Faculty of Resource Science and Technology

Complete Mitochondrial DNA Sequence of Rasbora: Mitogenome Characterization and Phylogenetic Implications

Jill Sela Roja Anak Entalai

Master of Science
2020

Complete Mitochondrial DNA Sequence of Rasbora: Mitogenome Characterization and Phylogenetic Implications

Jill Sela Roja Anak Entalai

A thesis submitted

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

(Molecular Biology)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2020

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

……

Signature

Name: Jill Sela Roja Anak Entalai

Matric No.: 15020868

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date : January 2020

ACKNOWLEDGEMENT

Above all, I would like to thank God for his blessing and give me strength, good health and wisdom throughout my years of study.

And millions thanks to my supervisor Dr Chung Hung Hui for his endless guidance and advices during my research. He consistently helped me to troubleshoot whenever I faced any problem in my research and writing this thesis. I am very lucky to have a supervisor like him because he freely teaches and shares his valuable knowledge and I learn a lot from him.

I would also like to thank my fellow lecturers and laboratory assistances who allowed me to used their equipment or instrument and provided some chemicals and reagents to run my experiments. My experiments were completely successful because of their contribution.

I would like to thank my fellow labmates especially post-graduate (Cynthia, Leonard and Aimi), undergraduates (Final year students) and friends who were directly and indirectly involved to help me throughout my years of research. I really appreciate their kindness and companionship during my study.

Finally, I would like to express my very profound gratitude to my parents and other family members for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

ABSTRACT

Rasbora are freshwater fishes that belongs to Cyprinidae family whereby most of its member share similar morphology which makes them difficult to be differentiated between species. In this study, both the morphometry analysis and mitochondrial DNA approach were used to resolve the phylogenetic relationship between these species. The complete sequence of mt-genome for *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp. were 16,574 bp, 16,581 bp, 16,709 bp and 16,518 bp respectively. The structure of all study species mtDNA was characterized using bioinformatics analysis. *Rasbora* mt-genome comprises of a standard set of 22 tRNAs, two rRNAs, 13 PCGs and two typical non-coding regions. The AT-skew and GC-skew values in the whole genome of *Rasbora* were 0.141 and -0.268 respectively, revealing that the H-strand had the amount of C and A were more prevalent. All the tRNA genes could be folded into cloverleaf secondary structures, while the secondary structure of Ser $^{(AGY)}$ lacked a D arm (DHU stem). By comparing *Rasbora* genome sequence with the recognition sites in teleost species, five conserved sequence blocks were identified in the control region. The common conserved motif sequence of origin of L-strand replication were also identified. The phylogenetic analyses were based on the concatenated of nucleotide sequence and amino acid sequence from 13 PCGs. The phylogenetic position of *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp. were evaluated from the Maximum likelihood trees. Overall, this study enriches the knowledge of *Rasbora* mt-genomes while partially resolving the ambiguity of the phylogenetic relationships of *Rasbora* taxonomy.

Keywords: *Rasbora argyrotaenia*, *Rasbora myersi*, *Rasbora sarawakensis*, mitochondrial DNA, phylogenetic analysis

DNA Mitokondria Rasbora Lengkap: Pencirian Mitogenom dan Implikasi Filogenetik

ABSTRAK

Rasbora adalah ikan air tawar yang termasuk dalam keluarga Cyprinidae di mana kebanyakan ahlinya berkongsi morfologi yang sama yang membuat mereka sukar dibezakan antara spesies. Dalam kajian ini, kedua-dua analisis morfometri dan pendekatan mitokondria DNA digunakan untuk menyelesaikan hubungan filogenetik antara spesies ini. Urutan lengkap mt-genom untuk R. argyrotaenia, R. myersi, R. sarawakensis *dan* Rasbora *sp. adalah 16,574 bp, 16,581 bp, 16,709 bp dan 16,518 bp masing-masing. Struktur mtDNA untuk semua spesies di bawah kajian disenaraikan dengan analisa bioinformatik. Mt-genome* Rasbora *terdiri daripada set standard 22 tRNAs, dua rRNAs, 13 PCGs dan dua kawasan tidak berkod khas. Nilai* AT-skew *dan* GC-skew *dalam seluruh mt-genom* Rasbora *adalah 0.141 dan -0.268 masing-masing, mendedahkan bahawa H-strand mempunyai jumlah C dan A secara lebih lazim. Semua gen tRNA boleh dilipat ke dalam struktur sekunder cloverleaf, sementara struktur sekunder Ser (AGY) tidak mempunyai lengan D (batang DHU). Dengan membandingkan urutan mt-genom* Rasbora *dengan tapak pengiktirafan dalam spesies teleost, lima blok urutan dipulih telah dikenal pasti di rantau kawalan. Urutan motif umum yang dipelihara dari asal replikasi L-strand juga dikenalpasti. Analisis filogenetik adalah berdasarkan urutan-urutan nukleotida dan urutan asid amino dari 13 PCGs. Kedudukan filogenetik* R. argyrotaenia, R. myersi, R. sarawakensis *dan* Rasbora *sp. telah dinilai dari pokok* Maximum Likelihood. *Keseluruhannya, kajian ini memperkaya pengetahuan tentang mt-genome* Rasbora *serta menyelesaikan sebahagian kekaburan hubungan filogenetik taxonomi* Rasbora.

Kata kunci: Rasbora argyrotaenia, Rasbora myersi, Rasbora sarawakensis, *mitokondria DNA, analisis filogenetik*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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

CHAPTER 1

INTRODUCTION

1.1 Background research

In recent years, increased evidence suggests the adaptation of next generation sequencing techniques to investigate functional genomics of the mitochondrial genome. Mitochondria are the most important organelle in the cell as it supplies energy to carry out many cellular activities. This membrane-bound organelle possessing genetic material that are evolutionarily separated from the nuclear genome. Mitochondrial DNA (mtDNA) in most vertebrate is circular and double-stranded. Mitochondrial genome (mt-genome) content less than 40 genes and 2 common non-coding regions. The mtDNA has been widely used in molecular phylogeny because mtDNA has unique properties such as unique maternal inheritance with high evolutionary rates, as well as contains conserve genes that lack genetic recombinantion.

Interestingly, mtDNA has been used for several decades as molecular marker. The mtDNA based approach provide powerful information to determine accurate identity of the organism and even to reveal the relationship among related species. Additionally, molecular approach is more precise, practical and provides in-depth understanding to the variations and similarities of organisms, and delivers better explanations especially for very close related species.

Traditionally, an organism is identified via morphological features and classify based on their similarity (Lumbantobing, 2014). Each species having morphological differences which is used to recognize what species they are and also in what group they are. It is important to know the morphology characters of certain species, so that there are no

misidentification occurs. Therefore, in this study, both morphological characteristics and molecular methods were devised to differentiate among *Rasbora* species.

It is no doubt that morphological method is effectively used to identify and classified each organism into certain groups. However, this method has limitations when it comes to closely related species. Usually, closely related species have high morphological similarities up to the point that is impossible to tell their differences which poses problems in determining their true identity. Previously, many researchers have stated Rabora as a catch-all group (Kottelat and Vidthayanon, 1993; Liao et al., 2010; Lumbantobing, 2014). *Rasbora* accommodates small *cyprinid* fish which were unable to be assigned into a particular group due to the lack of unique morphological characteristics. Thus far, *Rasbora* constitute the most species-rich genus in the *Cyprinidae* family with 147 nominals but only 87 recorded as a valid species (Lumbantobing, 2014).

Furthermore, the phylogenetic relationship of *Rasbora* is uncertain and poorly understood, Howes (1991) stated that *Rasbora* group form non-monophyletic relationship. In 1990's, Kottelat and Vidthayanon (1993) and Kottelat and Witte (1999) suggested the splitting of *Rasbora* splitting into several groups based on morphological characteristics without taking into consideration molecular phylogeny. This modification was then revised by Liao et al. (2010) and upon revision, some species were removed based on such phylogenetic analysis. Difficulties to recognize organisms based on physical appearance have led to the use of the molecular datasets. In this advance molecular era, the full set of mtDNA of any living things are able to be extracted, copied and sequenced. The mt-genome contain much data which is able to reveal the true identity of an organism.

Therefore, in this study, complete mtDNA sequences from four species of *Rasbora* has been studied and presented. All four species of *Rasbora* originated from Sarawak and

there is no data of complete mtDNA sequence existed before. 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. It is hypothesized that using both the former and latter approach, the phylogenetic relationship of the numerous *Rasbora* species found in Sarawak can be eventually resolved. This work contributes towards the genetic resource enrichment for peat swamp conservation and comprehensive in-depth comparisons across other phylogenetic researches done on the *Rasbora*-related genus especially in the state of Sarawak.

1.2 Objectives 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 *R. argyrotaenia, R. myersi, R. sarawakensis* and *Rasbora* sp. 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 Background of *Rasbora*

Scientific classification**:** *Rasbora* belongs to the kingdom Animalia and grouped in the phylum Chordata, which designates them as being vertebrates. The class Actinopterygii identifies them as being bony-fish with rays-fins. From there on, *Rasbora* is classified according to various characteristics. *Rasbora* is in the order called Cypriniformes where it shares this classification with similar species, like the carp, minnow and loaches.

Rasbora is a genus belonging to the most diverse fish family, the Cyprinidae family. According to Nelson et al. (2016) this family have about 367 genera and 3,006 species. Members of this family are widely distributed across Africa, Eurasia and North America but however, absent in Australia and South America. Various species of this family are very important sources of food (e.g. *Cyprinus carpio*, *Hypophthalmichthys nobilis*, *Cirrhinus molitorella* and etc.), ornamental fish (e.g. *Cyprinus robrofuscus* and *Carassius auratus*) and also molecular biology studies (e.g. *Danio rerio*).

Rasbora is a small freshwater fish which can be found throughout Indian Subcontinent, Southern China and abundantly in Southeast Asia. Also, this genus lives in various habitat from rivers, forest streams, lakes, ponds, swamps and even roadside ditches. Actually, this genus well-known as rich-species among Cyprinid fish. This genus has grown to consist of 147 nominal species, of which 87 are regarded as valid (Lumbantobing, 2014). Many species of *Rasbora* were discovered and still growing, perhaps out there several species has not yet described.

Genus *Rasbora* in the same subfamily of our important model organism zebrafish, which is Danioninae subfamily. The relationship of this subfamily has been unexplored until 1990s. According to Howes (1991), Danioninae formed uncertain relationship. The relationships of *Rasbora* have yet to be thoroughly analysed, and several researchers (Howes, 1991; Cavender and Coburn, 1992; Nelson, 2006) have suggested the Rasborinae may represent a non-monophyletic grouping.

The relationship of this group especially *Rasbora* lineage is poorly understood because of their lack of unique morphological characteristics. *Rasbora* is commonly understood as possessing polyphyletic lineage (Kottelat, 1999). The *Rasbora* identified according to the concept of Brittan (1954) normally referred as *Rasbora (sensu lato)*. The intra-relationship of *Rasbora (sensu lato)* is unknown and unexplored (Kottelat & Witte, 1999). Also, the non-monophyly of the genus *Rasbora* was supported by Mayden et al. (2007) and Fang et al. (2009).

Rasbora was first described by Bleeker (1860) to accommodate small cyprinid fishes. Table 2.1 show the morphological descriptions of *Rasbora* according to Bleeker (1860). Back then, only eleven species of *Rasbora* reported by Bleeker (1860). The eleven *Rasbora* species were namely, *R. cephalotaenia*, *R. einthovenii*, *R. lateristriata*, *R. kalochroma*, *R. dusonensis*, *R. leptosome*, *R. argyrotaenia*, *R. borneensis*, *R. buchanani*, *R. sumatrana* and *R. bankenensis*.

Although Bleeker (1860) was the first to describe *Rasbora*, the most recognized and well-studied research on *Rasbora* has to be Brittan (1954). Before Brittan (1954), several genera claimed having relationship with *Rasbora* lineage, namely, *Megarasbora* (Gunther, 1868), *Brevibora* (Dunker, 1904), *Trigonopoma* (Weber & Beaufort, 1915) and

Microrasbora (Annandale, 1918). Nevertheless, Brittan (1954) only recognized three subgenera, namely, *Rasbora*, *Megarasbora* and *Rasboroides*.

Then, Brittan (1972) further divided subgenera *Rasbora* into eight species group complexes. The eight species group complexes are, *Rasbora argyrotaenia*, *Rasbora caudimaculata*, *Rasbora daniconius*, *Rasbora einthovenii*, *Rasbora lateristriata*, *Rasbora pauciperforata*, *Rasbora trifasciata* and *Rasbora sumatrana-elegans*. Kottelat and Vidthayanon (1993) modified *Rasbora (sensu lato)* by reassignments of some species, but retained the 8 species group complexes. Then, *Rasbora (sensu lato)* were split into six genera, namely, *Boraras*, *Megarasbora*, *Rasbora (sensu stricto)*, *Rasboroides*, *Sundadanio* and *Trigonostigma* (Kottelat and Vidthayanon, 1993; Kottelat and Witte, 1999). The splitting of *Rasbora (sensu lato)* was done without conducting phylogenetic analysis.

Again, the splitting of *Rasbora (sensu lato)* was and revised by Liao et al. (2010) by conducting phylogenetic analysis based on morphological data. This time, the rearrangement was done based on the phylogenetic analysis result. Formerly, *Rasbora (sensu lato)* have eight species group complexes. But after the modification only six species group complexes were retained and some of the species were reassigned to different species groups (Liao et al., 2010).

Table 2.1: *Rasbora* morphological description according to Bleeker (1860).

Morphological description of *Rasbora*
Body slightly elongated or elongated, compressed, cover with large scales, back low;
Jaw enclosed in terete, simple lips upper lip slightly protrusible;
No barbels;

Table 2.1 continued

Snout acute, slightly depressed, not protruding anterior to the mouth;
Anterior suborbital bone pentagonal, tip acute, pointing upward;
Mouth anterior, gape strongly oblique, ending anterior to the eye or below the anterior part of the eye;
Upper jaw with a short process towards the symphysis, symphysis emarginate, taking in the hook of the lower jaw;
Lower jaw not shorter than upper jaw, emarginate towards the symphysis, symphysis hooked with a bony tubercle;
Postorbital groove on both sides parallel to the free margin of the jaw, not united with the groove on the opposite side;
Eye slightly superior, not covered by palpebral membrane;
Belly not kneeled;
Dorsal fin with few rays, starting behind ventral fin and ending anterior to anal fin;
Anal fin with few rays (3/5 or 3/6);
Lateral line strongly curved, much closer to ventral line than dorsal line;
Gill opening ending below the pre-operculum;
Pharyngeal teeth slightly spoon-shaped to hooked.

2.2 Characteristics of *Rasbora*

The common name of *Rasbora* is ‘ikan seluang’ or ‘ikan lansi’ or ‘ikan bada’ in Sarawak. *Rasbora* fish has slim and small body which make them fast swimmer. They are considered mid-water dweller and live in moderately flow or stagnant water. Besides, the appropriate conditions for *Rasbora* to lives are temperature between 18-26 °C and water pH range 5-7.

To differentiate male and female, the male has brighter color and slimmer bodies, while the female *Rasbora* has a rounder lower profile and at breeding time become quite plump. There is not much information on how *Rasbora* breed. Basically, the male fish will continuously push the female to the mating place. The mating places for *Rasbora* are usually crowded with dead leaves, small branches or stone bed, where there is safe place to hide their egg. The mating begins when the female turn upside down and rubbing her belly with her partner.

Basically, the support structure of *Rasbora* body is made of bone and their fins composed of soft spines called rays. Their small body covered with cycloid scales. Several species of *Rasbora* having horizontal stripe starting either from the tip of the snout, operculum or at the middle of the body to the caudal base, sometimes the stripe may be edge with gold, silver or red, and black spot either at the base of caudal or at the middle of the body or anal fin base streak.

Comprehensive study on *Rasbora* morphology was done by Brittan (1954). The descriptions of *Rasbora* morphology according to Brittan (1972) are shows in Table 2.2 where it has led to the concept of the genus *Rasbora (sensu lato)*. Although, there are much information on morphological characteristics for identification of each *Rasbora* species, yet the identity of many species from this genus remain ambiguous. Many researchers

considered *Rasbora (sensu lato)* as catch-all group (Kottelat and Vidthayanon, 1993; Liao et al., 2010), where it is relatively difficult to assigned them into certain species group under certain definite morphological similarity, so, they are placed on a single species group.

The identification of certain species of *Rasbora* are confusing, since most of them having high similarity on morphology characteristics. For example, the *R. dusonensis (sensu lato)* is actually *R. tornieri* and *R. dusonensis* (Kottelat, 1991) is *R. myersi* with the sole differences between them being the color pattern on caudal fin and caudal peduncle.

Table 2.2: Morphological identification character of *Rasbora* (Brittan, 1972).

Morphological description of genus *Rasbora*
Laterally compressed, elongated body;
Presence of a symphyseal knob on the lower jaw;
3 pharyngeal tooth rows;
1 or 2 scales between lateral line scale row and pelvic insertion;
Dorsal fin insertion opposite to pelvic fin insertion or slightly behind it;
2 simple and 7 branched dorsal fin rays
3 simple and 5 branched anal fin rays;
And absence of barbels.

Brittan (1984) stated a single species of *Rasbora* can vary considerably in size, shape and in particular color due to the ecology and the food available, as well as temperature and fluctuations in pH of water, all have their effect. For example, the *R. harlequins* from

Singapore and Sumatra having different color pattern and shape. Besides, Brittan (1984) also mentioned identification of *Rasbora* is not easy especially when evaluating extra-small specimens.

2.2.1 *Rasbora argyrotaenia*

R. argyrotaenia was found by Dr Pieter Bleeker in 1859. This fish is widely distributed across Asia across Thailand, Malaysia and Indonesia. Also, this species can live in large murky river and also in small clear forest stream.

According to Brittan (1954) *R. argyrotaenia* has brilliant silver and beautifully slim elongated body. This species can be considered as large-sized among *Rasbora* species which can reach up to 110 mm in standard length. Figure 2.1 show *R. argyrotaenia*, this species having dark lateral stripe running from the opercle to the caudal base and together with the gold or silver line along the lateral stripe.

The colour of the caudal fin is bright yellow with blackish edge and hyaline or clear pectorals, pelvic and anal fin. The caudal fin deeply forked with pointed lobes; the dorsal fin broadly pointed; and pointed pectorals, pelvic and anal fin. All those colour and fins features based on characteristics stated by Brittan (1954).

The morphological characteristics of *R. argyrotaenia* according to Kottelat et al. (1993) and Inger and Chin (1962) are the dorsal 7; pectoral 13-14; ventral 7-8; anal 5; lateral line scales 27-29; pre-dorsal scales 11; circumpeduncular scales 14. Standard length 52.4-85.7 mm; total length 71.4-115.0 mm. As the name indicates, this species is placed under their own group the Argyrotaenia.

Figure 2.1: *Rasbora argyrotaenia*.

2.2.2 *Rasbora myersi*

R. myersi was initially found by Brittan (1954). This species is abundantly distributed in Thailand and Borneo including Peninsular Malaysia and Sumatra Indonesia. This fish can live in clear forest streams and even in large river with murky water. *R. myersi* has stout elongated body. Figure 2.2 show the picture of *R. myersi*, the colour of the body is silvery or grey and some has a bit yellowish stripe around the body contour. Upon preservation in absolute ethanol, a fine and faint stripe appear along the lateral line, while alive, the stripe remains largely invisible. The caudal fin is deeply forked, greyish with blackish edge. The other fins are clear or greyish with pointed lobes.

As according to Kottelat (2013), this large-sized *Rasbora* is placed under the Argyrotaenia group as well because of their dorsohypural distance equal to or slightly less than the predorsal length, fewer scale rows between the lateral line and mid-ventral row in front of the pelvic fins. Usually, the Argyrotaenia group members present a dark mid-lateral stripe running from the opercle to the caudal base, but, *R. myersi* has only dusky like stripe along the mid-lateral.

The validity status of *R. myersi* has been doubted previously, some claimed this species as *R. dusonensis*. The confusion occurs because of the morphological characteristics of this species is the same as *R. dusonensis* as stated by Kottelat (1991) where the main difference between them is their body color. Then, Ng and Kottelat (2013) revised these 3

species *R. dusonensis (sensu lato)*, *R. dusonensis (sensu kottelat)* and *R. myersi (sensu lato)*, where they found out *R. dusonensis (sensu lato)* is *R. tornieri*, *R. myersi (sensu lato)* is *R. dusonensis* while *R. dusonensis (sensu kottelat)* is *R. myersi*.

Figure 2.2: *Rasbora myersi*.

2.2.3 *Rasbora sarawakensis*

R. sarawakensis was originally recognized by Brittan in 1951. Although this small sized fish is named after Sarawak, this does not mean that it can only be found in Sarawak. This species is endemic to Borneo where it has been sighted in west Kalimantan, Indonesia. Mostly lives in clear or weakly coloured brown slow-moving stream with shaded dense forest canopy above.

According to Siebert and Guiry (1996), *R. sarawakensis* belong to the Trifasciata group. The Trifasciata group are said to show a dark lateral stripe that is markedly less intense on the anterior half of the body. The *R. sarawakensis* being distinguishable by its distinctive longitudinal stripe with very intense dark blue running from the operculum toward the caudal base as show at Figure 2.3.

As for the characteristics of *R. sarawakensis*, the fins are clear to bright orange-yellow; the caudal fin forked; the dorsal, pectorals, pelvic and anal fin broadly pointed. And it is relatively stout-bodied with large and pointed head; complete lateral line with 24-26 scales; 11 predorsal scales; 12 circumpenduncular scales; $7\frac{1}{2}$ branches dorsal rays; $5\frac{1}{2}$

branches anal rays; distinct supra-anal streak; present of dark pigment at dorsal and anal fins (Brittan, 1972; Kottelat et al., 1993; Robert, 1989).

Figure 2.3: *Rasbora sarawakensis*.

2.2.4 *Rasbora sumatrana*

R. sumatrana found by Dr Pieter Bleeker in 1859. This species can be found across Southeast Asia from Thailand, Malaysia and Indonesia. This species inhabits various habitat such as clear or muddy rivers, forest streams, black water swamp and lakes.

R. sumatrana has slender, elongated body and laterally compressed. Greatest body depth located between verticals through pelvic-fin insertion and dorsal-fin origin. A narrow dark stripe runs from the operculum to the caudal peduncle where it stops as a dark spot. Reticulate pattern of the scales is distinct. There, black streak at the supra-anal and dark line at the base of caudal peduncle. The fins with tinted yellow, caudal lobes back margins and tips are black.

Color in life is silvery, darker above; a black mid-lateral line extending from behind operculum to base of caudal; a black mid-dorsal line, most pronounced behind dorsal; a wide black streak at base of anal, drawn out posteriorly and running in mid-ventral behind anal; distal third or more of caudal lobes black; no black on other fins; caudal orange; dorsal yellow; ventral yellow or clear; pectoral without pigment (Inger and Chin, 1962).

As the name is *R. sumatrana*, this species belongs to the Sumatrana-elegans group. The Sumatrana-elegans group are said to have a unique black pigment pattern consisting of a reduced or modified mid-lateral stripe alongside other such as the supra-anal pigmentation and the blotch on the caudal peduncle (Brittan, 1954; Kottelat and Vidthayanon, 1993; Kottelat and Tan, 2012).

2.3 Structure of Mitochondrial DNA

Mitochondria are the powerhouse of a cell that plays a crucial role in many cellular activities. This organelle possesses their own genetic material and relatively autonomous where they can manufacture their own proteins, enzymes and RNAs. Figure 2.4 show the illustration of the structure of the human mt-genome.

Animal mtDNA is extremely compact, encode only the important components of polypeptides synthesis. Animal mt-genome are circular and double-stranded DNA, and encoding 37 genes comprises of 13 protein coding genes (PCGs), 22 tRNAs and 2 rRNAs. The 13 PCGs are main components of oxidative phosphorylation process, while 2 subunits of rRNAs (small-12S and large-16S) and 22 tRNAs are essential in translation process (Boore, 1999).

The mtDNA is divided into 2 strands, heavy (H) and light (L) strand. On the H-strand consist of 2 rRNAs (12S and 16S), 12 PCGs (ND1, ND2, ND3, ND4, ND4L, ND5, COI, COII, COIII, ATP8, ATP6 and Cytb) and 14 tRNAs (Phe, Val, LeuUAA, Ile, Met, Trp, Asp, Lys, Gly, Arg, His, SerGCU, LeuUAG, Thr) and the remaining genes were encoded on the L-strand (1 PCG**:** ND6, and 8 tRNAs**:** Gln, Ala, Asn, Cys, Tyr, SerUCN, Glu and Pro).

Besides, the mtDNA are divided into 2 main regions**:** coding and non-coding region. The coding region consist of tRNA, rRNA and protein genes. And the non-coding region consist of several regulatory elements for DNA regulation. Mitochondrial genome having 2

non-coding regions, the control region (CR) and origin of L-strand replication (OL). The CR contain several regulatory elements that is responsible in initiating and terminating the replication and transcription process. The CR is located between 2 tRNAs (Pro and Phe) and also act as promoter for both H- and L-strand. All the regulatory elements sequences seem highly conserved.

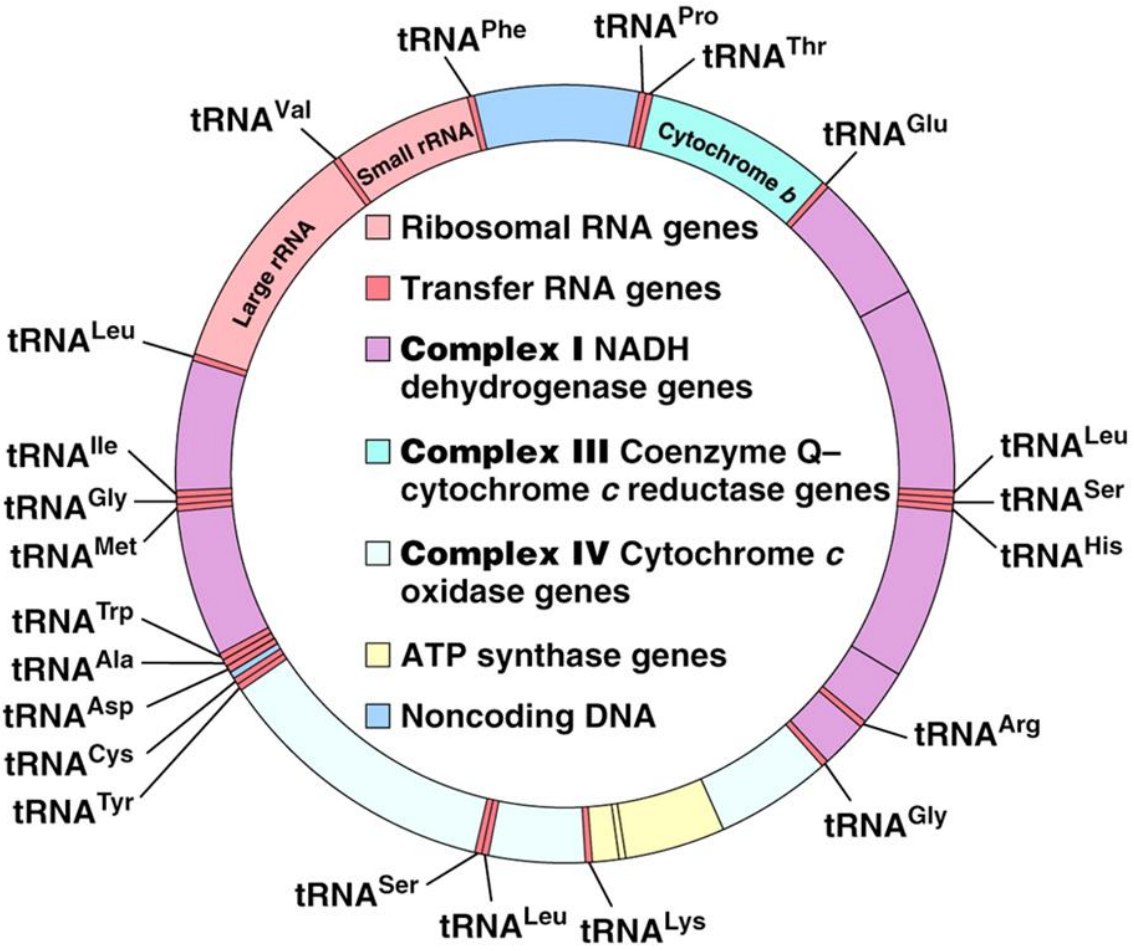


Figure 2.4: The structure of human mitochondrial genome (Becker et al., 2009).

2.4 Mitochondrial DNA as Molecular Marker

MtDNA have been used in molecular marker for a number of studies including species identification by DNA barcode and taxonomic studies. Various mtDNA markers have been used for phylogenetic study to a great extent. The common mtDNA marker used

for molecular analysis are PCGs, rRNA and CR. Each mtDNA marker has different characteristics and different group of researcher tends to champion their preferable group (Table 2.3).

MtDNA possess several favourable characteristics for phylogenetic study. The size of mt-genome is small and the genes content are similar across animal phyla. Also, the arrangement of major genes in mt-genome is highly conserved (Mandal, et al., 2014). Additionally, easy to recover from degraded DNA due to high copy number (Brown, 2008) in most cells. These characteristics give high possibility to amplify target genes and to construct the complete genome sequence.

Besides, mtDNA is maternally inherited (Sato & Sato, 2012). In other word, each generation share the same mtDNA from the maternal family line. Mitochondria of the sperm primarily designed for locomotion and do not involved in embryo development and hence not inherited in the subsequent generation. Furthermore, mtDNA is haploid whereby lack of DNA recombinantion or no genetic exchange between parent is involved. These characteristics enable it to be used to trace respective species ancestry and relatives.

Also, the evolution rate of mtDNA is much faster than nuclear DNA. According to Brown et al. (1979), mtDNA can evolve about 5-10 times faster than in a single-copy of nuclear DNA. Moreover, the changes of mtDNA is mainly because of mutation and inefficient of DNA repair system present in the small organelle. Modification of mtDNA pass down to the next generation may affects how it looks, how it behaves and its physiology. The variation in DNA might be used to define a particular species of an organism while the similarity in DNA may explained each species close relatedness.

