



Molecular Genetic and Taxonomic Studies of the
Swamp Eel (*Monopterus albus* Zuiew 1793)

Tuty Arisuryanti, B.Sc., M.Sc.

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Research Institute for the Environment and Livelihoods
Charles Darwin University, Northern Territory, Australia

Accepted February 2016



Candidate Declaration

I, Tuty Arisuryanti, hereby declare that the work herein, now submitted as a thesis for the degree of Doctor of Philosophy of the Charles Darwin University, is the result of my own investigations, and all references to ideas and work of other researchers have been specifically acknowledged. I hereby certify that the work embodied in this thesis has not already been accepted in substance for any degree, and is not being currently submitted in candidature for any other degree.

Full name : Tuty Arisuryanti

Signature :

Date : 12 February 2016

Abstract

The swamp eel *Monopterus albus* is restricted to fresh water and readily adapts to rice field environments. The swamp eel is popular as food, is readily transported live and has achieved a wide distribution in tropical and subtropical Asia. The principal aim of this study is to investigate the taxonomy and phylogeny of *M. albus* using molecular genetic data, by analysing genetic diversity throughout its distribution, with a special focus on populations from Indonesia.

Evidence for significant cryptic speciation within *M. albus* was found through DNA barcoding using the COI mitochondrial DNA fragment and a population genetic analysis using five microsatellite markers. These data reveal two distinct Indonesian forms of *M. albus*, which are significantly divergent from each other in sympatry and from other forms from more northerly Asian countries. The first form is indigenous to Indonesia occurring on the islands of Java and several adjacent islands to the east and the second form is wide spread in Indonesia and in other Southeast Asian countries to the north, but may have had its distribution modified through human-mediated introductions.

Integrated analyses of nucleotide sequences from multiple studies and gene regions indicate that *M. albus* is highly diverse genetically and comprises at least five cryptic species. Phylogenetic analyses indicate that the southern forms are more recently evolved from older lineages established in northern Asia (e.g. China and Japan).

Morphological variation in the two forms of Indonesian *Monopterus* indicate significant size-related variation and phenotypic plasticity. Characters were found that correlate with the molecular data thereby supporting the presence of two cryptic species of *Monopterus* in Indonesia and providing a basis for their identification in the field.

These findings contribute to the understanding of *M. albus* taxonomy through the discovery of cryptic species in what is best referred to as the *M. albus* species complex. Further, phylogenetic analyses provide a framework for understanding the evolutionary diversification and biogeography of this important group of freshwater fishes from eastern and Southeast Asia.

Contents

Candidate Declaration	ii
Abstract	iii
Contents	v
List of Tables	vii
List of Figures	ix
Appendices	xi
Acknowledgement	xii
CHAPTER 1	
General Introduction	1
1.1. Background	1
1.2. Taxonomic History of Swamp Eels	4
1.2.1. Family Synbranchidae	5
1.2.2. Genus <i>Monopterus</i>	12
1.2.3. The <i>Monopterus albus</i> species complex	18
1.3. Molecular Markers	22
1.3.1. DNA Sequencing	22
1.3.2. Denaturing Gradient Gel Electrophoresis	24
1.3.3. Restriction Fragment Length Polymorphism	25
1.3.4. Microsatellite	26
1.4. Application of Molecular Markers	27
1.4.1. Taxonomy and Evolution.....	27
1.4.2. Phylogenetic Reconstruction	31
1.4.3. Principles of Biological Classification and Numerical Taxonomy.....	33
1.4.4. Molecular Genetic Studies of <i>Monopterus albus</i> species complex..	39
1.5. Research Objectives and Thesis Format	44
CHAPTER 2	
Molecular genetic evidence for cryptic speciation in the <i>Monopterus albus</i> species complex in central Indonesia	46
2.1. Introduction	46
2.2. Materials and Methods	51
2.2.1. Sample collection for COI mitochondrial sequencing	51
2.2.2. DNA extraction, amplification and sequencing of cytochrome oxydase c subunit I (COI) mitochondrial gene: procedures and analysis....	51
2.2.3. Sample collection for Denaturation Gradient Gel Electrophoresis (DGGE) of COI variation.....	53
2.2.4. DNA amplification, optimization and procedures for Denaturation Gradient Gel Electrophoresis (DGGE) of COI variation.....	54
2.2.5. Sample collection and Restriction Fragment Length Polymorphism (RFLP) amplification procedures prior to microsatellite analysis	55
2.2.6. Microsatellite procedures and analysis	56
2.3. Results	58
2.3.1. COI mitochondrial sequence variation and phylogenetic relationships	58
2.3.2. Variation within clades	60
2.3.3. DGGE analysis of COI variation	61
2.3.4. Microsatellite variation and patterns of divergence	62
2.4. Discussion	66

CHAPTER 3	
Molecular taxonomy and evolution of the <i>Monopterus albus</i> complex: An integrated analysis of nucleotide data from multiple studies and gene regions	92
3.1. Introduction	92
3.2. Materials and Methods	95
3.2.1. Sample collection and storage	95
3.2.2. DNA extraction, PCR amplification and sequencing	96
3.2.3. Sequence alignment and phylogenetic analysis	97
3.3. Results	99
3.3.1. Molecular genetic variation and phylogenetic analysis of two 16SrDNA mitochondrial gene fragments	99
3.3.2. Molecular genetic variation and phylogenetic analysis of multiple gene regions	103
3.4. Discussion	105
CHAPTER 4	
Morphological variation and numerical taxonomy of <i>Monopterus albus</i> species in Indonesia	139
4.1. Introduction	139
4.2. Materials and Methods	142
4.2.1. Sample collections	142
4.2.2. Meristic and morphometric procedures	143
4.2.3. Morphological data and statistical analysis	145
4.2.4. Multivariate statistical analysis	147
4.3. Results	149
4.3.1. Vertebrae number	149
4.3.2. Morphometric data	150
4.3.2.1. Principal Component Analysis (PCA)	150
4.3.2.2. Discriminant Function Analysis (DFA)	153
4.4. Discussion	156
CHAPTER 5	
General Discussion	194
5.1. Global taxonomy and evolution of <i>Monopterus albus</i> complex	195
5.2. Genetic variation of the Indonesian <i>Monopterus albus</i> complex	197
5.3. Morphological variation and taxonomy	199
5.4. Further studies	200
Appendices	205
Bibliography	216

List of Tables

Table 1.1.	The currently recognised members of Synbranchidae and their distributions.....	6
Table 1.2.	Summary of the taxonomic classification of species of <i>Monopterus</i>	17
Table 2.1.	Sampling location, sample code, geographic reference and sample size of <i>M. albus</i> . Samples of <i>Ophisternon</i> are used as outgroups.....	71
Table 2.2.	Details of polymorphic loci, primer sequence, fragment size and dye.....	74
Table 2.3.	Mean percentage nucleotide divergence of a 514 bp fragment of the COI mitochondrial gene among four identified clades of <i>M. albus</i> in this study.....	77
Table 2.4.	Grouping of haplotype of <i>M. albus</i> based on the phylogenetic analysis of the COI mitochondrial gene fragment	78
Table 2.5.	Genetic variability of five microsatellite loci in <i>M. albus</i> haplotype A (18 populations).....	83
Table 2.6.	Genetic variability of five microsatellite loci in <i>M. albus</i> haplotype B (6 populations)	84
Table 2.7.	Genetic variability of five microsatellite loci in <i>M. albus</i> admixture populations (4 populations)	85
Table 2.8.	Pairwise F_{ST} values between <i>M. albus</i> populations in Indonesia based on five microsatellite loci	86
Table 2.9.	Summary of AMOVA	87
Table 2.10.	Result of assignment test (self classification) of <i>M. albus</i> based on five microsatellite loci	88
Table 3.1.	Synonyms of <i>M. albus</i> based on Rosen and Greenwood (1976), Fishbase and Eschmeyer (2014)	112
Table 3.2.	Samples used in this study and the previous studies with sampling location, population code and GenBank accession number	114
Table 3.3.	Primers used for mitochondrial and nuclear genes for <i>M. albus</i> in this study	119
Table 3.4.	Mean percentage nucleotide sequence divergence of a 510-bp fragment of the 16S mitochondrial gene among five identified clade of <i>M. albus</i> from this study combined with Collins <i>et al.</i> (2002) and Miya <i>et al.</i> (2001) data	121
Table 3.5.	Haplotype of <i>M. albus</i> based on 16S mitochondrial gene region inferred from this study combined with Collins <i>et al.</i> (2002) and Miya <i>et al.</i> (2001) data.....	122
Table 3.6.	Mean percentage nucleotide sequence divergence of an 539-bp fragment of the 16S mitochondrial gene among five identified clade of <i>M. albus</i> of this study combined with Matsumoto <i>et al.</i> (2010) and Miya <i>et al.</i> (2001) data	127
Table 3.7.	Haplotype of <i>M. albus</i> based on 16S mitochondrial gene region inferred from this study combined with Matsumoto <i>et al.</i> (2010) and Miya <i>et al.</i> (2001) data	128
Table 3.8.	Estimates of genetic diversity for <i>M. albus</i> based on two partial 16S mitochondrial sequence data set.....	132
Table 3.9.	Mean percentage nucleotide sequence divergence of a 3,366-bp fragment of the multiple gene regions among four identified clade of <i>M. albus</i>	134
Table 3.10.	Haplotype of <i>M. albus</i> based on mitochondrial and nuclear gene regions	135
Table 3.11.	Estimates of genetic diversity from <i>M. albus</i> on partial mitochondrial and nuclear sequences from the multigene data set	136

Table 4.1.	Sampling location, date collection, number of specimens, total length and voucher catalogue numbers.....	162
Table 4.2.	Definitions of meristic counts and morphometric measurements and their codes <i>M. albus</i> used in this study	165
Table 4.3.	Distribution of vertebrae number frequency in seven populations of <i>M. albus</i>	167
Table 4.4.	Summary of morphometric for data for 13 variables measured in millimetres (mean±SD) for seven populations of <i>M. albus</i>	168
Table 4.5.	Correlation coefficients for TL against 12 morphometric variables each haplotype of <i>M. albus</i> and for the pooled data	169
Table 4.6.	Principal component analysis of 13 log transformed morphometric characters for <i>M. albus</i> from seven populations	173
Table 4.7.	Correlation coefficients (r) for untransformed, Log10 transformed morphometric characters and ratios with Total Length for six population of <i>M. albus</i>	175
Table 4.8.	Principal Component analysis of 13 log10 transformed morphometric characters for six populations of <i>M. albus</i>	176
Table 4.9.	Results for a Principal Components analysis of 12 ratio-transformed morphometric characters for <i>M. albus</i> from six populations	178
Table 4.10.	Principal Component analysis of 12 size-transformed morphometric variables for six populations of <i>M. albus</i>	180
Table 4.11.	Discriminant Function analysis of 13 log10 transformed morphometric variables for six <i>M. albus</i> populations	182
Table 4.12.	Discriminant function analysis of 12 ratio transformed morphometric characters for six populations of <i>M. albus</i>	184
Table 4.13.	Discriminant function analysis of 12 size transformed morphometric variables for six populations of <i>M. albus</i>	186
Table 4.14.	Two group Discriminant Function analyses of 13 log transformed and 12 ratio transformed morphometric characters for seven populations of <i>M.albus</i>	188
Table 4.15.	Classification results of discriminant function analysis using log and ratio transformed morphometric variable from seven populations of <i>M.albus</i> based on the grouping of the two haplotypes.....	190
Table 4.16.	The two-group Discriminant Function analyses of untransformed morphometric and meristic characters for 88 <i>M.albus</i> from seven populations of <i>M. albus</i>	191

List of Figures

Figure 1.1.	Distribution of the family Synbranchidae.....	8
Figure 1.2.	Type localities for valid <i>Monopterus</i> species in Asia.....	13
Figure 1.3.	Distribution map of <i>M.albus</i>	19
Figure 2.1.	Map sampling collection sites for <i>M.albus</i> samples in Indonesia, Vietnam and Taiwan.....	73
Figure 2.2.	Bayesian tree inferred from COI mitochondrial gene sequences....	75
Figure 2.3.	Neighbour-joining tree based on COI sequence data using Kimura two parameter (K2P) substitution model.....	76
Figure 2.4.	Summary of nucleotide variations in the partial COI-mtDNA of <i>M.albus</i> in this study	79
Figure 2.5.	Principal Coordinate Analysis (PCoA) of pairwise genetic distances of 19 haplotypes of <i>M.albus</i> populations based on COI mitochondrial sequences.....	80
Figure 2.6.	Distribution of haplotype patterns of DGGE of COI variation in 15 populations in Indonesia.....	81
Figure 2.7.	Different haplotype pattern (A and B) of <i>M.albus</i> showed by Spel restriction enzyme using RFLP.....	82
Figure 2.8.	Relationships between populations defined by mitochondrial haplotype using Principal Coordinate Analysis of <i>M.albus</i> population microsatellite frequencies.....	89
Figure 2.9.	Principal Coordinate Analysis of 588 <i>M.albus</i> individuals genotyped at five microsatellite loci	90
Figure 2.10.	Results from the admixture analysis obtained with the program STRUCTURE under the admixture and correlated allele frequencies model.....	91
Figure 3.1.	Map distribution of <i>M.albus</i> species complex based on museum records and previous studies	113
Figure 3.2.	Bayesian tree inferred from 16S mitochondrial gene sequences combining between this study with Collins <i>et al.</i> (2002) and Miya <i>et al.</i> (2001) data	120
Figure 3.3.	Summary of nucleotide variations in the partial 16SrDNA of <i>M. albus</i> in this study combined with the previous study (Collins <i>et al.</i> , 2002 and Miya <i>et al.</i> , 2001).....	124
Figure 3.4.	Principal Coordinat Analysis (PCoA) of pairwise genetic distances of 23 haplotypes of <i>M. albus</i> populations based on 16S mitochondrial sequences	125
Figure 3.5.	Bayesian tree inferred from 16S mitochondrial gene sequences combining between this study with Matsumoto <i>et al.</i> (2010) and Miya <i>et al.</i> (2001) data.....	126
Figure 3.6.	Summary of nucleotide variations in the partial 16SrDNA of <i>M. albus</i> in this study combined with the previous study (Matsumoto <i>et al.</i> , 2010 and Miya <i>et al.</i> , 2001)	130
Figure 3.7.	Principal Coordinate Analysis (PCoA) of pairwise genetic distances of 24 haplotypes of <i>M. albus</i> populations based on 16S mitochondrial sequences	131
Figure 3.8.	Bayesian tree inferred from mitochondrial and nuclear gene sequences.....	133
Figure 3.9.	Principal Coordinate Analysis (PCoA) of pairwise genetic distances of 29 haplotypes of <i>M. albus</i> populations based on mitochondrial and nuclear sequences.....	137
Figure 3.10.	Map distribution of putative species of the <i>M. albus</i> species complex from integrated studies and gene regions	138
Figure 4.1.	Morphology of <i>M. albus</i> haplotype A	163
Figure 4.2.	Morphology of <i>M. albus</i> haplotype B	164

Figure 4.3.	Morphometric measurements of <i>M. albus</i>	166
Figure 4.4.	Scatterplot for total length versus vertebrae count for seven populations of <i>M.albus</i>	170
Figure 4.5.	Vertebrae frequency histogram for <i>M. albus</i>	171
Figure 4.6.	Vertebrae of <i>M. albus</i>	172
Figure 4.7.	Plot of Principal Component scores for the first three axes for seven populations of <i>M.albus</i> based on 13 log ₁₀ transformed morphometric variables	174
Figure 4.8.	Plot of Principal Component scores for the first three axes for six populations of <i>M.albus</i> based on 13 log ₁₀ transformed morphometric variables	177
Figure 4.9.	Plot of Principal Component scores for the first three axes for six populations of <i>M.albus</i> based on 12 ratio-transformed morphometric variables	179
Figure 4.10.	Plot of Principal Component scores for the first three axes for six populations of <i>M. albus</i> based on 12 size transformed morphometric variables	181
Figure 4.11.	Plot of Discriminant Function scores for the first three axes for six populations of <i>M.albus</i> based on 13 log ₁₀ transformed morphometric variables	183
Figure 4.12.	Plot of Discriminant Function scores for the first three axes for six populations of <i>M.albus</i> based on 12 ratio transformed morphometric variables	185
Figure 4.13.	Plot of Discriminant Function scores for the first three axes for six populations of <i>M.albus</i> based on 12 size transformed morphometric variables	187
Figure 4.14.	Histograms of discriminant scores from two group Discriminant Function analyses of 13 log transformed (A) and 12 ratio transformed (B) morphometric characters for seven populations of <i>M. albus</i>	189
Figure 4.15.	Histograms of discriminant scores from two group Discriminant Function analyses of 14 log transformed (A) and untransformed (B) morphological variables, and for the two best discriminating variables (C) for seven populations of <i>M. albus</i>	192
Figure 4.16.	Bivariate plot of the two best discriminating variables, snout tip to eye length and vertebrae number identified from the DFA of 14 untransformed variables.....	193
Figure 5.1.	Principal geographic features of the region covered in Southeast Asia reconstruction	203
Figure 5.2.	Map of Pleistocene sea levels in Southeast Asia and Austral-Asia	204

Appendices

Appendix 1.	Allele frequency of <i>M. albus</i> haplotype A (18 populations)	205
Appendix 2.	Allele frequency of <i>M. albus</i> haplotype B (6 populations).....	207
Appendix 3.	Allele frequency of <i>M. albus</i> admixture populations (4 populations)	208
Appendix 4.	Sequence of 16S mitochondrial gene (primer 16Sar and 16Sbr) of <i>Ophisternon gutturale</i> and <i>Ophisternon sp.</i>	209
Appendix 5.	Sequence of 16S mitochondrial gene (primer L1567 and H2196) of <i>Ophisternon gutturale</i> and <i>Ophisternon sp.</i>	210
Appendix 6.	Sequence of COI mitochondrial gene of <i>Ophisternon gutturale</i> and <i>Ophisternon sp.</i>	211
Appendix 7.	Sequence of RAG1 nuclear gene of <i>Ophisternon gutturale</i> and <i>Ophisternon sp.</i>	212
Appendix 8.	Sequence of 1 st intron of S7 nuclear gene of <i>Ophisternon gutturale</i> and <i>Ophisternon sp.</i>	215

ACKNOWLEDGEMENTS

I am very grateful to the Directorate General of Higher Education (DITJEN DIKTI), Republic of Indonesia for granting me the scholarship to carry out my study. Financial support for this research was also provided by Charles Darwin University from grants awarded to Professor Christopher Austin. I am also indebted to Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia, and the Museum and Art Gallery of the Northern Territory (MAGNT), Australia for providing facilities to support my research.

This study would not have been possible without the support and encouragement of many people. Firstly, I would like to express my deep and sincere gratitude to my principal supervisor, Professor Christopher M. Austin, for his support, valuable advice, guidance, and encouragement throughout my entire course of PhD study. He has not only broadened my knowledge on the application of molecular genetics, especially in the fields of taxonomy and evolutionary studies but also inspired and motivated me to develop and improve my learning of research methods. I am also grateful to my former associate supervisors, Dr. Nu-Wei Vivian Wei and Dr. Mark Schultz, and my associate supervisor, Dr. Michael Hammer, for their valuable advice, guidance, and assistance while working in the laboratory, analysing the data, and preparing my thesis at the final stage of my study. Dr. Penny Wurm provided support as administrative supervisor in the final stage of my candidature.

Secondly, I would like to express my deep gratitude to Dean Faculty of Biology UGM (Dr. Suwarno Hadisusanto, S.U.), Dr. Retno Peni Sancayaningsih (former

Dean Faculty of Biology UGM), Dr. Budi Setiadi Daryono, M.Agr.Sc (Vice-Dean of Academic Affairs Faculty of Biology UGM), Professor Langkah Sembiring (former Vice-Dean of Academic Affairs Faculty of Biology UGM), Professor Sukarti Moeljopawiro, Professor Jusup Subagja, and Dr. Niken Satuti Nur Handayani who recommended and encouraged me to apply for 'DIKTI' scholarship to study at Charles Darwin University (CDU).

Appreciations and thanks go to Dr. Mark Schultz for his assistance and support in relation to laboratory work in Arafura Timor Research Facility (ATRF) and Dr. Barry Russell and Dr. Michael Hammer for his help and support in relation to access to research facilities at MAGNT. My sincere thanks to Dave Wilson, Dean Thorburn, Dr. Binh Thanh Thai, Dr. Mark Schultz, Dr. Nu-Wei Vivian Wei, Rury Eprilurahman, Intan Ratna Pratiwi, Tony Kurniawan, Anjar Tri Wibowo, Marlina, Dessy Nurul Astuti, Agista Zaziroh, Ign. Hardaningsih, Achmad Fanani, Edy Kurniawan, Eko Sigit Purnomo, Djoko Musiantoro, Evy Suprianto, and Bambang Setyadi who helped me to collect eel samples. I would also like to express my thanks to my colleagues, both past and present, from Professor Austin's research group, who provided me with various assistance and support in relation to laboratory work and discussion of systematic and evolutionary concepts. These included Mäia Berman, Shane Penny, Florencia Cerruti, Rury Eprilurahman, Emma Francis and Chelzie Darussalam.

My appreciations and my warm thanks go to Monica Turkey, Director of International Strategy and Development and all of the staffs at International Support CDU for their great work supporting International students including me. My sincere thanks extend to Jayshree Mamtora and all CDU library staffs for

their valuable advice and friendly help in finding articles and books needed in this study. I would like also to express my gratitude to John Logan, Tatiana Porter, all ACL (Australian Centre of Languages) staff who provide orientation and English language support. My warm thanks extend to Ruth Warwick and all academic support staff for providing excellent training for writing skills.

Next, I owe a special debt of gratitude for the support and encouragement of my colleagues at Faculty of Biology UGM, my DIKTI's friends who study at CDU (Aam, Indra, Deasy, Meika, Ratih, Ratri, Frederika, Uud, Silvia, Rahmat, Evi, Hanan, Razaq, Arum and Rury), my friends who stay together with me at International House Darwin at CDU, Darwin Muslim Community and Indonesian Community in Darwin. They let me be part of a happy family in Australia.

I would like to extend my gratitude to Dr. Bronwyn Myer and her family, Dr. Kiki Detmers, Sarah Hobgen, Yuwana Podin, Wajid Javed, Theofransus Litaay, Muhammad Salman Quddus, Rabia Tabassum, Aurelia, Ria Fitriana and Odete Guterres for their friendship and all help given to me while I am undertaking this study and during my period in Darwin, Australia.

I owe my loving thanks to my deceased father, Ilham Zainudin, and my mother, Mudjiati, my sisters and brother and their family (Siti Nuryaningsih, Jumadi Prasanto, Yulia Sekar Permata, Mohammad Priyambodo, Istaning Wardani, Nabilla Asyifa Putri, Faza, and Rahayuningtyas) for their prayers, encouragement and loving support. My special thanks extend to Farouk Ra'djai, Meirza Reynaldi and Sikka Laksitorini for their spiritual advice during difficult time periods in my study. I would also express my gratitude to all of the people whom I did not mention here but assisted me in many ways during my studies.

Lastly, I dedicate this thesis to my beloved daughter, **Astri Puspita Pertiwi Setyadi**, who was always patient during the difficult periods with me and who always prayed and encouraged me to finish my studies.

CHAPTER 1

General Introduction

1.1 Background

Indonesia, with nearly 18,000 islands, is one of the major mega biodiversity countries in the world, second only next to Brazil (Rhee *et al.*, 2004; Allen and Erdmann, 2012). The archipelago is on a crossroad between two oceans, the Pacific and the Indian, and bridges two continents, Asia and Australia. The country contributes significantly to two of the world's 25 biodiversity "hotspots", Sundaland and Wallacea, as recognised by Conservation International (<http://www.conservation.org/where/priorityareas/hotspots/asia/pacific/Pages/asiapacific.aspx>, accessed 15 March 2013). The Sundaland hotspot covers the western half of the Indo-Malayan Archipelago, and is dominated by Borneo and Sumatra. The Wallacea hotspot encompasses Sulawesi, the Moluccas, and the lesser Sundas. Both hotspots are considered to be the centres of species diversity for a number of major groups of plants, mammals, birds, amphibians and reptiles, insects, fishes and marine invertebrates (Rhee *et al.*, 2004). For fishes, the total number of freshwater species is about 1,400 or 7% of total global diversity (Kottelat *et al.*, 1993; Rhee *et al.*, 2004). The number of freshwater species could well be significantly increased with further systematic sampling and associated taxonomic studies.

Freshwater fishes are one of the most important sources of protein in Indonesia. Like other countries over exploitation of wild fish stocks and environmental degradation has meant that the demand for fish protein can no

longer be met from traditional fishing. Thus aquaculture production and sustainable management of fish stocks that have not collapsed are of critical importance for food security and enhanced income (Dunham, 2004; Chauhan and Rajiv, 2010; Morin *et al.*, 2010).

Molecular genetic information is now widely used to study aspects of the ecology or population structure of many species and being considered essential to assist with the sustainable exploitation, management and conservation of commercially important fish species. Such information is also valuable for genetic improvement programs, to further understand population genetic diversity and identify cryptic species, and to help ensure that correct taxonomic nomenclature is applied (Dunham, 2004; Bertorelle *et al.*, 2009; Gum *et al.*, 2009; Aung *et al.*, 2010; Singh *et al.*, 2010; Saltgiver *et al.*, 2012; Thaulow *et al.*, 2012; White and Last, 2012).

Species of the family Synbranchidae are a group of economically important freshwater fishes worldwide. Within this family, species of the genus *Monopterus* are especially popular due to their reputation as delicious food, their ability to survive and grow in poorly oxygenated waters, and as they can be transported live. Approximately 13 species of *Monopterus* are currently recognized, with the majority of species from Africa and Asia, especially India (Rosen and Greenwood, 1976; Bailey and Gans, 1998; Menon, 1999; Gopi, 2002; Nguyen, 2005; Britz *et al.*, 2011; Kottelat, 2013; Eschmeyer, 2015). Of the 12 species presently recognized in Asia, two species are recognised to be native to Indonesia, the swamp eel *Monopterus albus* Zuiew and *M. javanensis* Lacépède. The species, *M. albus*, is character poor and phenotypically plastic, which has lead to an extensive and confusing taxonomic nomenclature. Rosen and Greenwood (1976) found that this species has been described under 13 species names and variously placed within six different genera. However, the

current opinion is that only one polytypic species is present across its large geographic range encompassing western Indonesia, including Sumatra, Java, Bali, Lesser Sundas, Sulawesi, and the Mollucas, and in Southeast Asia extending into India and China (Kottelat *et al.*, 1993; Berra, 2007).

Monopterus albus is found mainly in warm, freshwater environments such as muddy creeks, swamps, canals, and rice fields. It is easily recognised by its distinctive cylindrical snake-like body with tapered tail and small eyes and the absence of scales and fins. The species reaches about 30-100 cm in length with brown or black colouring dorsally and light-brown colouring ventrally. The gill openings are merged into a single slit underneath the head while the mouth is large and protractile and both upper and lower jaws have tiny teeth. The skin produces a thick mucous layer making these eels difficult to hold (Rosen and Greenwood, 1976; Kottelat *et al.*, 1993; Berra, 2007).

In its native habitat *M. albus* is epigeal and mainly nocturnal, consuming a variety of invertebrates such as crustaceans and insects, and sometimes vertebrates including small fish and tadpoles. However, their food habits depend on the presence of other predators. If no other predators are present they become primary predators in ponds, when other predators are present they subsist on smaller aquatic invertebrates (Hill and Watson, 2007; Shafland *et al.*, 2010).

Reproductively the species is a protogynous hermaphrodite and has a distinctive process of sexual development. *Monopterus albus* usually spawn throughout the year. After spawning their eggs are commonly laid in a bubble nest located in shallow waters. All young *M. albus* hatch as females. After spending the first part of their life as females, some eels transform into large males. The change from female to male can take up to a year (Liem, 1963).

Monopterus albus has the ability to crawling on land to migrate. However, their migration are restricted due to the fish often become confined in mud burrow until rains come (Graham, 1997). Therefore, *M. albus* is considered as a non-migratory species with limited tolerance of brackish water, resulting in only limited powers of dispersal other than potentially within continuous catchments (Cai *et al.*, 2008; Matsumoto *et al.*, 2010). A consequence of this life cycle is an increased likelihood of reproductive isolation between populations leading to genetic divergence, and over sufficiently long periods of time the possibility of speciation (Ferguson, 2002). Indeed, this life cycle is thought to be an important factor in accounting for high diversity of the swamp eels worldwide. Thus, even within the Indonesian archipelago, there is a possibility that *M. albus* may consist of several cryptic species. Alternatively, the species may have high levels of genetic similarity among widely separated populations due to recent dispersal facilitated by humans due to their significance for food and their hardiness in relation to transportation and translocation to new aquatic environments. It is also possible that both factors, cryptic speciation and translocation, have created complicated geographic patterns of genetic variation and diversity. Thus, the current taxonomy of *M. albus* may not represent the true biodiversity of this “species” which could represent a complex of cryptic species.

1.2. Taxonomic History of Swamp Eels

The first scientific descriptions of synbranchid swamp eels were by Zuiw (1793) and Bloch (1795). The nomenclature, number of species, and the inter- and intra-generic relationship have long been uncertain and confused. The

taxonomic classification of swamp eels was revised by Regan (1912), Rosen and Rumney (1972), and Rosen and Greenwood (1976) based on morphological and anatomical studies. Although these authors investigated in detail a combination of morphological and osteological features, the taxonomic disputation and confusion relating to these fish has remained. Over more recent years the taxonomic status of swamp eels has been investigated by different authors using molecular markers. However, these studies focused on the status of populations occurring on separate landmasses and islands and used different genes making them difficult to compare (Collins *et al.*, 2002; Perdices *et al.*, 2005; Cai *et al.*, 2008; Valdez-Moreno *et al.*, 2009; Matsumoto *et al.*, 2010; Cai *et al.*, 2012).

1.2.1. Family Synbranchidae

The Synbranchidae is the only family of the Suborder Synbranchoidae established by Swainson (1838) and is currently considered to be represented by 2 subfamilies, 4 genera and 22 species (Table 1.1) (Rosen and Greenwood, 1976; Talwar and Jhingran, 1992; Bailey and Gans, 1998; Menon, 1999; Gopi, 2002; Nguyen, 2005; Britz *et al.*, 2011, Kottelat, 2013; Eschmeyer, 2015). The recognised members of Synbranchidae are widely distributed in tropical and subtropical regions ranging from northern South America, West Africa, East Indies, Indo-Malayan Archipelago, eastern Asia north to Japan, to northern Australia (Figure 1.1) (Rosen and Greenwood, 1976; Kottelat *et al.*, 1993; Pusey *et al.*, 2004; Berra 2007).

Table 1.1 The currently recognised members of Synbranchidae and their distributions.

Subfamily	Genus	Species	Distribution
<i>Macrotreminae</i> Rosen and Greenwood 1976	<i>Macrotema</i> Regan 1906	<i>Macrotema caligans</i> Cantor 1849	Thailand and Malay Peninsula
<i>Synbranchinae</i> Swainson 1838	<i>Synbranchus</i> Bloch 1795	<i>Synbranchus marmoratus</i> Bloch 1795	Northern South America
		<i>Synbranchus madeira</i> Rosen and Rumney 1972	Bolivian Amazon
<i>Ophisternon</i> M'Clelland 1845	<i>Ophisternon</i> M'Clelland 1845	<i>Ophisternon bengalense</i> M'Clelland 1845	Indo-Malayan region, Philippine Islands, New Guinea, India, Indochina
		<i>Ophisternon gutturale</i> Richardson 1844	Northern Australia and Southern New Guinea
		<i>Ophisternon candidum</i> Mees 1962	North West Cape, Western Australia
		<i>Ophisternon afrum</i> Boulenger 1909	Portuguese Guinea, along the west African coast from Guinea Bissau to the Nigler delta
		<i>Ophisternon infernale</i> Hubbs 1938	Caves of Yucatan, Mexico
		<i>Ophisternon aegnigmaticum</i> Rosen and Greenwood 1976	Northern South America
<i>Monopterus</i> Lacépède 1800	<i>Monopterus</i> Lacépède 1800	<i>Monopterus albus</i> Zuiew 1793	East Indies, Indo- Malayan Archipelago, Eastern Asia north to Japan
		<i>Monopterus eapeni</i> Talwar and Jhingran 1992	India
		<i>Monopterus boueti</i> Pellegrin 1922	Liberia, West Africa

Table 1.1. Continued

Subfamily	Genus	Species	Distribution
		<i>Monopterusuchia</i> Hamilton 1822	India, Pakistan, Nepal, Burma
		<i>Monopterusfossorius</i> Nayar 1952	India
		<i>Monopterusindicus</i> Silas and Dawson 1961	India
		<i>Monopterusdesilvai</i> Bailey and Gans 1998	Srilangka
		<i>Monopterusroseni</i> Bailey and Gans 1998	India
		<i>Monopterusdigressus</i> Gopi 2002	India
		<i>Monopterusichthyopoides</i> Britz, Lalremsanga, Lalrotluanga, and Lalramliana 2011	India
		<i>Monopterusbicolor</i> Nguyen and Nguyen 2005	Vietnam
		<i>Monopterusdienbienensis</i> Nguyen and Nguyen 2005	Vietnam
		<i>Monopterusjavanensis</i> Lacépède 1800	Indonesia (Sunda Trait), China, India

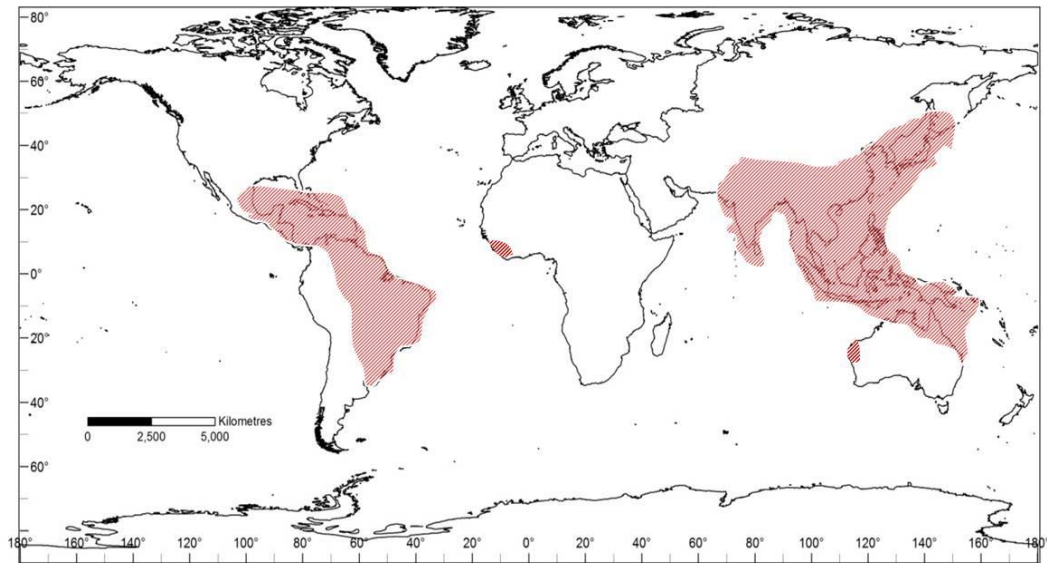


Figure 1.1. Distribution of the family Synbranchidae (red shading) (after Rosen and Greenwood, 1976; Kottelat *et al.* 1993; Pusey *et al.*, 2004; Berra 2007).

One of the earliest systematic studies of the family Synbranchidae was by Regan (1912). He used *Symbranchidae* for the name of the family, in which he placed three genera, *Symbranchus* Bloch, *Monopterus* Lacépède and *Macrotema* Regan. The three genera were separated based on the shape of the gill opening, eye shape, and the number of vertebrae. Three species were originally placed in the genus *Symbranchus*, whereas the other two genera, *Monopterus* and *Macrotema*, were represented by single species. Regan (1912) did not report the name of the three species of *Symbranchus*. However, he mentioned that the first species of *Symbranchus* was from the brackish water in Central and South America while the second and the third species of *Symbranchus* were from West Africa and East Indies respectively. He stated that the species of *Monopterus* was *Monopterus javanensis* Lacépède from the rivers of southern and eastern Asia and he did not explicitly refer to *M. albus*. He created the genus *Macrotema* for a species that was formerly placed in *Symbranchus* (*S. caligans* Cantor). The genus *Macrotema* was distinguished from *Symbranchus* on the basis of

morphological characters, body size and habitat, as it favours more saline environments.

The next taxonomic study of the family Synbranchidae focusing on the genus *Synbranchus* (previously *Symbranchus*) was conducted by Rosen and Rumney (1972). Their study identified a new species of *Synbranchus* from the Rio Madeira system in South America. This new species was distinguished from *Synbranchus marmoratus* Bloch in meristic and morphometric, pigment and osteological features. In addition, Rosen and Rumney (1972) reported that the new species may be endemic from the Rio Madeira system and was named *Synbranchus madeirae* Rosen and Rumney.

The current understanding of the systematics of the Synbranchidae is largely derived from the revision by Rosen and Greenwood (1976) based on anatomical studies. These authors removed the family Alabetidae from the synbranchoids and combined, based on osteological features, three families (Amphipnoidae, Flutidae, and Monopteridae) into the Synbranchidae. In their study, they proposed a new classification with 2 subfamilies, 4 genera, and 15 species of which 4 species are from the New World and 11 species are from the Old World. The division of the family Synbranchidae into 2 subfamilies, the Macroteminae and Synbranchinae, was based on the type of gill opening, the position of posterior nares, the presence of a caudal fin, and vertebral count. In addition, Rosen and Greenwood (1976) split the subfamily Synbranchinae into three genera, *Synbranchus* Bloch, *Monopterus* Lacépède and *Ophisternon* Regan based on skull anatomy, the fine scale structure of hyoid and gill arches, the type of gill opening, the branchial circulatory system and skeleton, and the number of vertebrae.

On the basis of their osteological studies, Rosen and Greenwood (1976) supported Rosen and Rumney's (1972) findings by placing *Synbranchus*

marmoratus and *Synbranchus madeirae* into the genus *Synbranchus*. Next, Rosen and Greenwood (1976) divided the genus *Monopterus* into six species, *M. albus* Zuiew, *M. boueti* Pellegrin, *M. cuchia* Hamilton, *M. fossorius* Nayar, and two forms originally described under the name *M. indicus* Silas and Dawson and *M. "indicus"* Eapen.

The original studies by Regan (1912) and Rosen and Rumney (1972) did not place the genus *Ophisternon* M'Clelland in the family Synbranchidae. However, based on the shape of gill opening and osteological features, Rosen and Greenwood (1976) proposed that six species formerly placed in the genera *Synbranchus*, *Furmastix* and *Anomatophasm* be transferred to the genus *Ophisternon*. Two species of this genus are from the New World and four from the Old World. This Gondwanan distribution pattern maybe a reflection of distributional patterns before continental drift (Rosen, 1975). The six species of the genus *Ophisternon* (*O. bengalense* M'Clelland, *O. afrum* Boulenger, *O. candidum* Mees, *O. infernale* Hubbs, *O. aenigmaticum* Rosen and Greenwood, and *O. gutturale* Richardson) proposed by Rosen and Greenwood have six or seven branchiotegals that ossify to the tips at all sizes and extend backward beyond the ventral tip of cleithrum. These species also have the shoulder girdle connected to the skull by a forked post-temporal bone.

Even though Rosen and Greenwood (1976) investigated in detail the anatomy of synbranchoid fishes, they reported that some undescribed species collected from the Atlantic slope of Mexico, Guatemala, Honduras, Cuba and north eastern South America could not be placed taxonomically without further investigations. Their position was subsequently supported by Perdices *et al.* (2005) and Valdez-Moreno *et al.* (2009) who used molecular genetic data to investigate genetic relationships within and between species within the genera *Synbranchus* and *Ophisternon*.

Perdices *et al.* (2005) reported evidence that the species of the genera *Synbranchus* and *Ophisternon* have long resided in Mesoamerica. They collected species of both genera from 45 South and Central American drainages, and Cuba. Their investigation based on nuclear and mitochondrial genetic sequences supported the recognition of the Mesoamerican species *O. aegnimaticum*, but their study did not support the monophyly of the species of *Synbranchus*. Their results revealed that the species of *Synbranchus* from Las Perlas (Pacific Panama) appeared to be highly distinct from other *Synbranchus* species. In addition, within the *Synbranchus* clade the *S. marmoratus* samples from the River Bayano (Pacific region) and South America are very highly divergent indicating cryptic speciation. Their study also indicated that *S. marmoratus* is otherwise genetically distinct and also supports what appears to be a cryptic species complex consisting of several Brazilian populations. This conclusion is also supported by studies on chromosome and genome size variation by Torres (2000). For the *Ophisternon* clade, Perdices *et al.* (2005) reported that their study supported the results of Rosen (1975), who indicated that *Ophisternon* samples from Cuba were morphologically distinct. Based on their findings, Perdices *et al.* (2005) therefore proposed a revision to the taxonomy of *S. marmoratus* and *O. aenigmaticum*. However, Perdices *et al.* (2005) recommended more samples of *Synbranchus* species be examined not only to investigate in more detail the morphology and anatomy but also to test their taxonomic hypotheses using molecular genetic data. A DNA barcoding study undertaken by Valdez-Moreno *et al.* (2009) showed that *O. aegnimaticum* from Catemaco Lake (Mexico) and Chisec River (Guatemala) differed genetically by 9%, which greatly exceeds that normally observed between fish species (Ward *et al.*, 2005), again suggesting the presence of cryptic species and the limitation to traditional morphologically-based methods for this group of fishes.

In summary the results of different studies discussed above, not only demonstrate taxonomic uncertainties and disagreements using morphology and anatomy but also indicate that more recent molecular genetic studies on species of the family Synbranchidae suggest the presence of cryptic speciation and significant limitation to the conventional morphology-based taxonomic approach to understand the diversity within this group of fishes (Ward *et al.*, 2005).

1.2.2. Genus *Monopterus*

The genus *Monopterus* is widely distributed in Asia, except for one species, which is found in Liberia (West Africa), and has an especially high diversity in India (Figure 1.2.) (Rosen and Greenwood, 1976; Bailey and Gans, 1998; Menon, 1999; Gopi, 2001; Nguyen, 2005; Berra, 2007; Britz *et al.*, 2011). Species of *Monopterus* are adapted to mainly warm, largely stagnant freshwaters such as muddy ponds, swamps, canals, and rice fields (Hill and Watson, 2007; Shafland *et al.*, 2010). Most of species of the genus *Monopterus* are epigean while two species (*M. eapeni* Talwar and Jhingran and *M. roseni* Bailey and Gans) are cavernicolous (Bailey and Gans, 1998).

Much disputation and confusion has centred on the generic name for the fish currently placed in the genus *Monopterus* since it was erected by Lacépède (1800) (Rosen and Greenwood, 1976). Rosen and Greenwood (1976) list a total of eight generic names associated with this group of eels and they argued for *Monopterus* Lacépède as the valid name of this genus.

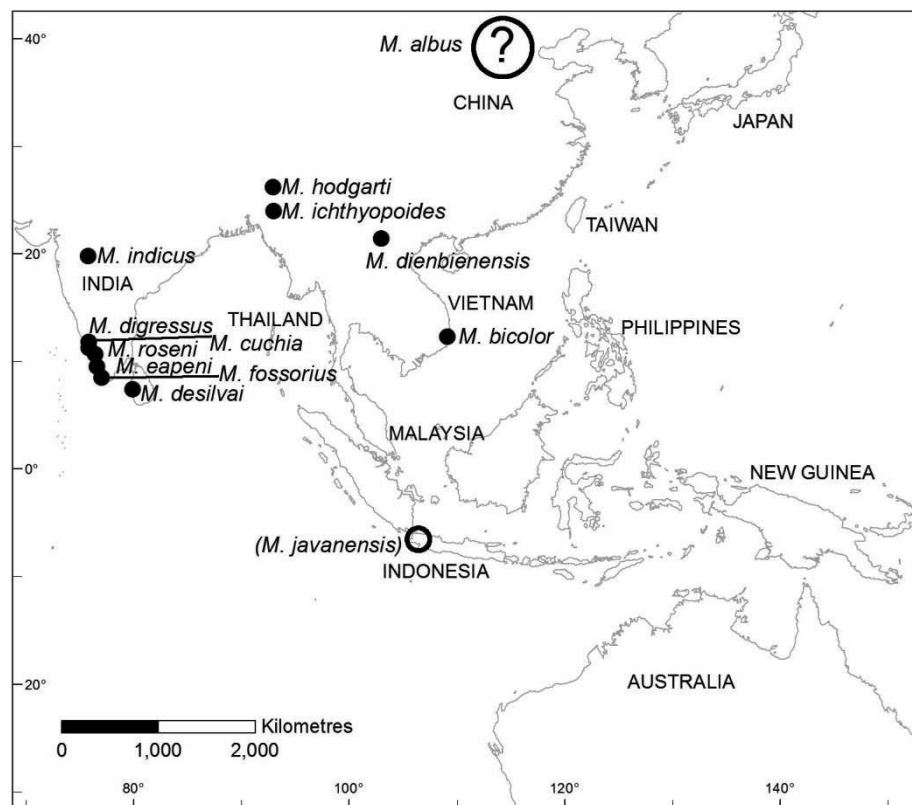


Figure 1.2. Type localities for valid *Monopterus* species in Asia. An additional species *M. boueti* (not shown) has the type locality of Monvoria, Liberia, West Africa. One important synonym relevant to the current study, *M. javanensis*, is also shown. The large circle and question mark for *M. albus* indicates uncertainty over the exact type locality (Rosen and Greenwood, 1976; Bailey and Gans, 1998; Menon, 1999; Gopi, 2002; Nguyen, 2005; Britz *et al.*, 2011)

The earliest major taxonomic study of swamp eels was conducted by Regan (1912) who considered that the genus *Monopterus* was monotypic, comprising *M. javanensis* Lacépède. Next, Rosen and Greenwood (1976) placed *M. javanensis* Lacépède as a synonym of *M. albus* Zuiew based on an examination of specimens from Java, Indonesia.

