学位の種類	博士（理学）
報告番号	乙第１８７０号
学位記番号	論　第10号
氏名	WAHYU ENDRA KUSUMA
授与年月日	平成 28年 12月 8日
学位論文の題名	Molecular phylogeny and historical biogeography of the Indonesian freshwater fish Rasbora lateristriata species complex (Actinopterygii: Cyprinidae)
論文審査担当者	主査：　熊澤 慶伯 副査：　鈴木 善幸，村瀬 香，渡辺 勝敏

名古屋市立大学 博士学位論文

Molecular phylogeny and historical biogeography of the Indonesian freshwater fish *Rasbora lateristriata* species complex (Actinopterygii: Cyprinidae)

2016 年

Wahyu Endra Kusuma

名古屋市立大学大学院システム自然科学研究科

Abstract

Indonesia is one of megadiverse countries that hold a huge number of world's biodiversity. Indonesia is a home for more than 1200 species of freshwater fish. Owing to human activities, this biodiversity is under serious threats. Due to their inability to disperse through non-freshwater environments, freshwater fish are highly vulnerable to pollution and environmental changes. Basic scientific information on, e.g., taxonomy, phylogeny, ecology, and genetic diversity should be urgently collected before they disappear.

Small freshwater fishes in genus *Rasbora* (87 valid species) are distributed in a large geographical area, ranging from western India to Lesser Sunda Islands of Indonesia, as far as Lombok and Sumbawa Islands. Among them, 66 species occur in Indonesia. *Rasbora lateristriata* was described from Java Island but its taxonomy, phylogeny, and distributional boundary have not been fully studied. This species occur in both western (Java and Bali Islands) and eastern (Lombok and Sumbawa Islands) sides of Wallace's Line, a geographical barrier between Indomalaya Ecozone in the west with fauna of the Asian origin and Australasia Ecozone in the east with those of the Australian origin. When and how this species crossed Wallace's Line is one of the biggest enigmas in the historical biogeography of this region. *Rasbora baliensis* was described as a species endemic to Balinese lakes but its taxonomic status has been controversial in relation to *R. lateristriata*.

I collected rasbora samples from 47 freshwater localities of Java and five neighboring Islands, which included 236 individuals assignable to *R. lateristriata* or *R. baliensis*. I extracted DNA from several individuals representing each locality, amplified a part of 4 genes (mitochondrial COI and Cytb genes and nuclear RAG1 and opsin genes), and sequenced them. These genes were also sequenced for a representative individual of other *Rasbora* species that I collected. Molecular phylogenetic analyses were conducted with these DNA sequences and those obtained from public databases for additional species. I also conducted

morphological analyses with many meristic and morphometric characters, including ones that were used to describe *R. baliensis* by Brittan (1954).

Molecular analyses using four genes, as well as morphological analyses featuring the body color pattern consistently supported that the currently recognized *R. lateristriata* forms a species complex including at least four major lineages that possibly represent different species. In one of the major lineages, Balinese individuals clustered tightly with those from East Javanese, Lombok and Sumbawa localities, calling for taxonomic revision on *R. baliensis*. The other three major lineages occur in distinct regions of central, west-central, and western Java and they can be clearly discriminated by the combination of melanophore pigment patterns in the basicaudal blotch and the supra anal pigment.

Molecular phylogeny of this study suggested west-to-east divergences of the *R. lateristriata* species complex. *R. lateristriata* likely had an origin in Sumatra or western parts of Java and then migrated to the east before it crossed Wallace's Line, colonizing Lombok and Sumbawa Islands. Based on the relaxed-clock Bayesian estimation of divergence times using the nuclear gene sequences, the divergences of this species complex in Java Island probably occurred from the late Miocene to Plio-Pleistocene. The dispersal over Wallace's Line occurred very recently (less than five hundred thousand years ago) either naturally or by human introduction.

In conclusion, the present study revealed some hidden biodiversity on *Rasbora* fish in Java, provided new molecular and morphological evidence to revise the taxonomy of *R. lateristriata* and *R. baliensis*, and proposed a new hypothesis on the origin and migrational pathway of the *R. lateristriata* species complex. As their natural habitats are rapidly deteriorated by human activities, many freshwater fish species other than the rasboras await molecular and morphological investigations. As demonstrated in this study, multidisciplinary approaches by field sampling, morphological investigations, molecular experiments, and computational analyses will be effective to tackle complex evolutionary issues and provide basic scientific knowledge necessary to design conservation plans on Indonesian fauna.

Glossary of specialized terms

Bayesian method: a method using Bayesian principle for estimating posterior probabilities of phylogenetic trees based on observed molecular data under a particular model of sequence change and its optimized parameters. Markov chain Monte Carlo (MCMC) process is usually employed to heuristically search for better trees in multi-dimensional space of trees and parameters. MrBayes is a program for phylogenetic inference which performs the Bayesian analysis

Biodiversity (Biological diversity): variability among living organisms, including plants, animals and microbes. The biodiversity includes species diversity, genetic diversity within species, and ecosystem diversity

Bootstrap: a procedure to assess the precision in estimating the phylogenetic tree by resampling a random subset of the original data matrix (DNA sequences)

Clade (monophyletic group): a group of organisms (taxa) that include a common ancestor and all descendants of that ancestor

Conspecific: an organism belonging to the same species

Cryptic species: two or more species with very similar morphological appearance (and thus classified into a single species conventionally) but are genetically distinct in the species level

Haplotype (haploid genotype): a group of genes within a chromosome of an organism which is inherited from a single parent. Mitochondrial DNA genotype is sometimes called a haplotype because mtDNA inherits maternally only from a mother to her children

Historical biogeography: a study of species distributions in a geographical region through the geological time, elucidated by the phylogenetic study and distribution information

Lineage sorting: the random process of fixation of gene lineages along a species lineage

Maximum likelihood: the maximum likelihood method in phylogenetic analyses infers the maximum likelihood tree that has the highest probability of realizing observed molecular data under a particular model of sequence change and its parameters. One of the popular softwares to perform phylogenetic inference based on the maximum likelihood criterion is GARLI (Genetic Algorithm for Rapid Likelihood Inference)

MEGA (Molecular Evolutionary Genetic Analysis): a software package for conducting various tasks in molecular evolutionary studies, e.g., sequence alignment, inferring phylogenetic trees, estimating rates of molecular evolution, calculating genetic distances and testing evolutionary hypotheses

Meristic: countable traits that can be used for identifying or describing a species, e.g., number of fins, number of scales or number of gills

Morphometric: measurable or quantitative traits, e.g., size, distance or proportion in morphological characters

Node: a connecting point in the phylogenetic tree, which represents a common ancestor of descendants

Outgroup: an outgroup in phylogenetic analyses is a group of taxa (genes or organisms) that can be used as a reference to determine the root position of ingroup taxa and infer the evolutionary direction of character changes (i.e., ancestral vs. derived characters)

Phylogeny (phylogenetic tree): is a diagrammatic hypothesis about the evolutionary relationships of a group of organisms or genes

Phylogeography: a study concerned with the principles and processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species

Polymerase Chain Reaction (PCR): a technique in molecular biology to amplify a segment of DNA into multiple copies

Quaternary glaciation: a cooling event in Quaternary period when the earth experienced extreme cooling and the ice sheet was expanded globally. The Quaternary glacial maxima appeared periodically from 2.58 million years ago to 11,000 years ago

Species complex: a group of closely related species with very similar morphological appearance, so that the species delimitation between them is obscure

Sympatric: a distributional state of multiple organisms occurring within a locality or an area

Taxon (taxa in plural): a taxonomic group or unit, a group of one or more populations of an organism or organisms seen by taxonomists to form a unit

Topology: a branching pattern of a phylogenetic tree

Type locality: a location where the designated type specimen was originally collected

Type specimen: is particular organismal individuals to which the scientific name of a species was permanently attached according to the description based on these specimens. Holotype is a main specimen which is designated in the original description of a species by the original author. Paratype refers to several additional specimens designated together with the holotype. Usually, a taxonomist describes new species by observing several specimens. One specimen is assigned to the holotype and the rest is designated as the paratype

Wallace's Line: is a hypothetical faunal boundary which separates Asian and Australasian faunas with different origins. Wallace's Line runs through the Lombok Strait between Bali and Lombok Islands and through the Makassar Strait between Borneo and Sulawesi Islands in Indonesia

Table of Contents

Chapter 1: Introduction

1.1 Indonesia: a megadiverse country

Biological diversity or "biodiversity" was defined by United Nations of Environment Programme as the variability among living organisms from all sources including terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part, which includes the species diversity, genetic diversity, and ecosystem diversity (UNEP, 1992). Understanding the biodiversity is certainly related to the need of knowing fundamental issues, e.g., how many living organisms (species) currently inhabit our planet. There is no certain number on how many species are on earth and the comprehensive view on the global biodiversity is far from complete. Several researchers proposed an approximate number of global biodiversity ranging from 3 to 100 million species (Hamilton et al., 2010; May, 2010). More recently, Mora et al. (2011) suggested a more precise number (8.7 million eukaryotic species) from which only 1.2 million have been successfully described as valid species. In other words, >85% of the earth's species are unknown.

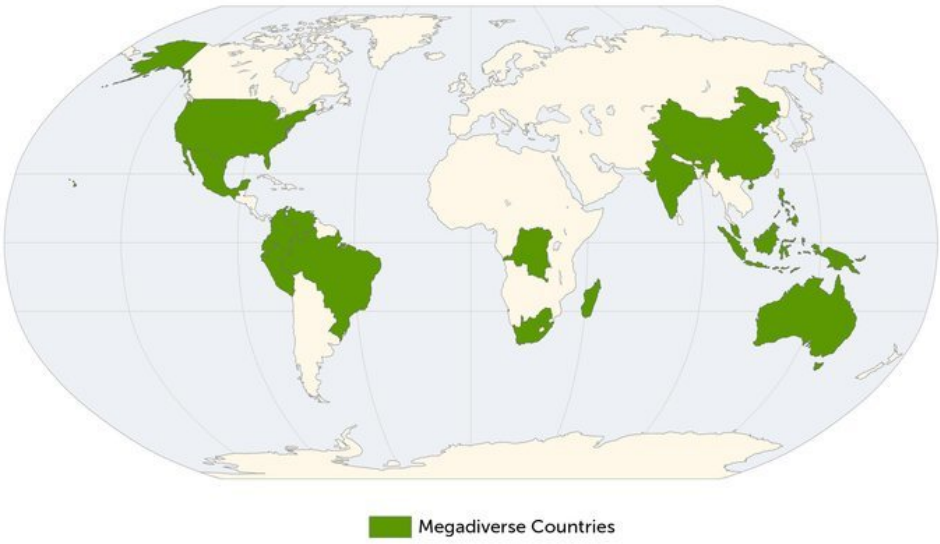

Fig. 1. Seventeen megadiverse countries based on the diversity and number of endemism of higher plants and vertebrates as proposed by Mittermeier et al. (1997). Indonesia was placed in the second position after Brazil (see Table 1). The image was obtained from Mittermeier et al. (1997).

More than 70% of the above-mentioned earth's biodiversity concentrate on 17 countries known as "Megadiverse Countries" (Mittermeier, 1988) (Fig. 1). A megadiverse country can therefore be regarded as a country that holds a large portion of earth's species. To be classified as a megadiverse country, a country must hold two important criteria. First, a country must have more than 5000 plant species that are considered to be endemic. Second, a country must have marine ecosystem within its border (Mittermeier et al., 1997). Indonesia unquestionably is extremely rich in biodiversity. Indonesia is also known as an archipelagic island country. This country is a home to ~ 37000 higher plant species (~18500 of them are endemic species) and more than 7000 vertebrates species (~2100 species are endemic), thus being as one of the megadiverse countries (Table 1).

The next question is, what makes Indonesia possess huge biodiversity? Indonesia is an archipelagic country located in an intriguing geographical location. This country includes multiple ecozones: Indomalaya Ecozone in the west with fauna of the Asian origin and Australasia Ecozone in the east with those of the Australian origin (Wallace, 1860; Metcalfe et al., 2001). Wallacea (Sulawesi, Lombok, Sumbawa, Flores and other islands) is situated in the middle between Sundaland (the Malay Peninsula, Sumatra, Borneo, Java and Bali) and Near Oceania including Australia and New Guinea (Monk et al., 1997; Hall, 2009; Hall et al., 2011). Sundaland is characterized by shallow depth (less than 200 m) in most of the area with Wallace's Line as an eastern boundary (Woodruff, 2003; Hall and Morley, 2004; Hall, 2008). These geographical regions have long been associated with complex geological, biogeographical, climatic and environmental histories. The unique and complex features of this geographic region contribute to the high species richness and endemism of its biota (Woodruff, 2003, 2010; Lohman et al., 2011; de Bruyn et al., 2014).

Table 1. Data for number of species and endemism of higher plant and vertebrate species in megadiversity countries according to Mittermeier et al. (1997)

No	Country	Higher plants	Mammals	Birds	Reptilians	Amphibians	Vertebrate except fish	Freshwater fish[a]	% endemism in global diversity
1	Brazil	50000-56000 (16500-18500)	524 (131)[b]	1622 (>191)	468 (172)	517 (294)	3131 (788)	>3000	3.3
2	Indonesia	37000 (14800-18500)	515 (201)	1531 (397)	511 (150)	270 (100)	2827 (848)	1400 (456)	3.5
3	Colombia	45000-51000 (15000-17000)	456 (28)	1815 (>142)	520 (97)	583 (367)	3374 (634)	>1500	2.6
4	Mexico	18000-30000 (10000-15000)	450 (140)	1050 (125)	717 (368)	284 (169)	2501 (802)	468	3.3
5	Australia	15638 (14458)	282 (210)	751 (355)	755 (616)	196 (169)	1984 (1350)	183	5.6
6	Madagascar	11000-12000 (8800-9600)	105 (77)	253 (103)	300 (274)	178 (176)	836 (630)	75	2.6
7	China	27100-30000 (10000)	499 (77)	1244 (99)	387 (133)	274 (175)	2404 (484)	1010	2
8	Philippines	8000-12000 (3800-6000)	201 (116)	556 (183)	193 (131)	63 (44)	1013 (474)	330 (109)	1.98
9	India	>17000 (7025-7875)	350 (44)	1258 (52)	408 (187)	206 (110)	2222 (393)	750	1.6
10	Peru	18000-20000 (5356)	344 (46)	1703 (109)	298 (98)	241 (>89)	2586 (342)	855	1.4
11	Papua New Guinea	15000-21000 (10500-16000)	242 (57)	762 (85)	305 (79)	200 (134)	1509 (355)	282	1.5
12	Ecuador	17600-21100 (4000-5000)	271 (21)	1559 (37)	374 (114)	402 (138)	2606 (310)	>44	1.3
13	United States of America	18956 (4036)	428 (101)	768 (71)	261 (90)	194 (126)	1651 (388)	790	1.6
14	Venezuela	15000-21070 (5000-8000)	228 (11)	1360 (45)	293 (57)	204 (76)	2145 (189)	1250	0.8
15	Malaysia	15000 (6500-8000)	286 (27)	738 (11)	268 (68)	158 (57)	1459 (163)	600 (55)	0.7
16	South Africa	23420 (16500)	247 (27)	774 (7)	299 (76)	95 (36)	1415 (146)	153	0.6
17	Dem. Rep. of Congo	11000 (3200)	415 (28)	1094 (23)	268 (33)	80 (53)	1857 (137)	962	0.6

[a] Data for number of endemic freshwater fish species available only for Indonesia, Philippines and Malaysia. Data were taken from ASEAN Member States (AMS) Species Inventory by Taxa (2008)

[b] Number inside parenthesis indicate endemic species

1.2 Threats to Indonesian biodiversity

Although Indonesia holds a significant proportion of world's biodiversity, its biodiversity is currently under serious threats owing to human activities and/or environmental change in an unprecedented speed. High economic growth in Indonesia is often associated with the high rate of biodiversity loss. Increasing numbers of human populations and their economic activities lead to the loss of biodiversity in all the three categories of its definition (species, gene and ecosystem) through, e.g., habitat degradation by exploitation, pollution and climatic change, overfishing (overhunting), and introduction of alien and invasive species (Groom, 2005). For example, a century ago 90% of Borneo Island was covered with tropical rainforests (Fig. 2). Since 1997, one million ha of Indonesian rainforests have been destroyed every year (World Bank, 2001).

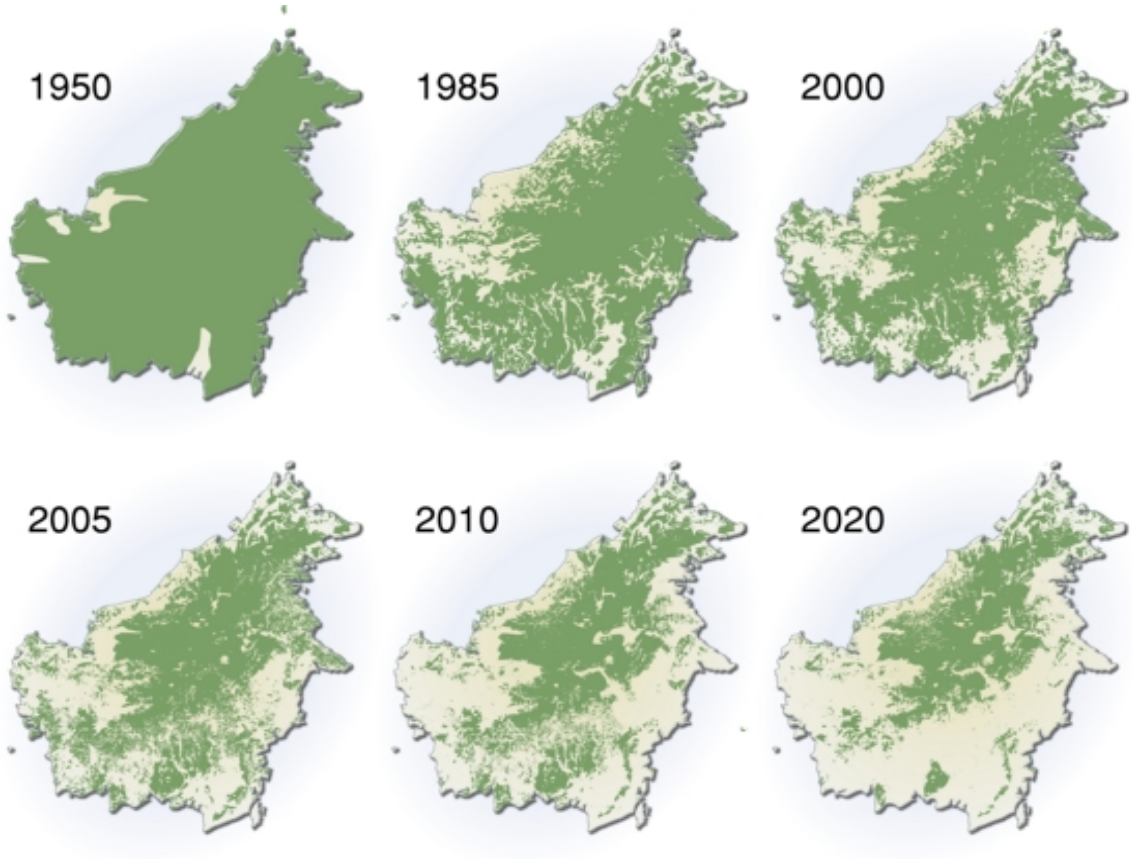

Fig. 2. A map showing the speed of deforestation in Borneo Island. The image is available under permission from GRID-Arendal (http://www.grida.no/graphicslib/detail/extent-of-deforestation-in-borneo-1950-2005-and-projection-towards-2020_119c).

Recent data by International Union for Conservation of Nature (Baillie et al., 2004) showed high numbers of threatened species in Indonesia. In mammals, there are 146 threatened species and 88 species of them are endemic. In birds, among 119 threatened species, 69 are endemic. In amphibians and turtles, the numbers of threatened species and endemic threatened species are 33/23 and 24/4, respectively. In addition, Indonesian government, through the Government Regulation of the Republic Indonesia No. 7/1999 (PP No.7/1999), identified 294 species that should be prioritized for conservation. Therefore, it is not surprising that Indonesia is categorized as one of the biodiversity hotspots with high abundance of endemic species and considerable loss of habitats (Myers et al., 2000). If there are no effective conservation actions, the biodiversity crisis in this country will be unrecoverable in future.

Among vertebrate groups (e.g., mammals, birds, amphibians, turtles and marine fishes), freshwater fishes are considered to be the most endangered. Rates of biodiversity loss or extinction in freshwater fish are greater than those in other animal groups and this makes conservation actions in freshwater fish become a priority (Burkhead, 2012; Reid at al., 2013). There are at least five major threats to global freshwater biodiversity: overexploitation, water pollution, flow modification, degradation of habitat, and invasion of exotic species (Dudgeon et al., 2006). What causes freshwater fish to have higher rates to extinction? Freshwater fish is a water-dependent organism with the limitation of dispersal constrained only within and among drainages. Habitat isolation and dispersal limitation for a long period may lead to a speciation event. In one hand, this situation is responsible for the high biodiversity and endemism of freshwater fish. On the other hand, freshwater fish become highly vulnerable to pollution and environmental change (Cambray and Bianco, 1998; Olden at al., 2010). A population of freshwater fish may be wiped-out when its environment is disrupted by pollution because of the inability to disperse through non-freshwater environments.

Basic scientific information on, e.g., taxonomy, phylogeny, ecology, and genetic diversity is important for designing conservation management for freshwater fish. In addition, historical biogeography illustrates the evolutionary history of a species on how past environmental and/or geological events influenced its contemporary distribution. Understanding how a species evolved and survived in the past in response to environmental changes is also informative to predict its future sustainability (Avise, 2000; Olden et al., 2010). As another example, population genetic studies play an important role in developing effective conservation plans. Populations with low genetic diversity will be under a high risk of extinction (Frankham, 2003, 2005).

Unfortunately, very few above-mentioned studies have been conducted in Indonesia so far. As Indonesian biodiversity is decreasing in an enormous speed, many species are waiting for the basic scientific research and subsequent conservation actions to ensure their sustainability. Otherwise, they may become extinct before we are aware of their existence. Especially, basic scientific research on Indonesian freshwater fish must be conducted as soon as possible.

1.3 Modern approaches for studying the biodiversity

The basic but very important aspect for describing the biodiversity is how to accurately recognize taxonomic status of a species. Most of the taxonomic information have been obtained based on morphological characters. In fishes, a species can be distinguished from its congeners (closely related organisms within the same genus) using either measurable (morphometric) or countable (meristic) characters or combination of them. However, in some cases, morphology-based approaches fail to distinguish individuals that have very similar morphological appearance but actually belong to separate species (cryptic species). For example, a morphologically similar Cuban freshwater fish Caribbean gambusia (*Gambusia puncticulata*) was actually composed of four different cryptic species as revealed by

molecular analysis (Lara et al., 2010). In addition, the morphology-based approaches are time-consuming in general and may be susceptible to misleading conclusions due to homoplasious or adaptive character changes in morphology. A recent trend is therefore to use molecular approaches together with morphological ones in investigating taxonomic and/or phylogenetic status of a species (Hebert et al., 2003; Hebert and Gregory, 2005; Ward et al., 2005).

DNA is a heritable material and a blueprint for making all organisms. The DNA sequence varies between individuals in a species, as well as between species. Based on an assumption that the degree of difference in DNA sequence between taxa (molecular divergence) reflects their relatedness, we can estimate how close their relationship is. If two taxa have very similar DNA sequences, it implicates that they are closely related with each other and their direct common ancestor diverged very recently. In contrast, more distantly related organisms in phylogeny will show higher molecular divergences as a consequence of more base substitutions on each lineage after speciation. Phylogenetic relationships among species can be inferred through a phylogenetic analysis using molecules (DNA or protein) and/or their morphological traits (Nei and Kumar, 2000; Avise, 2004). A result of the phylogenetic analyses is a tree-like pattern called a phylogeny or a phylogenetic tree, i.e., a diagrammatic hypothesis that depicts evolutionary relationships among organisms under study.

Molecular phylogenetic trees can be constructed by several methods involving computational analyses: e.g., neighbor joining method (Saitou and Nei, 1987), maximum parsimony method (Edwards and Cavalli-Sforza, 1963), maximum likelihood method (Felsenstein, 1981) and Bayesian method (Rannala and Yang, 1996). In recent years with highly developed computing environments, maximum likelihood and Bayesian methods have been preferred because their conclusions are based on explicit principles for choosing the best tree under certain models and parameters (reviewed in Yang and Rannala, 2012). The

maximum likelihood method is a method for inferring a tree that maximizes the probability of realizing the observed molecular data under a specific model of base substitution and its optimized parameters while Bayesian method finds topological relationships with the highest posterior probability based on the observed data, model and parameters (Felsenstein, 1981; Rannala and Yang, 1996; Nei and Kumar, 2000; Yang and Rannala, 2012).

The DNA barcoding was proposed by Herbert et al. (2003) for species identification by rapid, accurate and automated procedures using a short and standardized gene region as a molecular tag. A mitochondrial DNA region encoding cytochrome c oxidase subunit I (COI, ~ 655 bp) was selected as the barcode tag for animals due to its relatively high evolutionary rate and absence of insertion-deletion patterns. For delimiting a species, the "Barcode Gap" was proposed based on an idea that intraspecific variations in the COI sequences should have smaller molecular divergences than interspecific variations. A recent study showed that the species delimitation in many animal groups can be done with a threshold of 2% molecular divergence in the DNA barcoding region and that individual species thus recognized can be given a unique barcode index number (BIN, Ratnasingham and Hebert, 2013). Although the concept of the DNA barcoding raised some controversies in the scientific community (e.g., Ebach and Holdrege, 2005; Hickerson et al., 2006), this approach has demonstrated its effectiveness in many animal and plant groups, especially in freshwater fishes (Hubert et al., 2008; Lara et al., 2010; Collins et al., 2012: Rosso et al., 2012; Young et al., 2013).

1.4 Indonesian freshwater fishes: rasboras

How many freshwater fish species are distributed in Indonesia? A survey conducted by Kottelat et al. (1993) in western parts of Indonesia and Sulawesi discovered 964 species. Further survey conducted in 1996 reported 79 additional species (Kottelat and Whitten, 1996). Mittermeier et al. (1997) mentioned 1400 freshwater fish species is distributed in Indonesia. More recent data provided by FishBase (Froese and Pauly, 2015), which also cover eastern

regions of Indonesia including Papua, recorded 1228 freshwater fishes. Brazil (> 3000 species) and Colombia (> 1500 species) are the only countries that exhibit higher freshwater fish diversity than Indonesia (Mittermeier et al., 1997). One of the most species-rich freshwater fish groups in Indonesia is family Cyprinidae.

In Cyprinidae, genus *Rasbora* is one of the most species-rich genera, with 87 valid species so far recognized (Eschmeyer, 2015). *Rasbora* is distributed in a large geographical area, ranging from western India to Lesser Sunda Islands of Indonesia, as far as Lombok and Sumbawa Islands (Brittan, 1954, 1972, 1998; Kottelat et al., 1993; Froese and Pauly, 2015). Currently, sixty-six species of rasboras are naturally distributed in Indonesia and many of them are endemic species (Froese and Pauly, 2015). Recent studies conducted by Lumbantobing (2010, 2014) described eight new species from Sumatra Island. In addition, several new species are now ready to be described and most of them are endemic species (Lumbantobing, D.N., personal communication).

Rasbora is a schooling fish that swims together in the same direction for behavioral reasons. This species occurs strictly in freshwater habitats (i.e., primary freshwater fish) and can be easily found in large and small rivers, ponds, ditches, lakes, paddy field, and swamps. It rarely occurs in low oxygen waters and mountainous rivers with a swift current (Brittan, 1954, 1972, 1998). They breed only by sexual reproduction in which females lay semi-adhesive eggs on the underside of water plants and males quickly release the sperm for fertilization (Brittan, 1998). *Rasbora* is characterized by small to medium sizes (5-20 cm in the standard length) with a body elongated and compressed laterally, a symphyseal knob on the tip of lower jaw without barbels, and a dark lateral stripe that extends from the opercle to the caudal fin base.

One of broadly distributed species in the genus *Rasbora* is *R. lateristriata*. This species is supposed to be distributed from Borneo, Sumatra, Java, Bali, across Wallace's Line, to Lombok and Sumbawa Islands of Indonesia based on some literatures (Kottelat et al., 1993;

Froese and Pauly, 2015) but its exact distributional range is unclear. *R. lateristriata* was first described as *Leuciscus lateristriatus* by Bleeker (1854) using several specimens from Java and Sumatra, including materials from Bogor, West Java, collected by Kuhl and van Hasselt (van Hasselt, 1823). Bleeker (1860) revised *L. lateristriatus* to *Rasbora lateristriata*. However, some researchers (e.g., Brittan, 1954 and Alfred, 1963) later suggested that the Sumatran specimens do not match the original description of the species. They assigned only specimens from Java to *R. lateristriata*.