The rate of evolution among the different mitochondrial regions is different because CR is a non-coding DNA, while rRNA and PCGs are coding DNA and functional genes.

The changes of DNA sequence of a functional gene may affect its physiology, and has lethal effect to the organism. Thus far, complete mitochondrial genome of several *Rasbora* species have been published which includes *R. steineri*, *R. aprotaenia*, *R. lateristriata*, *R. argyrotaenia*, *R. trilineata*, *R. myersi* (Chang et al., 2013; Kusuma & Kumazawa, 2016; Kusuma et al., 2017; Ho et al., 2016; Lim et al., 2019).

Table 2.3: Characteristics of mt-DNA markers (Arif and Khan, 2009).

Characteristics of mitochondrial DNA
Control region**:**
Hypervariable sequences (used for identification of species and sub-species)
Protein coding genes**:**
Used in low categories such as families, genera and species
According to Zardoya and Meyer (1996) mitochondrial PCGs classified into 3 group**:**
Good**:** ND4, ND5, ND2, cyt b and COI
Medium**:** COB, COIII, ND1 and ND6
Poor**:** ATPase 6, ND3, ATPase 8 and ND4L
Ribosomal RNA**:**
12S rDNA (highly conserved and used in high category level such as phyla and subphyla)
16S rDNA (usually used in mid category differentiation such as families)

CHAPTER 3

METHODOLOGY

3.1 Sampling and Sample Maintenance

Four species of *Rasbora* were characterized in this study. These species originated from Sarawak and were identified according to attributes listed in Table 2.2. A total of 61 samples of *Rasbora* spp. were collected from July – September 2017 (Table 3.1). The samples were caught by several fishing equipment such as gill net, casting net, seining and scoop net depending on the sampling location and situation (Chang et al., 2013). The samples were fixed in 95% ethanol at the sampling sites and transported to laboratory. They were then allowed to be fixed in ethanol for 3-5 days to remove the water from the cell body. Upon fixation, the sample were ready for morphometry and molecular study (Lumbantobing, 2014).

Table 3.1: Sampling location according to its coordinate and the number of samples collected.

Sample type	Location	Coordinate	No. of sample
Rasbora argyrotaenia	Sungai Latong, Machan, Sibu	1°59'55.0" N, 112°00'55.3" E	7
Rasbora myersi	Sungai Latong, Machan, Sibu	1°59'55.0" N, 112°00'55.3" E	15
Rasbora sarawakensis	Gunung Gading downstream, Lundu	1°41'27.4" N, 109°50'43.5" E	11
Rasbora sarawakensis	Ranchan downstream, Serian	1°08'31.8" N, 110°34'58.5" E	4
Rasbora sarawakensis	Sungai Batu, Beratok, Siburan	1°19'09.8" N, 110°24'27.7" E	4
Rasbora sp.	Sungai Latong, Machan, Sibu	1°59'55.0" N, 112°00'55.3" E	20

3.2 Morphometry Methods

3.2.1 Morphometric Measurements

In this study, the morphometric characterization was followed according to Lumbantobing (2010). All the 24 characters are listed on Table 3.2 and the schematic measurements was followed (Figure 3.1). All the measurements were taken from the specimen facing to left side to avoid introducing technical variation. Twenty-four characters were measured to the nearest 0.005 mm using the Vernier calipers. The measurement was reported as a range of percentage of standard length and mean ± standard deviation.

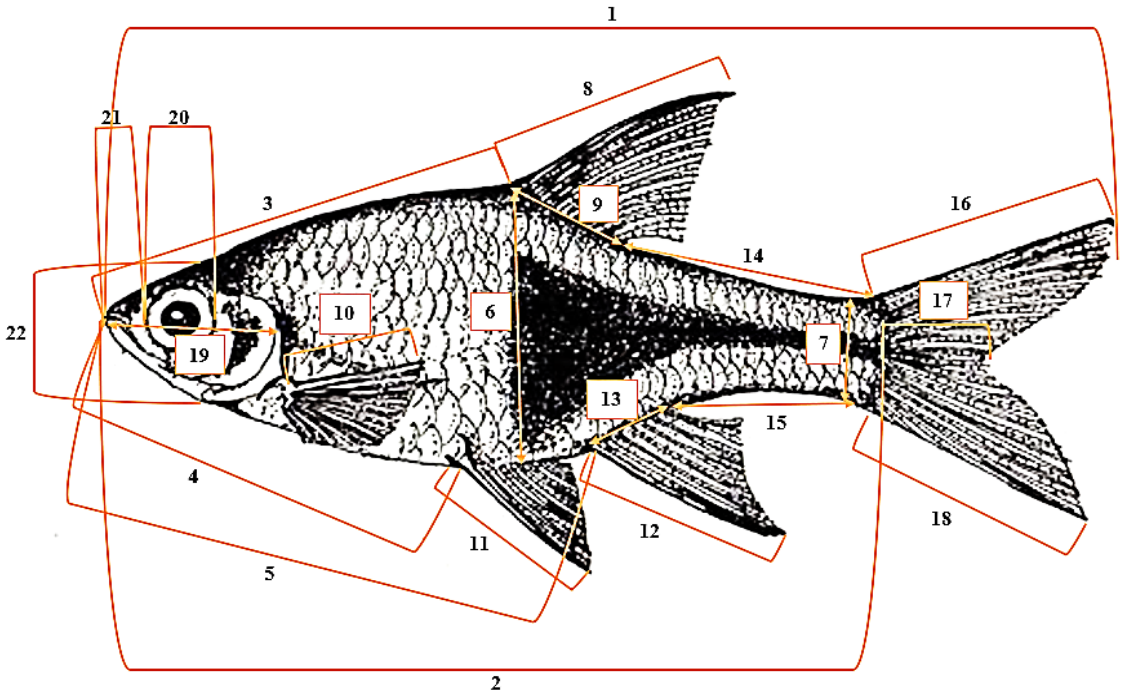


Figure 3.1: Lateral view of schematic body for morphometric measurements (modified Mohsin and Azmi, 1983). Guideline measurement for morphometric characters, 1**:** Total length, 2**:** Standard length, 3**:** Pre-dorsal length, 4**:**Pre-pelvic length, 5**:** Pre-anal length, 6**:** Body depth, 7**:** Caudal peduncle depth, 8**:** Dorsal fin depth, 9**:** Dorsal fin base length, 10**:** Pectoral fin length, 11**:** Pelvic fin length, 12**:** Anal fin depth, 13**:** Anal fin base length, 14**:** Dorsohypural distance, 15**:** Caudal peduncle length, 16**:** Upper caudal fin lobe length, 17**:** Middle caudal fin lobe length, 18**:** Lower caudal fin lobe length, 19**:** Head length, 20**:** Eye diameter, 21**:** Snout length, 22**:** Head depth.

Table 3.2: The list of 24 morphometric characters defined by Lumbantobing (2010) and adapted in this study.

#	Characters		Descriptions
1	TL	Total Length	Distance from the tip of snout to the tips of caudal fin
2	SL	Standard Length	Distance from the snout tip to the end of vertebral column
3	PrDL	Predorsal Length	Distance from the snout tip to the base of first dorsal fin
4	PrPvL	Prepelvic Length	Distance from the snout tip to the base of pelvic fin
5	PrAL	Preanal Length	Distance from the tip of snout to the base of first anal fin
6	BD	Body Depth	Maximal depth measured vertically from the base of dorsal fin to the base of pelvic fin
7	CpD	Caudal peduncle Depth	Depth measured at the caudal peduncle at the end of vertebral column
8	DD	Dorsal Depth	Maximal depth measured vertically from first dorsal fin base to the tip of first dorsal fin
9	DfbL	Dorsal fin base Length	Distance of the dorsal fin base
10	PcL	Pectoral Length	Maximal length of pectoral fin
11	PvL	Pelvic Length	Maximal length of pelvic fin
12	AD	Anal Depth	Maximal distance measured from the first anal fin base to the tip of anal fin
13	AfbL	Anal fin base Length	Distance of the anal fin base
14	DHD	Dorso-Hypural Distance	Length measured from the last of dorsal fin to the end of vertebral column
15	CpL	Caudal peduncle Length	Distance from the last anal fin to the end of vertebral column
16	UcLL	Upper caudal Lobe Length	Maximal length from base of caudal fin to the end of upper caudal fin
17	McLL	Median caudal Lobe Length	Maximal length from base of caudal fin to the end of median caudal fin
18	LcLL	Lower caudal Lobe length	Maximal length from base of caudal fin to the end of lower caudal fin
19	HL	Head Length	Distance from the anterior point of snout to the posterior point of operculum
20	ED	Eye Diameter	The diameter measured from the outer bony cavity
21	SnL	Snout Length	Distance measured from the anterior most point of snout tip to the posterior most of the snout
22	HD	Head Depth	Length measured from the head top, perpendicular to the middle eye, to the lower jaw base
23	HW	Head Width	Distance measured vertically between eye margin from dorsal view
24	IW	Inter-orbital Width	Distance measured vertically between snout margin from the dorsal view

3.2.2 Meristic Counts

There were 11 characters and were all listed in Table 3.4. In fin-ray counts, the simple and branched ray are counted together; the last 2 branches ray counted as 2; and all these fin-ray were counted under stereo microscope. While for the scale's counts, the small scale counted as one and not $\frac{1}{2}$. The counts were reported as range and mean ± SD (standard deviation).

Table 3.3: The list of characters for meristic counts.

#	Characters		Descriptions
1	DF	Dorsal fin	No. of rays counted at dorsal fin
2	PF	Pectoral fin	No. of rays counted at pectoral fin
3	PvF	Pelvic fin	No. of rays counted at pelvic fin
4	AF	Anal fin	No. of rays counted at anal fin
5	CF	Caudal fin	No. of rays counted at caudal fin
6	DpcF	Dorsal procurrent	No. of rays counted at dorsal part of caudal fin
7	VpcF	Ventral procurrent	No. of rays counted at ventral part of caudal fin
8	SLL	Scales of Lateral Line	No. of scales counted along the lateral line
9	PrDS	Predorsal Scale	No. of scales counted at the predorsal length
10	TrDS	Transverse Scale	No. of scales counted from the front of dorsal fin to the front of anal fin
11	Ccps	Circumpeduncular Scale	No. of scales counted around the caudal peduncle

3.2.3 Statistical Analyses

All continuous and discrete data in this study were tabled using Microsoft excel 2016. Then, these data were computed by using the statistical software, Statistical Package for Social Science (SPSS) version 22. The statistical analysis was carried out separately for morphometric and meristic characters.

Prior to the analysis, it was necessary to eliminate any size effect in the data set. Variation should be attributable to body shape differences, and not related to the relative size of the fish. In this study, there are significant linear correlations between all measured characters and the standard length of the fish. Therefore, it was necessary to remove size-

dependent variation for the morphometric characters. An allometric formula, equation (1) outlined by Elliott et al. (1995) was used to correct for length effects in samples.

$$M_{adj} = M\,(L_s / L_o)\,b \qquad (1)$$

Where M**:** original measurement, M_{adj}**:** size adjusted measurement, L_o**:** standard length of fish, L_s**:** overall mean of standard length for all fish from all samples in each analysis. Parameter b was estimated for each character from the observed data as the slope of the regression of log M on Log L_o, using all fish in all groups. The efficiency of size adjustment transformation was assessed by testing the significance of the correlation between transformed variables and standard length.

Standard length was excluded from the final analyses. Univariate (ANOVA) was carried out to test the significance of morphological differences. In addition, size adjusted data were standardized and summitted to discriminant function analysis (DFA) using SPSS. Population centroids with 95% confidence ellipses derived from the DFA were used to visualize the relationships among the individuals of groups. Individuals were assigned to the samples using the discriminant functions, and the percentage of correctly assigned fish was an additional measure of differentiation among samples.

To complement the multivariate analysis, the significant morphometric characters from one-way ANOVA test were used for principal component analysis (PCA). In order to determine which morphometric measurement most effectively differentiates species the contributions of variables to principal components (PC) were examined. To examine the suitability of the data for PCA, the Bartlett's Test of sphericity and Kaiser-Meyer-Olkin (KMO) measure was estimated. The KMO is a measure of sampling adequacy. The data is

acceptable for PCA when the value is greater than 0.6. The Bartlett's test of sphericity is the test for null hypothesis that the correlation matrix has an identity matrix.

3.3 Mitochondrial DNA

3.3.1 Total DNA Extraction

The total DNA of a specimen was extracted via conventional organic extraction method according to Wasko et al. (2013) with slightly modifications. Approximately 10 mm of muscle tissue of sample was taken and placed in a 2 ml tube that contained 910 µl of lysis buffer, 30 µl of 10 mg/ml of Proteinase K and 60 µl of 10% SDS. The tube was then incubated at 48°C for 45-50 mins in water bath. During incubation period, the tube was gently inverted several times for every 8-10 mins to homogenize the muscle cells.

Then, an equal volume of phenol**:** chloroform**:** isoamyl alcohol (25**:**24**:**1) was added to the tube that contained the lysed cells. The mixture was gently inverted and then left for 10 mins to settle down all the organic components. After that the tube was centrifuged for 10 mins at 13,500 x g. Subsequently, the top aqueous layer was pipetted and transferred into a new clean 1.5 ml tube.

Afterward, an equal volume of isopropanol was added into the aqueous solution and gently mixed by inverting the tube few times to recover the DNA. Then, the tube was again centrifuged at 13,000 x g for 10 mins. The DNA is precipitated at the bottom of the tube. Then, the supernatant is removed by gently pouring out.

Subsequently, the DNA pellet was washed using 1 ml of chilled 70% ethanol. Then, the tube was centrifuged at 12,500 x g for 5 mins. Again, the supernatant was removed by slowly pouring it out. Next, the tube was left for about 30 mins with open cap to allow the DNA to dry. Lastly, the DNA pellet was resuspended by adding 80 µl of sterile distilled water. The DNA was stored at -20°C for further use.

3.3.2 Primer Design

Several complete mtDNA of *Rasbora* species is available in the GenBank. Hence, the primers used for polymerase chain reaction were designed based on the consensus sequences of alignment of three mtDNA sequences of *Rasbora* in particular; the *R. steineri* (Ref. seq**:** NC_020005.1), *R. sumatrana* (Ref. seq**:** NC_031542.1) and *R. trilineata* (Ref. seq**:** NC_025336.1).

All 3 mtDNA were retrieved in FASTA format were aligned using the multiple sequence alignment tool, the Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) to determine the consensus sequences. From the identified consensus sequence, after that the forward and reverse primer were designed by using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Then, the designed primer is checked using Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html) to verify their features.

The designed primers were then synthesized by First Base Laboratories Sdn. Bhd. The properties of proper designed primer are; size of 17 – 25 bases in length; the GC content of the primer about 40 – 60%; and preferable melting temperature, T_m is between 40 - 60°C. Also, the differences in melting temperature of forward and reverse primer must not exceed 5°C.

3.3.3 Mitochondrial DNA Amplification

In this study, the mtDNA was amplified using the Takara LA-Taq polymerase with the reaction mixture prepared according to the cool start method. All the reaction reagents and DNA samples were placed on ice prior to adding each reagent following to the order as shown in Table 3.4. Next, all the PCR tubes were set in a Bio-Rad Thermal cycler. The conditions for PCR were set by following the Takara LA-Taq recommendation as shown in

Table 3.5. After the PCR cycle was completed, the tubes were stored at -20°C or directly subjected to agarose gel electrophoresis.

Table 3.4: The PCR reaction mixture.

Components	Final concentration	1× volume (µl)
Sterile distilled water	N/A	13
10× LA PCR Buffer II (Mg^{2+} plus)	1×	2
2.5 mM dNTP Mixture	0.40 mM	3.2
10 µM Forward Primer	0.20 µM	0.4
10 µM Reverse Primer	0.20 µM	0.4
5 U/µl Takara LA Taq	0.05 Units	0.2
DNA template	As required	0.8
Total volume		20

Table 3.5: The PCR cycle conditions.

Steps	Temperature (°C)	Duration		
Initial denaturation	94	1 min		
Denaturation	94	30 secs	}	30 cycles
Annealing	Depend on primers Tm	5 min	}	
Extension	72	(1 min/kb)	}	
Final extension	72	10 min		
Hold	4	∞		

T_m is annealing temperature

3.3.4 PCR Purification

The PCR product was purified using EasyPure® PCR Purification Kit (Transgen, China). Initially, five volumes of binding buffer are added to one volume of PCR product. Then, the sample mixture was mixed via vortexing. All the mixture was transferred into a spin column with a collection tube and incubated for a minute. After that, the sample mixture was centrifuge at 10,000×g for a minute. Subsequently, the filtrate was discarded. Next, 650 µl of washing buffer was added into the column and spun at the same speed for a minute. Then the filtrate was discarded and the empty column was spun to remove any residual washing buffer.

In the final step, the spin column was placed in a clean microcentrifuge tube. About 20 µl of elution buffer was added directly to the center of the column matrix and incubated for a minute. Again, the sample was centrifuge at the same speed for one minute to elute the PCR product (DNA). The purified PCR product was subsequently stored at -20°C and properly sealed with parafilm. Purified amplicons were sequenced pair-endedly with short reads (~200 bp) 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 (Bankevich et al., 2012).

3.4 Mitochondrial DNA Sequence Characterization

Each end of respective fragments was aligned to form a full length of mt-genome sequence. The location of protein-coding genes (PCGs) and ribosomal RNA (rRNA) genes were determined by comparison of their similarity with the published mitogenome of *Rasbora* species. Whereas the transfer RNA (tRNA) genes were identified using the program tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/).

The composition of whole genome, PCGs, rRNAs and tRNAs nucleotide was determined using MEGA X. The strand asymmetry was calculated using the following formulae (Perna and Kocher, 1995):

$$\textbf{AT-skew} = [A - T] / [A + T]$$

$$\textbf{GC-skew} = [G - C] / [G + C]$$

And then the whole DNA sequence was verified using MitoFish: MitoAnnotator (http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html). The circular map mt-genome was drawn using GenomeVx (http://wolfe.ucd.ie/GenomeVx/) (Conant and Wolfe, 2008). The predicted clover-leaf structure of tRNAs were determined using tRNAscan-SE. The

proposed secondary structure of the putative origin of light strand (OL) was analyzed using the Mfold Web Server (http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form).

3.5 Phylogenetic Analysis

To clarify the phylogenetic position of the present *Rasbora*, the complete mitogenome sequence of other *Rasbora* species were incorporated together for phylogenetic analysis. The sequences were aligned using Clustal W and adjustments were made manually. Phylogenetic analyses are based on the concatenated of 13 PCGs sequence. 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 bootsraps, and maximum likelihood method. Maximum likelihood method was used to constructed phylogenetic tree and the substitution model was assessed by MEGA X software. The confidence level at each branch was evaluated by performing bootstrapping with 1000 replicates.

CHAPTER 4

RESULTS

4.1 Morphology of *Rasbora*

The most noticeable colour changes in all specimens upon preservation in ethanol was the eyes. Black eyes turned to fully white colour upon storage for more than 30 minutes. However, some part of the specimen like caudal fin are able to retain their original colour. The specimens of *R. argyrotaenia* was collected from a forest stream at Machan, Sarawak (Figure 4.1). Diagnosis**:** Dorso-hypural distance almost equal with the dorsal fin origin and the posterior orbital margin. Dorsal fin origin starts at 13th predorsal scale. The tip of pectoral fin reaching 9th lateral-line scale, the pelvic fin insertion below 11th lateral-line scale. The tip of pelvic fin not reaching anal opening, and anal fin origin below 19th lateral-line scale. Having 13 circumpenduncular scales. Slender body with dark midlateral stripe start from operculum to the base of caudal peduncle. The dark stripe more intense on central portion toward caudal peduncle and edge with bright silver. Also present of a weak supra-anal pigmentation. Coloration in life**:** Yellowish above, lighter below. Having yellow dorsal and caudal fin. The caudal fin edge with black color, and clear pectoral, pelvic and anal fins.

Figure 4.1: *R. argyrotaenia*. A) Fresh specimen. B) Alcohol preserved specimen. Scale bar: 10 mm.

General appearance of *R. myersi* is shown in Figure 4.2. *R. myersi* was collected from a clear forest stream at Machan, Sarawak. Diagnosis**:** Dorso-hypural distance slightly less with the dorsal fin origin and the posterior orbital margin. Dorsal fin origin starts at 13th predorsal scale. The tip of pectoral fin reaching at 10th lateral-line scale which is very close to the pelvic fin insertion. The pelvic fin insertion located below 11th lateral-line scale. The tip of pelvic fin also almost reaches anal opening, and the anal fin originates below 18th lateral-line scale. Having 13 circumpenduncular scales. Stout body with absent of dark lateral stripe and supra-anal pigmentation, but, present of white like stripe running from central to caudal peduncle base. Also, present of silvery line at the axial body starting behind operculum to caudal peduncle base. Coloration in life**:** yellowish or ground coloration above and lighter below. Having ground coloration at the dorsal and caudal fins. Dark color at the margin of caudal fin and tip dorsal fin. Clear pectoral, pelvic and anal fins.

Figure 4.2: *R. myersi*. Scale bar: 10 mm.

General appearance of *R. sarawakensis* is shown in Figure 4.3. The specimens of *R. sarawakensis* were collected from several locations, Gunung Gading downstream (Lundu), Ranchan downstream (Serian) and Beratok (Siburan). Diagnosis**:** Dorsal fin originates at 10th predorsal scale. The tip of pectoral fin reaching at 9th lateral-line scale which the pectoral fin tip reaches the pelvic fin insertion. The pelvic fin insertion located below 9th lateral-line scale. The tip of pelvic fin reaches anal opening, and the anal fin originates below 15th lateral-line scale. Having 13 circumpenduncular scales. Compressed body with present of dark blue lateral stripe starting from operculum and terminate at the caudal base. Also present of supra-

anal pigmentation. Coloration in life**:** Slightly dusky above, lighter below. Having bright orangey-red caudal and dorsal fins, light orange on pectoral, pelvic and anal fins. On the first two rays of anal and dorsal fins present of dark pigmentation. Coloration in alcohol**:** at the centre of the body, present of the axial streak which is slightly above the dark stripe and diffuse with the stripe toward the caudal base.

Figure 4.3: *R. sarawakensis*. A) Fresh specimen. B) Alcohol preserved specimen. Scale bar: 10 mm.

General appearance specimen of *Rasbora sp.* shown in Figure 4.4. The specimens of *Rasbora sp.* was collected from a clear forest stream at Machan, Sarawak. Diagnosis**:** Dorso-hypural distance slightly less with the dorsal fin origin and the posterior orbital margin. Dorsal fin origin starts at 12th predorsal scale. The tip of pectoral fin reaching at 9th lateral-line scale, pelvic fin insertion below 11th lateral-line scale. The tip of pelvic fin almost reaching anal opening, and anal fin origin below 16th lateral-line scale. Having 13 circumpenduncular scales. Stout body with absent of dark lateral stripe and supra-anal pigmentation, but present of a fine line at the axial running from behind operculum to caudal base. Present of dark red mid-dorsal stripe, extending from nape to dorsal caudal peduncle. Coloration in life**:** Slightly yellowish above and lighter below. Having red caudal and dorsal

fin, both fin margin edge with dark color. The pectoral, pelvic and anal fins color yellow, the tip of pectoral and anal fin color black. Coloration in alcohol**:** The fine dark line running behind operculum to caudal base, term as axial streak.

Figure 4.4: *Rasbora sp.*. A) Fresh specimen. B) Alcohol preserved specimen. Scale bar: 10 mm.

4.2 Morphological Analyses

The original measurement of morphometric characters of studied *Rasbora* were presented as percentage of standard length for body characters and as percentage of head length for head characters. The data were recorded in range and mean ± standard deviation (Table 4.1). Whereas, the meristic characters, the counted data were presented as range and mode as shown in Table 4.2.

Table 4.1: Morphometric data of *Rasbora* present as percentage of SL (standard length) for body characters and as percentage of HL (head length) for head characters. Cell contains**:** range and mean ± standard deviation (mm).

	R. argyrotaenia n = 7	*R. myersi* n = 15	*R. sarawakensis* n = 19	*Rasbora sp.* n = 20
Standard Length (mm)	93.250 - 110.200	73.400 - 95.250	27.100 - 46.150	81.280 - 117.200
	102.330 ± 6.310	84.423 ± 6.464	35.178 ± 4.788	94.983 ± 8.474
Percentage of SL				
Total Length	125.000 - 133.100	126.415 - 136.240	129.298 - 137.475	119.217 - 132.658
	127.182 ± 2.754	131.619 ± 2.401	131.913 ± 1.946	129.130 ± 2.714
Head Length	21.495 - 23.200	22.546 - 25.236	24.768 - 28.576	22.526 - 25.956
	22.173 ± 0.645	23.580 ± 0.867	26.489 ± 0.978	23.860 ± 0.863
Pre-dorsal Length	48.439 - 51.250	49.692 - 53.648	49.207 - 54.907	47.026 - 52.257
	49.783 ± 0.888	51.502 ± 1.045	51.830 ± 1.361	50.911 ± 1.242
Pre-pelvic Length	44.535 - 47.084	46.258 - 50.000	45.077 - 51.034	44.738 - 49.688
	45.729 ± 0.913	47.829 ± 1.209	47.853 ± 1.837	47.556 ± 1.478
Pre-anal Length	67.495 - 70.250	67.860 - 73.059	63.101 - 92.795	64.209 - 70.973
	68.504 ± 1.113	70.365 ± 1.434	67.190 ± 6.442	68.205 ± 1.444
Dorsal fin Depth	19.419 - 21.050	21.598 - 24.188	26.375 - 32.435	20.488 - 23.815
	20.174 ± 0.683	22.717 ± 0.693	29.187 ± 1.433	22.489 ± 0.929
Dorsal fin base Length	9.920 - 10.450	9.964 - 11.817	11.593 - 15.456	9.354 - 12.343
	10.178 ± 0.203	10.826 ± 0.540	13.174 ± 1.124	11.187 ± 0.799
Anal fin Depth	15.833 - 16.450	16.026 - 22.781	18.374 - 23.016	16.568 - 19.720
	16.078 ± 0.267	17.973 ± 1.582	20.068 ± 1.190	18.219 ± 0.887
Anal fin base Length	9.173 - 11.234	9.531 - 12.338	10.166 - 13.473	10.168 - 12.392
	9.999 ± 0.669	10.467 ± 0.679	11.829 ± 0.979	11.071 ± 0.599
Pectoral fin Length	18.178 - 20.050	20.565 - 23.165	19.373 - 24.950	19.502 - 21.212
	18.918 ± 0.632	21.817 ± 0.747	21.793 ± 1.442	20.454 ± 0.456

Table 4.1 continued

	R. argyrotaenia	*R. myersi*	*R. sarawakensis*	*Rasbora sp.*
Pelvic fin Length	15.888 - 18.180	16.744 - 19.108	15.405 - 20.570	16.297 - 19.720
	16.772 ± 0.736	17.925 ± 0.610	18.156 ± 1.372	18.261 ± 0.858
Body Depth	17.879 - 26.343	22.139 - 25.527	27.702 - 33.135	21.383 - 26.727
	21.577 ± 2.540	23.916 ± 0.921	30.233 ± 1.506	23.520 ± 1.126
Caudal peduncle Length	9.748 - 11.000	11.281 - 13.056	11.348 - 14.753	10.842 - 12.990
	10.573 ± 0.444	12.052 ± 0.495	13.158 ± 0.866	12.055 ± 0.495
Upper caudal lobe Length	25.449 - 29.230	28.955 - 34.298	29.126 - 37.737	27.657 - 32.962
	27.061 ± 1.419	32.133 ± 1.619	33.897 ± 2.094	29.668 ± 1.436
Middle caudal lobe Length	8.933 - 10.636	9.386 - 13.624	11.701 - 15.867	10.752 - 13.573
	9.956 ± 0.707	11.133 ± 1.316	13.538 ± 1.036	12.151 ± 0.794
Lower caudal lobe Length	28.194 - 32.200	32.898 - 39.510	31.994 - 39.920	28.672 - 35.993
	30.250 ± 1.464	35.395 ± 2.124	34.347 ± 1.935	32.822 ± 1.929
Dorsohypural Distance	36.625 - 38.350	32.107 - 37.520	31.613 - 36.848	32.537 - 39.069
	37.736 ± 0.570	35.167 ± 1.484	33.968 ± 1.420	35.772 ± 1.378
Caudal peduncle Length	17.834 - 19.720	13.594 - 21.089	15.555 - 21.909	15.535 - 18.914
	18.561 ± 0.634	16.310 ± 2.047	18.619 ± 1.663	17.126 ± 0.915
Percentage of HL				
Eye diameter	23.773 - 27.236	26.404 - 29.955	26.002 - 38.545	25.149 - 30.979
	26.035 ± 1.225	28.072 ± 1.194	32.723 ± 3.814	28.492 ± 1.599
Snout Length	15.455 - 21.084	15.648 - 27.320	13.119 - 56.575	18.240 - 24.053
	18.499 ± 1.668	20.826 ± 3.400	21.366 ± 9.588	21.234 ± 1.532
Head width	36.409 - 40.049	35.262 - 41.047	32.877 - 45.745	36.870 - 43.966
	38.274 ± 1.266	37.872 ± 1.802	37.732 ± 3.007	40.459 ± 2.274
Interorbital width	17.802 - 26.331	21.696 - 27.100	19.589 - 27.500	24.277 - 30.893
	24.077 ± 3.037	24.802 ± 1.690	23.702 ± 2.299	28.219 ± 1.713
Head Depth	46.273 - 50.833	49.737 - 57.887	57.421 - 75.000	45.347 - 78.856
	48.672 ± 1.997	54.602 ± 2.971	63.573 ± 4.322	55.272 ± 7.662

Table 4.2: Meristic data of *Rasbora*. Cell contains**:** range and mode.