Rosen and Greenwood (1976) also reviewed a species placed in the genus *Typhlosynbranchus* and two species placed in the genus *Amphipnous* and transferred them to the genus *Monopterus*. In addition, Rosen and Greenwood (1976) transferred one species from the genus *Unibranchapertura* to the genus *Monopterus*. This resulted in the genus *Monopterus* being revised to include six

species based on the shape of gill opening without lateral folds and internally attached to isthmus, the characteristics of the head, the shape of the body, the structure of the upper lip, the number of vertebrae, the detailed structure of gill arch skeleton and the branchial circulatory system. This revision of the genus *Monopterus* undertaken by Rosen and Greenwood (1976) represent the most recent comprehensive taxonomic review of the genus and has been long considered the definitive taxonomic study.

More specifically, Rosen and Greenwood (1976) reviewed the status of *Unibranchapertura cuchia* formerly described by Hamilton (1822) from southern Bengal India, *Typhlosynbranchus boueti* previously described by Pellegrin (1922) from Sierra Leone, Monrovia, Liberia (West Africa), and *Amphipnous fossorius* formerly described by Nayar (1952) from Karamanai River, Trivandrum (Travancore State, India). Rosen and Greenwood (1976) transferred the three species into the genus of *Monopterus* based on the number of vertebrae, the type of gill opening and the characterization of the gill arch skeleton.

Next, Rosen and Greenwood (1976) reviewed the status of *Monopterus indicus* Eapen based on Eapen's depiction (Eapen, 1963) of the specimens from Kottayam, Kerala, India. This species was retained in the genus *Monopterus* due to the characteristics of the head and the structure of upper lip even though the number of vertebrae for this species (135 abdominal and 24 caudal) deviates from that typical of the genus *Monopterus*. However in their study, Rosen and Greenwood (1976) could not give a definitive opinion on the vertebral count until more samples became available for further study. Therefore, they retained the name "*M. indicus*" but this now led to confusion with the species formerly described by Silas and Dawson (1961) as *Amphipnous indicus* collected from Robbers Cave, Mahableshwar (Satara district, Maharashtra State, India) with its transfer to the genus *Monopterus*. While Rosen and Greenwood (1976)

considered that "*M. indicus*" Eapen and *M. indicus* Silas and Dawson were different species, based on the shape of the body and the characterization of dorsal gill arch skeleton, they did not address the nomenclatural issue. The confusion was resolved by Talwar and Jhingran (1992) who renamed *M. "indicus"* Eapen to *Monopterus eapeni* Talwar and Jhingran.

Talwar and Jhingran (1992) also distinguished two subgenera under the genus *Monopterus*, the subgenus *Amphipnous* Müller and the subgenus *Monopterus* Lacépède, based on the presence of scales on the posterior body and the presence of paired supratharyngeal pouches in the former subgenus. On the basis of these characters, the subgenus *Amphipnous* comprised *M. cuchia* Hamilton, *M. indicus* Silas and Dawson, *M. fossorius* Nayar, and *M. desilvai* Bailey and Gans and the subgenus *Monopterus* comprised *M. albus* Zuiew, *M. boueti* Pellegrin, *M. eapeni* Talwar, and *M. roseni* Bailey and Gans. However, Bailey and Gans (1998), who described two new species of *Monopterus* from Marawila, Sri Lanka (*M. desilvai*) and from Kerala, India (*M. roseni*), disagreed with the treatment of Talwar and Jhingran as they considered it at odds with the conclusions of Rosen and Greenwood (1976). Bailey and Gans (1998) considered that these grouping were not monophyletic. The phylogeny established by Rosen and Greenwood (1976), based on a range of characters, indicated that the African species, *M. boueti* of the subgenus *Monopterus*, is more closely related to the species of the subgenus *Amphipnous* than to the other species placed in the subgenus *Monopterus* even though the presence of posterior scales and a supratharyngeal pouch in these species is absent. The subgenus *Monopterus* would thus be paraphyletic. Until the phylogenetic relationships of these species have been further tested, Bailey and Gans (1998) recommend that Talwar and Jhingran's use of the concept of *Amphipnous* be used as a group name rather than as a formerly recognised subgenus.

Based on classical morphology-based treatments, Menon (1999) stated that *Moringua hodgarti* Chaudhuri from Abor Hills (Asam, India) is a valid species of *Monopterus*. However, Kottelat (2013) and Eschmeyer (2015) reported that this species has to be valid as *Moringua hodgarti* and considered as species incertae sedis in Synbranchidae. Later, Gopi (2002) described a new species, *M. digressus*, from Kerala India. This species is different from other *Monopterus* species in body size, body shape, absence of eyes, vertebral count, and the ratio of precaudal to caudal length. The species has a pair of suprapharyngeal pouches for aerial respiration but no scales on the posterior body. Due to this combination of characters, he could not place *M. digressus* under the subgenus or species group proposed by Talwar and Jingran (1992) without further study. Next, Nguyen (2005) reported two new *Monopterus* species, one from Dien Khanh, Khanh Hoa, South Vietnam (*M. bicolor*) and one from Thanh Luong, Dien Bien, North Vietnam (*M. diebienensis*). Very recently, Britz *et al.* (2011) reported another new species, *M. ichthyophoides*, from Mizoram, India. This species is distinguished from other *Monopterus* species by having only two, instead of five or six branchiostegal rays and a lower number of vertebrae (114-117). As this new species has scales in the caudal part of the body they included it in the subgenus *Amphipnous*. Next, Kottelat (2013) and Eschmeyer (2015) explained that *M. javanensis* is considered to be valid species for Southeast Asia region based on molecular genetic studies conducted by Collins *et al.* (2002) and Matsumoto *et al.* (2010).

A summary of the taxonomic treatments and conclusions of Regan (1912), Chaudhuri (1913), Pellegrin (1922), Nayar (1952), Silas and Dawson (1961), Eapen (1963), Rosen and Greenwood (1976), Talwar and Jhingran (1992), Menon (1999), Kottelat (2013) and Eschmeyer (2015) is given in Table 1.2, together with the six new species described by Bailey and Gans (1998),

Gopi (2002), Nguyen (2005), Britz *et al.* (2011) from India, Srilangka and Vietnam. In total, there are 13 species of *Monopterus* currently recognised mostly from India.

Table 1.2. Summary of the taxonomic classification of species of *Monopterus*

After Regan (1912), Chauduri (1913), Pellegrin (1922), Nayar (1952), Silas and Dawson (1961), Eapen (1963)	After Rosen and Greenwood (1976)	After Talwar and Jhingran (1992)	After Bailey and Gans (1998), Menon (1999), Gopi (2002), Nguyen (2005), Britz <i>et al.</i> (2011), Kottelat (2013), Eschmeyer (2015)
<i>M. javanensis</i>	<i>M. albus</i>	-	<i>M. javanensis</i> <i>M. albus</i>
<i>Typhlosynbranchus boueti</i>	<i>M. boueti</i>	-	-
<i>Unibranchapertura cuchia</i>	<i>M. cuchia</i>	-	-
<i>Amphipnous fossorius</i>	<i>M. fossorius</i>	-	-
<i>Amphipnous indicus</i>	<i>M. indicus</i>	-	-
<i>M. indicus</i>	<i>M. "indicus"</i>	<i>M. eapeni</i>	<i>M. desilvai</i> <i>M. roseni</i> <i>M. digressus</i> <i>M. ichthyopoides</i> <i>M. bicolor</i> <i>M. dienbienensis</i>

The ongoing discovery of new species of *Monopterus* suggests that the diversity within the genus is still poorly known. Therefore, more comprehensive and systematic studies of the genus *Monopterus* are needed to extend the morphology based reviews conducted by Rosen and Greenwood (1976) and Britz *et al.* (2011). These studies need to take advantage of molecular genetic data due to the morphological conservatism of the group with the very real possibility of cryptic speciation and to ensure taxonomic reviews are based on robust phylogenetic analyses.

1.2.3. The *Monopterus albus* species complex

Monopterus albus is an economically important freshwater fish in Asia. It is popular due to its reputation as delicious food and because it has the ability to survive and grow in poorly oxygenated inland waters and to be transported live out of water. It occurs widely across East Indies the Greater Sunda Islands, the Malay Peninsula, and the Indochinese Peninsula, the Philippines, the southern part of East Asia (southeastern China, the Korean Peninsula, the western Japanese Archipelago) as well as Southeast Asia (Malaysian and Indonesian Archipelago) (see Figure 1.3) (Rosen and Greenwood, 1976; Berra, 2007). The species has also been reportedly found in two rivers in north-eastern Australia, however these records need to be authenticated, and even if they do occur they may be the result of introductions (Allen *et al.*, 2002). In Indonesian *M. albus* is broadly distributed across the islands of Sumatra, Java, Bali, Lesser Sundas, Sulawesi, and the Moluccas (Kottelat *et al.*, 1993).

Since it was first described by Zuiew in 1793 there has been considerable taxonomic disputation and confusion centred on this species (Rosen and Greenwood, 1976). Firstly, the actual type location remains confused. The type specimen is lost and Nichols (1943) predicted that Zuiew's specimen came from Asiatic Russia. However, Eschmeyer in Catalog of Fish (<http://research.calacademy.org/redirect?url=http://researcharchive.calacademy.org/research/Ichthyology/catalog/fishcatmain.asp>, accessed 20 November 2015) has indicated Suriname, South America as the species type locality. Secondly, the fish species is phenotypically plastic, which has led to an extensive and confusing taxonomic nomenclature. *Monopterus albus* has 23 synonyms (within 11 Genera) listed on Fishbase (<http://www.fishbase.org/search.php>, accessed 15

March 2013) and it has been described under nine species names from seven different genera according to Eschmeyer (2015). In addition, Rosen and Greenwood (1976) considered that the fish species has been described under at least 13 species names within six different genera based on morphological information from across a number of countries. Rosen and Greenwood (1976) undertook morphological and osteological studies of *M. albus* based on the specimens from China (Fukien, Anhwei, Shansi, Hainan, and Yunan), Burma, Vietnam, Java (Indonesia), and Okinawa (Japan).

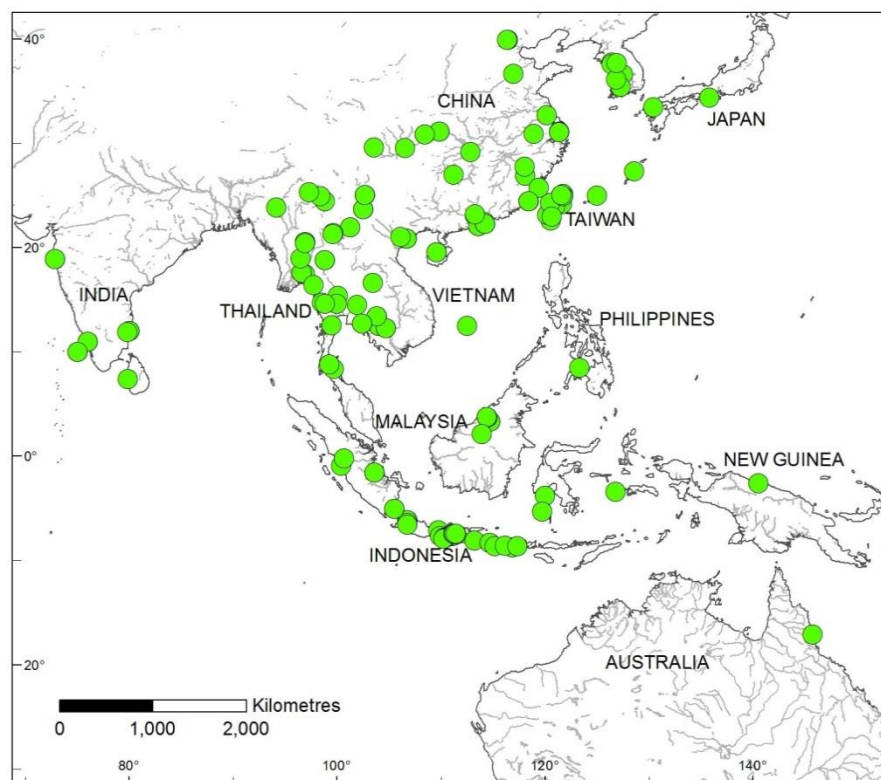


Figure 1.3. Distribution map of *M. albus* (green circles) based on museum records (www.lifemapper.org, accessed 15 March 2013), Allen *et al.* (2002), Kottelat *et al.* (1993); Matsumoto *et al.* (2010); Cai *et al.* (2012) and locations sampled in this study (Chapter 2).

Much of the taxonomic debate has centred on the status of populations occurring on separate landmasses and islands. Several studies consider *M. albus* to represent a species complex. Nichols (1943) classified the samples of “*M. albus*” from northern China as *Monopterus cinereus* Richardson and differentiated this fish from southerly Chinese populations on the basis of differences in tail length, eye size, coloration, and head shape. However, Rosen and Greenwood (1976) considered that the species assigned by Nichols (1943) to be a synonym of *M. albus* based on specimens they investigated from Yunan. This is due to Nichols (1943) falling to uncover the peculiar lower jaw and gill arch features to the species. More recently, the application of modern molecular genetic techniques to the systematics of this species suggest that the taxonomic issues involving this species are more complex and far from settled (Collins *et al.*, 2002; Cai *et al.*, 2008; Matsumoto *et al.*, 2010; Cai *et al.*, 2012) (see below).

Molecular genetic studies have been used to investigate variation among samples of *M. albus* collected from several countries including Japan, Taiwan, Yogyakarta (Indonesia) (Matsumoto *et al.*, 2010), Malaysia, Vietnam, Jakarta (Indonesia) and the USA, where the swamp eels has been introduced (Collins *et al.*, 2002). These researchers also collected *M. albus* from several regions in China for comparative purposes. These studies used different fragments of the 16srDNA mitochondrial gene: Collins *et al.* (2002) used universal primers 16Sar and 16Sbr, whereas Matsumoto *et al.* (2010) adopted the L1567 and H2196 primers. Results clearly support the existence of at least three cryptic species, however, each study has limitations in terms of sampling.

Cai *et al.* (2008) also studied genetic variation in *M. albus*, and based on an analysis of four populations from the Sichuan Basin in China using mitochondrial control region sequences, found little genetic differentiation among

samples. Subsequently, Cai *et al.* (2012) expanded the geographical sampling of *M. albus* to eight more locations in China. Their results revealed that *M. albus* fell into five genetic lineages based on mitochondrial control region sequences. One of these lineages (lineage C) inhabiting inland areas demonstrated the highest haplotype diversity (39 from 60) due to its wide distribution, which may have resulted from the historical effects of glaciation and inter-regional introduction. In addition, this study showed that the three lineages (A, B, and E) inhabiting coastal regions were distinct from lineage C. This was considered most likely due to isolation from inland areas blocking gene flow between inland and coastal populations. The authors did not draw any conclusions regarding the taxonomic implications of their findings.

In summary, the findings of different studies on molecular genetic variation support, somewhat tentatively, that populations of the putative species *M. albus* are sufficiently differentiated over a significant portion of its distribution to flag the existence of as many as three cryptic species. This therefore, justifies more detailed molecular taxonomic studies to clarify the status of the “*M. albus*” species complex, especially “*M. albus*” from Southeast Asia including Indonesia, using a greater range of molecular markers and more intensive geographic sampling. Additionally, DNA sequence data has the advantage that it can be compared with that obtained from previous studies to help clarify whether *M. albus* should be considered as a single variable species, one species with several subspecies, or several distinct biological or genealogical species.

1.3. Molecular Markers

Genetic information gained from molecular markers is now considered essential for effective sustainable exploitation, management, and conservation of wildlife and commercially important species (Awise, 2000). In addition, molecular genetic data can be used to address taxonomically questions in morphologically variable or conservative species and can contribute to the understanding of patterns of speciation, evolution and biogeography (Hillis *et al.*, 1996; Awise, 2000). Molecular genetic markers have been especially useful for the identification of genetically divergent wild stocks of conservation significance for a number of fish species (Ward and Greewe, 1994; Arif and Khan, 2009; Chauhan and Rajiv, 2010). Therefore, relevant techniques used to generate molecular markers in this study and relevant fields of genetic research are outlined and discussed in the following sections.

1.3.1. DNA Sequencing

DNA sequencing refers to a method for determining nucleotide differences among individuals from orthologous gene regions. DNA sequencing has become almost routine for numerous kinds of applied research including systematics, forensic biology and biotechnology since the first DNA sequences were obtained in the early 1970s (Hillis *et al.*, 1996). Following the development of the Polymerase Chain Reaction (PCR) procedure, many genetic techniques have been development to measure and explore genetic diversity based directly or indirectly on DNA sequence difference among many groups of animals, plants and microbes. The findings of these studies have been usefully applied to not

only genetic improvement programmes but also in conservation management (Dunham, 2004; Arif and Khan, 2009) as well as taxonomic and evolutionary studies (Hillis *et al.*, 1996; Avise, 2000).

Direct DNA sequencing has been used to analyse genetic and phylogenetic relationship within many groups of animals such as fish (Cowman and Bellwood, 2011; Bloom and Lovejoy, 2012; Alda *et al.*, 2013), shrimps (Anker and Baeza, 2012; Von Rintelen *et al.*, 2012), gastropods (Greve *et al.*, 2010; Modica *et al.*, 2011; Kotsakiozi *et al.*, 2012); frogs (Maciel *et al.*, 2010; Oliver *et al.*, 2013), turtles (Yilmaz *et al.*, 2011; Le *et al.*, 2013), birds (Bristol *et al.*, 2013; Packert *et al.*, 2013), and camels (Barreta *et al.*, 2013) to give just a few examples. Several DNA sequencing studies have been conducted on swamp eels. The entire mitochondrial genome of *M. albus* was sequenced by Miya *et al.* (2001). This study focused on mitogenome sequencing for the study of the higher Teleostean phylogeny and unfortunately did not report the sampling location for the *M. albus* specimen sequenced.

The gene region most commonly used in DNA sequencing studies of *M. albus* has been the 16S mitochondrial rRNA gene (Collins *et al.*, 2002; Chen, *et al.* 2003; Matsumoto *et al.* 2010). However studies used a range of 16S primers that amplified different fragments of this gene. Collins *et al.* (2002) used universal primer 16 Sar and 16 Sbr (Palumbi, 1996), whereas Chen *et al.* (2003) designed their own primers, and Matsumoto *et al.* (2010) used primer L1567 (Matsumoto *et al.*, 2010) and H2196 (Yamaguchi *et al.*, 2000). Some additional mitochondrial gene regions that have been used for *M. albus* are 12S rRNA (Chen *et al.*, 2003) and D-loop (Cai *et al.* 2008; Cai *et al.*, 2012). No data utilising sequences from nuclear gene regions have been published. The advantage of the use of different gene regions is that increases the robustness and reliability of systematic

conclusions and phylogenetic hypotheses and it is considered unwise to base analyses on information from a single gene region or kind (i.e. mitochondrial sequences). In fact it is now considered highly desirable, if not essential, to include sequences from both mitochondrial and nuclear genes regions to investigate species boundaries and genetic relationships between populations and taxa (Schonhuth *et al.* 2012; Santini *et al.*, 2013).

1.3.2. Denaturing Gradient Gel Electrophoresis

Denaturing Gradient Gel Electrophoresis (DGGE) is a very powerful genetic fingerprinting tool developed by Fischer and Lerman (1980). DGGE utilizes a PCR step to amplify the target DNA fragment, which is then applied to a denaturation gradient within a polyacrylamide gel electrophoresis to separate sequence variants among the PCR products based on their melting characteristics (Fischer and Lerman, 1980; 1983). This melting profile is totally dependent on the sequence of the DNA molecule, allows the identification of DNA fragments that may differ by as little as a single nucleotide (Myers *et al.*, 1985; Sheffield *et al.*, 1989).

Although the method has limitations such as low sensitivity for differentiating fragments larger than 500 bp, use of specialised equipment, and being time consuming due to the need for preliminary experiments including determination of melting behaviour before fragment analysis can take place, the method is commonly used for DNA profiling and detecting mutations. It is more sensitive than other related methods including single-strand conformation polymorphism (SSCP). The DGGE can also use nonradioactive visualization methods and is low-cost compared to other molecular genetic screening methods (Wallis, 2002; Knapp, 2009).

DGGE is widely used for microorganism profiling especially in the fields of clinical science (McAuliffe *et al.*, 2005; Baillie and Bouwer, 2011), systematics (Gong *et al.*, 2002; Doms *et al.*, 2007), environmental science (Thornhill *et al.* 2009; Kriwy and Uthicke, 2011), and aquaculture (Papakostas *et al.*, 2006). Even though the method has many advantages, it has been little used in plants and animal genetics. No published molecular genetic studies of swamp eels have used the DGGE method.

1.3.3. Restriction Fragment Length Polymorphism

Restriction Fragment Length Polymorphism (RFLP) is one of the most well established molecular techniques used to detect variation in a sequence of DNA through gel electrophoresis. In this method, a defined DNA fragment is amplified by PCR and then digested with selected restriction endonuclease enzymes to generate polymorphic profiles of the species or population investigated. Despite the fact that the method has become less used due the reduced cost of direct sequencing, this method is still useful for genetic profiling or identification of species or populations especially in relation to known variants (Dunham, 2004; Dudu *et al.*, 2011). For example the method has been carried out to analyse genetic diversity of fish (Cheng and Lu, 2005; Dudu *et al.*, 2011; Balaraj and Basheer, 2012), sheep (El-Hanafy and Salem, 2009), bats (Ferreira *et al.*, 2005), and aphids (Jalalizand *et al.*, 2012).

1.3.4. Microsatellites

Microsatellites consist of multiple copies of short tandem repeats (STRs) located in both coding and noncoding regions of eukaryotic genomes and range in size from 1 to 6 base pair long (e.g. ACACAC... and GATAGATA...). Microsatellites are typically co-dominant markers and are highly polymorphic due to a higher rate of mutation compared to other neutral regions of DNA (Arif and Khan, 2009; Chauhan and Rajiv, 2010). Microsatellites can be detected by the design of PCR primers, which uniquely flank a repeated region. Therefore, a single pair of microsatellite PCR primers will be specific for a single locus and work for each individual of a species or population and generate allelic data based on the amplification of products of different lengths (Ellegren, 2004).

Microsatellites have been widely used in many areas of genetics including population studies and conservation biology. This form of molecular marker can be used to describe population structure and population differentiation, and investigate evidence for genetic drift and genetic bottlenecks (Sekino *et al.*, 2010; Pujolar *et al.*, 2011; Kalinowski *et al.*, 2012). In addition, microsatellites can be used to detect sudden changes in population size, effect of population fragmentation and mixing of different populations, and to establish parentage and genealogical relationships and species boundaries (So *et al.*, 2006; Kanno *et al.*, 2010; Karaïskou *et al.*, 2011; Kitanishi *et al.*, 2012).

Microsatellite markers have been investigated in *M. albus* by Lu *et al.* (2005), Li *et al.* (2007), and Lei *et al.* (2011). Their studies found several markers that showed polymorphisms in *M. albus*. Lei *et al.* (2011) examined two sets of multiplex PCR reactions to facilitate rapid genotyping. The use of multiplexed loci using different fluorescent dyes combined with fragment separation on an

automated genetic analyser can greatly increase the efficiency microsatellite genotyping (Guichoux *et al.*, 2011).

1.4. Application of Molecular Markers

Molecular genetic markers are now considered a major tool or even essential for addressing taxonomically problematic organisms, to assist with the management and genetic improvement of livestock and crops including commercially fished and aquaculture species as well as conserving genetic diversity of endangered species, including fish (Chauhan and Rajiv, 2010). The application of molecular markers in the fields of genetic research relevant to this study are outlined and discussed in the following sections.

1.4.1. Taxonomy and Evolution

The accurate identification of species is important in all areas of biology including biodiversity and conservation-related research. Molecular genetic data is used routinely to resolve taxonomic uncertainties and establish hypotheses for systematic relationships at various taxonomic levels including the determination of species boundaries in phenotypically plastic or conservative organisms. Recently, there have been calls for the use of genetic information derived from sequences of the mitochondrial cytochrome c oxidase subunit I (mt-DNA COI) gene region, referred to as a DNA barcode, as primary data for taxonomic classification and identification of organisms (Hebert *et al.*, 2003a,b). For animals, the use of 650 base fragment of mt-DNA COI is now being used widely in DNA barcoding studies (Ma *et al.*, 2012; Ni *et al.*, 2012; Pèrez-Lozada *et al.*, 2012).

However, the use of a DNA barcoding approach for biological classification has both strong proponents and opponents. This is due to the view that there will always be limitations to an approach to DNA barcoding based on a single source of information and the reliance on a simplistic genetic-yardstick approach to determining species boundaries (Ferguson, 2002). Other problems that arise from reliance on information from a single mitochondrial gene include heteroplasmy, discrepancies between maternal and paternal lineages, incomplete lineage sorting and disconnections between genetic divergence and speciation (Hebert *et al.*, 2003a,b; Moritz and Cicero, 2004; Hajibabaei *et al.*, 2007; Asgharian *et al.*, 2011). Nevertheless, DNA barcoding can still be useful for resolving taxonomic anomalies, for confirming taxonomic identifications and identifying taxonomically complex groups, however the limitation of the technique need to be clearly understood (Cerutti-Pereyra *et al.*, 2012).

Regardless of one's position on the debate on the utility of DNA barcoding, which is still an efficient method for species level identification, there is agreement that taxonomic classification at all levels should be consistent with evolutionary relationships (Hebert *et al.*, 2003a,b). Thus taxonomic classification of organisms should be on the basis that they represent monophyletic groups based on rigorous phylogenetic analysis (Mishler and Theriot, 2000; Dornburg *et al.*, 2012; Sales *et al.*, 2013). The ready availability of molecular markers, especially nucleotide sequences, combined with increasingly sophisticated models and methods for phylogenetic reconstruction, means that increasingly robust classifications are being developed for many groups of organisms (Avisé, 2000; Knowlton, 2000; Makowsky *et al.*, 2013; McCormack *et al.*, 2013).

Studies of species relationships or species boundaries have to assume or refer to a species concept. However, the subject of species concepts has a large and complicated literature. The problem with a consistent species concept is

partly due to biological diversity itself, whereby systematists could be dealing with microbes, asexual or sexual species and of course the ongoing evolutionary process itself with many populations of many species in various stages of the speciation process. Biologists also bring to the debate different taxonomic philosophies (e.g. biological, classical taxonomic, cladistic or genealogical or phenetic), use different kinds of data, and focus on either conceptual or operational aspects for defining species. All these factors have led to significant dissent among biological systematists with one author having identified 24 different species concepts (Mayden, 1997). The classical taxonomic approach, which uses morphological or phenotypic dissimilarity for defining species, is still the most commonly used method. However, this concept faces a number of problems not least of those of phenotypically plastic species resulting in the identification of multiple taxa that may be fully compatible reproductively (Agapow, 2004; Bickford *et al.*, 2006). Conversely, this concept will treat organisms that are phenotypically similar as conspecific even though they maybe reproductive isolated, a situation that is rarely acknowledged let alone investigated (Mallet, 1995; Balakrishnan, 2005).

Conceptual limitations of the morphological species concept are overcome by the biological species concept. This concept defines species as interbreeding groups in natural populations, which are reproductively isolated from other such groups (Mayr, 2000). The biological species concept has been widely adopted by many taxonomists and ecologists as well as molecular biologists (Balakrishnan, 2005; Hausdorf and Hennig, 2010). However, the biological species concept has several disadvantages such as inapplicability to asexual taxa and fossils and the difficulties involved with the assessment of reproductive relationships among allopatric populations and its operation more generally. As a result, most taxonomists use a “blended” approach that combines

the morphological species concept and the biological species concept to delineate species boundaries, by assuming that observed phenotypic “gaps” results from reproductive isolation (Mallet, 1995; Wu, 2001a,b; Balakrishnan, 2005; Hausdorf, 2011).

The rise of phylogenetic methods and the increasing use of nucleotide data in systematic studies has resulted in the development of several phylogenetic or genealogical-based species concepts (Mishler & Theriot, 2000). These concepts define a species as a group of individuals derived from a common ancestor and identifiable on the basis of a set or combination of derived traits, that allows them to be defined as a monophyletic group. While many monophyletic groups within genera, may also be biological species, phylogenetic species concepts do not have a requirement that “their” species are reproductively isolated from one another. As a result any group of individuals or populations that can be diagnosed on the basis of a small number unique traits can qualify as “phylogenetic” species. Thus the application of this concept tends to encourage extreme division of species into a larger number of small groups (Balakrishnan, 2005; Hausdorf, 2011).

The use of molecular genetic data and phylogenetic methods are now commonplace for taxonomic studies. It is also now widely accepted that studies on the significance of morphological evolution, ecological or life history traits and biogeographical patterns should be done within a phylogenetic framework (Hillis, 1998; Page and Holmes, 2003).

1.4.2. Phylogenetic Reconstruction

Many methods have been developed for phylogenetic reconstruction using nucleotide data and the most common are distance-based, maximum parsimony, maximum likelihood, and Bayesian interference. Much has been written on the pros and cons of the different approaches and associated systematic philosophies (Hillis *et al.*, 1996; Felsenstein, 2003; Hall, 2011; Tamura *et al.* 2011).

Distance methods require the calculation of a matrix of genetic distance values between all pairs of taxa, and the phylogeny can then be estimated from these values using a variety of tree building methods (Hall, 2011, Tamura *et al.*, 2011). The calculation of genetic distance refers broadly to the mean number of nucleotide changes per site, but will vary based on the model of nucleotide evolution used, or assumed (Felsenstein, 2003; Van de Peer, 2009; Hall, 2011). Maximum parsimony is a character-based method that infers a phylogenetic tree by minimizing the total number of evolutionary steps required to explain a given data set. However, this method is restricted to smaller data sets, especially when bootstrapping is required, even with the availability of faster computers (Jin *et al.*, 2006).

The basic principle of the maximum likelihood method is to find the optimal tree based on an analysis of DNA sequence data with an assumed model of evolution. The advantage of the maximum likelihood approaches is that it allows robust statistical tests to be performed. Therefore, this method is attractive for phylogenetic analysis. However, like parsimony, estimation of optimal trees is computationally demanding for large data sets (McCormarck *et al.*, 2009). Lastly, an increasingly popular approach to phylogenetic analysis is by Bayesian methods. In general, Bayesian inference has similar statistical attributes to

maximum likelihood approaches. This is due to the Bayesian inference being based on a likelihood function and can use different models of nucleotide evolution. Nevertheless, the Bayesian approach is different from other methods as it allows prior information about the phylogeny to be specified. The advantage of the Bayesian methods is relatively rapid speed, which allows complex evolutionary models to be investigated. Bayesian phylogenetic inference uses the Monte Carlo Markov Chain method to estimate the best trees (Yang and Rannala, 2010; Zhang *et al.*, 2011).

All phylogenetic tree-building methods require that a robust alignment of nucleotides is available. This is usually straightforward for protein coding genes as alignments are based on alignment of codons, however this can become complex for RNA and non-coding sequences, especially where divergence levels are significant between taxa or OTUs. A number of methods and programs for multiple alignments are available and different programs and settings can produce different results and it is advisable to check alignments by eye to ensure there are no aberrant results (Hall, 2011). However, not all of phylogenetic methods require an alignment, This is due to limitation in Multiple Sequence Alignment (MSA) (Chan and Ragan, 2013). Alignment-free methods have been supported by some studies, such as the use of kSNPv2 method (Gardner and Hall, 2013), a D2 method (Chan *et al.*, 2014) and a POY method (Wheeler *et al.*, 2015). Nevertheless, the scalability and robustness of these methods to key evolutionary processes remain to be investigated (Chan and Ragan, 2013; Chan *et al.*, 2014).

1.4.3. Principles of Biological Classification and Numerical Taxonomy

Biological classification is the process by which biologists group and categorize extinct and living species of organisms within a taxonomic framework and is the oldest biological discipline (Abbott *et al.*, 1985; Ebach and Williams, 2004). The goal of classifying is to place an organism into an already existing group or to create a new group for it, based on its resemblances to and differences from known forms. Therefore, to this end, a hierarchy of categories is recognized for the taxon (Gordon, 1987; Mayr and Bock, 2002; Mishler, 2009). Biological classification also involves the formal description and naming of the organisms. Classification and naming of organisms in the oldest form of biology and is often referred to as classical taxonomy as it is based on the assessment and description of morphological characters (De Hoog, 1981; Cook, 2010). The current system of biological classification is based on the work of Carolus Linnaeus who grouped species according to shared physical characteristics. In the earlier artificial systems, only a few morphological characters were determined. Later on, in the natural system of classification, an increasingly large number of morphological characters were taken into consideration. As a result, classification of group of organisms was more satisfactory and stable as their arrangements increasingly reflected natural (evolutionary) relationship with each other (Stevens, 2002; Barkley *et al.*, 2004). Basically, similarities in morphological characters, especially those that are unique or derived, are used for grouping organisms together, and thereby indicating biologically meaningful relationships. On the other hand, dissimilarities or differences in characters can be used for separating different organisms within a nested framework of increasing dissimilarity. Organism with the greater differences are considered to be unrelated or distantly related and as a consequence placed into different higher

taxa on the basis of their level of divergence (Wilson, 2004; Dayrat, 2005; Roth, 2005).

Before Linnaeus, the use and application of a generic name and a “*differentia specifica*” for each species were not fixed (Barkley *et al.*, 2004; Reid, 2009). Linnaeus, then, took every effort to improve the process of the “composition” of a species delineation by abolishing unnecessary rhetorics, introducing new descriptive terms, and defining their meaning with an unprecedented precision. In the late 1740s, he began to use a parallel system of naming species with *nomina trivialia*, which is started with the genus name and the specific epithet (Schuh, 2003). Linnaeus consistently applied *nomina trivialia* to the species of plants in the 1st edition of *Species Plantarum* (1753) and to the species of animals in the 10th edition of *Systema Naturae* (1758). By consistently using these specific epithets, Linnaeus separated nomenclature from taxonomy (Ride *et al.*, 1999; Greuter *et al.*, 2000; McNeill *et al.*, 2006). In the nineteenth century, the new practise was codified in the first Rules and Laws of Nomenclature. The 1st edition of *Species Planetarum* and the 10th edition of *Systema Naturae* were chosen as starting points for the Botanical and Zoological nomenclature respectively. This convention for naming species is referred to as binomial nomenclature (Kress and DePriest, 2002; Greuter, 2004). The use of binomial nomenclature refers to the genus as a first name and the species as a second name. The name of the genus always starts with capital letter and the name of species starts with a lower case letter, which is frequently descriptive. Nomenclature is now regulated by Nomenclature Codes, which allow names divided into ranks (McNeill, 2000; Reid, 2009; Stuessy, 2009).

In the more recent systems of classification of organisms, it is now generally accepted that classification should reflect or be consistent with the Darwinian principle of common descent (Padian, 1999; Winsor, 2009;

Chakrabarty, 2010). The long-term impact of Darwinian evolution has been different and very important. It has led to the expectation that the basic hierarchical arrangement of organisms in a taxonomic schema should be aligned with their phylogenetic relationships rather than with a set of discrete and arbitrary classes (Serenó, 2005; Podani, 2009; Makowsky *et al.*, 2013).

A strict application of the Darwinian principle in taxonomy is called cladistic taxonomy. The cladistic approach to taxonomy arranges taxa in a cladogram, which reflects the evolutionary relationships of the organisms under consideration (Queiroz and Gauthier, 1992; Huson and Bryant, 2006; Stuessy, 2009; Van de Peer, 2009). If a taxon includes all descendants of some ancestral form, it is called monophyletic. If it does not it is considered paraphyletic, and if it includes other groups that do not share a hypothesised common ancestor it is called polyphyletic (Laurin *et al.*, 2005; Pickett, 2005; Queiroz, 2006; Hörandl and Stuessy, 2009). The arrangement of the organisms as clades rather than taxa means that the process of classification is now more complex and greater consideration needs to be given to the process of recognising and naming groups of organisms. A new formal code of nomenclature, the PhyloCode, has been proposed and is currently under development, however the use and value of this approach is still being debated (Berry, 2002; Forey, 2002; Queiroz, 2006; Dayrat *et al.*, 2008).

Biological taxonomy or classification is not static, and opinions about the correct status of taxa at all levels and their correct or appropriate placement are constantly revised as a result of new research (Grant, 2003; Forey *et al.*, 2004). Many aspects of classification will always remain a matter of scientific judgement (Niklas *et al.*, 2001; Harris, 2005). Biologists, however, have attempted to view all living organisms with equal thoroughness and thus have devised a formal classification. The formal classification provides the basis for relatively uniform

and internationally understood nomenclature, thereby simplifying cross-referencing and retrieval of information (Lance *et al.*, 2000; Barkley *et al.*, 2004; Ebach and Williams, 2004; Agnarsson and Kunther, 2007; Mishler, 2009).

Comparison of characters depend to some extent on the purposes of the comparison. For identification purposes, a suitable key with attention given only to the most important diagnostic characters for the taxon under study should be sufficient especially for well know groups (Padial, *et al.*, 2010; Puillandre *et al.*, 2012). If the taxon is likely to be a new one, its general position is determined by observing as many as characters as possible and by comparing them with the descriptions and keys in existing and past classifications (Balakrishnan, 2005; Maan and Seehausen, 2011). In making comparisons, resemblances resulting from convergence should be identified. In addition, inconspicuous characters, including internal anatomy could be of great importance in indicating affinities. Although, such characters may be of limited value for identification of a group, they may be of the utmost importance in understanding evolutionary relationships (Weissing *et al.*, 2011; Fujita *et al.*, 2012; Lim *et al.*, 2012). Therefore, characters are often weighted to some extent by taxonomists according to their utility for different purposes. However, weighting or non-weighting of characters can be the subject of great dispute. On the one hand it has been pointed out that weighting characters are often demonstrably arbitrary and imprecise (Agrawal, 2001; Podani, 2009; Nosil and Feder, 2012). On the other side, if characters were examined equally weighted (which is what people usually mean when they use term without weighting or unweighted), some obvious cases of extreme convergence would generate obviously abberant classifications. A classification based on equally weighted characters is generally called phenetics as opposed to phyletics, in which characters are weighted by presumed importance in

indicating lines of descent and defining monophyletic groups (Pisani *et al.*, 2007; Cook *et al.*, 2010)

With the advent of powerful computers, mathematical approaches known as numerical taxonomy were developed for estimating phenetic distance with the intention to bring a more objective approach to biological classification (Sneath and Sokal, 1973). This method has been revolutionary for taxonomy by introducing computer-based numerical algorithms and statistical procedures, which are now indispensable tools in modern taxonomy, including molecular taxonomy (Sneath, 1995; Jensen, 2009; Stuessy, 2009). Numerical taxonomy classifies organisms based on their overall similarity. Firstly, a representative set of characteristics of a group of organisms are chosen and measured. Then, these measurements are used to calculate similarity coefficients between all pairs of organisms. The similarity coefficient is a number between 0 and 1, where 1 indicates absolute identity, and 0 indicates absolute dissimilarity. Finally, the similarity coefficients are used to develop a classification system usually based on clustering (dendrograms) or ordination (Sneath and Sokal, 1973; Hill, 2005).

Popular ordination and classification techniques widely used by numerical taxonomists are Multiple Discriminant Analysis (MDA) and Principal Component Analysis (PCA). MDA is the appropriate statistical technique when the dependent variable is categorical such as a set of two or more species or population samples and the independent variables are quantitative (i.e. the measured set of characters). This approach can support classification by yielding an efficient representations of variation in multiple variables and samples in the form of highly reduced set of compound variables (dimensions) amenable to straight forward interpretation and classification (Duda *et al.* 2001). PCA is a way of identifying correlated patterns of variability among variables, and expressing the data in such a way as to highlight their similarities and differences on a reduced number

of components, dimensions, factors or axes. This useful statistical technique is a common technique for finding patterns in data of high dimensionality (Shlens, 2009). For example, both approaches have been applied to detect cryptic species in groups such as beetles (Arribas *et al.*, 2013) and harvestman (Arthofer *et al.*, 2013). However, the limitation of numerical taxonomy is that it will classify unrelated organisms together based on overall morphological similarity and does not distinguish between analogous and homologous features (Wayne, 1981; Warheit, 1992). According to pheneticists the unimportant analogous (i.e. convergent) features are usually numerically over-whelmed by the larger number of homologous (derived) features (Sneath and Sokal, 1973).

Some biologists consider that numerical taxonomy can produce quantitative measures of overall differences among groups that can be used to establish the maximal difference allowed at each taxonomic level (De Hoog, 1981; Sneath, 1995). Although such an approach may be possible, many difficulties exist (Price *et al.*, 2003; Jensen, 2009). For example an "Order" in one authority's classification may be a superorder or class in another, and it is unlikely that agreed upon numerical criteria can be developed to establish a particular level of the taxonomic hierarchy. In reality most of the established classifications of the better known groups result from a general consensus among practicing taxonomists. It also follows that no complete definition of a group can be made until the group itself has been recognized, after which common (or defining) characters can be formally stated. In addition, as further information is obtained about the group, it is subject to taxonomic revision (Lance *et al.*, 2000; Kuntner and Agnarsson, 2006; Patterson, 2006).

In order to address the problem of dealing with gaps and taxonomic rank in classification including delimitation of species, integrative taxonomy should be applied. Ideally, morphological, molecular and possibly other data such as

reproductive, behavioural and ecological data should be combined (Schlick-Steiner *et al.*, 2010; Gebiola *et al.*, 2012). This is due to none of the methods being 'superior' but all having different intrinsic sources of error. For example, character convergence (homoplasy) is much more common in morphological data than in molecular sequence data, but character reversions are more common in the latter (Hickerson *et al.*, 2006). Recently, many taxonomists combine morphological, ecological and molecular data to resolve taxonomic problems in many organisms including those which are considered as species complexes such as the existence of two sympatric species in *Cyphopharynx furcifer* (Cichlidae) from Lake Tanganyika investigated by Takahashi and Hori (2012), the reinstatement of *Pahonia luteva* (Muscidae-Diptera) as a species distinct from *Phaonia errans* by Renaud *et al.* (2012), the recognition of two undescribed species of the genus *Tetragonopterus* (Charidae) from Central Brazil examined by Silva *et al.* (2013), and the evaluation of taxonomic status of leuciscine fish *Squalius lucumonis* collected from Tiber River (Italy) by Tancioni *et al.* (2013).

1.4.4. Molecular Genetic Studies of *Monopterus albus* species complex

A variety of molecular genetic techniques have been applied to *M. albus* due mostly to its commercial importance or for ecological reasons relating to biological invasion. These studies focused on a range of questions including taxonomy, phylogenetic relationships, origin of translocated populations, population structure and mitochondrial genomics. Many molecular genetic studies of *M. albus* have been conducted in China based on fish sampled from several regions of this country. These include traditional molecular markers such as isoenzymes (Yang *et al.*, 2005) and Random Amplified Polymorphic DNA (RAPD)

(Liu *et al.*, 2005; Yin *et al.*, 2005) and more recently to include direct sequencing of mitochondrial DNA (Cai *et al.*, 2008; Cai *et al.*, 2012) and an analysis of microsatellite markers (Lu *et al.* 2005; Li *et al.*, 2007; Lei *et al.*, 2011). Esterase isozymes were used by Yang *et al.* (2005) to determine the genetic relationships of three kinds of *M. albus* differentiated by body colour from Poyang Lake Region (China). This study revealed that esterase isozyme of *M. albus* in every population consisted of two polymorphic loci and exhibited high levels of polymorphism. Similarly RAPD was used to investigate polymorphism of *M. albus* in Dongting Lake with 12 out of 20 primers providing reproducible fingerprints (Liu *et al.*, 2005). Yin *et al.* (2005) also used RAPD profiles based on 24 random markers to investigate genetic structure of wild and cultured populations in several regions in China. While all of these studies reported polymorphisms none drew any conclusions relating to taxonomy or species boundaries. Next, direct sequencing of the mitochondrial D-loop was conducted on *M. albus* from several locations in Sichuan basin, China (Mianyang, Ya'an, and Longchang populations from Sichuan province, and Zhongxian from Chongqing province) by Cai *et al.* (2008). These authors generated 553 base pair sequences consisting of 13 haplotypes with eight polymorphic sites. They reported that populations from Zhongxian were genetically differentiated, but that all the populations showed some degree of genetic heterogeneity. The authors suggested that the populations should be treated as a monophyletic group and did not make any other taxonomically related conclusions (Cai *et al.* 2008). Cai *et al.* (2012) expanded their earlier study to eight additional geographical sampling locations for *M. albus* in China. These authors generated 501 base pair sequences from 167 individuals, which comprised 60 haplotypes with 85 variable sites. Their results also revealed phylogenetic differentiation of *M. albus*, based on the mitochondrial control region, which fell into five genetic lineages. Furthermore,

some researchers examined microsatellite markers for the *M. albus* collected from several regions in China (Lu *et al.* 2005; Li *et al.*, 2007; Lei *et al.*, 2011). Lu *et al.* (2005) described 31 pairs of primers for microsatellites and found 11 markers that showed polymorphisms with 3-13 alleles in a study of swamp eels from two Chinese populations (Hunan and Guandong) and one population from Bengal. Li *et al.* (2007) investigated 30 microsatellites from a (GT)_n-enriched genomic library and found polymorphism in 13 microsatellite markers with 2-13 alleles in a test swamp eel population from China. Next, Lei *et al.* (2011) developed two sets of multiplex loci for PCR to fast genotype 11 loci that showed 2-20 alleles. In this study they compare genetic diversity of *M. albus* collected from two populations (Zhejiang and Hanan) based on these microsatellite markers.