*Rasbora baliensi*s was described by Brittan (1954) from a small crater lake, Lake Bratan in Bali Island as the closest relative of *R. lateristriata*. *R. baliensis* is supposedly a species endemic to Bali (Brittan, 1954, 1972; Kottelat et al., 1993; Whitten et al., 1996). However, *R. baliensis* was described based on small numbers of specimens and relatively indistinct segregating characters (Brittan, 1954). Some researchers (Kottelat and Vidthayanon, 1993; Whitten et al., 1996) suspected that *R. baliensis* from Bali might be indistinguishable from *R. lateristriata* from eastern parts of Java, pointing out that the taxonomic status of *R. baliensis* needs to be re-evaluated. Although molecular phylogeny involving many *Rasbora* species has been studied (Mayden et al., 2007; Rüber et al., 2007; Britz et al., 2009; Fang et al., 2009; Tang et al., 2010), *R. lateristriata* and *R. baliensis* were not included in these studies. Thus, taxonomic and phylogenetic status on these species is still uncertain.

The evolutionary relationships and contemporary distributions of Indonesian freshwater fauna have been likely associated with the recurrent sea level changes in the Quaternary glaciation that occurred 2.6 to 0.01 million years ago (Mya). By the recurrent sea level changes in the Quaternary glaciation, the sea level repeatedly fell up to 120 m and rose up to 20 m from the present level. The Sunda shelf became dried and exposed throughout marine regression, forming a land bridge connecting Sumatra, Borneo, Java and Bali Islands (and some other intervening islands) with Indo-China and created a massive landmass called Sundaland. On the contrary, during the period of marine transgression, these islands became

disconnected and isolated (Rainboth, 1996; Voris, 2000; Woodruff, 2010) (Fig. 3). In the period of lower sea levels, the freshwater fauna may have expanded their geographical distribution by traversing the paleo-drainage systems. On the other hand, during periods of higher sea levels, when the time is sufficient enough to prevent the genetic admixture, it may have promoted intraspecific diversification and/or allopatric speciation (Yap, 2002; de Bruyn and Mather, 2007; Lohman et al., 2011; de Bruyn et al., 2013).

Wallace's Line runs through the Lombok Strait between Bali and Lombok Islands and through the Makassar Strait between Borneo and Sulawesi Islands (Fig. 3). Lombok and Makassar Straits are considered to be deep enough not to allow migration of terrestrial animals across them even during the Quaternary glaciation (Moss and Wilson, 1998; Hall, 2009, 2013; de Bruyn et al., 2014). Thus, most freshwater fish fauna of the Asian origin are not distributed in the east of Wallace's Line, with some exceptions including the cyprinid *Rasbora* and anabantid *Anabas* (Briggs, 1987; Berra, 2001). When and how these fishes migrated across Wallace's Line remain to be an enigma.

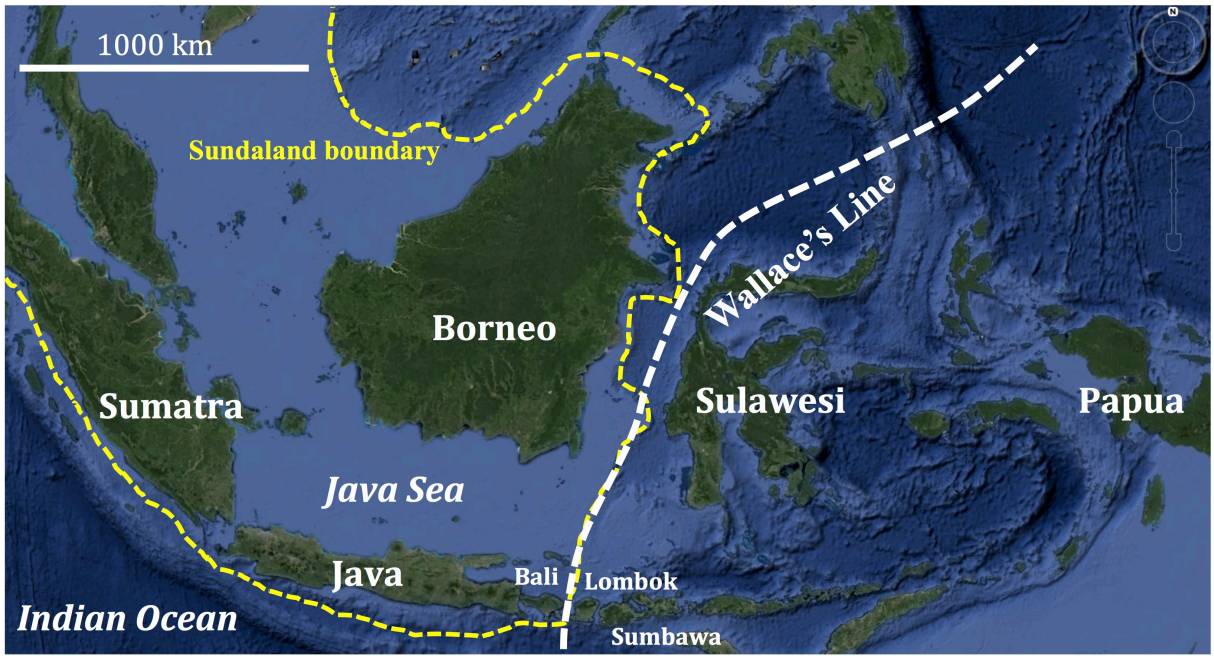

Fig. 3. A map showing Indonesian islands and geographic position of Wallace's Line. The boundary of Sundaland was adopted from Bird et al. (2005).

In this study, I conducted several approaches including molecular phylogenetic analyses using mitochondrial and nuclear gene sequences, coupled with some morphological investigations, to elucidate phylogenetic relationships of *R. lateristriata* and its allies collected from various localities in Java and neighboring islands. I also used these data to resolve taxonomic uncertainties of *R. lateristriata* and *R. baliensis*. Finally, I discuss the historical biogeography, especially on when and how the rasboras diversified and finally crossed the hypothetical barrier, Wallace's Line.

Chapter 2: Material and Methods

2.1 Specimen collection

I conducted field sampling to collect *Rasbora* individuals in many freshwater localities from Sumatra, Borneo, Java, Bali, Lombok and Sumbawa Islands of Indonesia. Fish samples were collected using various non-destructive fishing gears, i.e., fyke net, cast net, hand net and backpack electrofishing (Fig. 4). Geographic coordinates of each sampling site were recorded using a handheld GPS unit. Whenever possible, live specimens were photographed from the left lateral side immediately after the kill (Fig. 5). The collected specimens were tentatively identified based on morphological characters in the field before subsequently reconfirmed in the laboratory. A small portion of the right pectoral fin was excised from fresh individuals in the field and preserved in TNESU8 buffer for molecular studies (Asahida et al., 1996). Whole body specimens were later preserved in 99% ethanol and deposited to the Specimen Depository, Faculty of Fisheries and Marine Science, University of Brawijaya with voucher numbers listed in Table 2. I obtained research permission from Indonesian Institute of Sciences (LIPI) to bring the fin clip samples and the whole body specimens to Japan.

The taxonomic history of *R. lateristriata* is somewhat convoluted. In Buitenzorg (currently known as Bogor) of West Java in 1820-1823, two Netherland scientists Heinrich Kuhl and Johan Coenraad van Hasselt collected several specimens of freshwater fish and named one of them as *Barbus leuciscus* Cuv. *lateristriatus* (van Hasselt, 1823). However, due to inadequate descriptions and lack of references, this name failed to become a formal scientific name and was unavailable at that time (*nomen nudum*) (Roberts, 1993). Afterward, Bleeker (1854) examined several materials from Java and Sumatra, including materials collected by Kuhl and van Hasselt (van Hasselt, 1823), gave adequate descriptions, and named these specimens *Leuciscus lateristriatus*. Later, Bleeker (1860) in his subsequent

Table 2. *Rasbora* samples collected and analyzed for the present study

Locality No.[a]	Species	No. of individuals[b]	Locality	Locality details	Geographic coordinate		Voucher No.[c]
					South (S)	East (E)	
1	*R. lateristriata*	14 (4)	Sukabumi	Cileuleuy River, Sukabumi, West Java	06°46'55.4"	106°49'32.5"	UB.1.143.1-14 (1,2,6,9)
2	species complex	2 (2)	Tegal	Bogares River, Tegal, Central Java	06°59'40.3"	109°10'23.0"	UB.1.142.1-2 (1,2)
3		12 (4)	Sleman	Masanan village, Sleman, Central Java	07°42'23.4"	110°27'44.1"	UB.1.119.1-12 (1,2,3,8)
4		9 (4)	Salatiga	Rowoganjar River, Salatiga, Central Java	07°19'12.6"	110°26'43.5"	UB.1.141.1-7,10,12 (1,2,6,12)
5		11 (4)	Jepara	Kalang River, Jepara, Central Java	06°30'09.9"	110°49'01.6"	UB.1.127.1-10,12 (2,3,6,10)
6		30 (4)	Pasuruan	Umbulan, Pasuruan, East Java	07°45'28.7"	112°56'04.0"	UB.1.117.1-30 (1,7,14,20)
7		4 (4)	Lumajang A	Lake Ranu Klakah, Lumajang, East Java	07°59'06.2"	113°16'07.8"	UB.1.125.1-4 (1-4)
8		21 (4)	Lumajang B	Kedungjajang River, Lumajang, East Java	08°03'52.0"	113°14'47.7"	UB.1.116.1-3,7-13,15-25 (1,2,20,24)
9		15 (4)	Banyuwangi A	Fish market, Banyuwangi, East Java	08°35'48.7"	114°13'23.9"	UB.1.115.1-4,6-9,11,14-18, 20 (9,14,18,20)
10		9 (4)	Banyuwangi B	Bomo River, Banyuwangi, East Java	08°21'12.5"	114°16'54.9"	UB.1.126.25-33 (25,26,27,30)
11		55 (4)	Bratan	Lake Bratan, Tabanan, Bali	08°17'01.7"	115°10'27.4"	UB.1.111.1-55 (1,2,5,7)
12		3 (2)	Buyan	Lake Buyan, Tabanan, Bali	08°15'06.4"	115°07'36.3"	UB.1.112.1-3 (2,3)
13		3 (2)	Penet	Penet River, Tabanan, Bali	08°19'19.8"	115°11'36.3"	UB.1.113.1-3 (2,3)
14		11 (4)	Batur	Lake Batur, Kintamani, Bali	08°15'08.9"	115°24'01.1"	UB.1.114.1-3,6,8-11,18,25,27 (1,3,9,27)
15		13 (4)	Lombok	Sasaot River, Narmada, Lombok	08°32'01.8"	116°14'08.3"	UB.1.118.1-13 (1,2,9,11)
16		12 (4)	Serange	Serange village, Sumbawa, West Nusa Tenggara	-	-	UB.1.139.1-12 (4,5,7,9)
17		12 (4)	Sekokat	Sekokat village, Sumbawa, West Nusa Tenggara	-	-	UB.1.140.1-12 (1,5,7,8)
-	*R. aprotaenia*[d]		Bogor	Katulampa Dam, Bogor, West Java	06°38'00.5"	106°50'13.0"	UB.1.116.20
-	*R. aprotaenia*		Serang	Jidol River, Serang, West Java	06°03'56.8"	106°07'11.6"	UB.1.145.7
-	*R. argyrotaenia*		Bandung	Cijenuk River, Bandung Barat, West Java	06°56'41.1"	107°21'50.1"	UB.1.129.5
-	*R. aurotaenia*		Banjarmasin	Batang River, Banjarmasin, South Kalimantan	02°57'45.6"	114°45'22.2"	UB.1.148.3
-	*R. einthovenii*		Kutai Kartanegara	Downstream of Mahakam River, Kutai Kartanegara, East Kalimantan	00°39'54.1"	117°14'30.8"	UB.1.147.4
-	*R. elegans*		Samarinda	Bayur River, Samarinda, East Kalimantan	-	-	UB.1.144.5
-	*T. gracile*		Riau	Muda Setia village, Pelalawan, Riau	-	-	UB.1.150.1
-	*R. myersi*		Banjarmasin	Downstream of Barito River, Banjarmasin, South Kalimantan	02°57'16.7"	114°44'38.3"	UB.1.149.6
-	*R. tornieri*		Banjarmasin	Downstream of Barito River, Banjarmasin, South Kalimantan	02°57'16.7"	114°44'38.3"	UB.1.146.3

[a] Numbers corespond to locality numbers in Fig. 8

[b] Number of individuals used for morphological analyses (outside parenthesis) and molecular analyses (inside parenthesis)

[c] Voucher number of specimens deposited to Specimen Depository, Faculty of Fisheries and Marine Science, University of Brawijaya. Numbers in parentheses indicate those used for molecular analyses

[d] Specimen used only for mitochondrial genome study

Fig. 4. Different types of fishing gears used in this study: (A) fyke net, (B) cast net, (C) hand net, and (D) backpack electrofishing.

work revised the name of *L. lateristriatus* and replaced it by *Rasbora lateristriata*. Some researchers (e.g., Brittan, 1954; Alfred, 1963) noticed that only specimens from Java can be assigned to the pertinent species (*R. lateristriata*). Given that the holotype and paratype specimens of *R. lateristriata* were unavailable for an unknown reason, Alfred (1963) assigned specimens collected from Batavia (currently known as Jakarta), Ciampea, Bandung and Garut of West Java as the lectotype. I therefore conducted field sampling in several sites of West Java and finally collected this species in a small river in Sukabumi near Bogor, West Java.

For *R. baliensis*, the sampling was conducted in the type locality of this species, in Lake Bratan, an enclosed-crater lake about 1231 m above sea level in the Buyan-Bratan caldera complex, Bali Island (Brittan, 1954). I also conducted field sampling in three additional sites. One of the sites, Lake Batur, was distantly located from other localities. In total, *R. lateristriata*-like specimens were successfully collected from 17 localities from, Java, Bali,

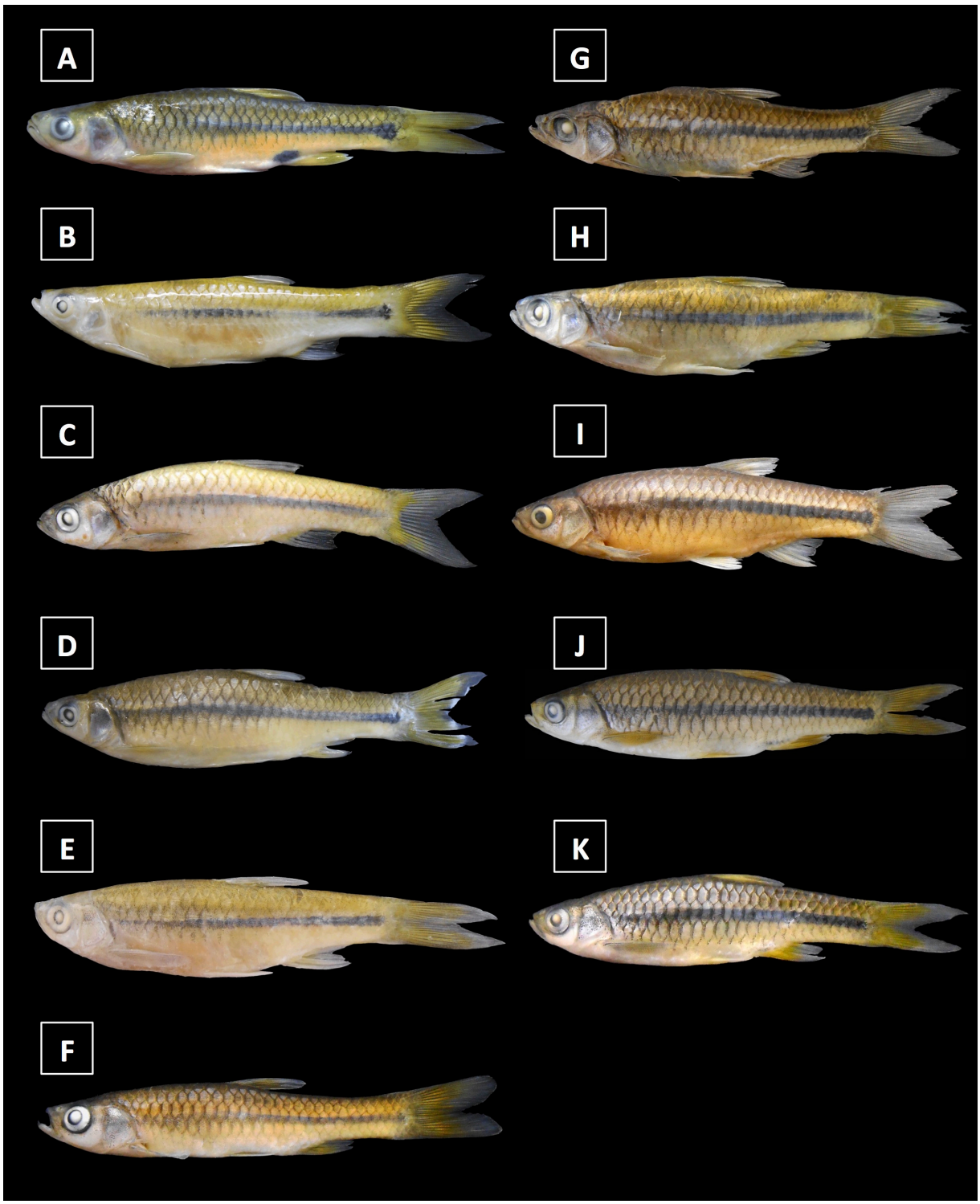

Fig. 5. Left lateral images of *R. lateristriata*-like specimens collected from eleven localities. The museum voucher number for each specimen is (A) UB.1.143.1 (Sukabumi, 75.0 mm SL), (B) UB.1.142.1 (Tegal, 57.1 mm SL), (C) UB.1.119.2 (Sleman, 61.3 mm SL), (D) UB.1.141.4 (Salatiga, 46.1 mm SL), (E) UB.1.127.9 (Jepara, 45.2 mm SL), (F) UB.1.117.8 (Pasuruan, 66.1 mm SL) (G) UB.1.125.3 (Lumajang, 57.5 mm SL), (H) UB.1.115.20 (Banyuwangi, 57.6 mm SL), (I) UB.1.111.13 (Bratan, 73.6 mm SL), (J) UB.1.118.7 (Lombok, 54.8 mm SL) and (K) UB.1.139.4 (Serange, 59.9 mm SL).

Lombok and Sumbawa Islands. *R. aprotaenia* was collected similarly and I tried to collect this species in its type locality. Brittan (1954) described *R. aprotaenia* using seven individuals collected from a type locality that was mentioned to be in 'Tjilowaeng River' or 'Ciliwung River', West Java. I therefore conducted field sampling in this river and finally collected *R. aprotaenia* in Katulampa Dam of Ciliwung River (Table 2).

I noticed from morphological appearance, *R. lateristriata*-like fish from Java and Bali Islands were very similar to each other. Thus, I regarded them as the *R. lateristriata* species complex as discussed later in more details.

2.2 DNA amplification and sequencing

Genomic DNA was extracted from the fin samples following a procedure described by Asahida et al. (1996). For protein digestion, 20 µl of Protaeinase K (20mg/ml) was added to a 1.5 ml tube containing the preserved tissue of right pectoral fin in TNESU8 buffer. The mixture was then incubated at 37°C for 15-20 hours (or at room temperature for several days). After incubation, the mixture was extracted with an equal volume of phenol-chloroform (1:1). After the extraction, cold ethanol was added in the mixture for DNA precipitation. Finally, precipitated DNA was dissolved in TE-buffer. DNA samples were then stored in a freezer at -30°C until use.

Two mitochondrial genes coding for cytochrome oxidase subunit I (COI, 655 bp) and cytochrome *b* (Cyt*b*, 1091 bp), as well as two nuclear genes coding for recombination activating gene I (RAG1, 1557 bp) and opsin (a protein part of rhodopsin) (866 bp) in a total 4169 bp were amplified by polymerase chain reaction (PCR) using sets of primers listed in Table 3. These four genes have been frequently used for elucidating both interspecific and intraspecific phylogenetic analyses in fishes, especially by Tang et al. (2010) who conducted phylogenetic analysis using many species from *Rasbora*.

The PCR was performed in a 10 μl reaction mixture with a SpeedSTAR HS DNA polymerase (Takara) according to the manufacture's protocol. The PCR was done by 30 cycles of 98°C for 5 s, 55°C for 15 s and 72°C for 20 s. PCR products were treated with an ExoSAP-IT (Affymetrix) and directly sequenced in both directions using a Big Dye terminator v3.1 cycle sequence kit (Life Technologies) on the ABI 3500 DNA sequencer. Obtained sequences for both directions were edited and assembled with Sequencher 4.8 (Gene Codes). Because there were very few indels in the determined sequences, sequences were manually aligned by eye using MacClade 4.08 (Sinauer Associates). Possible heterozygotic sites in the RAG1 and opsin genes were treated following IUPAC ambiguity codes.

Table 3. Sets of primers for PCR amplification and/or sequencing

Region	Name	Sequence (5' to 3')	Source
COI	FishF1	TCAACCAACCACAAAGACATTGGCAC	Ward et al. (2005)
	FishR1	TAGACTTCTGGGTGGCCAAAGAATCA	Ward et al. (2005)
Cyt*b*	LA-cyp	ATGGCAAGCCTACGAAAAAC	Tang et al. (2010)
	HA-cyp	TCGGATTACAAGACCGATGCTT	Tang et al. (2010)
RAG1	RAG1F1	CTGAGCTGCAGTCAGTACCATAAGATGT	Lopez et al. (2004)
	RAG1R1	CTGAGTCCTTGTGAGCTTCCATRAAYTT	Lopez et al. (2004)
	Ras_RAG1F1	GCATCAGGCTCCACTTAC	this study
	Ras_RAG1R1	ATAGCGCTCGAGATTTTCC	this study
Opsin	Rh 28F	TACGTGCCTATGTCCAAYGC	Chen et al. (2008)
	Rh 1039R	TGCTTGTTCATGCAGATGTAGA	Chen et al. (2003)

In order to determine complete mitochondrial genome sequences, two individuals for *R. lateristriata* (voucher number: UB.1.116.20) and *R. aprotaenia* (UB.1.120.3) were randomly selected. Mitochondrial genome (mitogenome) sequences of both species were completely sequenced using the long PCR amplification and subsequent amplification, sequencing and assembly of shorter (650-950 bp) DNA regions, starting from the long PCR product as a template (Miya and Nishida, 1999; Inoue et al., 2001). The PCR and sequencing were

conducted as described above. Obtained sequences for both directions were edited and assembled with Sequencher 4.8 (Gene Codes). Gene characterization and annotation in determined mitogenome sequences were done using DOGMA (Wyman et al., 2004) followed by manual inspection.

2.3 Phylogenetic analyses

2.3.1 Analysis of mitogenomic sequences

Mitogenomic sequences for species other than *R. lateristriata* and *R. aprotaenia* were downloaded from DDBJ/EMBL/GenBank. Eight mitochondrial genome sequences from genus *Rasbora* have been deposited including *Rasbora borapetensis* (accession number AB924546), *Rasbora cephalotaenia* (AP011430), *Rasbora daniconius* (AP011285), *Rasbora steineri* (JX843769), *Rasbora trilineata* (KM200714), *Rasboroides vaterifloris* (AP011432), *Trigonostigma heteromorpha* (AP011421), *Trigonostigma espei* (AP011449). *Danio rerio* (AC024175) and *Acheilognathus typus* (AB239602) were selected as outgroup taxa. Getmitogenome (Jonniaux, 2014) was used to retrieve the sequences from the database. This software was also used to determine the boundaries between 37 genes (13 proteins, 22 tRNAs and two rRNAs) by aligning gene sequences of the above-mentioned taxa with the currently available alignment for other taxa (unpublished data). Maximum likelihood analysis was conducted by using 3757 amino acid sites of 13 mitochondrial protein genes to determine the phylogenetic position of *R. lateristriata* and *R. aprotaenia* among other *Rasbora* species. Garli v2.0 (Zwickl, 2014) was used to conduct the analyses under the mtREV+IG4 model. The nodal support was assessed by 500 non-parametric bootstrap resamplings.

2.3.2 Interspecies phylogenetic analyses

Phylogenetic analyses to determine the phylogenetic position of the *R. lateristriata* species complex among other *Rasbora* species were conducted using one randomly selected

individual representing each locality or each species. DNA sequences for many rasboras that were not sampled during my fieldwork, including two outgroup taxa of *Chromobotia macracanthus* and *Catostomus commersonii* were downloaded from DDBJ/EMBL/GenBank. These sequences were mainly reported by Tang et al. (2010). The list of species and their voucher numbers are shown in Table 4. DNA sequences from mitochondrial COI and Cyt*b*, as well as nuclear RAG1 and opsin or their combination, were used to construct phylogenetic trees using Bayesian and maximum likelihood (ML) methods. The dataset to infer the phylogenetic position of the *R. lateristriata* species complex using only one representative individual per location is named dataset 1. I conducted phylogenetic analyses using five different combinations of genes in dataset 1, i.e., concatenated four genes and individual genes of COI, Cyt*b*, RAG1 and opsin. The best partition schemes and its evolutionary models for first, second and third codon positions of genes in each dataset were estimated using PartitionFinder (Lanfear et al., 2012).

I conducted Bayesian analyses using MrBayes v3.12 (Ronquist and Huelsenbeck, 2003). Due to the limitation of evolutionary models available in this software, I used the most complex evolutionary model of GTR+I+G with 4 gamma categories for all partitions. Starting from randomly generated trees, the Markov chain Monte Carlo (MCMC) process was initially set at 2,000,000 generations and continued until the Average Standard Deviation of Split Frequency became less than 0.01. Two independent runs with four simultaneous MCMC chains at temperature 0.20 were conducted by default. The first 25% generations were discarded as "burnin" after the likelihood score reached the stationarity. Trees were sampled every 100 generations and a 50% majority consensus tree with Bayesian posterior (Bayes-P) probabilities at nodes was constructed based on trees from the remaining generations. ML analyses were conducted using GARLI v2.0 (Zwickl, 2014) using the estimated partition schemes and models. Five independent runs were conducted using the default search setting (5,000,000 generations) from a randomly generated initial tree. The statistical support at each node was assessed by 1000 non-parametric bootstrap resamplings.