	R. argyrotaenia	*R. myersi*	*R. sarawakensis*	*Rasbora* sp.
Dorsal fin rays	8-9	8-9	7-8	7-9
	8	8	7	8
Pectoral fin rays	14-16	15-16	12-13	13-16
	15	15	12	15
Pelvic fin rays	9	9	7-8	8-10
	9	9	8	9
Anal fin rays	7	6-7	7	7
	7	7	7	7
Caudal fin rays	19	19	19	18-21
	19	19	19	19
Dorsal procurrent rays of caudal fin	5	4-6	5-6	4-5
	5	5	5	4
Ventral procurrent rays of caudal fin	4-5	5-6	5-7	4-6
	5	6	6	5
Scales of lateral line	31-33	27-30	24-26	26-30
	32	29	26	28
Predorsal scales	12-13	11-14	9-11	11-13
	12	13	10	12
Transverse scales	6-8	6-7	6-7	6-7
	7	7	6	7
Circumpeduncular scales	13-15	13-14	12-14	12-14
	13	14	13	13

4.2.1 Morphometric Statistical Analyses

Prior to statistical analysis, it has to be noted that the original data was transformed using allometric formula to remove the size effect of organism before subjected to one-way analysis of variance (ANOVA). The results from one-way ANOVA in Table 4.3 showed 20 out of 23 morphometric characters are having $p < 0.05$, which means 20 morphometric characters are significantly different among four species of *Rasbora*. Among all the morphometric characters, the preanal length, body depth and snout length were observed to be not significantly different.

Table 4.3: Analysis of variance (ANOVA) of morphometric characters among *Rasbora*.

	Wilks' lambda	F value	P value
Total length	0.715	7.586	0.000
Head length	0.723	7.286	0.000
Predorsal length	0.843	3.535	0.020
Prepelvic length	0.829	3.914	0.013
Preanal length	0.950	1.005	0.397
Dorsal fin depth	0.743	6.580	0.001
Dorsal fin base length	0.799	4.783	0.005
Anal fin depth	0.747	6.425	0.001
Anal fin base length	0.786	5.182	0.003
Pectoral fin length	0.581	13.696	0.000
Pelvic fin length	0.807	4.542	0.006
Body depth	0.889	2.373	0.080
Caudal peduncle length	0.648	10.317	0.000
Upper caudal lope length	0.625	11.413	0.000
Median caudal lope length	0.653	10.078	0.000
Lower caudal lope length	0.625	11.413	0.000
Dorsohypural distance	0.819	4.199	0.009
Caudal peduncle length	0.800	4.743	0.005
Eye diameter	0.718	7.471	0.000
Snout length	0.953	0.937	0.429
Head width	0.613	12.005	0.000
Inter-orbital width	0.508	18.385	0.000
Head depth	0.776	5.498	0.002

In this study for morphometric characters, the value of KMO for overall matrix was 0.744. The Bartlett's Test of sphericity was significant at $p < 0.05$ (Table 4.4). The results from KMO and Bartlett's suggested that the sampled data was appropriate to proceed with a factor analysis procedure. In PCA, the characters with an eigenvalue exceeding 1 were included while the characters with an eigenvalue less than 1 were discarded. Principal component analysis of 20 morphometric characters extracted five components with eigenvalue greater than 1 (Table 4.5). The five components explaining 65.92% of total variability of morphometric traits of *Rasbora* species (Table 4.5).

Table 4.4: The results of Kaiser-Meyer-Olkin measure and Bartlett's test of Sphericity for morphometric characters.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.744
Bartlett's Test of Sphericity	Approx. Chi-Square	587.168
	df	190
	Sig.	0.000

The first PC have the most variance and highest eigenvalue. The next component accounted as much of the left-over variance as it can and the same continued till the last factor. The eigenvalue of first five components of morphometric characters (Table 4.5) was 6.28, 2.50, 1.65, 1.57, and 1.19 respectively, which accounted for 31.41%, 12.49%, 8.24%,7.82% and 5.95% of variations respectively.

The component loading of the PC were presented in Table 4.6. The most important component to differentiate species is the component with high variances. From the result, the most significant loading contributing on PC1 are lower caudal lobe length, upper caudal lobe length, total length, pectoral fin length and anal fin length while on PC2 are eye diameter, middle caudal lobe length, interorbital width, head depth and head length.

Visual examination of plots of PC1 and PC2 scores (Figure 4.5) for the morphometric characters revealed all the 61 samples randomly scattered together. The

observation from the plot in Figure 4.5 show that the *R. sarawakensis* (green ring) disperse randomly among other species of *Rasbora*. In brief, the scatters plot of principal component did not show any clear cluster among the *Rasbora* species populations.

Table 4.5: Eigenvalues, percentage of variance and percentage of cumulative variance for the principal component of morphometric characters among the *Rasbora*.

	Eigenvalue		
Component	Total	% of Variance	Cumulative %
1	6.283	31.414	31.414
2	2.498	12.491	43.905
3	1.647	8.237	52.142
4	1.565	7.823	59.965
5	1.191	5.954	65.918

Table 4.6: Component loading of each morphometric character after rotation on the first 5 principal components.

	Component				
	1	2	3	4	5
Lower caudal lobe length	0.817				
Upper caudal lobe length	0.816				
Total length	0.790				
Pectoral fin length	0.667				
Anal fin depth	0.548				
Pelvic fin length	0.468				
Eye diameter		0.862			
Middle caudal lobe length		0.767			
Interorbital width		0.745			
Head depth		0.596			
Head length		0.517			
Anal fin base length			0.756		
Dorsal fin depth			0.679		
Head width			0.586		
Dorsal fin base length			0.565		
Caudal peduncle depth			0.454		
Prepelvic length				0.804	
Predorsal length				0.726	
Caudal peduncle length					0.780
Dorsohypural distance					0.657

Extraction Method**:** Principal Component Analysis.
Rotation Method**:** Varimax with Kaiser Normalization.

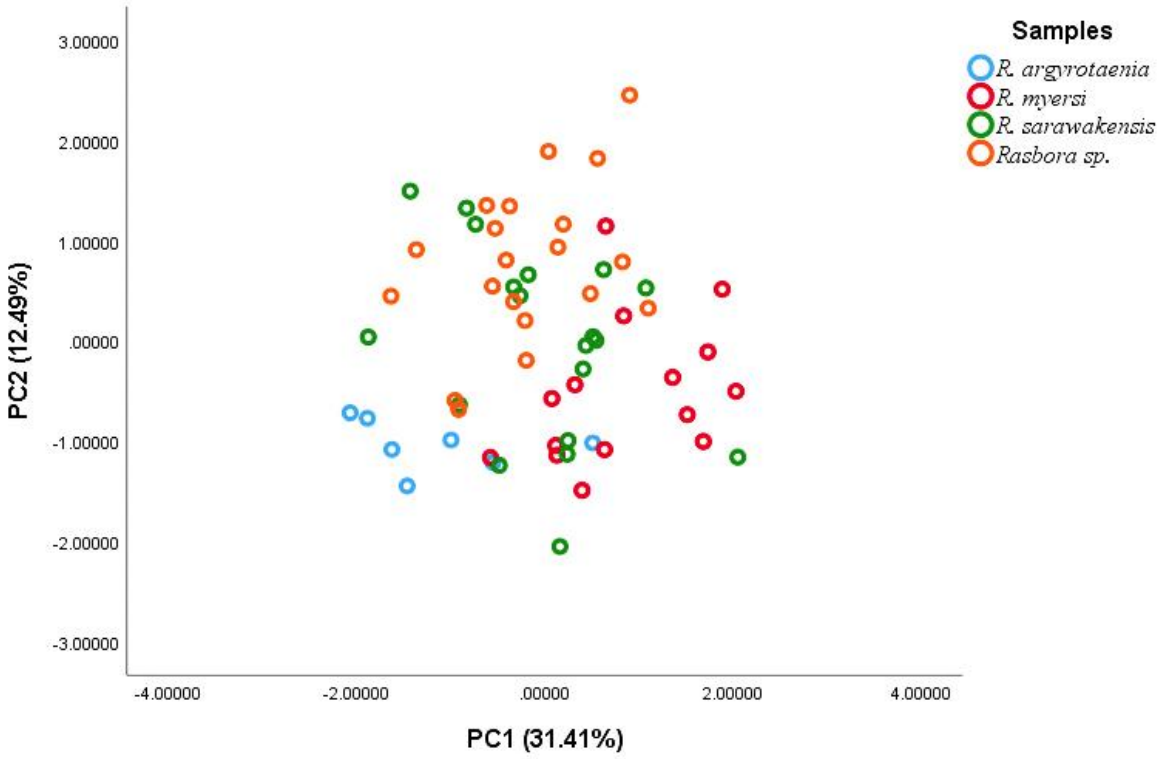


Figure 4.5: Plot of the factor scores for PC1 and PC2 of morphological characters of four *Rasbora* species.

The discriminant function analysis was subsequently used to determine which morphometric measurements can function to discriminate the four *Rasbora* species and also used to predict each *Rasbora* fish into correct group membership. The Wilk's lambda test is to test which variable contribute significantly in discriminant function. The smaller the value of Wilks' lambda is better. In the present case the value of Wilk's lambda is 0.049 (Table 4.8). The significance of Wilk's lambda was tested using chi-square, the chi-square is recorded at 144.77 and the p-value less than 0.05 which means that the function explains the group membership well.

Discriminant function analysis produced three discriminant functions but only functions with high eigenvalue are of interest. The larger the eigenvalue, the more variance in the dependent variable is explained by the function. The eigenvalue for the first two functions is 4.351 (DF1) and 2.602 (DF2) (Table 4.7), explaining 62.1% and 37.1% of

variance accounted for DF1 and DF2 respectively. The canonical relation is a correlation between the discriminant scores and the level of these dependent variables. The higher the correlations value, the better the function that discriminates the value, 1 is considered as perfect. Here, the correlation of 0.902 (DF1) and 0.850 (DF2) is comparatively high.

Table 4.7: Eigenvalue, percentage of variance, percentage of cumulative of variance and canonical correlation for the discriminant function.

Eigenvalue

Functions	Eigenvalue	% of variance	Cumulative %	Canonical Correlation
1	4.351	62.1	62.1	0.902
2	2.602	37.1	99.2	0.850
3	0.059	0.8	100.0	0.236

Table 4.8: Result of Wilks' lambda test for verifying differences among *Rasbora* species when morphological measurements are separately compared using discriminant function analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	0.049	144.777	60	0.000
2 through 3	0.262	64.268	38	0.005
3	0.944	2.761	18	1.000

Plotting DF1 and DF2 showed a clear discrimination between dependent variable (Figure 4.6). The group centroids were positioned far from each other, showing a well group discrimination. All the samples were clustered near their respective group centroid. From Figure 4.6, the *R. argyrotaenia*, *R. myersi* and *Rasbora sp.* totally isolated from each other. *R. sarawakensis* located at the middle of the other 3 species of *Rasbora*, and very close with *Rasbora sp.*.

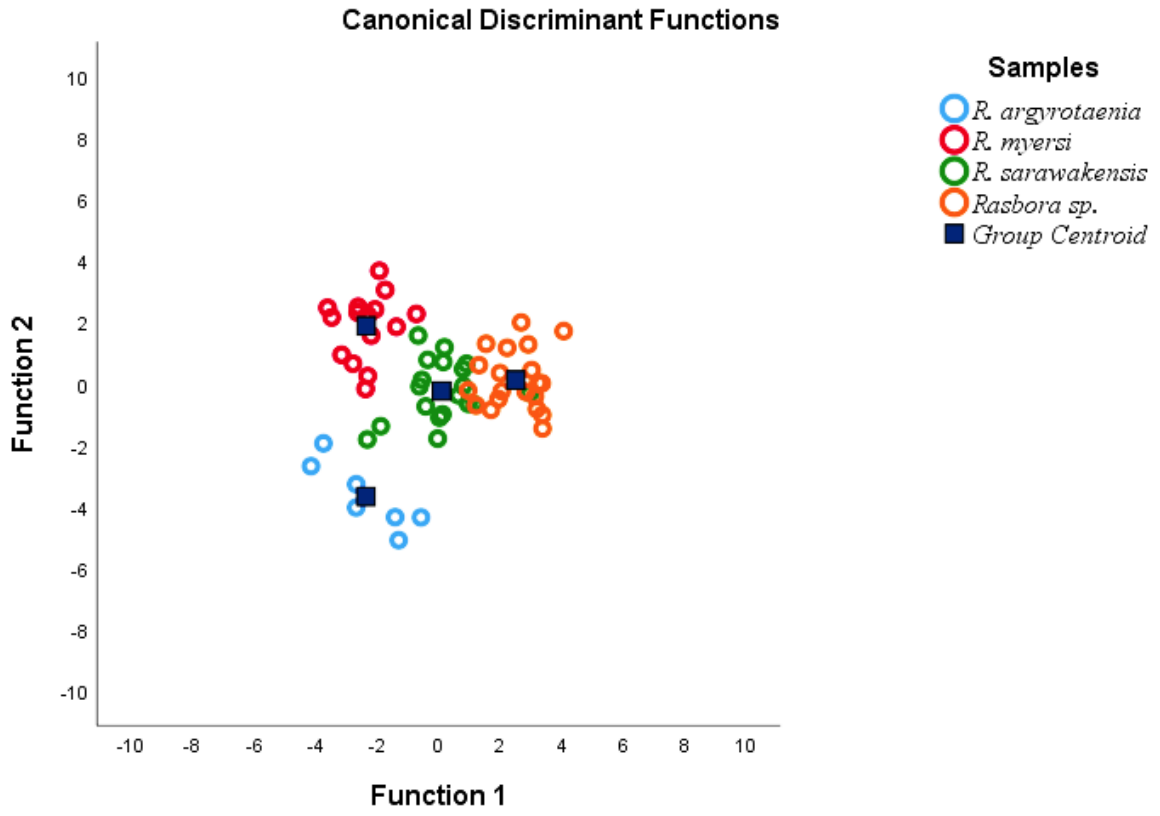


Figure 4.6: Coordinate plot of *Rasbora* samples according to the first two discriminant functions from morphometric data analysis.

Pooled within-group correlations between discriminating variables and DFs revealed that the four head measurements (interorbital width, head width, eye diameter and head length) and a tail measurement (middle caudal lobe length) contributed to the first DF (Table 4.9). The measurements that contributed to the second DF (Table 4.9) were the pectoral fin length, upper caudal lobe length, lower caudal lobe length, total length, caudal peduncle depth, caudal peduncle length, dorsohypural distance, predorsal length, prepelvic length, anal fin depth and pelvic fin length.

A correct classification of individual into their original population varied from 89.5 - 100% by discriminant analysis and 93.4% of individuals were classified in their correct grouping (Table 4.10). The proportion of correctly classified *R. argyrotaenia* and *R. myersi* into their original group was 100% correct.

Table 4.9: Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.

	Discriminant Function		
	1	2	3
Interorbital Width	0.450*	0.180	0.180
Head Width	0.373*	0.099	-0.027
Middle caudal lobe Length	0.329*	0.150	-0.027
Eye Diameter	0.261*	0.193	-0.107
Head Length	0.256*	0.193	-0.052
Pectoral fin Length	-0.061	0.517*	-0.395
Upper caudal lobe Length	-0.058	0.475*	0.038
Lower caudal lobe Length	-0.076	0.470*	0.063
Total Length	-0.039	0.388*	0.134
Caudal peduncle Depth	0.208	0.369*	-0.136
Caudal peduncle Length	0.035	-0.306*	-0.131
Dorsohypural Distance	-0.025	-0.288*	0.204
Predorsal Length	0.045	0.261*	0.068
Prepelvic Length	0.093	0.254*	-0.038
Anal fin Depth	0.202	0.247*	-0.177
Pelvic fin Length	0.170	0.207*	-0.150
Dorsal Depth	0.206	0.238	-0.484*
Anal fin base Length	0.242	0.066	-0.323*
Head depth	0.204	0.200	-0.274*
Dorsal fin base Length	0.232	0.072	0.268*

Asterisk denotes the largest correlation between each variable and discriminant functions.

From the discrimination classification, only 17 out of 19 samples from *R. sarawakensis* was correctly classified into their original group, and the other 2 samples were predicted classified under *R. argyrotaenia* and *Rasbora sp.* each. The same goes to *Rasbora sp.* only 18 of individuals were correctly classified into their original group, while 2 individuals were predicted classified under *R. sarawakensis.*

Table 4.10: Count and percentage of samples classified in each group for morphometric data.

	R. argyrotaenia	*R. myersi*	*R. sarawakensis*	*Rasbora sp.*	Total
R. argyrotaenia	7	0	0	0	7
	100.0%	0.0%	0.0%	0.0%	100.0%
R. myersi	0	15	0	0	15
	0.0%	100.0%	0.0%	0.0%	100.0%
R. sarawakensis	1	0	17	1	19
	5.3%	0.0%	89.5%	5.3%	100.0%
Rasbora sp.	0	0	2	18	20
	0.0%	0.0%	10.0%	90.0%	100.0%

93.4% of original grouped cases correctly classified.

4.2.2 Meristic Statistical Analyses

These meristic characters were mainly based on rays and scales counted. The Kruskal Wallis test (Table 4.11) shows 8 out of 11 meristic characters was significantly difference at $p < 0.05$. Among all the meristic characters, the anal fin rays, caudal fin rays and transverse scales shows no significant difference at $p > 0.05$ (Table 4.11). All the significant meristic characters were used in multivariate analysis, PCA and DFA. The suitability of meristic data for PCA was examined by measuring the Bartlett's Test of Sphericity and Kaiser-Meyer-Olkin (KMO). Here for meristic data, the value of KMO for overall matrix was 0.826. And the Bartlett's Test of Sphericity was significant $p < 0.05$ (Table 4.12). The results (KMO and Bartlett's) suggested that the sampled data was appropriate to proceed with a PCA procedure.

The PCA of eight meristic characters extracted two principal components with eigenvalues greater than 1. The Table 4.13 shows the eigenvalue, percentage of variance and cumulative percentage of total variance. The eigenvalue of PC1 was found to be 4.246 and PC2 was 1.292. The PC1 and PC2 was accounted for 53.072% and 16.150% of the variation respectively, explaining 69.222% of the total variability.

Table 4.11: Univariate analysis (Kruskal Wallis rank test) of meristic character among *Rasbora* species.

	Kruskal Wallis H	P value
Dorsal fin rays	28.174	0.000
Pectoral fin rays	43.026	0.000
Pelvic fin rays	50.055	0.000
Anal fin rays	3.067	0.381
Caudal fin rays	0.706	0.872
Dorsal pro-current rays	24.410	0.000
Ventral pro-current rays	29.495	0.000
Scales of lateral line	48.126	0.000
Predorsal scales	45.664	0.000
Transverse scales	6.091	0.107
Circumpenduncular scales	8.136	0.043

Table 4.12: The results of Kaiser-Meyer-Olkin measure and Bartlett's test of Sphericity for meristic characters.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.826
Bartlett's Test of Sphericity	Approx. Chi-Square	256.874
	df	28
	Sig.	0.000

Table 4.13: Eigenvalues, percentage of variance and percentage of cumulative variance for the principal component of meristic characters among the *Rasbora* species.

Eigenvalues

Component	Total	% of variance	Cumulative %
1	4.246	53.072	53.072
2	1.292	16.150	69.222

Table 4.14 shows the component loading after rotation of meristic characters on extracted PC. The most significant loading on PC1 after rotation were predorsal scales, scales of lateral line, pectoral fin rays, pelvic fin rays and dorsal fin rays. While for PC2 the significant loadings were ventral procurrent rays, dorsal procurrent rays and circumpenduncular scales.

Visual examination of plots of PC1 and PC2 scores (Figure 4.7) in the meristic characters revealed that overlapping some of the samples from *R. argyrotaenia*, *R. myersi* and *Rasbora sp., R. sarawakensis* scattered separately from other 3 species of *Rasbora*. The scatter plot of PC (Figure 4.7) show moderate separation among the *Rasbora* species samples.

Table 4.14: Component loading after rotation of each meristic character on the first 2 principal components.

	Component	
	1	2
Predorsal scales	0.901	
Scales of lateral line	0.849	
Pectoral fin rays	0.845	
Pelvic fin rays	0.835	
Dorsal fin rays	0.738	
Ventral procurrent rays	-0.603	
Dorsal procurrent rays		0.821
Circumpenduncular scales		0.587

Extraction Method**:** Principal Component Analysis.
Rotation Method**:** Varimax with Kaiser Normalization.

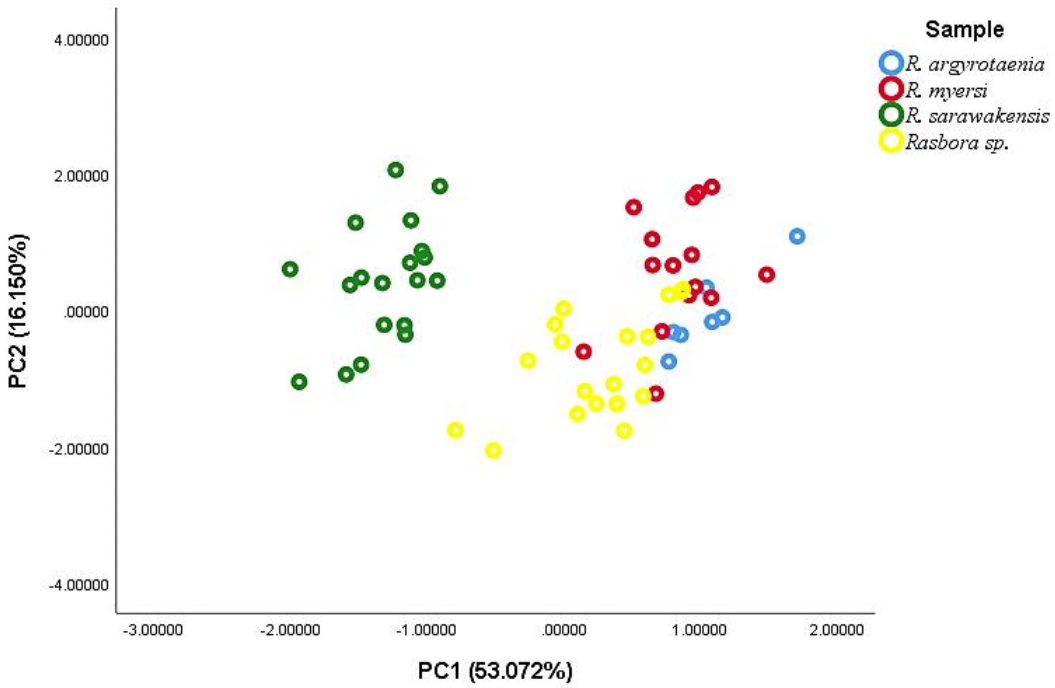


Figure 4.7: Plot of the factor scores for PC1 and PC2 of meristic characters of four *Rasbora* species.

The DFA was used to determine which meristic character can effectively discriminate the four species of *Rasbora*. The good model for DF was measured by the eigenvalue, the bigger the eigenvalue, the stronger discriminating power of the function. Table 4.15 shows the eigenvalue, percentage of variance and canonical correlation of each DFs.

In this study, the first two functions that show strong discriminating power, where the eigenvalue was greater than 1, DF1 was 12.171 which accounted for 80.2% of the between-group variability, and DF2 was 2.298 which accounted for 15.1% of the between-group variability. And the degree of relationship between meristic counts of *Rasbora* and the function (canonical correlation) was found to be strong at 0.961 (DF1) and 0.835 (DF2).

The Wilks' lambda test of DFA indicates how these significant meristic function discriminates the four species of *Rasbora*. The DFA shows significant Wilks' lambda statistic (0.014, chi-square 232.342; 0.178, chi-square 93.128; 0.588, chi-square 28.692) (Table 4.16) shows that the model differentiated score among the group significantly.

Figure 4.8 shows the plotting of DF1 and DF2, which explained 95.4% of the total between-group variation and revealed between population differences. The group centroid of each species of *Rasbora* was obviously separated. From the illustration plotting, *R. argyrotaenia* and *R. sarawakensis* isolated from other species, while *Rasbora sp*. and *R. myersi* scattered randomly close with each other.

Table 4.17 shows the pooled within groups correlation between discriminant variables and discriminant functions revealed that among the eight meristic characters**:** pectoral fin rays, pelvic fin rays and dorsal fin rays contributed to the first DF, and only scales of lateral line count contributed to the second DF, and the rest of meristic count contributed to the third DF.

A correct classification of individual into their original population varied from 80.0% - 100.0% by discriminant analysis and 90.2% of individuals were classified in their correct group membership (Table 4.18). The proportion of correctly classified *R. argyrotaenia* and *R. sarawakensis* into their original group was 100% correct.

From the discrimination classification predicted, 13 out of 15 samples from *R. myersi* were correctly classified into their original group, and the other two samples were predicted to be classified under *Rasbora sp.*. Also, 16 samples from *Rasbora sp.* was correctly classified into their original group, a sample was classified under *R. argyrotaenia*, and another three were classified under *R. myersi.*

Table 4.15: Eigenvalue, percentage of variance, percentage of cumulative of variance and canonical correlation for the discriminant function.

Eigenvalues

Functions	Eigenvalue	% of variance	Cumulative %	Canonical correlation
1	12.171	80.2	80.2	0.961
2	2.298	15.1	95.4	0.835
3	0.701	4.6	100.0	0.642

Table 4.16: Result of Wilks' lambda test for verifying differences among *Rasbora* species population when morphological measurements are separately compared using discriminant function analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	0.014	232.342	24	0.000
2 through 3	0.178	93.128	14	0.000
3	0.588	28.692	6	0.000

Table 4.17: Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.

	Function		
	1	2	3
Pectoral fin rays	0.509*	-0.155	0.293
Pelvic fin rays	0.489*	-0.217	-0.144
Dorsal fin rays	0.252*	-0.107	-0.074
Scales of lateral line	0.594	0.646*	0.148
Dorsal procurrent rays	-0.135	0.195	0.740*
Predorsal scales	0.492	-0.193	0.532*
Ventral procurrent rays	-0.253	-0.075	0.511*
Circumpenduncular scales	0.076	0.072	0.351*

Asterisk denotes largest absolute correlation between each variable and any discriminant function

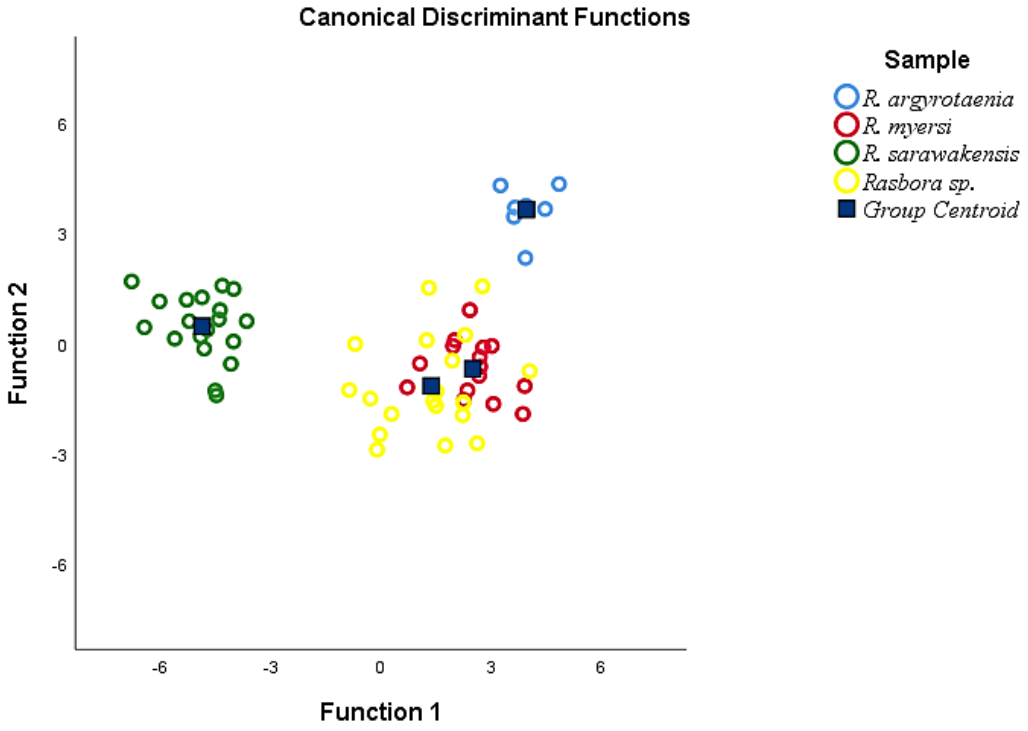


Figure 4.8: Coordinate plot of *Rasbora* samples according to the first two discriminant functions from meristic data analysis.

Table 4.18: Percentage of samples classified in each group for meristic data. Cell contains**:** Original count and percentage.

	R. argyrotaenia	*R. myersi*	*R. sarawakensis*	*Rasbora sp.*	Total
R. argyrotaenia	7 100.0%	0 0.0%	0 0.0%	0 0.0%	7 100.0%
R. myersi	0 0.0%	13 86.7%	0 0.0%	2 13.3%	15 100.0%
R. sarawakensis	0 0.0%	0 0.0%	19 100.0%	0 0.0%	19 100.0%
Rasbora sp.	1 5.0%	3 15.0%	0 0.0%	16 80.0%	20 100.0%

90.2% of original grouped cases correctly classified

4.3 Mitochondrial DNA Analyses

4.3.1 Amplification of Mitochondrial DNA

The mtDNA of *R. argyrotaenia* was amplified by using two pairs of primer as shown in Table 4.19. So, predictedly two DNA fragments will be produced whereby the expected size was approximately 9,398 bp and 9,187 bp respectively. Those two fragments were enough to be used to construct a complete sequence of *R. argyrotaenia* mt-genome. Temperature optimization of the primer was conducted to determine the optimum temperature of the primer. From Figure 4.9, the DNA band appear at approximately 9 kb and 10 kb position as expected.

A complete mtDNA sequence of *R. myersi* mt-genome was constructed using two pair of primers (Table 4.19). The expected size of the first and second fragment was about 10,000 bp and 8,924 bp in length. Here, it should be noted that the primers used to construct *R. myersi* mt-genome was different from primer used on *R. argyrotaenia* mt-genome. Since the primer used were designed as well, the annealing temperature of the primers was optimized (Figure 4.10). From Figure 4.10, the DNA produced from both primers were located at the expected position.