Other molecular genetic studies of *M. albus* have been conducted by Collins *et al.* (2002) and Matsumoto *et al.* (2010). Collins *et al.* (2002) investigated 16S mitochondrial rRNA sequence variation and identified three distinct lineages from four populations in the continental United States (Atlanta, Georgia; Tampa, Florida, North Miami Florida; and Homestead, Florida). This research yielded 18 unique haplotype-locality combinations supporting several independent introduction of the species and indicated that at least two or possibly three cryptic species are present among the introduced *M. albus* populations in the US. Thus genetic analyses support multiple introductions originating from unknown, but different geographic locations from within the species' natural distribution in Asia. The "*M. albus*" from two populations (Tampa, Florida and North Miami, Florida) were similar to samples from Nanning (China) being identified as a *Monopterus* clade C. The "*M. albus*" from Homestead, Florida which is about 40 km from North Miami (Florida) were quite similar to samples from Jakarta (Indonesia), Ca Mau (Vietnam), and Kulala Lumpur (Malaysia), and

classified as *Monopterus* clade B. Surprisingly, the samples from Atlanta (Georgia) were identified as a third form and denoted as *Monopterus* clade A, most likely originating from Japan or Korea. Based on these results Collins *et al.* (2002), suggested that *Monopterus* clade B, including specimen from Java (Indonesia), should be referred to as *Monopterus javanensis* while *Monopterus* clade C might represent the true *Monopterus albus*. It was proposed that *Monopterus* clade A may represent a distinct and possibly unnamed species.

Next, Matsumoto *et al.* (2010) undertook a comprehensive molecular phylogenetic analysis of mitochondrial 16S rDNA sequence variation in this species focusing largely on populations from the north-eastern part of its distribution. Based on this study this fish species was found to comprise three clades that correlate with geographical location and reproductive behaviour, and the authors suggested that the fish species is composed of at least three distinct species. They proposed the Japanese main island populations (Honshu and Kyushu), Chinese populations (Shanghai) and Taiwan population (Taipei) as a single species, and they identified Ryuku Island populations as an endemic native species. In addition, they suggested that the Southeast Asia populations Puli and Hengchuen (Taiwan), Fuzhou and Haikou (China), and Yogyakarta (Indonesia), which have the highest genetic diversity, may include more than one species. In this study, Matsumoto *et al.* (2010) also stated that based on previous field observations, the reproductive behaviour of "*M. albus*" differs significantly among the three geographical groups. In the China-Japan populations, the eggs are spawned into a foam mass where they are fertilized and are subsequently cared for by the male who keeps the larvae in his buccal cavity until the larvae begin to respire (Matsumoto and Iwata, 1997). In contrast the Ryuku population, the breeding males do not keep the larvae in his buccal cavity, which is narrower

(Matsumoto *et al.*, 2007). Lastly, in Southeast Asian populations, the eggs are spawned on plant leaves under water where the embryos develop into the larvae without parental care (Matsumoto *et al.*, 2010). This data, therefore, supports the hypothesis that the swamp eel "*M. albus*" is a species complex consisting of species with differing reproductive biologies. Moreover, the result strongly supported Collins *et al.* (2002) data that "*M. albus*" introduced into the southeastern United States comprises at least three species, two of which are similar to forms from Southeast Asia based on their molecular phylogenetic relationships.

The complete mitochondrial DNA nucleotide sequence of *M. albus* was described by Miya *et al.* (2001, 2003) (GenBank accession number: NC_003192). This provides a valuable resource for the design of primers to investigate genetic relationship within or between populations of *M. albus*. For instance, the molecular genetic analysis of *M. albus* populations collected from four regions in Sichuan basin (China) analysed by Cai *et al.* (2008) used primers designed from Miya *et al.*'s data. Unfortunately the origin of the specimen sequenced by Miya *et al.* (2001) was not given or was unknown.

The above results, although derived from data from different gene fragments and incomplete or inconsistent sampling raises the possibility that multiple species may be present in what is currently referred to as *M. albus*. The studies indicate significant divergence can be present over small geographic scales, and so far sampling in the southern part of the species distribution especially across the Indonesian archipelago has been limited.

There have been a number of studies on the reproductive biology of *M. albus*. This is due to *M. albus* having a small genome size and natural sex reversal. Therefore, *M. albus* is considered to be an ideal vertebrate model for

studying molecular mechanisms of sex-determination and evolution of sex reversal. Several genes potentially contributed in sexual development of *M. albus* have been investigated such as Sox11 and Sox19 (Liu and Zhou, 2001), SRY-related genes (Zhou *et al.*, 2002), Sox9a (Lu *et al.*, 2003; Zhou *et al.*, 2003), CPY17 (Yu *et al.*, 2003), Sox 17 (Wang *et al.*, 2003), osteoclast-stimulating factor (OSF-like gene) (Xia *et al.*, 2004), *dmrt1* genes (Huang *et al.*, 2005), *vasa*-like gene (Ye *et al.*, 2007), aromatase (P450arom) and 11 β -hydroxylase (P45011 β) (Yu *et al.*, 2008; Liu *et al.*, 2009). These genes being studies for sex development may also be useful for comparative genomic and phylogenetic studies and to understand the evolutionary biology of this and related fish species.

1.5. Research Objectives and Thesis Format

The overall aim of this study is to advance the understanding of the taxonomy and evolution of the *Monopterus albus* species complex with a special focus on Indonesian populations. Following from this objective there are three major components to this thesis. First, this study aims to clarify species boundaries in the swamp eel focussing on Indonesian populations. This section takes a barcoding and population genetic approach to investigating species boundaries. Second a broad based molecular taxonomic and phylogenetic study of swamp eel (*M. albus*) is undertaken using sequences from several gene regions, with comprehensive geographic sampling and utilisation of published nucleotide data sets. Lastly, an investigation of variation in morphological characters within and between selected Indonesian populations is undertaken making use of numerical taxonomic methods for comparison with previous morphological-based taxonomic treatments and the genetic data.

More specifically, in relation to the format of the thesis, the first research chapter (Chapter 2) presents evidences of cryptic species of *M. albus* using barcoding (COI mtDNA) and population genetics (five polymorphic microsatellite loci). A more extended study was conducted of COI variation among Indonesian populations using Denaturation Gradient Gel Electrophoresis (DGGE). Chapter 3 extends this study with several phylogenetic analyses of swamp eel populations using multi gene sequences from mitochondrial 16SrDNA (two fragments) and cytochrome c oxydase subunit I (COI) gene regions and from nuclear RAG-1 and 1st intron of S7 gene regions. A number of separate analyses were conducted in order to utilise sequences from existing studies that used sequenced from different 16SrDNA regions by Collins *et al.* (2002) and Matsumoto *et al.* (2010). Finally, Chapter 4 presents an analysis of morphological variation within and between genetically differentiated Indonesian swamp eel populations studied in the preceding chapter using univariate, bivariate and multivariate analyses, with the broader implications of this study and the prospects for future research discussed in Chapter 5.

CHAPTER 2

Molecular genetic evidence for cryptic speciation in the *Monopterus albus* species complex in central Indonesia

2.1. Introduction

Cryptic species complexes are found in many taxonomic groups including fishes and provide challenges for both ecologists and taxonomists (Hebert *et al.*, 2004; Cerutti-Pereyra *et al.*, 2012). Accurate identification of species is fundamental not only to biology in general but especially for biodiversity conservation and fishery management (Dunham, 2004; Bortolus, 2008; Ladner and Palumbi, 2012). As a result taxonomists are increasingly using molecular tools routinely to reveal cryptic species and more confidently define species boundaries (Zemlak *et al.*, 2009; Ni *et al.*, 2012; Puillandre *et al.*, 2012).

Over recent years a popular molecular-based approach to taxonomic investigations has been DNA barcoding which uses sequences from an approximate 650-bp segment of the cytochrome c oxidase subunit I (COI) mitochondrial gene. This has been promoted by the foundation of the Consortium for the Barcode of Life (<http://www.barcodeoflife.org>), which advocates this approach as a standard method to support both taxonomic identification, with reference to their online database (<http://www.boldsystems.org>), and the discovery of new species of animals (Hebert *et al.*, 2003a,b). This COI-based identification system has been applied to a variety of fish taxa including rays and sharks (Holmes *et al.*, 2009; Cerruti-Pereyra *et al.*, 2012), salmon and trout (Rasmussen *et al.*, 2009), and giant perches (John *et al.*, 2010)

Even though DNA barcoding using the relatively fast evolving COI gene has been demonstrated to be successful in the identification of closely related species and also revealing cryptic speciation and phylogeographic structures within a species, it has been a target of criticism (Moritz and Cicero, 2004; Lee, 2004; Ebach and Holdrege, 2005; Rubinoff *et al.*, 2006). This is due to proponents often adopting an artificial or arbitrary distance measure to distinguish within and among species variation, failing to consider stochasticity in these distance estimates and heterogeneity in evolutionary rates and lacking sensitivity to incomplete sampling (Meier *et al.*, 2006; Rubinoff *et al.*, 2006; Whitworth *et al.*, 2007). An additional criticism relates to the simplistic use of genetic distance to construct neighbour-joining trees from these sequences for taxonomic interpretation and the problem of applying the biological species concept within this context (Cognato, 2006; Nunes *et al.*, 2014). Other issues that can limit the effectiveness of DNA barcoding is the occurrence of heteroplasmy and numts (nuclear mitochondrial-like sequences), incomplete lineage sorting and mitochondrial lineages being discordant from the nuclear genome, sequencing, sequence-editing and database errors and of course misidentification or misapplication of names for samples and lack of voucher specimens (Hebert *et al.*, 2003a,b; Moritz and Cicero, 2004; Hajibabaei *et al.*, 2007; Buhay, 2009; Asgharian *et al.*, 2011).

As a consequence, it is beneficial if the results of DNA barcoding can be supported by other molecular data from rapidly evolving nuclear markers that are commonly used in population genetics such as allozyme electrophoresis, microsatellite and single nucleotide polymorphisms (SNPs), which are used to demonstrate species differentiation including evidence for reproductive isolation. The former two techniques are the most commonly used and have the advantage of being codominant markers (Austin and Ryan, 2002; Thai *et al.*, 2006;

Fassatoui *et al.*, 2009). However, microsatellites are more informative due more rapid rates of evolution and higher allelic diversity per locus compared to allozymes. SNPs are less informative as they are mostly bi-allelic and therefore require the genotyping of far more loci to generate an equivalent amount of information (Shaffer and Thompson, 2007; Hausdorf and Hennig, 2010). The combination of the use of DNA barcoding and microsatellite markers to confidently identify cryptic speciation have been applied in several animals including spruce budworm (Lumley and Sperling, 2011), galaxiid fish (Vanhaecke *et al.*, 2012), and bees (Hurtado-Burillo *et al.*, 2013)

Freshwater fish make up one of the most diverse groups of vertebrates with over 13,000 species recognised worldwide, and many more species are waiting to be discovered and described in the tropics, especially in countries where exploratory surveys are still incomplete such as Brazil and Indonesia (Lévêque *et al.*, 2008, Allen and Erdmann, 2012). In Indonesia, the total number of freshwater fishes is currently about 1,400 or 7% of total global freshwater fish species diversity (Rhee *et al.*, 2004). The actual number of species is likely to be much higher however, as the application of systematic sampling and associated taxonomic studies together with the use of molecular methods for the discovery of cryptic species has been relatively limited compared to other parts of the world.

In Indonesia, it is generally accepted that the common swamp eel belongs to the species described as *Monopterus albus* Zuiew 1793 (Saparinto, 2009; Roy, 2013). This fish species is one of an important group of freshwater food fish in the country with a wide distribution extending across Sumatra, Jawa, Bali, Lesser Sundas, Sulawesi and the Moluccas (Kottelat *et al.*, 1993; Berra, 2007). However, much debate has centred on delimitation and distribution of the species since it was described by Zuiew 1793 (Rosen and Greenwood, 1976). The nominal species has a very wide distribution extending from Indonesia in the

south to Japan and China in the north. The taxonomy of *M. albus* in Indonesia is largely unknown, as no comprehensive revision has been performed since the early taxonomic treatments. As a consequence of recent genetic research it is becoming clear that *M. albus* represents a species complex and the number and identify of swamp eel species in Indonesia is an open question.

No COI-based DNA barcoding studies have been conducted for *M. albus* even though this marker has become widely used for fish groups over the last 10 years. The relatively limited genetic research on species boundaries and population structure of the *M. albus* species complex has been primarily based on nucleotide sequence data from the mitochondrial 16SrDNA gene (Collins *et al.*, 2002; Matsumoto *et al.*, 2010) and D-loop (Cai *et al.*, 2008; Cai *et al.*, 2012). This research supports clearly the existence of cryptic species of *M. albus* but each study has limitation with respect to sampling especially in Indonesia. The studies by Collins *et al.* (2002) and Matsumoto *et al.* (2010) only included swamp eels from Jakarta and Yogyakarta thus the knowledge of genetic variation of this species from Indonesia is quite limited and population genetic variation is poorly understood. Furthermore, while microsatellite variation has been investigated in swamp eels by Li *et al.* (2007) and Lei *et al.* (2011) who developed primers for several microsatellite loci they only examined samples collected from China. No microsatellite studies have been done for *M. albus* from any other country.

Biologically, swamp eels are considered to be non-migratory species and are likely to have highly restricted dispersal ability over their geographic range. This would favour reproductive isolation between populations leading to genetic divergence and over long periods of time, could lead to speciation (Fergusson, 2002). Thus within the Indonesian Archipelago, consisting of multiple islands, *M. albus* may consist of several endemic cryptic species. Alternatively, the species may have high levels of genetic similarity among widely separated populations if

dispersal has been facilitated by humans as a result of their popularity as a food fish and for aquaculture and their tolerance to transportation. It is also possible that both factors, cryptic speciation and translocation, are significant factors leading to complicated geographic patterns of genetic diversity within this species complex.

Given the limited observed morphological variability within swamp eels (Rosen and Greenwood, 1976), here, I focus on a comprehensive genetic analysis of *Monopterus albus* species complex collected from Indonesia with reference to samples from Vietnam and Taiwan. In the first part of this study, I use DNA barcoding based on sequence variation at a fragment of the mitochondrial cytochrome c oxidase subunit I gene (COI) to determine if there is evidence for more than one species within the *M. albus* species complex over its geographic range in Indonesia.

In a second part to this study, I examined in more detail a complex geographic patterns of genetic variation within swamp eels from central Indonesia using a combination of Denaturing Gradient Gel Electrophoresis (DGGE), mitochondrial cytochrome c oxidase subunit I (COI) gene, and analysis of variation at microsatellite loci. DGGE is very powerful genetic fingerprinting tool developed by Fischer and Lerman (1980). DGGE was used to efficiently explore and assess genetic patterns of variation within and between populations for two divergent COI haplotypes revealed in the first part of the study. In addition, I carried out an analysis of genetic diversity and population differentiation of Indonesian *M. albus*, using five microsatellite markers for the same individuals genotyped for COI using DGGE.

2.2. Materials and Methods

2.2.1. Sample collection for COI mitochondrial sequencing

Tissue samples of swamp eels were collected or obtained from Indonesia (29 sites), Vietnam (1 site) and Taiwan (1 site) (Figure 2.1). Samples of *Ophisternon* species including *Ophisternon bengalense* from Maros, South Sulawesi (Indonesia), *Ophisternon gutturale* from Rapid Creek, Darwin (Australia) and *Ophisternon* sp. from Angurugu River, Groote Eylandt (Australia), were used as out-groups. Sampling location, sample code, and sample size details are provided in Table 2.1. The tissue samples consisted of 50-100 mg of muscle tissue, which was dissected with a sterilized surgical scissor, placed into a 1.5 ml screw top cryogenic vial, and preserved in 95% ethanol in the field and stored at -20°C in laboratory.

2.2.2. DNA extraction, amplification and sequencing of cytochrome oxidase c subunit I (COI) mitochondrial gene: procedures and analysis

Total genomic DNA was extracted from muscle tissue of each specimen using DNeasy blood and tissue kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. Two individuals from each population, except for Lake Sentani (Papua) for which only one individual was available, were analysed by direct sequencing of the COI mitochondrial gene. The COI mitochondrial gene was amplified using primers FishF2 (5'-TCGACTAATCATAAAGATATCGGCAC-3') and FishR2 (5'-ACTTCAGGGTGACCGAAGAATCAGAA-3') (Ward *et al.*, 2005). The KAPA2G Robust PCR kit (Kapa Biosystems) was used for the polymerase chain reaction (PCR). PCR amplifications were conducted in 25 µL

reaction volumes containing 10-100 ng of genomic DNA, 0.2 mM of each dNTP, 2mM MgCl₂, 0.014U Tag Polymerase, 0.6 μM of each primer and 1x PCR reaction buffer. The thermal cycler profile consisted of a 2 min denaturation at 94°C followed by 35 cycles of 50 s at 94°C, 2 min at 49°C, and 1.5 min at 72°C. A final extension of 6 min at 72°C was performed. Then, PCR products were visualized using 1% agarose gels and cleaned using Viogene PCR purification kit (Viogene Inc.). Sequence reactions were performed in both directions using the Big Dye Terminator Ver. 3.3. sequencing kit (Applied Biosystems), 5-7 μL purified PCR product, and 0.8 μM of either primer per reaction. Sequence-reaction products were then loaded into an ABI 3130xL Genetic Analyzer (Applied Biosystems). Amplicons were sequenced in both forward and reverse directions.

COI sequences were visualized and edited using SeqMan and Editseq Pro Program Lasergene DNASTAR software package (DNASTAR Inc., Madison, USA). Sequences were transferred to fasta format and aligned with the opal multiple sequence alignment routine implemented by the MESQUITE 2.74 package (Maddison and Maddison, 2010) and ClustalW in MEGA5 (Tamura *et al.*, 2011). Subsequent inspection and editing of the alignments was done manually. Chromatograms were inspected for noisy and ambiguous base calling and translated to check for stop codons. Noisy tails were trimmed. For each individual, sequencing reactions were performed using both forward and reverse primers, resulting in a consensus fragment of 514 bp in length. Therefore the data set used for phylogenetic analysis was composed only of those sequences that consisted of a minimum of 514 bp after trimming. Intraspecific diversity was estimated as number of haplotype, haplotype diversity, number of polymorphic sites, and nucleotide diversity using the software DnaSP 5.10.01 (Librado and Rozas, 2009). Kimura-2-parameter model were used to estimate intraspecific and interspecific genetic distance (Kimura, 1980) and was summarised in an NJ tree

as this is the standard methodology used in barcoding studies (Hebert *et al.*, 2003a). The NJ tree with bootstrap values (1000 replication) was constructed using PAUP* version 4 (Swofford, 2003). Sequences of *M. albus* (NC_003192) and *M. cuchia* (FJ459508-FJ459511) from GenBank (www.ncbi.nlm.nih.gov) were used for comparative purposes. Phylogenetic relationships were also estimated using a Bayesian approach. The best-fit model of evolution was selected with jModelTest 0.1.1 (Posada, 2008) under the Akaike information criterion (AIC) suggested by Posada and Buckley (2004). MrBayes 3.2 (Ronquist *et al.*, 2012) was used for Bayesian inference under the best-fit model. The analyses were run for 2,000,000 generations with a sampling frequency set to every 1000 generation. The analysis was done until the standard deviation of split frequencies was below 0.01. The analysis used a relative burn-in of 25% for diagnostics. Consensus trees were visualised in FigTree 1.4.0 (Rambaut, 2012). Principal Coordinate Analysis (PCoA) based on genetic distance of COI mitochondrial gene was carried out in GenAlEx version 6.5 (Peakall and Smouse, 2012) to examine the patterns of divergence between haplotypes.

2.2.3. Sample collection for Denaturation Gradient Gel Electrophoresis (DGGE) of COI variation

As the initial DNA-barcoding part of the study suggested the possibility of cryptic speciation and population admixture a more detailed study of COI variation was undertaken using DGGE. Guided by the COI barcoding results fifteen populations of swamp eels were sampled from several regions in central Indonesia. The sample size used for the DGGE procedures was 20 individuals from each population. Details of sampling locations and codes are provided in Table 2.1. All fish were captured by farmers and fishermen using nets and

transported to the laboratory on ice and then frozen. Tissue samples were then dissected from each partially thawed fish and placed into 1.5 ml screw top cryogenic vials and preserved in 95% ethanol.

2.2.4. DNA amplification, optimization and procedures for Denaturation Gradient Gel Electrophoresis (DGGE) of COI variation

A short highly variable fragment of the COI was selected for DGGE analysis based on the results of the first part of the study. This fragment is approximately 280 bp and allowed discrimination of the two divergent forms of swamp eel found in the first part of the study. The DGGE procedure required one of the primers to be modified at the 5' end with a GC-Clamp. The two primers designed for DGGE analysis are: EELCOIF (5'-ATAGTMATGCCYAATATAAT YG-3') and EELCOIRGCCLAMP (5'-CGCCCGCCGCGCCCGCGCCCGTCC CGCCGCCCCCGCCCGTRTTAAGGTTTCGRTCRGTG-3'). The KAPA2GRobust PCR kit (Kapa Biosystems) was used for the polymerase chain reaction (PCR). PCR reactions were performed in 25 µl final volume containing 10-100 ng of genomic DNA, 0.2 mM of each dNTP, 2 mM MgCl₂, 0.014U Tag Polymerase, 0.6 µM of each primer and 1x PCR reaction buffer. Reactions were amplified using the following cycling conditions: an initial denaturation 94°C for 1 min; 2 cycles of 30 s at 94°C, 30 s at 48°C, 30 s at 72°C; 2 cycles of 30 s at 94°C, 30 s at 49°C, 30 s at 72°C; 33 cycles of 30 s at 94°C, 30 s at 50°C, 30 s at 72°C and a final extension of 10 min at 72°C. All PCRs were carried out in a Labnet Multigene Thermal Cycler (Labnet International Inc., USA).

DGGE was performed with the Ingenuity PhorU-2 apparatus (Ingenuity International, The Netherlands) as per manual instructions using a gradient mixer (Ingenuity International, The Netherlands) to form the linear denaturation gradient.

Gels were made from acrylamide: bis-acrylamide (ratio - 37.5:1) and were polymerized by adding 120 µl of 20% (w/v) ammonium persulphate and 12 µl of N,N,N,N,-tetramethylethylenediamine to 24 ml of polyacrylamide solution. Approximately 50 ng of each PCR amplicon mixed with an equal volume of loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol and 15% (v/v) glycerol) was loaded on 6.5% polyacrilamide gels in TAE (1x) buffer. Optimal separation was achieved with a parallel denaturing gradient of urea-formamide ranging from 20% to 40% (100% corresponded to 7M urea and 40% v/v formamide). Gels were run for 17 h at 60°C and 75V and stained with SYBR Gold (8 µl in 50 ml of 1xTAE) for 15 min. The banding patterns were then photographed under a UV transilluminator and analysed using Quantity One 1-D analysis software (Bio-Rad, USA). DGGE haplotype frequencies were analysed using GenAEx version 6.5 (Peakall and Smouse, 2012).

2.2.5. Sample collection and Restriction Fragment Length Polymorphism (RFLP) amplification procedures prior to microsatellite analysis

Prior to the microsatellite analysis, the Restriction Fragment Length Polymorphism (RFLP) method was used to genotype additional samples from selected populations. RFLP analysis was chosen as it was quicker and less expensive than the DGGE procedure.

The choice of populations and sample sizes for microsatellite analysis were based on the results obtained from phylogenetic analysis of COI mitochondrial data and DGGE results. Specifically, microsatellite analysis was used to study the population genetics of populations showing admixtures of two divergent haplotypes revealed in the first parts of the study. The RFLP procedure

was used to allocate additional swamp eel samples to one or other of two dominant COI haplotypes resolved by direct sequencing or DGGE analysis.

Amplified product of the COI mitochondrial gene from 588 samples from 20 selected populations were digested with SpeI restriction endonuclease in order to allocate each individual into one or other of the two dominant haplotypes. The SpeI restriction enzyme was chosen as it cuts one of the fragments at a site in one haplotype (Haplotype A) but leaves the other uncut (Fig. 2.6). The RFLP reaction was carried out in 10 μ L final volumes, containing 1 μ L of Restriction Buffer 10x, 0.1 μ L of 10 μ g/ μ L Bovine Serum Albumin (BSA), 2 μ L of PCR product and 0.25 μ L of 10u/ μ L SpeI Restriction Enzyme (Promega). The digestion was set up for 3 hours at 37°C. The restriction fragments were visualised by electrophoresis in 1% agarose gels, stained with gel red nucleic acid stain (Biotium). The haplotypes for individuals were scored directly from the gel.

2.2.6. Microsatellite procedures and analysis

Five microsatellite loci *Mal01*, *Mal 07*, *Mal13* (Li *et al.*, 2007), *Mal007* and *Mal 008* (Lei *et al.*, 2011) were used in this study. These microsatellite markers were selected based on the observed level of polymorphism in previous published studies (Li *et al.*, 2007; Lei *et al.*, 2011). Primer pairs were also chosen on the basis that annealing temperature were similar thereby allowing multiplexed PCR. The forward primer of each locus was “tailed” to allow annealing of a fluorescently labelled oligonucleotide following the method of Schuelke (2000). For each sample, two PCR reactions were performed for two different primer combinations: 1) *Mal01*, *Mal07*, and *Mal13*; 2) *Mal007* and *Mal008*. Details of five

polymorphic loci, primer sequence, fragment size and dye are provided in Table 2.2.

PCR of microsatellites was performed using a reaction volume of 12.5 μ L containing 10-100 ng of genomic DNA, 2xQiagen Multiplex PCR Master Mix and 0.3 μ M of each forward and reverse primer. Thermal cycles employed for multiplex (1) were initial denaturation at 94°C for 1 min, followed by 33 cycles of 94°C for 30 s, 57°C for 30 s and 72°C for 40 s, and then a final extension at 72°C for 6 min; and that of reaction (2) was similar, except that the annealing temperature was 52°C. The amplified products were mixed with internal size standard (GeneScan™500 LIZ®, Applied Biosystems) and deionised formamide. The mixture was subjected to capillary electrophoresis on an ABI 3130x1 Genetic Analyzer (Applied Biosystems). Fragments were sized using GeneMapper version 4.0 software (Applied Biosystems) and allele scores were verified manually.

Microsatellite Tool Add-in for excel (<http://animalgenomics.ucd.ie/sdepark/ms-toolkit/>) was used to convert data in an excel file into format suitable for the statistical software packages that I used. GENEPOP version 4.2. (Rousset, 2008) was used to test genotypic distributions for conformance to Hardy-Weinberg expectations, and to test the loci for genotypic disequilibrium. Allele frequency, observed and unbiased expected heterozygosities under Hardy-Weinberg expectation were calculated using ARLEQUIN version 3.1. (Excoffier *et al.*, 2006). Number of alleles at each locus for each sample was estimated using FSTAT version 2.9.3. (Goudet, 2001). Significance levels were determined using the Markov chain method (dememorisation number=1000, batches=100, iteration=1000). Levels of significance were adjusted for the number of simultaneous tests using the sequential Bonferroni procedure (Rice, 1989). An assignment test was implemented by GenAIEx version 6.5 (Peakall and Smouse,

2012) and GeneClass Version 2.0 (Piry *et al.*, 2004). An assignment test was used in order to determine the extent to which individuals could be correctly assigned to their population of origin (Paetkau *et al.*, 2004).

Principal Coordinate Analysis (PCoA) based on the multilocus genotypes was carried out in GenAlEx version 6.5 (Peakall and Smouse, 2012) and GENETIX version 4.05 (Belkhir *et al.*, 2004) in order to examine patterns of genetic variation and identify intermediate genotypes (if any) resulting from population admixture. The microsatellite genotypes for the 588 individuals were also analysed using STRUCTURE 2.3 (Pritchard *et al.*, 2000) under the admixture model with no prior information and $K=2$ to define appropriate Q for individual assignment. Burn-in period was 50,000 steps and 100,000 MCMC iterations were used, together with the correlated allele frequencies and the admixture model options.

2.3. Results

2.3.1. COI mitochondrial sequence variation and phylogenetic relationships

Sixty one COI sequences were obtained from *M. albus* individuals from the 31 sample locations (Figure 2.1 and Table 2.1), which represented 19 distinct haplotypes. The 514-bp alignment contained 117 polymorphic sites with 114 being parsimony-informative (Figure 2.4) and amino acid sequence translations (vertebrae mitochondrial code) were unambiguous as there were no gaps or nonsense codons among the 61 sequences. Sequences representing each divergent COI haplotype have been deposited in GenBank with accession number: KP729549-KP729589.

The optimal model of nucleotide sequence substitution for the COI matrix including the outgroup samples was the HKY model with gamma-distributed rate variation across sites and proportion of invariable sites ($\alpha=0.953$, proportion of invariant sites=0.047) as inferred by jModelTest 0.1.1. The Bayesian analysis of COI mitochondrial data, together with additional sequences of *M. albus* and *M. cuchia* taken from GenBank, revealed that the swamp eels fall into four distinct clades (designated as Clade A, B, C, D), with the associated nodes supported by posterior probabilities of 0.88, 0.85, 1.00, and 1.00 respectively (Figure 2.2). Similarly, the neighbour-joining tree exhibited four distinct clades (Clade A, B, C, and D) with bootstrap value of 89, 99, 100, 100% respectively (Figure 2.3).

The K2P distance among all COI haplotypes within *M. albus* was quite variable ranging from 0.0 to 18.61% (mean= 3.57%). The average nucleotide sequence divergence between haplotypes within clade A, clade B, and clade D is 0.97% (range 0-3.60%), 0.84% (range 0-2.79%) and 0.52% (range 0-0.78%) respectively. The two samples representing clade C had identical haplotypes. Levels of between-clade divergences are shown in Table 2.3. The highest genetic divergence among clades was observed between individuals from clades B and D ranging from 17.08% to 18.59% (mean 18.31%), whereas least genetic divergence was detected between individuals from clades A and B ranging from 6.13% to 8.78% (mean 7.45%). Levels of differentiation within and between *M. albus* clades is summarised by Principal Coordinates Analysis (PCoA) in Figure 2.5. This analysis indicates that the degree of differentiation between clades is much greater than within clades.

The four clades or OTUs generally show strong geographic patterns of differentiation. The most strongly differentiated clade (Clade D) comprises the northern samples from Taiwan and an *M. albus* GenBank sample most likely from

Japan. The two Vietnamese samples from further south form a separate distinct clade (Clade C). This clade is in turn distinct from Clades A and B, which were only found from the Indonesia archipelago well to the south. The samples forming Clade B have a relatively wide distribution, which includes the island of Sumatra, the western and central part of the Island of Java and outlying populations in southern Sulawesi and Lake Sentani in Papua in the far west. Representatives of Clade A have a restricted distribution, which extends from central Java to the islands of Lombok and Bali. While no sampling sites were found to be polymorphic for haplotypes from clades A and B, the geographic distribution of these clades show considerable overlap in the central part of east Java. Within this region, sampling sites PKY, BRS, GSK, NGT, PLG, SRG, TWG, CPK, KYR, WLK, SKP, DMP, and PJK had clade A haplotypes and sampling sites KMR, PLB, and GMP had clade B haplotypes (Figure 2.1, map 3).

2.3.2. Variation within clades

Eleven COI haplotypes were found in clade A from 38 individuals and five haplotypes were observed in clade B from 19 individuals. Clade C comprised one haplotype from two samples and clade D has two haplotypes, one from Taiwan and one representing the GenBank sample of *M. albus* (Table 2.4). Excluding the outgroup samples and *M. cuchia* there were 28 variable sites of which 24 were parsimony-informative sites, and for clades A and B there were 17 variable sites with 16 parsimony-informative sites. In clade C there were four variable sites none of which were parsimony informative. The haplotype diversities (h) within clades ranged from 0.0 to 0.89 and the nucleotide diversities (π) from 0.0 to 0.0096. The most genetically diverse clade was clade A ($h=0.89$, $\pi=0.0096$),

whereas the least variable clade was clade C ($h=0$, $\pi=0$) with intermediate values for clade B ($h=0.69$, $\pi=0.0084$) and clade D ($h=0.67$, $\pi=0.0052$).

Within Clade B the level of divergence amongst most haplotypes was low with haplotypes differing between 1–6 bp. The one exception being sample PYK from Sumatra with haplotype (B2) which differed from all other haplotypes within clade B by 13 to 15 base pairs. Similarly, most haplotypes within clade A only differed by a few bp. One exception was finding one relatively divergent haplotype – A9 (differing by 12 to 18 bp), in just three individuals from two populations (NGR and LMJ). Population LMJ showed the highest within population variation with the two haplotypes differing by 14 bp. For the majority of populations each of the two individuals showed identical haplotypes. The most wide spread haplotype (B3) was found in six populations distributed from West Java to Papua. Differentiation among haplotypes within and among clades is summarised by the Principal Coordinate Analysis (PCoA) (Figure 2.5).

2.3.3. DGGE analysis of COI variation

On the basis of the phylogenetic analysis of the COI mitochondrial gene demonstrating four clades and evidence of cryptic species of *M. albus*, this part of the study focused on a more detailed analysis of *M. albus* from Indonesia. For this study, I selected seven populations from clade A and eight populations from clade B with each population represented by 20 individuals, with DGGE analysis used to allocated individuals to each COI haplotype. Twelve populations were chosen as they represented both haplotypes from a restricted geographic area on Java and nearby islands to the east. Two populations were also chosen from the island of Sumatra to the west and a single population from southern Sulawesi to the east for comparative purposes. The seven populations representing clade

A were TWG, PKY, PLG, BRS, WLK, NGR and NRM and the eight populations representing clade B were PDG, PYK, CMS, DPK, KMR, PLB, GMP and RPG.

Three hundred individuals from the 15 populations were effectively genotyping using DGGE (Figure 2.6) as either haplotypes A or B corresponding to Clade A and Clade B in the phylogenetic analysis (Figure 2.2 and Figure 2.3). Five populations (PKY, PLG, TWG, WLK, and NRM) were monomorphic for haplotype A, six populations (PDG, PYK, DPK, CMS, KMR, and RPG) were monomorphic for haplotype B and four populations (BRS, GMP, PLB, and NGR) were polymorphic for haplotypes A and B, herein referred to as the 'admixture' populations (Figure 2.6). In the admixture populations, the frequency of haplotype A ranged from a high of 0.95 in population BRS to a low of 0.10 in population PLB (0.10) (Figure 2.6). While haplotype A was only present in the eastern populations (excluding Sulawesi) and haplotype B was predominate in the western samples (Sumatra and western Java) there was no consistent geographic pattern in haplotype frequency among the populations from central Java (Figure 2.6). Haplotype frequencies were often quite different between adjacent populations and the polymorphic populations ranged from central Java (PLB, BRS, and GMP) to the island of Bali (NGR).

2.3.4. Microsatellite variation and patterns of divergence

The twenty populations surveyed for variation at five microsatellite loci included the 15 samples previously subject to DGGE analysis and five additional populations from central Java. The majority of sampling was focussed on the region of highest haplotype diversity in central Java. As a larger sampling of eels was used, prior to microsatellite genotyping, all individuals were first assigned to COI mitochondrial haplotype using RFLP analysis (Figure 2.7). A total of 588

individuals were genotyped with 415 belonging to haplotype A and 173 haplotypes B. A total of 14 populations were monomorphic for haplotype A and two populations were monomorphic for haplotype B. The same four populations as determined by DGGE analysis showed an admixture of the two haplotypes. All individuals genotyped by both methods (300) were assigned the same haplotype. Individuals from the populations showing an admixture of mitochondrial haplotypes were assigned to a subpopulation on the basis of mitochondrial haplotype (Tables 2.5 and 2.6) and as pooled populations (Tables 2.7).

A total of 58 alleles were detected at the five microsatellite loci among the 588 individuals, ranging in size from 184 to 364 bp. Population allelic frequencies are provided in Appendix 1, 2, and 3 and summaries of genetic variability are provided for each population and subpopulation categorised by mitochondrial haplotype (Tables 2.5 and 2.6) and for the four mixed populations pooled (Table 2.7). For the populations categories by haplotype the number of alleles ranged from four alleles at locus *Ma107* to 16 alleles at locus *Ma1008* with from 1 to 14 alleles per population per locus. One locus (*Ma107*) was monomorphic in population haplotype A (DMP and CPK) whereas two loci (*Ma107* and *Ma1008*) were monomorphic in the haplotype B subpopulation RPG.

Within haplotype A populations, the lowest mean number of allele per locus (2.0) was observed in Dempel, Ngawi, East Java (DMP), while the highest mean number of allele per locus (5.60) was found in Gamping, Yogyakarta (GMP-A). Average observed heterozygosity at the five loci ranges from 0.40 to 0.76 (Table 2.5). The lowest mean number of alleles per locus (2.60) within population haplotype B was observed in Rappang, South Sulawesi (RPG), while the highest mean number of alleles per locus (4.20) was found in Brosot, Yogyakarta (BRS-B). Average observed heterozygosity at the five loci ranged from 0.16 to 0.47 (Table 2.6). Within the admixture populations the variability

statistics were consistently higher. The lowest mean number of allele per locus was 5.80 (BRS), while the highest mean number of allele per locus was 7.80 (GMP). Average observed heterozygosity at the five loci ranges from 0.31 to 0.56 (Table 2.7).

A much greater proportion of significant deviations from HWE expectations occurred within the admixture populations (20 of 20 tests) compared with haplotype B populations (9 of 30 tests) and haplotype A (13 of 90 tests). Based on average F_{IS} values it can be seen that there was a consistent pattern of heterozygote deficiencies within populations that was most pronounced in the admixture populations (0.27-0.42) compared with the haplotype A (0.11-0.33) and haplotype B (0.07-0.33) populations (Table 2.5, Table 2.6 and Table 2.7).

Pairwise F_{ST} analyses indicate significant genetic heterogeneity among populations with the majority of pairwise comparisons yielding significant differences (Table 2.8). The populations categorised by the two haplotypes were well differentiated from each other ($F_{ST} = 0.25-0.58$). The haplotype B group showed more inter-population divergence ($F_{ST} = 0.06-0.34$) compared to the haplotype A group ($F_{ST} = 0-0.28$). The AMOVA analysis revealed that genetic variation among the haplotype groups is high (42%) compared to the degree of differentiation with groups (11%) (Table 2.9)

The results of the assignment test conducted on the 24 haplotype-defined populations (18 populations from haplotype A and 6 populations from haplotype B) is shown in Table 2.10. Populations show contrasting patterns with some having very high or 100% correct classification (e.g. DMP and CPK), indicating a high level of differentiation. While others show a high proportion of misclassification ($\geq 75\%$) reflecting a high level of similarity among other populations due to recent divergence or population mixing (e.g. NGT, SRG, and

WLK). The most noteworthy finding was that there were no misclassifications between samples with differing mitochondrial haplotypes.

Principal Coordinate Analysis (PCoA) of the microsatellite allelic frequencies for the *M. albus* populations revealed a clear divergence between populations with haplotype A and population haplotype B (Figure 2.8). The PCoA based on the individual genotype revealed a similar pattern but also a small number of potential hybrid individuals (12) showing intermediate scores (Figure 2.9). Most of the intermediate individuals were from a single population BRS and consisted of 9 out of 59 individuals and were all haplotype B. The other potential hybrids consisted of single individuals from populations NGR, GMP which also had haplotype B and PLB which has haplotype A. Results from admixture analysis using STRUCTURE assuming $K=2$ assigned all but 1 of 588 individuals to their correct haplotype. BRS 29 (haplotype B) was the only sample incorrectly allocated to its haplotype group and its nuclear genetic affinity to haplotype A samples can also be seen from Figures 2.9 and 2.10. The group membership coefficients for eight individuals were below 0.8 (labelled in Figure 2.10) and showed intermediate genotype scores. These corresponded to a subset of the same individuals identified in the PCoA analysis (Figure 2.9) as being of potential hybrid origin and again mostly consisted of individuals from population BRS and all were haplotype B except for the same individual from population PLB which has haplotype A. Excluding these eight individuals membership coefficients (Q1) for haplotype A individuals ranged between 0.847 and 0.997 (mean: 0.992 ± 0.018) and haplotype B membership coefficients (Q2) ranged from 0.840 and 0.997 (mean: 0.99 ± 0.021). The eight potential hybrids from the admixture analysis that corresponded to those individuals previously identified as being of hybrid origin by PCoA analysis, had the following relative admixture values of $Q1/Q2=0.322/0.678$ (PLB-13), $Q1/Q2=0.403/0.597$ (BRS-29),

Q1/Q2=0.641/0.359 (BRS-37), Q1/Q2=0.628/0.372 (BRS-41),
 Q1/Q2=0.749/0.251 (BRS-42), Q1/Q2=0.759/0.244 (BRS-50),
 Q1/Q2=0.518/0.482 (BRS-61), and Q1/Q2=0.654/0.341 (GMP-61).

2.4. Discussion

Using DNA-barcoding combined within a population genetic approach I have found strong evidence that the genus *Monopterus* is represented in Indonesia by two cryptic species, neither of which corresponds to *Monopterus albus* sensu stricto from northern Asia (Collins *et al.*, 2002; Matsumoto *et al.*, 2010; Kottelat, 2013). This conclusion is robust to different conceptualisations of species and varying approaches to species identification and consistent with recent molecular studies on *Monopterus*, other genera of the Synbranchidae and freshwater fish more generally.

Based on data from 588 individual fish, sampled from almost the full length of the Indonesian archipelago from central Sumatra to Papua, it is apparent that there are two species of *Monopterus* that have overlapping distributions and occur in sympatry in eastern Java and neighbouring islands. Further, these two species are also differentiated from each other and reference samples from Vietnam, Taiwan, and the sample sequenced for the complete mitogenome of *M. albus* on GenBank presumably from Japan. This conclusion is based on multiple lines of evidence, starting with DNA barcoding.

The DNA barcoding approach is used to support species delimitations when interspecific genetic variation exceeds intraspecific variation to such an extent that a “barcoding gap” exists, and that sibling species pairs show reciprocal monophyly (Wiens and Penkrot, 2002; Hebert *et al.*, 2004). The extent

of the barcoding gap is often used as a simple yardstick for delineating and identifying species. Hebert *et al.* (2004) proposed that for species recognition there should be a minimum threshold where by the divergence between putative species should be 10 times the mean intraspecific variation. Since the average nucleotide sequence divergence within clade A and within clade B is 0.97% and 0.84% respectively (average 0.94%), the barcoding gap threshold would need to be 9.4%. The average nucleotide sequence divergence between clade A and D (17.37%) and between clade B and D (18.31%) greatly exceed this value and also form highly divergent and well supported monophyletic groups based on the phylogenetic analyses. Thus it can be confidently concluded that Indonesian swamp eels belong to one or more species that are well differentiated from the northern forms of *M. albus* represented by clade D contrary to what has been commonly assumed (Saparinto, 2009; Roy, 2013). However, as the average nucleotide sequence divergence between clade A and clade B (7.45%) is lower than the 9% threshold, these two clades would not qualify as discrete species even though they form discrete, relatively well supported clades, based on a strict application of the recommendations of Herbert.

Alternatively if the criterion used by Zemlak *et al.* (2009) is adopted, who considered that nucleotide sequence divergences exceeded 3.5% could be used as a rule of thumb for discriminating species, then the recognition of clade A and clade B as species would be more than justified given the average divergence level of 7.45%. Similarly, for clade C, representing the Vietnamese samples, the average nucleotide sequence divergence from clade A (8.13%) and between clade B and clade C (7.88%) would fall short of the criteria of Herbert *et al.* (2004) but would constitute a valid species based on the 3.5% divergence level of Zemlak *et al.* (2009). There is no doubt that clade C from Vietnam and clade D with an average divergence of 17.71% represent different swamp eel species.

While the conclusions based on COI distances for separate species status for clade A and clade B is somewhat inconclusive, the finding that variation within clades is much less than between clades (Figure 2.5) and that the clades remain distinct over significant geographic space, which includes a zone of geographic overlap, provide support for speciation of swamp eels in Indonesia. However as these data are based on a single mitochondrial locus, with maternal inheritance, it does not exclude the possibility that these two mitochondrial forms of swamp eel could freely interbreed where they come into contact. Finding that the genetic relationships among swamp eels derived from nuclear (microsatellite) loci correlates very strongly with the mitochondrial results very strongly supports the conclusion that clade A and clade B represent distinct species. The limited number of individuals identifiable as having a hybrid origin and the very limited evidence for any significant introgression in the four populations with an admixture of the two haplotypes from Yogyakarta, Central Java (BRS, PLB and GMP) and Bali (NGR) provides very strong support for reproductive isolation between members of these two clades thereby satisfying the criterion for biological species (Richardson *et al.*, 1986; Austin and Ryan, 2002). Extensive hybridisation and introgression is not uncommon in fishes, including between genera (Epifanio and Nielsen, 2001; Scribner *et al.*, 2001) due to external fertilisation so it is actually quite surprising to find so little evidence of hybridisation between two closely related sibling species of *Monopterus*.

The discovery of cryptic species of *Monopterus* in Indonesia and Vietnam is a novel finding but consistent with recent molecular studies of the *Monopterus albus* species complex by Collins *et al.* (2002) and Matsumoto *et al.* (2010). While their research supports clearly the existence of cryptic species of *M. albus* each study has limitations with respect to sampling especially with respect to Indonesia. These studies focussed mostly on swamp eels sampled from northern

Asia and only included a limited number of specimens from Jakarta and Yogyakarta. The studies are also difficult to compare with this study and each other as they used different regions of the 16S gene. Nevertheless their results were sufficient for Kottelat (2013) to conclude that *Monopterus* from Indonesia should be referred to *M. javanensis*, which was previously considered to be a synonym of *M. albus*. This position now needs to be re-evaluated again, given the evidence of two species of *Monopterus* on the islands of Java.

As more molecular genetic studies are undertaken it is emerging that cryptic speciation is not limited to *Monopterus* but may be common in the family Synbranchidae. Perdices *et al.* (2005) and Valdez-Moreno *et al.* (2009) found molecular evidence for cryptic speciation in the genera *Synbranchus* and *Ophisternon* based on studies in central and South America. A recent extreme example of cryptic speciation in freshwater fish is the identification of 15 species in what was thought to be a single species of *Galaxias* (Raadik, 2014).

Molecular taxonomic studies of freshwater fishes of significant scope in Asia are still relatively uncommon, however most support significant cryptic speciation in several diverse groups including *Tor* (Nguyen, *et al.*, 2008), *Pangio* (Bohlen *et al.*, 2011), air sac catfish (Ratmuangkhwang *et al.*, 2014), and *Channa* (Barman *et al.*, 2014). These studies together with my study suggest freshwater fish diversity in Asia, especially those with poorly studied faunas such Indonesia, may be significantly underestimated.