Table 4. Accession numbers of COI, Cyt*b*, RAG1 and opsin gene sequences used in dataset 1

Species	Locality	DDBJ/EMBL/GenBank Accession Number			
		COI	Cyt*b*	RAG1	opsin
Rasbora lateristriata	Sukabumi	LC130642	LC131148	LC130784	LC130916
species complex	Tegal	LC130651	LC131157	LC130793	LC130920
	Sleman	LC130656	LC131162	LC130798	LC130925
	Salatiga	LC130665	LC131171	LC130807	LC130926
	Jepara	LC130673	LC131179	LC130814	LC130930
	Pasuruan	LC130681	LC131187	LC130821	LC130934
	Lumajang A	LC130694	LC131200	LC130833	LC130938
	Lumajang B	LC130698	LC131204	LC130837	LC130942
	Banyuwangi A	LC130709	LC131215	LC130848	LC130949
	Banyuwangi B	LC130710	LC131216	LC130849	LC130950
	Bratan	LC130717	LC131223	LC130856	LC130954
	Buyan	LC130727	LC131233	LC130866	LC130958
	Penet	LC130730	LC131236	LC130869	LC130960
	Batur	LC130732	LC131238	LC130871	LC130962
	Lombok	LC130743	LC131249	LC130881	LC130966
	Serange	LC130756	LC131262	LC130894	LC130972
	Sekokat	LC130765	LC131271	LC130901	LC130974
Rasbora aprotaenia	Serang	LC130776	LC131282	LC130908	LC130978
Rasbora argyrotaenia	Bandung	LC130777	LC131283	LC130909	LC130979
Rasbora aurotaenia	Banjarmasin	LC130778	LC131284	LC130910	LC130980
Rasbora einthovenii	Kutai Kartanegara	LC130779	LC131285	LC130911	LC130981
Rasbora elegans	Samarinda	LC130780	LC131286	LC130912	LC130982
Rasbora myersi	Banjarmasin	LC130782	LC131288	LC130914	LC130984
Rasbora tornieri	Banjarmasin	LC130783	LC131289	LC130915	LC130985
Trigonopoma gracile	Riau	LC130781	LC131287	LC130913	LC130983
Rasbora bankanensis	-	HM224220	HM224337	EU292709	FJ531357
Rasbora borapetensis	-	HM224222	HM224342	HM224100	HM223985
Rasbora caudimaculata	-	EF452870	HM224339	-	-
Rasbora cephalotaenia	-	AP011430	AP011430	HM224099	HM223984

Table 4. (Continued)

Species	Locality	DDBJ/EMBL/GenBank Accession Number			
		COI	Cyt*b*	RAG1	opsin
Rasbora daniconius	Koshi Barrage, Nepal	EF452872	HM224345	HM224103	HM223988
Rasbora dusonensis	Surat Thani, Thailand	HM224225	HM224348	HM224105	HM223990
Rasbora hobelmani	Phayao, Thailand	HM224229	HM224354	HM224110	HM223994
Rasbora jacobsoni	-	HM224230	HM224355	HM224111	HM223995
Rasbora kalbarensis	-	FJ753502	EF151116	FJ753538	-
Rasbora kalochroma	-	HM224231	HM224356	HM224112	HM223996
Rasbora kottelati	Brunei Darussalam	HM224232	HM224357	HM224113	HM223997
Rasbora meinkeni	-	HM224233	-	HM224115	HM223998
Rasbora paviana	Chiang Mai, Thailand	HM224223	HM224344	HM224101	HM223986
Rasbora rasbora	-	HM224238	HM224364	HM224121	HM224003
Rasbora steineri	-	HM224241	HM224368	EU409631	EU409662
Rasbora sumatrana	-	EF452882	HM224369	EF452837	EF452908
Rasbora trilineata	-	EF452883	HM224370	HM224124	HM224006
Rasbora tubbi	Brunei Darussalam	HM224242	HM224371	HM224125	HM224007
Rasbora vulgaris	Phang-nga, Thailand	HM224243	HM224373	HM224126	HM224008
Boraras merah	-	EF452884	HM224358	EF452838	EF452909
Catostomus commersonii	Illinois, USA	AB127394	AB127394	EU409612	FJ197032
Chromobotia macracanthus	-	AB242163	AB242163	EU711137	FJ197037
Kottelatia brittani	-	EF452869	HM224338	HM224098	HM223983
Rasbosoma spilocerca	Mekong River, Thailand	HM224240	HM224367	HM224123	HM224005
Trigonopoma pauciperforatum	-	HM224236	HM224362	HM224119	HM224001
Trigonostigma hengeli	-	HM224228	HM224353	HM224109	HM223993

Table 5. Accession numbers of COI, Cyt*b*, RAG1 and opsin gene sequences used in dataset 2

Species	Locality	DDBJ/EMBL/GenBank Accession Number			
		COI	Cyt*b*	RAG1	opsin
Rasbora lateristriata	Sukabumi	LC130642-645	LC131148-151	LC130784-787	LC130916-919
species complex	Tegal	LC130651-652	LC131157-158	LC130793-794	LC130920-921
	Sleman	LC130653-656	LC131159-162	LC130795-798	LC130922-925
	Salatiga	LC130665-668	LC131171-174	LC130807-810	LC130926-929
	Jepara	LC130673-676	LC131179-182	LC130814-817	LC130930-933
	Pasuruan	LC130681-684	LC131187-190	LC130821-824	LC130934-937
	Lumajang A	LC130694-697	LC131200-203	LC130833-836	LC130938-941
	Lumajang B	LC130698-701	LC131204-207	LC130837-840	LC130942-945
	Banyuwangi A	LC130706-709	LC131212-215	LC130845-848	LC130946-949
	Banyuwangi B	LC130710-713	LC131216-219	LC130849-852	LC130950-953
	Bratan	LC130717-720	LC131223-226	LC130856-859	LC130954-957
	Buyan	LC130727-728	LC131233-234	LC130866-867	LC130958-959
	Penet	LC130730-731	LC131236-237	LC130869-870	LC130960-961
	Batur	LC130732-735	LC131238-241	LC130871-874	LC130962-965
	Lombok	LC130743-746	LC131249-252	LC130881-884	LC130966-969
	Serange	LC130754-757	LC131260-263	LC130892-895	LC130970-973
	Sekokat	LC130765-768	LC131271-274	LC130901-904	LC130974-977
Rasbora aprotaenia	Serang	LC130776	LC131282	LC130908	LC130978
Rasbora elegans	Samarinda	LC130780	LC131286	LC130912	LC130982
Rasbora hobelmani	Phayao, Thailand	HM224229	HM224354	HM224110	HM223994
Rasbora paviana	Chiang Mai, Thailand	HM224223	HM224344	HM224101	HM223986
Rasbora sumatrana	-	EF452882	HM224369	EF452837	EF452908
Rasbora vulgaris	Phang-nga, Thailand	HM224243	HM224373	HM224126	HM224008
*Barbatula barbatula**	-	-	-	EU711107	FJ650476
*Danio rerio**	-	-	-	U71093	L11014
*Hypentelium nigricans**	Maryland, USA	-	-	EU711134	FJ197033
*Micronemacheilus pulcher**	-	-	-	EU409611	EU409637
*Mylopharyngodon piceus**	-	-	-	GU217831	GU218587
*Notemigonus crysoleucas**	Ontario, Canada	-	-	EF452831	FJ197062
*Rasboroides vaterifloris**	-	-	-	HM224127	HM224009
*Semotilus atromaculatus**	Alabama, USA	-	-	EU409629	EU409658
*Squaliobarbus curriculus**	Guangxi, China	-	-	HM224069	HM223951
*Tanakia lanceolata**	-	-	-	KF417865	KF429468

* species other than genus *Rasbora* used only for divergence time estimation

2.3.3 Analyses within the R. lateristriata species complex

Phylogenetic reconstructions for elucidating relationships within the *R. lateristriata* species complex were done using maximum likelihood and Bayesian methods. Methods for conducting phylogenetic analyses were the same as explained in the section 2.3.2. I conducted phylogenetic analyses using dataset 2 that included randomly selected 2 to 4 individuals per locality (mostly 4 individuals). Three different combinations of genes were used in the analyses: concatenation of four genes (COI+Cyt*b*+RAG1+opsin), concatenated mitochondrial genes (COI+Cyt*b*) and concatenated nuclear genes (RAG1+opsin). Accession numbers of gene sequences used in the dataset 2 are shown in Table 5.

2.4 Genetic divergence estimation

Using COI gene sequences of all available individuals of each locality (ranging from 2 to 13 individuals), molecular divergence was estimated. A standardized threshold of 2% sequence divergence as suggested by Ward (2009) and Ratnasingham and Hebert (2013) was used as a reference for the species delimitation. Kimura's 2-parameter (K2P) model implemented in MEGA v6.06 (Tamura et al., 2013) was used to calculate the pairwise divergences.

2.5 Divergence time estimation

Divergence times among major lineages within the *R. lateristriata* species complex were estimated using the relaxed-clock Bayesian method implemented in BEAST v1.8.2 (Drummond et al., 2012). The XML input file was generated using BEAUti v1.8.2 (Drummond et al., 2012). The dataset using RAG1 and opsin gene sequences was created by randomly selecting one haplotype representing each major lineage. I applied the Uncorrelated Lognormal Clock model (Drummond et al., 2006) with no *a priori* correlation of evolutionary rates between a lineage's rate and that of its ancestor and the Yule tree prior assuming a

constant speciation rate per lineage. jModelTest v2.1.5 (Darriba et al., 2012) was used to select the best substitution model of each partition. SYM+I+G and GTR+I+G were selected for RAG1 and opsin genes, respectively. A gamma-distributed substitution rate with 4 categories was selected with the base frequency estimated from the data. A user-specified starting tree was set as the ML tree topology resulting from the four genes analysis. Two independent MCMC processes for two hundred million generations were performed and trees were sampled every 1000 generations. Tracer v1.6.0 (Rambaut et al., 2015) was used to confirm more than 200 effective sample sizes for parameters and the convergence of two independent runs after the first 10% samples were removed as 'burnin'. LogCombiner v1.8.2 was used to combine sampled trees. A single ultrametric tree with a median posterior divergence time estimate and 95% highest posterior density intervals (95% HPD) was created using TreeAnnotator v1.8.2.

Seven calibration points were used as priors for the divergence time estimation. The calibration points were based on Betancur-R et al. (2013) that used 18 genes (17 nuclear and 1 mitochondrial genes) to estimate divergence times between major lineages of bony fishes, including cypriniform fishes. I referred to estimated divergence times from this work at 7 nodes which are within or close to Cyprinidae. Because time estimates by Betancur-R et al. (2013) were point estimates without confidence intervals, I used these values as means of the prior time distribution at the corresponding nodes and arbitrarily set 20% of the means as standard deviations of the normal distributions. I also conducted the dating analyses using the first and second codon positions of concatenated mitochondrial gene sequences (COI and Cyt*b*). Parameters for performing this analysis in BEAST were set to be the same as in nuclear gene sequence analysis.

2.6 Historical biogeography reconstruction and haplotype network analysis

Historical biogeography inference of the *R. lateristriata* species complex was done using RASP v3.2 (Yu et al., 2015) under the Lagrange (Dispersal-Extinction-Cladogenesis, DEC) model (Ree et al., 2005; Ree and Smith, 2008). A condensed time-calibrated phylogenetic tree produced by BEAST v1.8.2 (Drummond et al., 2012) for divergence time estimation was used after removing most outgroup taxa as an input tree to reconstruct ancestral geographic distributions of the *R. lateristriata* species complex. To assign current distributions of the *R. lateristriata* species complex in Java Island, I divided Java into three regions: West, Central and East Java. The assignment of Java Island into three different regions is in accordance with the paleogeographical history of this island from Late Miocene to Early Pliocene (Hall, 2009, 2013).

To elucidate relationships among COI haplotype sequences of the *R. lateristriata* species complex, an unrooted haplotype network was reconstructed using NETWORK v5.0.0.0 (http://www.fluxusengineering.com). I applied a median-joining algorithm (Bandelt et al., 1999) with the default settings.

2.7 Morphological analyses

Morphological analyses were conducted using all collected specimens of the *R. lateristriata* species complex. I also examined museum specimens labeled as *R. lateristriata* and deposited at the Museum Zoologicum Bogoriense (MZB), Cibinong, Indonesia. Methods for measuring morphometric characters and counting meristic characters basically followed those of Brittan (1954) and Lumbantobing (2014). Briefly, standard length (SL) was measured from the anterior tip of the mouth to the end of the hypural plate. Dorsal-hypural distance (DHD) was measured as a distance from the origin of the dorsal fin to the end of hypural plate. Head length (HL) is a distance taken from the anterior tip of the mouth to the posterior edge of the opercle. A lateral line scale is a series of pored scales along the lateral

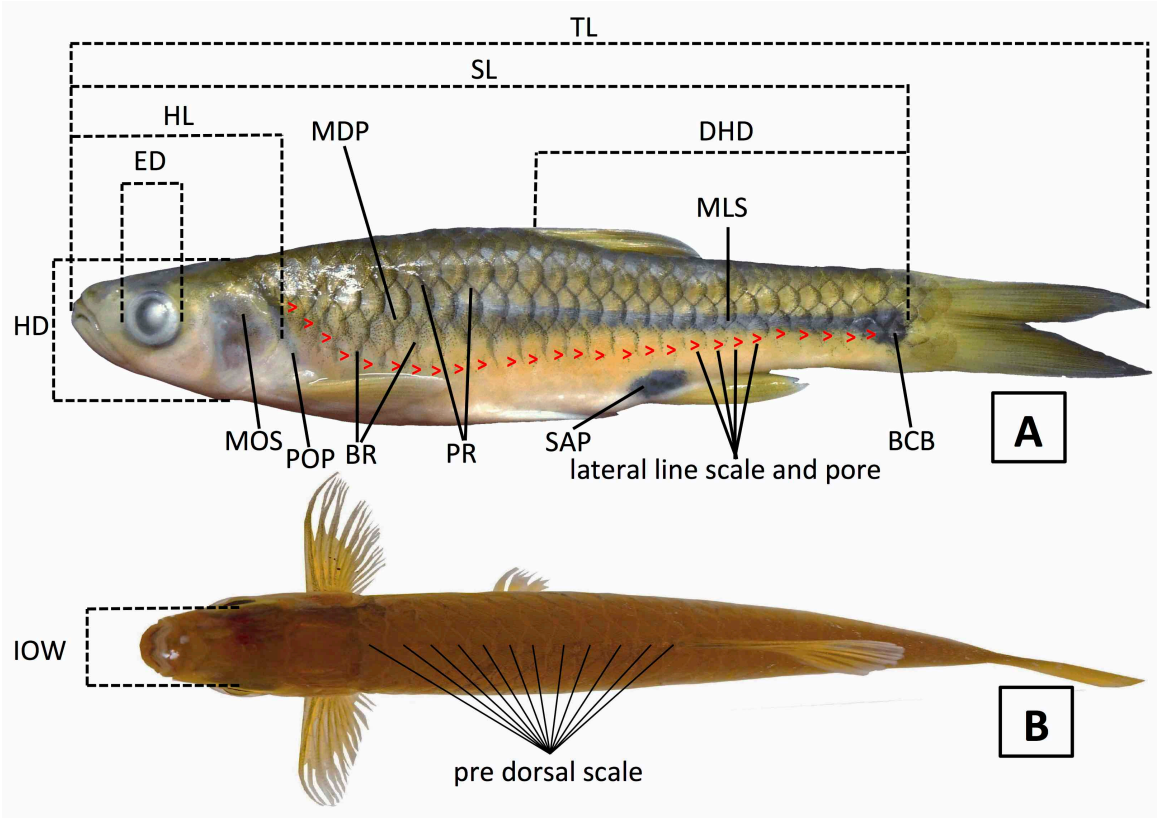

Fig. 6. Morphological characters analyzed in this study. (A) A left lateral and (B) a dorsal view of *R. lateristriata* from Sukabumi. Morphometric and meristic characters are derived from Brittan (1954) and Lumbantobing (2014). The terminology of the body color pattern follows Brittan (1954) with some additional features from Lumbantobing (2010, 2014). The abbreviation of each character is: BCB = Basicaudal Blotch; BR = Basal Reticulation; DHD = Dorsal Hypural Distance; ED = Eye Diameter; HD = Head Depth; HL = Head Length; IOW = Inter-Orbital Width; MDP = Midhumeral Diffuse Patch; MLS = Midlateral Stripe; MOS = Midopercular Stripe; POP = Postopercular Pigmentation; PR = Peripheral Reticulation; SAP = Supra Anal Pigment; SL = Standard Length and TL: Total Length.

line starting just behind the upper end of the gill opening to the base of the caudal fin. Finally, the pre-dorsal scale is a series of scales along the midline starting from the origin of the dorsal fin forward to the skull.

I followed Brittan (1954) for the terminology of the body color pattern and Lumbantobing (2010, 2014) for some additional pigmentation features, including peripheral reticulation (PR), basal reticulation (BR), basicaudal blotch (BCB), midhumeral diffuse patch (MDP), midopercular stripe (MOS) and postopercular pigmentation (POP). Supra anal pigment (SAP) is defined as melanophore pigmentation located above the base of the anal fin.

BCB is the melanophores spot observable at the base of the hypural plate. MDP is characterized by melanophore pigmentation starting from the gill opening to the dorsal fin origin in the midlateral region. In total, I observed 38 meristic or morphometric characters featuring the body color pattern in this study. More details on the morphological examination used in this study are shown in Fig. 6.

Chapter 3: Results

3.1 Collection of specimens

I collected *Rasbora* samples from 47 freshwater localities (Fig. 7). Two hundred thirty six individuals were identified as *R. lateristriata* or *R. baliensis* and the others were identified as one of *R. aprotaenia*, *R. argyrotaenia, R. aurotaenia*, *R. einthovenii, R. elegans, R. myersi*, *R. tornieri,* and *T. gracile*. Among the 236 *R. lateristriata* or *R. baliensis* individuals, 127, 72, 13 and 24 individuals were collected from Javanese, Balinese, Lombok and Sumbawa localities, respectively (Table 2 and Fig. 8). Among them, 24 individuals (10%) were young with SL less than 35 mm. *R. lateristriata*-like samples were not collected from Sumatran and Bornean localities where I conducted field samplings (Fig. 7). The number of collected individuals varied from locality to locality, ranging from 2 to 55 individuals (Table 2).

A single *Rasbora* species usually occurred at each locality, with several exceptions (Fig. 7). *R. tornieri* and *R. myersi* were co-distributed in Jambi while *R. aurotaenia* and *R. myersi* occurred in Palembang of Sumatra. In Borneo, *R. tornieri*, *R. myersi* and *R. argyrotaenia* sympatrically occurred in a location of Banjarmasin. In addition, *R. elegans* and *R. argyrotaenia* co-existed in Samarinda and *R. einthovenii* and *R. argyrotaenia* occurred in Kutai Kartanegara. All but one Javanese localities had a single *Rasbora* species. Only Tegal of Central Java had both *R. lateristriata* and *R. argyrotaenia* together. Because rasboras were scarce at Lake Buyan, Penet River, Bomo River, Lake Ranu Klakah, Rowoganjar River and Bogares River, only less than 10 individuals were caught at these localities (Table 2).

3.2 Phylogenetic positions of *R. lateristriata* and *R. aprotaenia* revealed using mitogenome sequences

The lengths of mitochondrial genomes for *R. lateristriata* and *R. aprotaenia* which I sequenced were 16,539 bp (DDBJ/EMBL/GenBank accession No. LC021505) and 16,541 bp

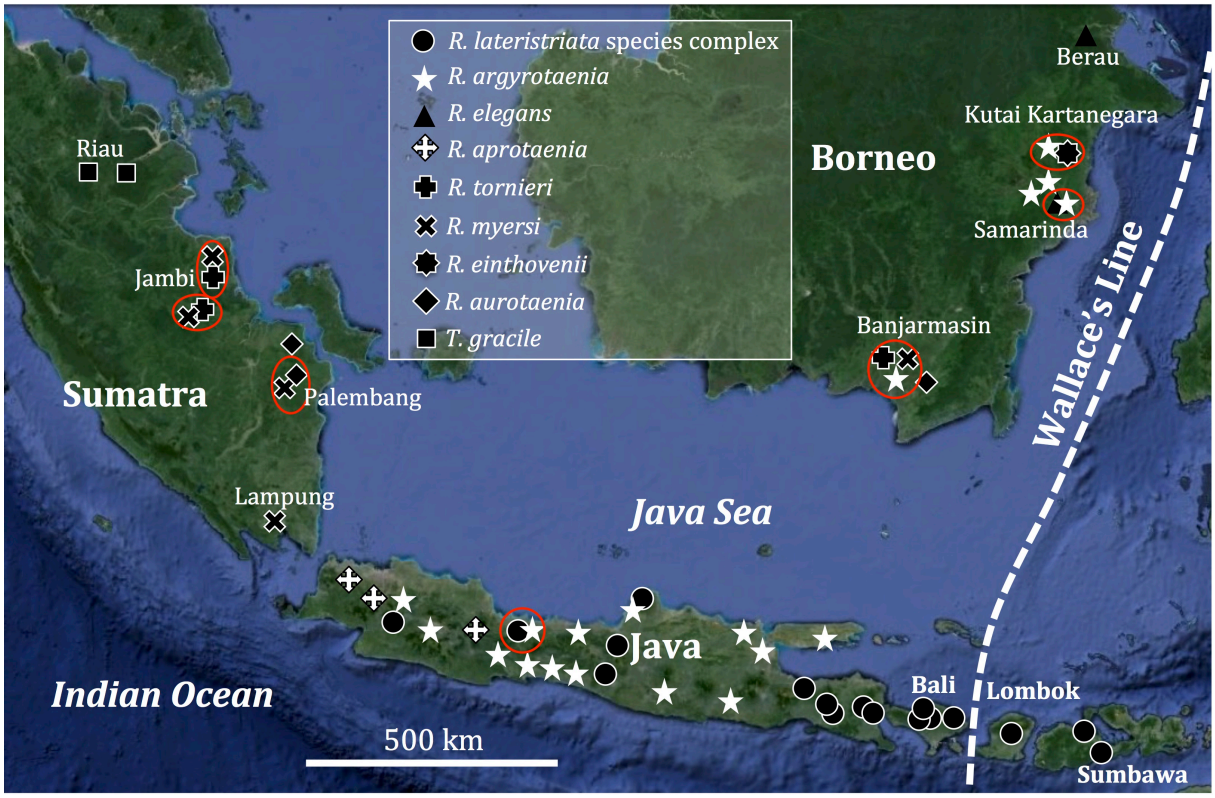

Fig. 7. Sampling localities for rasboras. All 47 sampling locations for *Rasbora*, ranging from Sumatra, Borneo, Java, Bali, to Lombok and Sumbawa Islands in the eastern side of Wallace's Line are shown with one or more symbols that correspond to identified *Rasbora* species. Species identification is based primarily on morphological features complemented by molecular information. Locations with sympatric distribution of multiple *Rasbora* species are highlighted with a red circle. The map is produced based on a satellite image from Google Earth v7.1.5.1557.

Fig. 8. Seventeen sampling localities for the *R. lateristriata* species complex. Locality of *R. aprotaenia* for mitochondrial genome study is highlighted with a white circle, collected from its type locality. Locality names and numbers correspond to those in Table 2. The map is produced based on a satellite image from Google Earth v7.1.5.1557.

Table 6. Features of the mitochondrial genome of *R. lateristriata*

Features	Code	Start	Stop	Size (bp)	Spacer (+) or Overlap (-)	Strand	Start Codon	Stop Codon[1]
tRNA-Phe	F	1	69	69	0	F		
12S rRNA		70	1022	953	0	F		
tRNA-Val	V	1023	1093	71	0	F		
16S rRNA		1094	2774	1681	0	F		
tRNA-Leu (UUR)	L	2775	2848	74	0	F		
ND1		2850	3824	975	1	F	ATG	TAA
tRNA-Ile	I	3829	3900	72	4	F		
tRNA-Gln	Q	3899	3969	71	-2	R		
tRNA-Met	M	3971	4039	69	1	F		
ND2		4040	5084	1045	0	F	ATG	T - -
tRNA-Trp	W	5085	5157	73	0	F		
tRNA-Ala	A	5161	5228	68	3	R		
tRNA-Asn	N	5231	5303	73	2	R		
L-strand Origin		5304	5340	37	0			
tRNA-Cys	C	5338	5405	68	-3	R		
tRNA-Tyr	Y	5407	5476	70	1	R		
CO1		5478	7028	1551	1	F	GTG	TAA
tRNA-Ser (UCN)	S	7029	7099	71	0	R		
tRNA-Asp	D	7101	7170	70	1	F		
CO2		7176	7866	691	5	F	ATG	T - -
tRNA-Lys	K	7867	7941	75	0	F		
ATPase 8		7944	8108	165	2	F	ATG	TAA
ATPase 6		8102	8781	680	-7	F	ATG	TA -
CO3		8782	9566	785	0	F	ATG	TA -
tRNA-Gly	G	9567	9637	71	0	F		
ND3		9638	9986	349	0	F	ATG	T - -
tRNA-Arg	R	9987	10056	70	0	F		
ND4L		10057	10353	297	0	F	ATG	TAA
ND4		10347	11728	1382	-7	F	ATG	TA -
tRNA-His	H	11729	11797	69	0	F		
tRNA-Ser (AGY)	S	11805	11865	61	7	F		
tRNA-Leu (CUN)	L	11868	11940	73	2	F		
ND5		11941	13770	1830	0	F	ATG	TAA
ND6		13767	14288	522	-4	R	ATG	TAA
tRNA-Glu	E	14289	14357	69	0	R		
Cytb		14367	15507	1141	0	F	ATG	T - -
tRNA-Thr	T	15508	15578	71	0	F		
tRNA-Pro	P	15588	15657	70	9	R		
Control region		15685	16539	855	0			
CSB-2		16309	16325	17				
CSB-3		16350	16368	19				

[1]Hyphens indicate an incomplete stop and imply subsequent addition of A residues to the 3' end of the mRNA by polyadenilation

(LC021504), respectively. Features of the mitochondrial genome of *R. lateristriata* and *R. aprotaenia* are shown in Tables 6 and 7, respectively. Both mitogenomes encode 37 genes for 13 proteins, 22 tRNAs, and 2 rRNAs with a major noncoding region in the typical vertebrate gene arrangement (Anderson et al. 1981). All protein genes start with an ATG initiation codon, except for COI gene which uses GTG as a start codon. Seven protein genes have a stop codon in the mitogenome sequences, whereas the remaining six protein genes appear to

Table 7. Features of the mitochondrial genome of *R. aprotaenia*

Features	Code	Start	Stop	Size (bp)	Spacer (+) or Overlap (-)	Strand	Start Codon	Stop Codon[1]
tRNA-Phe	F	1	69	69	0	F		
12S rRNA		70	1021	952	0	F		
tRNA-Val	V	1022	1092	71	0	F		
16S rRNA		1093	2774	1682	0	F		
tRNA-Leu (UUR)	L	2775	2848	74	0	F		
ND1		2850	3824	975	1	F	ATG	TAA
tRNA-Ile	I	3829	3900	72	4	F		
tRNA-Gln	Q	3899	3969	71	-2	R		
tRNA-Met	M	3971	4039	69	1	F		
ND2		4040	5084	1045	0	F	ATG	T - -
tRNA-Trp	W	5085	5157	73	0	F		
tRNA-Ala	A	5161	5228	68	3	R		
tRNA-Asn	N	5231	5303	73	2	R		
L-strand Origin		5304	5340	37				
tRNA-Cys	C	5338	5405	68	-3	R		
tRNA-Tyr	Y	5406	5475	70	0	R		
CO1		5477	7027	1551	1	F	GTG	TAA
tRNA-Ser (UCN)	S	7028	7098	71	0	R		
tRNA-Asp	D	7100	7169	70	1	F		
CO2		7175	7865	691	5	F	ATG	T - -
tRNA-Lys	K	7866	7940	75	0	F		
ATPase 8		7943	8107	165	2	F	ATG	TAA
ATPase 6		8101	8780	680	-7	F	ATG	TA -
CO3		8781	9565	785	0	F	ATG	TA -
tRNA-Gly	G	9566	9636	71	0	F		
ND3		9637	9985	349	0	F	ATG	T - -
tRNA-Arg	R	9986	10055	70	0	F		
ND4L		10056	10352	297	0	F	ATG	TAA
ND4		10346	11727	1382	-7	F	ATG	TA -
tRNA-His	H	11728	11796	69	0	F		
tRNA-Ser (AGY)	S	11804	11864	61	7	F		
tRNA-Leu (CUN)	L	11867	11940	74	2	F		
ND5		11941	13770	1830	0	F	ATG	TAA
ND6		13767	14288	522	-4	R	ATG	TAG
tRNA-Glu	E	14289	14357	69	0	R		
Cytb		14367	15507	1141	0	F	ATG	T - -
tRNA-Thr	T	15508	15578	71	0	F		
tRNA-Pro	P	15588	15657	70	9	R		
Control region		15685	16541	857	0			
CSB-2		16311	16327	17				
CSB-3		16352	16370	19				

[1]Hyphens indicate an incomplete stop and imply subsequent addition of A residues to the 3' end of the mRNA by polyadenilation

have a mechanism in which their stop codons are posttranscriptionally created by polyadenylation. All tRNA genes can be folded into the standard cloverleaf secondary structures for mitochondrial tRNAs (Kumazawa and Nishida 1993). Phylogenetic analyses (Fig. 9) showed, with a 100% bootstrap support, that *R. lateristriata* is more closely related to *R. aprotaenia* than to any other *Rasbora* species examined, pointing to their phylogenetic closeness. These species then clustered with *R. steineri* with a 100% bootstrap support.

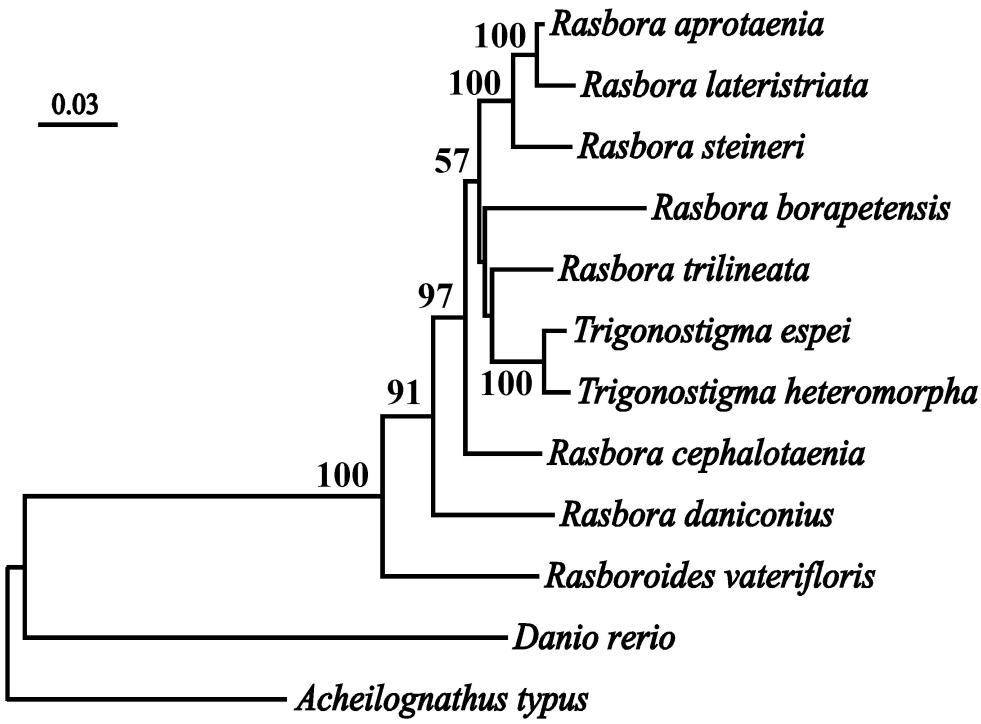

Fig. 9. Maximum likelihood tree constructed using 3757 amino acid sites of 13 mitochondrial protein genes. *R. lateristriata* and *R. aprotaenia* are shown to be closely related to each other pointing to their phylogenetic affinity. Bootstrap probabilities by 500 replications are shown for each node when they are 50% or larger.