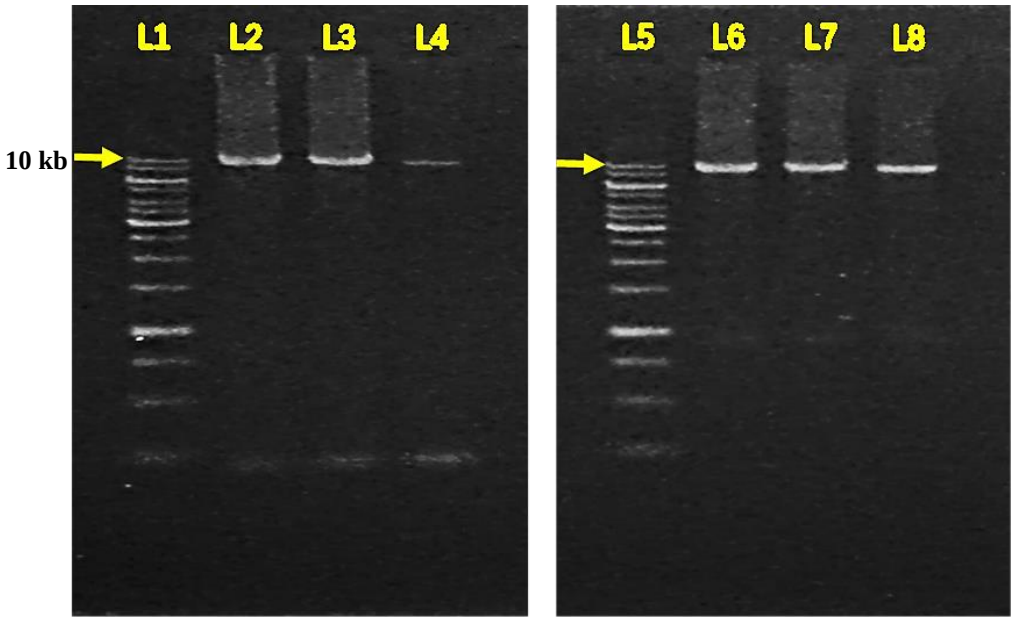


Figure 4.9: Agarose gel electrophoresis results of mtDNA PCR for *R. argyrotaenia*. L1 and L5 is 1 kb ladder. Fragment A**:** L2 (48.6°C), L3 (50.9°C) and L4 (53.0°C); Fragment B**:** L6 (48.6°C), L7 (50.9°C) and L8 (53.0°C). The (→) indicate the reference band for 10 kb.

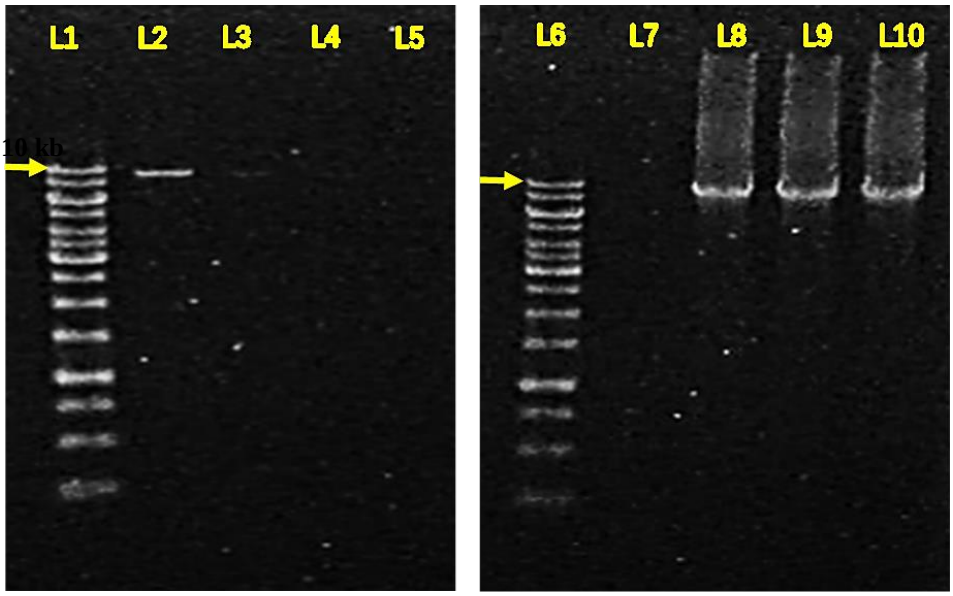


Figure 4.10: Agarose gel electrophoresis results of mtDNA PCR for *R. myersi*. L1 and L6 is 1 kb ladder. Fragment A**:** L2 (50.2°C), L3 (51.0°C), L4 (52.9°C) and L5 (54.2°C); Fragment B**:** L7 (54.2°C), L8 (52.9°C), L9 (51.0°C) and L10 (50.2°C). The (→) indicate the reference band for 10 kb.

A complete mtDNA sequence of *Rasbora sp.* was also constructed from two pairs of designed primer. Coincidentally, the two pair of primers used to amplify *Rasbora sp.* mt-

genome was the same primer used to amplify *R. argyrotaenia*. The temperature optimization of primer also done as in optimization at *R. argyrotaenia*. Although the test temperature was the same, but on *Rasbora sp.* mtDNA amplification, both primers produced multiple bands (Figure 4.11) whereby the correct band were gel extracted for sequencing purpose.

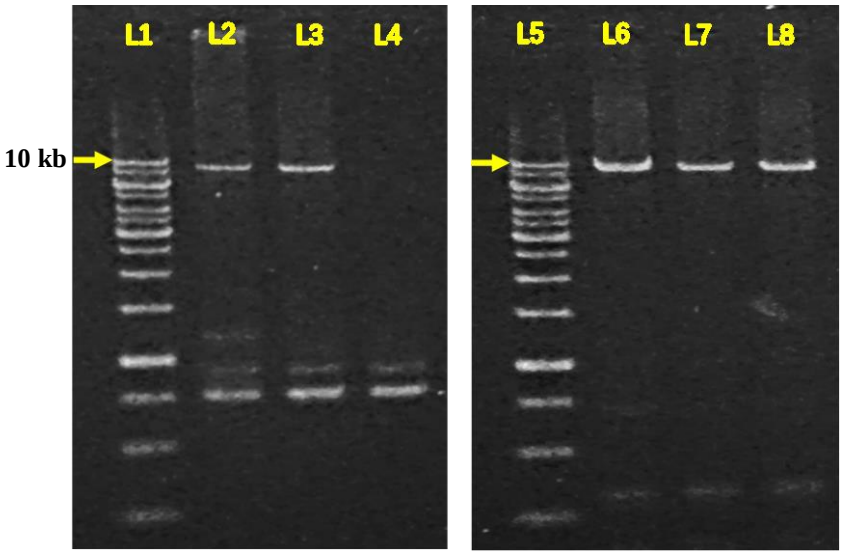


Figure 4.11: Agarose gel electrophoresis results of mtDNA PCR for *Rasbora sp.*. L1 and L5 is 1 kb ladder. Fragment A: L2 (48.6°C), L3 (50.9°C) and L4 (53.0°C); Fragment B: L6 (48.6°C), L7 (50.9°C) and L8 (53.0°C). The (→) indicate the reference band for 10 kb.

The mtDNA of *R. sarawakensis* was amplified using 3 set of primers (Table 4.19). The expected size of amplicons produced by primer A, B and C was 8,179 bp, 6,500 bp and 5,458 bp respectively. The temperature optimization for all pair of primer was done as well. The results of temperature optimization of the three pair of primer is shown in Figure 4.12. All three primers produced multiple bands whereby the bands with correct size were gel extracted and subjected to sequencing.

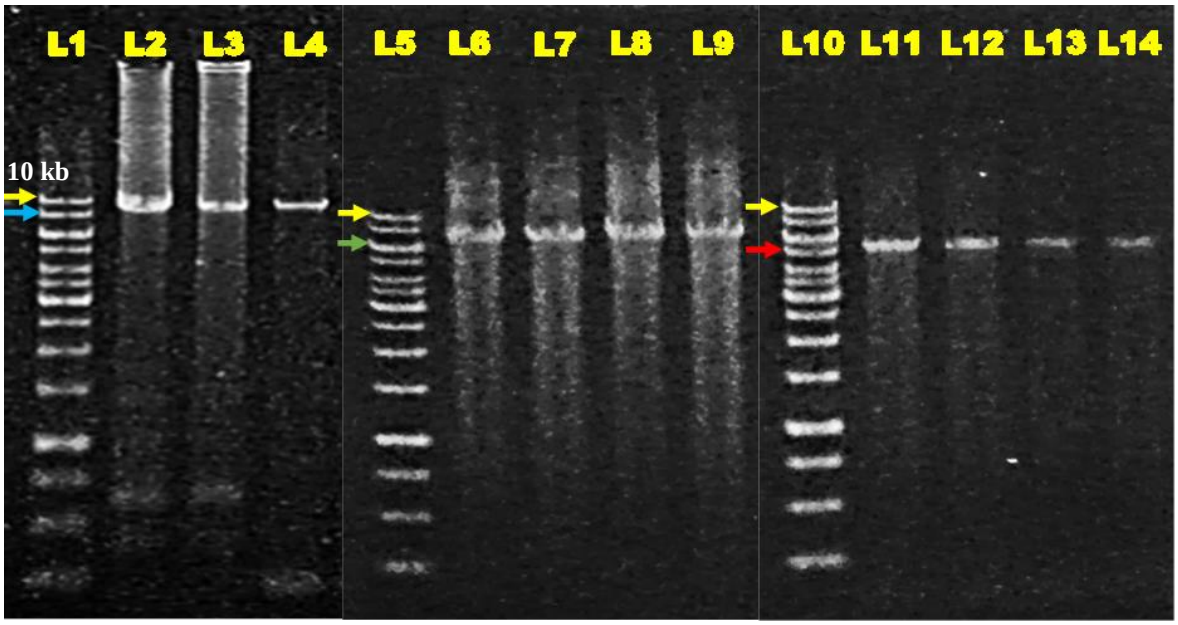


Figure 4.12: Agarose gel electrophoresis results of mtDNA PCR for *R. sarawakensis*. L1, L5 and L10 is 1 kb ladder. Fragment A**:** L2 (50.1°C), L3 (52.3°C) and L4 (54.3°C); Fragment B**:** L6 (50.7°C), L7 (52.3°C), L8 (53.6°C) and L9 (54.5°C); Fragment C**:** L11 (50.0°C), L12 (51.4°C), L13 (52.8°C) and L14(53.5°C). The arrow indicates the reference band, ( ) for 10 kb, ( ) for 8 kb, ( ) for 6 kb and ( ) for 5 kb.

Table 4.19: List of primers used to amplify *Rasbora* mtDNA.

Species		Primer code	Nucleotide sequence (5'-3')	Length size (bp)	T_m (°C)
R. myersi	A	1A-F	CCTATCTTACCGAGAAAG	10000	50.2
		1A-R	GAGGCCTTCCCATCTAGA		
	B	1B-F	GTGCTTCCTCTACACCAC	8924	50.2
		1B-R	TGATGTTGAGAAGGCTAC		
R. argyrotaenia	A	2A-F	ATTCGCCATTATAGCAGG	9398	48.6
Rasbora sp.		2A-R	AGATGCCGCTTACCACTC		
	B	2B-F	CACACCTCAAAACAACGAG	9187	48.6
		2B-R	AGGTGGTCGTTAATCAC		
R. sarawakensis	A	3A-F	CTAGAGGACGCCTCACTA	8179	52.2
		3A-R	AGATGCCGCTTACCACTC		
	B	3B-F	CACACCTCAAAACAACGAG	6500	53.6
		3B-R	GAGGCCTTCCCATCTAGA		
	C	3C-F	AAACCTAGCAGAAACCAA	5458	50.0
		3C-R	GTGGAATCAGATTGCTAG		

4.3.2 Mitochondrial Genome Organization

In general, the complete mt-genome of *Rasbora* was circular DNA molecule which is approximately 16 kb long. The complete mtDNA sequence of *R. argyrotaenia*, *R. myersi* (MK875250), *R. sarawakensis* (MK733839) and *Rasbora sp.* were 16,574 bp, 16,581 bp, 16,709 bp and 16,518 bp in length, respectively. The complete mt-genome of *Rasbora* contains 13 protein-coding genes (PCGs), 22 transfer RNA genes, 2 ribosomal RNA genes and 2 non-coding regions (Control region and origin of L-strand replication).

The *Rasbora* mt-genome are divided into two strands, heavy (H-strand) and light (L-strand) strand. Most of the genes were encoded on the heavy strand (H-strand) including 12 PCGs (ND1, ND2, ND3, ND4L, ND4, ND5, ATP8, ATP6, COI, COII, COIII and *Cyt* b), two rRNAs (12s rRNA and 16s rRNA) and 14 tRNAs, while on the light strand (L-strand) encoded only one PCG (ND6) and eight tRNAs (Gln, Ala, Asn, Cys, Tyr, Ser, Glu and Pro).

Several genes were detected to be overlapping each other in all respective mitogenome (Table 4.20 - Table 4.23). The overlapped regions frequently occurred between adjacent PCGs and less likely between tRNAs. Overlapped region detected between PCGs were ATP8–ATP6, ND4L–ND4 and ND5–ND6. Only a single overlapping region found between tRNAs which is Ile–Gln overlap. The overlapping occurred was not found to be in large fragment but rather a few nucleotide bases overlapped at the end of a gene sequence and at the front sequence of subsequent gene.

Intergenic spacers were present in the *Rasbora* mt-genome. In all studied *Rasbora* the common intergenic spacer was from 1 bp to 12 bp. Most of the spacer were detected between tRNA - PCG, and also at the adjacent tRNA genes. The intergenic spacer with length of approximately 34 bp was spotted in all studied *Rasbora*. This intergenic spacer was found in a cluster of 5 tRNA genes and known as OL (origin of L-strand replication). Interestingly, there were spacer spotted other than OL on some studied *Rasbora* mt-genome.

From the mtDNA analyses, a big space with 70 bp (Table 4.23) was detected between ND4-His of *Rasbora sp.*, and also a gap about 27 bp was noticed between COII and Lys gene (Table 4.22) of *R. sarawakensis*. However, the presence of these gaps (between ND4–His or COII–Lys) was not detected in other *Rasbora* mt-genome.

The mtDNA sequence of *Rasbora* started with tRNAPhe and ended with tRNAPro gene followed by the control region. The arrangement of mitochondrial PCGs in *Rasbora* was ND1-ND2-COI-COII-ATP8-ATP6-COIII-ND3-ND4L-ND4-ND5-(ND6). The position of 2 rRNA genes can be consistently found in between tRNAPhe, tRNALeu (UAA) and tRNAVal. The arrangement of genes cluster Cytb–Thr–Pro–CR was also consistently observed in all *Rasbora* mt-genome.

Table 4.20: Characteristics of the mitochondrial genome of *R. argyrotaenia*.

Gene	Position		Length (bp)	Anti-codon	Codon		Intergenic Nucleotide*	Strand
	From	To			Start	Stop		
tRNAPhe	1	69	69	GAA			0	H
12S rRNA	70	1026	957				0	H
tRNAVal	1027	1097	71	TAC			0	H
16S rRNA	1098	2786	1689				0	H
tRNALeu (UUR)	2787	2862	76	TAA			0	H
ND1	2864	3838	975		ATG	TAA	+1	H
tRNAIle	3843	3914	72	GAT			+4	H
tRNAGln	3983	3913	71	TTG			-2	L
tRNAMet	3985	4053	69	CAT			+1	H
ND2	4054	5098	1045		ATG	T--	0	H
tRNATrp	5099	5171	73	TCA			0	H
tRNAAla	5174	5241	68	TGC			+2	L
tRNAAsn	5243	5315	73	GTT			+1	L
tRNACys	5350	5415	66	GCA			+34	L
tRNATyr	5417	5486	70	GTA			+1	L
COI	5488	7038	1551		GTG	TAA	+1	H
tRNASer (UCN)	7039	7109	71	TGA			0	L
tRNAAsp	7111	7180	70	GTC			+1	H
COII	7185	7875	961		ATG	T--	+4	H
tRNALys	7876	7950	75	TTT			0	H
ATP8	7953	8117	165		ATG	TAA	+2	H
ATP6	8111	8790	680		ATG	TA-	-7	H
COIII	8791	9575	785		ATG	TA-	0	H
tRNAGly	9576	9646	71	TCC			0	H
ND3	9647	9995	349		ATG	T--	0	H
tRNAArg	9996	10065	70	TCG			0	H
ND4L	10066	10362	297		ATG	TAA	0	H
ND4	10356	11737	1382		ATG	TA-	-7	H
tRNAHis	11738	11806	69	GTG			0	H
tRNASer (AGY)	11807	11874	68	GCT			0	H
tRNALeu (CUN)	11877	11949	73	TAG			+2	H
ND5	11950	13779	1830		ATG	TAA	0	H
ND6	13776	14297	522		GTG	TAG	-4	L
tRNAGlu	14298	14366	69	TTC			0	L
Cytb	14373	15513	1141		ATG	T--	+6	H
tRNAThr	15514	15584	71	TGT			0	H
tRNAPro	15596	15665	70	TGG			+11	L
D-loop	15666	16574	908				0	-

*Positive number corresponding to the nucleotides separating different gene, whereas negative numbers indicate overlapping nucleotides between adjacent genes; H and L denote heavy and light strands.

Figure 4.13: The gene map of *R. argyrotaenia* mitochondrial genome drawn using GenomeVx software. Transfer RNAs (tRNA) are labelled with their corresponding amino acids and are shown in red; COI, COII and COIII refer to subunits of cytochrome c oxidase; *Cyt* b refers to cytochrome b; 12S rRNA and 16S rRNA refer to ribosomal RNAs; ND1-ND6 refer to components of NADH dehydrogenase; ATPase 6 and ATPase 8 refers to classes of ATP synthase. Genes on the main circle are transcribed clockwise whereas those outside the main circle are transcribed anticlockwise.

Table 4.21: Characteristics of mitochondrial genome of *R. myersi*.

Gene	Position		Length (bp)	Anti-codon	Codon		Intergenic Nucleotide*	Strand
	From	To			Start	Stop		
tRNAPhe	1	69	69	GAA			0	H
12S rRNA	70	1022	953				0	H
tRNAVal	1023	1093	71	TAC			0	H
16S rRNA	1094	2779	1686				0	H
tRNALeu (UUR)	2780	2854	75	TAA			0	H
ND1	2856	3830	975		ATG	TAA	+1	H
tRNAIle	3835	3906	72	GAT			+4	H
tRNAGln	3905	3975	71	TTG			-2	L
tRNAMet	3977	4045	69	CAT			+1	H
ND2	4046	5090	1045		ATG	T--	0	H
tRNATrp	5091	5163	73	TCA			0	H
tRNAAla	5166	5233	68	TGC			+2	L
tRNAAsn	5235	5307	73	GTT			+1	L
tRNACys	5342	5407	66	GCA			+34	L
tRNATyr	5409	5478	70	GTA			+1	L
COI	5480	7030	1551		GTG	TAA	+1	H
tRNASer (UCN)	7031	7101	71	TGA			0	L
tRNAAsp	7103	7172	70	GTC			+1	H
COII	7177	7867	691		ATG	T--	+4	H
tRNALys	7868	7942	75	TTT			0	H
ATP8	7945	8109	165		ATG	TAA	+2	H
ATP6	8103	8782	680		ATG	TA-	-7	H
COIII	8783	9567	785		ATG	TA-	0	H
tRNAGly	9568	9638	71	TCC			0	H
ND3	9639	9987	349		ATG	T--	0	H
tRNAArg	9988	10057	70	TCG			0	H
ND4L	10058	10354	297		ATG	TAA	0	H
ND4	10348	11729	1382		ATG	TA-	-7	H
tRNAHis	11730	11798	69	GTG			0	H
tRNASer (AGY)	11799	11866	68	GCT			0	H
tRNALeu (CUN)	11869	11941	73	TAG			+2	H
ND5	11942	13771	1830		ATG	TAA	0	H
ND6	13768	14289	521		ATG	TAG	-4	L
tRNAGlu	14290	14358	69	TTC			0	L
Cyt b	14365	15505	1141		ATG	T--	+6	H
tRNAThr	15506	15576	71	TGT			0	H
tRNAPro	15588	15657	70	TGG			+11	L
D-loop	15658	16581	923				0	-

*Positive number corresponding to the nucleotides separating different gene, whereas negative numbers indicate overlapping nucleotides between adjacent genes; H and L denote heavy and light strands.

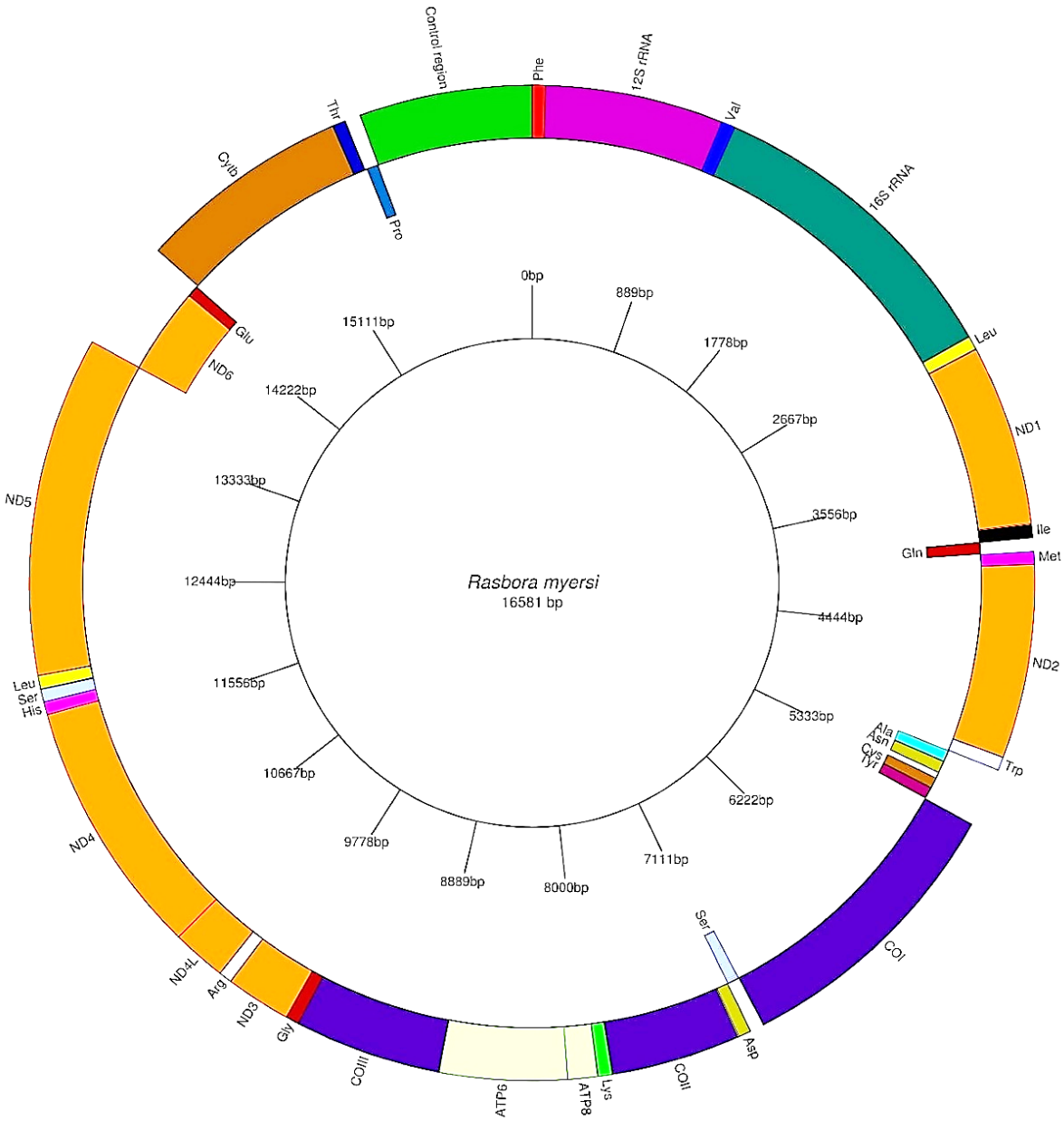


Figure 4.14: The complete mitochondrial genome organization of *R. myersi* drawn using GenomeVx software. Transfer RNAs (tRNA) are labelled with their corresponding amino acids and are shown in red; COI, COII and COIII refer to subunits of cytochrome c oxidase; *Cyt* b refers to cytochrome b; 12S rRNA and 16S rRNA refer to ribosomal RNAs; ND1-ND6 refer to components of NADH dehydrogenase; ATPase 6 and ATPase 8 refers to classes of ATP synthase. Genes on the main circle are transcribed clockwise whereas those outside the main circle are transcribed anticlockwise.

Table 4.22: Characteristics of mitochondrial genome of *R. sarawakensis*.

Gene	Position		Length (bp)	Anti-codon	Codon		Intergenic Nucleotide*	Strand
	From	To			Start	Stop		
tRNAPhe	1	69	69	GAA			0	H
12S rRNA	70	1023	954				0	H
tRNAVal	1024	1094	71	TAC			0	H
16S rRNA	1095	2771	1677				0	H
tRNALeu (UUR)	2772	2846	75	TAA			0	H
ND1	2848	3822	975		ATG	TAA	+1	H
tRNAIle	3827	3898	72	GAT			+4	H
tRNAGln	3897	3967	71	TTG			-2	L
tRNAMet	3969	4037	69	CAT			+1	H
ND2	4038	5084	1045		ATG	TAA	0	H
tRNATrp	5087	5156	70	TCA			+2	H
tRNAAla	5159	5226	68	TGC			+2	L
tRNAAsn	5228	5300	73	GTT			+1	L
tRNACys	5335	5400	66	GCA			+34	L
tRNATyr	5401	5470	70	GTA			0	L
COI	5472	7022	1551		GTG	TAA	+1	H
tRNASer (UCN)	7023	7093	71	TGA			0	L
tRNAAsp	7095	7164	70	GTC			+1	H
COII	7170	7871	691		ATG	TAA	+5	H
tRNALys	7899	7973	75	TTT			+27	H
ATP8	7976	8140	165		ATG	TAA	+2	H
ATP6	8134	8816	683		ATG	TA-	-7	H
COIII	8817	9601	784		ATG	TA-	0	H
tRNAGly	9602	9671	70	TCC			0	H
ND3	9672	10020	349		ATG	T--	0	H
tRNAArg	10021	10090	70	TCG			0	H
ND4L	10091	10387	297		ATG	TAA	0	H
ND4	10381	11762	1382		ATG	TA-	-7	H
tRNAHis	11763	11831	69	GTG			0	H
tRNASer (AGY)	11832	11898	67	GCT			0	H
tRNALeu (CUN)	11900	11972	73	TAG			+1	H
ND5	11973	13802	1830		ATG	TAA	0	H
ND6	13799	14317	522		ATG	TAA	-4	L
tRNAGlu	14321	14389	69	TTC			+3	L
Cyt b	14396	15532	1141		ATG	TAA	+6	H
tRNAThr	15537	15607	71	TGT			+4	H
tRNAPro	15620	15688	69	TGG			+12	L
D-loop	15689	16709	1020				0	-

*Positive number corresponding to the nucleotides separating different gene, whereas negative numbers indicate overlapping nucleotides between adjacent genes; H and L denote heavy and light strands.

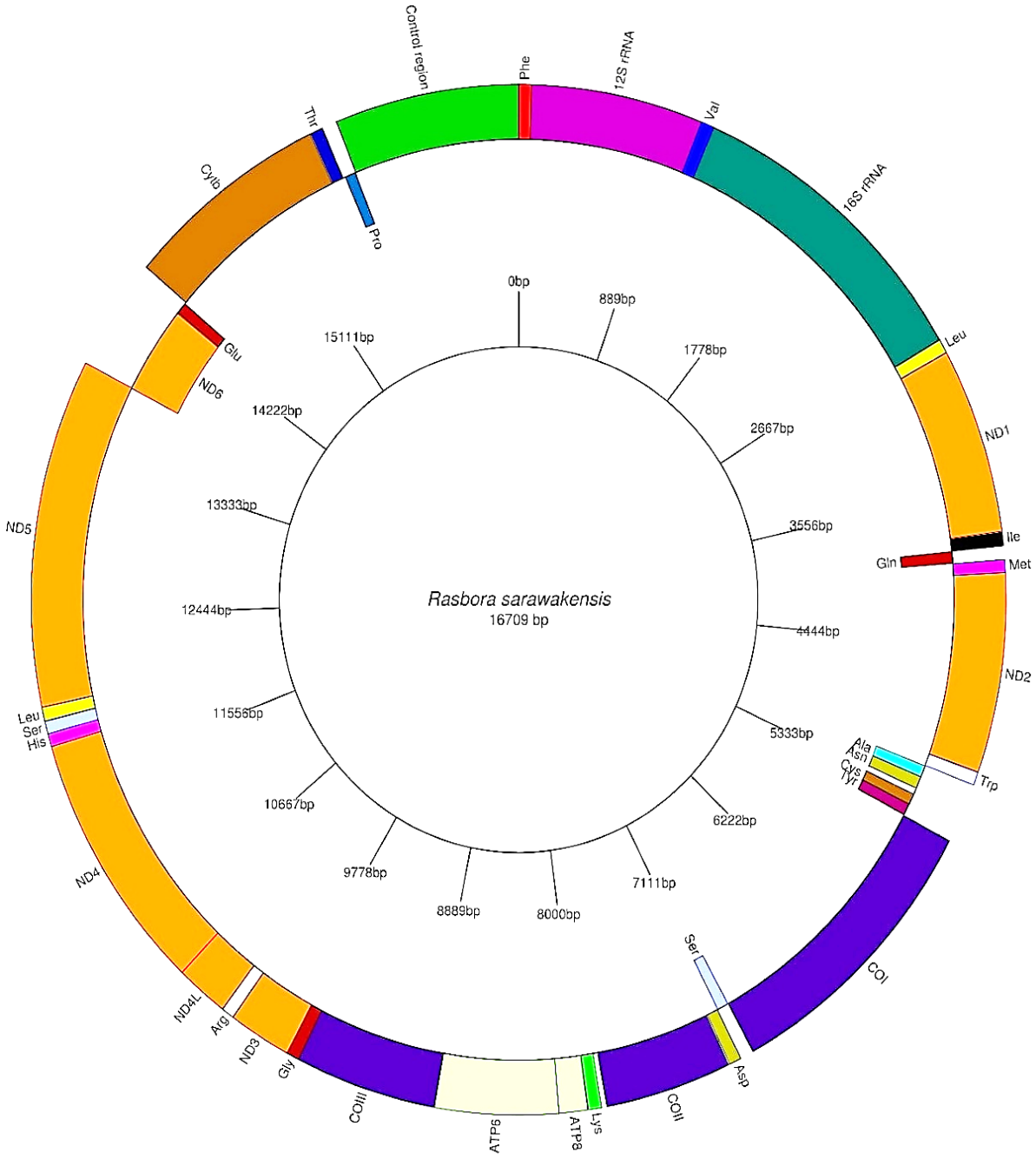


Figure 4.15: The complete mitochondrial genome organization of *R. sarawakensis* drawn using GenomeVx software. Transfer RNAs (tRNA) are labelled with their corresponding amino acids and are shown in red; COI, COII and COIII refer to subunits of cytochrome c oxidase; *Cyt* b refers to cytochrome b; 12S rRNA and 16S rRNA refer to ribosomal RNAs; ND1-ND6 refer to components of NADH dehydrogenase; ATPase 6 and ATPase 8 refers to classes of ATP synthase. Genes on the main circle are transcribed clockwise whereas those outside the main circle are transcribed anticlockwise.