Species identification and discovery is crucial for the resolution of many biological issues such as biodiversity assessment and species conservation (Frézal and Leblois, 2008). However, species identification by morphology alone is commonly confronted with difficulties such as phenotypic plasticity, life history variation, and deficiency of taxonomic knowledge and expertise (Hebert *et al.*, 2003a). It is clear that morphological variation in the *M. albus* is complex in terms

of conservatism in the few available external characteristics including the overall head and body shape, coloration, the opening shape of gill aperture, and ventral surface of the head, which together make the delineation and identification of species difficult even with input from an expert taxonomist. Thus DNA barcoding can overcome these difficulties but, as demonstrated in this study, does not always lead to clear cut answers and has met with varying degrees of success in different groups of fish (Keskin and Atar, 2011; Cerruti-Pereyra *et al.*, 2012; Zhu *et al.*, 2013; Jo *et al.*, 2014; Sembiring *et al.*, 2015).

The strength of this study, which represents one of the most comprehensive molecular genetic studies of freshwater fish in Indonesia, is the use of multiple genetic data sets and a variety of analytical methods. Thus the outcome of this study is the confident identification of two cryptic species of *Monopterus* in Indonesia, which is strongly established. The next steps required are firstly to extend this study to include more samples from northern Asia and additional gene regions to examine confidently the deeper level relationships, biogeography and systematics of this group in a rigorous fashion. Secondly it is necessary to re-evaluate morphological variation and taxonomy of *Monopterus* in Indonesia to determine if diagnostic characters can be found that correlate with the molecular data and to review the nomenclature of *Monopterus* species for this region. Now that the existence of these two species has been established it will also be of particular interest to undertake comparative studies of the life history and ecology of the species, especially in relation to reproductive biology and behaviour, to understand how they maintain reproductive isolation in sympatry.

Table 2.1. Sampling location, sample code, geographic reference and sample size of *Ophisternon* are used as outgroups

Location	Code	Latitude (S)	Longitude (E)	Sample Size (N)	
				COI	Microsatellite DGGE
Indarung-Lubuk Kilangan, Padang, West Sumatra	PDG	00° 56' 44.36"	100° 25' 43.79"	2	20
Payakumbuh Selatan, Payakumbuh, West Sumatra	PYK	00° 12' 54.53"	100° 39' 00.46"	2	20
Depok-Pancoran Mas, Depok, West Java	DPK	06° 23' 30.25"	106° 46' 59.18"	2	20
Ciomas, Bogor, West Java	CMS	06° 33' 54.94"	106° 44' 31.97"	2	20
Soko Kembang-Kayupuring, Petungkriyono, Pekalongan, Central Java	PKY	07° 08' 48.52"	109° 43' 24.96"	2	20
Tawangmangu, Karanganyar, Central Java	TWG	07° 39' 37.98"	111° 08' 59.02"	2	28
Ngantenan-Canan, Wedi, Klaten, Central Java	NGT	07° 45' 35.10"	110° 34' 56.00"	2	20
Gesikan, Gantiwarno, Klaten, Central Java	GSK	07° 40' 35.80"	110° 33' 56.30"	2	20
Prayan-Piangu, Trucuk, Klaten, Central Java	PLG	07° 44' 46.33"	110° 41' 06.39"	2	20
Kemiri Kidul, Kemiri, Purwokerto, Central Java	KMR	07° 41' 10.84"	109° 53' 34.22"	2	25
Tunggul, Gondang, Stagen, Central Java	SRG	07° 27' 17.53"	111° 07' 22.29"	2	20
Gejawang Kilen, Balecat, Gamping, Sleman, Yogyakarta	GMP	07° 48' 07.33"	110° 18' 27.20"	2	77
Palbapang, Bantul, Yogyakarta	PLB	07° 54' 24.15"	110° 19' 24.34"	2	51
Brosot, Galur, Kulon Progo, Yogyakarta	BRS	07° 56' 29.16"	110° 13' 28.32"	2	59
Citrodiwangsan, Lumajang, East Java	LMJ	08° 06' 54.80"	113° 13' 16.86"	2	20
Pojok, Kwadungan, Ngawi, East Java	PJK	07° 30' 09.24"	111° 28' 32.04"	2	20
Cepoko, Ngrambe, Ngawi, East Java	CPK	07° 28' 07.94"	111° 12' 54.94"	2	20
Walikukun, Widodaren, East Java	WLK	07° 25' 09.83"	111° 14' 13.12"	2	25
Kayutreja, Widodaren, Ngawi, East Java	KYR	07° 26' 21.00"	111° 13' 31.00"	2	20
Sekarputih, Widodaren, Ngawi, East Java	SKP	07° 27' 35.87"	111° 15' 50.74"	2	20
Dempel, Geneng, Ngawi, East Java	DMP	07° 27' 18.79"	111° 26' 05.11"	2	20
Pohsanten, Mendaya, Negara, Bali	NGR	08° 19' 04.58"	114° 41' 08.50"	2	61
Kerobokan, Kuta Utara, Badung, Bali	BDG	08° 19' 04.58"	114° 41' 08.50"	2	-

Table 2.1. Continued

Location	Code	Latitude (S)	Longitude (E)	Sample Size (N)	
				COI	Microsatellite DGGE
Narmada, Lombok Barat, West Nusa Tenggara (NTB)	NRM	08° 39' 19.00"	115° 10' 13.07"	2	20
Sayang-Sayang, Cakranegara, Mataram, NTB	SYG	08° 34' 14.56"	116° 08' 05.10"	2	-
Seketeng, Sumbawa, NTB	SKT	08° 39' 11.65"	117° 21' 52.14"	2	-
Rappang, Panca Rijang, Sindendreng Rappang, South Sulawesi	RPG	03° 47' 00.60"	120° 00' 52.72"	2	22
Gowa, South Sulawesi	GOW	05° 18' 35.31"	119° 44' 31.63"	2	-
Lake Sentani, Sentani, Jayapura, Papua	LST	02° 36' 50.92"	140° 33' 22.18"	1	-
Thuan Thanh District, Back Ninh Province, North Vietnam	THT	21° 02' 37.75"	106° 04' 32.71"	2	-
Kaohsiung, Taiwan	KHS	23° 00' 36.44"	120° 39' 48.82"	2	-
Maros, South Sulawesi*	MRS	05° 00' 45.63"	119° 37' 46.35"	2	-
Rapid Creek, Darwin, NT, Australia**	RCK	12° 24' 53.48"	130° 55' 16.74"	1	-
Angurugu River, Groote Eyland, NT, Australia***	AGR	13° 58' 00.19"	136° 25' 35.52"	2	-

* *Ophisternon bengalense*

** *Ophisternon gutturale*

*** *Ophisternon* sp.

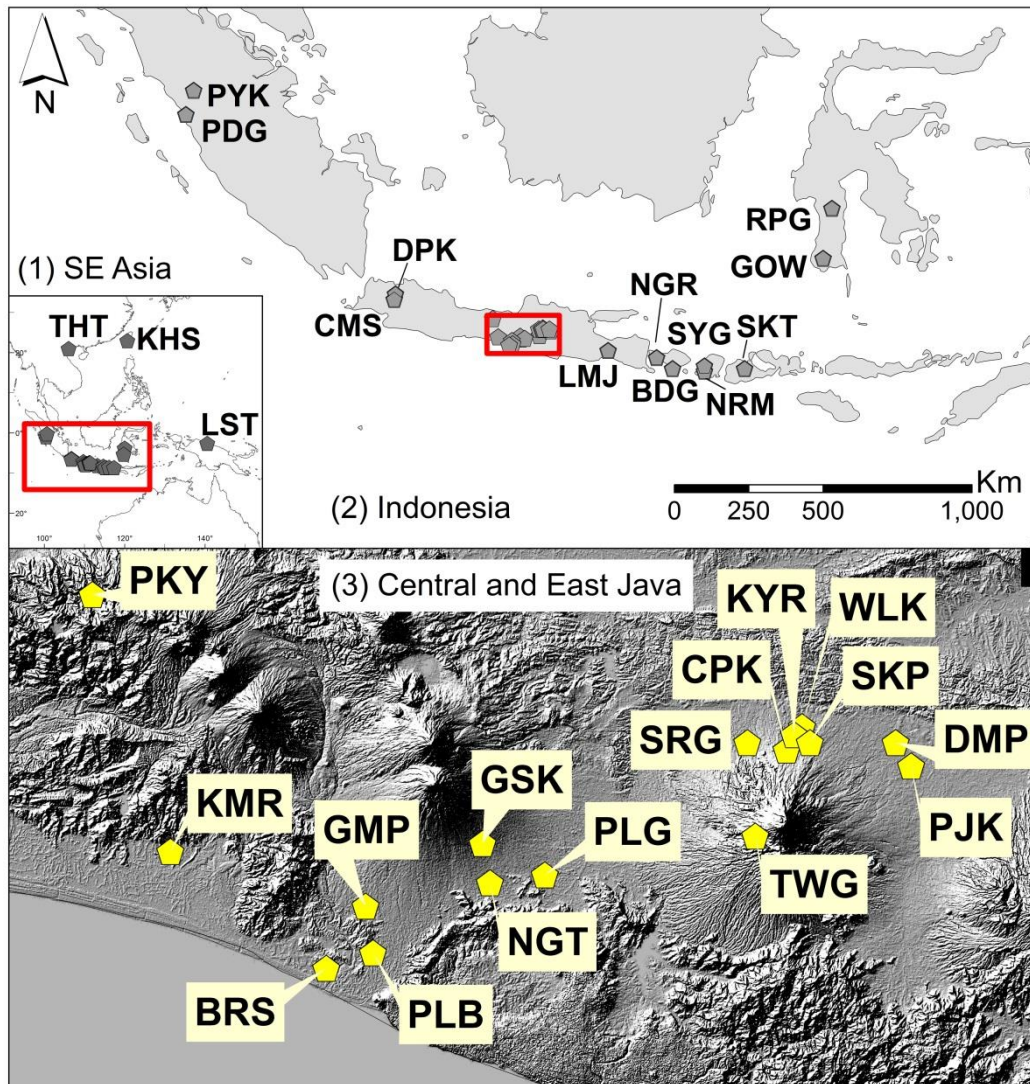


Figure 2.1. Map of collection sites for *M. albus* samples in Indonesia, Vietnam and Taiwan. The population codes are given in Table 2.1.

Table 2.2. Details of Five polymorphic loci, primer sequence, fragment size and dye

Locus	Primer Sequence	Fragment Size	Dye	Original References
<i>Mal01</i>	F-ATATTTGCAAGCGAGCCCTGT R-GTGTCTTCTCTTCAATGCAGGCACAAA	204-225	Fam	Li <i>et al.</i> (2007)
<i>Mal07</i>	F-CCCATACAGACATTTGCCACT R-GTGTCTTCCAGACAACTCTTTCCACA	184-196	Ned	Li <i>et al.</i> (2007)
<i>Mal13</i>	F-CAGCAAAGTAAAGCCGACTA R-GTTTTGTGCGCATTCTGCAAGTTT	207-226	Pet	Li <i>et al.</i> (2007)
<i>Mal007</i>	F-CATCAGGGCTAAAGAAAATGTCCA R-GTTTCCAGCCATCAGCTGAGAAAATCC	327-363	Fam	Lei <i>et al.</i> (2011)
<i>Mal008</i>	F-AGCTCACTATCCCGTCTGTCTGA R-GTTTCCTGCCCTTCTCACATTACAAAC	320-364	Vic	Lei <i>et al.</i> (2011)

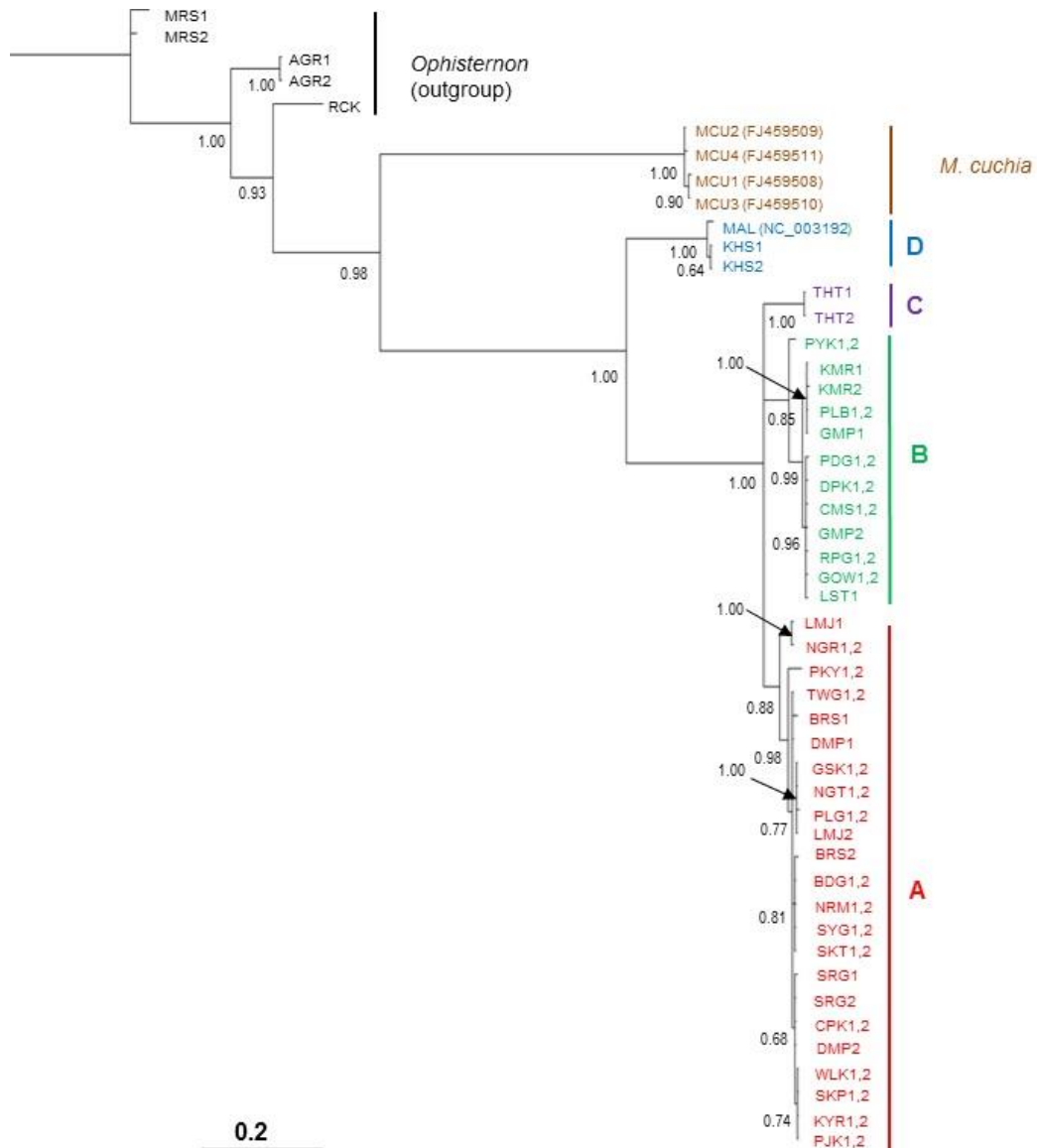


Figure 2.2. Bayesian tree inferred from COI mitochondrial gene sequences. Tree produced from 2×10^6 generations using HKY+G model. Number of each node represent posterior probabilities and scale correspond to substitution/site. MAL is a sample code for *Monopterus albus* and MUC1-MUC4 are sample codes for *M. cuchia* taken from GenBank with accession number NC_003192 and FJ459508-FJ459511 respectively for comparative purposes

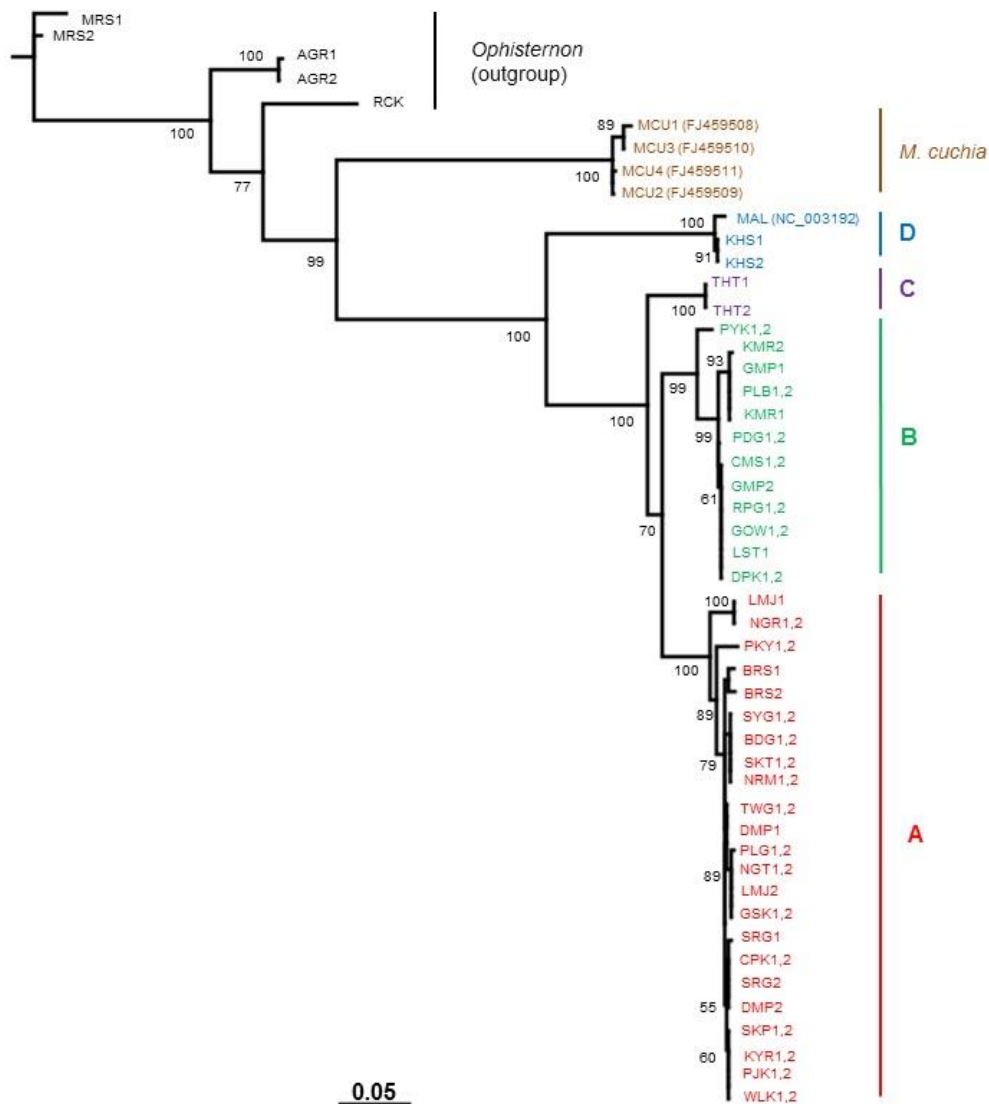


Figure 2.3. Neighbour-joining tree based on COI sequence data using Kimura-two parameter (K2P) substitution model. Number at each node represents bootstrap values and scale corresponds to substitution/site. MAL is a sample code for *Monopterus albus* and MUC1-MUC4 are sample codes for *M. cuchia* taken from GenBank with accession number NC_003192 and FJ459508-FJ459511 respectively for comparative purposes

Table 2.3. Mean percentage nucleotide sequence divergence of a 514-bp fragment of the COI mitochondrial gene among four identified clades of *M. albus* in this study

	Clade A	Clade B	Clade C	Clade D
Clade A	0.97 (0-3.60)			
Clade B	7.45 (6.13-8.78)	0.84 (0-2.79)		
Clade C	8.13 (7.89-8.56)	7.89 (7.66-8.35)	0	
Clade D	17.37 (16.77-18.61)	18.31 (17.08-18.59)	17.71 (17.53-17.79)	0.52 (0-0.78)

Table 2.4. Grouping of haplotype of *M. albus* based on the phylogenetic analysis of the COI mitochondrial gene fragment

Clade	Haplotype	Frequency	Individual code	Gen accession number
A	HA1	2	PKY1, PKY2	KP729553
	HA2	5	GSK1, GSK2,	KP729556
			NGT1, NGT2	KP729557
			LMJ2	KP729568
	HA3	2	PLG1, PLG2	KP729558
	HA4	3	TWG1, TWG2,	KP729559
			DMP1	KP729574
	HA5	1	SRG1	KP729560
	HA6	4	SRG2	KP729561
			CPK1, CPK2,	KP729572
			DMP2	KP729575
	HA7	1	BRS1	KP729562
	HA8	1	BRS2	KP729563
	HA9	3	LMJ1	KP729567
NGR1, NGR2			KP729576	
HA10	8	WLK1, WLK2	KP729569	
		SKP1, SKP2	KP729570	
		KYR1, KYR2	KP729571	
		PJK1, PJK2	KP729573	
HA11	8	BDG1, BDG2	KP729577	
		NRM1, NRM2	KP729578	
		SYG1, SYG2	KP729579	
		SKT1, SKT2	KP729580	
B	HB1	2	PDG1, PDG2	KP729549
	HB2	2	PYK1, PYK2	KP729550
	HB3	10	DPK1, DPK2	KP729551
			CMS1, CMS2	KP729552
			GMP2	KP729566
			RPG1, RPG2	KP729581
			GOW1, GOW2	KP729582
			LST1	KP729583
	HB4	4	KMR1	KP729554
			PLB1, PLB2	KP729564
GMP1			KP729565	
HB5	1	KMR2	KP729555	
C	HC1	2	THT1	KP729584
			THT2	KP729585
D	HD1	2	KHS1	KP729586
			KHS2	KP729587
	HD2	1	MAL	NC_003192

The population codes are given in Table 2.1.

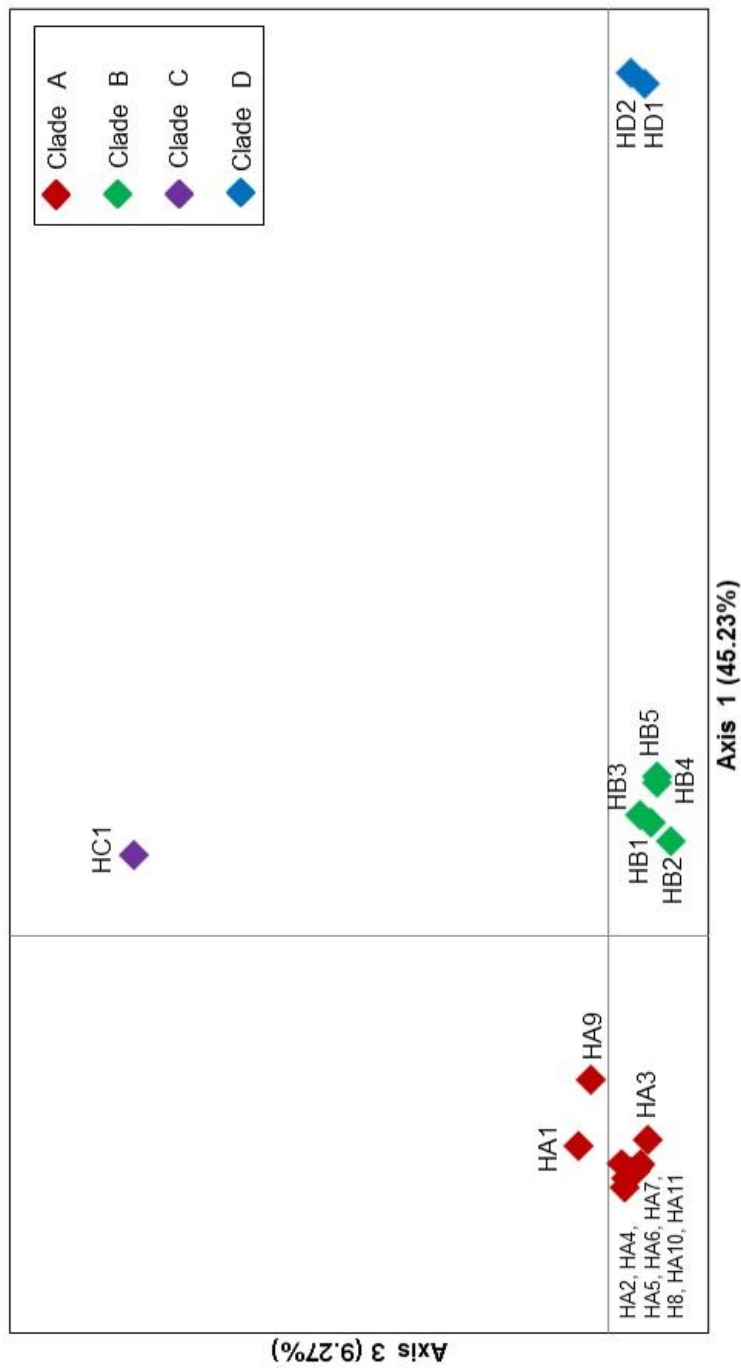


Figure 2.5. Principal Coordinates Analysis (PCoA) of pairwise genetic distances for 19 haplotypes of *M. albus* based on COI mitochondrial sequences. This analysis included MAL (HD2) taken from GenBank with accession number NC_003192 as a reference

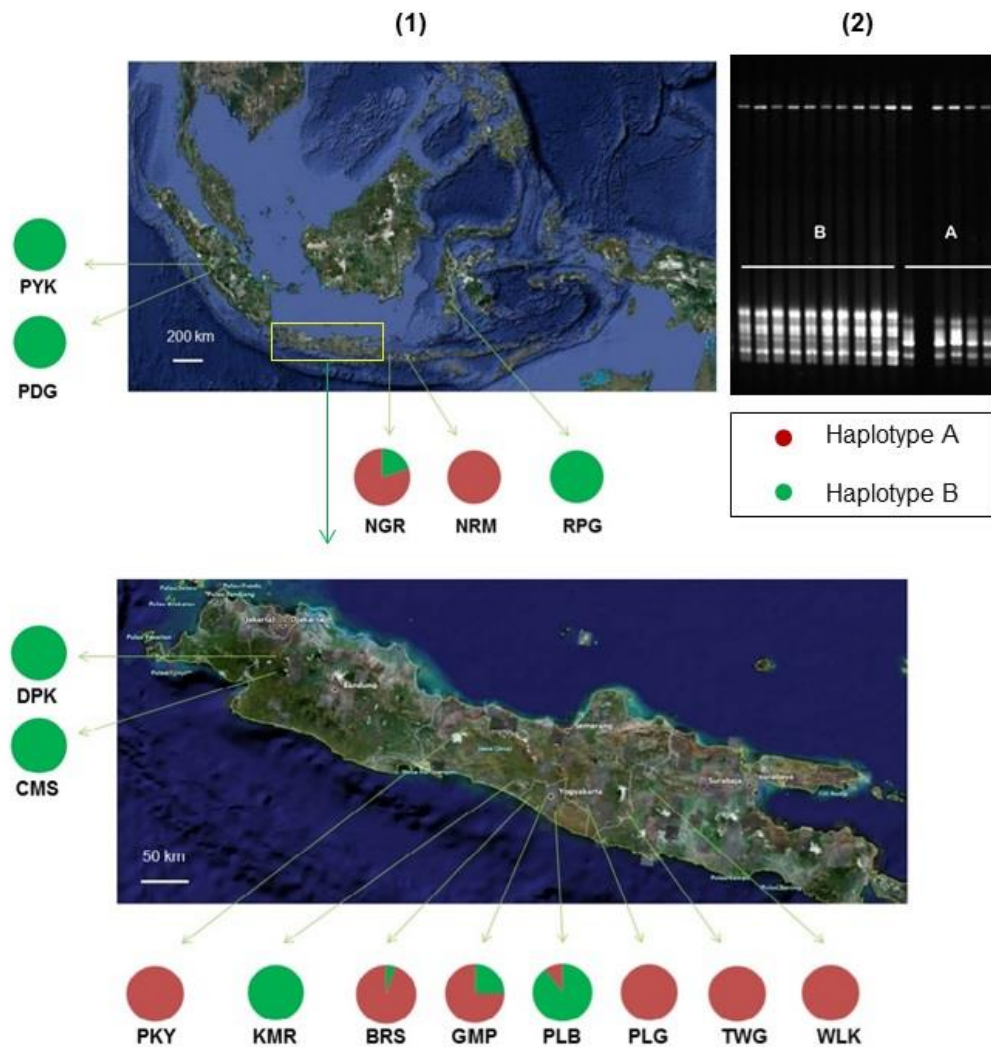


Figure 2.6. Distribution of haplotype patterns of DGGE of COI variation in 15 populations in Indonesia (1) and different haplotype pattern showed by SYBR gold staining (2). The population codes are given in Table 2.1.

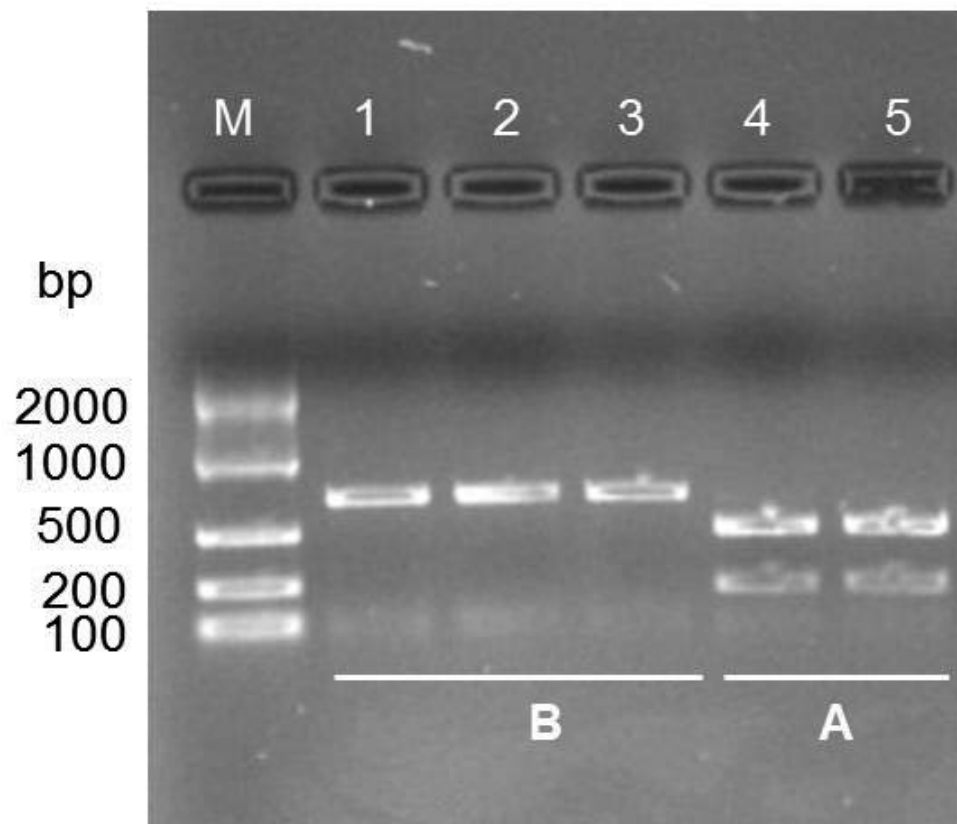


Figure 2.7. Different haplotype patterns (A and B) of *M. albus* showed by *SpeI* restriction enzyme using RFLP (M is marker and bp is base pairs)

Table 2.5. Genetic Variability of five microsatellite loci in *M. albus* haplotype A (18 populations)

Locus	Parameter	NRM	TWG	PKY	PLG	GSK	NGT	SRG	DMP	WLK	SKP	KYR	FJK	CPK	LMJ	NGR-A	PLB-A	GMP	BRS-A	Average cross population	
Mal01	N	20	28	20	20	20	20	20	20	25	20	20	20	20	20	41	20	40	21	23.06	
	A	3	3	3	2	2	2	2	2	2	2	2	2	2	3	3	5	4	4	4	2.72
	He	0.59	0.59	0.65	0.33	0.38	0.33	0.33	0.47	0.48	0.45	0.43	0.41	0.43	0.64	0.42	0.69	0.66	0.67	0.50	
	Ho	0.60	0.75	0.60	0.20	0.30	0.20	0.20	0.50	0.44	0.35	0.20	0.35	0.40	0.40	0.37	0.40	0.50	0.48	0.40	
Mal07	P_{HW}	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	
	Fs	-0.02	-0.27	0.08	0.42	0.23	0.40	0.23	-0.07	0.09	0.23	0.54	0.15	0.07	0.38	0.13	0.43	0.24	0.30	0.21	
	N	20	28	20	20	20	20	20	20	25	20	20	20	20	20	41	20	20	21	21.94	
	A	4	5	3	2	3	4	3	1	4	3	4	3	1	2	5	5	5	4	4	3.39
Mal007	He	0.72	0.71	0.64	0.48	0.52	0.61	0.68	0	0.75	0.66	0.70	0.66	0	0.51	0.60	0.76	0.70	0.73	0.58	
	Ho	0.40	0.50	0.55	0.35	0.40	0.50	0.35	0	0.56	0.40	0.50	0.50	0	0.35	0.29	0.50	0.52	0.48	0.40	
	P_{HW}	*	**	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.	**	n.s.	n.s.	*	n.s.	
	Fs	0.45	0.30	0.14	0.28	0.23	0.18	0.49	0	0.26	0.40	0.29	0.24	0	0.32	0.51	0.35	0.26	0.36	0.28	
Mal008	N	20	28	20	20	20	20	20	20	25	20	20	20	20	20	41	20	20	21	21.94	
	A	5	4	2	2	2	2	2	2	2	3	3	3	3	2	4	4	5	3	2.94	
	He	0.79	0.70	0.51	0.51	0.47	0.49	0.49	0.49	0.51	0.66	0.61	0.64	0.62	0.51	0.66	0.64	0.78	0.67	0.60	
	Ho	0.55	0.54	0.45	0.40	0.30	0.40	0.40	0.30	0.44	0.40	0.45	0.40	0.45	0.60	0.54	0.60	0.65	0.48	0.46	
Mal008	P_{HW}	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	
	Fs	0.31	0.24	0.12	0.23	0.36	0.19	0.19	0.40	0.14	0.40	0.27	0.38	0.28	-0.19	0.18	0.06	0.17	0.29	0.22	
	N	20	28	20	20	20	20	20	20	25	20	20	20	20	20	41	20	20	21	21.94	
	A	8	3	3	5	7	7	6	3	6	5	4	4	3	7	6	4	10	4	4	5.28
Mal13	He	0.81	0.67	0.65	0.77	0.83	0.82	0.84	0.63	0.81	0.73	0.68	0.73	0.66	0.85	0.83	0.76	0.86	0.75	0.76	
	Ho	0.70	0.50	0.50	0.55	0.70	0.50	0.60	0.40	0.52	0.40	0.50	0.50	0.50	0.55	0.59	0.60	0.60	0.43	0.54	
	P_{HW}	n.s.	n.s.	n.s.	n.s.	n.s.	*	*	n.s.	***	*	n.s.	n.s.	n.s.	n.s.	***	n.s.	***	**	n.s.	
	Fs	0.13	0.26	0.24	0.29	0.16	0.39	0.29	0.37	0.36	0.46	0.26	0.32	0.24	0.36	0.30	0.22	0.31	0.44	0.30	
Mean (all Loci)	N	20	28	20	20	20	20	20	20	25	20	20	20	20	20	41	20	20	21	21.94	
	A	2	3	3	3	2	2	2	2	3	2	3	3	2	3	3	2	4	3	3	2.61
	He	0.41	0.59	0.61	0.55	0.49	0.49	0.47	0.50	0.61	0.49	0.59	0.55	0.47	0.65	0.64	0.51	0.67	0.43	0.54	
	Ho	0.35	0.57	0.55	0.50	0.60	0.50	0.50	0.45	0.48	0.40	0.45	0.35	0.30	0.60	0.56	0.45	0.60	0.52	0.49	
Mean (all Loci)	P_{HW}	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	
	Fs	0.15	0.04	0.10	0.10	-0.23	-0.02	-0.07	0.10	0.22	0.19	0.25	0.37	0.36	0.08	0.13	0.12	0.10	-0.22	0.10	
	A	4.40	3.60	2.80	3.00	3.20	3.20	3.20	2.00	3.40	3.00	3.20	3.00	2.20	3.40	4.20	4.00	5.60	3.60	3.60	
	He	0.66	0.65	0.61	0.53	0.52	0.59	0.56	0.42	0.63	0.60	0.60	0.60	0.44	0.63	0.63	0.67	0.74	0.65	0.65	
Mean (all Loci)	Ho	0.52	0.57	0.53	0.40	0.44	0.44	0.41	0.33	0.49	0.39	0.42	0.42	0.33	0.50	0.47	0.51	0.58	0.48	0.48	
	P_{HW}	**	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	**	n.s.	*	n.s.	n.s.	***	*	***	***	***	
	Fs	0.20	0.11	0.13	0.26	0.18	0.20	0.26	0.16	0.21	0.33	0.32	0.29	0.19	0.19	0.21	0.21	0.19	0.23	0.23	
	A	0.20	0.11	0.13	0.26	0.18	0.20	0.26	0.16	0.21	0.33	0.32	0.29	0.19	0.19	0.21	0.21	0.19	0.23	0.23	

The population codes are given in Table 2.1.

(N=sample size; A=total number of alleles; He=expected heterozygosity; Ho=observed heterozygosity; P_{HW} =Hardy-Weinberg probability test; * $P<0.05$; ** $P<0.01$; *** $P<0.001$; Fs=fixation indices; n.s.=non-significant; n.a.=no variation)

Table 2.6. Genetic Variability of five microsatellite loci in *M. albus* haplotype B (6 populations)

Locus	Parameter	KMR	RPG	NGR-B	PLB-B	GMP-B	BRS-B	Average cross population
<i>Mal01</i>	<i>N</i>	25	22	20	31	37	38	28.83
	<i>A</i>	3	3	4	2	5	4	3.50
	<i>He</i>	0.37	0.67	0.59	0.34	0.56	0.68	0.54
	<i>Ho</i>	0.36	0.23	0.60	0.36	0.70	0.76	0.50
	<i>P_{HW}</i>	n.s.	***	n.s.	n.s.	**	***	
	<i>F_{is}</i>	0.04	0.66	-0.02	-0.05	-0.26	-0.12	0.04
<i>Mal07</i>	<i>N</i>	25	22	20	31	37	38	28.83
	<i>A</i>	2	1	3	2	3	3	2.33
	<i>He</i>	0.25	0	0.52	0.09	0.05	0.40	0.22
	<i>Ho</i>	0.12	0	0.45	0.10	0.05	0.23	0.16
	<i>P_{HW}</i>	n.s.	n.a.	n.s.	n.s.	n.s.	*	
	<i>F_{is}</i>	0.51	0	0.14	-0.04	-0.01	0.41	0.17
<i>Mal007</i>	<i>N</i>	25	22	20	31	37	38	28.83
	<i>A</i>	3	4	5	4	4	5	4.17
	<i>He</i>	0.50	0.71	0.61	0.70	0.74	0.61	0.65
	<i>Ho</i>	0.32	0.36	0.45	0.48	0.60	0.60	0.47
	<i>P_{HW}</i>	n.s.	**	n.s.	***	*	n.s.	
	<i>F_{is}</i>	0.36	0.49	0.27	0.31	0.20	0.01	0.27
<i>Mal008</i>	<i>N</i>	25	22	20	31	37	38	28.83
	<i>A</i>	4	1	3	3	4	6	3.50
	<i>He</i>	0.53	0	0.68	0.66	0.68	0.75	0.55
	<i>Ho</i>	0.36	0	0.55	0.48	0.32	0.68	0.40
	<i>P_{HW}</i>	n.s.	n.a.	n.s.	n.s.	***	n.s.	
	<i>F_{is}</i>	0.32	0	0.19	0.27	0.52	0.08	0.23
<i>Mal13</i>	<i>N</i>	25	22	20	31	37	38	28.83
	<i>A</i>	3	4	3	4	4	3	3.50
	<i>He</i>	0.56	0.71	0.63	0.65	0.64	0.61	0.63
	<i>Ho</i>	0.32	0.50	0.50	0.58	0.49	0.63	0.50
	<i>P_{HW}</i>	n.s.	*	n.s.	n.s.	n.s.	n.s.	
	<i>F_{is}</i>	0.44	0.29	0.21	0.11	0.25	-0.03	0.21
Mean (all Loci)	<i>A</i>	3.00	2.60	3.60	3.00	4.00	4.20	
	<i>He</i>	0.44	0.42	0.61	0.49	0.53	0.61	
	<i>Ho</i>	0.30	0.22	0.51	0.40	0.43	0.58	
	<i>P_{HW}</i>	**	***	n.s.	*	***	***	
	<i>F_{is}</i>	0.33	0.29	0.16	0.12	0.14	0.07	

The population codes are given in Table 2.1.

(*N*=sample size; *A*=total number of alleles; *He*=expected heterozygosity; *Ho*=observed heterozygosity; *P_{HW}*=Hardy-Weinberg probability test : **P*<0.05; ***P*<0.01; ****P*<0.001; *F_{is}*=fixation indices; n.s.=non-significant; n.a.= no variation)

Table 2.7. Genetic Variability of five microsatellite loci in *M. albus* admixture populations (4 populations)

Locus	Parameter	NGR	PLB	GMP	BRS	Average cross population
<i>Mal01</i>	<i>N</i>	61	51	77	59	62
	<i>A</i>	5	6	5	6	5.50
	<i>He</i>	0.67	0.70	0.72	0.75	0.71
	<i>Ho</i>	0.44	0.37	0.60	0.66	0.52
	<i>P_{HW}</i>	***	***	***	***	
<i>Mal07</i>	<i>F_{is}</i>	0.34	0.47	0.17	0.12	0.28
	<i>N</i>	61	51	77	59	62
	<i>A</i>	6	5	5	5	5.25
	<i>He</i>	0.76	0.58	0.66	0.67	0.67
	<i>Ho</i>	0.34	0.26	0.30	0.32	0.31
<i>Mal007</i>	<i>P_{HW}</i>	***	***	***	***	
	<i>F_{is}</i>	0.55	0.56	0.55	0.52	0.55
	<i>N</i>	61	51	77	59	62
	<i>A</i>	8	7	9	6	7.50
	<i>He</i>	0.80	0.83	0.88	0.77	0.82
<i>Mal008</i>	<i>Ho</i>	0.51	0.53	0.62	0.56	0.56
	<i>P_{HW}</i>	***	***	***	***	
	<i>F_{is}</i>	0.37	0.37	0.29	0.28	0.33
	<i>N</i>	61	51	77	59	62
	<i>A</i>	9	7	14	8	9.50
<i>Mal13</i>	<i>He</i>	0.89	0.84	0.89	0.85	0.87
	<i>Ho</i>	0.57	0.53	0.47	0.59	0.54
	<i>P_{HW}</i>	***	***	***	***	
	<i>F_{is}</i>	0.36	0.37	0.47	0.31	0.38
	<i>N</i>	61	51	77	59	62
Mean (all Loci)	<i>A</i>	6	5	6	4	5.25
	<i>He</i>	0.80	0.79	0.80	0.69	0.77
	<i>Ho</i>	0.54	0.53	0.55	0.59	0.55
	<i>P_{HW}</i>	***	***	***	**	
	<i>F_{is}</i>	0.33	0.33	0.32	0.14	0.28
Mean (all Loci)	<i>A</i>	6.80	6.00	7.80	5.80	
	<i>He</i>	0.78	0.75	0.79	0.75	
	<i>Ho</i>	0.48	0.44	0.51	0.55	
	<i>P_{HW}</i>	***	***	***	***	
Mean (all Loci)	<i>F_{is}</i>	0.39	0.42	0.36	0.27	

The population codes are given in Table 2.1.

(*N*=sample size; *A*=total number of alleles; *He*=expected heterozygosity; *Ho*=observed heterozygosity; *P_{HW}*=Hardy-Weinberg probability test : **P*<0.05; ***P*<0.01; ****P*<0.001; *F_{is}*=fixation indices; n.s.=non-significant; n.a.= no variation)

Table 2.8. Pairwise F_{ST} values between *M. albus* populations in Indonesia based on five microsatellite loci

	NRM	TWG	PKY	PLG	GSK	NGT	SRG	DMP	WLK	SKP	KYR	PJK	CPK	LMJ	NGR-A	PLB-A	GMP-A	BRS-A	NGR-B	PLB-B	GMP-B	BRS-B	KMR	
NRM																								
TWG	0.11*																							
PKY	0.18*	0.13*																						
PLG	0.17*	0.10*	0.15*																					
GSK	0.17*	0.11*	0.19*	0.03																				
NGT	0.15*	0.10*	0.18*	0.03	0.00																			
SRG	0.13*	0.07*	0.18*	0.04	0.01	0.00																		
DMP	0.22*	0.17*	0.26*	0.21*	0.17*	0.18*	0.17*																	
WLK	0.11*	0.09*	0.19*	0.09*	0.07*	0.05*	0.03	0.17*																
SKP	0.11*	0.10*	0.20*	0.11*	0.11*	0.10*	0.10*	0.07*	0.02															
KYR	0.12*	0.10*	0.20*	0.10*	0.08*	0.08*	0.06*	0.13*	0.01	0.01														
PJK	0.14*	0.12*	0.18*	0.12*	0.12*	0.11*	0.07*	0.21*	0.06*	0.04	0.03													
CPK	0.22*	0.14*	0.22*	0.14*	0.10*	0.10*	0.08*	0.18*	0.17*	0.18*	0.16*	0.15*												
LMJ	0.15*	0.09*	0.14*	0.07*	0.06*	0.05*	0.05*	0.16*	0.04*	0.08*	0.07*	0.08*	0.15*											
NGR-A	0.09*	0.13*	0.18*	0.15*	0.16*	0.14*	0.12*	0.28*	0.09*	0.10*	0.12*	0.13*	0.25*	0.14*										
PLB-A	0.10*	0.08*	0.14*	0.09*	0.07*	0.05*	0.04	0.21*	0.05*	0.09*	0.09*	0.10*	0.11*	0.07*	0.10*									
GMP-A	0.04*	0.10*	0.17*	0.15*	0.16*	0.14*	0.11*	0.23*	0.08*	0.08*	0.10*	0.10*	0.24*	0.11*	0.04*	0.09*								
BRS-A	0.11*	0.08*	0.13*	0.09*	0.13*	0.12*	0.09*	0.21*	0.09*	0.10*	0.11*	0.10*	0.20*	0.06*	0.12*	0.10*	0.08*							
NGR-B	0.36*	0.36*	0.38*	0.42*	0.42*	0.41*	0.41*	0.49*	0.37*	0.39*	0.39*	0.39*	0.47*	0.37*	0.37*	0.34*	0.29*	0.34*						
PLB-B	0.43*	0.43*	0.45*	0.49*	0.49*	0.48*	0.48*	0.54*	0.44*	0.46*	0.46*	0.46*	0.53*	0.44*	0.43*	0.41*	0.35*	0.41*	0.09*					
GMP-B	0.38*	0.39*	0.42*	0.45*	0.45*	0.44*	0.43*	0.50*	0.40*	0.41*	0.41*	0.42*	0.49*	0.40*	0.39*	0.37*	0.32*	0.37*	0.10*	0.06*				
BRS-B	0.32*	0.32*	0.35*	0.36*	0.37*	0.36*	0.35*	0.43*	0.32*	0.34*	0.34*	0.34*	0.42*	0.33*	0.32*	0.30*	0.25*	0.29*	0.07*	0.10*	0.09*			
KMR	0.45*	0.45*	0.48*	0.51*	0.52*	0.50*	0.50*	0.57*	0.47*	0.48*	0.48*	0.48*	0.56*	0.47*	0.45*	0.43*	0.37*	0.43*	0.10*	0.04	0.09*	0.07*		
RPG	0.46*	0.46*	0.49*	0.52*	0.53*	0.51*	0.51*	0.58*	0.47*	0.49*	0.49*	0.49*	0.57*	0.48*	0.46*	0.44*	0.38*	0.44*	0.27*	0.25*	0.28*	0.25*	0.34*	

Population codes are given in Table 2.1.