3.3 Phylogenetic position of the *R. lateristriata* species complex using multilocus gene sequences

COI, Cyt*b*, RAG1 and opsin gene sequences determined from a representative individual of each locality for the *R. lateristriata* species complex and 8 other *Rasbora* species were aligned with those downloaded from the database as the dataset 1 to conduct ML and Bayesian analyses. The four gene sequences had different sizes, ranging from 655 bp (COI) to 1557 bp (RAG1). As expected, mitochondrial genes had more variable and informative sites per determined base than nuclear genes but the latter genes still provided a number of variable and informative sites (Table 8).

Figure 10 shows an ML tree constructed using four concatenated gene sequences of the dataset 1. Phylogenetic relationships between various *Rasbora* species were largely in

Table 8. Some characteristics of molecular data used in this study

Gene	N[a]	Sites			Scheme of partition[c]	
		Alignable	Variable	Informative[b]	Number of partition	Subset partitions and the best substitution models (in parentheses)
Dataset 1 (one individual per locality)						
COI	51	655	253	216	3	1st (TrNef+G); 2nd (HKY+I+G); 3rd (TrN+G)
Cyt*b*	50	1091	533	444	3	1st (SYM+I+G); 2nd (HKY+I+G); 3rd (TIM+I+G)
RAG1	50	1557	479	280	3	1st (TVMef+I+G); 2nd (F81+IG); 3rd (SYM+G)
opsin	49	866	240	142	2	1st+2nd (TrNef+I+G); 3rd (TVM+G)
COI+Cyt*b*+RAG1+opsin	51	4169	1506	1083	9	COI_3rd (TrN+G); COI_1st (TrN+I+G); COI_2nd+Cyt*b*_2n+opsin_2nd (TIM+I+G); Cyt*b*_3rd (TIM+I+G); Cyt*b*_1st (SYM+I+G); RAG1_1st+RAG1_2nd (TVMef+I+G); RAG1_3rd (SYM+G); opsin_3rd (TVM+G); opsin_1st (TrNef+I+G)
Dataset 2 (multiple individuals per locality)						
COI+Cyt*b*	68	1746	264	206	4	COI_3rd (TrN); COI_1st+Cyt*b*_1st (K80+I); COI_2nd+Cyt*b*_2nd (F81+I); Cyt*b*_3rd (TrN)
RAG1+opsin	68	2423	109	78	3	RAG1_1st+RAG1_2nd+opsin_1st+opsin_2nd (K80+I); RAG1_3rd (K80+I+G); opsin_3rd (TVM+G)
COI+Cyt*b*+RAG1+opsin	68	4169	374	289	5	COI_3rd+Cyt*b*_3rd (TrN+G); COI_1st+ RAG1_1st+RAG1_2nd+opsin_1st (K80+I); COI_2nd+Cyt*b*_2nd+opsin_2nd (HKY+I); Cyt*b*_1st+RAG1_3rd (K80+I); opsin_3rd (TVM+G)

[a] Number of individuals including outgroups

[b] Parsimony-informative sites

[c] The best partition scheme and substitution models were estimated by PartitionFinder.

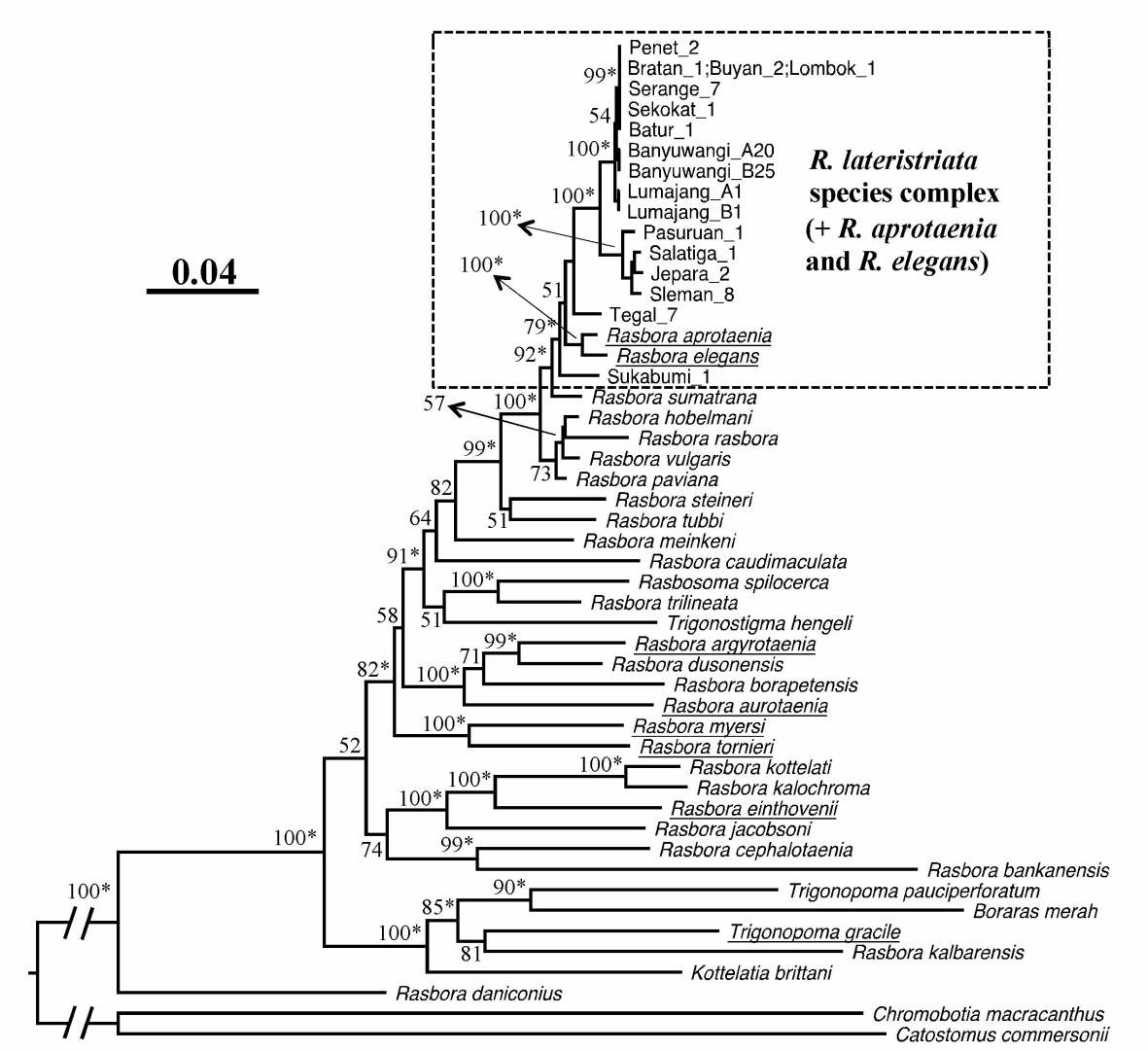

Fig. 10. A maximum likelihood tree among *Rasbora* species constructed using COI, Cyt*b*, RAG1 and opsin gene sequences. Values at nodes show bootstrap probabilities (>50 % only) and an asterisk shows that the corresponding node received a Bayes-P probability of 1.00. Refer to Table 4 for used taxa and individuals with accession numbers. Individuals of the *R. lateristriata* species complex are shown with their locality name and the number of individuals from that locality (e.g., Bratan_1). The *Rasbora lateristriata* species complex boxed with dotted lines is defined in text. As explain in more details in section 4.1, I propose to regard the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* as the *R. lateristriata*-group. In underlined taxa other than the *R. lateristriata* species complex, the corresponding specimen was collected, identified and sequenced by me.

agreement with Tang et al. (2010) who used neither of *R. lateristriata* nor *R. baliensis* in the molecular analyses. This ML tree indicates that individuals of the *R. lateristriata* species complex together with those of *R. aprotaenia* and *R. elegans* make a monophyletic group with relatively high support values (79% bootstrap and 1.00 Bayes-P probabilities). I will later

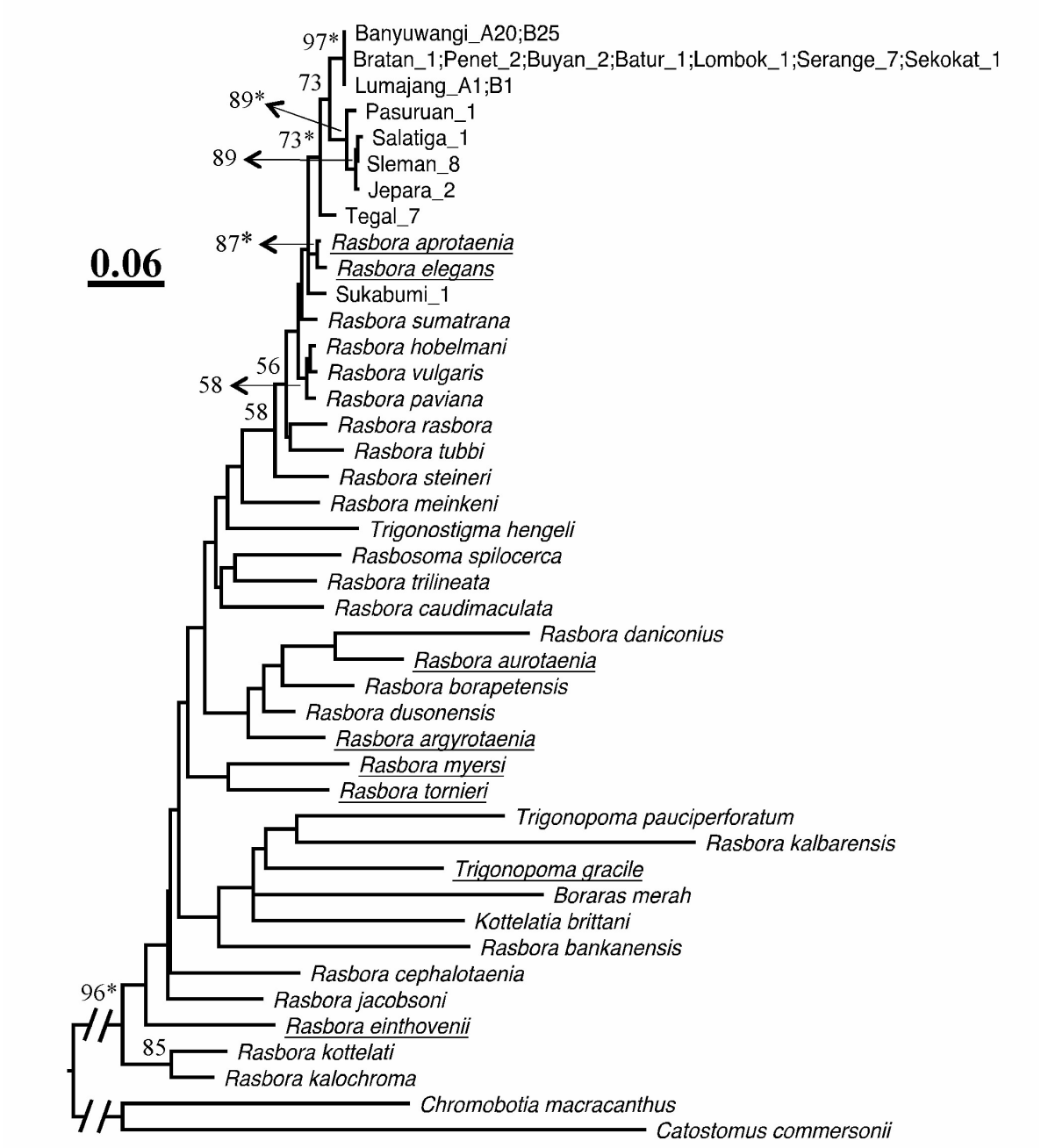

Fig. 11. A maximum likelihood tree among *Rasbora* species constructed using COI gene sequences. For other details, see the legend of Fig. 10.

name this clade the *R. lateristriata*-group (see Chapter 4). This clade has a sister-group relationship with *R. sumatrana* with strong bootstrap (92%) and Bayes-P (1.00) probabilities. In contrast, *R. argyrotaenia*, another commonly occurring species in Java Island, turned out to be distantly related to the *R. lateristriata* species complex (Fig. 10).

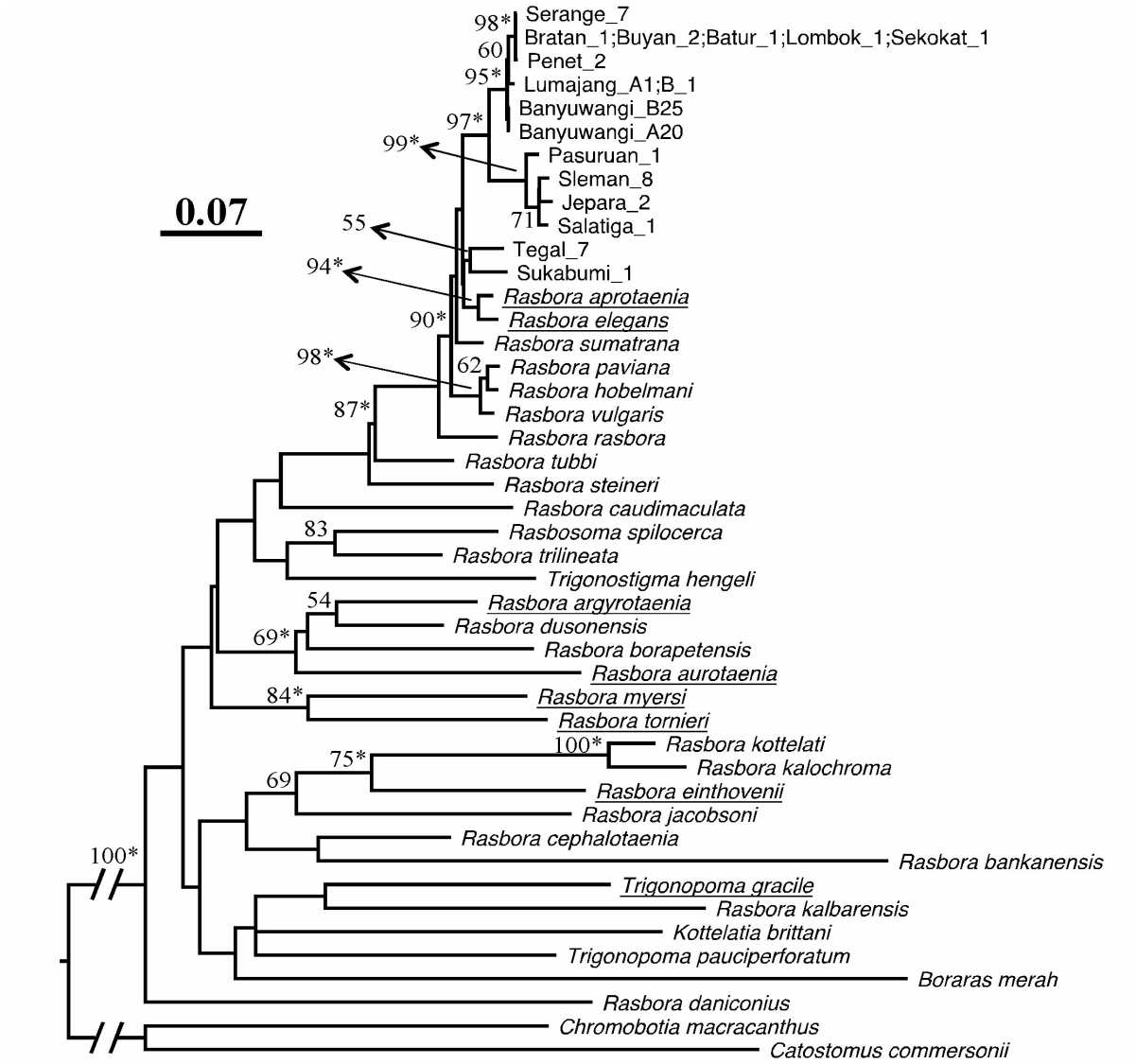

Fig. 12. A maximum likelihood tree among *Rasbora* species constructed using Cyt*b* gene sequences. For other details, see the legend of Fig. 10.

I also conducted ML and Bayesian analyses based on each of the COI, Cyt*b*, RAG1 and opsin gene sequences (Figs. 11-14). The monophyly of the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* and its sister-group relationship to *R. sumatrana* were commonly seen for two ML trees based on COI and Cyt*b* genes (Figs. 11-12). Although ML trees based on RAG1 and opsin genes did not necessarily support these conclusions, these trees lacked high resolution in general without strong bootstrap and Bayes-P supports at nodes (Figs. 13-14). Thus, I judged that no strongly competing phylogenetic information with respect to interspecific relationships exists among the four genes.

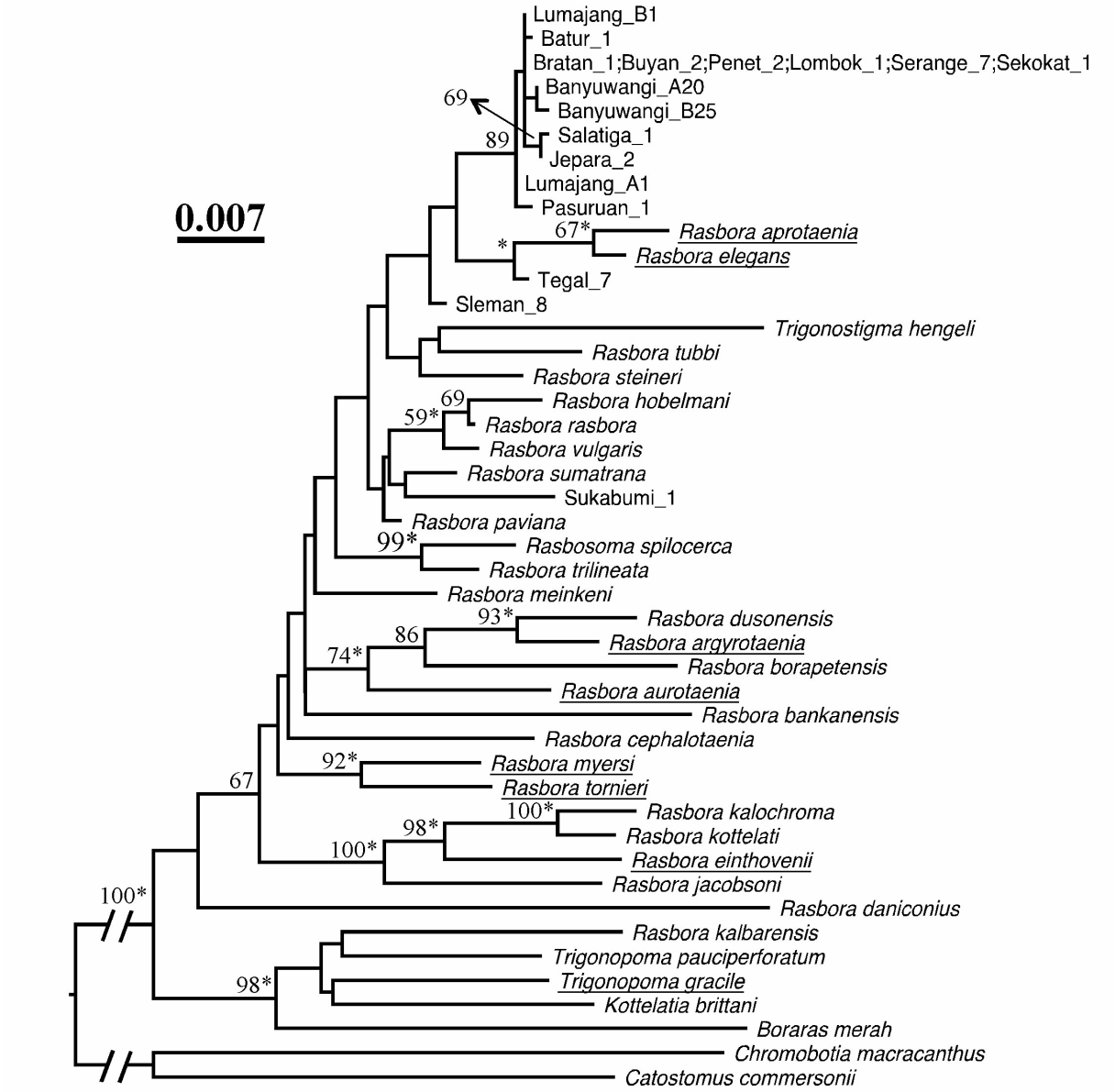

Fig. 13. A maximum likelihood tree among *Rasbora* species constructed using RAG1 gene sequences. For other details, see the legend of Fig. 10.

3.4 Phylogenetic relationships within the *R. lateristriata* species complex

In order to elucidate phylogenetic relationships within the *R. lateristriata* species complex, I conducted phylogenetic analyses using the dataset 2 including more (usually 4) numbers of individuals from each locality but less numbers of outgroup taxa than in the dataset 1. Figure 15 shows an ML tree constructed using four gene sequences. Four major clades were identified all with strong bootstrap (99 or 100%) and Bayes-P (1.00) probabilities.

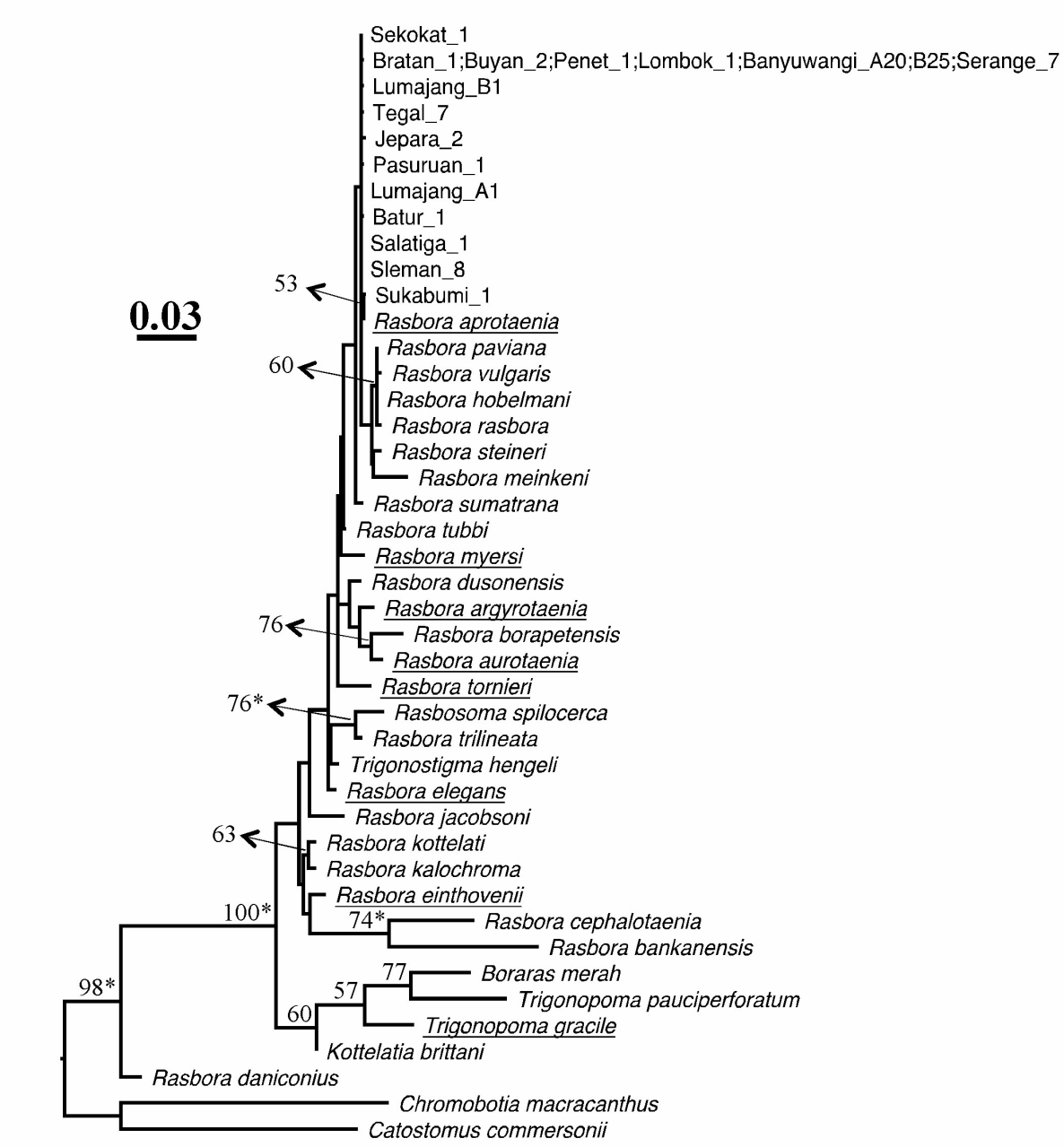

Fig. 14. A maximum likelihood tree among *Rasbora* species constructed using opsin gene sequences. For other details, see the legend of Fig. 10.

The first clade (Clade 1) is composed of individuals from eastern parts of Java, Bali, Lombok and Sumbawa. The second clade (Clade 2) consists of individuals from central parts of Java (Pasuruan, Sleman, Salatiga and Jepara) while the third clade (Clade 3) consists of individuals from a single Javanese locality (Tegal). Individuals from a West Javanese locality (Sukabumi) form the fourth clade (Clade 4). Among these major clades, Clade 1 and Clade 2

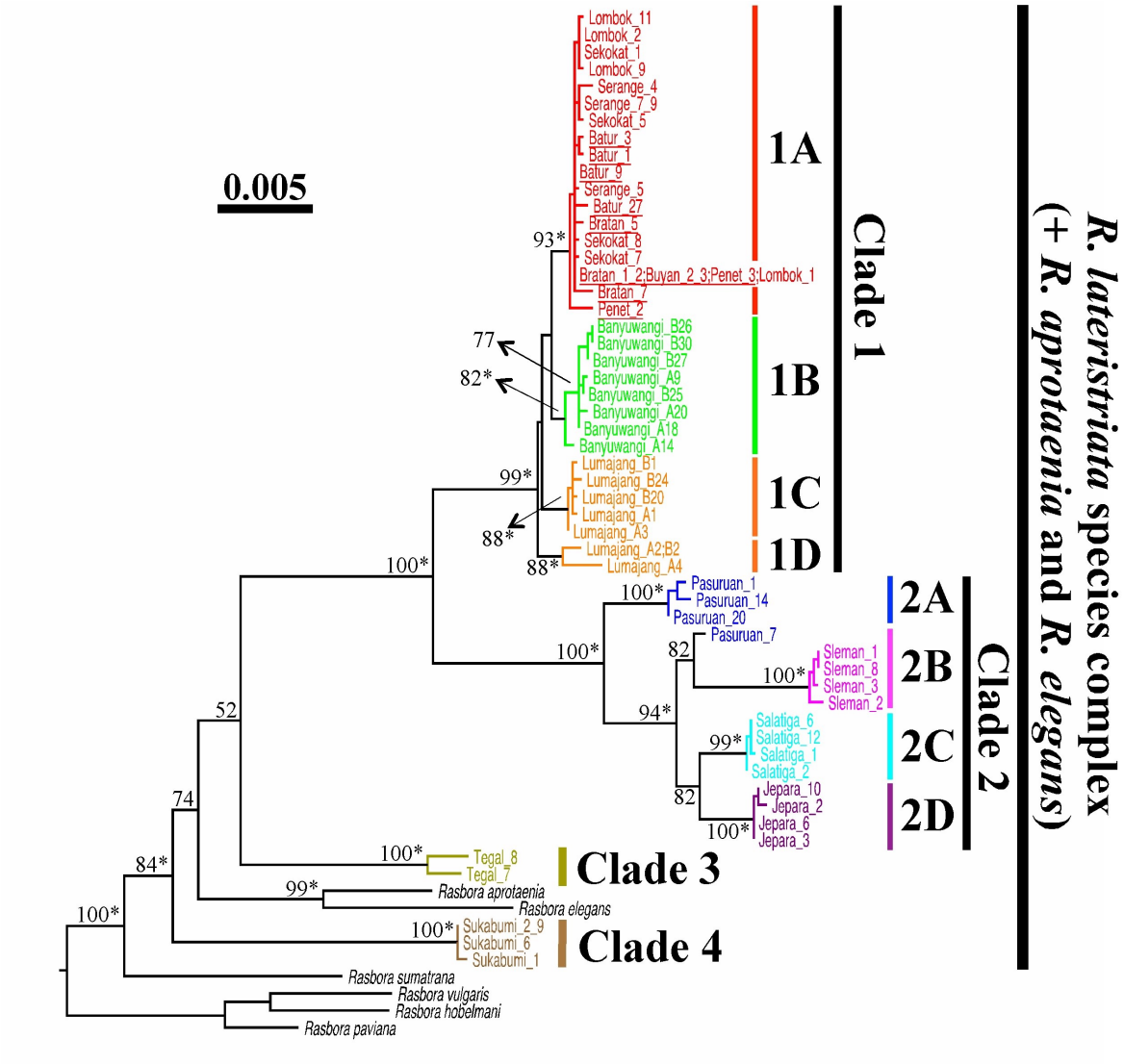

Fig. 15. A maximum likelihood tree of the *R. lateristriata* species complex with multiple individuals from each locality. The tree was constructed with nucleotide sequences of 4 genes (COI, Cyt*b*, RAG1 and opsin). Values at nodes show bootstrap probabilities (> 50% only) and an asterisk shows that the corresponding node received a Bayes-P probability of 1.00. Refer to Table 5 for used taxa and individuals with accession numbers. Individuals of the *R. lateristriata* species complex are shown with their locality name and the number of individuals from that locality (e.g., Bratan_1). Taxa from Balinese localities are highlighted with an underline. Different colors match the locality information in Fig. 8.

are more closely related to each other than they are to Clade 3 or Clade 4. The clustering of Clade 1 and Clade 2 accompanied strong support values (100% bootstrap and 1.00 Bayes-P probabilities). On the other hand, relationships among Clade 1+2, Clade 3, Clade 4 and the *R. aprotaenia-R. elegans* clade were supported with low or only moderately high bootstrap values and thus may not be fully resolved in this study.