Table 4.23: Characteristics of mitochondrial genome of *Rasbora* sp.

Gene	Position		Length (bp)	Anti-codon	Codon		Intergenic Nucleotide*	Strand
	From	To			Start	Stop		
tRNAPhe	1	69	69	GAA			0	H
12S rRNA	70	1022	953				0	H
tRNAVal	1023	1094	72	TAC			0	H
16S rRNA	1095	2773	1679				0	H
tRNALeu (UUR)	2774	2848	75	TAA			0	H
ND1	2850	3812	963		ATG	TAA	+1	H
tRNAIle	3817	3888	72	GAT			+4	H
tRNAGln	3887	3957	71	TTG			-2	L
tRNAMet	3959	4027	69	CAT			+1	H
ND2	4028	5072	1045		ATG	T--	0	H
tRNATrp	5073	5146	74	TCA			0	H
tRNAAla	5150	5217	68	TGC			+3	L
tRNAAsn	5219	5291	73	GTT			+1	L
tRNACys	5326	5392	67	GCA			+34	L
tRNATyr	5394	5463	70	GTA			+1	L
COI	5465	7015	1551		GTG	TAA	+1	H
tRNASer (UCN)	7016	7086	71	TGA			0	L
tRNAAsp	7088	7157	70	GTC			1	H
COII	7163	7853	691		ATG	T--	+5	H
tRNALys	7854	7928	75	TTT			0	H
ATP8	7931	8095	165		ATG	TAA	+2	H
ATP6	8089	8768	680		ATG	TA-	-7	H
COIII	8769	9553	784		ATG	TA-	0	H
tRNAGly	9554	9624	71	TCC			0	H
ND3	9625	9973	349		ATG	T--	0	H
tRNAArg	9974	10043	70	TCG			0	H
ND4L	10044	10340	297		ATG	TAA	0	H
ND4	10334	11653	1390		ATG	TAA	-7	H
tRNAHis	11724	11792	69	GTG			+70	H
tRNASer (AGY)	11793	11860	68	GCT			0	H
tRNALeu (CUN)	11863	11935	73	TAG			+2	H
ND5	11936	13765	1830		ATG	TAA	0	H
ND6	13762	14283	521		ATG	TAG	-4	L
tRNAGlu	14284	14352	69	TTC			0	L
Cyt b	14359	15499	1141		ATG	T--	+6	H
tRNAThr	15500	15570	71	TGT			0	H
tRNAPro	15579	15648	70	TGG			+8	L
D-loop	15649	16518	869				0	-

*Positive number corresponding to the nucleotides separating different gene, whereas negative numbers indicate overlapping nucleotides between adjacent genes; H and L denote heavy and light strands.

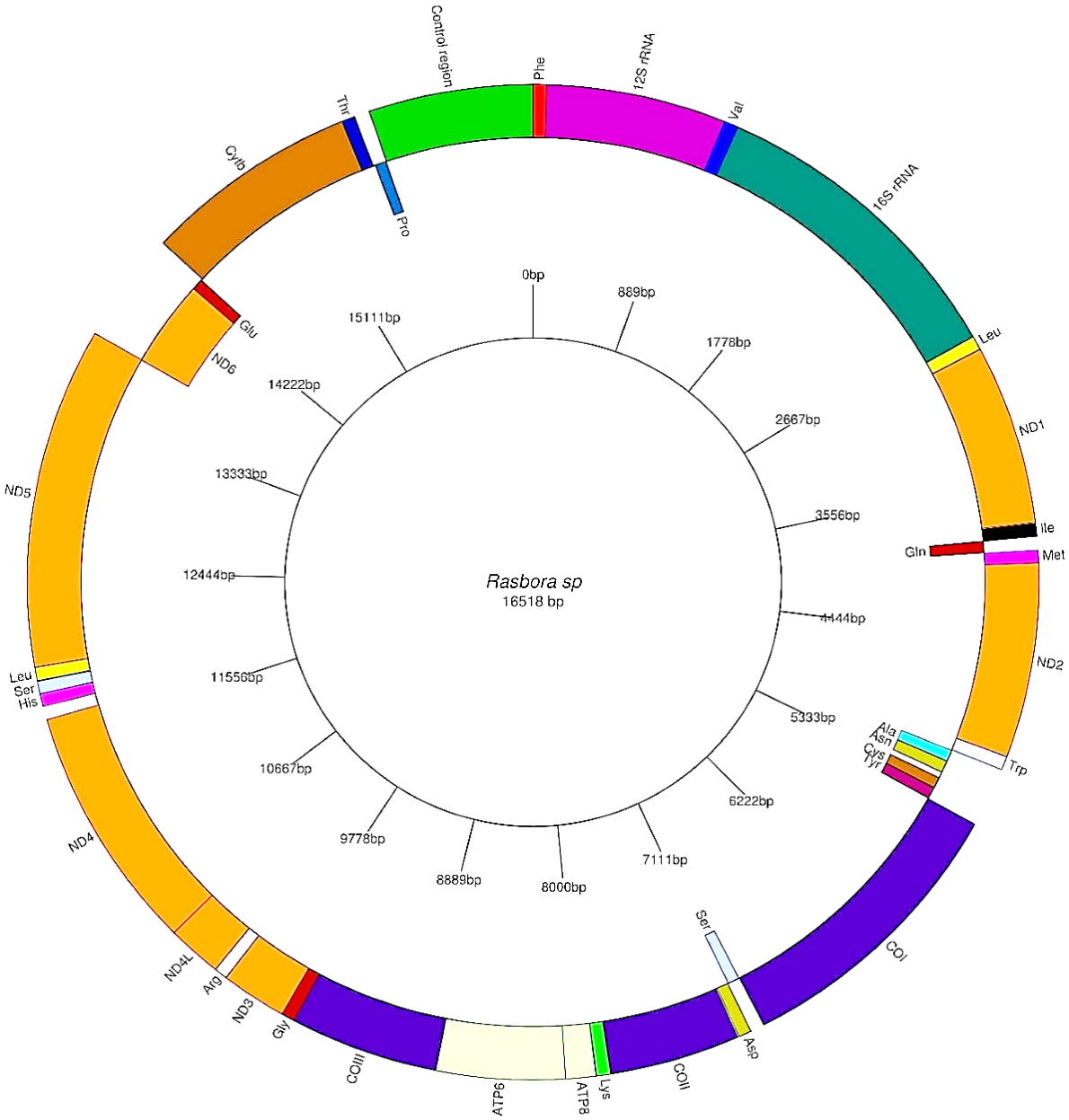


Figure 4.16: The complete mitochondrial genome organization of *Rasbora* sp. drawn using GenomeVx software. Transfer RNAs (tRNA) are labelled with their corresponding amino acids and are shown in red; COI, COII and COIII refer to subunits of cytochrome c oxidase; *Cyt* b refers to cytochrome b; 12S rRNA and 16S rRNA refer to ribosomal RNAs; ND1-ND6 refer to components of NADH dehydrogenase; ATPase 6 and ATPase 8 refers to classes of ATP synthase. Genes on the main circle are transcribed clockwise whereas those outside the main circle are transcribed anticlockwise.

4.3.3 Nucleotide Composition

The overall mean base composition of the whole genome of four studied *Rasbora* were estimated to be 33.9% A, 14.9% G, 25.5% T and 25.8% C (Table 4.24), with mean 59.4% of A+T content, indicating obvious anti-guanine bias in *Rasbora* mt-genome. The sequence analyses also show the highest A+T content was observed at the control region with mean 67.5% (range between 65.0 – 69.0%) in all *Rasbora*.

The strand bias was measured by calculating nucleotide skewness (GC-skew and AT-skew). The overall mean of GC- and AT-skews of the whole genome was 0.14 and -0.27, respectively. The negative and positive skewness (Table 4.24) indicating a strand compositional bias characterized. The positive value of GC-skew indicates the strong excess of C (cytosine) over G (guanine) nucleotides, while the negative value of AT-skew indicates the strong excess of A (adenine) over T (thymine) nucleotides.

The GC-skew and AT-skew values of PCGs, rRNAs and tRNAs shown in Table 4.24 were negative and positive respectively. The skew values show that the amount of C(s) and A(s) was more prevalent at the coding regions of *Rasbora*. The same observation for control region of *R. argyrotaenia, R. sarawakensis* and *Rasbora* sp. except for *R. myersi*. The value of GC-skew and AT-skew of control region for *R. myersi* were both negative, which mean that (C)s and (T)s were more prevalent.

4.3.4 Protein-Coding Genes

All the 13 PCGs found in other vertebrates were also present in *Rasbora* mt-genome. The 13 PCGs comprised of three subunits of the cytochrome c oxidase (COI, COII and COIII), seven subunits of the NADH ubiquinone oxidoreductase complex (ND1, ND2, ND3, ND4L, ND4, ND5 and ND6), one subunit of ubiquinol cytochrome b oxidoreductase complex (*Cyt* b) and two subunits of ATP synthases (ATP6 and ATP8).

Table 4.24: Summary of nucleotides compositions and skewness of mt-genome.

Species	Size (bp)	A%	G%	T%	C%	A+T %	AT skewness	GC skewness
Whole genome								
R. argyrotaenia	16574	33.5	15.1	25.1	26.3	58.6	0.143	-0.271
R. myersi	16581	33.5	14.9	25.7	26.0	59.2	0.132	-0.271
R. sarawakensis	16709	34.7	14.7	26.1	24.6	60.8	0.141	-0.252
Rasbora sp.	16518	33.7	14.9	25.1	26.3	58.8	0.146	-0.277
Mean		33.9	14.9	25.5	25.8	59.4	0.141	-0.268
PCGs								
R. argyrotaenia	11395	33.0	13.7	26.0	27.3	59.0	0.119	-0.332
R. myersi	11395	33.2	13.4	26.4	27.0	59.6	0.114	-0.337
R. sarawakensis	11404	34.1	13.4	27.3	25.3	61.4	0.111	-0.307
Rasbora sp.	11319	33.2	13.6	26.0	27.2	59.2	0.122	-0.333
Mean		33.4	13.5	26.4	26.7	59.8	0.116	-0.327
rRNAs								
R. argyrotaenia	2646	35.9	19.7	19.7	24.8	55.6	0.291	-0.115
R. myersi	2639	35.4	19.9	19.6	25.0	55.0	0.287	-0.114
R. sarawakensis	2631	38.0	19.2	19.1	23.6	57.1	0.331	-0.103
Rasbora sp.	2632	36.5	19.3	19.5	24.7	56.0	0.304	-0.123
Mean		36.5	19.5	19.5	24.5	55.9	0.303	-0.113
tRNAs								
R. argyrotaenia	1553	32.8	18.3	24.3	24.6	57.1	0.149	-0.147
R. myersi	1552	32.7	18.0	25.1	24.2	57.8	0.131	-0.147
R. sarawakensis	1546	33.6	18.0	23.9	24.5	57.5	0.169	-0.153
Rasbora sp.	1555	32.9	18.0	24.1	25.1	57.0	0.154	-0.165
Mean		33.0	18.1	24.4	24.6	57.4	0.151	-0.153
Control Region								
R. argyrotaenia	909	34.3	12.4	32.8	20.5	67.1	0.022	-0.246
R. myersi	924	33.2	12.8	35.8	18.2	69.0	-0.038	-0.174
R. sarawakensis	1019	34.6	12.6	34.3	18.4	68.9	0.004	-0.187
Rasbora sp.	869	33.3	13.2	32.5	21.1	65.0	0.012	-0.230
Mean		33.9	12.8	33.9	19.6	67.5	0.000	-0.209

The 13 PCGs of *Rasbora* mtDNA size ranged from 165 bp (ATP8) to 1830 bp (ND5). In total, the 13 PCGs comprises of 11,319 bp in *Rasbora sp.*, 11,404 bp in *R. sarawakensis* and the same length in *R. argyrotaenia* and *R. myersi* with 11,395 bp. And this PCGs accounting about 68% (varying 68.8% in *R. argyrotaenia*, 68.7% in *R. myersi*, 68.3% in *R. sarawakensis* and 68.3% in *Rasbora sp.*) of the entire mt-genome.

As shown in Table 4.25, all the 12 genes encoded on the H-strand, and a gene (ND6) encoded on the L-strand exhibited a marked similarity in nucleotide composition with an anti-guanine bias. Among PCGs, A+T content of *Rasbora* between range 53.9% (ND4L *R. argyrotaenia*) and 66.7% (ATP8 *R. sarawakensis*) (Table 4.25).

The AT- and GC-skew values of PCGs of all *Rasbora* are as shown in Table 4.25. In all studied *Rasbora*, the GC-skew values for PCGs was negative. The negative value in GC-skew means C are more than G nucleotide. While for AT-skew, most of the PCGs were positive value, only a few genes were negative (COI and ND3 in *R. argyrotaenia*; COI and ATP8 in *R. myersi*; COI and ND4L in *R. sarawakensis*; only COI in *Rasbora* sp.) values which means that T was slightly more than A nucleotide.

Start and stop codon was determined based on alignments with corresponding PCGs of other *Rasbora*. For *R. myersi*, *R. sarawakensis* and *Rasbora sp.*, (Table 4.20-4.23) most of the PCGs starts with ATG (methionine) codon except COI gene which initiates with the codon of GTG. Interestingly, for *R. argyrotaenia* there were 2 PCGs utilized GTG as a start codon (COI and ND6) (Table 4.20).

Two types of complete stop codons were detected in open reading frames of *Rasbora*. The 2 type of stop codon were TAA and TAG. Both stop codon were utilized by PCGs of 3 *Rasbora* (*R. argyrotaenia*, *R. myersi* and *Rasbora* sp.), but, *R. sarawakensis* only utilized TAA as a stop codon (Table 4.22). Additionally, not all the PCGs terminate with complete stop codon. Some of the PCGs stop with incomplete stop codon such T-- or TA- (Table 4.20–4.23) was observed.

4.3.5 Ribosomal and Transfer RNAs

There were two genes encoding for ribosomal RNAs (12S rRNA and 16S rRNA) that were identified in *Rasbora* mt-genome. For small subunit (12S rRNA) the size varied from 953 bp (*R. myersi* and *Rasbora* sp.), 954 bp (*R. sarawakensis*) and 957 bp (*R.*

argyrotaenia) in length, whereas the size for large subunit (16S rRNA) from 1677 bp (*R. sarawakensis*), 1679 bp (*Rasbora* sp.), 1686 bp (*R. myersi*) and 1689 bp (*R. argyrotaenia*) in length. These two ribosomal RNA genes were located on the H-strand between Phe and Leu (UUR), and these two rRNA genes was separated by the Val. The rRNA genes exhibited less A+T content with 55.9% (Table 4.24) compared to PCGs, tRNA genes and control region.

The tRNAs structure of *Rasbora* were constructed using tRNAscan-SE. The secondary structure of tRNAs of *Rasbora* were shown in Figure 4.21–4.24. Majority of postulated tRNA secondary structures had common features with 7 bp in the amino acid arm, 3-4 bp in the T arm (TΨC stem), 4-5 bp in the anticodon arm and 3-4 bp in D arm (DHU stem). Most of the *Rasbora* tRNA secondary structure was like cloverleaf, except Ser (AGY). The secondary structure of Ser (AGY) was different from other tRNA. The secondary structure of Ser (AGY) was missing the D arm (DHU stem).

Table 4.25: Summary of nucleotide composition for all protein-coding genes.

R. argyrotaenia

	Size (bp)	A%	G%	T%	C%	A+T%	AT skew	GC skew
Protein genes								
ND1	975	33.5	13.7	25.3	27.4	58.8	0.139	-0.333
ND2	1045	36.8	10.9	21.1	31.2	57.9	0.271	-0.482
COI	1551	28.3	17.4	29.6	24.7	57.9	-0.022	-0.173
COII	691	31.5	17.2	27.5	23.7	59.0	0.068	-0.159
ATP8	165	32.1	10.9	32.1	24.8	64.2	0.000	-0.389
ATP6	680	31.9	11.3	28.4	28.4	60.3	0.058	-0.431
COIII	785	29.8	16.4	26.1	27.6	55.9	0.066	-0.255
ND3	349	29.8	13.8	30.1	26.4	59.9	-0.005	-0.313
ND4L	295	29.0	14.1	24.9	32.0	53.9	0.076	-0.388
ND4	1382	34.1	13.4	25.3	27.2	59.4	0.148	-0.34
ND5	1830	36.1	11.7	26.2	26.0	62.3	0.159	-0.379
ND6	522	43.3	10.9	14.4	31.4	57.7	0.501	-0.485
Cyt b	1141	30.8	14.1	27.4	27.7	58.2	0.058	-0.325
Protein coding								
1st codon		34.9	13.3	24.9	26.9	59.8		
2nd codon		30.2	17.4	26.6	25.8	56.8		
3rd codon		34.0	10.6	26.4	29.1	60.4		

R. myersi

Protein genes								
ND1	975	34.1	13.1	26.9	25.9	61.0	0.118	-0.328
ND2	1045	36.9	10.7	22.1	30.2	59.0	0.251	-0.477
COI	1551	28.7	17.0	29.5	24.8	58.2	-0.014	-0.187
COII	691	32.3	16.4	27.8	23.6	60.1	0.075	-0.18
ATP8	165	31.5	10.9	33.3	24.2	64.8	-0.028	-0.379
ATP6	680	32.2	11.2	29.1	27.5	61.3	0.051	-0.421
COIII	785	30.2	16.3	25.6	27.9	55.8	0.082	-0.262
ND3	349	30.4	13.2	29.5	26.9	59.9	0.015	-0.342
ND4L	297	28.3	14.1	25.9	31.6	54.2	0.044	-0.383
ND4	1382	34.2	12.7	26.3	26.8	60.5	0.131	-0.357
ND5	1830	35.5	12	25.8	26.7	61.3	0.158	-0.38
ND6	522	44.3	10.3	15.7	29.7	60.0	0.477	-0.485
Cyt b	1141	30.9	13.6	27.9	27.7	58.8	0.051	-0.341
Protein coding								
1st codon		35.2	12.9	25.3	26.6	60.5		
2nd codon		30.0	17.1	27.0	25.8	57.0		
3rd codon		34.3	10.3	26.9	28.6	61.2		

Table 4.25 continued

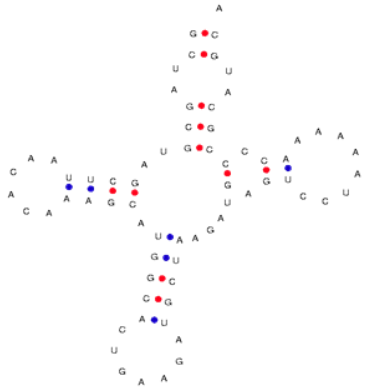
R. sarawakensis

	Size (bp)	A%	G%	T%	C%	A+T%	AT skew	GC skew
Protein genes								
ND1	975	35.3	13.3	26.4	25	61.7	0.144	-0.305
ND2	1047	38.1	10.7	24.4	26.8	62.5	0.219	-0.429
COI	1551	28.6	17.1	31.1	23.2	59.7	-0.042	-0.151
COII	702	34.2	15	27.5	23.4	61.7	0.109	-0.219
ATP8	165	38.2	7.3	28.5	26.1	66.7	0.145	-0.563
ATP6	683	33.4	11.1	29.7	25.8	63.1	0.059	-0.398
COIII	785	30.2	16.9	28.2	24.7	58.4	0.034	-0.188
ND3	349	30.1	14.3	29.5	26.1	59.6	0.01	-0.292
ND4L	297	29.0	13.5	29.6	27.9	58.6	-0.01	-0.348
ND4	1382	34.8	12.7	26.8	25.6	61.6	0.13	-0.337
ND5	1830	36.5	11.9	27.2	24.5	63.7	0.146	-0.346
ND6	519	45.5	10.4	14.8	29.3	60.3	0.509	-0.476
Cyt b	1137	32.3	13.6	28.2	25.9	60.5	0.068	-0.311
Protein coding								
1st codon		31.8	17.0	25.2	26.0	57.0		
2nd codon		29.2	12.8	31.8	26.3	61.0		
3rd codon		41.4	10.3	24.8	23.5	66.2		

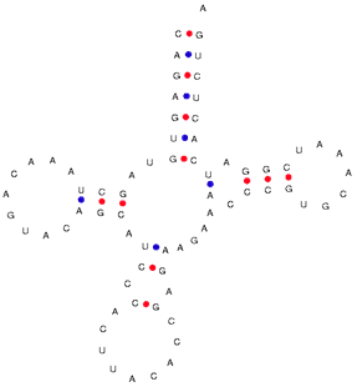
Rasbora sp.

Protein genes								
ND1	963	34.0	13.1	25.4	27.5	59.4	0.145	-0.355
ND2	1045	37.3	10.3	21.3	31.0	58.6	0.273	-0.501
COI	1551	27.7	17.7	29.7	25.0	57.4	-0.035	-0.171
COII	691	32.3	16.2	27.6	23.9	59.9	0.078	-0.192
ATP8	165	33.3	10.3	31.5	24.8	64.8	0.028	-0.413
ATP6	681	33.2	10.1	30.5	26.1	63.7	0.042	-0.442
COIII	785	28.3	18.3	26.5	26.9	54.8	0.033	-0.19
ND3	349	29.5	14.0	27.5	28.9	57.0	0.035	-0.347
ND4L	297	29.3	14.1	27.9	28.6	57.2	0.024	-0.34
ND4	1320	34.5	13.1	25.4	27.0	59.9	0.152	-0.347
ND5	1828	35.6	11.9	25.5	27.0	61.1	0.165	-0.388
ND6	522	44.6	10.7	14.2	30.5	58.8	0.517	-0.481
Cyt b	1141	32.1	13.5	27.2	27.3	59.3	0.083	-0.338
Protein coding								
1st codon		32.7	12.3	27.2	27.9	59.9		
2nd codon		32.6	15.0	24.6	27.8	57.2		
3rd codon		34.3	13.6	26.4	25.8	60.7		

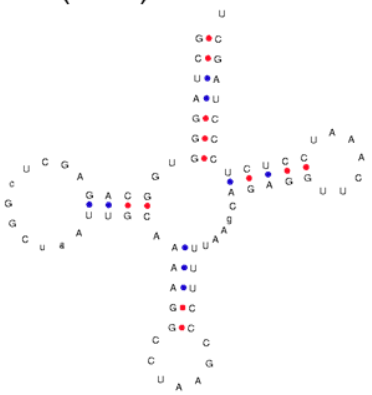
Phe



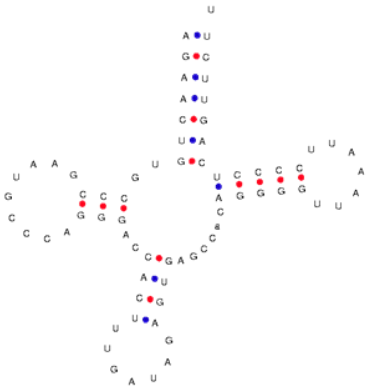
Val



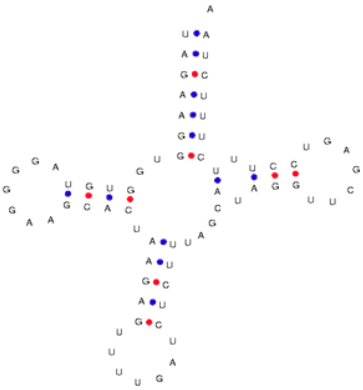
Leu (UUR)



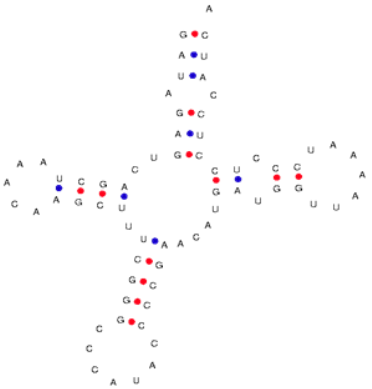
Ile



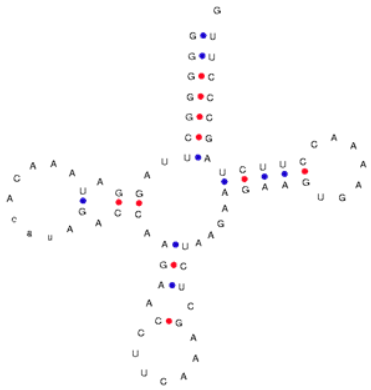
Gln



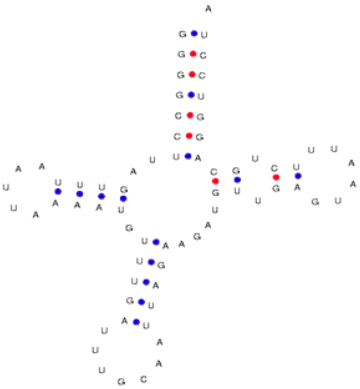
Met



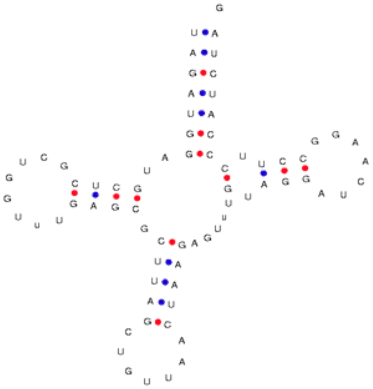
Trp



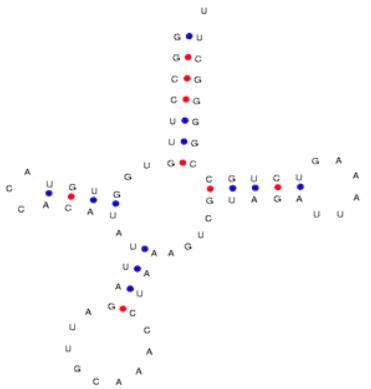
Ala



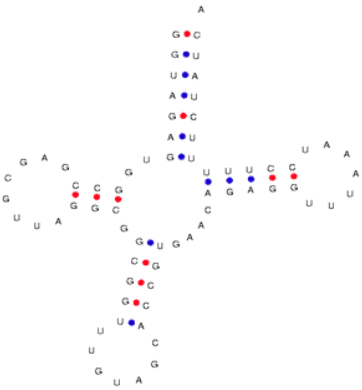
Asn



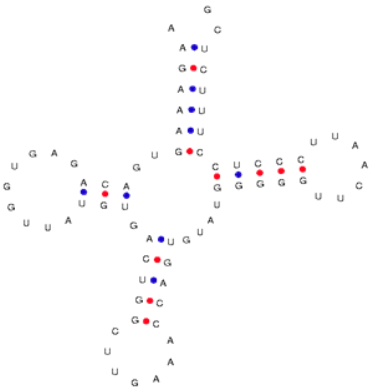
Cys



Tyr



Ser (UCN)



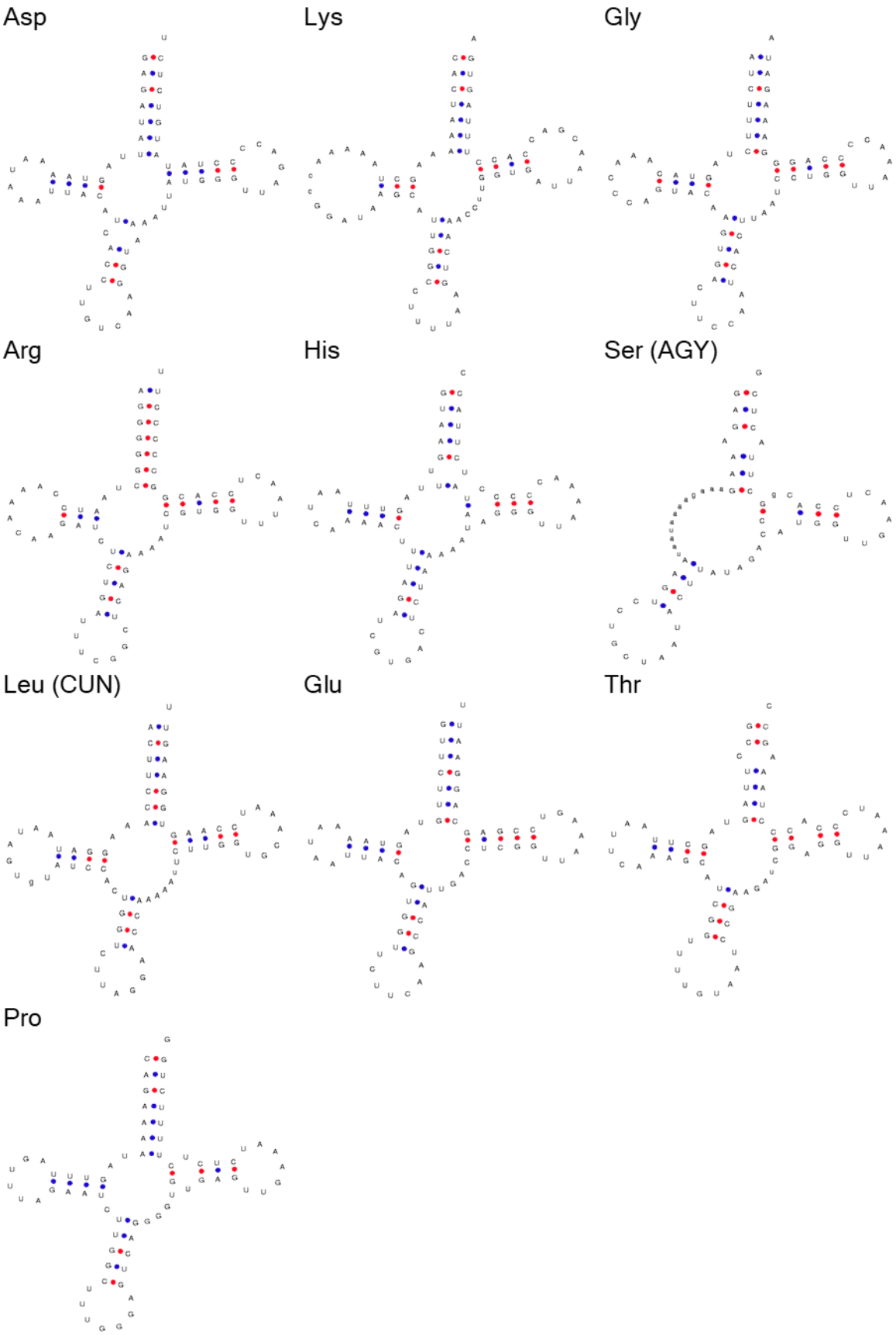
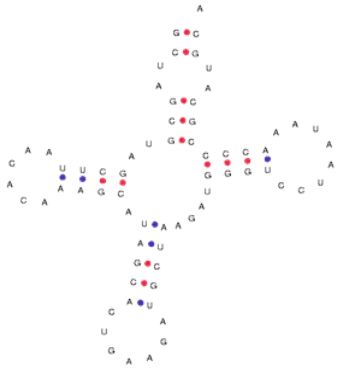
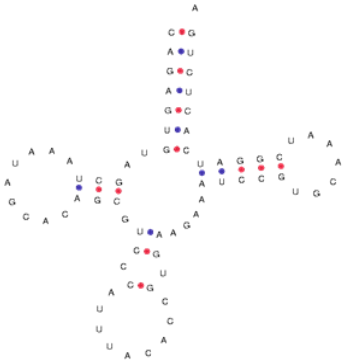


Figure 4.17: Putative secondary structures for the tRNA genes of the *R. argyrotaenia*. The tRNAs are labelled with the abbreviations of their corresponding amino acids.

