTCTCCGCTGTGAAGCAAGCACAGGCCTTGCACTATTAGTAGCTAC
 AGCCCGAACCCATGGTACAGACCACTTACAAAACCTCAACCTACTACAATGCTAAATGCTAAAAG
 TACTAATCCCAACAATTATACTATTTCCCACAATTTGACTTACCCCTGCAAACTTCTATGAACAAC
 TACTACAATACACAGCTCCTTAATTGCCCTAATTAGTCTGGCATGGTTAAATGAACATCAGAAAC
 AGGGTGGACTACCTCTAATACATATATAGCTACAGACCCATTATCAACCCCATTTCTAACACTAAC
 ATGTTGATTACTTCCCCTAATAATTTTAGCAAGTCAAAATCATATTAACCCAGAACCTATTAACCGT
 CAACGTCTATACATCTCACTGCTAGCATCTCTCCAACTTTCTAATTATAGCATTGGGCGCCACAG
 AAATCATTATATTTTATATTATATTTGAAGCCACACTAATTCCAACCTCTAATTATTATTACCCGGTG
 AGGAAATCAGGCTGAGCGATTAAACGCAGGAATTTACTTTTTATTTTATACACTCGCAGGCTCCCT
 TCCTCTATTGATCGCACTATTAATTTTACAGCAGACCACCGGAACACTATCTATATTAATTCTCCAA
 TATACACAACCACTCGCTCTAGATTCTGAGGCCATAAAATCTGATGGGCAGGATGTCTAATTGCA
 TTCCTAGTAAAAATACCACTTTATGGGGTACACCTTTGACTACCCAAAGCACATGTTGAAGCTCCA
 GTAGCAGGATCAATAGTACTAGCAGCCATCTTACTAAACTAGGAGGTTACGGCATAATACGAAT
 AATAATTGTATTAGACGCATTTAATAAAGAACTGGCATAACCCATTTATTATCTTAGCCCTTTGGGG
 GATTATTATAACAGGGTCAATTTGCCTGCGGCAAACTGACCTTAAATCTTTGATCGCTTACTCATCT
 GTAAGTCACATAGGCCTAGTAGCAGGAGGAATTTCTAATCCAAACCCCATGAGGATTTTCAGGGGC
 AATCGCCTTAATAATTGCCACGGACTAACATCCTCAATACTATTCTGCCTAGCCAACACAACATA
 CGAACGGGTACACAGTCAACCATATCCTTATTTCGAGGAATACAAATAATTTTTCCATTAACAGC
 AACATGATGATTCATTGCCAACTTAGCCAACCTAGCACTCCCACCACTACCTAATCTAGTAGGAGA
 GCTAACAATTATTACAACAATATTTAACTGATCCCCATGAACCATCGCACTTACAGGAATGGGAAC
 ACTAATTACAGCAAGTTACTCCCTATACATATTCCTAATAACACAACGAGGACAAATACCAACCA
 TATTATAAATCTGTACCTTATCACACCCGAGAACCTTATTAATAGCCCTTCACCTAATTCCTATT
 ATTCTACTCGTAGCAAAACCAGAACTCCTGCTAGGATGATGTTATTAATGACATTTATATTAACAT
 CACTTATACTATTATCAATCACAACCTTGATTTATCCACTACTAAAAACACTCAGCCCTAAACTAA
 AAAAACCAAAACCAACTAGCTATAGATATTAATAAACGCCGTAAAGCCATTCAATTCTACATCAGCCTTA
 CCGCACTACTAATCTTTATTAATCAAGGAGTGCAAAACATCTCTACCAGCTGATACTGAACAAGTA
 TTTATACACTTGATATCTTTGTAGGATTTAACTTCGACCCTACTCTCTCATCTTTACACCCGTTGCA
 CTTTTGTGCGCTGATCAATTCTAGAGTTTGCACTCTGATATATGCACGCAGACCCAAACAAAGAC
 CGATTCTTTAAATACCTCCTCCTATTTCCTGTAGCTATAATTATCCTAGTTACAGCTAACAACATGT
 TTCAGCTATTCATTGGCTGAGAAGGGGTTGGAATTATATCTTTCTATTAATCGGATGATGGTCCG
 GACGAGCAGACGCTAATACATCAGCCCTCCAAGCCATCATTTATAATCGAGTAGGGGACATCGGA
 TTCATTATGGCCATCGCATGATTCGCCACACAAATAAATACATGAAATATTCAGGAAATCTTAACA
 CAATCAGAAACACATAATTCAATGCTCCATTGGCAGGAATTATCCTAGCAGCCATGGGAAAATC
 AGCCCAATTTACACTTCACCCATGACTCCCAGCAGCCATAGAAGGTCCAACGCCCCGTATCTGCCCT
 ACTACACTCCAGCACCATAGTAGTTGCTGGAATTTTCTACTAGTTTCGCTCCACCCAATAATACA
 AAATAACAATTCAGCACTAATAGCTTGCTATACCTTGGCCTAGCAACAACGCTGTTCTCCGCTGC
 CTGCGCCCTGACCCAAAATGATATCAAAAAATCGTAGCCTTCTCAACATCAAGCCAACTAGGCCT
 AATAATAATAGCCATTGGACTAAATCAACCTCAACTAGCATTCTTCCACATCTGTACCCATGCCTTT