Within Clade 1, four subclades were recognized (Fig. 15). Subclade 1A consists of individuals from Bali, Lombok and Sumbawa Islands, and subclades 1B and 1C are composed of individuals from Banyuwangi and Lumajang, respectively. Monophyly of these subclades was supported with moderate to high bootstrap probabilities (82-93%). However, three individuals from Lumajang (Lumajang_A2, Lumajang_A4 and Lumajang_B2) did not cluster with the other Lumajang individuals. They basally diverged from all other Clade 1 individuals and formed subclade 1D. Within Clade 2, four subclades were recognized. Subclades 2A, 2B, 2C, and 2D are composed of individuals from Pasuruan, Sleman, Salatiga and Jepara, respectively. One individual from Pasuruan (Pasuruan_7) appeared within subclade 2B, rendering Pasuruan individuals to be non-monophyletic. Otherwise, all the subclades 1B, 1C, 2A, 2C and 2D comprise individuals from a distinct small geographical area. In other words, lineage sorting has likely operated to make haplotypes in these areas distinct from each other. In contrast, individuals in subclade 1A are not clustered based on their geographical origin, such as Bali, Lombok and Sumbawa Islands. Thus, the lineage sorting has not operated enough to segregate haplotypes in these islands although Lombok and Sumbawa Islands are located in the eastern side of Wallace's Line.

With respect to phylogenetic relationships between the subclades, subclades 1A and 1B are more closely related to each other than they are to subclade 1C (Fig. 15). However, this relationship does not accompany high bootstrap and Bayes-P probabilities. Within Clade 2, subclades 2C and 2D make a monophyletic group, to which subclade 2B clusters. The monophyly of subclades 2B-2D in relation to subclade 2A is supported with high bootstrap (94%) and Bayes-P (1.00) probabilities.

I also conducted phylogenetic analyses based on each of the four genes of the dataset 2 and concatenated mitochondrial and nuclear genes. Figures 16 and 17 show ML trees constructed using mitochondrial (COI+Cyt*b*) and nuclear genes (RAG1+opsin), respectively. These two ML trees and the ML tree based on 4 genes (Fig. 15) commonly supported

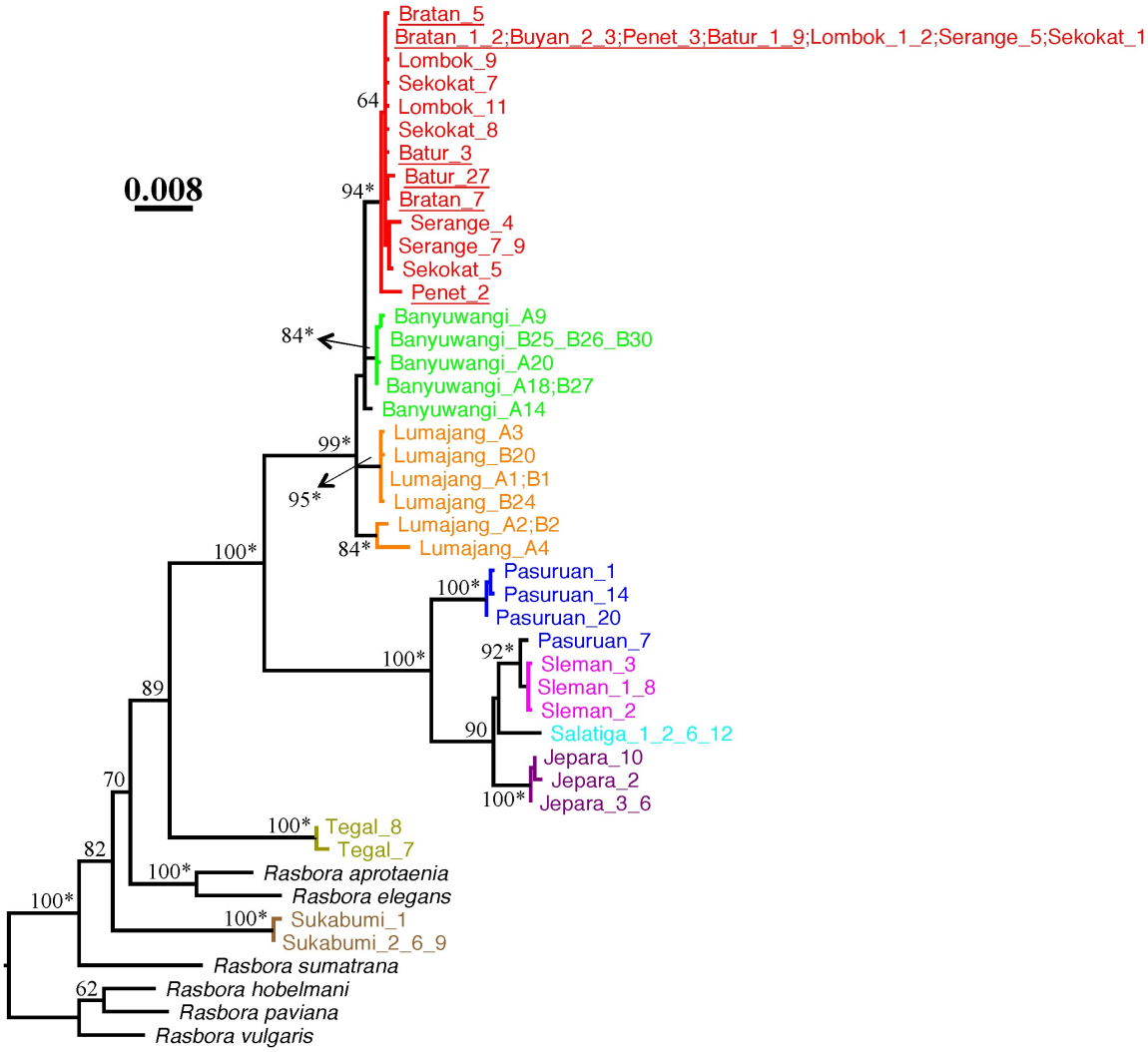

Fig. 16. A maximum likelihood tree of the *Rasbora lateristriata* species complex constructed using concatenated COI and Cyt*b* gene sequences. For other details, see the legend of Fig. 15.

a sister-group relationship of the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* to *R. sumatrana* and the clustering of *R. aprotaenia* and *R. elegans*. The monophyly of individuals in each of Sukabumi, Jepara and Sleman was also commonly seen. Whereas the ML trees based on the four genes and two mitochondrial genes were similar to each other in many topological relationships (e.g., recognition of Clades 1-4 and basal divergence of Clade 4 followed by the divergence of the *R. aprotaenia*+*R. elegans* clade and Clade 3), the ML tree based on two nuclear genes largely lacked the resolution on these relationships, as seen by generally low bootstrap and Bayes-P probabilities especially in relationships within Clades 1 and 2 (Fig. 17). However, there was one noticeable difference in the placement

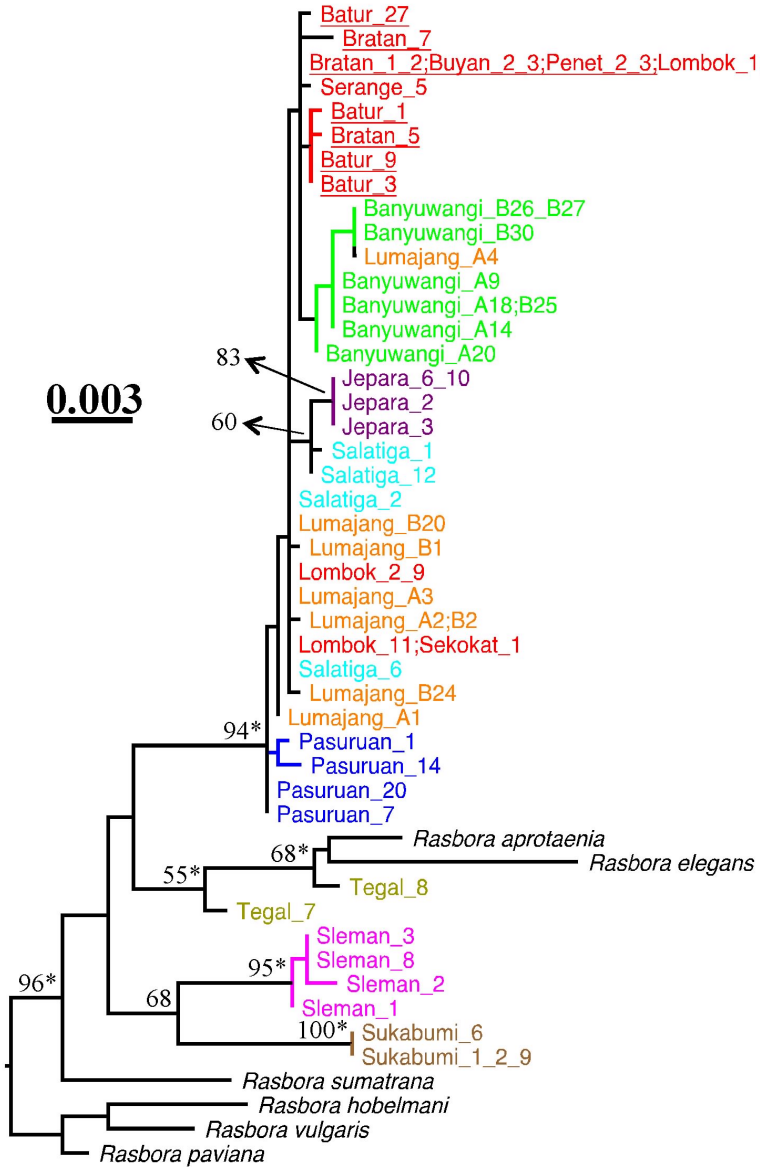

Fig. 17. A maximum likelihood tree of the *Rasbora lateristriata* species complex constructed using concatenated RAG1 and opsin gene sequences. For other details, see the legend of Fig. 15.

of Sleman individuals. In the 4-gene and mitochondrial ML trees, they are closely related to individuals from other Central Javanese localities (i.e., Pasuruan, Salatiga and Jepara) in Clade 2. On the other hand, they cluster with Sukabumi individuals outside Clade 2 in the nuclear ML tree. However, this new relationship is not supported with strong bootstrap and Bayes-P probabilities and I judged that there is no strongly competing phylogenetic information between mitochondrial and nuclear genes.

3.5 Molecular divergences between clades

I then estimated molecular divergences between Clades 1-4 using the COI barcoding region sequences. For this purpose, the COI gene was sequenced from all available individuals from each locality. I first confirmed that 4 major clades recognized in Fig. 15 also appeared in an ML tree based on all the available COI gene sequences (Fig. 18). The average pairwise K2P distances were then calculated between these clades to show that all of them exceed 2% (an empirical distance at the boundary of closely related species) (Ward, 2009; Ratnasingham and Hebert, 2013), supporting that Clades 1-4 represent different species (Table 9). On the other hand, the average pairwise K2P distances between subclades of Clade 1 and Clade 2 were much less than 2%, suggesting that these subclades represent intraspecific lineages (data not shown). Voucher numbers for all individuals and accession numbers of determined sequences are shown in Table 10.

3.6 Estimation of divergence times

To the best of my knowledge, prior to this study, only a few attempts have been made for divergence time estimation of lineages within the genus *Rasbora*. Studies conducted by Rüber et al. (2007) and Britz et al. (2009) estimated divergence times within Cyprinidae and used several species of rasboras in their analysis. However, both studies only used a single calibration point with the oldest known fossil of the Cyprinidae. More recently, Betancur-R et

Table 9. Pairwise divergences between major clades of the *R. lateristriat*a species complex

	N	Clade 1	Clade 2	Clade 3	Clade 4
Clade 1	82		0.007	0.007	0.007
Clade 2	41	0.035		0.007	0.008
Clade 3	2	0.031	0.038		0.006
Clade 4	9	0.035	0.050	0.026	

Distances were calculated using Kimura's 2-parameter distances for COI gene sequences (655 bp). Values above diagonal indicate standard deviation.

Table 10. Individuals in the *R. lateristriata* species complex used for estimating molecular divergences

Clade	Locality	No. of specimens	Voucher number	Accesion Number
				COI
Clade 1	Lumajang A	4	UB.1.125.1-4	LC130694-697
	Lumajang B	8	UB.1.116.1,2,9,11,19,20,24,25	LC130698-705
	Banyuwangi A	4	UB.1.115.9,14,18,20	LC130706-709
	Banyuwangi B	7	UB.1.126.25-27,29,30,32,34	LC130710-716
	Bratan	10	UB.1.111.1,2,5-7,10,13,27-29	LC130717-726
	Buyan	2	UB.1.112.2,3	LC130727-728
	Penet	3	UB.1.113.1-3	LC130729-731
	Batur	11	UB.1.114.1-3,6,8-11,18,25,27	LC130732-742
	Lombok	11	UB.1.118.1-11	LC130743-753
	Serange	11	UB.1.139.1-11	LC130754-764
	Sekokat	11	UB.1.140.1-11	LC130765-775
Clade 2	Sleman	12	UB.1.119.1-12	LC130653-664
	Salatiga	8	UB.1.141.1-4,6,7,10,12	LC130665-672
	Jepara	8	UB.1.127.2-8,10	LC130673-680
	Pasuruan	13	UB.1.117.1,2,7,9,11,14,18,20,24,26,28-30	LC130681-693
Clade 3	Tegal	2	UB.1.142.1-2	LC130651-652
Clade 4	Sukabumi	9	UB.1.143.1-7,9,10	LC130642-650

al. (2013) and Broughton et al. (2013) conducted the molecular dating for lineages within Cyprinidae. However, none of these studies used rasboras in their analyses. I thus decided to use seven calibration points reported by Betancur-R et al. (2013), which seemed to provide reliable prior dates in nodes close to *Rasbora*.

Figure 19 shows divergences times estimated between clades and subclades in the *R. lateristriata* species complex using RAG1 and opsin gene sequences. The divergence of the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* from *R. sumatrana* was estimated to be at 8.6 Mya in median (5.8-11.9 Mya in 95% HPD). This timing in mean corresponds to the late Miocene. The basal divergence between Clade 4 and the remaining taxa in the *R. lateristriata* species complex was estimated to be around 6.8 (4.6-9.4) Mya. The separation of the *R. aprotaenia*+*R. elegans* lineage from Clade 1+Clade 2+Clade 3 occurred at 5.4 (3.6-7.6) Mya of the Miocene-Pliocene boundary. The splits between Clade 3 and Clade 1+Clade 2 and

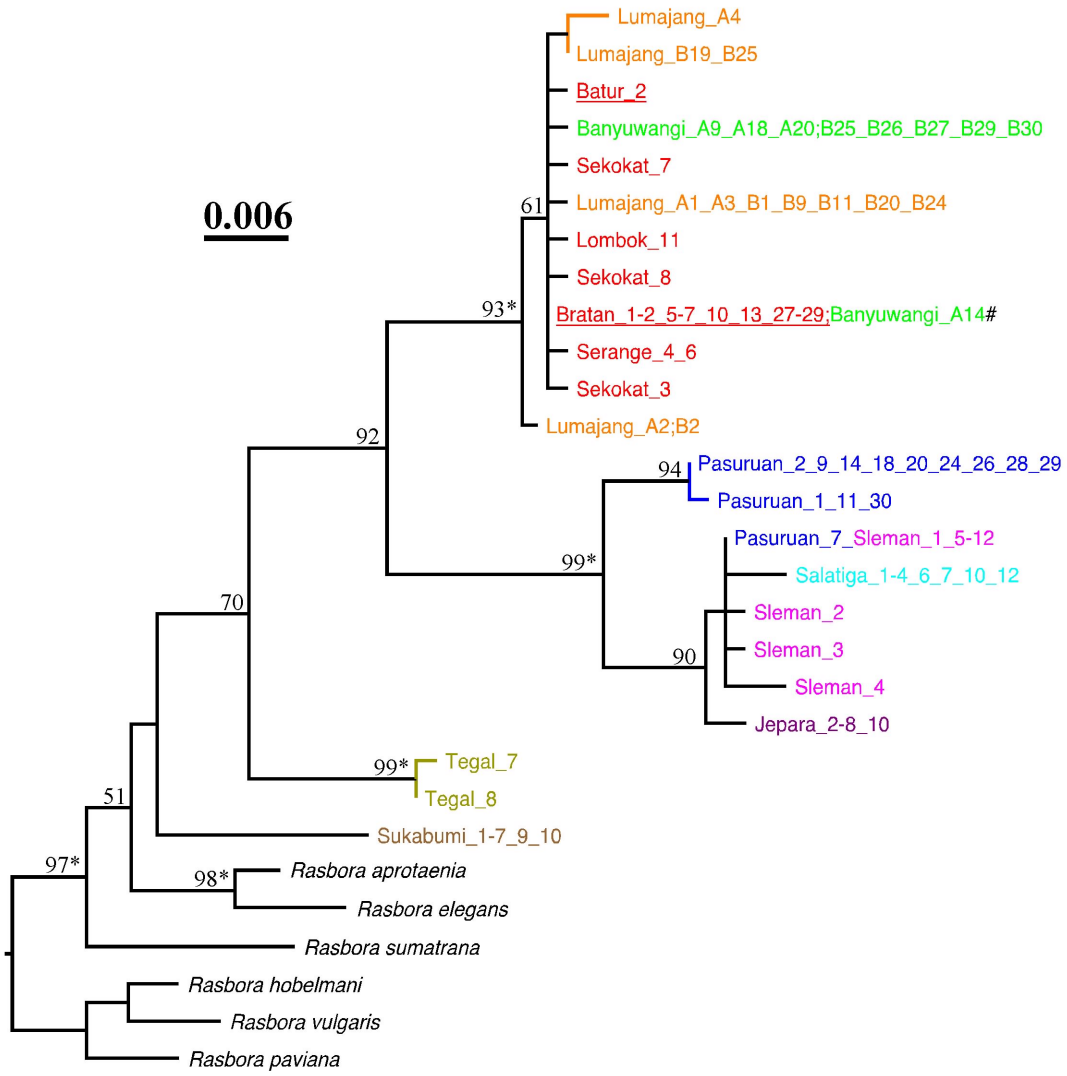

Fig. 18. A maximum likelihood tree of the *Rasbora lateristriata* species complex constructed using COI gene sequences. Values at nodes show bootstrap probabilities (> 50 % only) and an asterisk shows that the corresponding node received a Bayesian posterior probability of 1.00. For other details, see the legend of Fig. 15. Refer to Table 10 for the number of individuals used from each locality and their voucher numbers. Symbol of "#" indicates haplotypes identical with the following: Penet_1-3; Buyan_2_3; Batur_6_8-11_18_25_27; Lombok_1-10; Serange_1-3_5_7-11; and Sekokat_1_2_4-6_9-11.

between Clade 1 and Clade 2 were estimated, respectively, to be the middle Pliocene at 4.0 (2.4-6.0) Mya and the Pliocene-Pleistocene boundary at 1.6 (0.9-2.5) Mya. Finally, subclades within Clades 1 and 2 were estimated to have diverged in the Pleistocene at 0.5-1.4 Mya. Figure 20 shows the dating result based on two mitochondrial genes. The molecular dating using the mitochondrial gene sequences yielded slightly older ages in many nodes than the

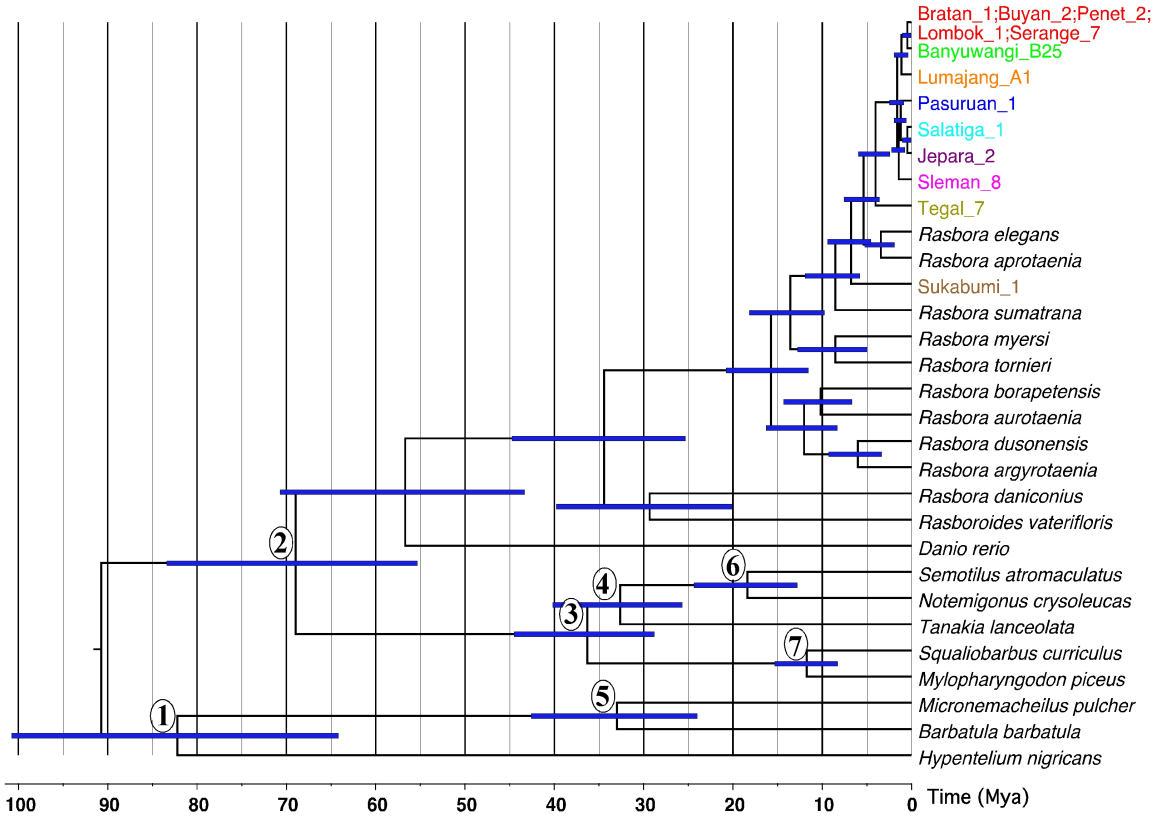

Fig 19. Divergence times estimated using RAG1 and opsin gene sequences. Circled numbers indicate seven calibration points used as priors for the relaxed-clock time estimation: 1) 78.8 Mya for the *H. nigricans* vs. *M. pulcher*+*B. barbatula* divergence; 2) 63.3 Mya for the *Danioninae* vs. *Leuciscinae*+*Acheilognathinae*+*Xenocypridinae* divergence; 3) 39.74 Mya for the *Xenocypridinae* vs. *Acheilognathinae*+*Leuciscinae* divergence; 4) 35.01 Mya for the *T. lanceolata* vs. *Leuciscinae* divergence; 5) 35.04 Mya for the *M. pulcher* vs. *B. barbatula* divergence; 6) 22.93 Mya for the *S. atromaculatus* vs. *N. crysoleucas* divergence; and 7) 11.48 Mya for the *S. curriculus* vs. *M. piceus* divergence (Betancur-R et al., 2013). Bars at nodes indicate 95% HPD intervals.

dating using the nuclear genes. For example, the divergence time between the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* and *R. sumatrana* was estimated to be 9.1 Mya (mitochondrial) vs. 8.6 Mya (nuclear). As another example, separation of Clade 4 from the remaining taxa in the *R. lateristriata* species complex occurred at 8.1 Mya (mitochondrial) vs. 6.8 Mya (nuclear) (Fig. 20). Some studies reported significant overestimation of divergence times using mitochondrial genes due to their rapid evolutionary rates (e.g., Steppan et al., 2005; Zheng et al., 2011). I thus used dating results primarily from nuclear genes for biogeographic discussion (see Chapter 4).

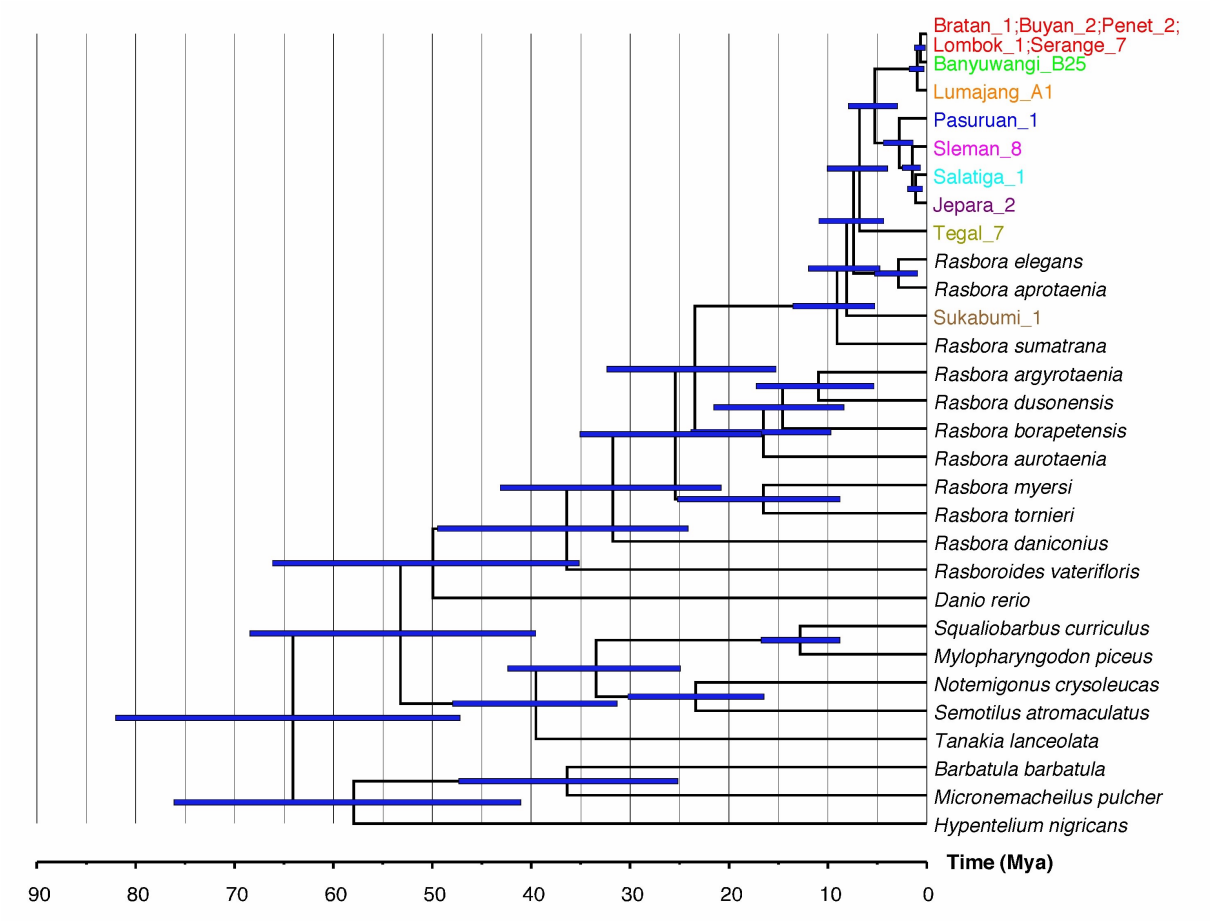

Fig. 20. Divergence time estimated using the first and second codon positions of mitochondrial COI and Cyt*b* gene sequences. Parameters in the BEAST analysis were set to be the same as in the nuclear genes analysis.