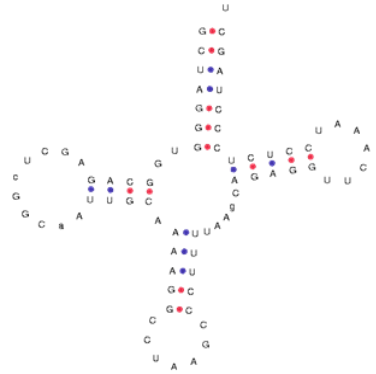
Phe



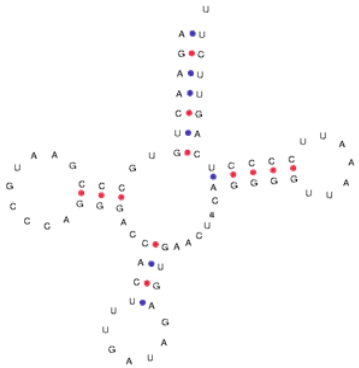
Val



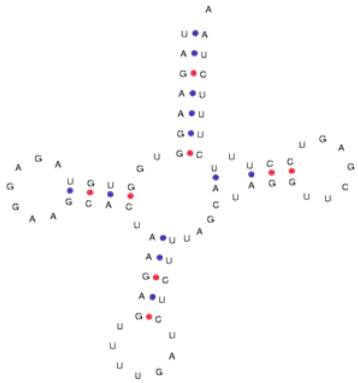
Leu (UUR)



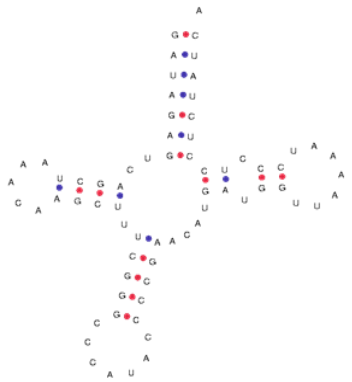
Ile



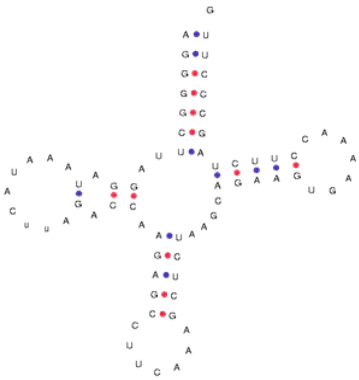
Gln



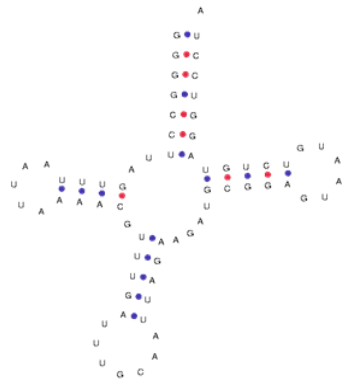
Met



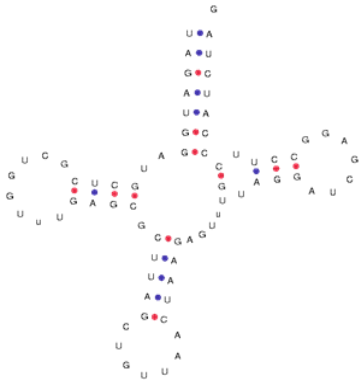
Trp



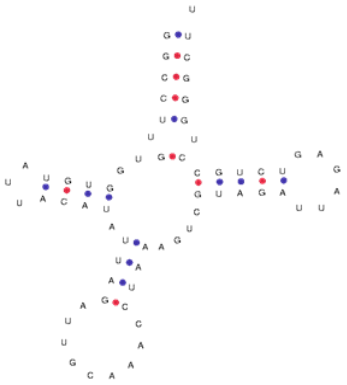
Ala



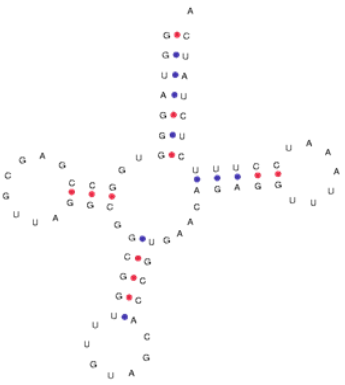
Asn



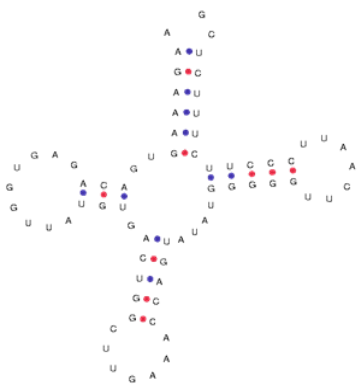
Cys



Tyr



Ser (UCN)



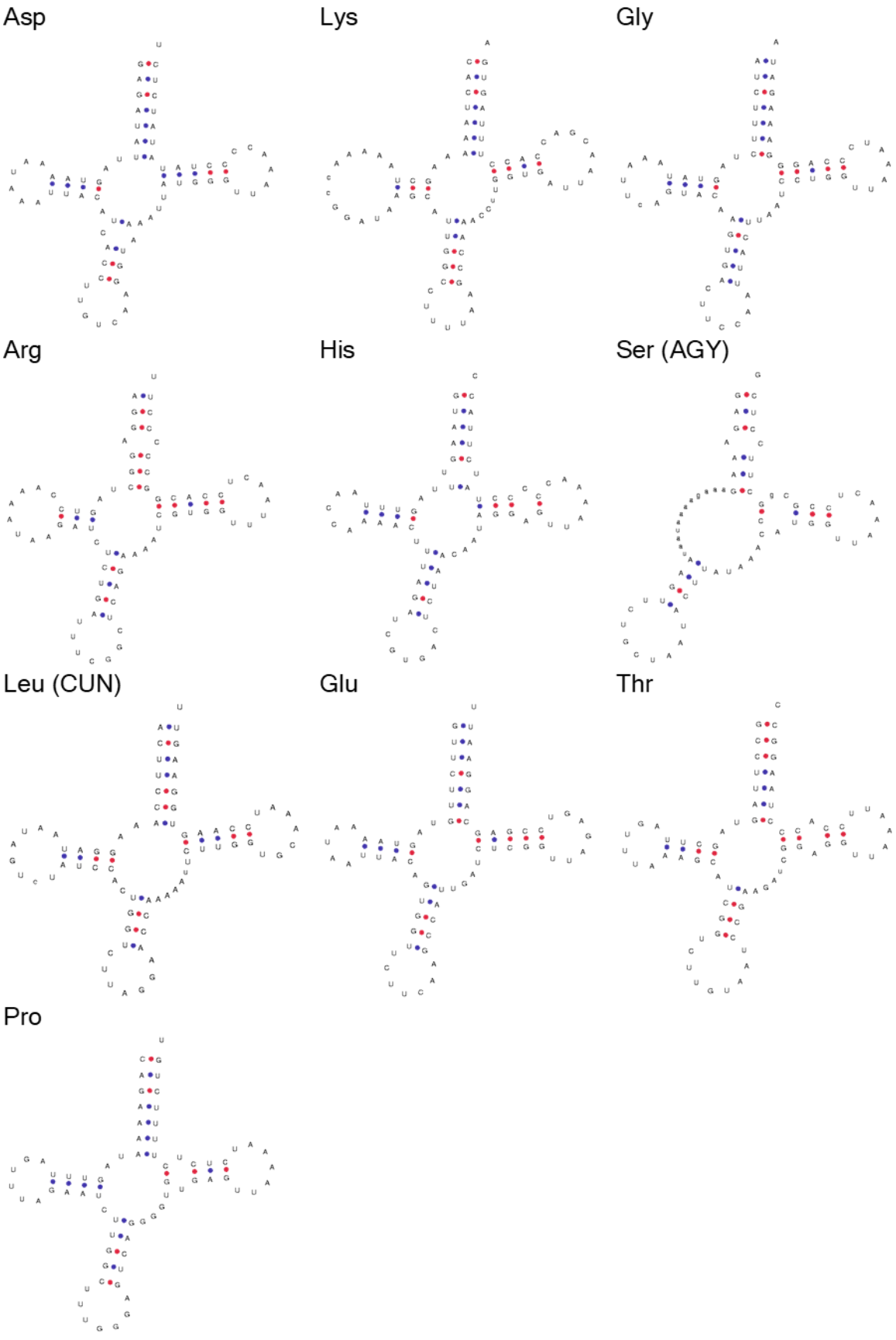
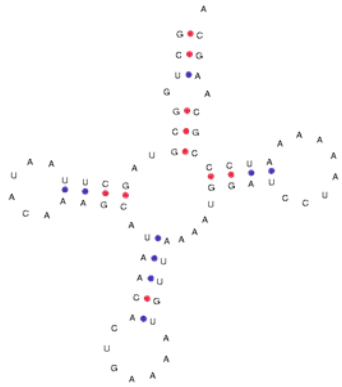
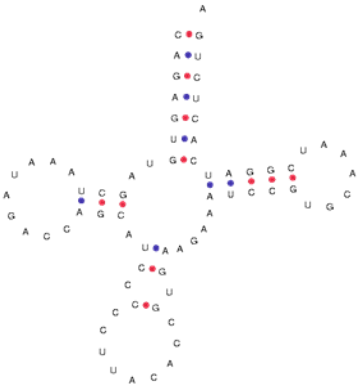


Figure 4.18: Putative secondary structures for the tRNA genes of the *R. myersi*. The tRNAs are labelled with the abbreviations of their corresponding amino acids.

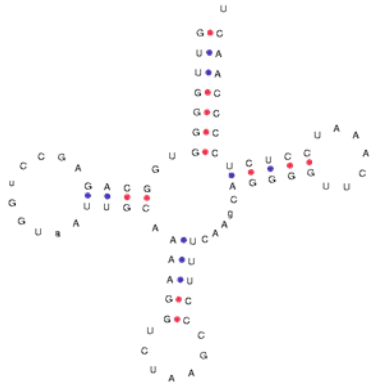
Phe



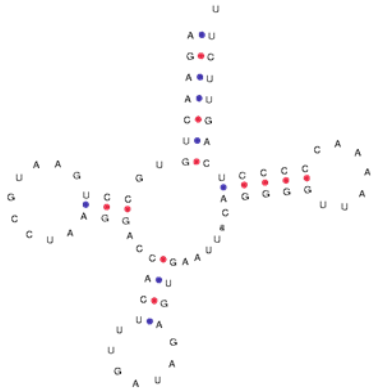
Val



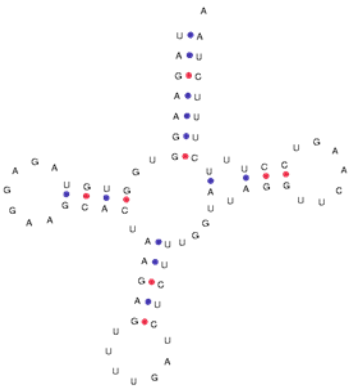
Leu (UUR)



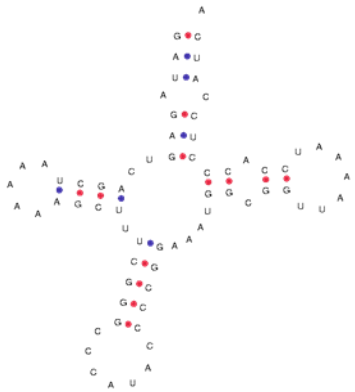
Ile



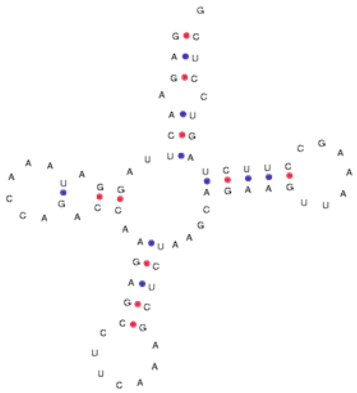
Gln



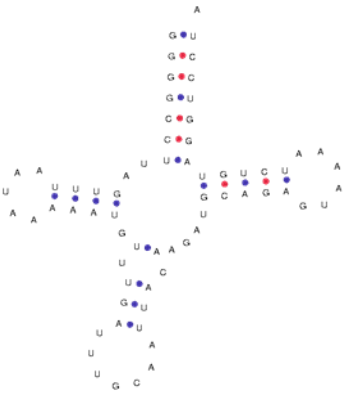
Met



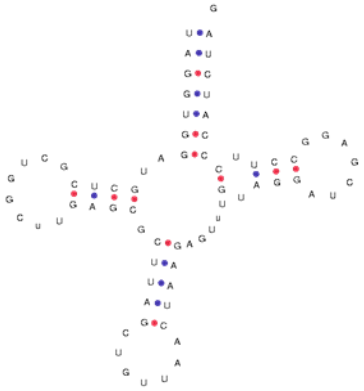
Trp



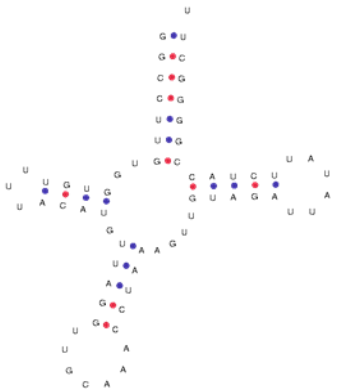
Ala



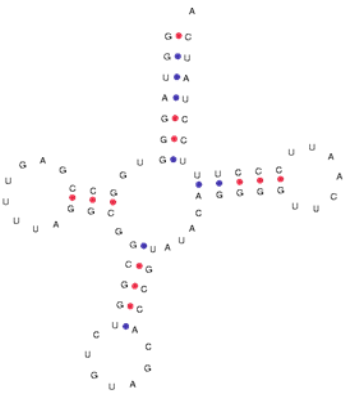
Asn



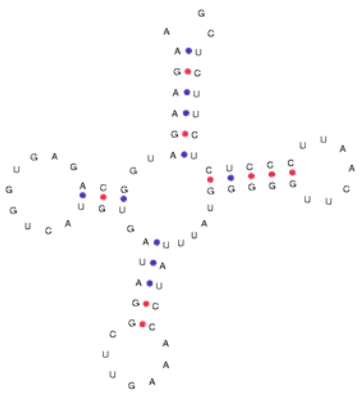
Cys



Tyr



Ser (UCN)



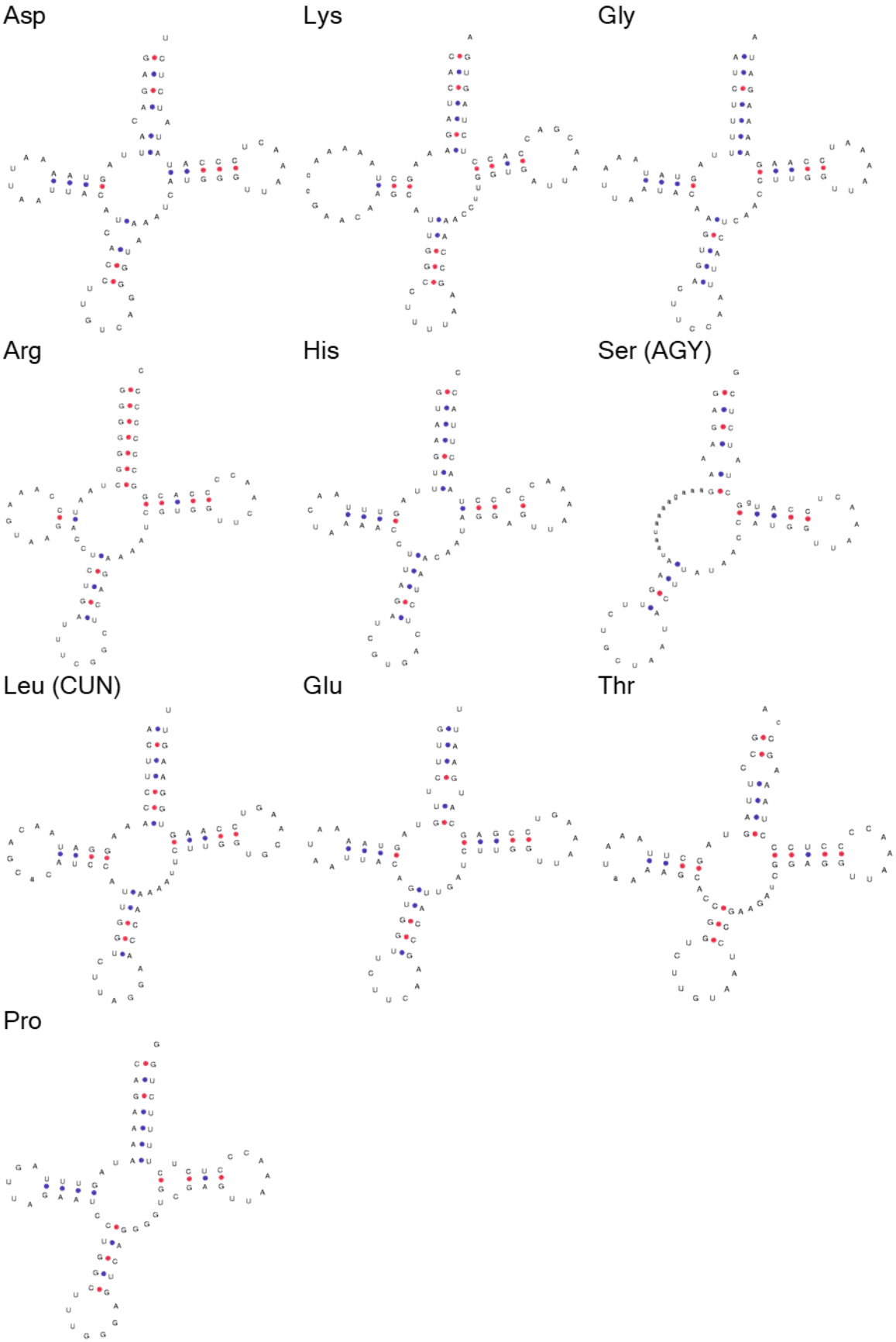
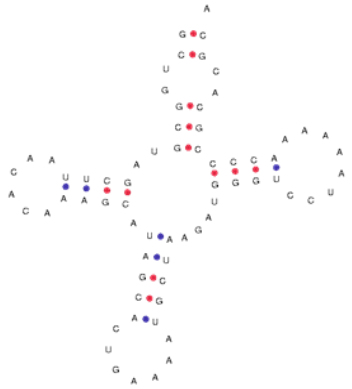
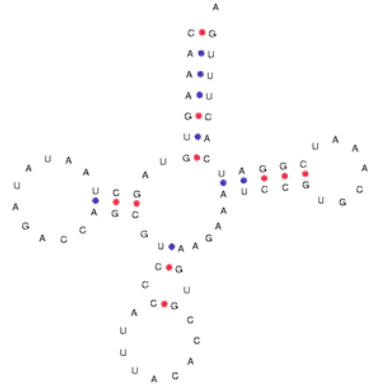


Figure 4.19: Putative secondary structures for the tRNA genes of the *R. sarawakensis*. The tRNAs are labelled with the abbreviations of their corresponding amino acids.

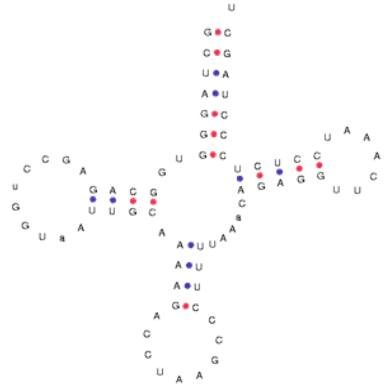
Phe



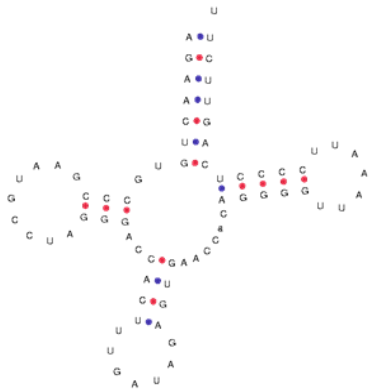
Val



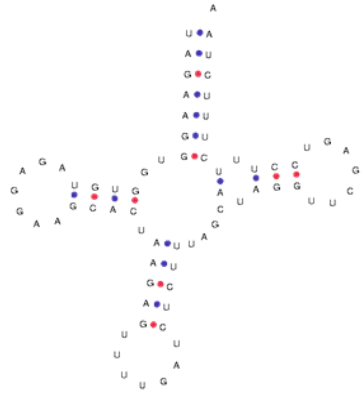
Leu (UUR)



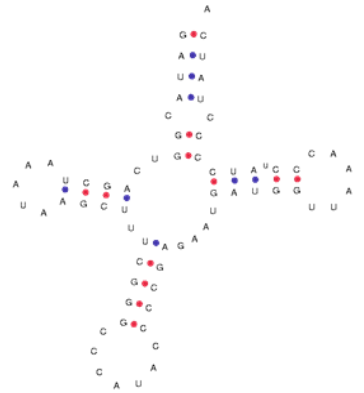
Ile



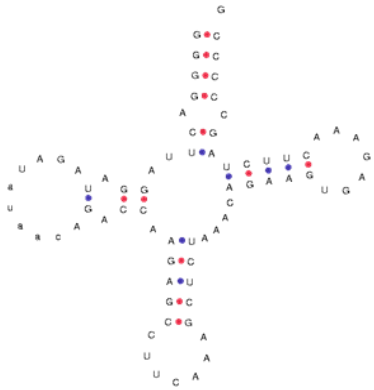
Gln



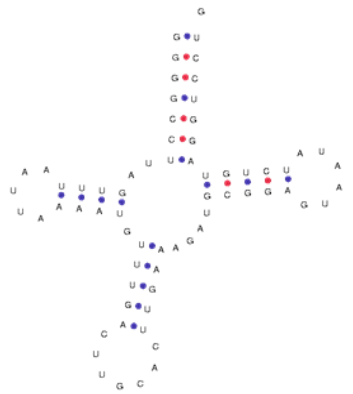
Met



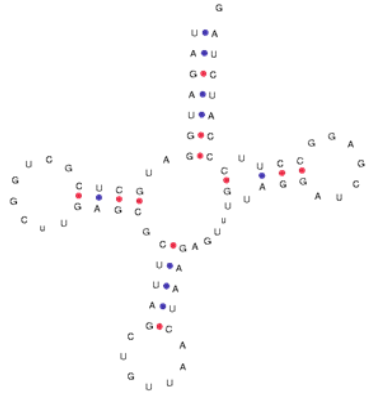
Trp



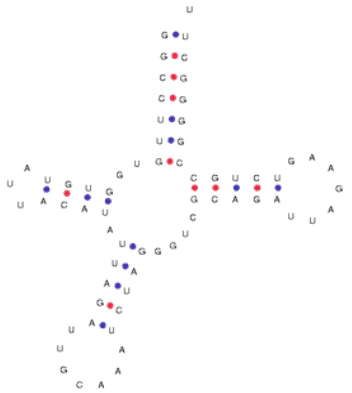
Ala



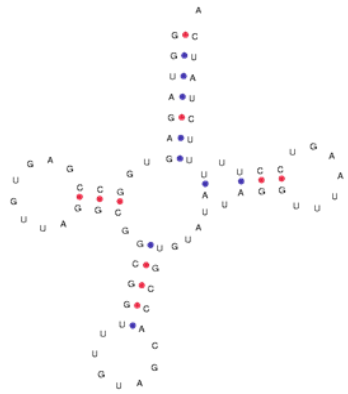
Asn



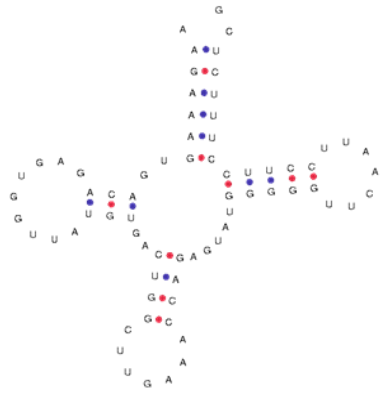
Cys



Tyr



Ser (UCN)



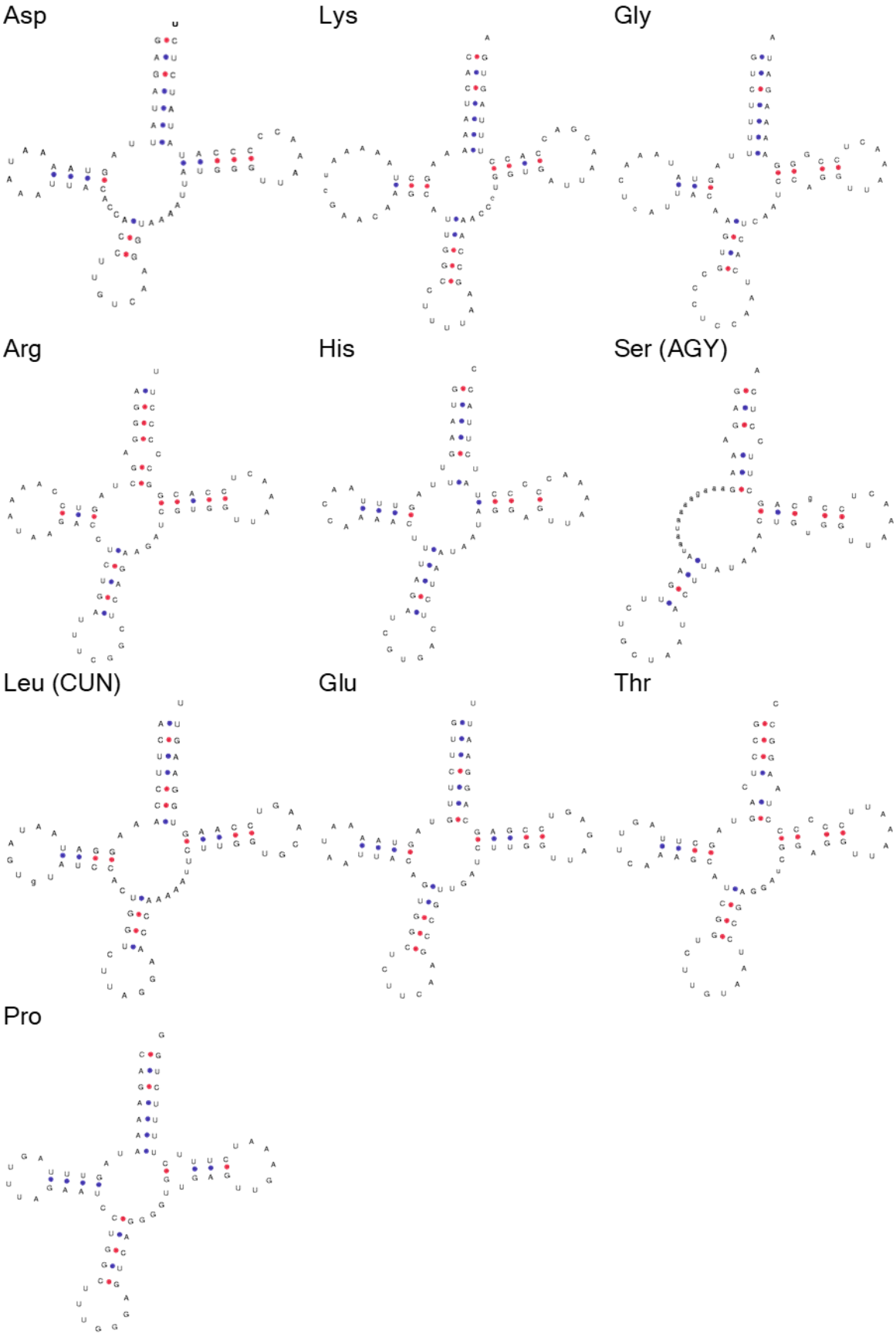


Figure 4.20: Putative secondary structures for the tRNA genes of the *Rasbora* sp.. The tRNAs are labelled with the abbreviations of their corresponding amino acids.