* $P < 0.05$ following sequential Bonferroni correction

Table 2.9. Summary of AMOVA

Source	df	SS	MS	Est. Var.	%	P-value
Among haplotype groups	1	801.477	801.477	3.169	42%	<0.001
Among Populations	22	521.641	23.711	0.832	11%	<0.001
Within Populations	564	2016.206	3.575	3.575	47%	<0.001
Total	587	3339.325		7.576	100%	

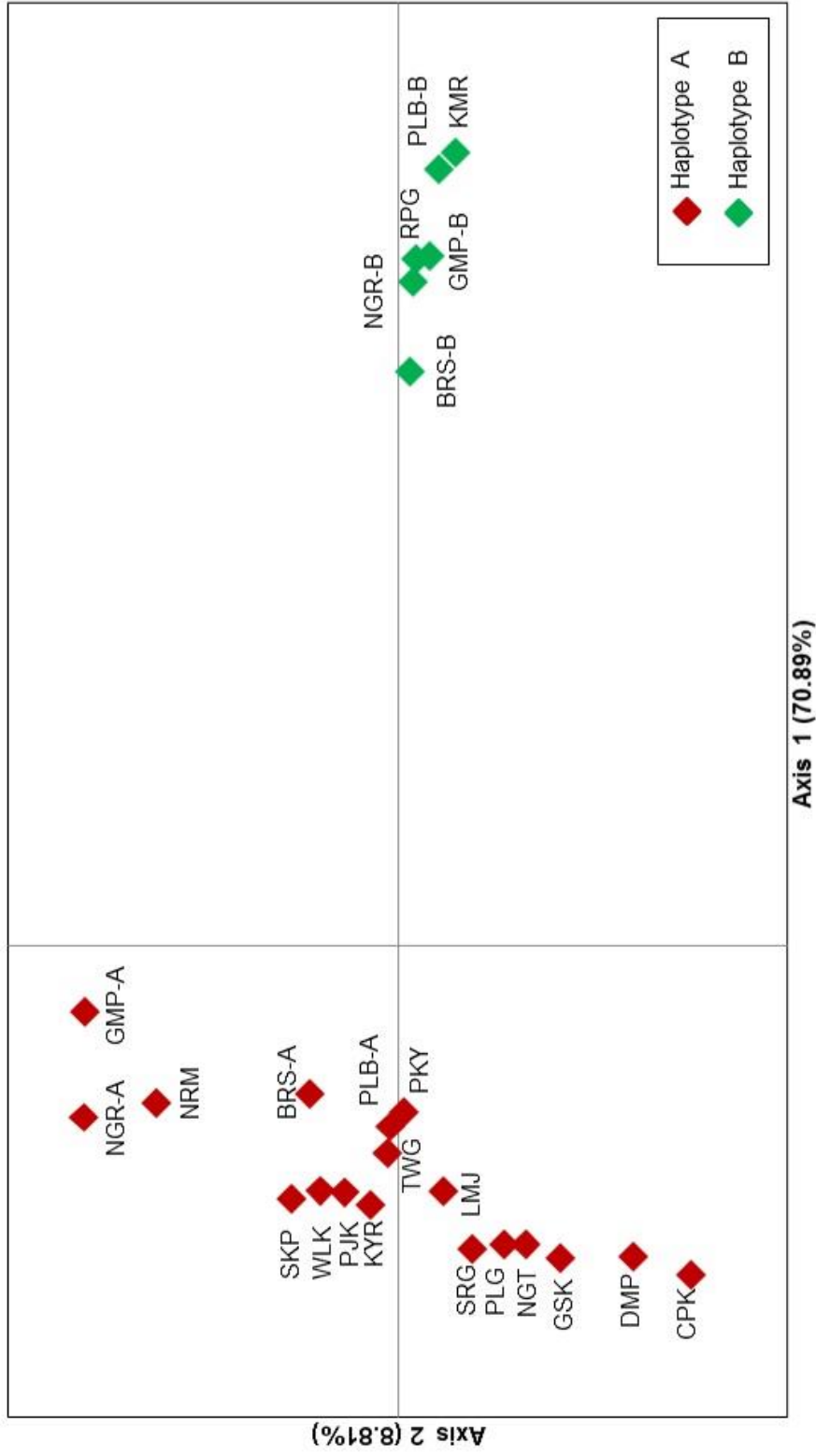


Figure 2.8. Relationships between populations defined by mitochondrial haplotype using Principal Coordinate Analysis of *M. albus* population microsatellite frequencies. Population codes are given in Table 2.1.

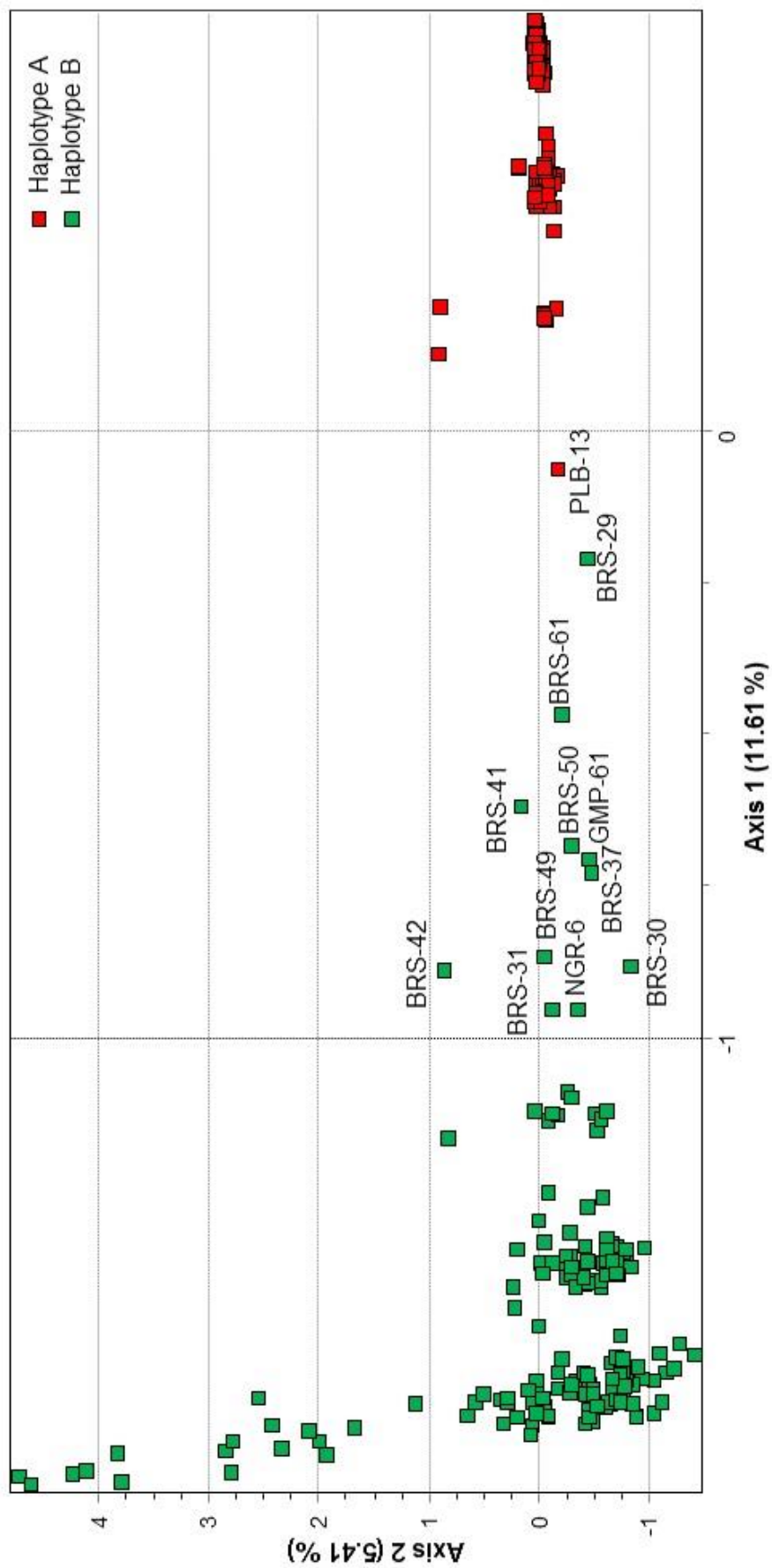


Figure 2.9. Principal Coordinate Analysis of 588 *M. albus* individuals genotyped at five microsatellite loci. Twelve individuals of potential hybrid origin are indicated by their ID sample codes

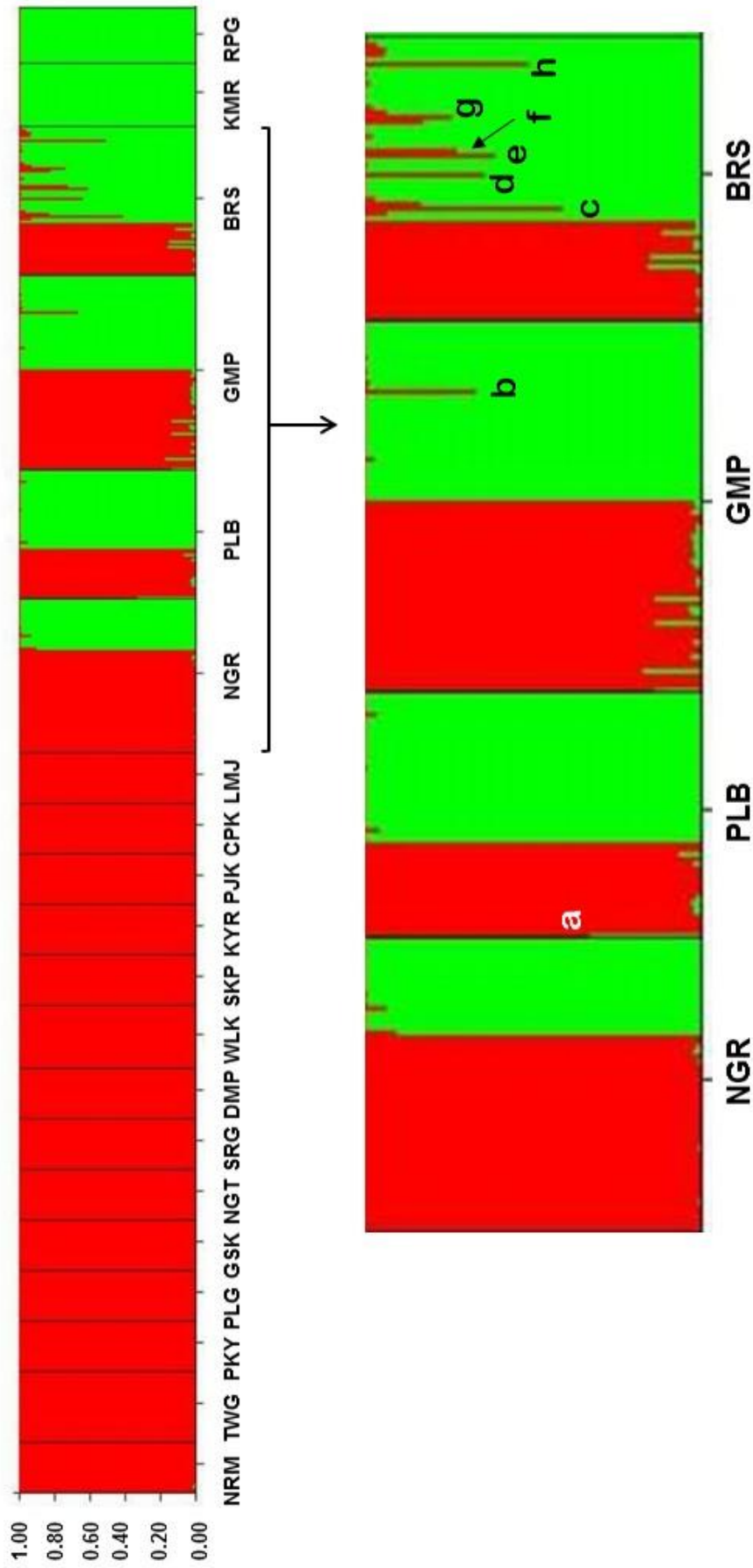


Figure 2.10. Results from the admixture analysis obtained with the program STRUCTURE under the admixture and correlated allele frequencies model. Individuals of potential hybrid origin are labelled as: a (PLB-13) – haplotype A; and, b (GMP-61); c (BRS-29); d (BRS-37); e (BRS-41); f (BRS-42); g (BRS-50); h (BRS-61) – haplotype B. Sample BRS-29 (c) was the only individual assigned to the “wrong” haplotype. Population codes are given in Table 2.1.

CHAPTER 3

Molecular taxonomy and evolution of the *Monopterus albus* complex: An integrated analysis of nucleotide data from multiple studies and gene regions

3.1. Introduction

A robust taxonomic or phylogenetic framework for significant fish groups is essential for progressing knowledge in relation to ecology, biogeography conservation and fishery management (Agapow, 2004; Schlick-Steiner *et al.*, 2010). Due to increased access to molecular data over the last twenty years, molecular based taxonomy and phylogenetic studies have had exponential growth. Therefore, many taxonomic studies now routinely use molecular tools to elucidate species boundaries, uncover cryptic species, and assess phylogenetic relationship of organisms (Fujita *et al.*, 2012; Gebiola *et al.*, 2012).

Swamp eels are economically important as one of the Asia's most popular and recognizable inland commercial fish species due to their reputation as being delicious eating and their ability to survive and grow in poorly oxygenated, shallow and muddy water (i.e. swamps including modified habitats like rice paddies) (Zhou *et al.*, 2002). They are widely distributed, being found in East India, the Philippines, East Asia and Southeast Asia, and are easily recognized by their dark brown or black cylindrical snake-like body with tapered tail, small eyes, single gill opening on the underside of the body and the absence of scales and fins (Rosen and Greenwood, 1976; Berra, 2007).

Over the last three decades, there has been much disputation regarding the taxonomic status of swamp eels commonly identified as *Monopterus albus* Zuiw 1793, with Rosen and Greenwood (1976) recognising 13 synonyms placed in six genera for this species (Table 3.1). Fishbase (<http://www.fishbase.org/search.php>) list 23 synonyms (within 11 genera) and Escmeyer (2014) considered that *M. albus* has been described under nine species names from seven different genera (Table 3.1). It is still uncertain whether this widespread fish consists of multiple species or a single variable species (Rosen and Greenwood 1976). The early studies based on morphological and anatomical characters conducted by Rosen and Greenwood (1976) provided the basis for the modern conceptualisation of the systematics of *M. albus*. These authors considered *M. albus* to be a single widespread variable species. However, in recent years there has been a number of new species of *Monopterus* identified and described, mostly from the Indian subcontinent (Bailey and Gans, 1998; Menon, 1999; Gopi, 2002; Nguyen, 2005; Britz *et al.*, 2011)

Recently, analyses of DNA sequences have been used to investigate the taxonomic status and genetic relationships among *M. albus* populations from East Asia and Southeast Asia including Indonesia and several resulting from translocation into North America (Collins *et al.*, 2002; Matsumoto *et al.*, 2010; Cai *et al.*, 2012). Collins *et al.* (2002) examined samples from Kuala Lumpur (Malaysia), Ma Cau (Vietnam), Jakarta (Indonesia), and USA (Tampa, Homestead, North Miami, and Atlanta) where the swamp eel is an alien species. Matsumoto *et al.* (2010) studied samples from several countries including Japan, Taiwan, and Indonesia (Yogyakarta). These researchers also obtained *M. albus* from several regions in China for comparative purposes. These studies are unfortunately difficult to compare as they used different fragments of the 16SrDNA mitochondrial gene. Collins *et al.* (2002) used universal primers 16Sar

and 16Sbr, whereas Matsumoto *et al.* (2010) applied the L1567 and H2196 primers, which amplify no overlapping fragments of the 16S gene.

Phylogenetic analyses conducted by Collins *et al.* (2002) revealed three distinct lineages within the *M. albus* complex. These authors found that “*M. albus*” samples from Tampa and North Miami (Florida) were very similar to samples from Nanning (China) and they suggested that they should be referred to *Monopterus albus*. In contrast their “*M. albus*” samples from Homestead (Florida) were very similar to samples from Southeast Asia including Jakarta (Indonesia) and therefore may be consistent with the taxon *Monopterus javanensis* described by Regan (1912) from this region. A third divergent “*M. albus*” lineage from Atlanta (Georgia) was considered a possible new species of unknown origin. Similarly, Matsumoto *et al.* (2010), who investigated 16S variation in *M. albus* from the north-eastern Asia and Southeast Asia also identified at least three possible cryptic species within “*M. albus*”. They also found that their samples of “*M. albus*” from Southeast Asia, which included samples from Yogyakarta (Indonesia) were genetically divergent and may include more than one species. On the basis of Collins *et al.* (2002) and Matsumoto *et al.* (2010) findings, Kottelat (2013) argued that the name of *M. albus* should be retained for the East Asian species and the oldest name for the Southeast Asian species is *M. javanensis*, a species which has also had a complex taxonomic history having been described under eight species names within five different genera (Eschmeyer, 2015). Lastly, Cai *et al.* (2012) studied *M. albus* from several regions in China using sequences from the mitochondrial control region. Their results revealed that *M. albus* fell into five genetic lineages demonstrating a fragmented population structure. However, the authors did not consider the differences detected to be of taxonomic significance. A map summarising the known sampling localities for *M. albus*, the

distribution of the main lineages discovered by molecular studies, and information on the type localities for *M. albus* and related species is provided in Figure 3.1.

The genetic studies discussed above suggest the strong possibility that multiple species may be present in what is currently referred to as *M. albus*. However, these studies are limited by the use of single mitochondrial gene fragments, the use of different gene fragments, and incomplete or inconsistent or limited sampling. Therefore more detailed molecular taxonomic studies using a greater range of molecular markers and more intensive geographic sampling are needed to fully investigate the taxonomy and evolutionary status of the *M. albus* species complex, especially the status for “*M. albus*” from poorly sampled regions in Southeast Asia focusing on Indonesia (Figure 3.1).

In this chapter I undertake a molecular genetic analysis of species boundaries of swamp eels in Indonesia with reference to other countries in Southeast Asia (Vietnam) and East Asia (Taiwan). I make use of multi-gene sequences including both mitochondrial, 16SrDNA (2 fragments) and cytochrome c oxidase subunit I (COI), and nuclear, RAG-1 and 1st intron of S7 gene regions. Wherever possible I integrate the data generated in my study with sequences available on GenBank from previous molecular studies of *M. albus* discussed above.

3.2. Materials and Methods

3.2.1. Sample collection and storage

A total of 31 populations of swamp eels (*M. albus*) were sampled from several regions in Indonesia, Thuan Thanh (North Vietnam) and Kaohsiung (Taiwan) as detailed in Table 3.2. Samples of *Ophisternon* species (*Ophisternon*

bengalense from Maros, South Sulawesi, *Ophisternon gutturale* from Rapid Creek, Darwin, Australia and *Ophisternon* sp. from Angurugu River, Groote Eylandt, Australia) were used as outgroup samples. One individual from every population was sequenced except for Kaohsiung (Taiwan) for which n=2. Specimens were obtained from farmers and fishermen, and were placed on ice and brought to the laboratory and frozen. Subsequently muscle tissue was dissected from each partially thawed fish and placed into 1.5 ml screw top cryogenic vials and preserved in 95% ethanol until required.

3.2.2. DNA extraction, PCR amplification and sequencing

Four DNA regions were selected for sequencing: the mitochondrial 16S gene (16SrDNA), the mitochondrial cytochrome c oxidase subunit I (COI), the non-coding first intron of S7 nuclear gene (1st intron of S7), and the coding nuclear RAG gene (RAG1). Total genomic DNA was extracted from the ethanol preserved muscle tissue using the DNeasy blood and tissue kit (QIAGEN, Valencia, CA, USA) following manufacture's protocols. The KAPA2G Robust PCR kit (Kapa Biosystems) was used for the polymerase chain reaction (PCR). The total volume of each PCR reaction was 25 μ L. The reaction mixture consisted 10-100 ng of genomic DNA, 0.2 mM of each dNTP, 2mM MgCl₂, 0.014U Tag Polymerase, 0.6 μ M of each primer and 1x PCR reaction buffer. The primers used for the amplification of the mitochondrial and nuclear gene regions are given in Table 3.3. For the RAG1 nuclear gene, the following internal primer RAG3F (5'-GGGTGATGTCAGYGAGAAGCA-3') was used to complete the sequences. To evaluate the reliability of the DNA amplification, a negative control was set up by omitting template DNA from the reaction mixture. The reaction mixture was initially pre-denatured at 94°C for 2 min followed by 35 cycles of

denaturation at 94°C for 50 s, annealing at 49°C (for 16SrDNA and COI mitochondrial genes, and 1st intron of S7 nuclear gene) or 60°C (for the RAG1 nuclear gene) for 2 min, and extension at 72°C for 1.5 min. Reactions were then subjected to a final extension at 72°C for 6 minutes. All PCR products were visualized using 1% agarose gel electrophoresis buffered with Tris-Acetate-EDTA (TAE), stained with gel red nucleic acid stain (Biotium), and visualized under UV light. PCR products were then purified using Viogene PCR purification kit (Viogene Inc.). All amplicons were sequenced and analysed in both forward and reverse directions using the Big Dye Terminator Ver. 3.3. protocol (Applied Biosystems) and the ABI 3130xL Genetic Analyzer (Applied Biosystems). All chromatograms were checked and assembled using SeqMan, and edited using EditSeq Pro Program Lasergene DNASTAR software package (DNASTAR Inc., Madison, USA). Consensus sequences of the coding genes (COI mtDNA and RAG1 nuclear genes) were translated into amino acids to check for stop codons and obvious sequence errors. Sequences obtained in this study were then submitted in GenBank under the accession numbers that are provided in Table 3.2.

3.2.3. Sequence alignment and phylogenetic analysis

In addition to sequences obtained in this study swamp eel sequences for the same gene fragments were retrieved from GenBank (www.ncbi.nlm.nih.gov) (Miya *et al.*, 2001; Collins *et al.*, 2002; and Matsumoto *et al.*, 2010). Details of sampling locations and codes are provided in Table 3.2. All sequence data were converted to fasta format, including the outgroup sequences from *Ophisternon bengalense* (MRS), *Ophisternon gutturale* (RCK) and *Ophisternon* sp. (AGR1 and AGR2), and aligned using the opal (a multiple sequence alignment program)

routine implemented by the MESQUITE 2.74 package (Maddison and Maddison, 2010) and ClustalW in MEGA5 (Tamura *et al.*, 2011). The final alignments were checked manually. End regions where the alignment was ambiguous were excluded from the analyses.

Three data sets were constructed for analysis. The first was a dataset containing sequences for the partial 16SrDNA mitochondrial gene using universal primer 16Sar and 16Sbr for this study and combined with the sequences obtained study by Collins *et al.* (2002) using the same primers (GenBank accession numbers: AF512841-AF512856). The second dataset contained sequences for the partial 16SrDNA mitochondrial gene region using primer L1567 and H2196 for this study and combined with the sequences obtained by Matsumoto *et al.* (2010) using the same primers (GenBank accession number AB494967-AB494995). Both data sets included the corresponding sequences from the *M. albus* complete mitogenome sequenced by Miya *et al.* (2001) (GenBank accession number NC_003192). A third dataset included all individuals from this study that were sequenced for all the mtDNA and nDNA gene regions (Table 3.2).

Phylogenetic relationships were inferred for each dataset separately. The Aike Information Criterion implemented in jModelTest ver.0.1.1 (Posada and Buckley, 2004; Posada, 2008) was used to determine the best fit evolutionary model for each dataset. Bayesian inference (BI) was performed with MrBayes v. 3.2 (Ronquist *et al.*, 2012) under the best-fit model. Two simultaneous Markov chain analyses (MCMC) were run for 2×10^6 generation for a first and a third datasets, and 4×10^6 generation for a second dataset, to estimate the posterior probabilities distribution. Topologies were sampled every 1000 generations. The analysis used a relative burn-in of 25% for diagnostics. Consensus trees were visualised in FigTree 1.4.0 (Rambaut, 2012).

Kimura-2-parameter distances were calculated within and among the main lineages recovered for the partial 16S mtDNA genes and multiple gene regions using MEGA5 (Tamura *et al.*, 2011). From the sequences, variable and parsimony informative (PI) sites were determined for the first and the second data sets of the partial 16S mtDNA. Standard diversity indices for the partial 16S mtDNA included haplotype (h) and nucleotide diversity (π) and neutrality (Tajima, 1989; Fu, 1997) and were obtained using DnaSP 5.10.01 (Librado and Rozas, 2009).

Distance based analyses were conducted to complement the phylogenetic analyses using principles coordinate analysis (PCoA) (Peakall and Smouse, 2012). The matrix of pairwise distance was analysed using the program GenAlEx version 6.5 under the PCoA option.

3.3. Results

3.3.1. Molecular genetic variation and phylogenetic analysis of two 16SrDNA mitochondrial gene fragments

Forty nine and sixty two 16S mtDNA sequences were assembled for the first and second datasets respectively. For the first data set 17 sequences were obtained from GenBank and 32 were sequenced by this study. For the second data set 30 sequences were obtained from GenBank and 32 were sequenced by this study. The first dataset consisted of a total of 25 haplotypes (Table 3.5) whereas the second dataset had a total of 39 haplotypes (Table 3.7). For the 510 bp sequences in the first dataset, 74 sites were variable, of which 63 were parsimony informative, 47 of which differed by transitional substitutions, 13 by transversional changes and there were eight insertions/deletions. Six mutational

sites exhibited multiple substitutions (Table 3.8 and Figure 3.3). The haplotype diversity (h) and nucleotide diversity (π) in the first dataset were 0.96 and 0.034 respectively (Table 3.8). For the second dataset (sequences = 539 bp), there were 144 variable nucleotide sites with 56 parsimony informative, 45 of which differed by transitional substitutions, 17 by transversional changes and six multiple substitutions (Table 3.8 and Figure 3.6). Deletions/insertions were high in the second dataset compared to the first dataset. In addition, transition mutation occurred more frequently than transversion mutation in both 16S datasets. The haplotype diversity (h) and nucleotide diversity (π) in the second dataset were 0.978 and 0.046 respectively. No significantly positive or negative values of Tajima's D or Fu's F were observed in either of the two datasets (Table 3.8), supporting selective neutrality of the sequences in both datasets.

The GTR model with gamma-distributed rate variation across sites as inferred by the jModelTest 0.1.1 under the Akaike information criterion (AIC) was found to be the best-fit nucleotide substitution model for the first and the second datasets. The length of aligned sequences of the partial 16S mtDNA gene of the first dataset and the second dataset after excluding the ambiguous positions was 510 bp and 539 bp respectively. The Bayesian analyses from the first and second dataset together indicate that five lineages can be identified within *M. albus* that are geographically correlated. These are designated as clade A, B, C, D, E (Figure 3.2. and Figure 3.5) and are coded as follows: blue – Clade D (China, Japan and Taiwan), purple – Clade C (North Vietnam and southern China), brown - Clade E (southern China and Taiwan), green – Clade B (South Vietnam, Taiwan, Malaysia and wide spread in Indonesia), red – Clade A (Indonesia, confined to the adjacent islands of Java, Bali and Lombok). The Bayesian analysis of the first data set identifies four distinct lineages which form well

supported monophyletic groups that correspond to Operational Taxonomic Units (OTUs). The most basal lineage (Clade D) consists of the *M. albus* sequences from the whole mitogenome, samples from Taiwan and Collins *et al.*'s samples from Atlanta (USA) with a Bayesian posterior probability (pp) of 1.00. Clade D's sister clade, also supported by a pp of 1.00, has three distinct clades consisting of Clade E (7 samples from southern China and Florida, USA) with a pp of 0.97, Clade C (2 samples from North Vietnam and southern China) with a pp of 0.97 and Clade B (17 samples from South Vietnam, Malaysia and Indonesia) with pp of 1.00. The 19 samples from central Indonesia (Clade A) show limited divergence and fall into basal positions in the second main lineage and do not form a coherent monophyletic group. The Bayesian analysis of the second 16S data set supports some of the phylogenetic relationships resolved in the first dataset and some new and unresolved relationships. The tree (Figure 3.5) supports Clades D, C, B, congruent with first 16S data set, but generally with lower support with pp = 0.62, 0.95, 0.72 respectively. In contrast to the first analysis this second 16S gene region provides strong support for a monophyletic clade A from central Indonesia (pp= 1.00), but does not support a monophyletic clade E with samples from this clade from southern China and nearby Taiwan paraphyletic with respect to Clade A.

Levels of within-clade and between-clade (OTUs) divergences estimated using the Kimura-2-parameter (K2P) genetic distances are summarised in Tables 3.4 and Table 3.6. The average nucleotide sequence divergence in the first dataset within clades A, B, C, D and E are 0.66%, 0.26%, 0.60%, 0.00% and 0.40% respectively. The average percentage sequence divergence in the second dataset within clades A, B, C, D and E is 0.97%, 0.43%, 1.33%, 1.06%, and 1.23% respectively. Levels of between-clade divergences in the *M. albus* OTUs for the first dataset are shown in Table 3.4 and levels the second dataset are

shown in Table 3.6. The highest genetic divergence between OTUs in the first dataset is between clades B and D (Table 3.4) which ranged from 9.80% to 10.50% (mean 10.22%), whereas the least genetic divergence is between clades A and E which ranged from 1.70% to 2.50% (mean 2.29%). Levels of differentiation within and between clade A, B, C, and E is summarised by Principal Coordinate Analysis (PCoA) in Figure 3.4. This analysis demonstrates significant clear differentiation between clades, including the members of clade A, which were not resolved as a monophyletic group in the corresponding Bayesian analysis (Figure 3.2). Thus this analysis also indicates that the degree of differentiation between clades was conspicuously greater than that within clades.

In the second dataset, the highest divergence is between clades A and D (Table 3.6), ranging from 7.20%-9.90% (mean 8.68%). The least genetic divergence in the second dataset is between clades B and C ranging from 1.80% to 3.40% (mean 2.57%). Levels of differentiation within and between *M. albus* clades summarised by Principal Coordinate Analysis (PCoA) in Figure 3.7 supported the differentiation between clades A, B and C, however the members of clade E showed only modest differentiation from clade A consistent with the corresponding Bayesian tree (Figure 3.5).

For the first 16S mtDNA dataset the highest haplotype number is within clade B, with nine haplotypes from 17 individuals, followed by clade A with eight haplotypes from 19 individuals and clade E with four haplotypes from seven individuals. Clades C and D both have two haplotypes from two and four individuals respectively (Table 3.5). The haplotype diversities (h) within clades ranged from 0.67 to 1.00 and the nucleotide diversities (π) from 0.0 to 0.0099. The most genetically diverse clade is clade C, with the highest value of haplotype and nucleotide diversities ($h=1.00$, $\pi=0.0099$), whereas the least variable clade is

clade D ($h=0.67$, $\pi=0$). A summary of genetic diversity estimates within and between clades for the first 16S data set is summarised in Table 3.8.

In the second 16S dataset, clade D has the most haplotypes (15 from 20 individuals) whereas the lowest number of haplotypes is in clades C and E, which have three haplotypes from three and five individuals respectively. Clade B has six haplotypes from 12 samples and clade A has 12 haplotypes from 22 individuals (Table 3.7). The haplotype diversities (h) within clades ranged from 0.70 to 1.00 and the nucleotide diversities (π) from 0.0043 to 0.015. The most genetically diverse clade was clade C, demonstrating the highest value of haplotype and nucleotide diversities ($h=1.00$, $\pi=0.015$). The least haplotype diversity was observed in clade E ($\pi=0.070$) whereas the least of nucleotide diversity was found in clade B ($\pi=0.0043$). A summary of genetic diversity within and between clades for the second 16S data is given in Table 3.8.

3.3.2. Molecular genetic variation and phylogenetic analysis of multiple gene regions

The multigene data set was assembled from 32 individuals obtained to represent as wide a geographic representation of the *M. albus* species complex as possible together with four outgroup samples. The data set consisted of data from the two fragments of 16S mtDNA (502 bp and 627 bp) presented separately above, COI mtDNA (526bp), RAG1 nDNA (1,384bp), and 1st intron of S7 nDNA (327 bp) giving a concatenated data set of 3,366 bp. A total of 31 haplotypes were identified from the 32 individuals (Table 3.10). The dataset comprises 384 variable sites of which 295 are parsimony-informative. The haplotype diversity (h) and nucleotide diversity (π) of the multigene data set are 0.998 and 0.0237 respectively (Table 3.11).

The best-fit model of nucleotide substitution for the multigene data set was identified as GTR+I+G as inferred by the jModelTest 0.1.1 under the Akaike information criterion (AIC). The Bayesian analysis revealed four distinct genetic lineages corresponding to clades or OTUs A, B, C, and D as previously identified. Each clade with multiple samples was recovered with a Bayesian posterior probability of 1.00 (Figure 3.8).

Average genetic distance between *M. albus* lineages is 2.42%. The highest genetic distance between samples within clade B is 0.68% within clade A 0.45% and within clade D is 0.20%. The highest divergence is between clades B and D, ranging from 7.09%-7.50% (mean 7.29%). The least genetic divergence is between clades B and C ranging from 2.46% to 2.67% (mean 2.59%). Average within and between clade divergence levels are summarised in Table 3.9. In addition, average sequence divergence revealed high levels between three clades (clade A, clade B, clade C) to clade D (>7.21%). Levels of differentiation within and between the more closely related *M. albus* clades A, B, and C was also summarised by Principal Coordinate Analysis (PCoA) in Figure. 3.9. This analysis indicated that the degree of differentiation between clade A, B, and C was much greater than that within clades.

The haplotype diversity (h) within clades ranged from 0.978 to 1.00 and the nucleotide diversity (π) from 0.0018 to 0.0070. The most genetically diverse clade was clade A with the highest value of haplotype and nucleotide diversities ($h=1.00$ and $\pi=0.0070$). The least haplotype diversity was detected in clade B ($h=0.978$) whereas the least nucleotide diversity was observed in clade D ($\pi=0.0018$). A summary of genetic diversity within and between clades for multigene dataset is given in Table 3.11.

For each gene region investigated for the 32 individuals the one with the most variable sites was the COI mt-DNA gene region (118 variable sites with 106 parsimony informative sites from 526 bp) and the least variable sites was observed in 1st intron of S7 nDNA gene region (41 variable sites with 26 parsimony informative sites from 327 bp). The highest haplotype number was observed within RAG1 nDNA region (24 haplotypes) and the least haplotype number was detected in 16S mt-DNA region using primer 16Sar and 16Sbr (14 haplotypes). The highest haplotype diversity (h) and nucleotide diversity (π) was found in RAG1 nDNA region ($h=0.962$) and COI mt-DNA gene region ($\pi=0.055$) respectively. The least haplotype diversity (h) and nucleotide diversity (π) was detected in 16S mt-DNA region using primer using primer H1567 and L2196 ($h=0.927$) and RAG1 nDNA region ($\pi=0.0064$). A summary of genetic diversity of each gene region is given in Table 3.11.

3.4. Discussion

The increasing application of molecular tools to study systematic and evolutionary questions at the species and population levels is showing that, taxonomies based on morphological characters are frequently inadequate if not misleading (Perdices *et al.*, 2005; Agorreta *et al.*, 2013). This is especially true for freshwater fish species including the synbranchid fish to which *M. albus* belongs, which is especially problematic and lacking resolution due to a lack of reliable external diagnostic characters (Rosen and Greenwood, 1976). The results of this study, using a robust approach that takes advantages of multiple gene regions to address questions about the systematics and evolution of the *M. albus* species complex in Southeast Asia and eastern Asia adds to the growing list of

taxonomically and morphologically challenging groups of freshwater fish from this region.

My phylogenetic results (both from 16S mtDNA and multiple gene region datasets) agree with and extend the finding published by Collins *et al.* (2002) and Matsumoto *et al.* (2010) in supporting that *M. albus* is not a single species, but consists of multiple species in what is appropriately referred to as a species complex. Integrated phylogenetic analysis of 16S mtDNA from this study and previous studies (Miya *et al.*, 2001; Collins *et al.*, 2002; Matsumoto *et al.*, 2010) revealed that for the populations of swamp eels so far sampled, five distinct genetic lineages are apparent that have diverged from each other to varying degrees. The most divergent lineage (D) is from the most northerly portion of the geographic range of the samples studied and consists of samples from Japan, China and Taiwan. Therefore, the four clades (clade A, clade B, clade C, and clade E), which are all significantly divergent from clade D are supported as separate taxonomic units by at least one or more datasets or analyses. Thus these four putative taxa must have names found for them among the taxa places previously as synonyms of *M. albus* or be described as new species.

The analyses using 16S mtDNA for the first and second datasets revealed two genetically distinct groups of *M. albus* from Indonesia. The two genetic forms of Indonesian *Monopterus* (clade A and clade B) were also distinct from other forms of the *M. albus* complex from North Asia. In addition, the integrated analysis using 16S mtDNA from the second dataset supports Matsumoto's finding regarding Yogyakarta populations (Figure 3.5). Based on limited sampling Matsumoto *et al.* (2010) suggested that at least two cryptic species may occur in Yogyakarta (YGI). My research based on much more comprehensive sampling showed that Yogyakarta populations do comprise two distinct groups (YGI1 grouped together with PLB and GMP in clade B and YGI2, YGI3, YGI4 grouped

together with BRS in Clade A). The integrated studies using 16S mtDNA from the first dataset also revealed that Jakarta (JKI) samples collected by Collins *et al.* (2002) grouped together with samples from Depok (DPK) and Ciomas, Bogor (CMS) collected in this study and included in clade B (Figure 3.2). The Depok location is in the middle of West Java between Jakarta and Ciomas (Bogor), and it is about 22 km from Jakarta and Bogor. These conclusions are fully supported by the results of the COI barcoding and microsatellite analyses presented in chapter 2.

The phylogenetic analysis of multiple gene regions provided further support for the recognition of additional species of *Monopterus*. Even though this part of the study did not include samples from clade E, the phylogenetic analysis demonstrates strong divergence among the four well supported lineages consisting of clade A, clade B, clade C and clade D. The results also showed that the level of divergence (>7.21%) across all gene regions between clades A, B, C and clade D representing northern samples was especially high.

I consider it appropriate to refer to the most divergent lineage (clade D) from North Asia as *M. albus* sensu stricto. This is consistent with Rosen and Greenwood (1976) who suggested that type locality of swamp eel is possibly from Asiatic Russia and appears to be the centre for intensive sampling (Figure 3.10). This is also consistent with the usage of the name by other authors (Miya *et al.*, 2001; Matsumoto *et al.*, 2010) and the recent review by Kottelat (2013).

Finding samples from North Vietnam collected from Thuan Thanh, Back Ninh to also to be a divergent and well supported clade (clade C) is consistent with the recognition of *Monopterus dienbienensis* by Nguyen (2005) from North Vietnam. The type locality of *M. dienbienensis* is located at Thanh Luong, Dien Bien (North Vietnam) about 30 km from Thuan Thanh, Back Ninh (North Vietnam). These samples grouped together with samples from Hainan Island

(China) based on Matsumoto's *et al.* findings and Buobai (China) from Collins' *et al.* investigation indicating that this species has a wider distribution beyond North Vietnam. A second species from Vietnam, *M. bicolor* (Nguyen, 2005), has also been described, however no samples were available from this region so the status of this species could not be investigated.

Finding two distinct clades of swamp eel in Indonesia, which have been confirmed to be behaving as reproductively isolated species based in the finding of chapter 2, means that species described as *Monopterus javanensis* Lacépède is likely to be valid. According to Rosen and Greenwood (1976) who examined specimens from Java (with no exact locality) considered *M. javanensis* to be a synonym of *M. albus*. Eschmeyer (2014) considered the specimen of *M. javanensis* were most likely from Sunda Strait located in West Java. On the basis of this study the swamp eel comprising clade B is distributed mostly in West Indonesia including West Java and East Indonesia. Therefore, the species named *M. javanensis* is mostly likely the appropriate name for swamp eel grouped as clade B. This is consistent with Kottelat (2013) who also reported that the name *M. albus* should be retained for the East Asian species and the oldest name for the Southeast Asian species is *M. javanensis*. However the finding of additional species of *Monopterus* in this study from Java, makes Kottelat's assessment less straightforward.

Kottelat's assessments are also complicated by the discovery of an additional lineage, Clade E should which also constitutes a valid species from North Asia, but with information lacking on the extent of its distribution makes the application of a name more problematic. In the study of Matsumoto *et al.* (2010), the swamp eel samples from Fuzhou, China (FZC) were more closely related to samples from Puli, Taiwan (PLT) and Hengchuen, Taiwan (HCT) than to the samples from Haikou, Hainan Island, China (HNC). Collins *et al.* (2002) also

supported the separation between samples from Nanning, China (NGC) which is grouped in clade E and samples from Buobai, China (BUC) which are included in clade C in my integrated studies. These data therefore all indicate that samples from Fuzhou (China), Puli (Taiwan), Hengchuen (Taiwan) and Nanning (China) should be considered as a distinct species. However it is possible that this putative species represented by clade E may correspond to other previously described species including *M. albus*, *M. bicolor* or *M. dienbienensis* and it may be that one or more of the other clades may need to be described as new species. Thus a more expanded and integrated morphological and molecular genetic study is needed that references known type material, where it exists.

Finding two genetic distinct groups of Indonesian swamp eel, with one most likely corresponding to *M. javanensis* (clade B) means that the lineage (clade A) needs to be described as a new species as no other named species of *Monopterus* have been described from this region. This new species mostly likely represents a relatively new speciation event as it is a sister species to *M. javanensis* (clade B) and has a more restricted distribution to the islands of Java, Bali and Lombok. Confusing the interpretation of the biogeographic patterns is that the samples of clade B extend further westward being found in Sulawesi and Papua. As this distribution extends well eastward of the Wallace line it is highly likely that this species has been translocated to New Guinea by humans due to its popularity as a food item (Nico *et al.*, 2011). It is also possible that this has happened to a lesser extent to the species corresponding to clade A as this is also found on either side of the Wallace line. The occurrence of both forms of Indonesian swamp eel either side of Wallace line is a very unusual pattern for an obligate freshwater species.

There is also a possibility that the species corresponding to clade B (*M. javanensis*) may be not native to Indonesia but introduced from countries to the

north. However, conversely, Indonesian swamp eels placed in clade B could be native to Indonesia and introduce to other countries through fish trading and marketing. To resolve the likely native distributions of these forms and the impacts of translocations further studies need to be conducted on eels captured from the wild and, where possible natural water courses. A limitation to this study was that all samples from Asian countries were obtained from farmers, fishermen or markets.

The wider implication of the finding of a cluster of cryptic species within the *M. albus* species complex is that more broad scale studies of molecular genetic variation within the genus *Monopterus* are justified, and these should be extended more broadly to the family Synbranchidae. This is supported by the findings of cryptic species in other genera (*Synbranchus* and *Ophisternon*) within this family by Perdices (2005) and Valdez-Moreno *et al.* (2009) using morphological and molecular genetic approaches. They also recommended that a more detailed taxonomic revision of *Synbranchus* and *Ophisternon* was justified given the discrepancies between their molecular data and the existing taxonomy.

In conclusion this comprehensive study of the *M. albus* species complex using multiple gene regions and multiple analyses that are integrated with earlier studies where possible revealed the presence of at least five cryptic species. Thus the systematics of this species is much more complex than previously thought. From a geographic perspective, it is apparent that Indonesia has two cryptic species (clade A and B), China has three cryptic species (clade C, clade D and clade E), Taiwan has three cryptic species (clade B, clade D, and clade E), Vietnam has two cryptic species (clade B and clade C), and Japan and Malaysia each have one cryptic species corresponding to clade D and clade B respectively (Table 3.5, Table 3.7, and Fig. 3.10). Clade A, native to Indonesia, and clade E, native to China, are most likely new species. It is recommended that specimens

matching clade B should be referred to as *M. javanensis* Lacépède, those matching clade C should be considered as *M. dienbienensis* Nguyen and those corresponding to clade D are *M. albus* Zuiw sensu stricto, until a more thorough taxonomic study can be undertaken. The taxonomic complexity of *M. albus* in Asia is mirrored in other recent molecular studies of other Asian freshwater fish species including *Tor* (Nguyen et al., 2008), *Pangio* (Bohlen et al., 2011), air sac catfish (Ratmuangkhwang et al., 2014), and *Channa* (Barman et al., 2014; Serrao et al., 2014).