3.7 Reconstruction of historical biogeography

I conducted the historical biogeography analysis to infer ancestral geographic ranges at each major node in the *R. lateristriata* species complex by the Lagrange (Dispersal-Extinction-Cladogenesis, DEC) model. Results suggested that Sumatra or West Java is the most likely origin for this species complex (Fig. 21). This view was supported with a moderately high probability value (63.7%) at a node where the species complex separated from *R. sumatrana*. Direct common ancestors between Clade 4 (Sukabumi) and *R. aprotaenia*+*R. elegans*+Clade 3+Clade 2+Clade1 and between *R. aprotaenia*+*R. elegans* and Clade 3+Clade 2+Clade 1 were estimated to be in West Java and West-or-Central Java, respectively. The ancestral area of Clade 3+Clade 2+Clade1 was estimated to be Central Java

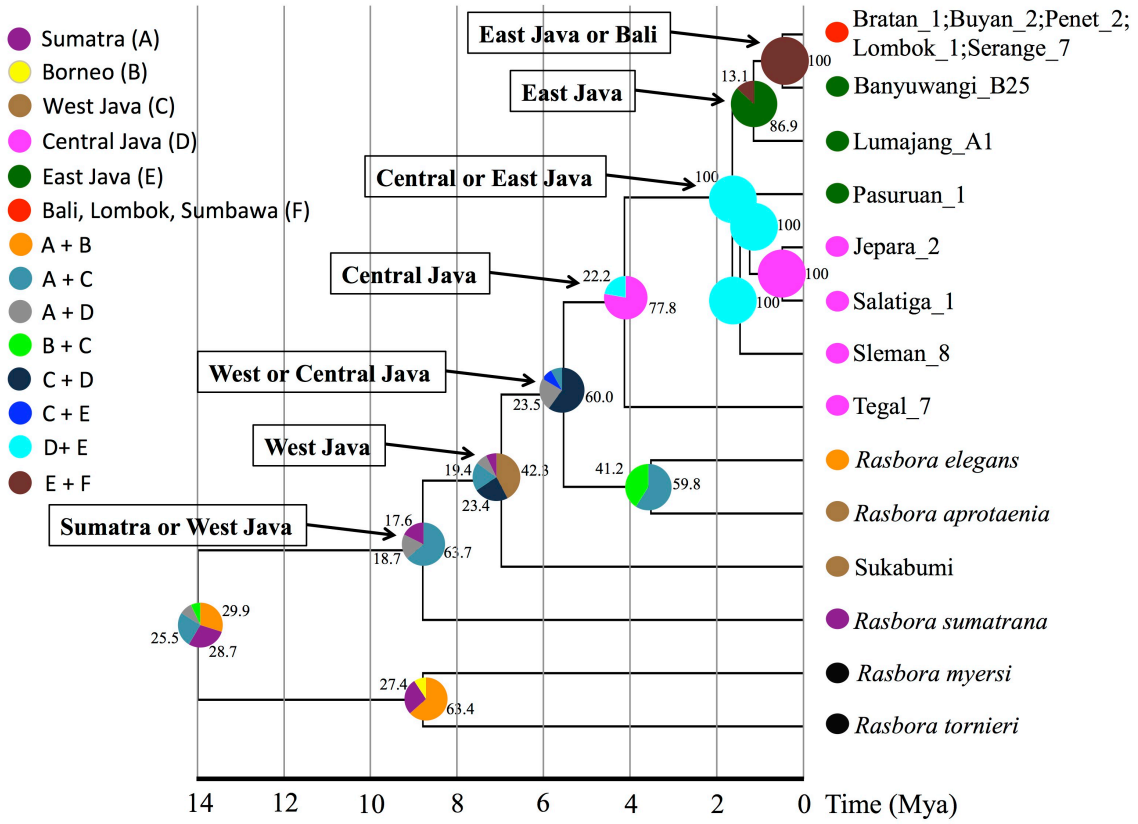

Fig. 21. Biogeographic reconstruction of the *R. lateristriata* species complex by RASP v3.2 (Yu et al., 2015) under the Lagrange (Dispersal-Extinction-Cladogenesis, DEC) model (Ree et al., 2005; Ree and Smith, 2008). Different colors at the tip of each branch represent current distributional areas. The first six colors in the left panel represent single region distribution at present while the next eight colors represent combination of two regions possibly assigned to ancestral nodes. Pie diagram at each node corresponds to relative probabilities of the estimated ancestral area reconstructions and only values above 10% are presented. An area or a combination of two areas which has the highest probability at the node is shown in square.

as supported with a relatively high probability value (77.8%). Central or East Java was strongly supported (100%) as the ancestral area of Clade 2+Clade 1. Finally, East Java or Bali was suggested as the ancestral area of subclade 1A+subclade 1B within Clade 1.

3.8 Reconstruction of haplotype network

Figure 22 shows a haplotype network or genealogical relationships among COI gene sequences of the *R. lateristriata* species complex. This network corroborates results from the

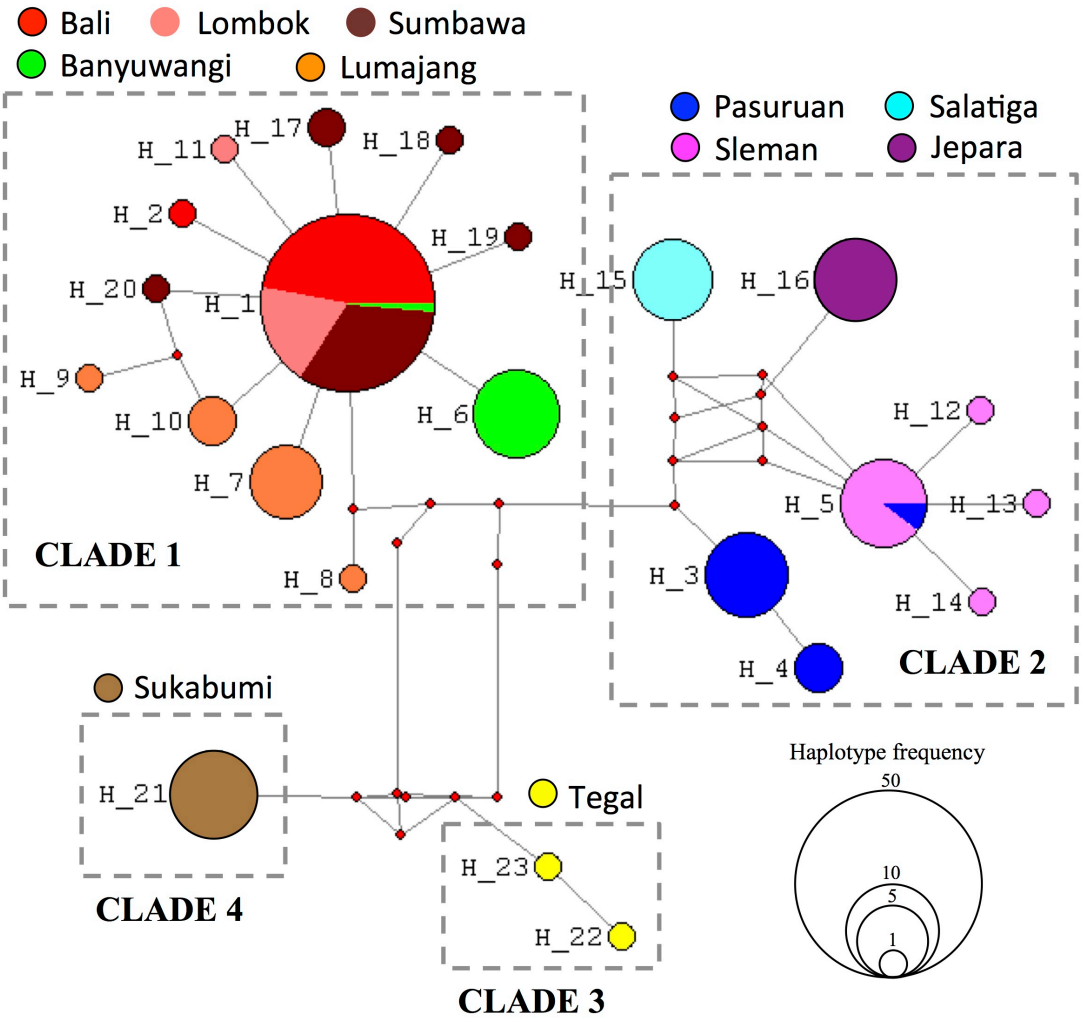

Fig. 22. A network showing relationships among COI haplotypes of the *R. lateristriata* species complex. Dotted boxes represent four haplogroups corresponding to the major clades recognized in the phylogenetic tree of Fig. 15. A circle represents each unique haplotype and its size is proportional to the corresponding frequency. In the case of a shared haplotype, the pie is divided in proportion to the relative haplotype frequency from each locality. Small red dots represent missing intermediate haplotypes.

phylogenetic analyses (Fig. 15) for the existence of four major clades. Within Clade 2, four geographically distinct groups of haplotypes were identified. Consistent with the phylogenetic analysis, a single individual of Pasuruan shared a haplotype (H_5) with most individuals of Sleman. Within Clade 1, haplotype H_1 was shared by many Balinese, Lombok and Sumbawa individuals commonly. Most of minor haplotypes within Clade 1 were separated by single base substitutions from the major haplotype H_1.

3.9 Analysis of morphological characters

Morphological investigations using selected morphometric and meristic diagnostic characters were conducted to test whether they could discriminate clades and subclades recognized in the *R. lateristriata* species complex from others. I first examined whether individuals from Bali show different morphological features from those of other localities in Java, Lombok and Sumbawa (Table 11). As a result, none of morphometric and meristic diagnostic characters suggested by Brittan (1954)(e.g., head and eye size, relative position when dorsal-hypural distance is carried forward, number of predorsal scales, and number of lateral line scales and pores) showed clear separation between the Balinese individuals and the others with considerable overlaps between them. Similarly, the four major clades of the *R. lateristriata* species complex were also indistinguishable from each other based on these characters (Table 11).

However, I noticed that several characters based on the body color pattern (Lumbantobing, 2014) can discriminate between the four major clades. The combination of pigmentation patterns of the SAP and BCB separated the *R. lateristriata* species complex into four groups (Fig. 23 and Table 12). The presence of SAP and the absence of BCB were shared by Clade 1 individuals. Clade 2 individuals shared the absence of both SAP and BCB. The absence of SAP and the presence of BCB were shared by Clade 3 individuals. Finally, Clade 4 individuals, *R. aprotaenia* and *R. elegans* shared the presence of both SAP and BCB. I examined 38 morphological characters in total, either from Brittan (1954) or Lumbantobing (2010, 2014). However, except for these two characters in the body color pattern, there appeared to be no conspicuous characters that can clearly distinguish individuals of some localities from the others. The details on the shape and melanophore pigmentation intensity of the body color pattern in the *R. lateristriata* species complex are shown in Table 13.

Table 11. Comparison of key morphological characters proposed by Brittan (1954) among samples collected by me

	Brittan (1954)		This study				
Characters[a]	*R. lateristriata*	*R. baliensis*	Clade 1[b]		Clade 2[b]	Clade 3[b]	Clade 4[b]
			Balinese	Non-Balinese			
Number of individual	n=25	n=4	n=72	n=86	n=62	n=2	n=14
Standard length (SL) (mm)	38.0-71.0	20.0-35.0	43.4-86.7 [59.7±12.9]	21.6-65.4 [44.71±7.5]	25.1-62.7 [47.2±8.3]	57.1-65.6 [61.4±6.0]	27.6-76.7 [50.2±13.6]
Head length % standard length (HL%SL)	3.4-4.5[c] [3.9]	3.1-3.3[c] [3.2]	22.0-28.1 [25.0±1.3]	23.4-29.3 [25.5±1.2]	22.8-29.0 [25.5±1.2]	22.6-24.2 [23.4±1.1]	25.3-30.1 [26.7±1.2]
Head depth % standard length (HD%SL)	4.6-5.4[c] [5.0]	4.0-4.3[c] [4.2]	15.8-19.8 [18.4±1.1]	15.3-19.2 [17.7±0.8]	15.9-20.2 [18.1±0.8]	16.9-18.4 [17.7±1.0]	16.2-18.8 [17.75±0.8]
Head length (HL) (mm)	-	-	11.3-21.0 [14.8±2.7]	6.1-16.0 [11.4±1.8]	7.0-15.7 [12.0±2.0]	13.8-14.8 [14.3±0.7]	8.3-20.3 [13.3±3.3]
Inter-orbital width % head length (IOW%HL)	2.4-3.5[d] [2.6]	2.6-3.0[d] [2.8]	38.0-46.9 [42.7±2.5]	32.9-45.2 [39.5±2.5]	31.5-45.2 [39.3±2.7]	42.8-47.3 [45.0±3.2]	32.6-44.9 [38.4±3.1]
Eye diameter % head length (ED%HL)	3.2-3.9[d] [3.5]	2.8-3.0[d] [2.9]	28.6-36.9 [33.2±2.4]	29.9-42.8 [35.2±2.5]	28.7-41.8 [35.1±2.6]	33.3-34.5 [33.9±0.8]	29.6-38.5 [34.4±2.6]
Lateral line scale	29-33	28	26-30 [27.8±0.9]	25-31 [28.3±1.2]	21-32 [28.5±2.3]	29-32 [30.5±2.1]	29-31 [30.6±0.6]
Lateral line pore	≥ 27	26	26-30 [27.65±1.0]	25-31 [28.3±1.2]	21-32 [28.4±2.3]	29-32 [30.5±2.1]	29-31 [30.6±0.6]
Pre-dorsal scale	12-14	11-12	10-12 [11.2±0.7]	11-14 [12.1±0.4]	10-14 [12±0.7]	12-13 [12.5±0.7]	12-14 [12.7±0.6]
Relative position when dorsal-hypural distance (DHD) carried forward, falling at	Between nostril and anterior rim of eye	Anterior rim of eye	Varied; ranging from before nostril to anterior rim of eye	Varied; ranging from before nostril to anterior rim of eye	Varied; ranging from before nostril to anterior rim of eye	Varied; ranging from before nostril to anterior rim of eye	Varied; ranging from before nostril to anterior rim of eye

Values in brackets show mean or mean±standard deviation

[a] Morphometric and meristic diagnostic key characters used by Brittan (1954) to discriminate *R. baliensis* from *R. lateristriata*

[b] See Table 2 for details of used specimens for each clade

[c] Proportional values, by dividing the character by the total length

[d] Proportional values, by dividing the character by the head length

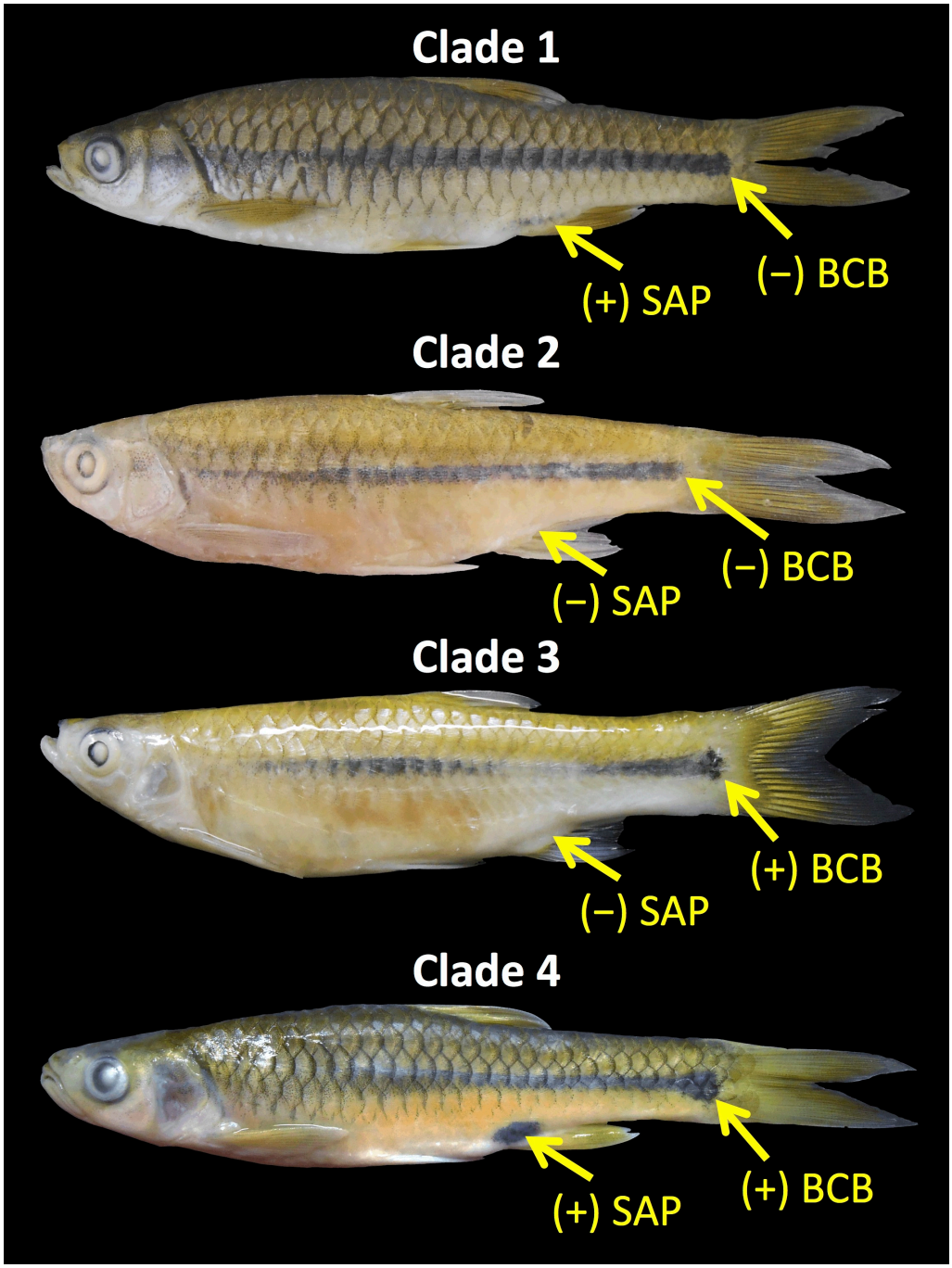

Fig. 23. Left lateral images of a representative individual from each major clade in the *R. lateristriata* species complex. The museum voucher number for each specimen is UB.1.118.7 (Lombok, 54.8 mm SL) from Clade 1, UB.1.127.9 (Jepara, 45.2 mm SL) from Clade 2, UB.1.142.1 (Tegal, 57.1 mm SL) from Clade 3, and UB.1.143.1 (Sukabumi, 75.0 mm SL) from Clade 4 (see Table 2 for details).

Table 12. Selected characters on the body color pattern for major clades of the *R. lateristriata* species complex, *R. aprotaenia* and *R. elegans*

Characters	Clade 1		Clade 2[a]	Clade 3	Clade 4	*R. aprotaenia*	*R. elegans*
	Balinese	Non-Balinese					
	(n=72)	(n=86)	(n=62)	(n=2)	(n=14)	(n=59)	(n=12)
Supra anal pigment (SAP)	+	+	-	-	+	+	+
Basicaudal blotch (BCB)	-	-	-	+	+	+	+
Midopercular stripe (MOS)	+	+	+ (43) / - (19)	-	-	-	-
Peripheral reticulation (PR)	+	+	+ (38) / - (24)	-	+	+	+
Basal reticulation (BR)	+	+	+ (51) / - (11)	-	+	+	+
Postopercular pigmentation (POP)	+	+	+ (42) / - (20)	-	+	+	+
Midlateral stripe (MLS)[b]	+	+	+	+	+	+	+
Midhumeral diffuse patch (MDP)	+	+	+ (48) / - (14)	-	+	+	-
Subdorsal blotch (SDB)	-	-	-	-	-	+	+

Plus (+) and minus (-) in the table stand for the presence and absence of the corresponding character, respectively. See Table 2 for more details of my specimens for each clade

[a] Individuals from Clade 2 have variations of body color patterns in MOS, PR, BR, POP and MDP. Numbers in parenthesis represent those for individuals having the corresponding state

[b] The pigmentation intensity as well as the shape of MLS somewhat differ between clades. See Table 13 for the details

Table 13. Detailed descriptions of selected characters of the body color pattern that can be used to delineate among major clades in the *R. lateristriata* species complex, *R. aprotaenia* and *R. elegans*

Characters / Clade	Supra anal pigment (SAP)	Basicaudal blotch (BCB)	Midopercular stripe (MOS)	Peripheral reticulation (PR) and Basal reticulation (BR)	Postopercular pigmentation (POP)
Clade 1[a,b] (n=158)	Dark mark, the intensity ranging from faint to prominent with elongate streak shaped	Absent	Distinct, the intensity is ranging from faint to dark and intense stripe melanophores	Distinct with dark and very prominent melanophores	Distinct, with dark and prominent intensity of melanophores
Clade 2[a] (n=62)	Almost all individuals are absent. A few individuals possess faint elongate streak shaped of SAP	Absent	Varied, ranging from absent, faint to dark intensity of melanophores	Varied, ranging from absent, faint to prominent intensity of melanophores	Varied, ranging from absent, faint to dark intensity of melanophores
Clade 3[a] (n=2)	Absent	Conspicuous dark mark, with triangular shaped	Absent, only scattered melanophores on the operculum	Indistinct, with light and obsolescent melanophores	Absent
Clade 4[a] (n=14)	Conspicuous dark mark, dense and bold melanophores with oval shaped	Conspicuous dark mark, with triangular or oval shaped	Absent, only scattered melanophores on the operculum appear	Distinct, with dark and prominent melanophores	Distinct, with dark and prominent intensity of melanophores
R. aprotaenia (n=59)	Conspicuous dark mark, dense and bold melanophores with oval shaped	Conspicuous dark mark, with triangular or oval shaped	Absent, only scattered melanophores on the operculum appear	Distinct, with dark and prominent melanophores	Distinct, with dark and prominent intensity of melanophores
R. elegans (n=12)	Conspicuous dark mark, dense and bold melanophores with oval shaped	Conspicuous dark mark, with triangular shaped. The size is larger than in *R. aprotaenia*	Absent, only scattered melanophores on the operculum appear	Distinct, with dark and prominent melanophores	Distinct, with dark and prominent intensity of melanophores

[a] See Table 2 for details of used specimens for each clade

[b] Balinese and non-Balinese individuals possess similar body color pattern

Table 13. (Continued)

Characters / Clade	Midlateral stripe (MLS)	Midhumeral diffuse patch (MDP)	Subdorsal blotch (SDB)
Clade 1[a,b] (n=158)	Dark and prominent of MLS, starting from caudal peduncle and terminating at operculum, resembles a solid line with relatively uniform overall appearance	Starting on the area at the middle between caudal peduncle and operculum just below the dorsal fin origin with somewhat diffuse melanophores, extending anteriorly and reaching the operculum. The onset of MDP overlap with the end part of MLS. MLS confluent towards MDP, resembles a solid dark line starting from caudal peduncle to the operculum	Absent
Clade 2[a] (n=62)	Ranging from dark to diffuse of melanophores intensity with relatively uniform overall appearance. It is extending from the base of the caudal fin and terminating at operculum	Varied, ranging from absent, faint to prominent intensity of melanophores	Absent
Clade 3[a] (n=2)	Dark and prominent appearing resembles bold line with very intense melanophores, extending from the base of the caudal fin but terminating at 3rd-4th lateral line scale	Absent	Absent
Clade 4[a] (n=14)	Dark and prominent, starting from the caudal peduncle and terminating at the midhumeral region. The melanophores more intense at the caudal peduncle extending anteriorly with the intensity gradually paled, become narrowing shaped appearing like a thin line when reaching the midhumeral region	Melanophores mostly concentrate on the middle of the side at below the dorsal fin origin. The intensity then somewhat sparse or diffuse anteriorly when reaching the operculum	Absent
R. aprotaenia (n=59)	Dark and prominent, starting from the caudal peduncle and connected with the subdorsal blotch (SDB)	Absent, only sparse of melanophores if present	Conspicuous dark blotch with oval or rectangular shaped, located in the midlateral region between below the dorsal fin and above the pelvic fin
R. elegans (n=12)	Faint and reduced melanophore intensity of MLS. Resembles thin line running from the caudal peduncle to the midhumeral region, connected with the subdorsal blotch (SDB)	Absent	Conspicuous dark blotch with oval or rectangular shaped, located in the midlateral region between below the dorsal fin and above the pelvic fin

[a] See Table 2 for details of used specimens for each clade

[b] Balinese and non-Balinese individuals possess similar body color pattern

Chapter 4: Discussion

4.1 Phylogenetic position of the *R. lateristriata* species complex

Phylogenetic analyses to determine the phylogenetic position of *R. lateristriata* among major species from genus *Rasbora* was conducted based on complete mitochondrial gene sequences and multilocus gene sequences involving COI, Cyt*b* RAG1 and opsin. Phylogenetic analyses using the mitogenomic dataset (Fig. 9) showed, with a strong bootstrap support (100%), that *R. lateristriata* is more closely related to *R. aprotaenia* than to any other *Rasbora* species examined, pointing to their phylogenetic closeness. Based on a view by Kottelat and Vidthayanon (1993), *R. aprotaenia* would be expected to cluster with *R. steineri* in the phylogenetic tree but this was not the case. Brittan (1954) separated species in genus *Rasbora* into eight groups. Both of Brittan (1954) and Liao et al. (2010) suggested the phylogenetic closeness between *R. lateristriata* and *R. aprotaenia* but they placed these species in a different group (the *R. lateristriata*-group and the *R. argyrotaenia*-group, respectively). Because my molecular experiments using the mitogenomes did not include *R. argyrotaenia*, it seems difficult to infer the phylogenetic position of *R. argyrotaenia* based on the mitogenomic dataset.

According to the multilocus dataset 1, the molecular phylogeny (Figs. 10 and 15) showed that individuals in Clades 1-4 together with *R. aprotaenia* and *R. elegans* make a monophyletic group. Individuals from Clades 1-4 are morphologically very similar to each other (Tables 11-13) but separated with large genetic distances (> 2%, Table 9). Therefore, I propose to regard the assemblage of Clades 1-4 as the *R. lateristriata* species complex until full taxonomic evaluations can formally define multiple species in it. While Brittan (1954) and Kottelat and Vidthayanon (1993) placed *R. lateristriata* in the *R. lateristriata*-group, a more recent phylogenetic reconstruction using 41 morphological characters transferred *R. lateristriata* into the *R. argyrotaenia*-group (Liao et al., 2010). My molecular phylogeny (Fig.

10) showed that the *R. lateristriata* species complex is distantly related with *R. argyrotaenia*, which does not support Liao et al. (2010) who transferred *R. lateristriata* into the *R. argyrotaenia*-group. More recently, Lumbantobing (2014) proposed another view by placing *R. lateristriata* within the Sumatrana group, mainly based on the shared presence of the MLS, SAP, and BCB pigmentations. However, as individuals of Clade 1 and Clade 2 lack BCB and those of Clade 2 and Clade 3 do not have SAP (Table 12), inclusion of *R. lateristriata* in the Sumatrana group may now be questionable. Taken all together, I propose to categorize the *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* into the *R. lateristriata*-group.

4.2 Taxonomic status of *R. baliensis*

According to Brittan (1954), *R. baliensis* can be distinguished from *R. lateristriata* by the combination of the following diagnostic characters: much bigger head and eye (vs. much smaller head and eye), dorsal-hypural distance when carried forward falling at anterior rim of eye (vs. falling at between nostril and anterior rim of eye), 11 to 12 pre-dorsal scales (vs. 12-14), 28 lateral line scales (vs. 29-33), 26 lateral line pores (vs. 27 or more), faint midlateral stripe (vs. darker), obsolescent SAP (vs. prominent) and obsolescent subpeduncular streak (vs. prominent). However, based on my observation using these characters, the delimitation between Balinese (possibly *R. baliensis*) and Javanese (possibly *R. lateristriata*) individuals was very difficult to recognize with clear morphological segregation (Table 11). Ranges in the examined morphometric and meristic diagnostic characters for 72 Balinese individuals considerably overlapped those for non-Balinese individuals in Clades 1-4. Namely, all diagnostic characters suggested by Brittan (1954) were unable to discriminate the Balinese (or Bratan) individuals from non-Balinese individuals.

Brittan (1954) described *R. baliensis* using 4 individuals (holotype with 35 mm in SL, range of SL= 20-35 mm; Fig. 24) collected from Lake Bratan, Bali in comparison with 25 *R. lateristriata* individuals obtained from Cikunir River (13 individuals), tributary of Cikunir

River (2) of West Java; Telaga Teroes (3) of Central Java; Lake Ranu Klakah (3), Lake Ranu Pakis (4) of East Java. A plausible reason for the discrepancy between Brittan's (1954) and my morphological analyses lies in the difference in the number of analyzed Balinese individuals. Since Brittan (1954) used very few Balinese individuals, their morphological characters may not have overlapped with those of the Javanese individuals accidentally. Another possible reason is that Brittan (1954) used only young individuals from Bali, judging from their body size (20-35 mm SL). Juvenile fishes tend to have larger head and eye than adults when standardized by the standard length (Loy et al., 1998; Reichard and Jurajda, 1999). An additional possible explanation for the discrepancy is that *R. baliensis* described by Brittan (1954) has gone extinct during the last several decades or that I simply could not collect them in Balinese lakes. Although I cannot strictly deny these possibilities, I consider that the last explanation is unlikely. I visited Balinese freshwater localities multiple times to collect 72 samples in total. None of these individuals showed distinct molecular or morphological features from those of East Javanese individuals (Fig. 15, Tables 11-12).