4.3.6 Non-coding Regions

There were two non-coding regions spotted in *Rasbora* mt-genome, (1) an origin of L-strand replication (O_L) and (2) control region (CR). The *Rasbora* O_L was located in a cluster of five tRNA genes (WANCY region) between Asn and Cys, and about 34 bp in length. This region was predicted to be capable of folding into a stable stem-loop secondary structure (Figure 4.21).

The predicted stem-loop structure of *Rasbora* O_L was built using Mfold web server. From the results, the stem-loop structure of *R. argyrotaenia*, *R. myersi* and *Rasbora* sp. was found similar with 12 bp on the stem and 13 bp on the loop (Figure 4.21**:** a, b and d). The stem-loop structure for *R. sarawakensis* is slightly different from the other *Rasbora*, which has longer stem with 16 bp with 1 bp mismatch, and formed a small loop with 5 bases (Figure 4.21c).

At the 5' side, 3 bp from the base of stem was the sequence of tRNACys gene, which mean the OL overlapped with tRNACys gene. And the sequence at 5' side being almost composed of pyrimidines (Thymine and Cytosine). The common motif sequence 5'-GCCGG-3' usually found at the base of stem within tRNACys gene. This motif sequence was detected on *R. argyrotaenia* (Figure 4.21a) and *Rasbora* sp. (Figure 4.21d), but absent on *R. myersi* and *R. sarawakensis*. The sequence noticed on *R. myersi* was (5'-ACCGG-3') (Figure 4.21b) and *R. sarawakensis* was (5'-GCCTG-3') (Figure 4.21c).

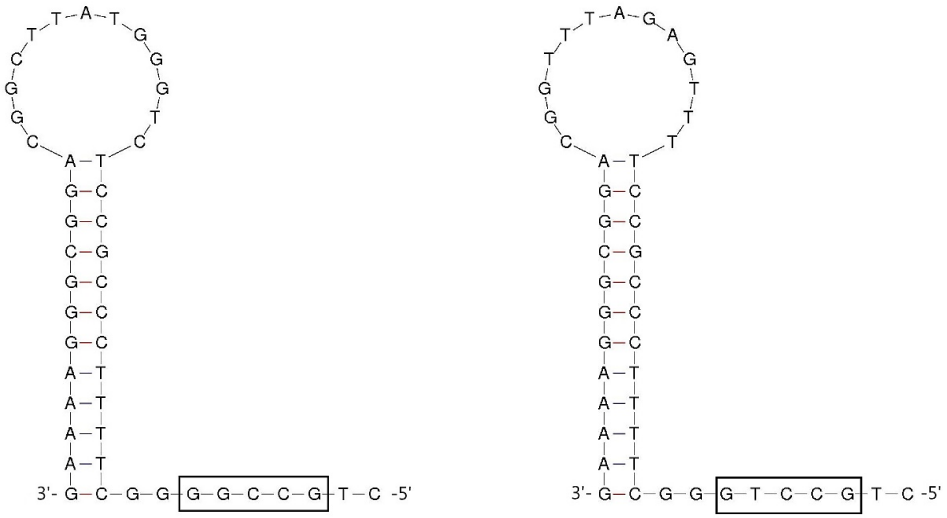
Another non-coding region was found in *Rasbora* mt-genome was the control region (CR). The CR of *Rasbora* mt-genome was identified between tRNAPro gene and tRNAPhe gene. The size of control regions was varied in different species, *R. argyrotaenia***:** 909 bp, *R. myersi***:** 924 bp, *R. sarawakensis***:** 1021 bp and *Rasbora* sp.**:** 870 bp. Control region is the highest A + T content on the *Rasbora* mt-genome. The A + T content greater than 64.0% (*R. argyrotaenia*, 67.1%; *R. myersi*, 69.0%; *R. sarawakensis*, 68.9%; and *Rasbora* sp., 65.0%).

Basically, CR comprised of three regulation domains**:** (1) extended termination-associated sequence domain (ETAS), (2) central conserved sequence block domain (CCD-CSB) and (3) conserved sequence block domain (CSB). Within the *Rasbora* control region, several conserved sequence blocks were detected. The ETAS domain is located at the upstream of 5' of CR. Within this domain contains several TAS motif sequence 5'-TACAT-3' and its palindromic cTAS motif sequence 5'-ATGTA-3'. In *R. argyrotaenia*, TAS and cTAS at the position 144 bp and 152 bp respectively. *Rasbora* sp., TAS (130 bp) and cTAS (138 bp). *R. sarawakensis*, TAS (138 and 284 bp) and cTAS (146 and 292 bp). *R. myersi* TAS (120, 141, 162 and 183 bp) and cTAS (128, 149, 170 and 191 bp).

In central conserved sequence blocks domain, three different types of CSBs commonly found in fish were CSB-F, CSB-E and CSB-D. However, only two central conserved sequence blocks (CSB-E and CSB-D) were identified in *Rasbora* central conserved sequence blocks domain. The consensus of CSB-E and CSB-D were shown in Table 4.26 and Figure 4.22. The CSB-F unidentifiable in the central conserved block domain of *Rasbora* CR. The order of central conserved sequence block in *Rasbora* was CSB-E follow by CSB-D.

In the conserved sequence block domain, 3 common conserved sequence blocks, CSB-1, CSB-2 and CSB-3 were identified. And these CSBs were found at the 3' side of the *Rasbora* CR. The CSB-1 was identified by sequence 5'-GACATA-3', the most conserved sequence for CBS-1. The CSB-2 and CSB-3 were easily identified compared to other conserved sequence block. The consensus sequence of CSB-2**:** AAACCCCCTTACCCCCT and CSB3**:** TGCCAAACCCCTAAACTCA. The putative promoter was also identified in *Rasbora* mt-CR. The promoter was position at the 3' side after the CSB-3. The consensus promoter sequence found on *Rasbora* mt-CR was 5'-GCATATA-3'.

a . b



c . d

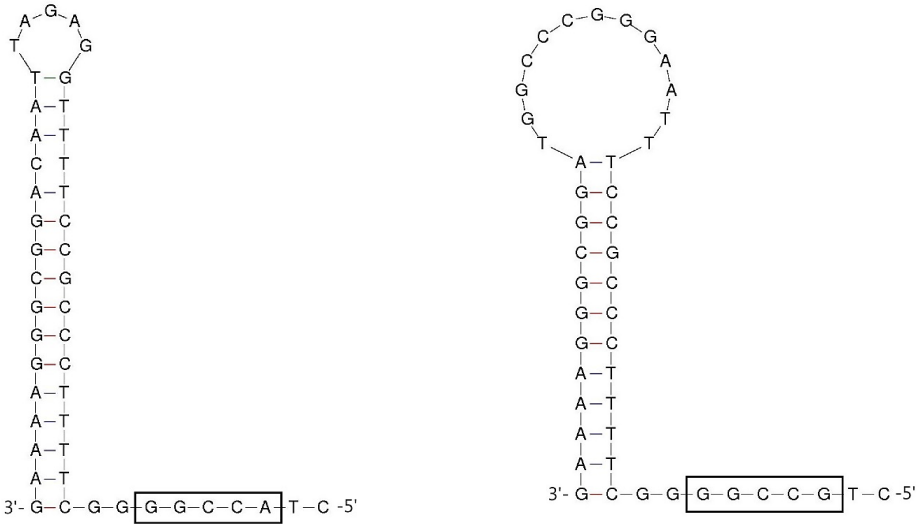
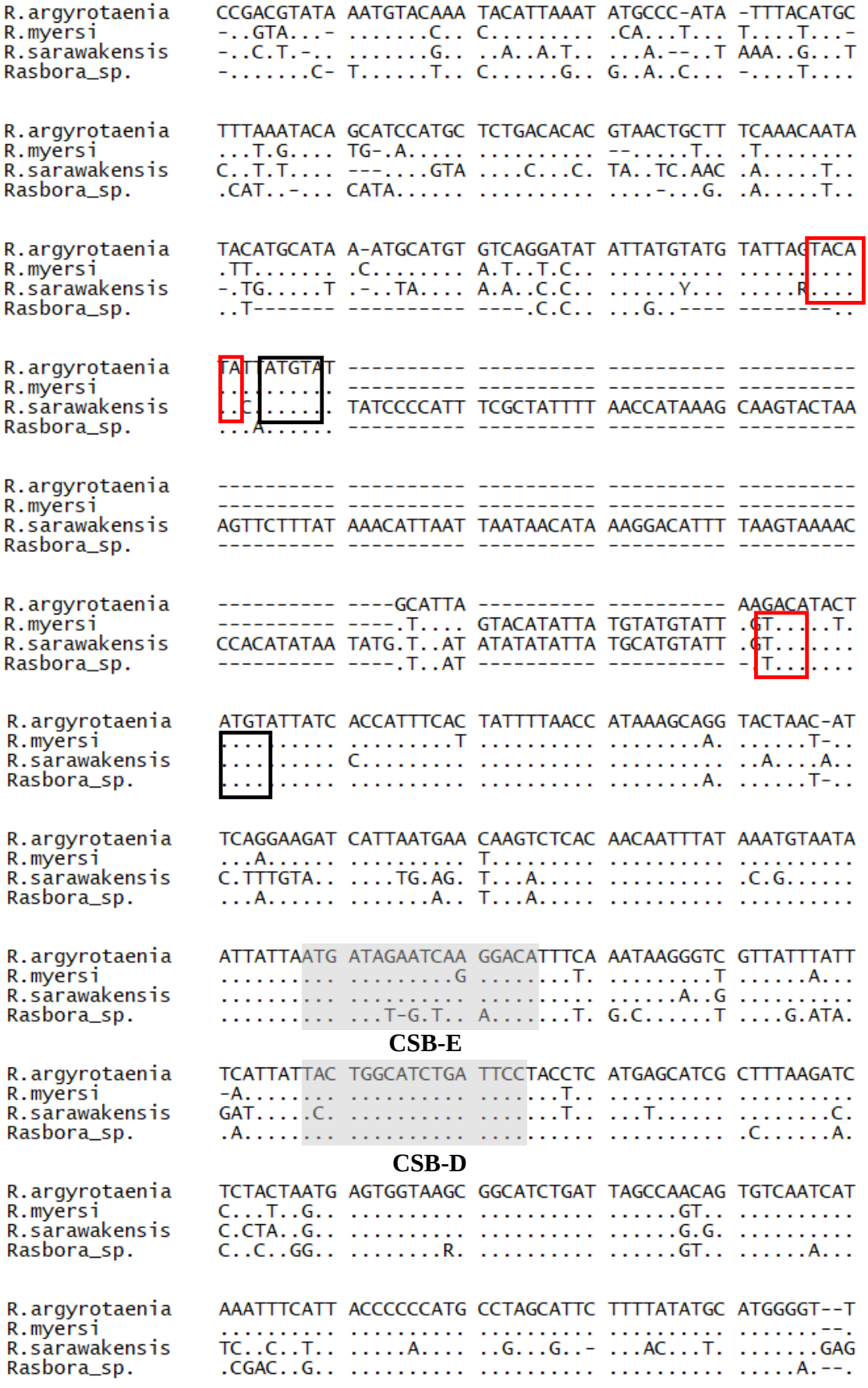


Figure 4.21: Potential secondary structure of the origin of L-strand replication (OL) mtDNA. a. O_L for *R. argyrotaenia*, b. O_L for *R. myersi*, c. O_L for *R. sarawakensis* and d. O_L for *Rasbora* sp.. Box indicates conserved motif.

Table 4.26: Conserved sequence of *Rasbora* mtDNA regulatory elements.

Element	Species	Sequence
CSB-E	*R. argyrotaenia*	ATGATAGAATCAAGGACA
	R. myersi	ATGATAGAATCAGGGACA
	R. sarawakensis	ATGATAGAATCAAGGACA
	Rasbora sp.	ATGATAT–GTTAAAGACA
CSB-D	*R. argyrotaenia*	TACTGGCATCTGATTCC
	R. myersi	TACTGGCATCTGATTCC
	R. sarawakensis	TCCTGGCATCTGATTCC
	Rasbora sp.	TACTGGCATCTGATTCC
CSB-1	*R. argyrotaenia*	ATAAATCCAATCTTGGAAGACATA
	R. myersi	ATAAATTCAATCTTGGAAGACATA
	R. sarawakensis	ATAAATGAATTCTTGGAAGACATA
	Rasbora sp.	ATAAATCCATTCTTGGTAGACATA
CSB-2	*R. argyrotaenia*	CAAACCCCCTTACCCCC
	R. myersi	CAAACCCCCTTACCCCC
	R. sarawakensis	CAAACCCCCTTACCCCC
	Rasbora sp.	CAAACCCCCTTACCCCC
CSB-3	*R. argyrotaenia*	TGCCAAACCCCTAAACT
	R. myersi	TGCCAAACCCCTAAACT
	R. sarawakensis	TGCCAAACCCCTAAAA-
	Rasbora sp.	TGCCAAACCCCTAAACT
Promoter	*R. argyrotaenia*	GCATATATATATA
	R. myersi	GCATATATATATA
	R. sarawakensis	GCATATATATATA
	Rasbora sp.	GCATATATATATA

Dash inferred gap.


R.argyrotaenia CCGACGTATA AATGTACAAA TACATTAAAT ATGCCC-ATA -TTTACATGC
R.myersi -..GTA...- ....C.. C......... .CA...T... T....T...-
R.sarawakensis -..C.T.-.. ....G.. ..A..A.T.. ...A.--..T AAA..G...T
Rasbora_sp. -.......C- T......T.. C......G.. G..A..C... -....T....
R.argyrotaenia TTTAAATACA GCATCCATGC TCTGACACAC GTAACTGCTT TCAAACAATA
R.myersi ...T.G.... TG-.A..... .... --.....T.. .T........
R.sarawakensis C..T.T.... ---....GTA ....C...C. TA..TC.AAC .A....T..
Rasbora_sp. .CAT..-... CATA...... ....-...G. .A....T..
R.argyrotaenia TACATGCATA A-ATGCATGT GTCAGGATAT ATTATGTATG TATTACTACA
R.myersi .TT....... .C........ A.T..T.C.. ....
R.sarawakensis -.TG.....T .-..TA.... A.A..C.C.. ......Y... .....R....
Rasbora_sp. ..T------- ---------- ----.C.C.. ...G..---- ---------..
R.argyrotaenia TATTATGTAT ---------- ---------- ---------- ----------
R.myersi .... ---------- ---------- ---------- ----------
R.sarawakensis ..C..... TATCCCCATT TCGCTATTTT AACCATAAAG CAAGTACTAA
Rasbora_sp. ...A...... ---------- ---------- ---------- ----------
R.argyrotaenia ---------- ---------- ---------- ---------- ----------
R.myersi ---------- ---------- ---------- ---------- ----------
R.sarawakensis AGTTCTTTAT AAACATTAAT TAATAACATA AAGGACATTT TAAGTAAAAC
Rasbora_sp. ---------- ---------- ---------- ---------- ----------
R.argyrotaenia ---------- ----GCATTA ---------- ---------- AAGACATACT
R.myersi ---------- ----.T.... GTACATATTA TGTATGTATT .GT.....T.
R.sarawakensis CCACATATAA TATG.T..AT ATATATATTA TGCATGTATT .GT.....
R.sarawakensis ---------- ----.T..AT ---------- ---------- -.T.....
R.argyrotaenia ATGTATTATC ACCATTTCAC TATTTTAACC ATAAAGCAGG TACTAAC-AT
R.myersi ....T ....A. ......T-..
R.sarawakensis .... C......... ....A....A..
Rasbora_sp. ....A. ......T-..
R.argyrotaenia TCAGGAAGAT CATTAATGAA CAAGTCTCAC AACAATTTAT AAATGTAATA
R.myersi ...A...... .... T.......... ....
R.sarawakensis C.TTTGTA.. ....TG.AG. T...A..... .... .C.G......
Rasbora_sp. ...A...... ....A.. T...A..... ....
R.argyrotaenia ATTATTAATG ATAGAATCAA GGACATTTCA AATAAGGGTC GTTATTTATT
R.myersi ....G ....T ....T ......A...
R.sarawakensis ....A..G ....
Rasbora_sp. ....T-G.T.. A.....T. G.C......T ....G.ATA.
CSB-E
R.argyrotaenia TCATTATTAC TGGCATCTGA TTCCTACCTC ATGAGCATCG CTTTAAGATC
R.myersi -A..... ....T.. ....
R.sarawakensis GAT.....C. ....T.. ...T...... ....C.
Rasbora_sp. .A..... .... .C......A.
CSB-D
R.argyrotaenia TCTACTAATG AGTGGTAAGC GGCATCTGAT TAGCCAACAG TGTCAATCAT
R.myersi C...T..G.. ....GT.. ....
R.sarawakensis C.CTA..G.. ....G.G. ....
Rasbora_sp. C..C..GG.. ....R. ....GT.. ......A...
R.argyrotaenia AAATTTCATT ACCCCCCATG CCTAGCATTC TTTTATATGC ATGGGGT--T
R.myersi ....--.
R.sarawakensis TC..C..T.. .....A.... ..G...G..- ...AC...T. ....GAG
Rasbora_sp. .CGAC..G.. ....A.--.

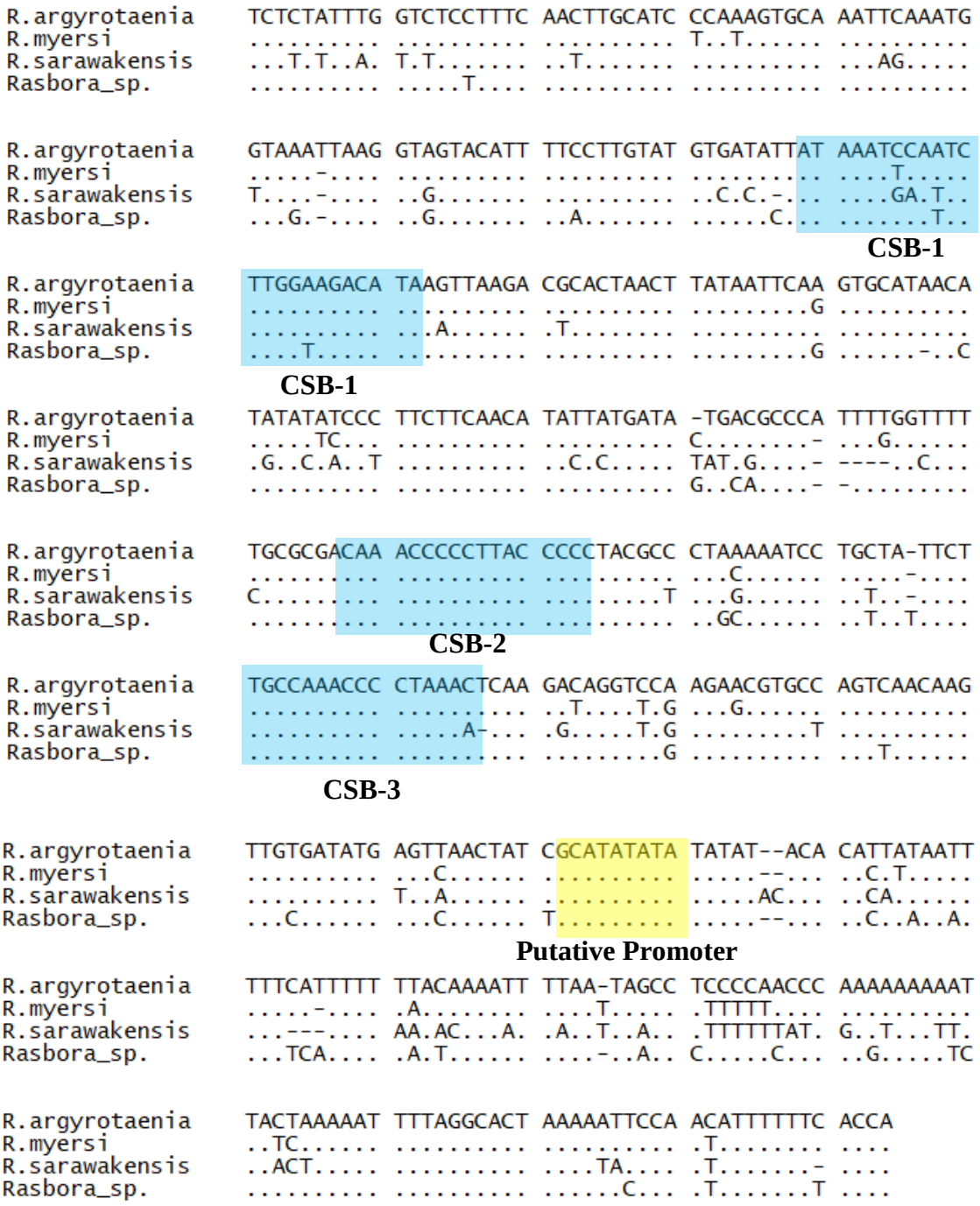


Figure 4.22: Aligned sequence of the mtDNA control region in *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp.. Dots indicate nucleotide identity. Dashes indicate indels. The TAS red frame, cTAS black frame, central conserved sequence block highlighted in grey, conserved sequence block highlighted in blue and promoter highlighted in yellow.

4.4 Phylogenetic Analyses

The phylogenetic analyses were based on the concatenated sequences of 13 mitochondrial PCGs. To determine the phylogenetic position of the *Rasbora* species, the mitogenomes of 18 species (Table 4.27) that are closely related to *Rasbora* and an outgroup *Clinocottus globiceps* were clustered together to construct ML tree. The ML tree based on nucleotide sequence was constructed using the Maximum likelihood method based on the GTR + G + I model in MEGA X (Nei and Kumar, 2000; Kumar et al., 2018), while the ML tree based on amino acid sequence was constructed using the Maximum likelihood method based on the mtREV24 + G + I + F model in MEGA X software (Adachi & Hasegawa, 1996; Kumar et al., 2018).

The results from phylogenetic analyses was shown in Figure 4.23 (ML tree based on nucleotide sequence) and Figure 4.24 (ML tree based on amino acid sequence). The trees inferred from 1000 replicates was taken to represent the phylogeny of the species analysed in this study. The bootstraps support values were presented as percentage and shown near the tree branches. The scale bar indicates the length measure in number of substitutions per site.

The ML analyses (Figure 4.23 and Figure 4.24) shows that *R. argyrotaenia* and *R. myersi* were positioned on the same clade as *R. borapentensis* (AB924546) and *R. argyrotaenia* (LC269105). *R. argyrotaenia*, *R. myersi* and *R. borapentensis* formed a monophyletic group and the molecular data shows very well bootstrap value (100%).

The other mitogenome produced from this study was *Rasbora sp.*. The phylogenetic analyses presented *Rasbora sp.* was equally close related to *R. steineri* (JX843769), *R. sumatrana* (AP011221), *R. lateristriata* (LC021505) and *R. aprotaenia* (LC021504), and they formed a monophyletic group. The molecular data based on nucleotide sequence of 13 mitochondrial PCGs for *Rasbora* sp. show a strong bootstrap value supported (100%) in ML

analysis. But the amino acid sequence shows a moderate bootstrap supported (89%) in ML analysis.

The phylogenetic position of *R. sarawakensis* was not consistent (Figure 4.23 and Figure 4.24). The result from ML analysis based on nucleotide sequence the *R. sarawakensis* formed a clade with a moderate bootstrap value supported 78%. Whereas in the ML analysis based on amino acid sequence, *R. sarawakensis* formed a monophyletic with *B. maculatus* with no bootstrap supported (<50%).

Table 4.27: List of species included in molecular phylogenetic analyses.

Species	Mt-genome size (bp)	GenBank accession no.
Rasbora argyrotaenia	16,574	-----
Rasbora myersi	16,581	-----
Rasbora sarawakensis	16,709	-----
Rasbora sp.	16,518	-----
Rasbora aprotaenia	16,541	LC021504
Rasbora argyrotaenia	16,740	LC269105
Rasbora borapentensis	16,510	AB924546
Rasbora cephalotaenia	15,647	AP011430
Rasbora daniconius	16,886	AP011285
Rasbora lateristriata	16,539	LC021505
Rasbora stenieri	16,530	JX843769
Rasbora sumatrana	16,539	AP011221
Rasbora trilineata	16,747	KM200714
Rasboroides vaterifloris	16,439	AP011432
Trigonostigma espei	17,162	AP011449
Trigonostigma heteromorpha	16,650	AP011421
Boraras. maculatus	15,640	AP011420
Horadandia atukorali	16,735	AP011400
Clinocottus globiceps	17,137	KF751382

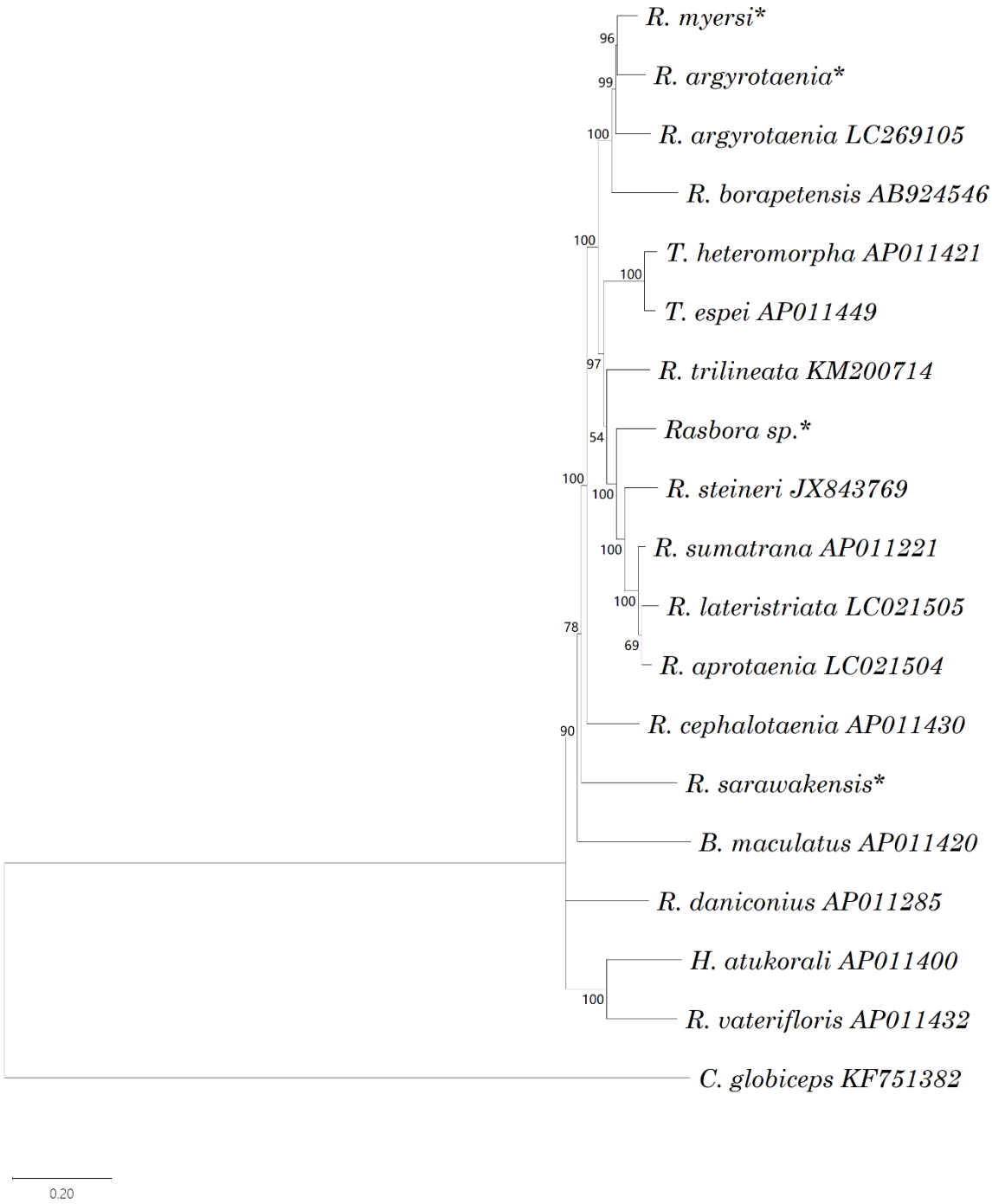


Figure 4.23: Maximum likelihood tree constructed using 9,967 nucleotide sequences of mitochondrial 13 protein coding genes. The asterisk indicates the sequence generated in this study.