Table 3.1. Synonyms of *M. albus* based on Rosen and Greenwood (1976), Fishbase, and Eschmeyer (2015)

Rosen and Greenwood (1976)	Fishbase (http://www.fishbase.org/search.php)	Eschmeyer (2015)
<i>Muraena alba</i> Zuiew 1793	<i>Muraena alba</i> Zuiew 1793	<i>Muraena alba</i> Zuiew 1793
<i>Monopterus javanensis</i> Lacepède 1800	<i>Fluta alba</i> Zuiew 1793	<i>Monopterus marmoratus</i> Richardson 1846
<i>Monopterus javanicus</i> Shaw 1803	<i>Monopterus alba</i> Zuiew 1793	<i>Pneumabranchnus cinereus</i> McClelland 1843
<i>Monopterus xanthognathus</i> Richardson 1845	<i>Monopterus javanensis</i> Lacepède 1800	<i>Symbranchus grammicus</i> Cantor 1842
<i>Monopterus cinereus</i> Richardson 1845	<i>Monopterus laevis</i> Lacepède 1803	<i>Unibranchapertura laevis</i> Lacepède 1803
<i>Monopterus helvolus</i> Richardson 1846	<i>Monopterus cinereus</i> McClelland 1844	<i>Cryptophthalmus robustus</i> Franz 1910
<i>Monopterus marmoratus</i> Richardson 1846	<i>Monopterus xanthognathus</i> Richardson 1845	<i>Apterigia immaculata</i> Basilewski 1855
<i>Ophiocardia phayriana</i> McClelland 1845	<i>Monopterus helvolus</i> Richardson 1846	<i>Apterigia nigromaculata</i> Basilewski 1855
<i>Unibranchapertura laevis</i> Lacepède 1803	<i>Monopterus marmoratus</i> Richardson 1846	<i>Apterigia saccogularis</i> Basilewsky 1855
<i>Symbranchus grammicus</i> Cantor 1842	<i>Symbranchus grammicus</i> Cantor 1842	
<i>Apterigia saccogularis</i> Basilewski 1855	<i>Symbranchus grammicus</i> Cantor 1842	
<i>Apterigia nigromaculata</i> Basilewski 1855	<i>Symbranchus xanthognathus</i> Richardson 1845	
<i>Apterigia immaculata</i> Basilewski 1855	<i>Apterigia immaculata</i> Basilewski 1855	
	<i>Apterigia nigromaculata</i> Basilewski 1855	
	<i>Apterigia immaculata</i> Basilewski 1855	
	<i>Apterigia nigromaculata</i> Basilewski 1855	
	<i>Apterigia saccogularis</i> Basilewski 1855	
	<i>Ophiocardia xanthognatha</i> Basilewski 1855	
	<i>Ophiocardia phayriana</i> McClelland 1844	
	<i>Ophiocardia phayriana</i> McClelland 1844	
	<i>Gymnotus albus</i> Zuiew 1789	
	<i>Unibranchapertura laevis</i> Lacepède 1803	
	<i>Pneumabranchnus cinereus</i> McClelland 1844	

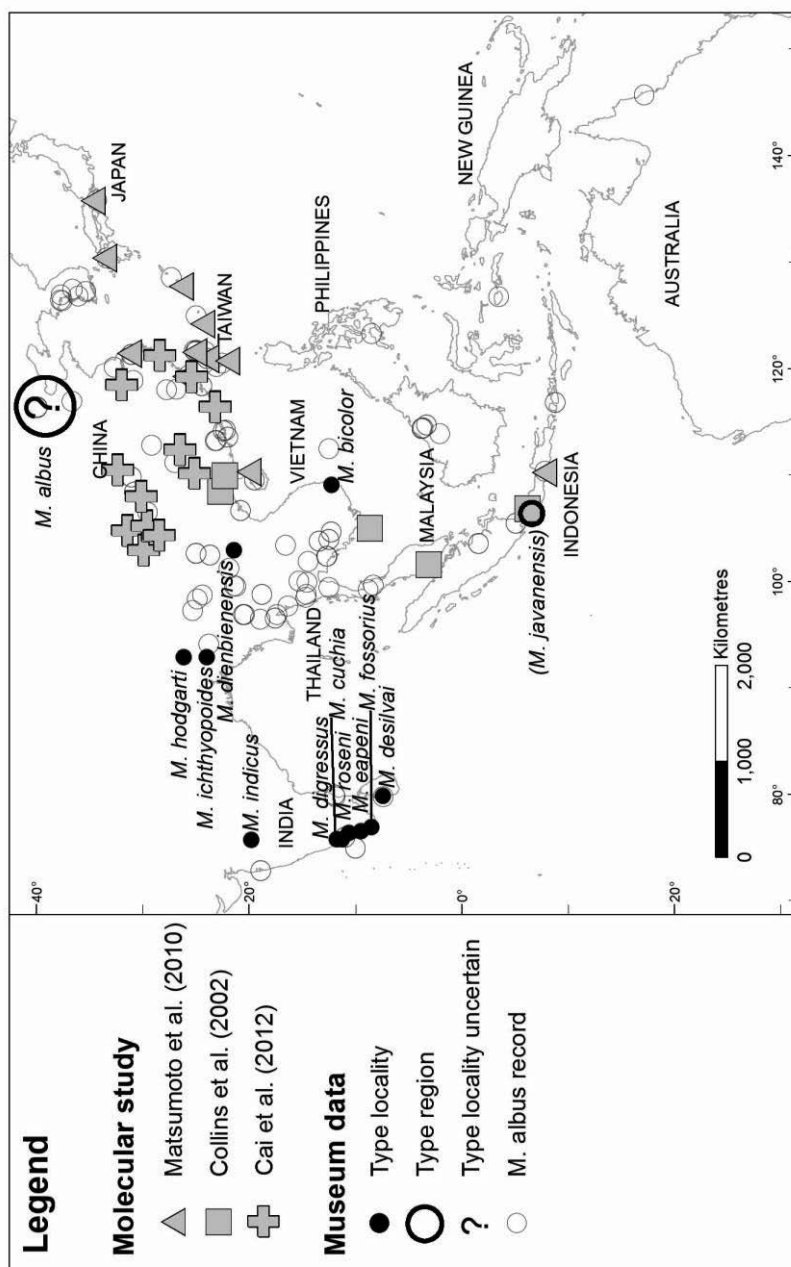


Figure 3.1. Map distribution of *M. albus* species complex based on museum records (www.lifemapper.org, accessed 15 March 2013) and previous studies (Collins *et al.*, 2002; Matsumoto *et al.*, 2010 and Cai *et al.* 2012). The large circle with a question mark for *M. albus* indicates uncertainty over the exact type locality

Table 3.2. Samples used in this study and the previous studies with sampling location, population code, and GenBank accession number

Location	Code	Country	GenBank Accession Number				References	
			16S*	16S**	COI	RAG1		1 st of S7
Indarung-Lubuk Kilangan, Padang, West Sumatra	PDG	Indonesia	KP729483	KP729516	KP729549	KP738612	KP738645	This study
Payakumbuh Selatan, Payakumbuh, West Sumatra	PYK	Indonesia	KP729484	KP729517	KP729550	KP738613	KP738646	This study
Depok-Pancoran Mas, Depok, West Java	DPK	Indonesia	KP729485	KP729518	KP729551	KP738614	KP738647	This study
Ciomas, Bogor, West Java	CMS	Indonesia	KP729486	KP729519	KP729552	KP738615	KP738648	This study
Soko Kembang-Kayupuring, Petungkriyono, Pekalongan, Central Java	PKY	Indonesia	KP729493	KP729526	KP729553	KP738622	KP738655	This study
Tawangmangu, Karanganyar, Central Java	TWG	Indonesia	KP729497	KP729530	KP729559	KP738626	KP738659	This study
Ngantenan-Canan, Wedi, Klaten, Central Java	NGT	Indonesia	KP729495	KP729528	KP729557	KP738624	KP738657	This study
Gesikan, Gantiwarno, Klaten, Central Java	GSK	Indonesia	KP729494	KP729527	KP729556	KP738623	KP738656	This study
Prayan-Planggu, Trucuk, Klaten, Central Java	PLG	Indonesia	KP729496	KP729529	KP729558	KP738625	KP738658	This study
Kemiri Kidul, Kemiri, Purwokerto, Central Java	KMR	Indonesia	KP729487	KP729520	KP729554	KP738616	KP738649	This study
Tunggul, Gondang, Sragen, Central Java	SRG	Indonesia	KP729498	KP729531	KP729560	KP738627	KP738660	This study
Gejawang Kilen, Balecatur, Gamping, Sleman, Yogyakarta	GMP	Indonesia	KP729489	KP729522	KP729565	KP738618	KP738651	This study

* primers: 16Sar and 16Sbr

** primers: L1567 and H2196

Table 3.2. Continued

Location	Code	Country	GenBank Accession Number				References	
			16S*	16S**	COI	RAG1		1 st of S7
Palbapang, Bantul, Yogyakarta	PLB	Indonesia	KP729488	KP729521	KP729564	KP738617	KP738650	This study
Brosot, Galur, Kulon Progo, Yogyakarta	BRS	Indonesia	KP729499	KP729532	KP729562	KP738628	KP738661	This study
Citrodiwangsan, Lumajang, East Java	LMJ	Indonesia	KP729500	KP729533	KP729567	KP738629	KP738662	This study
Pojok, Kwadungan, Ngawi, East Java	PJK	Indonesia	KP729505	KP729538	KP729573	KP738634	KP738667	This study
Cepoko, Ngrambe, Ngawi, East Java	CPK	Indonesia	KP729504	KP729537	KP729572	KP738633	KP738666	This study
Walikukun, Widodaren, East Java	WLK	Indonesia	KP729501	KP729534	KP729569	KP738630	KP738663	This study
Kayutrejo, Widodaren, Ngawi, East Java	KYR	Indonesia	KP729503	KP729536	KP729571	KP738632	KP738665	This study
Sekarputih, Widodaren, Ngawi, East Java	SKP	Indonesia	KP729502	KP729535	KP729570	KP738631	KP738664	This study
Dempel, Geneng, Ngawi, East Java	DMP	Indonesia	KP729506	KP729539	KP729574	KP738635	KP738668	This study
Pohsanten, Mendaya, Negara, Bali	NGR	Indonesia	KP729507	KP729540	KP729576	KP738636	KP738669	This study
Kerobokan, Kuta Utara, Badung, Bali	BDG	Indonesia	KP729508	KP729541	KP729577	KP738637	KP738670	This study
Narmada, Lombok Barat, West Nusa Tenggara (NTB)	NRM	Indonesia	KP729509	KP729542	KP729578	KP738638	KP738671	This study

* primers: 16Sar and 16Sbr

** primers: L1567 and H2196

Table 3.2. Continued

Location	Code	Country	GenBank Accession Number				References	
			16S*	16S**	COI	RAG1		1 st of S7
Sayang-Sayang, Cakranegara, Mataram, NTB	SYG	Indonesia	KP729510	KP729543	KP729579	KP738639	KP738672	This study
Seketeng, Sumbawa, NTB	SKT	Indonesia	KP729511	KP729544	KP729580	KP738640	KP738673	This study
Rappang, Panca Rijang, Sindendreng Rappang, South Sulawesi	RPG	Indonesia	KP729490	KP729523	KP729581	KP738619	KP738652	This study
Gowa, South Sulawesi	GOW	Indonesia	KP729491	KP729524	KP729582	KP738620	KP738653	This study
Lake Sentani, Sentani, Jayapura, Papua	LST	Indonesia	KP729492	KP729525	KP729583	KP738621	KP738654	This study
Thuan Thanh District, Back Ninh Province, North Vietnam	THT	Vietnam	KP729512	KP729545	KP729584	KP738641	KP738674	This study
Kaohsiung, Taiwan	KHS1 KHS2	Taiwan	KP729513 KP729514	KP729546 KP729547	KP729586 KP729587	KP738642 KP738643	KP738675 KP738676	This study This study
Maros, South Sulawesi ¹	MRS	Indonesia	KP729515	KP729548	KP729588	KP738644	KP738677	This study
Rapid Creek, Darwin, NT, Australia ²	RCK	Australia	Appendix 4 ^a	Appendix 5 ^a	Appendix 6 ^a	Appendix 7 ^a	Appendix 8 ^a	This study
Angurugu River, Groote Eyland, NT, Australia ³	AGR1 AGR2	Australia	Appendix 4 ^a Appendix 4 ^a	Appendix 5 ^a Appendix 5 ^a	Appendix 6 ^a Appendix 6 ^a	Appendix 7 ^a Appendix 7 ^a	Appendix 8 ^a Appendix 8 ^a	This study This study

* primers: 16Sar and 16Sbr

** primers: L1567 and H2196

1) *Ophisternon bengalense*2) *Ophisternon gutturale*3) *Ophisternon sp*

a) "awaiting processing by GenBank"

Table 3.2. Continued

Location	Code	Country	GenBank Accession Number		References
			16S*	1 st of S7	
Tampa, Florida	TPF1	USA	AF512841	RAG1	Collins <i>et al.</i> (2002)
Tampa, Florida	TPF2	USA	AF512843		Collins <i>et al.</i> (2002)
North Miami, Florida	NMF	USA	AF512844		Collins <i>et al.</i> (2002)
Homestead, Florida	HMF1	USA	AF512849		Collins <i>et al.</i> (2002)
Homestead, Florida	HMF2	USA	AF512852		Collins <i>et al.</i> (2002)
Atlanta, Georgia	ALG	USA	AF512856		Collins <i>et al.</i> (2002)
Nanning	NGC1	China	AF512842		Collins <i>et al.</i> (2002)
Nanning	NGC2	China	AF512845		Collins <i>et al.</i> (2002)
Nanning	NGC3	China	AF512846		Collins <i>et al.</i> (2002)
Nanning	NGC4	China	AF512847		Collins <i>et al.</i> (2002)
Buobai	BUC	China	AF512848		Collins <i>et al.</i> (2002)
Jakarta	JKI	Indonesia	AF512850		Collins <i>et al.</i> (2002)
Ca Mau	VTM	Vietnam	AF512851		Collins <i>et al.</i> (2002)
Kuala Lumpur	KLM1	Malaysia	AF512853		Collins <i>et al.</i> (2002)
Kuala Lumpur	KLM2	Malaysia	AF512854		Collins <i>et al.</i> (2002)
Kuala Lumpur	KLM3	Malaysia	AF512855		Collins <i>et al.</i> (2002)
Shanghai	SHC1	China		AB494967	Matsumoto <i>et al.</i> (2010)
Shanghai	SHC2	China		AB494969	Matsumoto <i>et al.</i> (2010)
Shanghai	SHC3	China		AB494970	Matsumoto <i>et al.</i> (2010)
Shanghai	SHC4	China		AB494974	Matsumoto <i>et al.</i> (2010)
Shanghai	SHC5	China		AB494975	Matsumoto <i>et al.</i> (2010)
Shanghai	SHC6	China		AB494976	Matsumoto <i>et al.</i> (2010)
Haikou, Hainan Island	HNC1	China		AB494984	Matsumoto <i>et al.</i> (2010)
Haikou, Hainan Island	HNC2	China		AB494985	Matsumoto <i>et al.</i> (2010)

* primers: 16Sar and 16Sbr

** primers: L1567 and H2196

Table 3.2. Continued

Location	Code	Country	GenBank Accession Number			References
			16S*	16S**	COI	
Fuzhou	FZC1	China		AB494992		Matsumoto <i>et al.</i> (2010)
Fuzhou	FZC2	China		AB494993		Matsumoto <i>et al.</i> (2010)
Fuzhou	FZC3	China		AB494995		Matsumoto <i>et al.</i> (2010)
Oki Tawn, Fukuoka,	KSJ1	Japan		AB494968		Matsumoto <i>et al.</i> (2010)
Kyushu						
Oki Tawn, Fukuoka,	KSJ2	Japan		AB494971		Matsumoto <i>et al.</i> (2010)
Kyushu						
Sakurai City, Nara, Honshu	HSJ1	Japan		AB494972		Matsumoto <i>et al.</i> (2010)
Sakurai City, Nara, Honshu	HSJ2	Japan		AB494977		Matsumoto <i>et al.</i> (2010)
Sakurai City, Nara, Honshu	HSJ3	Japan		AB494978		Matsumoto <i>et al.</i> (2010)
Okinawa, Ryukyu Island	OKJ1	Japan		AB494979		Matsumoto <i>et al.</i> (2010)
Okinawa, Ryukyu Island	OKJ2	Japan		AB494981		Matsumoto <i>et al.</i> (2010)
Okinawa, Ryukyu Island	OKJ3	Japan		AB494982		Matsumoto <i>et al.</i> (2010)
Ishigaki, Ryukyu Island	ISJ1	Japan		AB494980		Matsumoto <i>et al.</i> (2010)
Ishigaki, Ryukyu Island	ISJ2	Japan		AB494983		Matsumoto <i>et al.</i> (2010)
Ishigaki, Ryukyu Island	ISJ3	Japan		AB494986		Matsumoto <i>et al.</i> (2010)
Puli	PLT1	Taiwan		AB494991		Matsumoto <i>et al.</i> (2010)
Puli	PLT2	Taiwan		AB494994		Matsumoto <i>et al.</i> (2010)
Taipei	TPT	Taiwan		AB494973		Matsumoto <i>et al.</i> (2010)
Hengchuen	HGT	Taiwan		AB494994		Matsumoto <i>et al.</i> (2010)
Yogyakarta	YGI1	Indonesia		AB494987		Matsumoto <i>et al.</i> (2010)
Yogyakarta	YGI2	Indonesia		AB494988		Matsumoto <i>et al.</i> (2010)
Yogyakarta	YGI3	Indonesia		AB494989		Matsumoto <i>et al.</i> (2010)
Yogyakarta	YGI4	Indonesia		AB494990		Matsumoto <i>et al.</i> (2010)
No location data	MAL	No data	NC_003192	NC_003192	NC_003192	Miya <i>et al.</i> (2001)

*) primers: 16Sar and 16Sbr

**) primers: L1567 and H2196

Table 3.3. Primers used for mitochondrial and nuclear genes for *M. albus* in this study

Gene	Primer	Primer sequence (5' – 3')	References
16S mitochondrial gene (16SrDNA)	16Sar	F: CGCCTGTTTATCAAAAACAT	Palumbi <i>et al.</i> (1996)
	16Sbr	R: CCGGTCTGAACTCAGATCACGT	
16S mitochondrial gene (16SrDNA)	L1567	F: AAGGGGAGGCAA CTCGTA	Matsumoto <i>et al.</i> (2010) Yamaguchi <i>et al.</i> (2000)
	H2196	R: GTCTGAGCTTTAACGCTTTCT	
Cytochrome c oxidase subunit I (COI)	FishF2	F: TCGACTAATCATAAAGATATCGGCAC	Ward <i>et al.</i> (2005)
	FishR2	R: ACTTCAGGGTGACCCGAAGAATCAGAA	
Nuclear RAG gene (RAG1)	RAG1F	F: AGCTGTAGTCAGTAYCACAAARATG	Quenouille <i>et al.</i> (2004)
	RAG9R	R: GTGTAGAGCCAGTGRTGYTT	
First intron of S7 nuclear gene (1 st intron of S7)	S7RPEX1F	F: TGGCCTCTTCCTTGGCCGTC	Chow and Hazama (1998)
	S7RPEX2R	R: AACTCGTCTGGCTTTTCGCC	

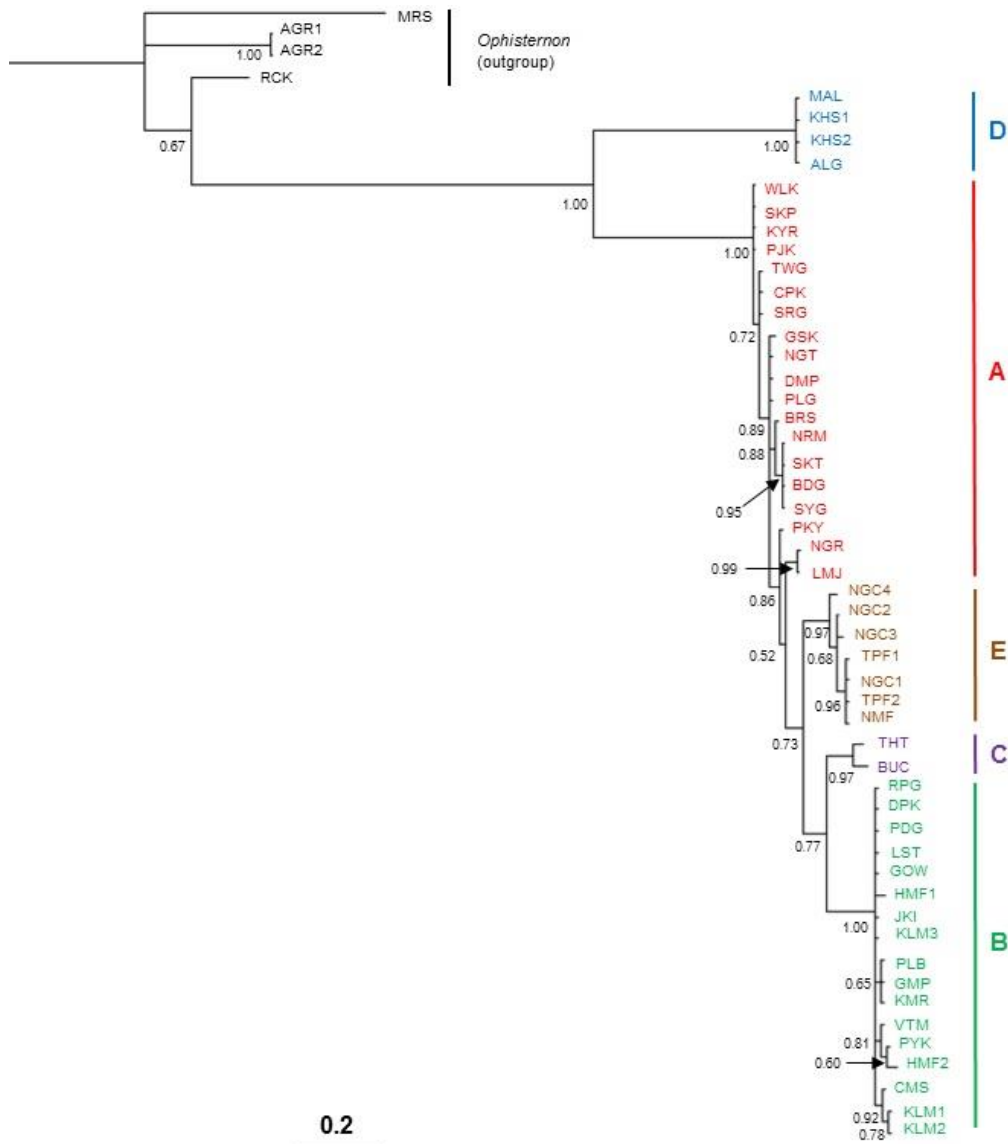


Figure 3.2. Bayesian tree inferred from 16S mitochondrial gene sequences combining between this study with Collins *et al.* (2002) and Miya *et al.* (2001) data. Tree produced from 2×10^6 generations using GTR+G model. Number of each node represent posterior probabilities and scale correspond to substitutions/site

Table 3.4. Mean percentage nucleotide sequence divergence of a 510-bp fragment of the 16S mitochondrial gene among five identified clade of *M. albus* from this study combined with Collins *et al.* (2002) and Miya *et al.* (2001) data

	Clade A	Clade B	Clade C	Clade D	Clade E
Clade A	0.66 (0-1.70)				
Clade B	3.38 (2.3-4.3)	0.26 (0-0.8)			
Clade C	2.59 (2.1-3.2)	2.79 (2.5-3.2)	0.6		
Clade D	10.00 (9.8-10.3)	10.22 (9.8-10.5)	9.55 (9.3-9.8)	0	
Clade E	2.29 (1.7-2.5)	3.52 (3-4.1)	2.69 (2.1-3)	9.91 (9.6-10.1)	0.4 (0-0.8)

Table 3.5. Haplotype of *M. albus* based on 16S mitochondrial gene region inferred from this study combined with Collins *et al.* (2002) and Miya *et al.* (2001) data

Clade	Haplotype	Sample number	Individual code	Gen accession number	Country
A	HA1	1	PKY	KP729493	Indonesia
	HA2	1	GSK	KP729494	Indonesia
	HA3	3	NGT	KP729495	Indonesia
			PLG	KP729496	Indonesia
			DMP	KP729506	Indonesia
	HA4	3	TWG	KP729497	Indonesia
			SRG	KP729498	Indonesia
			CPK	KP729504	Indonesia
	HA5	1	BRS	KP729499	Indonesia
	HA6	2	LMJ	KP729500	Indonesia
			NGR	KP729507	Indonesia
	HA7	4	WLK	KP729501	Indonesia
			SKP	KP729502	Indonesia
			KYR	KP729503	Indonesia
	HA8	4	PJK	KP729505	Indonesia
			BDG	KP729508	Indonesia
NRM			KP729509	Indonesia	
SYG			KP729510	Indonesia	
B	HB1	6	PDG	KP729483	Indonesia
			DPK	KP729485	Indonesia
			RPG	KP729490	Indonesia
			GOW	KP729491	Indonesia
			LST	KP729492	Indonesia
			JKI	AF512850	Indonesia
	HB2	1	PYK	KP729484	Indonesia
	HB3	1	CMS	KP729486	Indonesia
	HB4	3	KMR	KP729487	Indonesia
			PLB	KP729488	Indonesia
			GMP	KP729489	Indonesia
	HB5	1	HMF1	AF512849	USA
	HB6	1	HMF2	AF512852	USA
	HB7	1	VTM	AF512851	Vietnam
	HB8	2	KLM1	AF512853	Malaysia
			KLM2	AF512854	Malaysia
HB9	1	KLM3	AF512855	Malaysia	
C	HC1	1	THT	KP729512	Vietnam
	HC2	1	BUC	AF512848	China
D	HD1	2	KHS1	KP729513	Taiwan
			KHS2	KP729514	Taiwan
	HD2	2	MAL	NC_003192	No data
			ALG	AF512856	USA

Table 3.5. Continued

Clade	Haplotype	Sample number	Individual code	Gen accession number	Country
E	HE1	4	NGC1,	AF512842	China
			TPF1	AF512841	USA
			TPF2	AF512843	USA
			NMF	AF512844	USA
	HE2	1	NGC2	AF512845	China
	HE3	1	NGC3	AF512846	China
	HE4	1	NGC4	AF512847	China

Individual codes based on population codes are given in Table 3.2.

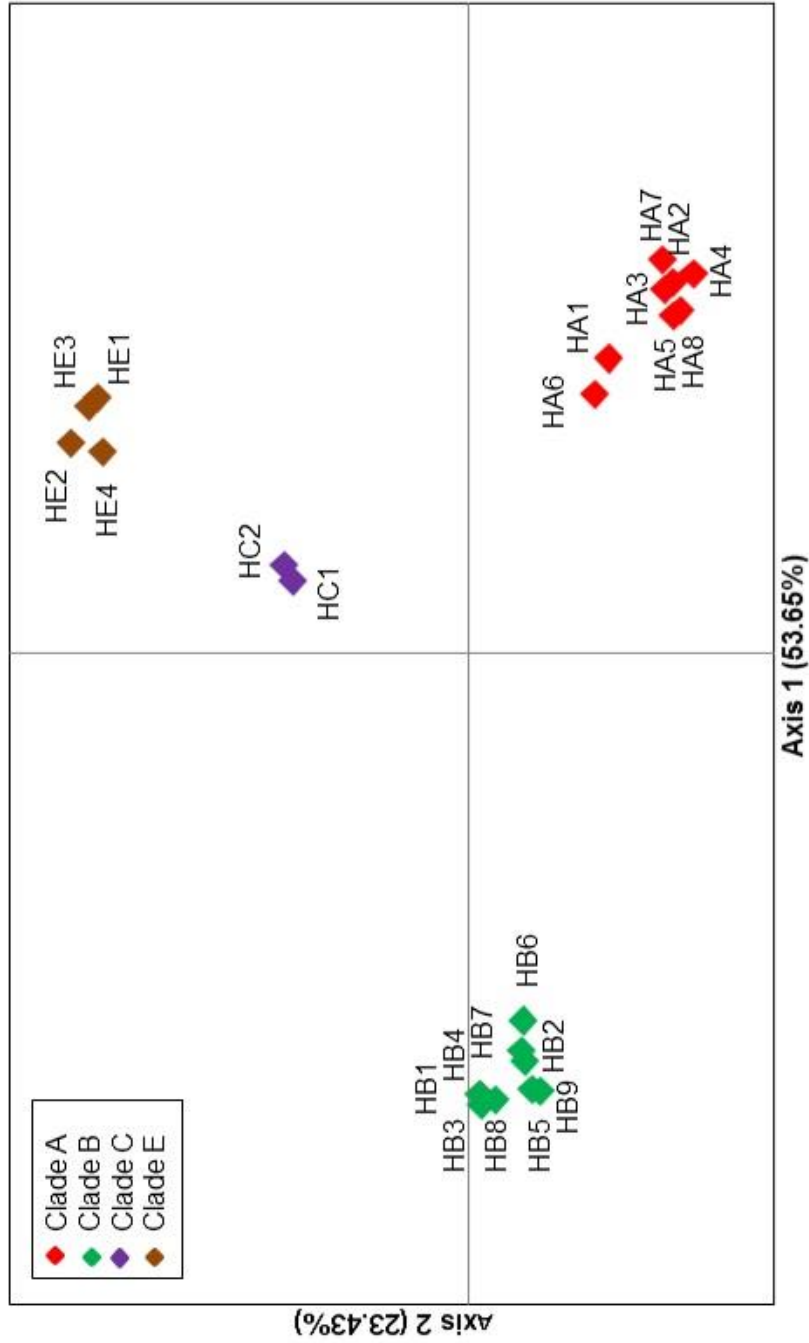


Figure 3.4. Principal Coordinates Analysis (PCoA) of pairwise genetic distances of 23 haplotypes of *M. albus* populations based on 16S mitochondrial sequences. This analysis incorporates data from Collins *et al.* (2002) taken from GenBank given in Table 3.2.

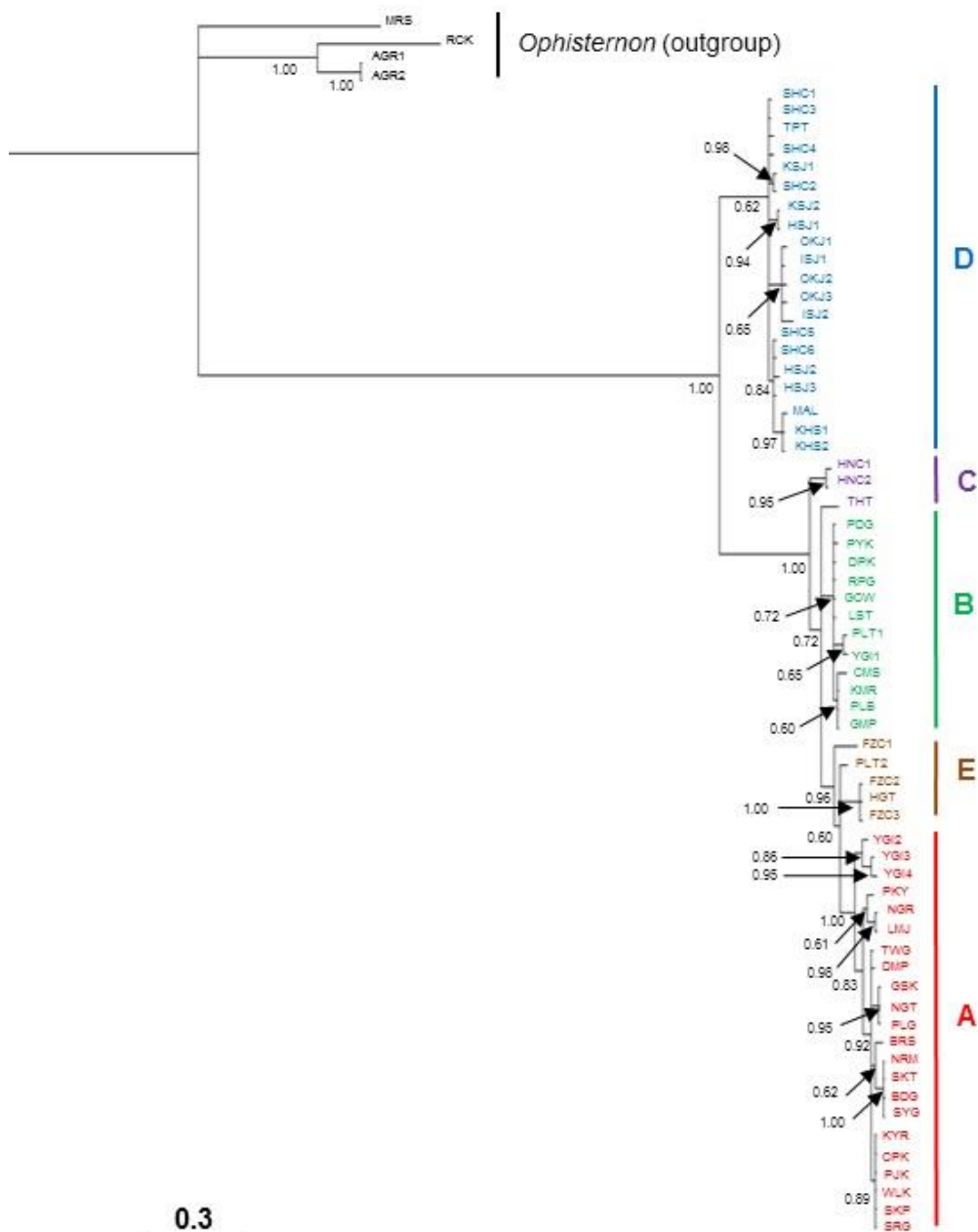


Figure 3.5. Bayesian tree inferred from 16S mitochondrial gene sequences combining between this study with Matsumoto *et al.* (2010) and Miya *et al.* (2001) data. Tree produced from 4×10^6 generations using GTR+G model. Number of each node represent posterior probabilities and scale correspond to substitutions/site

Table 3.6. Mean percentage nucleotide sequence divergence of an 539-bp fragment of the 16S mitochondrial gene among five identified clade of *M. albus* of this study combined with Matsumoto *et al.* (2010) and Miya *et al.* (2001) data

	Clade A	Clade B	Clade C	Clade D	Clade E
Clade A	0.97 (0-2)				
Clade B	3.32 (2-4.3)	0.43 (0-1.3)			
Clade C	3.76 (2.9-4.3)	2.57 (1.8-3.4)	1.33 (0.2-2)		
Clade D	8.68 (7.2-9.9)	7.69 (6.7-8.9)	7.48 (6.5-8.4)	1.06 (0-2.7)	
Clade E	2.84 (1.6-4.1)	3.03 (2.7-3.6)	3.13 (2.7-3.6)	8.01 (6.9-9.4)	1.23 (0-2.2)

Table 3.7. Haplotype of *M. albus* based on 16S mitochondrial gene region inferred from this study combined with Matsumoto *et al.* (2010) and Miya *et al.* (2001) data

Clade	Haplotype	Sample number	Individual code	Gen accession number	Country
A	HA9	1	PKY	KP729526	Indonesia
			GSK	KP729527	Indonesia
			NGT	KP729528	Indonesia
			PLG	KP729529	Indonesia
	HA11	1	TWG	KP729530	Indonesia
			SRG	KP729531	Indonesia
	HA12	1	BRS	KP729532	Indonesia
	HA14	2	LMJ	KP729533	Indonesia
			NGR	KP729540	Indonesia
	HA15	5	WLK	KP729534	Indonesia
			SKP	KP729535	Indonesia
			KYR	KP729536	Indonesia
			CPK	KP729537	Indonesia
			PJK	KP729538	Indonesia
			DMP	KP729539	Indonesia
	HA17	4	BDG	KP729541	Indonesia
			NRM	KP729542	Indonesia
			SYG	KP729543	Indonesia
			SKT	KP729544	Indonesia
	HA18	1	YGI2	AB494988	Indonesia
HA19	1	YGI3	AB494989	Indonesia	
HA20	1	YGI4	AB494990	Indonesia	
B	HB10	5	PDG	KP729516	Indonesia
			DPK	KP729518	Indonesia
			RPG	KP729523	Indonesia
			GOW	KP729524	Indonesia
			LST	KP729525	Indonesia
	HB11	1	PYK	KP729517	Indonesia
	HB12	1	CMS	KP729519	Indonesia
	HB13	3	KMR	KP729520	Indonesia
			PLB	KP729521	Indonesia
			GMP	KP729522	Indonesia
	HB14	1	PLT1	AB494986	Taiwan
HB15	1	YGI1	AB494987	Indonesia	
C	HC3	1	THT	KP729545	Vietnam
	HC4	1	HNC1	AB494984	China
	HC5	1	HNC2	AB494985	China

Table 3.7. Continued

Clade	Haplotype	Sample number	Individual code	Gen accession number	Country
D	HD3	2	KHS1	KP729546	Taiwan
			KHS2	KP729547	Taiwan
	HD4	1	MAL	NC_003192	No data
	HD5	1	SHC1	AB494967	China
	HD6	2	KSJ1	AB494968	Japan
			SHC2	AB494969	China
	HD7	1	SHC3	AB494970	China
	HD8	2	KSJ2	AB494971	Japan
			HSJ1	AB494972	Japan
	HD9	1	TPT	AB494973	Taiwan
	HD10	1	SHC4	AB494974	China
	HD11	3	SHC5	AB494975	China
			SHC6	AB494976	Japan
			HSJ3	AB494978	Japan
	HD12	1	HSJ2	AB494977	Japan
	HD13	1	OKJ1	AB494979	Japan
	HD14	1	ISJ1	AB494980	Japan
HD15	1	OKJ2	AB494981	Japan	
HD16	1	OKJ3	AB494982	Japan	
HD17	1	ISJ2	AB494983	Japan	
E	HE5	1	PLT2	AB494991	Taiwan
	HE6	1	FZC1	AB494992	China
	HE7	3	FZC2	AB494993	China
			FZC3	AB494995	China
			HGT	AB494994	Taiwan

Individual codes based on population codes are given in Table 3.2.

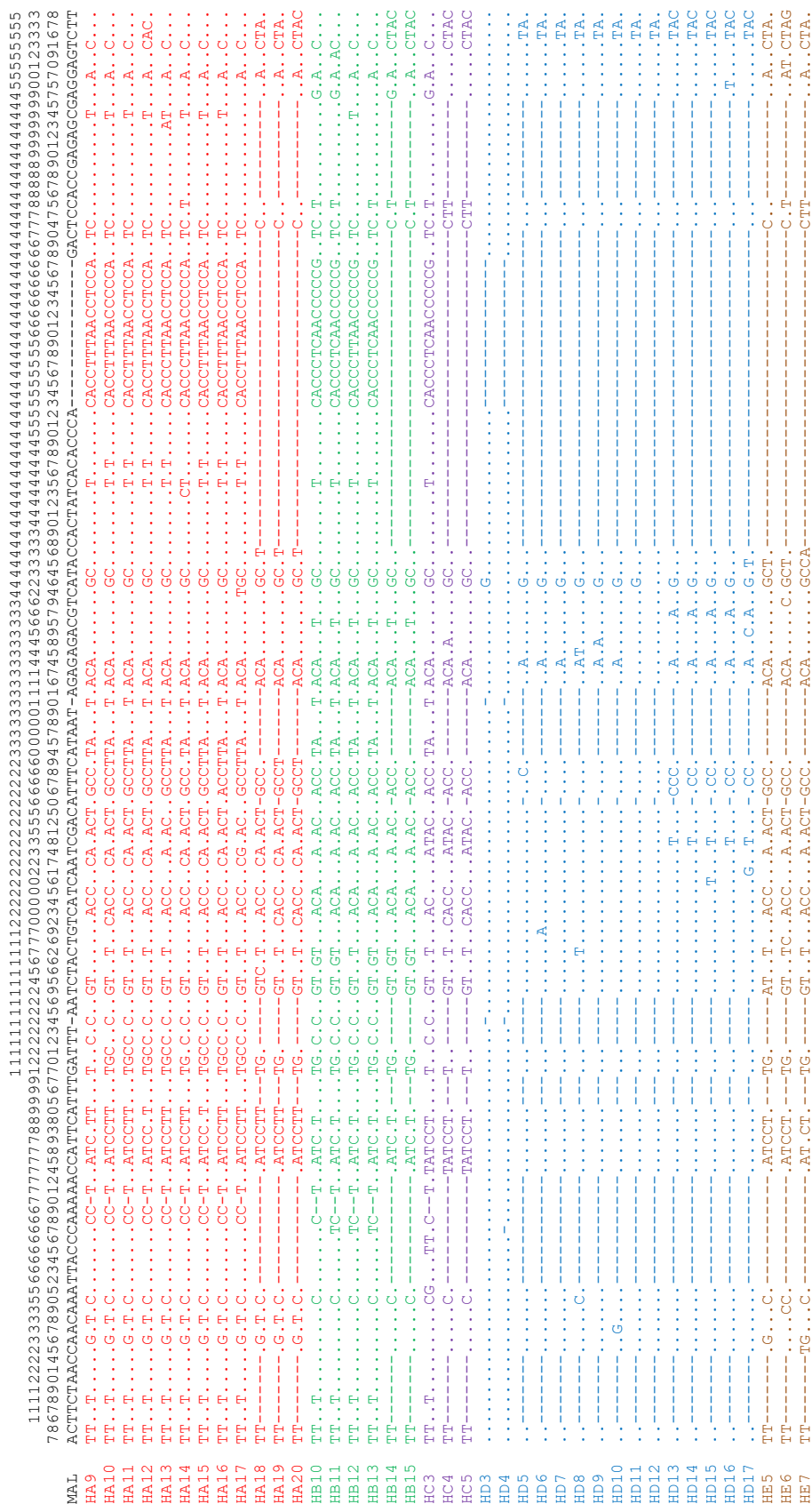


Figure 3.6. Summary of nucleotide variations in the partial 16S rDNA of *M. albus* in this study combined with the previous study (Matsumoto *et al.*, 2010 and Miya *et al.*, 2001). Only variable sites are shown. Haplotypes are named by letters referring to the clade and number. Dots indicate identity with the *M. albus* (MAL) sequence taken from GenBank with accession number NC_003192 as a reference (Miya *et al.*, 2001). A dash indicates insertions/deletions. The numbers shown above the reference sequence indicate the nucleotide positions in the alignment

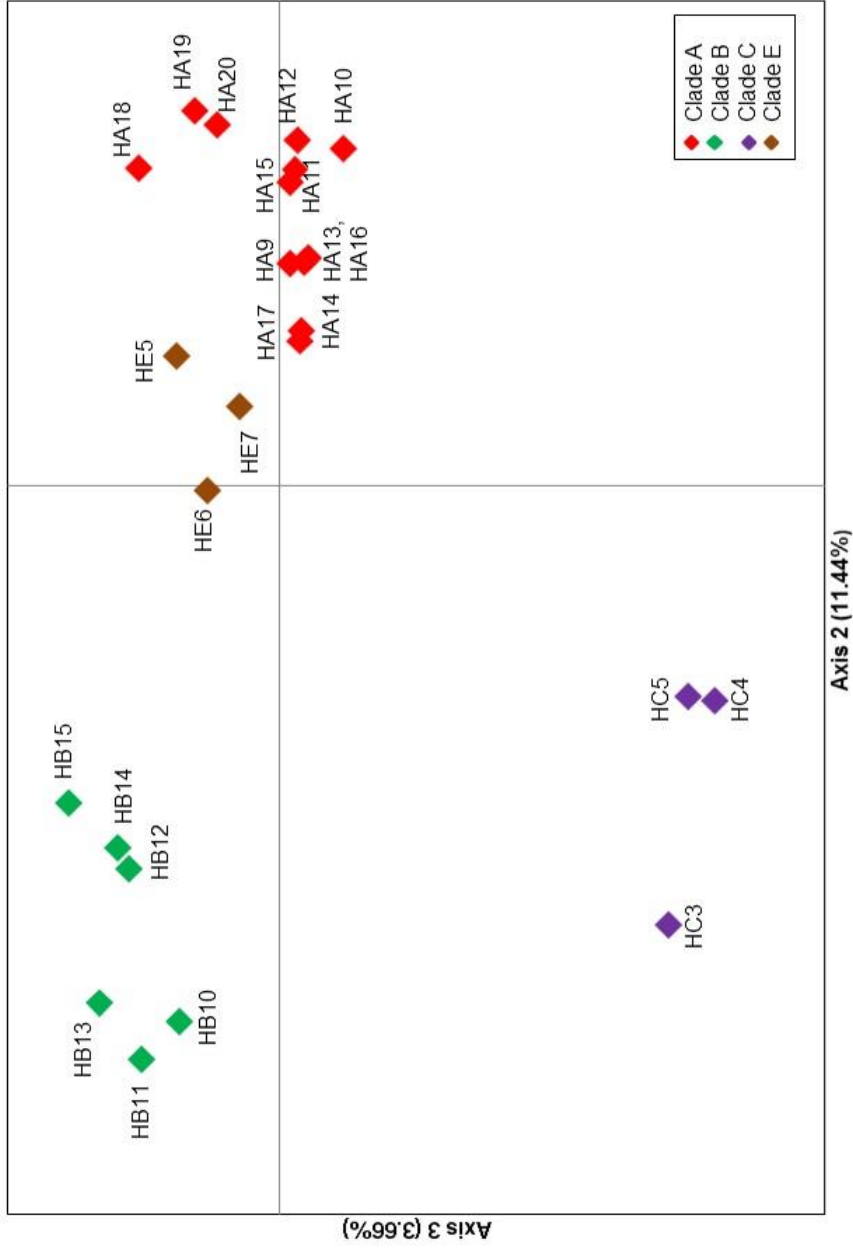


Figure 3.7. Principal Coordinate Analysis (PCoA) of pairwise genetic distances of 24 haplotypes of *M. albus* populations based on 16S mitochondrial sequences. This analysis incorporates data from Matsumoto *et al.* (2002) taken from GenBank given in Table 3.2.