Fig. 24. Images corresponding to the holotype (**A**, museum voucher number UMMZ157146) and the paratype (**B-D**, voucher numbers UMMZ157127_1, UMMZ157127_2 and UMMZ157127_3, respectively) specimens used by Brittan (1954) for describing *R. baliensis*. All specimens were collected in Lake Bratan, an enclosed-crater lake in Bali Island and currently deposited in the University of Michigan Museum of Zoology. The images are available through the courtesy of Dr. Douglas Nelson and Ms. Anna Barget of the University of Michigan Museum of Zoology (the Great Lakes Invasive Species Project, UMMZ).

Taken together, these results point to the need to revise the taxonomic status of *R. baliensis*. My molecular and morphological data suggest that Clade 1 consisting of individuals from Bali, Lombok, Sumbawa and East Java may represent a species. Thus, Clade 1 may be defined as *R. baliensis* in future by exploring further diagnostic characters in addition to the body color pattern shown in Table 12.

4.3 Cryptic species

Only three *Rasbora* species have previously been reported from Java Island (*R. argyrotaenia*, *R. lateristriata* and *R. aprotaenia*). Whereas *R. argyrotaenia* and *R. lateristriata* are widely distributed throughout Java Island, *R. aprotaenia* is known only from western parts of Java Island (Brittan, 1954, 1972; Kottelat et al., 1993; Froese and Pauly, 2015). My molecular (Fig. 15 and Table 9) and morphological (Table 12) analyses agreed in that Clade 2 and Clade 3 represent unknown cryptic species within the *R. lateristriata* species complex. Clade 2 consists of individuals from Central Javanese localities (Pasuruan, Sleman, Salatiga and Jepara) and Clade 3 consists of those from a west-central location (Tegal). In contrast to the very shallow divergences and non-monophyletic structure of haplotypes obtained from different localities (islands) within subclade 1A, haplotypes from the four central Javanese localities mostly make monophyletic groups corresponding to 4 subclades in Clade 2 (Fig. 15). This probably reflects sufficient times for the lineage sorting for each of the central Javanese localities, but not for localities in the subclade 1A. Two major rivers (Solo and Brantas) run in central Java (Fig. 8) but these rivers do not connect the four localities directly. Thus, populations in these four localities in central Java may have been disconnected without gene flow.

Clade 4 is composed of individuals from a west Javanese locality, Sukabumi. I compared the body color pattern of these individuals with the illustrated figure (original drawing with watercolor painting) of *R. lateristriata* specimens used by van Hasselt (1823)

and Bleeker (1854), which also appear as Fig. 20 in Roberts (1993) and Fig. 116 in Oijen and Loots (2012), respectively. As a result, Sukabumi samples had similar body color patterns as *R. lateristriata* in these papers, especially with respect to the presence of SAP and BCB pigmentations. As *R. lateristriata* was described using individuals collected from Bogor, West Java (Bleeker, 1854), this result is consistent with the geographical location of Sukabumi in West Java. Taken together, I propose that Clade 4 might be regarded as *R. lateristriata* in future revision, enabling Clade 2 and Clade 3 to be described under new species names.

4.4 Evolution of body color patterns

Body color patterns are important features not only for describing new species in the genus *Rasbora* (Kottelat, 2005; Kottelat and Tan, 2011; Lumbantobing, 2014) but also for elucidating their phylogenetic relationships (Liao et al., 2010, 2011). My observation on the morphological characters especially on the body color pattern revealed that combination of SAP and BCB serves to separate the *R. lateristriata* species complex into four groups (see section 3.9; Fig. 23 and Table 12*)*. Most species of *Rasbora* possess the SAP and only some don't: e.g., *R. aurotaenia* (Rainboth et al., 2012), *R. jacobsoni* (Brittan, 1954), *R. einthovenii* (Kottelat et al., 1993) and *R. tubbi* (Kottelat et al., 1993). My phylogenetic analysis suggested that these species without SAP are distantly related to each other (Fig. 10). On the other hand, BCB is absent in most of *Rasbora* species but present in *R. sumatrana* (Tan and Kottelat, 2009), *R. paviana* (Kottelat, 2005) *R. hobelmani* (Kottelat, 1984), *R. vulgaris* (Kottelat, 2005), *R. lateristriata* (Roberts, 1993), *R. elegans* (Kottelat et al., 1993), *R. aprotaenia* (Kottelat et al., 1993) and *R. kalbarensis* (Kottelat et al., 1993). Except for *R. kalbarensis*, species that have the BCB are closely related to each other (Fig. 10). The loss of SAP and the gain of BCB in *Rasbora* seem to have occurred independently multiple times.

Many species in the Sumatrana group and individuals from Clade 4 of the *R. lateristriata* species complex commonly have both SAP and BCB (Fig. 25; Table 12;

Lumbantobing, 2014). *R. sumatrana* has a sister-group relationship with the *R. lateristriata*-group species in which Clade 4 basally diverged from other clades (Figs. 10 and 15). Existence of SAP and BCB may therefore represent a plesiomorphic character state in the *R. lateristriata*-group (Fig. 26). Whereas SAP and BCB appear to have been retained during diversification of the Sumatrana group, they were lost differently among species (clades) of the *R. lateristriata*-group. The SAP is retained by members of *R. aprotaenia*, *R. elegans*, Clade 1 and Clade 4. Thus, SAP may have been lost independently at Clade 2 and Clade 3 or

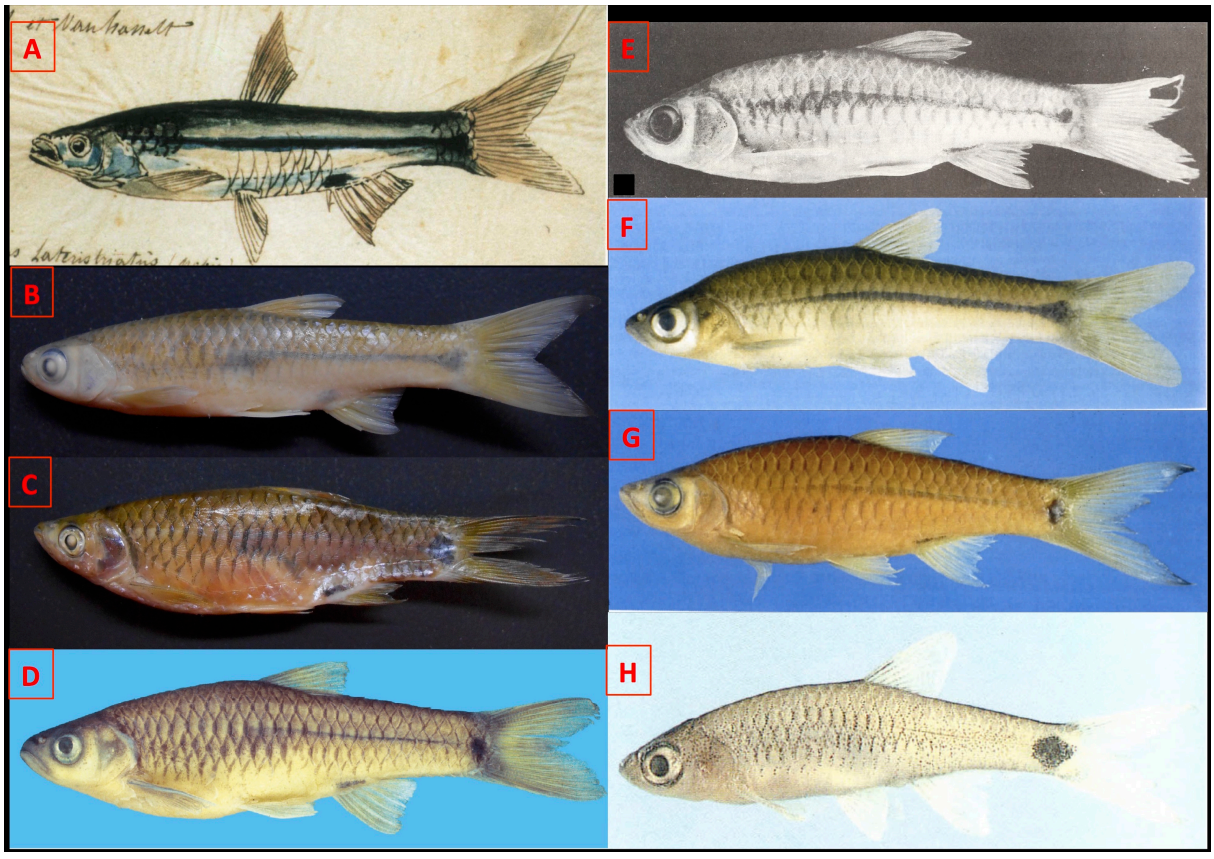

Fig. 25. Images of *R. lateristriata* (A, Bleeker, 1854 as in Roberts, 1993), *R. aprotaenia* (B, 64.5 mm SL; UB.1.120.25), *R. elegans* (C, 73.3 mm SL; UB.1.144.5), *R. sumatrana* (D, 66.6 mm SL; Kottelat et al., 1993); *R. hobelmani* (E, 50.1 mm SL; Kottelat, 1984); *R. paviana* (F, 60.0 mm SL; Kottelat, 2005*); R. vulgaris* (G, 59.7 mm SL; Kottelat, 2005) and *R. kalbarensis* (H, 20 mm SL; Kottelat, 1993). Most species of *Rasbora* possess the SAP. On the contrary, the BCB is absent in most of *Rasbora* except for these species.

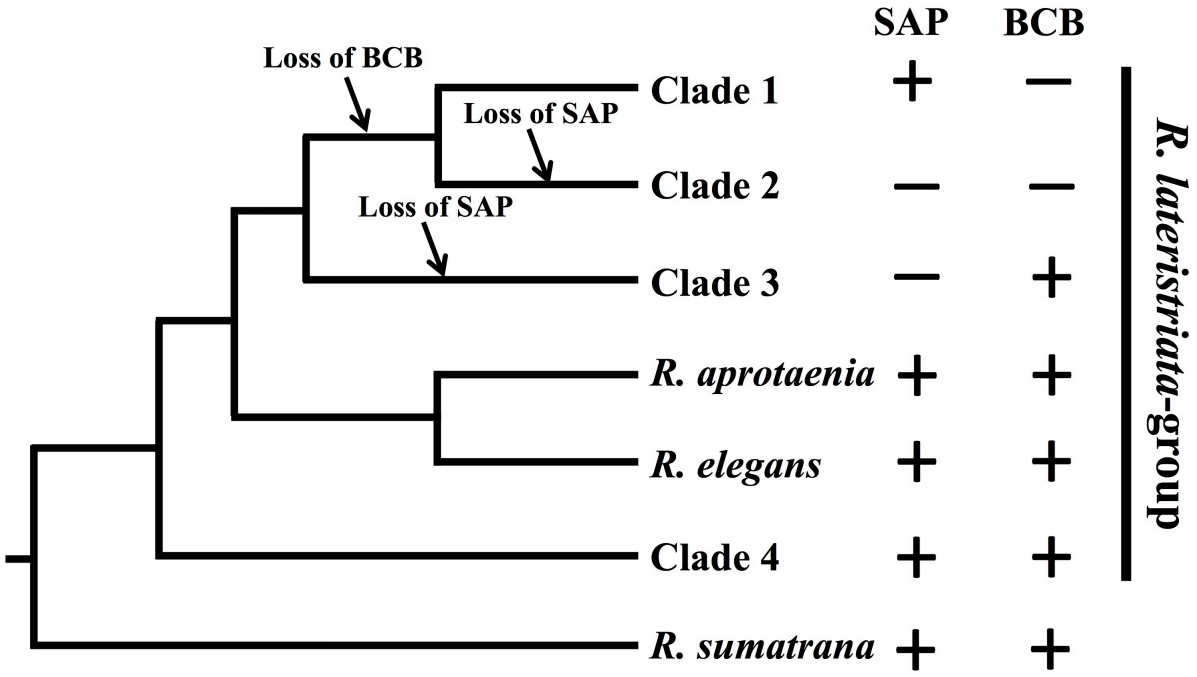

Fig. 26. Evolution of SAP and BCB pigmentation patterns in the *R. lateristriata*-group taxa. Based on the presence (+) and absence (-) of SAP and BCB in individual taxa, possible evolutionary changes in these characters were inferred by the parsimony criterion. Note that the parallel loss of SAP in lineages leading to Clade 2 and Clade 3 is equally parsimonious to the loss of SAP in the common ancestor of Clades 1-3 and its reversal gain in a lineage leading to Clade 1.

it may have been regained at Clade 1 (Fig. 26). On the other hand, members of *R. aprotaenia*, *R. elegans*, Clade 3 and Clade 4 retained BCB and its disappearance likely occurred in the common ancestor of Clade 1 and Clade 2 (Fig. 26).

Based on the similarity in the intensity and shape of melanophore pigmentation of MLS, Lumbantobing (2014) separated rasboras in the Sumatrana group into three subgroups: (1) the Hosii subgroup, which is characterized by the presence of wider anterior subdorsal portion of MLS than the posterior region (Fig. 27A); (2) the Lateristriata subgroup, which can be identified by the presence of MLS but having indistinct subdorsal blotch (Fig. 27B); (3) the Elegans subgroup, which can be illustrated by the undeveloped and subtle type of MLS (Fig. 27C). The division into three subgroups in the Sumatrana group does not seem to agree with

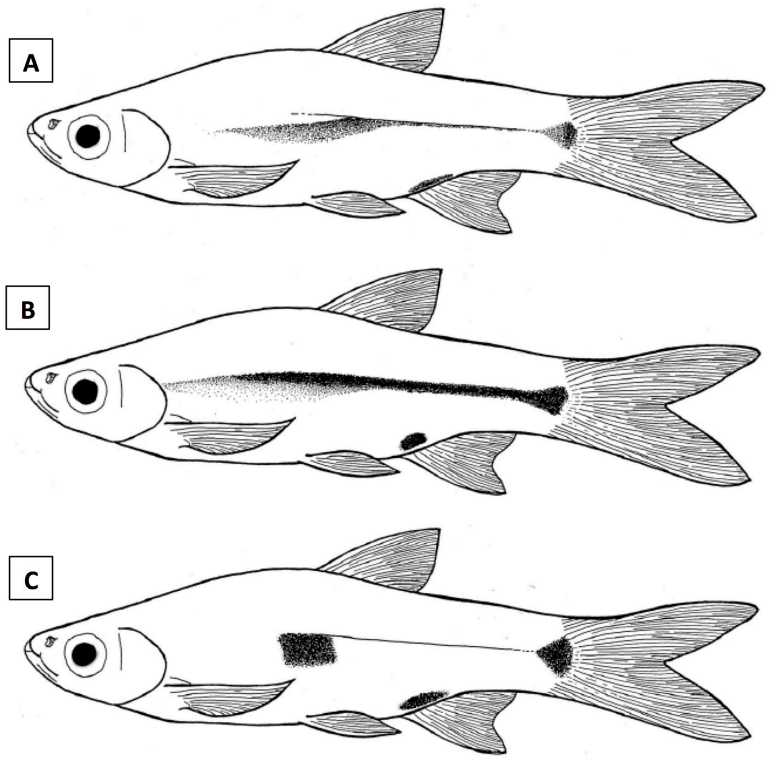

Fig. 27. Schematic drawing of three subgroups in the Sumatrana group: the Hosii subgroup (A), the Lateristriata subgroup (B) and the Elegans subgroup (C). The variation of intensity and shape of black midlateral stripe (MLS) running from caudal base extending to the anterior region serves as the key character in classifying the Sumatrana group into three subgroups. Images were obtained from Lumbantobing (2014).

my molecular phylogeny (Fig. 10). For example, *R. caudimaculata* and *R. trilineata* were categorized in the Lateristriata subgroup by Lumbantobing (2014) while both species are distantly related with the *R. lateristriata* species complex (Fig. 10). As another example, *R. aprotaenia* and *R. elegans*, which were categorized into different subgroups (i.e., the Hosii and Elegans subgroups, respectively), turned out to be closely related sister species (Fig. 10).

4.5 Historical biogeography

The *R. lateristriata*-group (i.e., *R. lateristriata* species complex+*R. aprotaenia*+*R. elegans* clade; see Fig. 10) consists of specimens mostly from Javanese localities but there are

two exceptions. First, *R. elegans* occurs in Peninsular Malaysia, Singapore, Sumatra and Borneo (Kottelat et al., 1993; Froese and Pauly, 2015). Second, Clade 1 extends from eastern Java and Bali to Wallacean Islands of Lombok and Sumbawa (Fig. 15). Rasboras outside the *R. lateristriata*-group all have non-Javanese distributions, except for *R. argyrotaenia* (Froese and Pauly, 2015). Thus, it is reasonable to deduce that rasboras colonized Java Island by at least two independent migrations: one by the ancestor of the *R. lateristriata*-group and the other by that of *R. argyrotaenia*.

The Sundaic region has a shallow continental shelf and the eustatic change of sea level has repeatedly connected major islands in this region to form Sundaland (Rainboth, 1996; Voris, 2000). Therefore, migrations of freshwater fishes into and out of Java may have been possible until the last glacial period (10-70 thousand years ago). However, Lombok Strait between Bali and Lombok Islands and Makassar Strait between Borneo and Sulawesi Islands were deep enough not to allow a land bridge across Wallace's Line even during the Quaternary glacial period (Moss and Wilson, 1998; Hall, 2009, 2013). This paleogeographical setting may have allowed the occurrence of the *R. lateristriata* species complex members also in Borneo and Sumatra. Weber and de Beaufort (1916) and Kottelat et al. (1993) noted that *R. lateristriata* is also distributed in Sumatra and Borneo. In order to explore this possibility in more details, I inspected 104 specimens collected from these islands and registered at the MZB as *R. lateristriata* (Table 14). As a result, characteristics of the Sumatran and Bornean individuals turned out to be different from those of Clade 1-4 individuals based on the body color pattern (Table 14). Unfortunately, these museum specimens have been stored in formalin and molecular characterization was unable to be executed. Thus, I do not have a definitive conclusion on the distributional area of the *R. lateristriata* species complex, especially on whether it exists in Borneo or Sumatra, too.

Table 14. Specimens of Museum Zoologicum Bogoriense inspected for their morphology

Calatog number	Registered at MZB as	Confirmed as	No. of specimens	Island	Detailed locality
MZB 10582[a]	*R. lateristriata*	*R. lateristriata*-like specimens	4	Java	Cileuleuy River, Kampung Gintung, Desa Nanggerang, Kecamatan Curug,Kabupaten Sukabumi, West Java
MZB 10584[a]	*R. lateristriata*	*R. lateristriata*-like specimens	7	Java	Cileuleuy River, Kampung Gintung, Desa Nanggerang, Kecamatan Curug,Kabupaten Sukabumi, West Java
MZB 16592[b]	*R. lateristriata*	*R. lateristriata*-like specimens	12	Java	Gua Gremeng, Central Java
MZB 4821[c]	*R. lateristriata*	*R. lateristriata*-like specimens	10	Sumatra	Desa Tes, Kecamatan Lebong Selatan, Bengkulu
MZB 5367[c]	*R. lateristriata*	*R. lateristriata*-like specimens	3	Sumatra	Lake Laut Tawar, Takengon, Central Aceh
MZB 5254[c]	*R. lateristriata*	*R. lateristriata*-like specimens	4	Borneo	Emil Timber, South Kalimantan
MZB 12457[c]	*R. lateristriata*	*R. lateristriata*-like specimens	4	Borneo	Mentesi River (station 16), Desa Tumbang Tohon, Sumber Barito, Murung Raya-Central Kalimantan
MZB 12262[c]	*R. lateristriata*	*R. lateristriata*-like specimens	3	Borneo	Semaja Drainage, Desa Samaenre Semaja, Nunukan Utara, Nunukan-East Kalimantan
MZB 12201[d]	*R. lateristriata*	*R. argyrotaenia*	6	Sumatra	Air terjun Batu Ampar, Kampung Air Salak, Singkep Selatan, Riau
MZB 13257[d]	*R. lateristriata*	*R. argyrotaenia*	3	Borneo	Anak Sungai Tabalar Kiri, , Kec. Tubasa, Berau, East Kalimantan
MZB 12243[d]	*R. lateristriata*	*R. argyrotaenia*	14	Borneo	Sungai Joloi, (St.1), Ds. Tumbang Tohon, Kec. Sumber Barito, Murung Raya, Central Kalimantan
MZB 2410[d]	*R. lateristriata*	*R. argyrotaenia*	2	Borneo	Tepi Sungai, Kec. Muara Ancolong, Kutai Tenggara, East Kalimantan
MZB 6898[d]	*R. lateristriata*	*R. argyrotaenia*	12	Borneo	Sungai Lubi, Anak Sungai Laung, Barito Utara, Central Borneo
MZB 12436[d]	*R. lateristriata*	*R. argyrotaenia*	3	Borneo	Sungai Tubulus, Ds. Tumbang Tohon, Sumber Barito, Murung Raya, Central Borneo
MZB 7191[d]	*R. lateristriata*	*R. argyrotaenia*	10	Borneo	Sungai Joloi, anak Sungai Barito, Central Borneo
MZB 6932[e]	*R. lateristriata*	*R. elegans*	5	Borneo	Sungai Laung, anak Sungai Barito, Laung Tutup, Barito Utara, Central Borneo
MZB 4494[e]	*R. lateristriata*	*R. elegans*	7	Sumatra	Sungai Bahorok, Langkat, North Sumatra
MZB 4177[e]	*R. lateristriata*	*R. elegans*	14	Borneo	Tempake Tanah Merah, East Kalimantan
MZB 4696[e]	*R. lateristriata*	*R. elegans*	4	Sumatra	Batang Aro, Bangkinan, Riau
MZB 5299[f]	*R. lateristriata*	*R. aprotaenia*	12	Java	Citamani Jaya, Kec. Sumur Pandeglang, West Java
MZB 19826	*R. sumatrana*	*R. sumatrana*	2	Sumatra	Sungai Lembang, Ds. Pasilembang, Aceh Selatan, Nangroe Aceh Darussalam
MZB 13063	*R. sumatrana*	*R. sumatrana*	4	Borneo	Sungai Batu Ngiring, Ds. Karendan, Central Kalimantan

[a] These specimens show the body color pattern as in Clade 4 individuals of Table 12 (data not shown)

[b] This specimen shows the body color pattern as in Clade 2 individuals of Table 12 (data not shown)

[c] These specimens show distinct body color pattern from Clades 1-4 (data not shown)

[d] Specimens of *R. argyrotaenia* misidentified as *R. lateristriata*

[e] Specimens of *R. elegans* misidentified as *R. lateristriata*

[f] Specimens of *R. aprotaenia* misidentified as *R. lateristriata*

If the *R. lateristriata* species complex did not migrate to Sumatra and Borneo Islands, why was it unable to disperse to these islands while *R. argyrotaenia* appears to be distributed in both Java and Borneo Islands (Fig. 7)?. My field observations suggested a possibility that these two fish prefer different habitats in which the *R. lateristriata* species complex tends to inhabit the upstream river whereas *R. argyrotaenia* tends to occur in the downstream area. To explore this hypothesis, I obtained the altitude data of each locality using Google Earth (Google Inc.) based on the GPS data recorded in the field. The average of the altitudes among 17 localities for the *R. lateristriata* species complex (435 m above the sea level) was higher than that among 18 localities for *R. argyrotaenia* (83 m above the sea level) (data not shown). This may possibly indicate that the *R. lateristriata* species complex adapted to the higher altitude had lower mobilities than *R. argyrotaenia* and that this prevented the species complex from migrating to Borneo even in the timing of land connection between Java and Borneo during the Quaternary glacial period.

My molecular phylogeny (Fig. 10) suggests that *R. sumatrana* of Sumatra is a sister taxon of the *R. lateristriata*-group. Some new species that appear to be closely related to *R. sumatrana* have also been described from Sumatra (Lumbantobing, 2014) although they were not sampled in my study. This is consistent with a view that an ancestor of the *R. lateristriata*-group diverged from *R. sumatrana* and its allies in the western part of Java or in Sumatra, after which the ancestor of the species complex started to colonize Java from the western side (Fig. 21). This divergence from *R. sumatrana* was dated to be around 8.6 Mya in the late Miocene (Fig. 19).

The molecular phylogeny suggests a west-to-east migrational history for the *R. lateristriata* species complex (Fig. 28). Clade 4 and *R. aprotaenia* occur only in west Java. Clade 3 occurs in west-central Java and Clade 2 has central Javanese localities. Finally, Clade 1 occurs in eastern Java and eastern islands. Within Clade 1, basal divergences occur between Lumajang individuals and the others, followed by a divergence between Banyuwangi

individuals and those from eastern islands. This series of divergence matches the west-to-east direction of migration (Fig. 28). This conclusion was corroborated by the historical biogeographic reconstruction in which geographic distribution at ancestral nodes of the *R. lateristriata* species complex was estimated to have changed from west to east in Java (Fig. 21).

What caused this west-to-east direction of divergence and migration? Geological evidence (Hall, 2009, 2013) suggests that Sumatra and Java were mostly submerged in the shallow sea in the mid-Miocene (~15 Mya) and that global cooling in the late Miocene (5-10 Mya) facilitated the emergence of some land areas in Sumatra and West Java. The Sunda Strait between Sumatra and Java started to open by the early Pliocene (~5 Mya) and active volcanic activities created land areas of East Java in the late Pliocene-Pleistocene (1-2 Mya). The estimated divergence time between *R. sumatrana* and the *R. lateristriata* species complex (8.6 Mya in mean and 5.8-11.9 Mya in 95% HPD; Fig. 19) corresponds to the timing for the emergence of Sumatra and West Java, somewhat earlier than the opening of the Sunda Strait. The estimated divergence time between Clade 2 in Central Java and Clade 1 in East Java and eastern islands (1.6 Mya in mean and 0.9-2.5 Mya in 95% HPD; Fig. 19) corresponds to the timing for the emergence of land areas in East Java. Thus, I consider that the *R. lateristriata* species complex diverged and migrated in association with the geological history of Java Island.

When and how did Clade 1 individuals cross the deep (> 250 m) and wide (> 20 km) Lombok Strait over Wallace's Line? Based on my molecular evidence (Fig. 15), individuals from Bali, Lombok and Sumbawa have very shallow molecular divergences without monophyletic structures for each island. It thus seems likely that Lombok and Sumbawa individuals in Subclade 1A originated from Balinese individuals. The haplotype network analysis using COI gene sequences also supported this conclusion, in which most Balinese, Lombok, and Sumbawa individuals shared a major COI haplotype (H_1) and the other minor

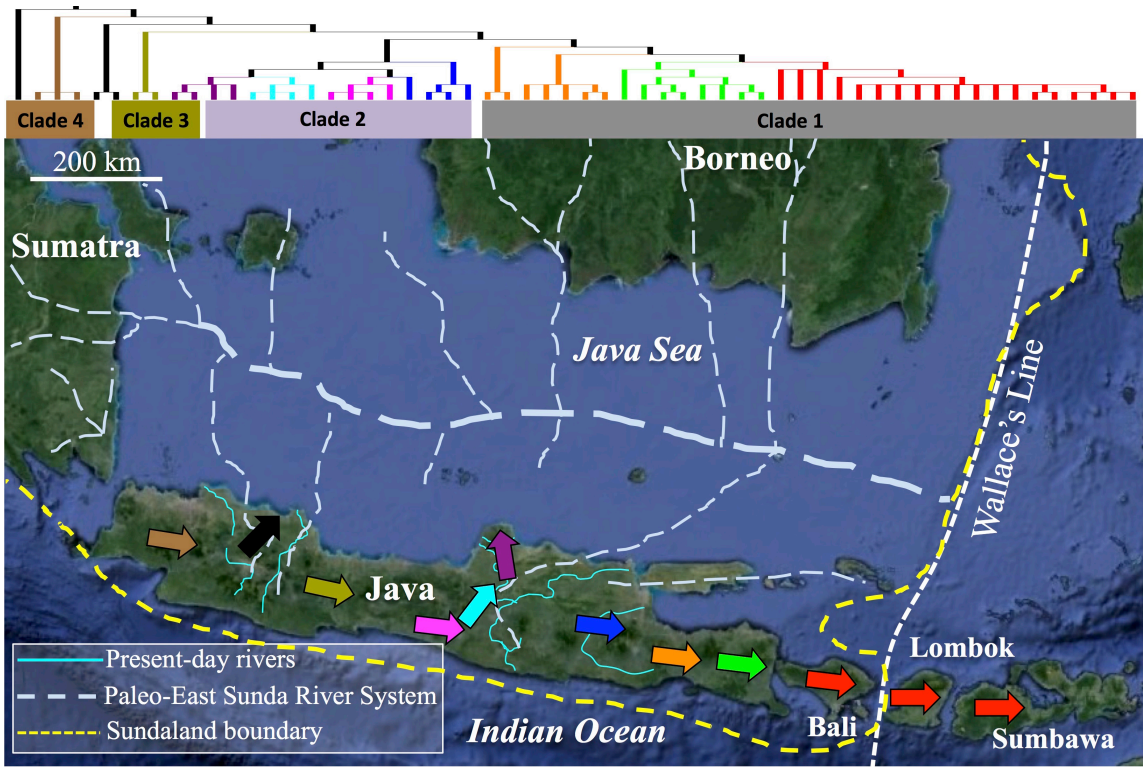

Fig 28. A schematic illustration on the hypothetical migrational pathway of the *R. lateristriata* species complex. On top, the phylogenetic tree shown in Fig. 15 is schematically depicted using different colors for lineages. The colors match geographic occurrence of the corresponding lineages, pointing to the hypothetical west-to-east divergence and migration. The paleo-drainage system (the East Sunda River System) was adapted from Voris (2000). The boundary of Sundaland is based on Bird et al. (2005). The map is produced based on a satellite image from Google Earth v7.1.5.1557.

haplotypes from these islands likely originated from the major haplotype by single base substitutions (Fig. 22). Because the estimated divergence time between Subclades 1A and 1B was around 0.48 Mya (Fig. 19), migration from Bali to Lombok and Sumbawa seems to have occurred much more recently than 0.48 Mya.