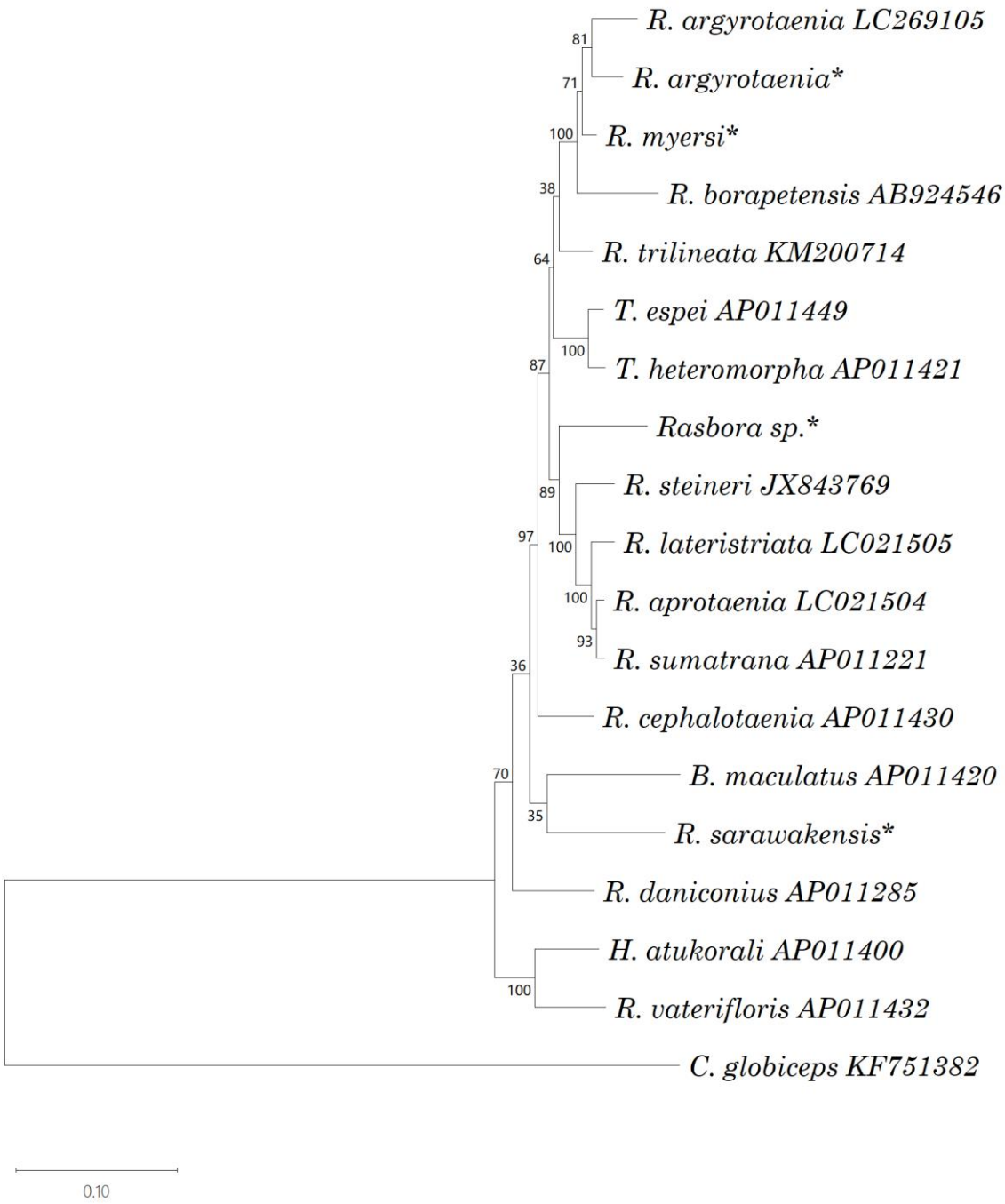


Figure 4.24: Maximum likelihood tree constructed using 3206 amino acid sequences of mitochondrial 13 protein coding genes. The asterisk indicates the sequence generated in this study.

CHAPTER 5

DISCUSSION

5.1 Morphological Analyses

In this study, the morphological analyses were conducted mainly to discriminate and identify *Rasbora* fishes. Morphological analyses are the traditional method in organism identification. This method has two techniques known as morphometric and meristic. Morphometric is the quantitative measurement where the specific body part was measured, while meristic is the quantitative counting features of a fish sample (Syaifullah et al., 2015; Muchlisin, 2013). The uneven number of samples for each Rasbora species were due to random sampling practices and subsequent removal of non *Rabora* species that were caught.

Many researchers were using statistical morphological analyses to determine the variation of morphological character of species, and also distinguish specimens into certain groups (Chandra and Prusty, 2011; Syaifullah et al., 2015). In this study, the results from discrimination function analyses were effectively used to discriminate the *Rasbora* fishes into four different group. The groups of clusters were proving that there were four different species of *Rasbora* (*R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp.). One-way ANOVA was used to reveal which morphological character was effective enough in distinguishing *Rasbora* species. Most of the morphological characters in this study can effectively distinguish the four *Rasbora* fishes, except for several morphological characters such as preanal length, body depth, snout length, anal fin rays, caudal fin rays and transverse scales.

The results from principal component analysis of morphometric, showed a low degree of variation (morphometric**:** 31.4% and meristic**:** 53.0%) among the fishes. And the PC score plot does not give a clear distinction of the *Rasbora* species (Osborne & Costello,

2004). That was because PCA was used to estimate morphometric variation among *Rasbora* species and not for species classification. Low morphological variation in PC result does not mean that there were no differences among *Rasbora* species. It simply means that the morphometric character contributed to the PC having less variance or not much variation to distinguish all fish groups (Lorenzo-Seva, 2013).

The other way to differentiate *Rasbora* species was assessed through discriminant function analysis. From the DFA result, the correct classification of *Rasbora* fishes was comparatively high (morphometric**:** 93% and meristic**:** 90%), greater and equal than 90% of the original group cases correctly classified. And the most significant morphological character for DF were interorbital width, head width, head length, eye diameter and middle caudal lobe length, and pectoral fin rays, pelvic fin rays, dorsal fin ray and scales of lateral line. All the contributing characters were important in *Rasbora* morphological study.

5.2 Mitogenome Analyses

The complete mtDNA sequence of *Rasbora* ranged from 16,709 bp to 16,518 bp. By comparing the present *Rasbora* mt-genome sequence the largest mtDNA sequence was *R. sarawakensis* with 16,709 bp followed by *R. myersi* with 16,581 bp, *R. argyrotaenia* with 16,574 bp and the smallest mt-genome sequence was *Rasbora* sp. with 16,518 bp length. The size of *Rasbora* species varied between 16,886 bp (*R. daniconius*) and 16,439 bp (*R. vaterifloris*).

The variation of mt-genomes size in these species was predominately due to the number and size of non-coding sequence. Small gap with size less than 20 bp between adjacent genes was commonly present in mt-genome. Sometimes a big gap with length greater than 60 bp also will be present to a vertain extent in the vertebrates mitogenome. The present *Rasbora* mt-genome, *Rasbora* sp. having an extra gap located between ND4 and

tRNA His with 70 bp length. There is no extra gap observed on *R. argyrotaenia*, *R. myersi* and *R. sarawakensis* mt-genome.

Basically, the complete *Rasbora* mtDNA is divided into 2 groups, coding region and non-coding region. The coding region contains 37 coding genes which including 13 protein coding genes, 22 tRNA genes and 2 rRNA genes. And there are 2 non-coding regions**:** (1) control region (CR) and (2) origin of light strand replication (OL). Although the non-coding regions were not synthesized any product, but these regions comprise of regulatory element that is important for DNA replication and transcription.

The PCGs present in *Rasbora* mt-genome was similar to other teleost fish and vertebrates mitogenome. All the PCGs in *Rasbora* was mainly responsible in energy production**:** ND1, ND2, ND3, ND4L, ND4, ND5 and ND6 were encoded for components of NADH dehydrogenase; COI, COII and COIII were encoded for subunit of cytochrome c oxidase; ATP8 and ATP6 were encoded classes of ATP synthase; and *Cyt* b was refer to as cytochrome b.

The *Rasbora* mt-genome genes arrangement was similar to typical vertebrates mt-genome. The two subunits mitochondrial ribosomal RNA, 12S rRNA and 16S rRNA were located between tRNAPhe and tRNALeu separated by tRNAVal, respectively (Prosdocimi et al., 2011). Most of PCGs were encoded on the H-strand together with 14 tRNA and 2 rRNA except ND6. ND6 was the only PCG encoded on the L-strand together with eight tRNA (Gln, Ala, Asn, Cys, Tyr, Ser, Glu and Pro) genes.

The common cluster of five tRNA, Trp-Ala-Asn-Cys-Tyr (WANCY region) found in vertebrates was also detected in the *Rasbora* mt-genome. However, the conserved gene arrangement of bony fishes mt-genome, ND6-Cytb-Glu-Pro-Thr-CR (Pereira, 2000) and ND6-Cytb-Glu-Thr-Pro-CR (Miya and Nishida, 1999) was not observed in *Rasbora*.

However, the gene arrangement of ND6-Glu-Cytb-Thr-Pro-CR was found in selected *Rasbora* mitochondrial genome.

The nucleotide composition of *Rasbora* show anti-guanine bias and strand asymmetry. In previous finding, mostly reported that fish mitogenome show anti-guanine bias (Wang et al., 2008) and strand asymmetry (Cheng et al., 2012). The highest A+T content of *Rasbora* mt-genome was detected in the CR, which consistent with the finding of previous report on another teleost (Shi et al., 2016). Briefly, the nucleotides compositions of *Rasbora* mt-genome was similar with other fish mt-genome reported.

As in other fish mt-genome, 2 common non-coding regions were found in *Rasbora* mt-genome. The 2 non-coding regions were control region (CR) and origin of L-strand replication (OL). The *Rasbora* CR was located between Phe and Pro, and the size of *Rasbora* CR ranged from 869 bp (*Rasbora* sp.) to 1020 bp (*R. sarawakensis*). The *Rasbora* OL was located between Asn and Cys, and the size of *Rasbora* OL was similar 34 bp.

Control region was characterized by discrete and conserved sequence blocks and exhibited the typical tripartite structure with termination-associated sequence domain (TAS), central conserved sequence blocks domain (CSB) and conserved sequence block domain (Lin et al., 2006). According to Broughton et al. (2010), the conserved sequences at the non-coding regions play important role in mitochondrial metabolism.

In fish families, 6 conserved sequence blocks have been reported in control region. The 6 common CSBs found in fish were CSB-1, CSB-2, CSB-3, CSB-D, CSB-E and CSB-F (Lee et al., 1995; Broughton and Dowling, 1994). However, only 5 CSB was identified in *Rasbora* mitochondrial CR. The 5 CSB found in *Rasbora* CR were CSB-1, CSB-2, CSB-3, CSB-D and CSB-E. The of CSB-F was not recognizable in *Rasbora* CR. Among the CSBs found in *Rasbora* CR, the CSB-2, CSB-3 and CSB-D were easily to identify. Because these 2 CSBs were show highly conserved sequence.

Also, the terminating site for H-strand found at the 5'-end of the *Rasbora* CR. The terminating site for H-strand were TAS motif and it palindromic cTAS motif. Besides, the putative transcription promoter also found in *Rasbora* CR. and this promoter was found at the 3'-end of the CR. the interesting fact about this promoter is functioning in bidirectional way (L'Abbe et al., 1991). In other words, this promoter is able to initiate the transcription for both L-strand and H-strand.

The other non-coding region is origin of L-strand replication (O_L). The position of *Rasbora* O_L is similar to the most vertebrate where located within the WANCY region, between tRNAAsn and tRNACys gene. At the base of the stem structure of O_L, a motif 5'-GCCGG-3' can be found. The 5'-GCCGG-3' motif seem to be involved in the transition from RNA synthesis to DNA (Cheng et al., 2012). However, not all *Rasbora* having the same sequence motif. The 5'-GCCGG-3'motif only present on O_L of *R. argyrotaenia* and *Rasbora* sp..

R. sarawakensis having 5'-ACCGG-3' instead of 5'-GCCGG-3'motif. The 5'-ACCGG-3' also reported present in some fishes such as in Red drum fish O_L (*Sciaenops ocellatus*) (Cheng et al., 2012), *Lophiogobius ocellicauda* (Quan et al., 2013) and many more. According to Cheng et al. (2012), the ACCGG may have the same function as GCCGG. *R. myersi* having motif 5'-GCCTG-3' instead of 5'-GCCGG-3'motif or 5'-ACCGG-3'. The 5'-GCCTG-3' also detected in *Conger myriaster* (Satoh et al., 2016). According to Satoh et al., (2016), motif 5'-GCCGG-3', motif 5'-ACCGG-3' and motif 5'-GCCTG-3'were necessary for *in-vitro* replication of the L-strand in mammals.

5.3 Phylogenetic Analyses

Mitogenome analysis has proven to be a powerful tool for species diagnosis and is essential for comprehensive evolutionary studies (Cameron, 2014). Therefore, in this study using mitogenome data to evaluate the phylogenetic position of four *Rasbora* (*R.*

argyrotaenia, *R. myersi*, *R. sarawakensis* and *Rasbora* sp.). The phylogenetic analyses were based on concatenated of 13 mitochondrial PCGs using maximum likelihood method.

The ML analyses producing slightly different topologies as they were generated based on nucleotide sequence as compared to amino acid sequence. Although they are presenting different topologies, but the position of *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp. are consistent within *Rasbora* clade on both analyses. And the mitochondrial 13 PCGs data show high confidence bootstrap value.

From the results, *R. argyrotaenia*, *R. myersi* and *R. borapentensis* are positioned within the same group. They are more closely related compared to other *Rasbora* species that were included in the analyses. Those three *Rasbora* are classified under the species group of the Argyrotaenia (Brittan, 1972), so, they are great representative of the Argyrotaenia species group. And the result of this study showing that these three species *R. argyrotaenia*, *R. myersi* and *R. borapentensis* are phylogenetically within the same clade, the group to represent the Argyrotaenia group species.

Besides, there were two mitogenome of *R. argyrotaenia* included in the phylogenetic analyses, one is generated from this study while the other is obtained from the GenBank. The relationship of both *R. argyrotaenia* were slightly different since the phylogenetic analyses were based on nucleotide sequence and amino acid sequence. However, the results from both analyses show that *R. argyrotaenia* and *R. myersi* is more closely related as compared to *R. borapentensis*, as they shared the most recent ancestor.

In nucleotide sequence analysis, the *R. argyrotaenia* is more closely related to *R. myersi* compared to GenBank *R. argyrotaenia* (LC269105). But the ML analysis based on amino acid sequence, the results show that studied *R. argyrotaenia* and GenBank *R. argyrotaenia* (LC269105) are more closely related compared to *R. myersi*. The position of the Argyrotaenia group in the ML tree based on nucleotide sequence is sister to the

Sumatrana group including two species from genus *Trigonostigma*. The Sumatrana group is sister to genus *Trigonostigma*. However, the position of the Argyrotaenia clade based on amino acid sequence is sister to the genus *Trigonostigma*, and they are sister to the Sumatrana group. From these phylogenetic analyses the Argyrotaenia group, Sumatrana group and *Trigonostigma* are closely related compared to other *Rasbora*. *Trigonostigma espei* and *Trigonostigma heteromorpha* were previously recognized as *Rasbora* species. But then, they are reassigned under a new genus known as *Trigonostigma* by Kottelat and Witten (1993) because they are having unique type of colour pattern and spawning behaviour compared to other *Rasbora* species. Nevertheless, based on our phylogenetic trees, the *Trigonostigma* is consistently placed within the *Rasbora* clade.

According to several researchers (Mayden et al., 2007; Fang et al., 2009; Tang et al., 2010) the *Rasbora* formed polyphyletic relationship. Also, their finding based on molecular phylogenetic studies presenting similar result that the evolutionary relationship of *Trigonostigma* and *Boraras* within the *Rasbora*. Well, the relationship recovered from this study also matched the previous studies in finding both *Trigonostigma* and *Boraras* embedded within *Rasbora*.

Apparently, the position of *Rasbora* sp. was also evaluated in this study. From the phylogenetic analyses results, the *Rasbora* sp. is more closely related to *R. steineri*, *R. sumatrana*, *R. lateristriata*, and *R. aprotaenia* compared to the rest of *Rasbora*. And the position of *Rasbora* sp. on the trees showed a great consistency with well bootstrap values. By referring to Lumbantobing (2014) species group classification, *R. sumatrana*, *R. aprotaenia* and *R. lateristriata* belong to the Sumatrana group.

According to Lumbantobing (2014), the Sumatrana group can be further divided into three subgroups (Hosii subgroup, Lateristriata subgroup and Elegans subgroup). The result from this study can show the monophyly Hosii subgroup (*R. sumatrana* and *R. aprotaenia*)

and monophyly Lateristriata subgroup (*R. lateristriata*). The presence of the Elegans subgroup within the Sumatrana group somehow cannot be clarified because there is no species from Elegans subgroup included in this analysis. Within the current clade of the Sumatrana group, it seems likely that *Rasbora* sp. and *R. steineri* nested within a big clade of the Sumatrana.

R. sarawakensis is slightly distant from 3 other studied *Rasbora* (*R. argyrotaenia*, R. *myersi* and *Rasbora* sp.), but still within *Rasbora*. From the result, *R. sarawakensis* formed a clade and sister to the rest of *Rasbora* (except *R. daniconius*) including *Trigonostigma* and *Boraras*. To date, *R. sarawakensis* have been assigned into the Trifasciata group, since there are no Trifasciata species included in the analyses, so, *R. sarawakensis* might be representing the clade of Trifasciata group. The ML analysis based on amino acid sequence show the *R. sarawakensis* and *B. maculatus* formed a clade. However, that phylogeny is not reliable because the ML bootstrap value less than 50%. In ML, the bootstrap value less than 50% the phylogeny is considered not supported at all.

It is hasty to say that *R. sarawakensis* and *B. maculatus* are not closely related because at some node on the evolutionary tree they are descendent from a common ancestor. Moreover, based on nucleotide sequence, it is clearly that *R. sarawakensis* and *B. maculatus* formed 2 separated clades with bootstrap supported. They are not categorized on the same group species because they are definitely distinct group species.

A big clade comprising of Argyrotaenia group, *Trigonostigma*, *R. trilineata*, and Sumatrana group is sister to *R. cephalotaenia*. And they (Argyrotaenia group, *Trigonostigma*, *R. trilineata*, and Sumatrana group + *R. cephalotaenia*) are sister to R. *sarawakensis*. *R. daniconius* is not within the big clade of *Rasbora*, but *R. daniconius* is sister to the rest of *Rasbora* instead, and sister to monophyly *Rasboroides* and monophyly *Horadandia*. From the analysis, it indicated that *Rasboroides* (*R. vaterifloris*) and

Horadandia (*H. atukorali*) are themselves monophyletic with well supported bootstrap values.

The phylogeny with weak bootstrap can be improved when their true close related species/group are included in phylogenetic analysis. Therefore, more mitogenomes data of *Rasbora species* are needed to clarify their phylogenetic position and their intra-relationship. Besides, it also explains the relationship of *Rasbora* with *Trigonostigma* and *Boraras* because some researchers (Tang et al., 2010) synonymize *Trigonostigma* and *Boraras* with *Rasbora*. The relationship recovered from these analyses match previous studies (Mayden et al., 2007; Tang et al., 2010) in finding both *Boraras* and *Trigonostigma* embedded within *Rasbora*. There is possibility that *Trigonostigma* and *Boraras* are synonym with *Rasbora*.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

All four species of *Rasbora* were successfully distinguished through morphological differences. The significant difference of morphological characters of *Rasbora* species was determined. In this study, *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp. were differentiate by interorbital width, head width, head length, eye diameter and middle caudal lobe length and the meristic counts of pectoral fin rays, pelvic fin rays, dorsal fin rays and scales of lateral line. Briefly, morphometric character and meristic character were contributing of about 60% and 80% variation of *Rasbora* species, respectively. And the correct classification of *Rasbora* were relatively high with 93% (morphometric) and 90% (meristic).

In this study, the complete mtDNA of *Rasbora* was successfully sequenced. The designed primers managed to amplify the target mtDNA. The amplified DNA fragments were assembled to form a complete sequence of mtDNA. And the complete sequence of mt-genome for *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora sp.* was 16,574 bp, 16,581 bp, 16,709 bp and 16,518 bp respectively and subsequently deposited into the GenBank.

Rasbora mt-genome comprised of a standard set of 22 tRNAs, 2 rRNAs, 13 PCGs and 2 typical non-coding regions. *Rasbora* mt-genome exhibited strand-specific asymmetry in nucleotide composition. The AT-skew and GC-skew values in the whole genome of *Rasbora* were 0.141 and -0.268 respectively, revealing that the H-strand consists mainly of C and A. All the tRNA genes could be folded into cloverleaf secondary structures, while the secondary structure of Ser $^{(AGY)}$ lacked a D arm (DHU stem). By comparing *Rasbora* genome

sequence with the recognition sites in teleost species, five conserved sequence blocks were identified in the control region. To conclude, the characteristics of *Rasbora* mt-genome were similar with other fish mitogenome and most vertebrate as well.

The phylogenetic analyses were based on the concatenated of nucleotide sequence and amino acid sequence from 13 PCGs. The phylogenetic position of *R. argyrotaenia*, *R. myersi*, *R. sarawakensis* and *Rasbora* sp. were evaluated using the Maximum Likelihood trees. The phylogenetic relationship of *Rasbora* species were explained based on the bootstrapping supported value. All the present *Rasbora* species was placed within the *Rasbora* clade, which mean they are closely related species and also derived from the same ancestor. Overall, this study enriches our knowledge of *Rasbora* mt-genomes and provides valuable information on the evolution of present *Rasbora* species.

6.2 Recommendations

Universal primers should be designed to sequence whole mtDNA of *Rasbora* in future. Then, it will be easily to sequence complete mtDNA of *Rasbora* with more mtDNA of *Rasbora* species to be available in the future. Reconstructing phylogenetic trees using various method (Neighbour joining and Bayesian inference) in order to verify the phylogenetic study of *Rasbora* and also with higher variety of Rasbora species once their mitogenome becomes available.

REFERENCES

Adachi, J., & Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution, 42*, 459-468.

Arif, I. A., & Khan, H. A. (2009). Molecular markers for biodiversity analysis of wildlife animals**:** a brief review. *Animal Biodiversity and Conservation, 32*(1), 9–17.

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.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research, 27*(8), 1767-1780.

Brittan, M. R. (1984). *Rasboras*. Jersey City, New Jersey**:** T.F.H. Publications.

Brittan, M. R. (1954). A revision of the Indo-Malayan freshwater fish genus *Rasbora*. *Monographs of the Institute of Science and Technology, Manila, 3*, 1–224.

Brittan, M. R. (1972) *Rasbora**:** a revision of the Indo-Malayan fresh-water fish genus Rasbora*. T.F.H. Publications, *Hong Kong*, 224 pp.

Bronstein, O., Kroh, A., & Haring, E. (2018). Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. *BMC Evolutionary Biology, 18*(1), 1–15.

Broughton, R. E., & Dowling, T. E. (1994). Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics*, *138*(1), 179–190.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *76*(4), 1967-1971.

Cameron, S. L. (2014). Insect mitochondrial genomics**:** implications for evolution and phylogeny. *Annual Review of Entomology*, *59*(1), 95-117.

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. H., Tsai, C. L., & Jang-Liaw, N. H. (2013). Complete mitochondrial genome of the Chinese rasbora, *Rasbora steineri* (Teleostei, Cyprinidae). *Mitochondrial DNA*, *24*(3), 183-185.

Cheng, Y., Shi, G., Xu, T., Li, H., Sun, Y., & Wang, R. (2012). Complete mitochondrial genome of the red drum, *Sciaenops ocellatus* (Perciformes, Sciaenidae)**:** Absence of the typical conserved motif in the origin of the light-strand replication. *Mitochondrial DNA*, *23*(2), 126–128.

Conant, G. C., & Wolfe, K. H. (2007). GenomeVx**:** Simple web-based creation of editable circular chromosome maps. *Bioinformatics*, *24*(6), 861-862.

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*, 202–220.

Fang, F., Norén, 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.

Hixson, J. E., Wong, T. W., & Clayton, D. A. (1986). Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *Journal of Biological Chemistry, 261*(5), 2384–2390.

Ho, C. W., Liu, M. Y., & Chen, M.H. (2016) Complete mitochondrial genome of *Rasbora trilineata* (Cypriniformes, Cyprinidae). *Mitochondrial DNA Part A, 27*(3), 1755-1757.

Howes, G., J. (1991). Systematics and biogeography**:** An overview. In**:** Winfield IJ, Nelson JS (eds). Cyprinid fishes**:** systematics, biology and exploitation. London**:** Chapman and Hall.

Huang, Z. H., & Tu, F. Y. (2016). Characterization and evolution of the mitochondrial DNA control region in Ranidae and their phylogenetic relationship. *Genetics and Molecular Research, 15*(3), 1-9.

Inger, R. F., & Chin, P. K. (1962). *The fresh-water fishes of North Borneo*. Chicago, U.S.A. : Chicago Natural History Museum.

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., & Nishida, M. (2013). MitoFish and MitoAnnotator**:** A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Molecular Biology and Evolution, 30*(11), 2531–2540.

Kottelat, M., & Tan, H., H. (2012). *Rasbora cryptica*, a new species of fish from Sarawak, Borneo (Teleostei**:** Cyprinidae). *Ichthyological Exploration of Freshwaters, 23*, 37–44.

Kottelat, M., & Vidthayanon, C. (1993). Boraras micros, a new genus and species of minute freshwater fish from Thailand (Teleostei**:** Cyprinidae). *Ichthyological Exploration of Freshwaters, 4*, 161–176.

Kottelat, M. (1999). Nomenclature of the genera *Barbodes*, *Cyclocheilichthys*, *Rasbora* and *Chonerhinos* (Teleostei**:** Cyprinidae and Tetraodontidae), with comments on the definition of the first reviser. *Raffles Bulletin of Zoology*, *47*(2), 591–600.

Kottelat, M. (2013). The Fishes of the Inland Waters of Southeast Asia**:** A Catalogue and Core Bibliography of the Fishes Known to Occur in Freshwaters, Mangroves and Estuaries. *Raffles Bulletin of Zoology, 27*, 1-663.

Kottelat, M. (1991). Notes on the taxonomy of some Sundaic and Indochinese species of *Rasbora*, with description of four new species (Pisces**:** Cyprinidae). *Ichthyological Exploration of Freshwaters, 2*, 177–191.

Kottelat, M., & Witte, K. (1999). Two new species of MicroRasbora from Thailand and Myanmar**:** with two new generic names for small Southeast Asian cyprinid fishes (Teleostei**:** Cyprinidae). *Journal of South Asian Natural History, 4*(1), 49–56.

Kottelat, M., Whitten, A.J., Kartikasari, S.N., & Wirjoatmodjo, S. (1993) Freshwater fishes of Western Indonesia and Sulawesi. Periplus, Hong Kong, 259 pp + 84 plates.

Kumar, S., Stecher G., Li M., Knyaz C., & Tamura K. (2018). MEGA X**:** Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution,* 35, 1547-1549.

Kusuma, W. E., & Kumazawa, Y. (2016). Complete mitochondrial genome sequences of two Indonesian rasboras (*Rasbora aprotaenia* and *Rasbora lateristriata*). Mitochondrial DNA Part A, *27*(6), 4222-4223.

Kusuma, W. E., Samuel, P. D., Wiadnya, D G.R., 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.

Lim, W. K. L., Roja, J. S., Kamar, C. K. A., Chung, H. H., Liao, Y., Lam, T. T., & Chong, Y. L. (2019). Sequencing and characterization of complete mitogenome DNA for *Rasbora myersi* (Cypriniformes**:** Cyprinidae**:** Rasbora) and its evolutionary significance. *Gene Reports*, *17*(2019), 100499.

Lin, G., Lo, L. C., Zhu, Z. Y., Feng, F., Chou, R., & Yue, G. H. (2006). The complete mitochondrial genome sequence and characterization of single-nucleotide polymorphisms in the control region of the asian seabass (*Lates calcarifer*). *Marine Biotechnology, 8*(1), 71 – 79.

Lorenzo-Seva, U. (2013). How to report the percentage of explained common variance in exploratory factor analysis? Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona.

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line**:** search and contextual analysis of transfer RNA genes. *Nucleic Acids Research*, *44*(1), 54-57.

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.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10-12.

Mayden, R. L., Tang, K. L., Conway, K. W., Freyhof, J., Chamberlain, S., Haskins, M., & He, S. (2007). Phylogenetic relationships of Danio within the order Cypriniformes**:** a framework for comparative and evolutionary studies of a model species. *Journal of Experimental Zoology Part B**:** Molecular and Developmental Evolution, 308B*(5), 642–654.

Miya, M., & Nishida, M. (1999). Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei**:** Stomiiformes)**:** First Example of Transfer RNA Gene Rearrangements in Bony Fishes. *Marine Biotechnology*, *1*(5), 416–426.

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.

Nei M., & Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nelson J.S. (2006). *Fishes of the world*. 4th Edition. John Wiley & Sons.

Nelson, J. S., Grande, T., & Wilson, M. V., H. (2016). *Fishes of the World* (5th ed.) Hoboken, New Jersey**:** John Wiley & Sons.

Ng, H. H., & Kottelat, M. (2013). The identity of the cyprinid fishes *Rasbora dusonensis* and *R. tonieri* (Teleostei**:** Cyprinidae). *Zootaxa, 3635*(1), 62-70.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation, 9*(11), 1-9.

Pereira, S. L. (2000). Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology*, *23*(4), 745–752.

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.

Quan, X., Jin, X., & Sun, Y. (2013). The complete mitochondrial genome of *Lophiogobius ocellicauda* (Perciformes, Gobiidae). *Mitochondrial DNA, 25*(2), 95 – 97.

Roberts, T.R. (1989) The freshwater fishes of Western Borneo (Kalimantan Barat, Indonesia). *Memoirs of the California Academy of Sciences*, 14, 1–210.

Sato, M., & Sato, K. (2012). Maternal inheritance of mitochondrial DNA**:** Degradation of paternal mitochondria by allogeneic organelle autophagy, allophagy. *Autophagy*, *8*(3), 424–425.

Satoh, T. P., Miya, M., Mabuchi, K., & Nishida, M. (2016). Structure and variation of the mitochondrial genome of fishes. *BMC Genomics, 17*(1), 1 – 20.

Shi, X., Tian, P., Lin, R., Huang, D., & Wang, J. (2016). Characterization of the complete mitochondrial genome sequence of the globose head whiptail *Cetonurus globiceps* (Gadiformes**:** Macrouridae) and its phylogenetic analysis. *PLOS ONE, 11*(4), 1 – 17.

Siebert, D.J., & Guiry, S. (1996) *Rasbora johannae* (Teleostei**:** Cyprinidae), a new species of the *R. trifasciata*-complex from Kalimantan, Indonesia. *Cybium,* 20, 395–404.

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.

Tang, K. L., Agnew, M. K., Hirt, M. V., Sado, T., Schneider, L. M., Freyhof, J., & Mayden, R. L. (2010). Systematics of the subfamily Danioninae (Teleostei**:** Cypriniformes**:** Cyprinidae). *Molecular Phylogenetics and Evolution, 57*(1), 189–214.

Wang, C., Chen, Q., Lu, G., Xu, J., Yang, Q., & Li, S. (2008). Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella,* Teleostei)**:** Insight into its phylogenic position within Cyprinidae. *Gene, 424*(1–2), 96–101.

Wasko, A. P., Martins, C., Oliveira, C., & Foresti, F. (2003). Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas,138*(3), 161-165.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415. Aravinda, R. K., Natarajamani, S., & Behera, S. K. (2012). Antipodal vivaldi antenna UWB antenna with 5.5GHz band-notch characteristics. In *2012 International Conference on Computing, Electronics and Electrical Technologies,* (pp. 821–824).