 TTTAAAGCCATACTATTCTTATGCTCAGGTATAATTATTCATAAGTTAAAAGATGAACAAGATATC
 CGAAAAATAGGAACTTAATAATACTAATAACCAACAACCACAACATATTTACTAATTGGAAACAT
 AGCATTATCAGGAATACCCCTCCTAGCTGGTTTCTACTCAAAAAGACGCCATCCTCGAATCTCTAGTT
 ACATCCAAACTAAGCATCCTCACCGTTATCCTTACACTCATCGCCGCATCATTCACCGCAGCTTATA
 GCTACCGGTTGATATATCACGTAACATTAGGACCAGGTTTAATTAACCCCAAAATTATACCAACCG
 AAAACGCAATAACAGTACTTAAGCCAATTAACGACTTGCTTGAGGAAGTATTTTTATAGGATTTG

TAATATGACACAACATACTACCTGAAAAAACACCAACACTGACAATACCTTTATACCTAAAACTA
ACAGCCATTGCAGTGATTATTATTGGATTTCTTACAGCAAAATGAATTACCACATTAGATGAAAAA
CAAATCAAAAACACACCAATGGCCATACTATTTTACTCATTGAGCATGTTAGGGTATTGTCCAACC
CTCCTCCACCGACTATTCCCAAACTAAAACCTTACACTAGGCCACACAATCGGAACAATACTAGAT
AAAGCATGACAAGTCTTATGGTTAGAAAAAATATTATGAACACTAACCAGCCATTACATACCTT
AACGACGCCCAACAAGCAAAAATTAACACATACTGACCATTTTTTCAACTTCCCTGGCCACACTA
ATCCTAGTATACATTATCCTCTAACTAACTGACCGTAAGCTGCCACGATCTAAGCCACGCGTTAA
TTCTAACACAACAACAGCGTCAAAAGCAGCACCCAAGCACAAATTACCAACATACCACCGCCAA
AAGAATAAACAACAGCTACACCGCTTACATCATTACGCAAAACAGAAAACCTCTTTTAAACCATCA
ACAACTACCCAGTAACCATCATATCACCCCCCTCAAAAAAACCCAGCCACTAAACCCACCCCAAA
TAAATATGCTAAAACATACTCTACAACAGAACGACCACCCCAAGCTTCAGGGAAAGGCTCAGCAG
CTAAAGCCGCTGAATAAGCAAAAACCTACAAGCATGCCACCAAGATAAATCAAAAAAAGGATTAA
AGATAAGAAAGAACCCCCACAACAACTAAAACCCCGCACCCACACCTGCCGTAACCTACCAAC
CAATGGCCGCAAAGTAAGGGGTTGGATTTGAGGCAACTGCAACAAGACCCATAATTAATAATTACC
AATAGTAAAGACGTTGTATATATCATATGGCAAGCCTACGAAAAACACACCCACTAATTAATAATT
GCTAACGACGCATTAATTGATCTACCAACCCAGCTAATATTTTCACTATGATGAAACTTTGGGTCT
CTACTAGGACTATGCTTAATTACCCAAATCCTGACAGGACTATTCTAGCTATACATTACACCTCTG
ACATCTCAACTGCCTTTTCATCAGTGGTTCACATCTGCCGAGACGTAAATTGTGGATGACTAATCC
GAAACATGCATGCCAACGGAGCATCATTCTTTTTTATTTGCCTATATATTCACATTGCCCGAGGCCT
ATATTACGGCTCCTACTTATACAAAGAACTTGAAACATCGGAGTGGTTCTATTTCTACTAGTCAT
AATAACAGCATTTCGTGGGCTATGTACTCCCATGAGGACAAATATCATTCTGAGGGGCCACAGTAAT
CACAAACCTATTATCCGCAGTACCATATGTAGGGGATATACTAGTACAATGAATTTGAGGCGGATT
CTCCGTAGATAATGCAACACTCACACGATTCTTTGCATTCCACTTCATCTTACCATTCAATTGTTGCA
GCTATAACAATTCTTCACCTGCTGTTCTTACACGAAACAGGGTCAAACAACCCAATAGGCCTTAAC
CCAAATGCAGATAAAATCTCCTTCCATCCATATTTTTCATACAAAGACCTTCTAGGATTTGTAATCA
TACTATCAGCCTTATCAATATTAGCATTATTCTCCCCAACCTCCTAGGAGATCCAGAAAACCTTCAC
CCGAGCCAACCCCTTAGTTACTCCCCACACATTAAACCAGAGTGATACTTTTTTATTGCTTACGCC
ATTCTACGTTCAATTCCAAATAAACTAGGGGGTGTTCCTTGCACTATTATTTTCAATCCTAGTCTTAT
TTGTAGTACCCCTACTACACACCTCAAAACAACGAGGAATAGTGTTCCGCCCATTAACCCAATTCC
TATTTTGAACCCTTGTAGCAGATATAATCATTCTCACGTGAATTGGAGGAATACCAGTAGAGCACC
CATATATTATTATTGGACAAATTGCATCTGTACTTTACTTCTCACTATTCCTGATTTTATTCCCACTA
GTAGGGTGACTAGAAAATAAAACACTATACCAGGACT