In the early 20th century, several cyprinid species, such as *Puntius gonionotus, Helostoma macrolepidota*, *Cyprinus carpio* and *Osteochilus hasseltii*, were introduced to Lombok by the local government for the aquaculture purpose (Monk et al., 1997). *Rasbora* might have been unintentionally transported on this occasion. However, there is no clear record on from which locality the cyprinids were introduced. Since Bali has never been known as a place for aquaculture activities (Sri Paryanti, 2006), it is not straightforward to

think that the introduced cyprinids originated from Balinese lakes. Ways for natural migration across Wallace's Line are more difficult to envisage. Flood may carry freshwater fishes in some distances, depending on their tolerance of salt water, and the Miocene seawater dispersal of the salt-tolerable ricefishes across the Makassar Strait has been proposed by Mokodongan and Yamahira (2015). However, there is no big river running in Bali (Whitten et al., 1996) and *Rasbora* is not tolerant of brackish water (Brittan, 1998). Taken together, whether this migration was natural or mediated by human being remains unclear.

Chapter 5: Conclusions and Future Prospects

Molecular phylogenetic analyses for rasboras from Java and neighboring islands have never been conducted before and the present study is eventually the first attempt. In this study, efforts to resolve molecular phylogeny and historical biogeography of the *R. lateristriata* species complex were made using the extensive sampling strategy and up-to-date molecular methods. As a result, I found evidence for the existence of possibly two new species represented by individuals from Clade 2 and Clade 3. It was also suggested that Clades 1-4 together with two other valid species (*R. aprotaenia* and *R. elegans*) form a monophyletic group, which I proposed to name the *R. lateristriata*-group. I also proposed to regard the morphologically homogeneous individuals of Clades 1-4 as the *R. lateristriata* species complex until full taxonomic investigation can be executed to redefine *R. lateristriata* and *R. baliensis* and describe new species in this group. Clade 4, consisting of individuals from Sukabumi, West Java, might be defined as *R. lateristriata* whereas Clade 1 comprising of individuals from eastern Java, Bali, Lombok and Sumbawa Islands may be defined as *R. baliensis* in future revision.

I discussed the historical biogeography of the *R. lateristriata* species complex with reference to geological history of Indonesian archipelago especially the islands of Sumatra, Borneo, Java and Bali. I proposed for the first time a hypothetical west-to-east migrational history of the *R. lateristriata* species complex. The *R. lateristriata* species complex was originated from Sumatra or western parts of Java Island and subsequently dispersed to central and east Java and Bali before it colonized Lombok and Sumbawa Islands over Wallace's Line. The estimation of divergence times suggested that the divergences in this species complex occurred from late Miocene to Plio-Pleistocene (8.6-0.48 Mya). Ancestors of Lombok and Sumbawa individuals likely originated from Balinese fresh waters and crossed Wallace's Line very recently (< 0.48 Mya). However, how Balinese individuals crossed wide

and deep Lombok Strait over Wallace's Line (i.e., an issue on natural vs. human-made introduction) was not fully resolved in this study. Further investigations involving population genetic approaches by either denser sampling in Bali, Lombok and Sumbawa Islands and/or additional genetic markers (e.g., microsatellite DNA) may be worth for deciphering the riddle.

Currently, several species commonly occur in both western (Java and/or Bali) and eastern (Lombok and/or Sumbawa) sides of Wallace's Line, e.g., *Puntius gonionotus*, *Puntius binotatus*, *Anabas testudineus*, *Channa gachua*, *Xiphophorus hellerii* and *Poecilia* sp. The occurrence of *X. hellerii* and *Poecilia* sp. in Lombok and Sumbawa Islands is certainly due to human introduction because they are native species in South America. Causes for the distribution of the remaining species over Wallace's Line are still uncertain. Studies on the historical biogeography of these species in future will be important to answer the critical question of natural vs. human-mediated introduction.

R. argyrotaenia is another widely distributed species in Indonesia. I collected individuals of this species from various localities in Java and Borneo (Fig. 7) and conducted preliminary phylogenetic analysis using three gene sequences (mitochondrial COI and Cytb genes and nuclear RAG1 gene). Interestingly, the result suggested no conspicuous phylogenetic pattern divided into geographical regions and the divergences within *R. argyrotaenia* individuals from wide localities (e.g., Java and Borneo) were much shallower than those within the *R. lateristriata* species complex (data not shown). Both *R. lateristriata* and *R. argyrotaenia* are small primary freshwater fishes with similar morphological appearance and no clearly different ecological characteristics between them are reported to my knowledge. What made the sharp difference in phylogenetic structures between the two species (species complex) remains an open question.

Indonesia is extremely rich in biodiversity. Unfortunately, this biodiversity has been rapidly declined and still under serious threats owing to human activities. Many species await molecular and morphological investigations before their natural habitats are critically

destroyed by human activities, which is now progressing in an enormous speed unfortunately. In the worst scenario, they may go extinct before we have a chance to study them. Thus, effective conservation efforts must be immediately conducted to ensure their sustainability. In order to do this, basic scientific information on taxonomy, phylogeny, ecology, and genetic diversity is critically needed but such studies are very rare in Indonesia.

As demonstrated in this dissertation, multidisciplinary approaches by field sampling, morphological investigations, molecular experiments, and computational analyses will be effective to tackle complex evolutionary issues. I thus believe that my study may be one of good examples to concord with the philosophy of "systematic natural sciences" at my graduate school.

Acknowledgments

I sincerely acknowledge the Directorate General of Higher Education Scholarship (DGHE/DIKTI Scholarship No.1757/E4.4/K/2012) from Ministry of Research, Technology and Higher Education of the Republic of Indonesia for allowing me to conduct my doctoral study at the Department of Information and Biological Science, Graduate School of Natural Sciences, Nagoya City University.

I would like to express my heartfelt gratitude to my supervisor Prof. Yoshinori Kumazawa for his guidance, discussion, and continuous help throughout my study. I am also indebted to Dr. Widodo, who introduced and guided me to the Ph.D study at Nagoya City University. Special thanks go to Drs. D. G. R Wiadnya, Anik M. Hariati, Ating Yuniarti and Mr. Muhammad Fakhri who always encouraged me to conduct research on Indonesian freshwater fishes and provided strong supports. I am also grateful to Mr. Sahat Ratmuangkhwang for his collaboration in the fieldwork and morphological analyses. I acknowledge Dr. Tetsuya Sado for his advice on fishing gears to collect rasboras. My sincere thanks also go to Dr. Prachya Musikasinthorn for instructions on morphological analysis of *Rasbora* and information on rasbora specimens collected in Sumatra and Indo-China.

I cordially thank Indonesian Institute of Sciences (LIPI, *Lembaga Ilmu Pengetahuan Indonesia*) for providing research permissions. I would also like to thank Drs. Renny Kurnia Hadiaty and Harsono for allowing me to inspect *Rasbora* specimens of the Museum Zoologicum Bogoriense (MZB) in Cibinong, Bogor-West Java. I cordially appreciate Dr. Daniel Lumbantobing who kindly gave a lecture on *Rasbora* taxonomy and morphological observation when I accidentally met him at MZB. He also kindly shared information of *R. lateristriata* and its geographical distribution. I sincerely acknowledge Dr. Douglas Nelson and Ms. Anna Barget of the University of Michigan Museum of Zoology (the Great Lakes Invasive Species Project, UMMZ) who kindly provided images of *R. baliensis* and *R.*

lateristriata specimens. My sincere gratitude is extended to Dr. Guillermo Ortí for providing raw data of calibration dating points for divergence time estimation.

This research required me to conduct extensive field sampling in many localities of Borneo, Java, Bali and Lombok Islands of Indonesia to collect more than 500 samples. Without generous assistance from the following persons, my fieldwork would not have been successful: Ms. Aryani Rahmawati; Messrs. Rian Kusuma, Saifur Rahman, Alvin Hamam, Dhimas Wicaksono, Rendy Irawan, Josev Marshal, Yonathan, Joshua, Robby Rahmat, Wujud Arif, Wayan Sudiartha, Kadek, and Komang. Many thanks go to my collaborators Messrs. Mahrus Ali and Supardan for providing fish specimens from Sumatra and Sumbawa Islands. I would like to thank all present and past Kumazawa Lab members, Messrs. Takeshi Fujitani, Masaki Kurisaki, Yao Sun, Ms. Mika Matsushima and Dr. Kazumi Matsubara, for their generous help in lab works and sincere friendship. Scientific advices and assistances from people at the Graduate School of Natural Sciences-Nagoya City University, especially from Prof. Akihiko Moriyama, Ms. Mieko Suzuki, Mr. Yukio Murase and Dr. Manabu Ogiso are gratefully acknowledged.

Finally, I am thankful to my wife, Ken Ratih Probosari and to my family for their continuous support and patience throughout my years in the doctoral course.

References

Alfred, E.R., 1963. Notes on a re-examination of some Bleeker type specimens of Indo-Malayan fresh-water fishes. Part 2. Abramidinae and Rasborinae. Bull. Nat. Mus. Singapore 32, 128-134.

Anderson, S., Bankier, A.T., Barrel, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457-465.

Asahida, T., Kobayashi, T., Saitoh, K., Nakayama, I., 1996. Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffer containing high concentration of urea. Fish. Sci. 62, 727-730.

Avise, J.C., 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge-Massachusetts.

Avise, J.C., 2004. Molecular markers, natural history, and evolution, 2nd edition. Sinauer Asscociates Inc. Sunderland-Massachusetts.

Baillie, J.E.M., Hilton-Taylor, C., Stuart, S.N., (Eds.) 2004. IUCN red list of threatened species^{TM}: a global species assessment. IUCN, Gland, Switzerland and Cambridge.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37-48.

Berra, T.M., 2001. Freshwater fish distribution. Academic Press, San Diego-California.

Betancur-R, R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton II, J.C., Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013. The tree of life and a new classification of bony fishes. PLoS Curr. 5.

Bird, M.I., Taylor, D., Hunt, C., 2005. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? Quatern. Sci. Rev. 24, 2228-2242.

Bleeker, P., 1854. Overzigt der ichthyologische fauna van Sumatra, met beschrijving van eenige nieuwe soorten. Natuurkundig Tijdschrift voor Nederlandsch Indië 7, 49-108.

Bleeker, P., 1860. Ichthyologiae archipelagi indici prodromus. Vol. 2. Cyprini. Batavia.

Briggs, J.C., 1987. Biogeography and plate tectonics: developments in palaeontology and stratigraphy, first ed. Elsevier, New York.

Brittan, M.R., 1954. A revision of the Indo-Malayan fresh-water fish genus *Rasbora*. Inst. Sci. Technol. Manila.

Brittan, M.R., 1972. Rasbora: a revision of the Indo-Malayan fresh-water fish genus *Rasbora*, with natural color photographs. T.F.H Publications, Hong Kong.

Brittan, M.R., 1998. Rasboras: keeping and breeding them in captivity. T.F.H Publications, Neptune City, New Jersey.

Britz, R., Conway, K.W., Rüber, L., 2009. Spectacular morphological novelty in a miniature cyprinid fish, *Danionella dracula* n. sp.. Proc. R. Soc. Lond. B 276, 2179–2186.

Broughton, R.E., Betancur-R, R., Li, C., Ortí, G., 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 5.

Burkhead, N.M., 2012. Extinction rates in North American freshwater fishes, 1900-2010. BioScience 62, 798-808.

Cambray, J.A., Bianco, P.G., 1998. Freshwater fish in crisis, a blue planet perspective. Ital. J. Zool. 65, 345-356.

Chen, W.-J., Bonillo, C., Lecointre, G., 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol. Phylogenet. Evol. 26, 262–288.

Chen, W.-J., Miya, M., Saitoh, K., Mayden, R.L., 2008. Phylogenetic utility of two existing

and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene 423, 125–134.

Collins, R.A., Armstrong, K.F., Meier, R., Yi, Y., Brown, S.D.J., Cruickshank, R.H., Keeling, S., Johnston, C., 2012. Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. PLoS ONE 7, e28281.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

de Bruyn, M., Mather, P.B., 2007. Molecular signatures of Pleistocene sea level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. Mol. Ecol. 16, 4295-4307.

de Bruyn M., Rüber L., Nylinder S., Stelbrink B., Lovejoy N.R., Lavoué S., Tan H.H., Nugroho E., Wowor D., Ng P.K.L., Siti Azizah M.N., von Rintelen T., Hall R., Carvalho G.R., 2013. Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast Asian biodiversity hotspots. Syst. Biol. 62, 398-410.

de Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R., von Rintelen, T., 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. Syst. Biol. 63, 879-901.

Drummond, A.J., Ho, S.Y.W., Philips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969-1973.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006.

Freshwater biodiversity: importance, threats, status and conservational challenges. Biol. Rev. 81, 163-182.

Ebach, M.C., Holdrege, C., 2005. DNA barcoding is no substitute for taxonomy. Nature 343, 697.

Edwards, A.W.F., Cavalli-Sforza, L.L., 1963. The reconstruction of evolution. Ann. Human Genetics 27, 105-106 (also published in Heredity 18, 553).

Eschmeyer, W.N., 2015. Catalog of fishes: genera, species, references. Available from http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp. Accessed on December 15, 2015.

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). Zool. Scr. 38, 237-256.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368-376.

Frankham, R., 2003. Genetics and conservation biology. C. R. Biol. 326, S22-S29.

Frankham, R., 2005. Genetics and extinction. Biol. Conserv. 126, 131-140.

Froese, R., Pauly, D., (Eds.) 2015. FishBase. World Wide Web electronic publication. Available from http://fishbase.org/. Accessed on December 20, 2015.

Groom, M.J., 2006. Threats to biodiversity, in: Groom, M.J., Meffe, G.K., Carroll, C.R. (Eds.) Principles of Conservation Biology, third ed. Sinauer Associates, Massachusetts, pp. 63-110.

Hamilton, A.J., Basset, Y., Benke, K.K., Grimbacher, P.S., Miller, S.E., Novotný, V., Samuelson, G.A., Stork, N.E., Weiblen, G.D., Yen, J.D.L., 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. Am. Nat. 176, 90-95.

Hall, R., 2008. Continental growth at the Indonesian margins of Southeast Asia, in: Spencer, J.E., Titley, S.R. (Eds.) Ores and Orogenesis: Circum Pacific Tectonics, Geologic Evolution and Ore Deposits. Arizona Geological Society Digest, pp. 245-258.

Hall, R., 2009. Southeast Asia's changing palaeogeography. Blumea 54, 148-161.

Hall, R., 2013. The palaeogeography of Sundaland and Wallacea since the late Jurassic. J. Limnol. 72, 1-17.

Hall, R., Cottam, M.A., Wilson, M.E.J., 2011. The SE Asian gateway: history and tectonics of the Australia-Asia collision. Geological Society, Special Publications 355, London, pp. 1-6.

Hall, R., Morley, C.K., 2004. Sundaland basins, in: Clift, P., Wang, P., Kuhnt, W., Hayes, D.E. (Eds.), Continent-Ocean Interactions Within the East Asian Marginal Seas. American Geophysical Union, Geophysical Monograph. Washington D.C., pp. 55-85.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identification through DNA barcodes. Proc. R. Soc. Lond. B. 270, 313-321.

Hebert, P.D.N., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 852-859.

Hickerson, M.J., Meyer, C.P., Moritz, C., 2006. DNA barcoding will often fail to discover new animal species over broad parameter. Syst. Biol. 55, 729-739.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3, e2490.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001. A mitogenomic perspective on the basal teleostean phylogeny: resolving higher level relationships with longer DNA sequences. Mol. Phylogenet. Evol. 20, 275-285.

Jonniaux, P., 2014. Getmitogenome: a software for manipulating animal mitochondrial gene sequences. Accessed on December 24, 2014.

Kottelat, M., 1984. A new *Rasbora* s.l. (Pisces: Cyprinidae) from northern Thailand. Rev. Suisse Zool. 91, 717-723.

Kottelat, M., 2005. *Rasbora notura*, a new species of cyprinid fish from the Malay Peninsula (Teleostei: Cyprinidae). Ichthyol. Explor. Freshw. 16, 265-270.

Kottelat, M., Tan, H.H., 2011. *Rasbora atranus*, a new species of fish from central Borneo (Teleostei: Cyprinidae). Ichthyol. Explor. Freshw. 22, 215-220.

Kottelat, M., Vidthayanon, C., 1993. *Boraras micros*, a new genus and species of minute freshwater fish from Thailand (Teleostei: Cyprinidae). Ichthyol. Explor. Freshw. 4, 161-176.

Kottelat, M., Whitten, A.J., 1996. Freshwater fishes of western Indonesia and Sulawesi: additions and corrections. Periplus Editions, Jakarta.

Kottelat, M., Whitten, A.J., Kartikasari, S.N., Wirjoatmodjo, S., 1993. Freshwater fishes of western Indonesia and Sulawesi. Periplus Editions, Hong Kong.

Kumazawa, Y., Nishida, M., 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37, 380-398.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695-1701.

Lara, A., de León, J.L.P., Rodríguez, R., Casane, D., Côté, G., Bernatchez, L., Garcia-Machado, E., 2010. DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. Mol. Ecol. Resour. 10, 421-430.

Liao, T.Y., Kullander, S.O., Fang, F., 2010. Phylogenetic analysis of the genus *Rasbora* (Teleostei: Cyprinidae). Zool. Scr. 39, 155-176.

Liao, T.Y., Kullander, S.O., Fang, F., 2011. Phylogenetic position of rasborin cyprinids and monophyly of major lineages among the Danioninae, based on morphological characters (Cypriniformes: Cyprinidae). J. Zool. Sys. Evol. Res. 49, 224-232.

Lohman, D. J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P. K. L., Shih, H.-T., Carvalho, G.R., von Rintelen, T., 2011. Biogeography of the Indo-Australian archipelago, Ann. Rev. Ecol. Syst. 42, 205–206.

López, J.A., Chen, W.-J., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Loy, A., Mariani, L., Bertelletti, M., Tunesi, L., 1998. Visualizing allometry: geometric morphometrics in the study of shape changes in the early stages of the two-banded sea bream, *Diplodus vulgaris* (Perciformes, Sparidae). J. Morphol. 237, 137-146.

Lumbantobing, D.N., 2010. Four new species of the *Rasbora trifasciata*-group (Teleostei: Cyprinidae) from northwestern Sumatra, Indonesia. Copeia 4, 644-670.

Lumbantobing, D.N., 2014. Four new species of *Rasbora* of the Sumatrana group (Teleostei: Cyprinidae) from northern Sumatra, Indonesia. Zootaxa 3764, 1-25.

May, R., 2010. Tropical arthropod species, more or less?. Science 329, 41-42.

Mayden, R.L., Tang, K.L., Conway, K.W., Freyhof, J., Chamberlain, S.J., Haskins, M., Schneider, L.M., Sudkamp, M., Wood, R.M., Agnew, M.K., Bufalino, A., Sulaiman, Z., Miya, M., Saitoh, K., He, S., 2007. Phylogenetic relationships of *Danio* within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. J. Exp. Zool. B: Mol. Dev. Evol. 308B, 642–654.

Metcalfe, I., Smith, J.M.B., Morwood, M., Davidson, I., 2001. Faunal and floral migrations and evolution in SE Asia-Australasia. Balkema, Lisse.

Mittermeier, R.A., 1988. Primate diversity and the tropical forest: case studies from Brazil and Madagascar and the importance of the Megadiverse Countries. In: Wilson, E.O., Peters, F.M. (Eds.) Biodiversity. National Academy Press, Washington D.C., pp. 145-154.

Mittermeier, R.A., Gil, P.R., Mittermeier, C.G., 1997. Megadiversity: earth's biologically wealthiest nations. CEMEX, Mexico City.

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. Mar Biotechnol. 1, 416-426.

Mokodongan, D.F., Yamahira, K., 2015. Origin and intra-island diversification of Sulawesi endemic Adrianichthyidae. Mol. Phylogenet. Evol. 93, 150-160.

Monk, K.A., De Fretes, Y., Reksodiharjo-Lilley, G., 1997. The ecology of Nusa Tenggara and Maluku. Periplus Editions, Hong Kong.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B., 2011. How many species are there on earth and in the ocean?. PLoS Biol. 9, e1001127.

Moss, S.J., Wilson, M.E.J., 1998. Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. in: Hall, R., Holloway, J.D. (Eds.) Biogeography and geological evolution of SE Asia. Backhuys, Leiden, pp. 133-163.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853-858.

Nei, M., Kumar, S., 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Oijen, M.J.P., Loots, G.M.P., 2012. An illustrated translation of Bleeker's fishes of the Indian Archipelago part II Cyprini. Zool. Med. 86, 1-469.

Olden, J.D., Kennard, M.J., Laprieur, F., Tedesco, P.A., Winemiller, K.O., Berthou, E.G., 2010. Conservation biogeography of freshwater fishes: recent progress and future challenges. Diversity Distrib. 16, 496-513.

Rainboth, W.J., 1996. The taxonomy, systematics and zoogeography of *Hypsibarbus*, a new genus of large barbs (Pisces, Cyprinidae) from the rivers of Southeastern Asia. University of California Press, La Jolla.

Rainboth, W.J., Vidthayanon, C., Yen, M.D., 2012. Fishes of the greater Mekong ecosystem with species list and photograph atlas. Misc. Publ. Mus. Zool. Univ. Mich. No 201.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2015. Tracer v1.6.0, Available from http://beast.bio.ed.ac.uk/Tracer. Accessed on December 29, 2015.

Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304-311.

Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS ONE 8, e66213.

Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59, 2299-2311.

Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4-14.

Reichard, M., Jurajda, P., 1999. Patterns of ontogenetic changes in relative growth in the precocial cyprinid, bitterling (*Rhodeus sericeus*). Neth. J. Zool. 49, 111-124.

Reid, G.M., MacBeath, T.C., Csatádi, K., 2013. Global challenges in freshwater-fish conservation related to public aquariums and the aquarium industry. Int. Zoo Yb. 47, 6-45.

Roberts, T.R., 1993. The freshwater fishes of Java, as observed by Kuhl and van Hasselt in 1820-23. Zool. Verh. 285, 1-94.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Rosso, J.J., Mabragaña, E., Castro, M.G., Díaz de Astarloa, M., 2012. DNA barcoding Neotropical fishes: recent advances from the Pampa Plain, Argentina. Mol. Ecol. Resour. 12, 999-1011.

Rüber, L., Kottelat, M., Tan, H.H., Ng, P.K.L., Britz, R., 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. BMC Evol. Biol. 7, 38.

Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

Secretariat Cabinet of the Republic of Indonesia, 1999. Government Regulations of the Republic of Indonesia number 7 year 1999 (PP No.7/1999) concerning preservation of plant and animal species. Jakarta.

Sri Paryanti, P., 2006. National aquaculture sector overview. Indonesia. National aquaculture sector overview fact sheets, in: FAO fisheries and aquaculture department (online). Rome. Updated 9 February 2006. Available from http://www.fao.org/fishery/countrysector/naso_indonesia/en. Acessed on January 20, 2016.

Steppan, S.J., Adkins, R.M., Spinks, P.Q., Hale, C., 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Mol. Phylogenet. Evol. 37, 370-388.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725-2729.

Tan, H.H., Kottelat, M., 2009. The fishes of the Batang Hari drainage, Sumatra, with description of six new species. Ichthyol. Explor. Freshw. 20, 13-69.

Tang, K.L., Agnew, M.K., Hirt, M.V., Sado, T., Schneider, L.M., Freyhof, J., Sulaiman, Z., Swartz, E. Vidthayanon, C., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2010. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). Mol. Phylogenet. Evol. 57, 189-214.

UNEP, 1992. Convention on biological diversity: environmental law and institution programme activity center. Nairobi.

van Hasselt, J.C., 1823. Uittreksel uit een' brief van den Heer J.C. van Hasselt, aan den Heer C.J. Temminck, geschreven uit Tjecande, Residentie Bantam, den 29sten December 1822.-Op. cit., 2, 130-133.

Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river system and time durations. J. Biogeogr. 27, 1153-1167.

Wallace, A.R., 1860. On the zoological geography of the Malay Archipelago. J. Proc. Linn. Soc. 4, 172-184.

Ward, R.D., 2009. DNA barcode divergence among species and genera of birds and fishes. Mol. Ecol. Resour. 9, 1077-1085.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B 360, 1847–1857.

Weber, M., de Beaufort, L.F., 1916. The fishes of Indo-Australian Archipelago III Ostariophysi: II Cyprinoidea, Apodes, Synbranchi. E.J. Brill, Leiden.

Whitten, A.J., Soeriaatmadja, R.E., Afiff, S.A., 1996. The ecology of Java and Bali. Periplus Editions, Hong Kong.

Woodruff, D.S., 2003. Neogene marine transgressions, palaeogeography and biogeographic transition on the Thai-Malay Peninsula. J. Biogeogr. 30, 551-567.

Woodruff, D.S., 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's pattern and the future of the remaining refugial-phase biodiversity. Biodivers. Conserv. 19, 919-941.

World Bank, 2001. Indonesia: environment and natural resources management in a time of transition. Washington D.C.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252-3255.

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13, 303-314.

Yap, S.Y., 2002. On the distributional patterns of the Southeast-East Asian freshwater fish and their history. J. Biogeogr. 29, 1187-119.

Young, M.K., Mckelvey, K.S., Pilgrim, K.L., Schwartz, M.K., 2013. DNA barcoding at riverscape scales: assessing biodiversity among fishes of the genus *Cottus* (Teleostei) in northern Rocky Mountain streams. Mol. Ecol. Resour. 4, 583-595.

Yu, Y., Harris, A.J., Blair, C., He, X.J., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46-49.

Zheng, Y., Peng, R., Kuro-o, M., Zeng, X., 2011. Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (Order Caudata). Mol. Biol. Evol. 28, 2521-2535.

Zwickl D.J., 2014. Genetic Algorithm for Rapid Likelihood Inference. Version 2.0. http://www.bio.utexas.edu/faculty/antisense/garli/garli.html. Accessed on August 4, 2014.

List of scientific publications

1. Wahyu Endra Kusuma, Sahat Ratmuangkhwang and Yoshinori Kumazawa. 2016. Molecular phylogeny and historical biogeography of the Indonesian freshwater fish *Rasbora lateristriata* species complex (Actinopterygii: Cyprinidae): Cryptic species and west-to-east divergences. Molecular Phylogenetics and Evolution 105, 212-223.
2. Wahyu Endra Kusuma and Yoshinori Kumazawa. Complete mitochondrial genome sequences of two Indonesian rasboras (*Rasbora aprotaenia* and *Rasbora lateristriata*). Mitochondrial DNA (in press).

List of conference presentations

1. Wahyu Endra Kusuma and Yoshinori Kumazawa
   Molecular phylogeny of Javanese freshwater rasbora fishes.
   Poster presentation at the 49th annual meeting of the Ichthyological Society of Japan, Gifu University, Gifu, Japan, 24-25 September, 2016.
2. Wahyu Endra Kusuma and Yoshinori Kumazawa
   Molecular phylogeography of the endemic freshwater fish *Rasbora baliensis* (Actinopterygii: Cyprinidae).
   Poster presentation at the International Symposium: "Biodiversity and environmental medicine in south and east Asia based on molecular biology 2015", Nagoya City University, Nagoya, Japan, 16-17 July, 2015.
3. Wahyu Endra Kusuma and Yoshinori Kumazawa
   Molecular phylogeography of the vulnerable *Rasbora baliensis* (Actinopterygii: Cyprinidae).
   Oral talk at the 47th annual meeting of the Ichthyological Society of Japan, Kanagawa Prefectural Museum of Natural History, Odawara, Japan, 15 November, 2014.