Table 3.8 Estimates of genetic diversity for *M. albus* based on two partial 16S mitochondrial sequence data set

Sequence	Clade	Number of individuals	Number of haplotypes	Variable sites	Parsimony Informative sites	Haplotype diversity (h)	Nucleotide diversity (π)	Tajima's (D)	Fu's (F)
¹ 16S mtDNA - 1	A	19	8	11	10	0.89	0.0065	0.48	0.13
	B	17	9	10	6	0.86	0.0039	-0.61	-2.74
	C	2	2	6	0	1.00	0.0099	0.00	1.61
	D	4	2	1	0	0.67	0.0000	0.00	0.00
	E	7	4	5	2	0.71	0.0040	-0.09	-0.13
	All	49	25	74	63	0.96	0.0340	0.13	1.14
² 16S mtDNA - 2	A	22	12	93	12	0.91	0.0096	-0.76	-2.77
	B	12	6	81	4	0.80	0.0043	-0.64	-1.41
	C	3	3	84	0	1.00	0.0150	0.00	0.70
	D	20	15	80	13	0.97	0.0105	-0.73	-6.88
	E	5	3	13	5	0.70	0.0136	-0.05	2.75
	All	62	39	144	56	0.98	0.0460	0.92	0.41

¹ Primers - 16Sar and 16Sbr² Primers - H1567 and L2196

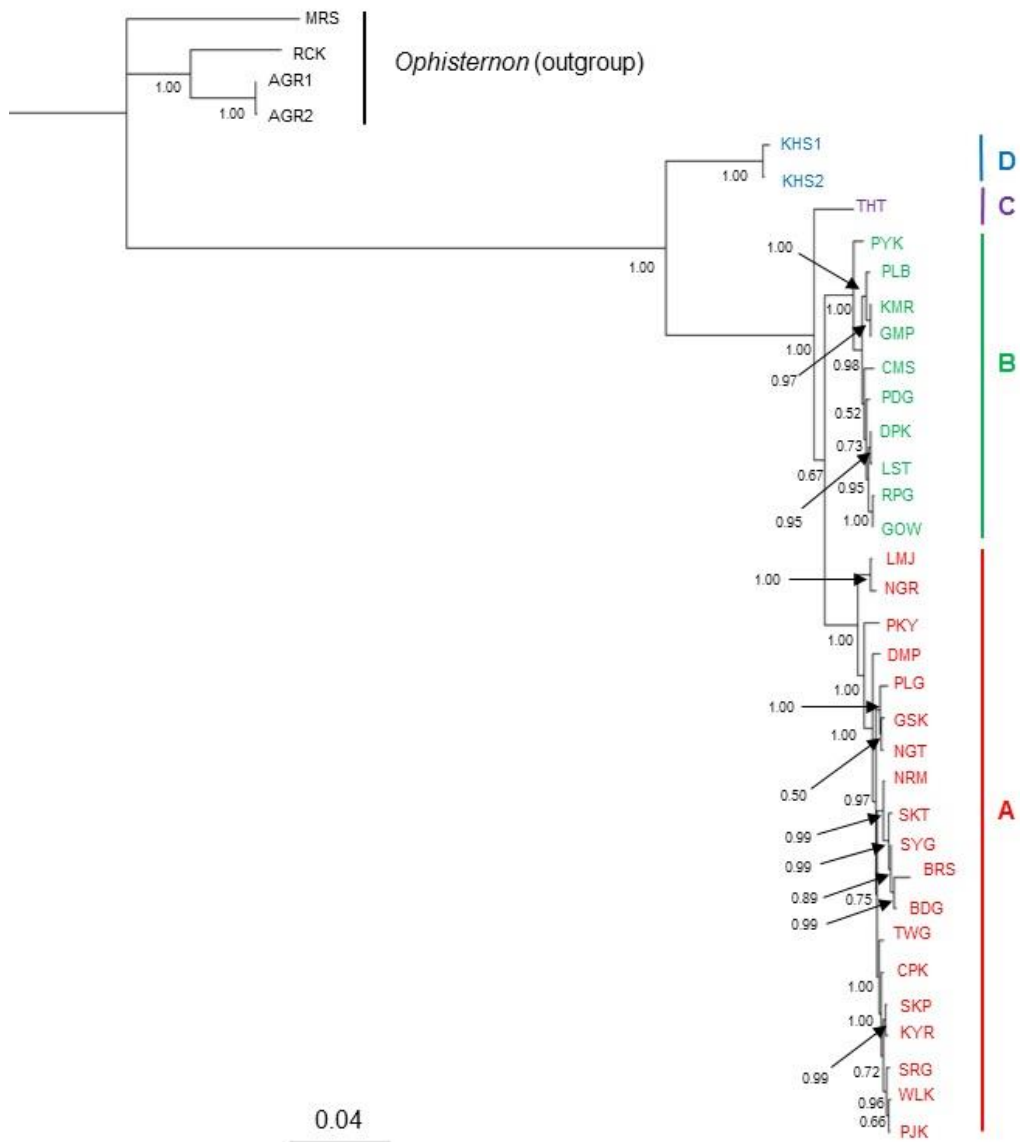


Figure 3.8. Bayesian tree inferred from mitochondrial and nuclear gene sequences. Tree produced from 2×10^6 generations using GTR+I+G model. Number of each node represent posterior probabilities and scale correspond to substitutions/site

Table 3.9. Mean percentage nucleotide sequence divergence of a 3,366-bp fragment of the multiple gene regions among four identified clade of *M. albus*

	Clade A	Clade B	Clade C	Clade D
Clade A	0.45 (0-0.88)			
Clade B	2.87 (2.56-3.19)	0.68 (0.06-1.3)		
Clade C	2.90 (2.81-3.13)	2.59 (2.46-2.67)	0 (0)	
Clade D	7.21 (6.99-7.54)	7.29 (7.09-7.50)	7.22 (7.13-7.30)	0.2 (0.2)

Table 3.10. Haplotype of *M. albus* based on mitochondrial and nuclear gene regions

Clade	Haplotype	Sample number	Individual code
A	HA1	1	PKY
	HA2	1	GSK
	HA3	1	NGT
	HA4	1	PLG
	HA5	1	TWG
	HA6	1	SRG
	HA7	1	BRS
	HA8	1	LMJ
	HA9	1	WLK
	HA10	1	SKP
	HA11	1	KYR
	HA12	1	CPK
	HA13	1	PJK
	HA14	1	DMP
	HA15	1	NGR
	HA16	1	BDG
	HA17	1	NRM
	HA18	1	SYG
	HA19	1	SKT
B	HB1	1	PDG
	HB2	1	PYK
	HB3	1	DPK
	HB4	1	CMS
	HB5	2	KMR, GMP
	HB6	1	PLB
	HB7	1	RPG
	HB8	1	GOW
	HB9	1	LST
C	HC	1	THT
D	HD1	1	KHS1
	HD2	1	KHS2

Individual codes based on population codes are given in Table 3.2.

Table 3.11. Estimates of genetic diversity for *M. albus* based on partial mitochondrial and nuclear sequences from the multigene data set

Sequence	Clade	Number of individuals	Number of haplotypes	Variable sites	Parsimony Informative sites	Haplotype diversity (h)	Nucleotide diversity (π)	Basepair (bp)
¹ 16S mt-DNA - 1	All	32	14	66	57	0.934	0.0288	502
² 16S mt-DNA - 2	All	32	15	103	75	0.927	0.0320	627
COI mt-DNA	All	32	17	118	106	0.950	0.0550	526
RAG1 nDNA	All	32	24	56	31	0.962	0.0064	1,384
1 st intron of S7 nDNA	All	32	20	41	26	0.948	0.0240	327
Multigene data set	A	19	19	107	63	1.000	0.0070	
	B	10	9	49	18	0.978	0.0044	
	C	1	-	-	-	-	-	
	D	2	2	6	0	1.000	0.0018	
	All	32	31	384	295	0.998	0.0237	3,366

¹ Primers - 16Sar and 16Sbr

² Primers - H1567 and L2196

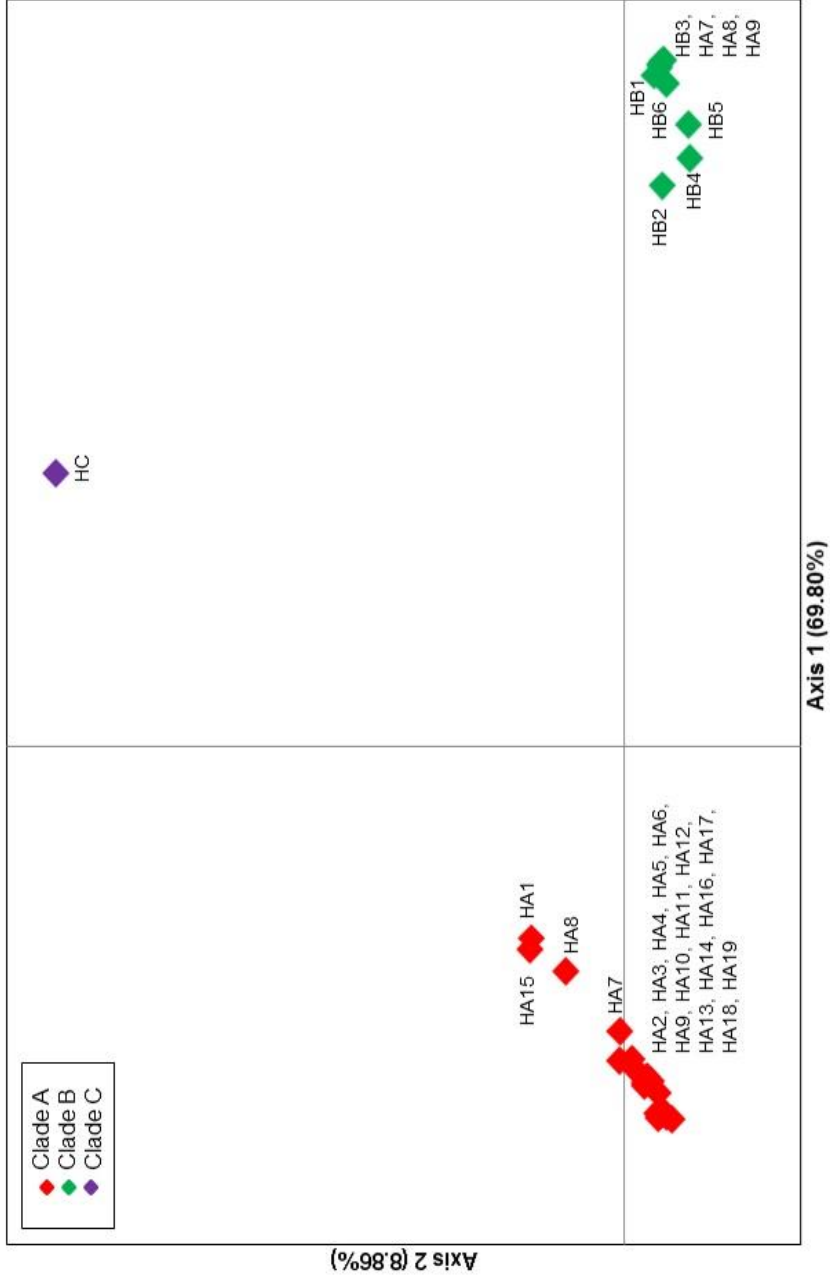


Figure 3.9. Principal Coordinate Analysis (PCoA) of pairwise genetic distances of 29 haplotypes of *M. albus* populations based on mitochondrial and nuclear sequences

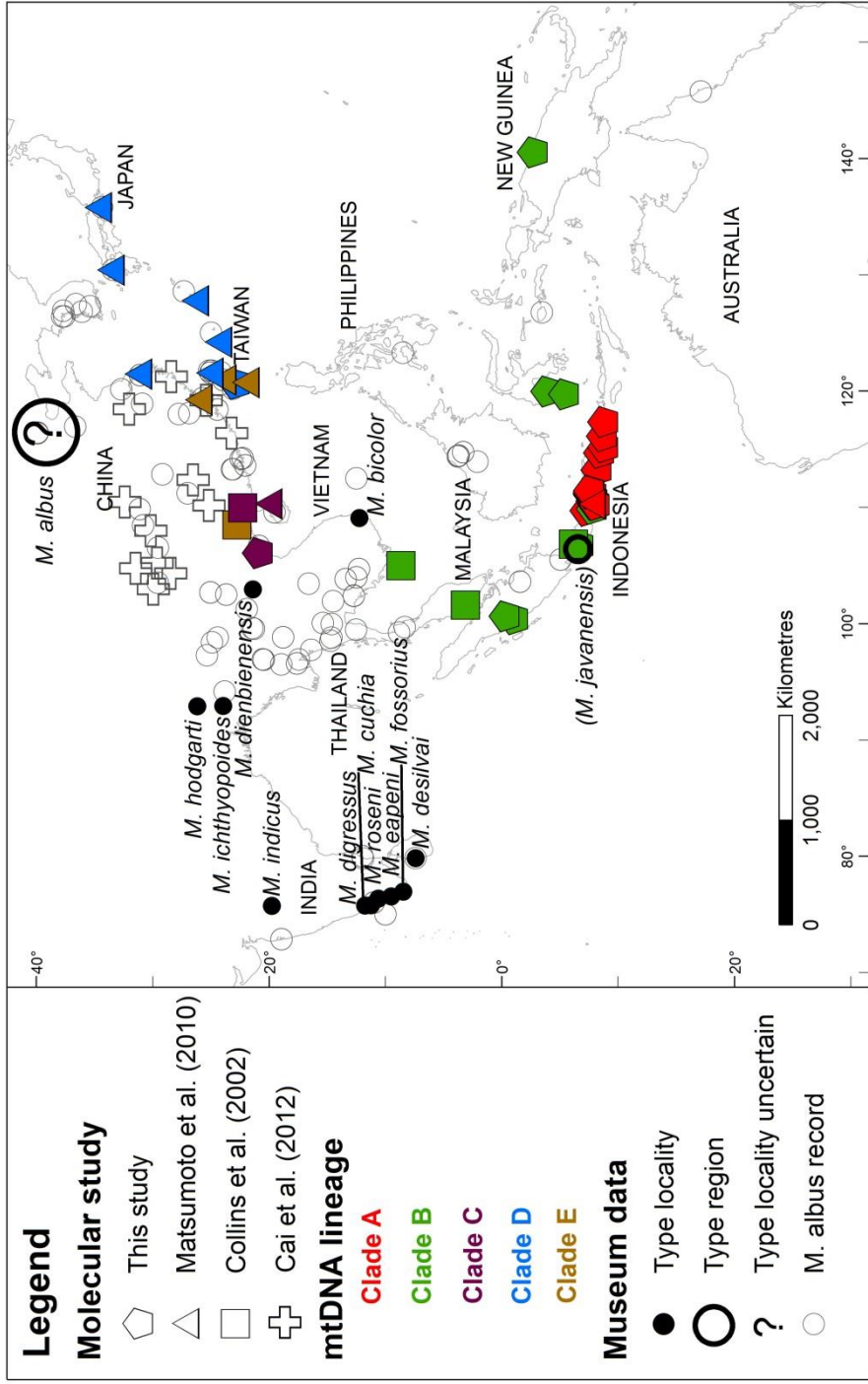


Figure 3.10. Map of distribution of putative species of the *M. albus* species complex from integrated studies and gene regions

CHAPTER 4

Morphological variation and numerical taxonomy of *Monopterus albus* species in Indonesia

4.1. Introduction

Taxonomy is the foundation for good and credible science, as accurate identification and the enumeration of species within a given experimental, systematic, ecological or biogeographical context underpins all research investigations and management decisions and is especially relevant for ecology and conservation (Blackwelder, 1967; Schuh, 2000; Wheeler, 2008). Morphological characters have been traditionally used in fish biology, as in other biological fields, to discriminate and measure relationships among various taxonomic categories including species. There are many well documented examples of morphological studies that have improved taxonomic understanding, and hence management, of some important commercial fishes such as sardine (Silva, 2003), mackerel (Bektas and Belduz, 2009), and tilapia (Hassanien *et al.*, 2011; Samaradivakara *et al.*, 2012).

Morphological taxonomy is based primarily on discrete morphological attributes, and morphometric and meristic characters (including internal anatomy and osteology). Morphometric characters are a set of measurements that represent size and shape variation of individuals and are continuous data (Bookstein, 1982). Conversely, meristic characters (i.e. vertebrae, fin rays, spines, scale rows, lateral line scales, gill rakers) are non-continuous data that can be counted. Therefore, the nature of morphometric and meristic characters

differ statistically and are often analysed separately in multivariate analyses (Ihssen *et al.*, 1981). Morphometric and meristic studies are widely used to detect differences between fish populations and have provided useful results for identifying fish species and stocks, both often matched to underlying biological or environmental variation that can then assist fisheries management and taxonomy (Murta, 2000; Pinheiro *et al.*, 2005).

In the face of an increasing number of phylogeographic, phylogenetic and DNA-barcoding studies it is becoming apparent that conventional morphological morphology-based studies, even using sophisticated methods of numerical taxonomy, frequently underestimate the number of biological species, sometimes grossly (Raadik, 2014). Nevertheless, it is important to integrate molecular and morphological analyses as everyday identification requires morphological diagnostic characters, and species need to be placed within existing nomenclatural frameworks, often requiring reference to type material (DeSalle *et al.*, 2005; Will *et al.*, 2005). Further, often the methodologies are complementary, with diagnostic morphological traits often emerging after groups are defined genetically (Austin and Knott, 1996; Lin *et al.*, 2005; Cheng, 2010; Abdurahman *et al.*, 2012; Blanton *et al.* 2013; Muchlisin, 2013).

A very high proportion of Indonesia's freshwater fishes, a fauna of approximately 1,400 species or 7% of total global species, lack rigorous taxonomic treatments using either or both morphological and molecular data thus impeding the understanding of the true diversity of these important vertebrates in the country (Rhee *et al.*, 2004). Within this context it may be possible that many presumed single species or morphotypes of obligate freshwater fish with wide geographic distributions potentially comprise several independent evolutionary lineages that could represent cryptic species, which is a significant issue world-

wide (Bickford *et al.*, 2006; Beheregaray and Caccone, 2007; Hellberg, 2009; Hammer *et al.*, 2013, Raadik, 2014).

The family Synbranchidae is one of the most important groups of freshwater food fish in tropical Asia, including Indonesia. A notable and popular species in Indonesia is the swamp eel *Monopterus albus* Zuiew 1793, however from a taxonomic viewpoint this is one of the most enigmatic and least understood freshwater species within the region. Lacking scales, most fins and any sort of external hard body parts, they are character poor, and hence this is likely to have hindered the taxonomy of the species beyond broad-based and often superficial morphological assessments. Since described by Zuiew in 1793, the swamp eel has remained a taxonomic puzzle with many researchers questioning the significance of morphological variation and the validity of many described species. A total of 13 synonyms have been listed for *Monopterus albus* across six genera (Rosen and Greenwood, 1976). The last comprehensive taxonomic revision of *M. albus* based on morphology and osteological features was conducted by Rosen and Greenwood (1976) who determined that this widespread polytypic species was best referred to as a single species, which has been largely accepted by the relevant taxonomic community. For example previous workers had recognised several species in China (e.g. Nichols, 1943) on the basis of differences in tail length, eye size, coloration, and head shape, but Rosen and Greenwood (1976) considered that there were no consistent and diagnostic patterns apparent, and therefore no justification for recognising more than a single species. Within this context a form of *Monopterus* described from the Sunda Strait in Java, Indonesia (*Monopterus javanensis* Lacepède 1800) was also considered a synonym by Rosen and Greenwood (1976) and therefore, it is commonly accepted that the only valid swamp eel species in Indonesia is *M. albus*.

In contrast to the prevailing morphological determination of a single widespread species referred to as *M. albus*, this study (chapter 2 and 3) and consistent with previous molecular genetic studies (Collins *et al.*, 2002; Matsumoto *et al.*, 2010; Cai *et al.*, 2012), highlights that *M. albus* contains multiple major lineages likely to represent a complex of cryptic species. Based on my findings from chapters 2 and 3 it is apparent that two cryptic species of *Monopterus* occur in Indonesia that have overlapping distributions and are reproductively isolated.

An important requirement for genetically defined species is to explore patterns of morphological variation and to seek diagnostic taxonomic traits. Thus the objective of this study was to investigate variation in morphology using morphometric and meristic characters within and between the genetically differentiated forms of *M. albus* I identified in Indonesia to support a comprehensive taxonomic treatment.

4.2. Materials and Methods

4.2.1. Sample Collections

Swamp eels were collected from seven locations from across Java, Lombok, West Sumatra and Sulawesi, Indonesia. The regions selected in this study were based on the results of the population genetic studies and analysis of the multigene data sets, which revealed that Indonesian swamp eels belong to one of two distinct genetic lineages that behave as biological species (see chapters 2 and 3). Four populations representing haplotype A and three populations representing haplotype B were chosen from locations previously sampled and genotyped (Chapters 2 and 3). Populations representing haplotype

A comprised sample from Planggu, Klaten (PLG), Tunggul, Sragen (SRG), Dempel, Ngawi (DMP) and Narmada, Lombok (NRM), and populations representing haplotype B comprised Indarung, Padang (PDG), Ciomas, Bogor (CMS), and Sindendreng, Rappang (RPG) (Figure 2.1 in chapter 2). Twenty individuals from each population were obtained from farmers or fishermen and preserved in 95% absolute ethanol with the exception of those from PLG, which were salt-preserved in the first instance. Each individual fish was genotyped and allocated to either haplotype A or haplotype B (see chapter 2). Preservation in the field was difficult due to the nature of the samples being donated by farmers who arranged for eels to be caught from their rice paddy fields in remote locations, combined with a lack of formalin as this chemical is banned by the Indonesian government due to misuse as a food preservative. Summary information of sampling location, population code, collection date, number of specimens and total length can be seen in Table 4.1.

Voucher specimens from each location, identified as belonging to haplotype A or haplotype B, have been submitted to Museum Biologi, Faculty of Biology Universitas Gadjah Mada, Indonesia (Museum Biologi-UGM) and the Museums and Art Galleries of the Northern Territory, Australia (MAGNT), and will also be deposited to Museum Zoologi Bogor, Indonesia (MZB). The photographs of a representative individual conforming to haplotype A and haplotype B can be seen in Figures 4.1. and 4.2.

4.2.2. Meristic and Morphometric Procedures

Only a single meristic character could be assessed in this species consisting of the number of vertebrae as these fish are lacking fins and scales. This also follows the treatments of previous workers on *Monopterus* taxonomy,

namely Rosen and Greenwood (1976), Bailey and Gans (1998), Gopi (2002), and Britz *et al.* (2011), who consider this to be an important diagnostic trait. The total number of vertebrae for each specimen was initially determined by exposure for 30 seconds at 50 kV of power in a radiographic unit (Torrex 150 D, Astrophysic Research Corporation) at MAGNT. The vertebrae number of each individual was counted from the radiographic negative film using a light box and a high-powered magnifying glass. The number of vertebrae was counted twice independently on the same day for each fish.

A total of 13 morphometric measurements were taken on each individual fish, including those variables that have been emphasized in previous taxonomic treatments (Rosen and Greenwood, 1976; Bailey and Gans, 1998; Gopi, 2000; Britz *et al.*, 2011). The details of these characters are given in Table 4.2 and Figure 4.3. Morphometric measurements were taken from the left lateral aspect, and measured with digital callipers to the nearest 0.01 mm (digital vernier callipers, Supatool, China) except total length which was measured using a ruler to the nearest 0.5 mm. Standard length is challenging to measure due to the difficulty in finding the end of the vertebral column or back of the hypural plate in swamp eels and is also error prone as fish often have damaged tails.

During counts of vertebrae, the tail tip and associated vertebrae were routinely encountered to have been damaged in the wild and healed. Further there was obvious damage to the tail of other fish evidently occurring during capture or in transport, handling and preservation, with the available method of preservation (ethanol only) resulting overall in fish specimens being quite brittle. As a result a number specimens with damaged tails could not be included in the morphological dataset (see below).

4.2.3. Morphological data and statistical analysis

To examine morphological variation among the *M. albus* samples morphometric and meristic characters were initially analysed separately since they have different statistical properties and distributions (Ihssen *et al.*, 1981). Further, preliminary statistical analyses indicated all variables were strongly correlated with total length with the exception of vertebrae number. As variation in size of fish from within and between populations can vary substantially and can overwhelm analyses, morphometric variables were statistically adjusted in relation to size using ratios and regression techniques (Thorpe, 1976). Non-size adjusted data sets were also analysed for comparative purposes and to identify potentially useful taxonomic characters.

After initial physical screening, the number of undamaged specimens suitable for inclusion in the data set were: PLG (n=12), NRM (n=11), SRG (n=16), DMP (n=17), PDG (n=14), CMS (n=12), and RPG (n=6). Therefore total number (N) of individuals for the morphometric analyses was 56 and 32 individuals representing haplotypes A and B respectively. The total number of samples for which vertebrae number could be obtained was less: PLG (n=8), NRM (n=5), SRG (n=14), DMP (n=15), PDG (n=7), CMS (n=7) and RPG (n=2). Every specimen in the meristic data set was also included in the morphometric data set. The total number of individuals for the meristic data sets was 42 and 16 for haplotype A and B respectively. The sample numbers are different between the morphometric data and the vertebrae number because in some specimens the tail tip can be detected but the vertebrae number cannot be unambiguously assessed.

Initial statistical screening after removing fish with obviously damaged tails or other deformities, was by examination of the distribution of each

morphometric character against total length using x-y scatterplots of untransformed variables to examine for obvious outlying values. Separate statistical analyses were conducted on morphometric and meristic data since morphometric data are continuous and more susceptible to age and environmentally induced variability, while meristic data are discrete and fixed early in development (Hermida *et al.*, 2005; Simon *et al.* 2010). Pearson's correlation coefficients were calculated for both morphometric and meristic characters against total length. Correlation analyses indicated a significant relationship between vertebrae number and total length (TL) for the pooled samples ($r = 0.561$, $p < 0.001$), but not with the samples grouped by haplotypes (A: $r = -0.049$, $p = 0.761$; B: $r = 0.185$, $p = 0.492$). Therefore, vertebrae count was not adjusted for size differences. Conversely, significant correlations were detected between size (TL) and all morphometric characters for both pooled and within grouped correlations (Table 4.3). Therefore, transformations were explored to reduce the effects of size differences within and between populations from overwhelming the analyses (Thorpe 1976). A ratio transformation of each metric variable against total length was used, which has been traditionally used by taxonomists to allow size free comparisons between samples (Atchley and Anderson, 1978). In addition, a method to generate size-independent shape characters using regression analyses was explored using the following formula (Senar *et al.*, 1994; Doherty and McCarthy, 2004; Simon *et al.*, 2010):

$$A'_{ij} = \text{Log}_{10}(A_{ij}) - (B(\text{Log}_{10}(TL_i) - \text{Log}_{10}(TL)))$$

where A'_{ij} is the adjusted value of character j for individual i , A_{ij} is the original value, B is pooled regression coefficient of $\log A$ on $\log TL$, TL_i is the total length of individual i , and TL is the overall mean of total length. This is referred to as the size-transformed data set. The efficacy of size transformation was determined

from correlation analyses (r-value) between the two kinds of transformed variables and Log₁₀ TL.

4.2.4. Multivariate statistical analysis

Two widely used multivariate procedures were used to explore phenotypic variation represented by the 14 variables measured within and between swamp eel populations and to identify the important distinguishing characters. Principal Component Analysis (PCA), using a correlation matrix, was used to reduce the data set to a reduced number of compound variables that efficiently represented the variation among samples. This approach makes no assumptions about group or population membership and allows the contribution of variables to patterns of variations to be assessed based on their correlation with the principal components (Thorpe 1976). Multiple Discriminant Analysis (MDA) is also designed to reduce complex multivariate datasets to a small number of compound variables. It requires that groups are predefined and, in taxonomic applications, it can be applied at the population level which allows assessment of morphological divergence between putative species. It is also a powerful method for taxonomic discrimination based on predefined groups representing species and identifies the most efficient discriminating variables (Albrecht *et al.* 1980; Austin and Knott, 1996).

All statistical routines and multivariate analyses were implemented using the XLSTAT software version 2015.1.02 (Addinsoft). PCA was conducted on several different data sets subject to different transformations. The procedure utilised the Pearson Correlation matrix and the Varimax rotation options. An initial exploratory analysis used the full data set of seven populations (n=88) for the 13 log₁₀ transformed data. Next, analyses were conducted on a reduced data set

consisting of six populations (n=76) comprising the same 13 log₁₀ transformed variables, 12 ratio variables (each variable divided by TL) and 12 size-transformed variables using the procedure described above. The effectiveness of the different analyses was evaluated based on the Eigen values and the percentage of variation explained on the first three components or discriminant functions. The relationships between swamp eels populations and their genetically defined lineage was evaluated by bivariate plots of the scores on the first three axes (components) and the contribution of variables was assessed on the basis of their correlations on each of the first three axes. The effectiveness of the two methods of data transformations was also examined by examining correlations of the transformed variables against TL.

Discriminant Function analysis (DFA) was performed on the same three data sets comprising the six populations of swamp eels examined by PCA. Discriminant Analysis maximises the separation of the predefined groups relative to that within groups, rather than simply maximising the overall variance among samples as in PCA. Thus DFA also identifies those variables that contribute most to the discrimination of the groups. The contribution of each character to discrimination among populations and groups was evaluated from the correlations between individual scores on each of the first three discriminate functions and each variable. The efficiency and usefulness of the DF analyses for separating and identifying populations was measured using the classification function (Albrecht *et al.*, 1980).

Two group Discriminate Function analyses were used to identify variables of potential taxonomic use for discriminating swamp eels from the two genetically defined haplotypes A and B. First, DFA was applied to the Log₁₀ transformed morphometric data and the ratio transformed data (the size transformed data set was not included as the method of transformation prohibits

its use for the identification of single specimens). Next DF was applied to the full data set inclusive of the previously removed PLG population and the meristic variable, vertebrae number. This resulted in a data set of 14 variables and 58 samples made up as follows: PLG (8), NRM (5), SRG (14) and DMP (15), representing haplotype A (42) and PDG (7), CMS (7) and RPG (2), representing haplotype B (16). For simplicity the analyses were based on the untransformed data.

4.3. Results

4.3.1. Vertebrae number

Radiographs were obtained for 140 individuals. However difficulties were encountered in discerning small distal vertebrae in smaller sized individuals and due to some having damage to tail tips. Thus the the final data set included those fish with confident counts and that had presumed complete tails (random checking of the presence of caudal rays under a dissecting microscope was conducted for confirmation), leaving a reduced number of individuals (n= 58) (Table 4.3). An inspection of a bivariate scatterplots with TL (Figure 4.2) and histogram (Figure 4.3) revealed that this single discrete character is diagnostic for haplotype A and B based on the samples examined and not influenced by the length of the fish. The variability in number of vertebrae within the two groups is high, ranging from 160 to 174 for haplotype A and from 175 to 182 for haplotype B, but the distributions are not overlapping (Table 4.3, Figures 4.4, 4.5 and 4.6)

4.3.2. Morphometric Data

A summary of the morphometric data is given in Table 4.4 and correlation analysis of each metric variable against total length of populations grouped by haplotype and the pooled data is given in Table 4.5. All variables had positive correlations with TL and all were significant ($P < 0.05$) with the exception of variable HL in the haplotype A populations.

4.3.2.1. Principal Component Analysis (PCA)

The PCA of the seven populations based on the log₁₀ transformed data efficiently summarised the variability within the data set. The first three components accounted for 94.8% of the variation but with almost all of this accounted for on the first component (89.9%) (Table 4.6). All variables had high positive correlations ($r > 0.89$) on the first component, indicating that this component represents size variation. The bivariate plots show that population PLG is the most divergent with high negative scores on component 1 (Figure 4.7). The other components suggest that other populations are divergent to a lesser extent such as NRM and SRG with positive scores on component 2 and with the exception of sample PLG members of the two haplotypes are separated to some extent on component 3. HD is the most significant variable contributing to the patterns of variation on PC2 and ED is the most important variable on PC3 (Table 4.6).

The preceding analysis indicates that in general swamp eels from haplotype B tend to be larger than those from haplotype A, which is consistent with the statistical summary for most variables (Table 4.4). Further, the analysis

indicates that swamp eels from site PLG are substantially smaller than all other samples. This may be due to environmental factors or sampling variation and possibly the method of preservation using salt. Whatever the cause, their much smaller size than the other samples means that size-related morphometric differences dominates the analyses and prevents more detailed analysis of variation among the remaining samples. Thus the eel samples from population PLG were removed from the next series of analyses.

Prior to the PCA and MDA analyses of the six populations, the effectiveness of transformations to reduce size-related variation was examined using correlation analysis (Table 4.7). From this table it can be seen that all variables are strongly positively correlated with TL for the untransformed and log₁₀ transformed data with the level of correlation being slightly higher for the untransformed data. Surprisingly, neither of the size transformations were effective at removing the effects of size, with the regression-based size-transformation being significantly less efficient than the ratio transformation data. For the size-transformed data all variables were positively correlated except for variable ED and 9 out of 12 were significant ($p < 0.05$) and the highest r value of 0.697 for variable SL. The ratio transformation was more efficient with four variables with negative correlations and with only five being significant ($p < 0.05$). Nevertheless some correlations were substantial, with variable ED having the highest correlation (-0.721) followed by variable SL (0.656).

Principal Component analysis applied to the six populations using the log₁₀ transformed morphometric variables is given in Table 4.8. The first three components account of 89.1% of the variation with the first component responsible for 78% with all variables having strong positive correlations on this axis. Components 2 and 3 have both positive and negative correlations indicating differences based on size. Variables HD, ED and GL with positive correlations

are contrasted with variables TL, WGA and GWB with negative correlations on component 2. For component 3, GBD and WGA are the most important variables. Inspection of the component scores (Figure 4.8) show that swamp eels from populations CMS and RPG (haplotype B) are consistently larger than all the other populations, with swamp eels from NRM and DMP (haplotype A) being the smallest. Samples from DMP, SRG (haplotype A) and PDG (haplotype B) have the widest overlap in their size ranges. PC3 indicates that these three populations differ in shape based on their haplotype, with PDG (haplotype B) samples having negative scores compared with DMP and SRG (haplotype A) with mostly positive scores. This indicates that these samples differ in shape with respect to HD, ED, GL with positive correlations and TL, WGA and GWB with negative correlations. While PC3 does not effectively separate all samples of the two haplotypes, in combination with component 1, reflecting size variation, there is almost no overlap between the two haplotypes for fish in the same size range (Figure 4.8).

The outcome of the Principal Component analysis on the ratio transformed variables is given in Table 4.9. The first three components account for less variation (69.3%) than the previous analysis, with the variation spread more evenly across the three axes. Surprisingly, all variables are positively correlated with component 1, indicating size-related variation still dominates the analysis, while components 2 and 3 have both positive and negative correlations indicating differences based on shape. Variables ED/TL (positive correlation) and SL/TL and STE/TL (negative correlation) are the two most important variables on component 2 and WGA/TL and GBD/TL with positive correlations are the most significant on PC3. Examination of the component scores (Figure 4.9) shows some separation of swamp eels from the two haplotypes on a combination of axes 2 and 3. PC3 highlights some differences between DMP + SRG (haplotype A) and PDG (haplotype B) with the latter having negative scores, but all other

populations have widely overlapping, mainly positive scores. While PC3 does not effectively separate all samples of the two haplotypes, in combination with component 2, there is minimal overlap between the two haplotypes (Figure 4.9).

A summary of the PCA of the size-transformed variables is given in Table 4.10 and the scores are plotted on Figure 4.10. The results are quite similar to the ratio-transformed analyses. The first three components account for slightly more variation (71.1%) than the previous analysis, with slightly more variation on component 1 (46.7%). Again all the variables have positive, mostly high, correlations, on component 1, indicating it represents mostly size variation. Component 2 and 3 both have positive and negative correlations indicating shape differences. Variables ED (positive correlation) and SL, WGA and STE (negative correlations) are the most important variables on component 2 and WGA and GBD with positive correlation are the most significant variables on PC3. This is also quite similar to the ratio-transformed analyses. Examination of the component scores (Figure 4.10) also give similar results to the ratio data, but there is generally a greater scatter of points within each population. The combination of components PC1 and PC3 and PC2 and PC3 indicates some differentiation between the swamp eels from the two haplotypes, however there is greater overlap between populations representing each haplotype in each scatter plot. Thus this data set seems to highlight within population variability, some of which may be related to the nature of the data transformation.

4.3.2.2. Discriminant Function Analysis (DFA)

The results of the DFA applied to the log₁₀ transformed morphometric variables are given in Table 4.11. The first three functions account for 95.1% of the variation with the first function accounting for 66.9% with many variables

having strong positive correlations. On Function 2 all variables have negative, mostly high correlations again indicating separation of population based on components of size. Variable correlations on DF3 have both negative and positive correlations indicating population separation based on shape. However no variables have strong correlations with GBD (0.286) and HL (-0.295), being the most important. The classification function correctly allocated all samples to their original population indicating that each population has a distinctive set of morphometric attributes. Examination of the DF scores shows that there is minimal overlap among populations with the combination DF1 and DF2 differentiating populations CMS and RPG (haplotype B) from the other populations. Similarly, the combination of DF1 and DF3 differentiates populations NRM and DMP (haplotype A) from the other populations (Figure 4.11). However none of first three functions clearly separates the three populations of haplotype A and the three populations of haplotype B (Figure 4.11), although the combinations of DF2 and DF3 minimises the overlap of individuals between the two haplotypes.

The details of the analysis of the ratio transformed morphometric variables are given in Table 4.12. The first three functions account for 95% of the variation with the first function accounting for 55.1% with variables having positive and negative correlations, indicating that size is not dominating the variation. On Function 2 all but one variable have positive, mostly, high correlations indicating separation of population based on components of size. Variable correlations of DF3, like DF1, have both negative and positive correlations indicating population separation based on shape. ED/TL and HD/TL are the most influential variables on DF1. The classification function correctly allocated all samples to their original populations. The DF1 and DF2 scores highlight the differences between populations CMS and RPG (haplotype B) and populations NRM and DMP

(haplotype A) (Figure 4.12). None of first three functions separates the three populations of haplotype A and the three populations from haplotype B (Figure 4.12), although the combinations of DF2 and DF3 is the most effective, outliers within several populations, prevents any clear separation of the two genetic forms.

The outcome of the DFA applied to the size transformed morphometric variables is given in Table 4.13. The first three functions account for 94.5% of the variation with the first function accounting for 38.5%, indicating that the variation is spread more evenly across the first three axes. Surprisingly, all variables having positive correlations, indicating that DF1 is based on size differences. On Functions 2 and 3 variables have both positive and negative correlations indicating that population separation is on the basis of shape. The most important variables on DF2 are ED and HD and on DF3 GBD and HL. The classification function again correctly allocated all samples to their original populations indicating that each population has a distinguishing set of morphometric attributes. The scatterplot of DF scores (Figure 4.13) shows that populations CMS and RPG (haplotype B) are well separated from DMP, but with the other populations having intermediate scores on DF1. The combination of DF1 and DF3 provides the best discrimination of haplotype A samples (NRM, SRG and DMP) from the haplotype B samples (PDG, CMS and RPG), however there are still several outliers from several populations that prevents a clear spatial separation (Figure 4.13).

Two group DFA applied to all seven populations (n=88) using the log₁₀ and ratio transformed data sets were largely successful at discriminating swamp eels from the two haplotypes (Tables 4.14 and 4.15 and Figure 4.14). The log₁₀ transformed data were slightly more successfully at discriminating the two haplotypes. The distributions of scores showed more overlap on the analysis of

the ratio transformed data (Figure 4.14) and the classification of individuals was not 100% successful (Table 4.15). The variable correlations were all positive for the log₁₀ transformed data set, with STE having the highest correlation (0.903). In contrast the ratio data set had mixed positive and negative correlations, but as with the previous analysis the variable STE (snout tip to posterior margin of orbit, STE/TL) had the highest correlation.

Two group DFA was also applied to the seven populations using a data set comprising all morphometric variables and the meristic variable (vertebrae number), with the results summarised in Table 4.16 and Figure 4.15. The log transformed and untransformed data set gave very similar results with 100% successful discrimination based on the classification function. Vertebrae number and STE had the highest correlations in both these analyses. The discriminant analysis based on just these two variables also gave 100% successful discrimination and classification (Figure 4.15C). A bivariate plot of these variables effectively separates the two haplotypes and represents the two best morphological variables for taxonomic identification (Figure 4.16) of the two species of swamp eels.

4.4. Discussion

An analysis of morphological variation of populations of swamp eels from the two genetically defined haplotypes from Indonesia was successful in demonstrating that the two forms can be phenotypically distinguished. This therefore provides further justification for their recognition as separate species. However, the analysis of morphological variation was not straightforward and, if conducted in isolation from the genetic analysis would not have led to the

independent evidence of two the forms or taxa of *Monopterus* that are clearly defined from genetic analyses. In other words, while phenotypic differences could be detected, there was no support for a morphological “gap” between the species as the degree of morphological variation within and between populations was substantial.

The strongest evidence for a correlation between the genetic results and the morphological data emerges from the meristic data (vertebral counts) and the multivariate analysis of ratio-transformed metric data. The vertebrae counts were quite variable within and between populations, but showed no overlap between haplotype A (range: 160-174) and haplotype B (range: 175-182) for the fish samples examined. While the average size of fish from haplotype B was significantly larger than haplotype A, this does not account for the observed difference. The vertebrae count is not correlated with size within each group and there is a region of significant overlap in the size of fish, where the differences between the two forms in vertebrae counts are maintained.

The multivariate analysis of morphometric data in fish (and other organisms with indeterminate growth) is highly contentious in relation to how environmental or age related variation leading to size differences should be managed (Thorpe, 1976; Bookstein, 1982). In this study I used two approaches, one using formula for eliminating size (Simon *et al.*, 2010) and a utilitarian approach using ratios. These were also compared with analyses based on the same data set without size transformations. While the use of ratio has attracted much criticism over the years (Thorpe, 1976; Atchley and Anderson, 1978; Bookstein, 1982), it is intuitively more meaningful relative to one of the purposes of the study (distinguishing genetically defined groups) and also allows comparisons with the fish taxonomic literature, which still widely employs ratio for characterising and distinguishing species (Lin *et al.*, 2005; Minos *et al.*, 2012;

Neto *et al.*, 2012). Interestingly, the less sophisticated method based on ratios was more successful in identifying differences between fish from the two haplotypes than the size-transformation method.

Multivariate methods of analysis of the data sets subject to the two kinds of data transformation and the untransformed data set emphasised the high degree of morphological variation within and between populations. Indeed the extent of this variation makes it impossible to distinguish the two genetically divergent, biological species identified in chapter 2 without reference to the prior genetic classification. The data sets without adjustment for size variation were more effective at distinguishing the populations and identifying differences between the two haplotypes. Surprisingly both the ratio and size-transformed data sets still exhibited size related variation based on correlations with TL and and correlations on the Principal Component or Discriminant Function axes. Further, the cruder method, using ratios tended to perform better in relation to the correlation analyses and the multivariate analyses.

The two group DFA was successful in that it demonstrated that difference between the two genetic forms of swamp eels do exist and can be used for taxonomic discrimination. Further, these analyses, when applied to all variables and all populations gave 100% discrimination. In addition, analyses of the two best discriminating variables, vertebrae number and snout tip to posterior margin of orbit (STE) successfully distinguishes all individuals of the two haplotypes.

A caveat for the use of DFA is that it is a very powerful method for distinguishing and the subsequent reclassification of pre-defined groups. Thus when the pre-defined groups are at the species level this can not be used as independent evidence for species recognition (Albrecht, 1980). Nevertheless success of the two group analysis is useful as the groups were defined on the basis of an independent genetic analysis and the outcome supports practical

taxonomic identification. Thus an unknown swamp eel individual could be measured for the same variables used in this study and using the DF equations assigned to one or other of the swamp eel populations with a high probability (Albrecht 1980, Austin and Knott, 1996; Khan *et al.*, 2012; Samaradivakara *et al.* 2012).

My data on vertebrae counts for haplotype A (160-174) is consistent with, but slightly higher than the data reported by Rosen and Greenwood (1976), for *M. albus*, who reported a range of 142-172, although most of the fish they investigated have a vertebrae number above ≥ 160 . Thus finding the fish from haplotype B populations (PDG, CMS, RPG) to have counts from 175-183 indicates they are quite divergent from the typical *Monopterus* phenotype. This result suggests that these populations, and therefore haplotype B, the more widespread of the two forms, may represent *M. javanensis* which also has a high vertebrae count (188) as reported by Regan (1912).

While the present study found significant variability in morphometric and meristic traits among the six populations of *M. albus* some characters were better for discriminating between fish from the two haplotypes. Head depth (HD), head length (HL), ratio head depth with total length (HD/TL) and ratio of head length with total length (HL/TL) were found to be important discriminating variables with populations NRM, SRG and DMP (haplotype A) in general having shorter HD, HL and HL/TL and longer HD/TL than populations PDG, CMS and RPG (haplotype B). This is consistent with the taxonomic literature as Eapen (1963) reported that, *M. albus* specimens have a total length 10-13 times that of the head length and 17-26 times body depth. In this study, length of the body in population NRM, SRG, DMP (haplotype A) is approximately 12-13 times in the total length of the head whereas the total head is 11-12 times that of the body population PDG, CMS, and RPG (haplotype B). Similarly, greatest body depth (GBD) in

populations NRM, SRG, DMP (haplotype A) is approximately 25-28 times that of the total length of the body while it is 24-32 for populations PDG, CMS, and RPG (haplotype B).

The findings of the present study suggest that the two haplotypes do differ in morphological traits, but these are obscured by a high degree of within and between population genetic variation. This is consistent with an observation by Allendorf *et al.* (1987) that fishes are known to demonstrate greater variance in morphometric traits both within and between populations than other vertebrates, and are more susceptible to environmentally induced morphological variation. Therefore different environmental factors such as water temperature, food availability, type of soils in paddy fields, and density may influence the phenotypic differentiation of these populations. It would therefore be of interest to conduct a morphological study of fish from the two haplotypes after they had been raised from the larval stage to adult hood under controled or communal conditions. It would also be important to control for other variables such as age/size and sex in such a study. For example it would be useful to morphologically compare individuals of the different haplotypes that were as closely matched as possible with respect to overall size.

Finally it is recommended that further morphological studies be conducted on additional populations of these species and those from the other haplotypes identified in chapter 2 to fully document and explore morphological variation within this group and to search for diagnostic characteristic that support or correlate with the molecular genetic patterns. Such studies would clearly benefit from the use of multivariate techniques and the sampling of fish from type localities and measuring of type specimens. In this regard the classification function of MDF is a very powerful as it can be used to predict group membership

of an unknown specimen (e.g. type specimen) with known (measured) populations (Austin and Knott, 1996).

Table 4.1. Sampling location, date collection, number of specimens, total length and voucher catalogue numbers

Sampling Location	Population Code	Haplotype	Latitude / Longitude	Date Collection	Number of specimens	Total length (mm)	Museum Biologi-UGM Voucher Catalogue Number
Prayan-Pianggu, Trucuk, Klaten, Central Java	PLG	A	07°44'46.33"S/110°41'06.39"E	27 June 2011	20	190-221	(47a)FW.R.III.1.
Tunggul, Gondang, Sragen, Central Java	SRG	A	07°27'17.53"S/111°07'22.29"E	8 July 2011	20	256-341	(47b)FW.R.III.1.
Dempel, Geneng, Ngawi, East Java	DMP	A	07°27'18.79"S/111°26'05.11"E	8 July 2011	20	277-374	(47c)FW.R.III.1.
Narmada, Lombok Barat, West Nusa Tenggara (NTB)	NRM	A	08°39'19.00"S/115°10'13.07"E	23 June 2011	20	228-279	(47d)FW.R.III.1.
Inderung-Lubuk Kilangan, Padang, West Sumatra	PDG	B	0°56'44.36"S/100°25'43.79"E	15 June 2011	20	297-381	(48a)FW.R.III.1.
Ciomas, Bogor, West Java	CMS	B	06°33'54.94"S/106°44'31.97"E	24 May 2011	20	362-414	(48b)FW.R.III.1.
Rappang, Panca Rijang, Sindendreg Rappang, South Sulawesi	RPG	B	03°47'00.60"S/120°00'52.72"E	13 June 2011	20	344-384	(48c)FW.R.III.1.

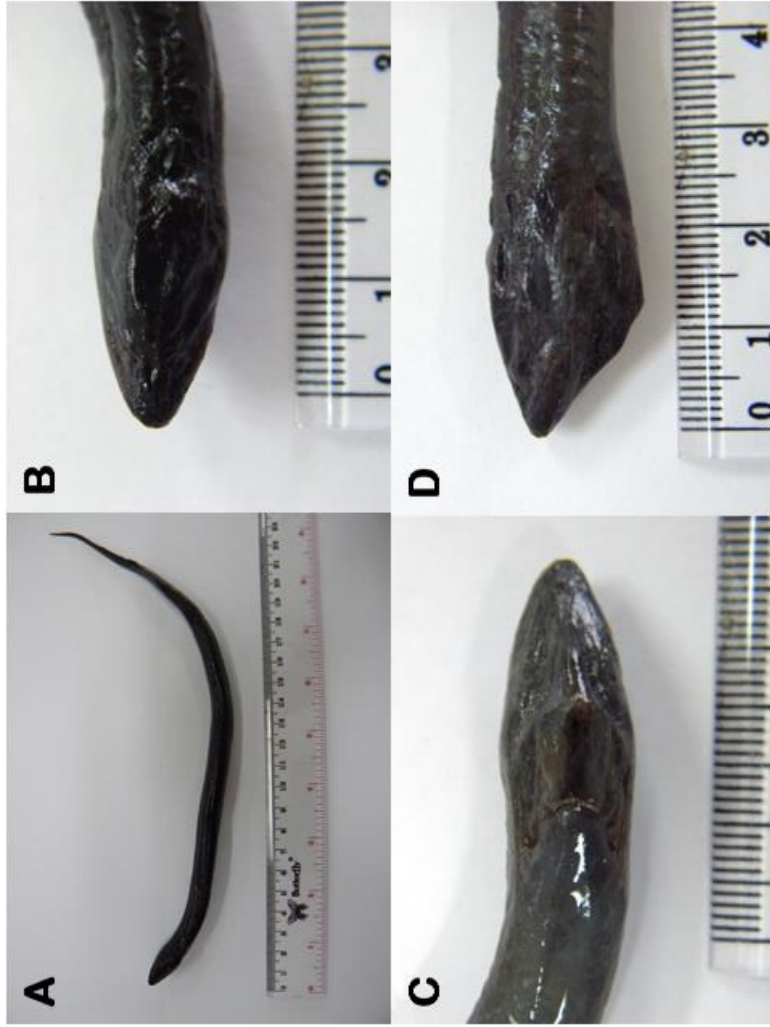


Figure 4.1. Morphology of *M. albus* haplotype A. (A) Whole body; (B) Dorsal head; (C) Ventral head; (D) Lateral head



Figure 4.2. Morphology of *M. albus* haplotype B. (A) Whole body; (B) Dorsal head; (C) Ventral head; (D) Lateral head

Table 4.2. Definitions of meristic counts and morphometric measurements and their codes *M. albus* used in this study

Character	Description	Code
Total vertebrae count	Total number of vertebrae in vertebral column	TV
Total length	Tip of the upper jaw to the end of tail base	TL
Head length	From the front upper lip to the posterior end of the gill aperture	HL
Head depth	Maximum depth measurement of the head	HD
Head width	The greatest width measurement of head	HW
Snout length	From the front of the upper lip to the fleshy anterior edge of the orbit	SL
Snout tip to occiput	From the front of the upper lip to occiput	STO
Snout tip to gill aperture at midline	From the front upper lip to the midline of gill aperture	STGA
Snout tip to posterior margin of orbit	From the front upper lip to the posterior edge of the orbit	STE
Gape length	From the front upper lip	GL
Eye diameter	The greatest bony diameter of the orbit	ED
Width gill aperture	Width between gill apertures	WGA
Greatest body depth	Maximum depth measured from the base of the dorsal spine	GBD
Greatest width of body	The greatest width of the body	GWB

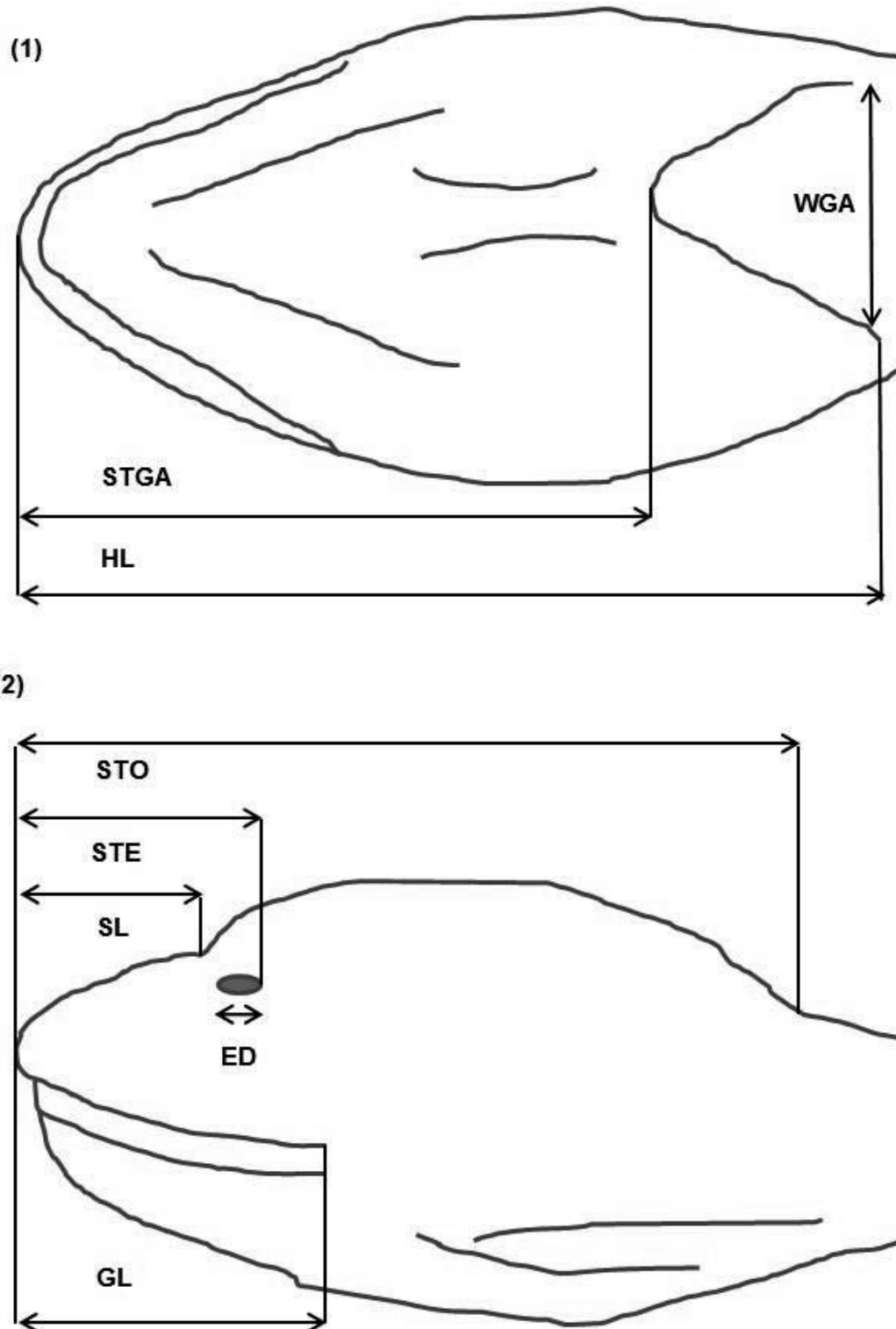


Figure 4.3. Morphometric measurements of *M. albus*
(1) Ventral view; (2) Lateral view

Table 4.3. Distribution of vertebrae number frequency in seven populations of *M. albus*

Population code	Haplotype	N	Vertebrae number																						
			160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
PLG	A	8			1			2	2	2	1														
NRM	A	5		1				1	1	1	1	2													
SRG	A	14	1	1	1		3	2	3	1				1											
DMP	A	15	2		2	2	1	2	1	2			1	1	1										
PDG	B	7														3				1	1	1	1		
CMS	B	7																2		1	2	1			1
RPG	B	2																1							1

Population codes are given in Table 4.1.

Table 4.4. Summary of morphometric for data for 13 variables measured in millimetres (mean±SD) for seven populations of *M. albus*. Sample size is given in parentheses

Variables	Haplotype A						Haplotype B						
	PLG (12)	NRM (11)	SRG (16)	DMP (17)	PDG (14)	CMS (12)	PLG (12)	NRM (11)	SRG (16)	DMP (17)	PDG (14)	CMS (12)	RPG (6)
TL	203.92±12.29	258.81±17.22	296.63±24.54	308.94±15.46	335.07±22.57	384.92±15.13	203.92±12.29	258.81±17.22	296.63±24.54	308.94±15.46	335.07±22.57	384.92±15.13	361.00±16.83
STO	13.95± 1.58	20.19± 1.71	22.06± 3.26	22.21± 1.90	25.62± 2.44	30.78± 1.71	13.95± 1.58	20.19± 1.71	22.06± 3.26	22.21± 1.90	25.62± 2.44	30.78± 1.71	29.77± 1.05
STGA	11.87± 1.10	18.55± 1.51	20.47± 3.05	18.09± 1.58	22.80± 2.00	27.03± 2.23	11.87± 1.10	18.55± 1.51	20.47± 3.05	18.09± 1.58	22.80± 2.00	27.03± 2.23	26.09± 2.11
HL	14.21± 1.60	21.68± 1.68	23.70± 2.98	23.37± 1.96	27.88± 2.58	32.31± 1.53	14.21± 1.60	21.68± 1.68	23.70± 2.98	23.37± 1.96	27.88± 2.58	32.31± 1.53	32.02± 1.48
SL	2.16± 0.15	3.54± 0.43	4.26± 0.45	4.30± 0.37	4.83± 0.43	6.52± 0.45	2.16± 0.15	3.54± 0.43	4.26± 0.45	4.30± 0.37	4.83± 0.43	6.52± 0.45	5.98± 0.43
STE	3.34± 0.40	4.85± 0.44	5.16± 0.55	5.36± 0.41	6.29± 0.37	8.15± 0.75	3.34± 0.40	4.85± 0.44	5.16± 0.55	5.36± 0.41	6.29± 0.37	8.15± 0.75	7.51± 0.49
GL	6.39± 0.79	9.31± 0.96	10.01± 1.48	8.53± 1.11	10.45± 0.66	11.74± 0.27	6.39± 0.79	9.31± 0.96	10.01± 1.48	8.53± 1.11	10.45± 0.66	11.74± 0.27	13.83± 0.96
ED	0.73± 0.07	1.71± 0.17	1.83± 0.18	1.66± 0.12	1.81± 0.14	1.89± 0.07	0.73± 0.07	1.71± 0.17	1.83± 0.18	1.66± 0.12	1.81± 0.14	1.89± 0.07	1.98± 0.07
WGA	3.63± 0.56	4.81± 0.73	5.73± 0.62	6.01± 0.62	5.77± 0.51	7.88± 0.49	3.63± 0.56	4.81± 0.73	5.73± 0.62	6.01± 0.62	5.77± 0.51	7.88± 0.49	7.13± 0.82
HD	6.79± 0.96	12.75± 1.61	13.97± 2.33	9.81± 0.86	11.97± 1.00	15.89± 1.18	6.79± 0.96	12.75± 1.61	13.97± 2.33	9.81± 0.86	11.97± 1.00	15.89± 1.18	16.18± 1.28
HW	5.11± 0.67	8.74± 1.28	9.15± 1.26	8.70± 0.89	9.38± 0.97	11.81± 0.83	5.11± 0.67	8.74± 1.28	9.15± 1.26	8.70± 0.89	9.38± 0.97	11.81± 0.83	11.65± 0.96
GBD	6.06± 0.50	9.78± 1.44	11.67± 1.41	10.64± 1.12	10.38± 0.59	15.59± 1.11	6.06± 0.50	9.78± 1.44	11.67± 1.41	10.64± 1.12	10.38± 0.59	15.59± 1.11	14.04± 1.25
GWB	5.09± 0.61	7.65± 0.80	8.24± 1.05	8.65± 0.88	9.58± 1.09	11.28± 0.84	5.09± 0.61	7.65± 0.80	8.24± 1.05	8.65± 0.88	9.58± 1.09	11.28± 0.84	9.99± 0.81

Population codes are given in Table 4.1.

Table 4.5. Correlation coefficients for TL against 12 morphometric variables each haplotype of *M. albus* and for the pooled data

Morphometric characters	r		Pooled N=76
	Haplotype A N=44	Haplotype B N=32	
STO	0.785	0.926	0.938
STGA	0.567	0.824	0.866
HL	0.854	0.944	0.957
SL	0.834	0.891	0.932
STE	0.659	0.786	0.882
GL	0.384	0.539	0.714
ED	0.404	0.639	0.584
WGA	0.808	0.776	0.839
HD	0.128	0.755	0.540
HW	0.559	0.760	0.788
GBD	0.705	0.790	0.782
GWB	0.701	0.683	0.847

*Values in bold are different from 0 with a significance level $P = 0.05$

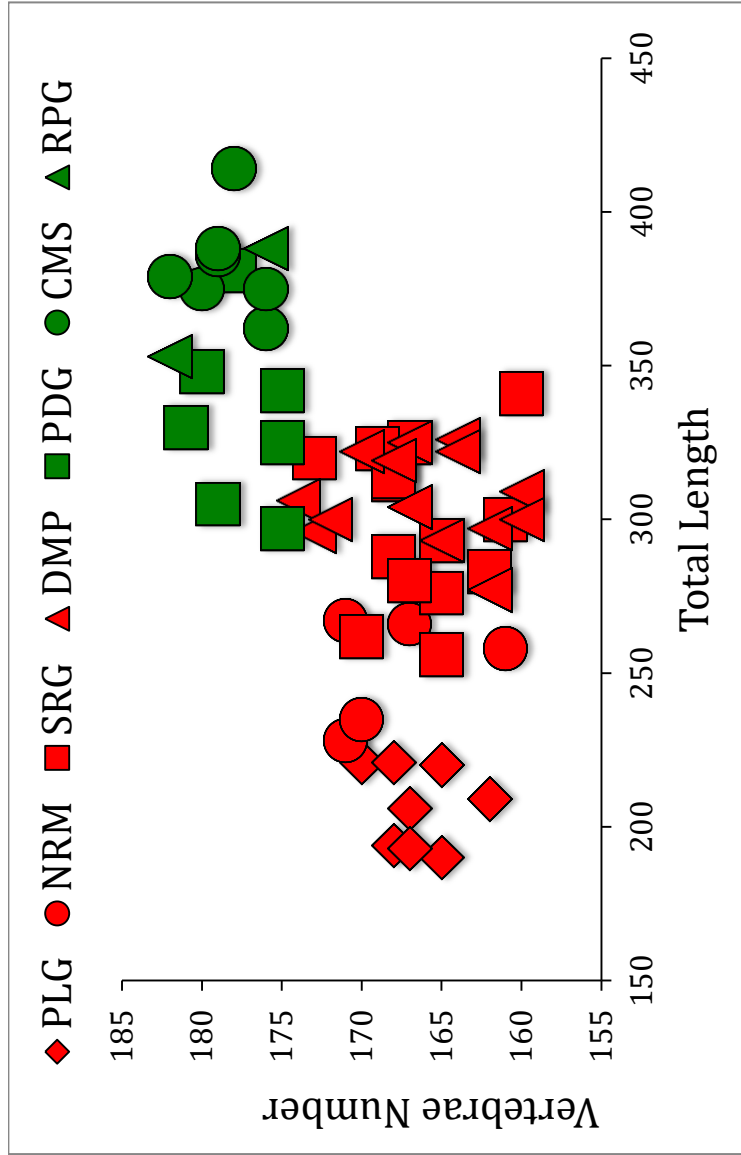


Figure 4.4. Scatterplot for total length versus vertebrae count for seven populations of *M. albus*. Haplotype A, red symbols (n=42), haplotype B, green symbols (n=16)

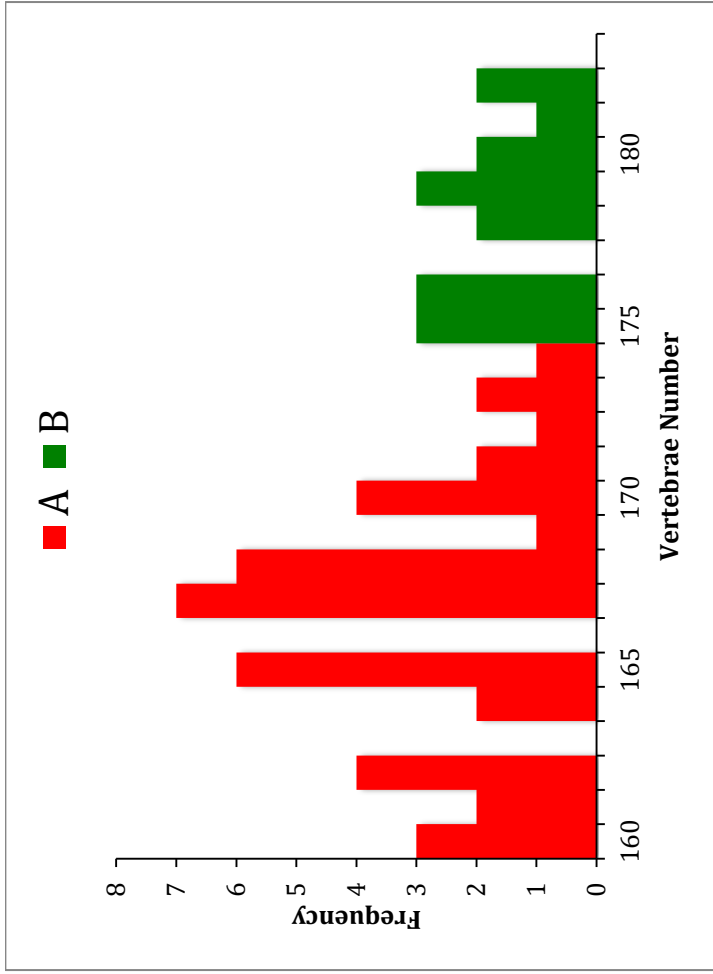


Figure 4.5. Vertebrae frequency histogram for *M. albus* for haplotype A (n=42) and haplotype B (n=16)



Figure 4.6. Vertebrae of *M. albus*. (A) haplotype A with vertebrae count: 161; (B) haplotype B with vertebrae count: 175

Table 4.6. Principal component analysis of 13 log transformed morphometric characters for *M. albus* from seven populations. Variable loadings are given for the first three components.

Variables/ PCA statistics	PC1	PC2	PC3
TL	0.969	-0.202	-0.004
STO	0.982	-0.057	-0.088
STGA	0.975	0.056	-0.093
HL	0.988	-0.059	-0.065
SL	0.972	-0.123	0.018
STE	0.947	-0.137	-0.108
GL	0.923	0.192	-0.251
ED	0.893	0.216	0.329
WGA	0.917	-0.233	0.069
HD	0.898	0.395	-0.059
HW	0.961	0.069	0.067
GBD	0.951	0.040	0.120
GWB	0.943	-0.115	0.085
Eigen value	11.683	0.398	0.238
% of variance	89.871	3.059	1.833
Cumulative % variance	89.871	92.930	94.763

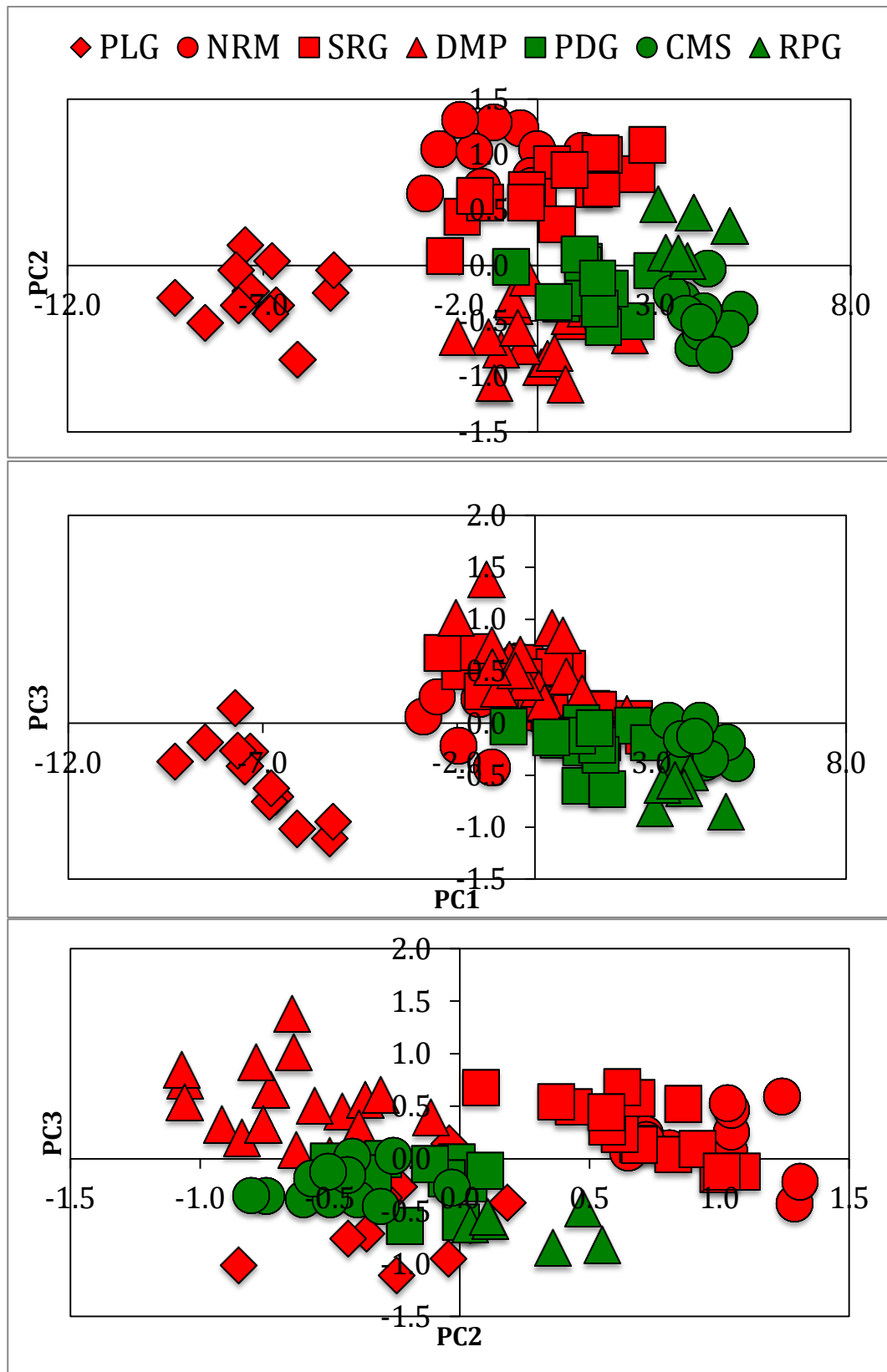


Figure 4.7. Plot of Principal Component scores for the first three axes for seven populations of *M. albus* based on 13 log₁₀ transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.7. Correlation coefficients (r) for untransformed, Log10 transformed morphometric characters and ratios with Total Length for six populations of *M. albus*.

Morpho- metric characters	r untrans- formed	r log10 transformed	r size- transformed morphometric characters	r ratios transformed
STO	0.936	0.925	0.448	0.339
STGA	0.866	0.848	0.358	0.189
HL	0.954	0.948	0.525	0.440
SL	0.931	0.931	0.697	0.656
STE	0.882	0.879	0.532	0.422
GL	0.714	0.703	0.206	-0.119
ED	0.585	0.585	-0.306	-0.721
WGA	0.836	0.839	0.366	0.170
HD	0.539	0.497	0.180	-0.198
HW	0.786	0.774	0.189	-0.106
GBD	0.783	0.779	0.328	0.113
GWB	0.844	0.842	0.242	0.007

*Values in bold are different from 0 with a significance level $\alpha=0.05$

Table 4.8. Principal Component analysis of 13 log₁₀ transformed morphometric characters for six populations of *M. albus*. Variable loadings are given for the first three components as correlations.

Variable	PC1	PC2	PC3
TL	0.928	-0.299	-0.065
STO	0.963	-0.066	-0.123
STGA	0.947	0.111	-0.132
HL	0.970	-0.085	-0.155
SL	0.925	-0.225	-0.025
STE	0.905	-0.208	-0.160
GL	0.859	0.323	-0.186
ED	0.738	0.458	-0.054
WGA	0.835	-0.291	0.340
HD	0.744	0.574	0.117
HW	0.892	0.103	0.181
GBD	0.880	0.048	0.386
GWB	0.853	-0.256	-0.058
Eigen value	10.135	1.014	0.439
% of variance	77.964	7.799	3.376
Cumulative % variance	77.964	85.763	89.140

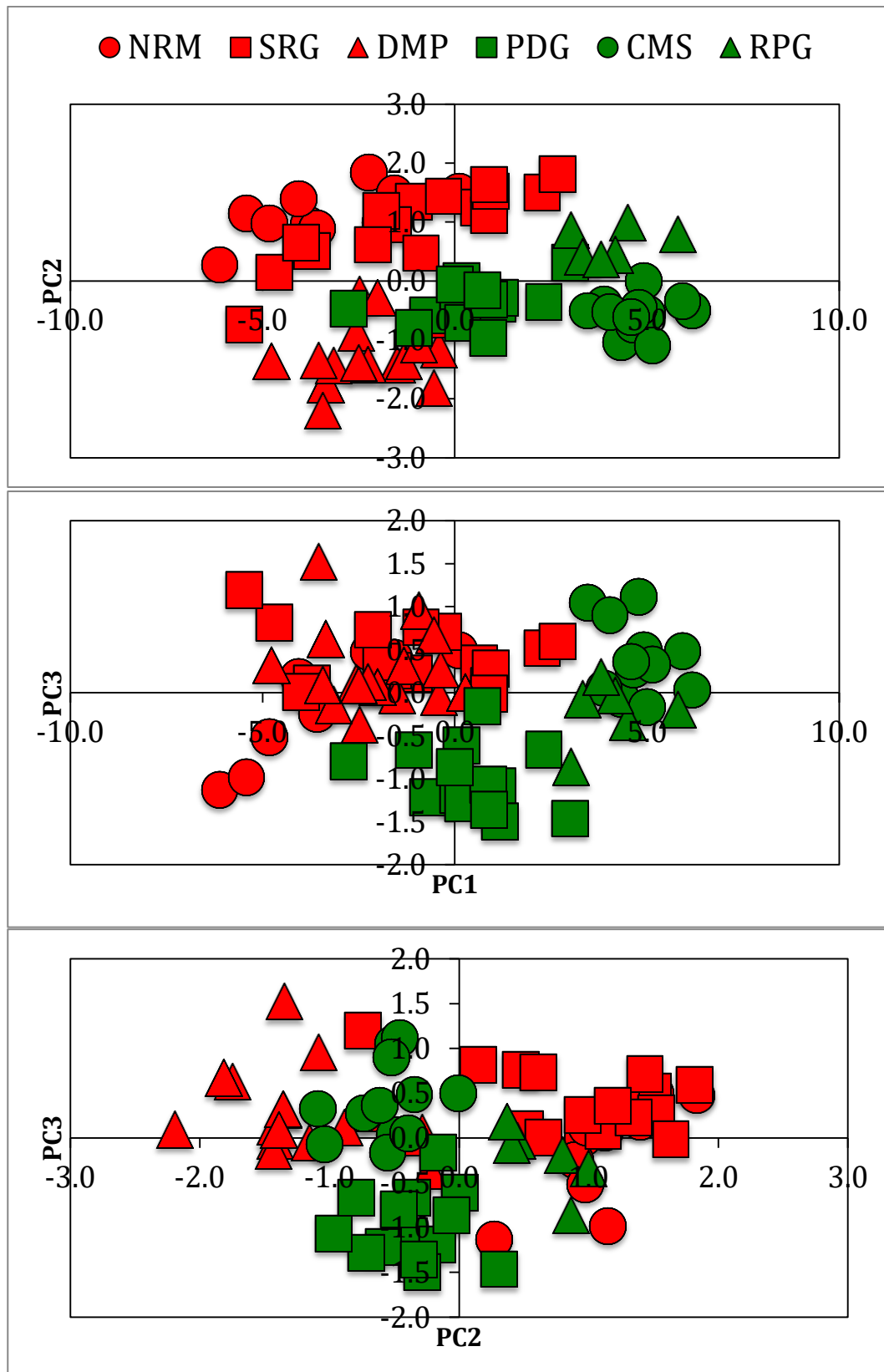


Figure 4.8. Plot of Principal Component scores for the first three axes for six populations of *M. albus* based on 13 log₁₀ transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.9. Results for a Principal Components analysis of 12 ratio-transformed morphometric characters for *M. albus* from six populations. Variable loadings are given for the first three components as correlations.

Variable	PC1	PC2	PC3
PCA statistics			
STO/TL	0.816	-0.184	-0.253
STGA/TL	0.857	0.002	-0.201
HL/TL	0.810	-0.249	-0.342
SL/TL	0.417	-0.743	0.158
STE/TL	0.524	-0.625	-0.109
GL/TL	0.768	0.279	-0.171
ED/TL	0.310	0.822	-0.009
WGA/TL	0.195	-0.177	0.802
HD/TL	0.809	0.408	0.046
HW/TL	0.693	0.314	0.277
GBD/TL	0.612	0.017	0.613
GWB/TL	0.316	-0.014	0.062
Eigen value	4.837	2.088	1.389
% of variance	40.310	17.400	11.573
Cumulative % variance	40.310	57.710	69.283

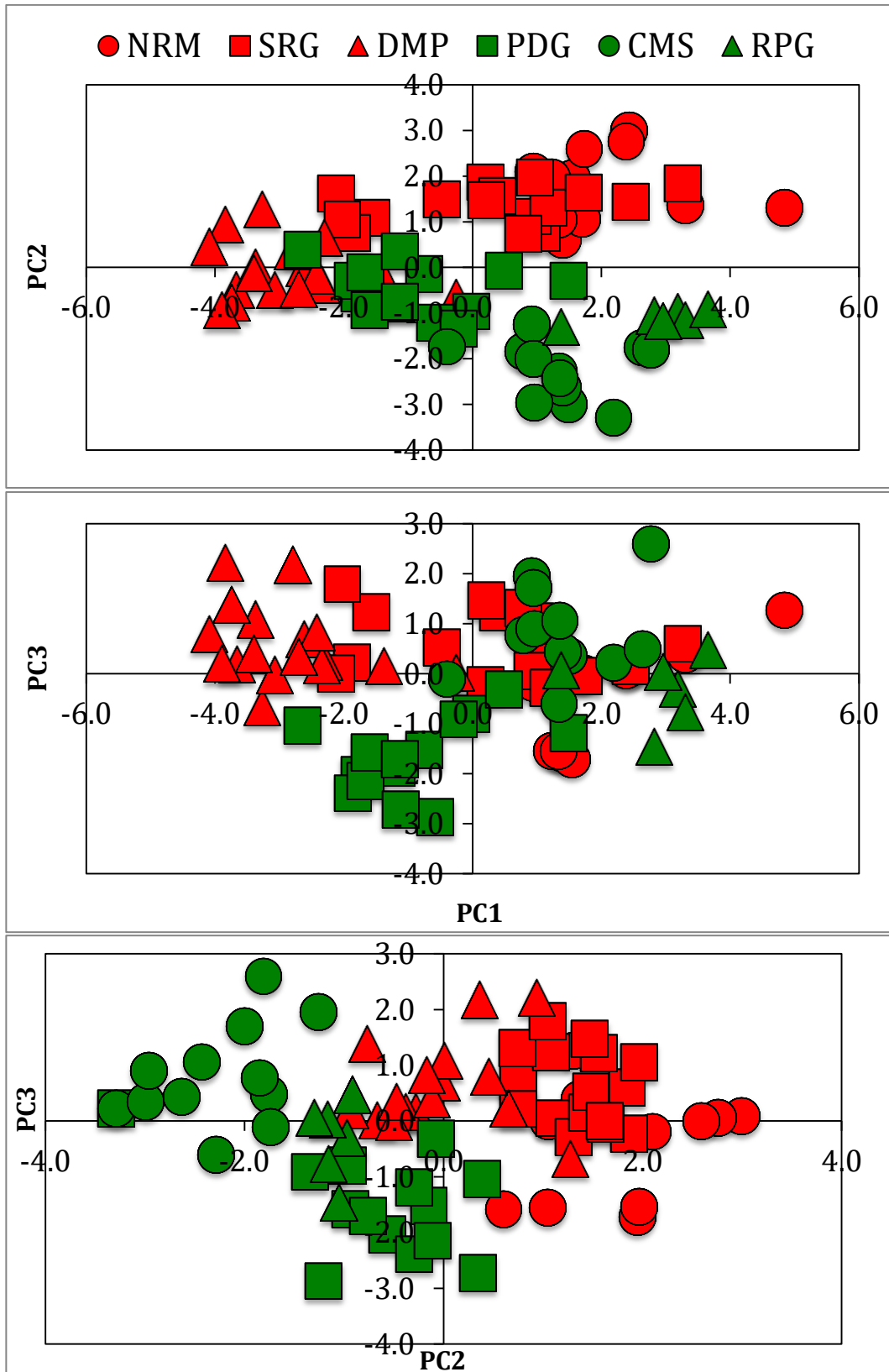


Figure 4.9. Plot of Principal Component scores for the first three axes for six populations of *M. albus* based on 12 ratio-transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.10. Principal Component analysis of 12 size-transformed morphometric variables for six populations of *M. albus*. Variable loadings are given for the first three components as correlations

Variables/ PCA statistics	PC1	PC2	PC3
STO	0.837	0.042	-0.241
STGA	0.865	0.156	-0.175
HL	0.864	0.061	-0.313
SL	0.587	-0.599	-0.084
STE	0.628	-0.486	-0.369
GL	0.793	0.299	-0.099
ED	0.384	0.719	0.248
WGA	0.318	-0.495	0.632
HD	0.847	0.281	0.139
HW	0.705	0.111	0.358
GBD	0.663	-0.215	0.539
GWB	0.379	-0.255	-0.109
Eigen value	5.608	1.679	1.250
% of variance	46.734	13.993	10.420
Cumulative % variance	46.734	60.727	71.147

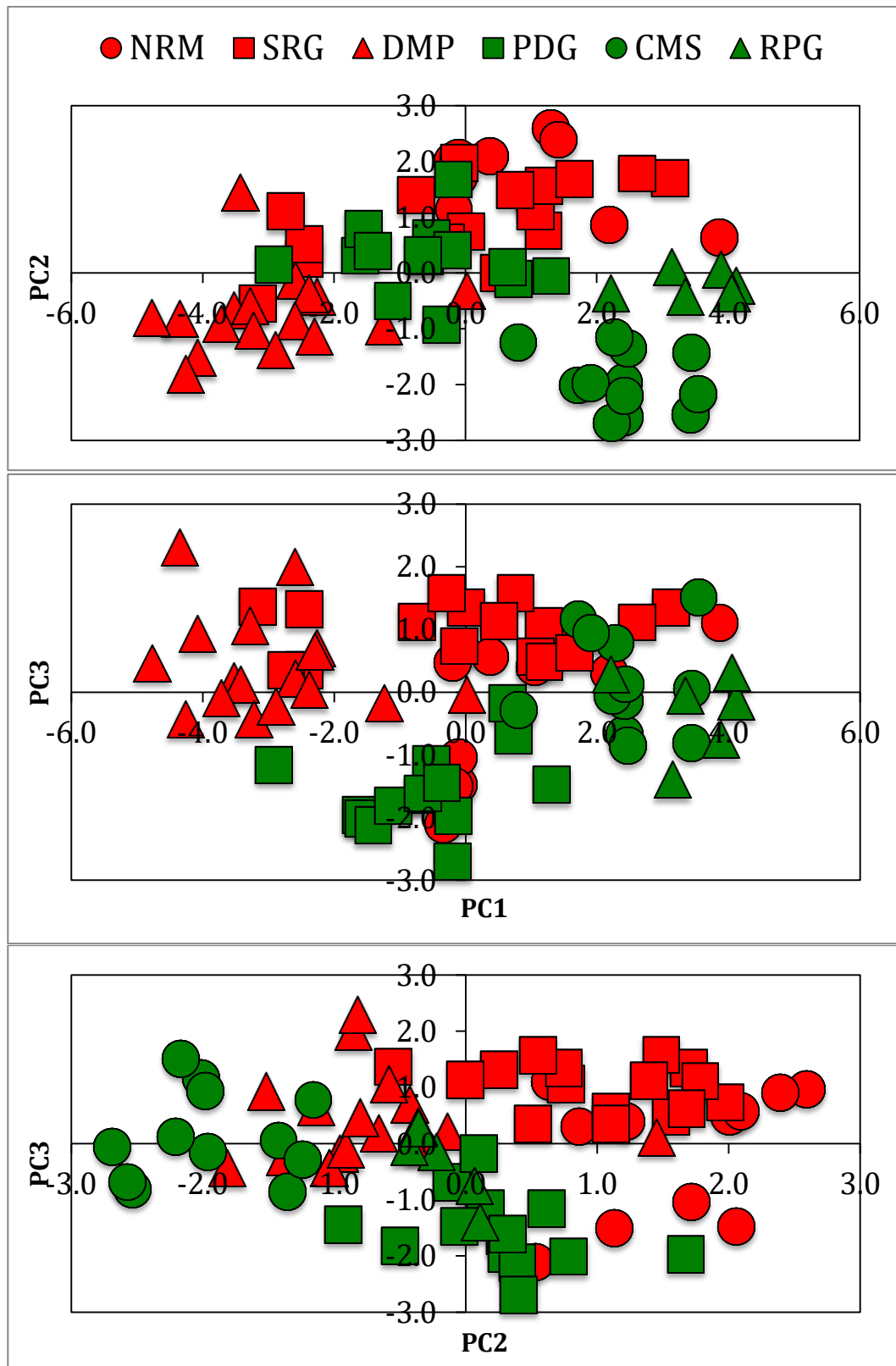


Figure 4.10. Plot of Principal Component scores for the first three axes for six populations of *M. albus* based on 12 size transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.11. Discriminant Function analysis of 13 log₁₀ transformed morphometric variables for six *M. albus* populations. The variable loadings are given for the first three functions as correlations

Variables/ DFA statistics	DF1	DF2	DF3
TL	0.766	-0.502	-0.151
STO	0.553	-0.648	-0.262
STGA	0.386	-0.766	-0.241
HL	0.563	-0.660	-0.295
SL	0.690	-0.652	-0.085
STE	0.648	-0.678	-0.262
GL	0.163	-0.803	-0.241
ED	0.077	-0.571	-0.050
WGA	0.682	-0.506	0.153
HD	-0.078	-0.916	0.079
HW	0.375	-0.692	-0.053
GBD	0.467	-0.733	0.286
GWB	0.637	-0.473	-0.192
Eigen value	18.814	5.124	2.788
% of variance	66.922	18.226	9.919
Cumulative % variance	66.922	85.149	95.067

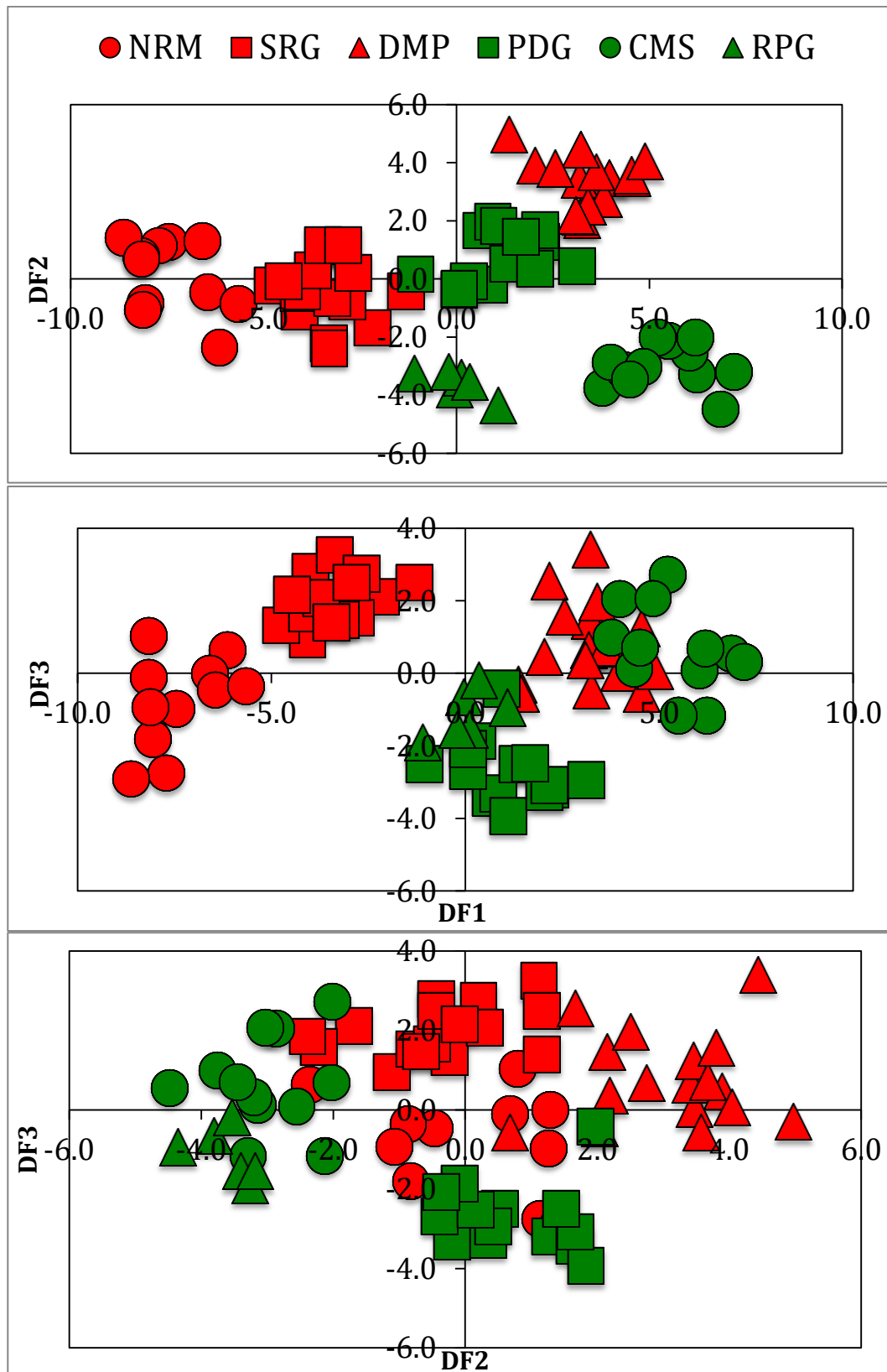


Figure 4.11. Plot of Discriminant Function scores for the first three axes for six populations of *M. albus* based on 13 log₁₀ transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.12. Discriminant function analysis of 12 ratio transformed morphometric characters for six populations of *M. albus*. Variable loadings are given for the first three functions as correlations

Variables	DF1	DF2	DF3
DFA statistics			
STO/TL	0.106	0.676	0.246
STGA/TL	0.382	0.704	0.212
HL/TL	0.128	0.797	0.444
SL/TL	-0.393	0.749	-0.160
STE/TL	-0.284	0.713	0.138
GL/TL	0.637	0.551	0.209
ED/TL	0.881	-0.164	-0.015
WGA/TL	-0.146	0.224	-0.552
HD/TL	0.747	0.558	-0.189
HW/TL	0.437	0.403	-0.133
GBD/TL	0.220	0.540	-0.674
GWB/TL	-0.014	0.104	0.066
Eigen value	9.993	4.697	2.550
% of variance	55.078	25.886	14.055
Cumulative % variance	55.078	80.964	95.019

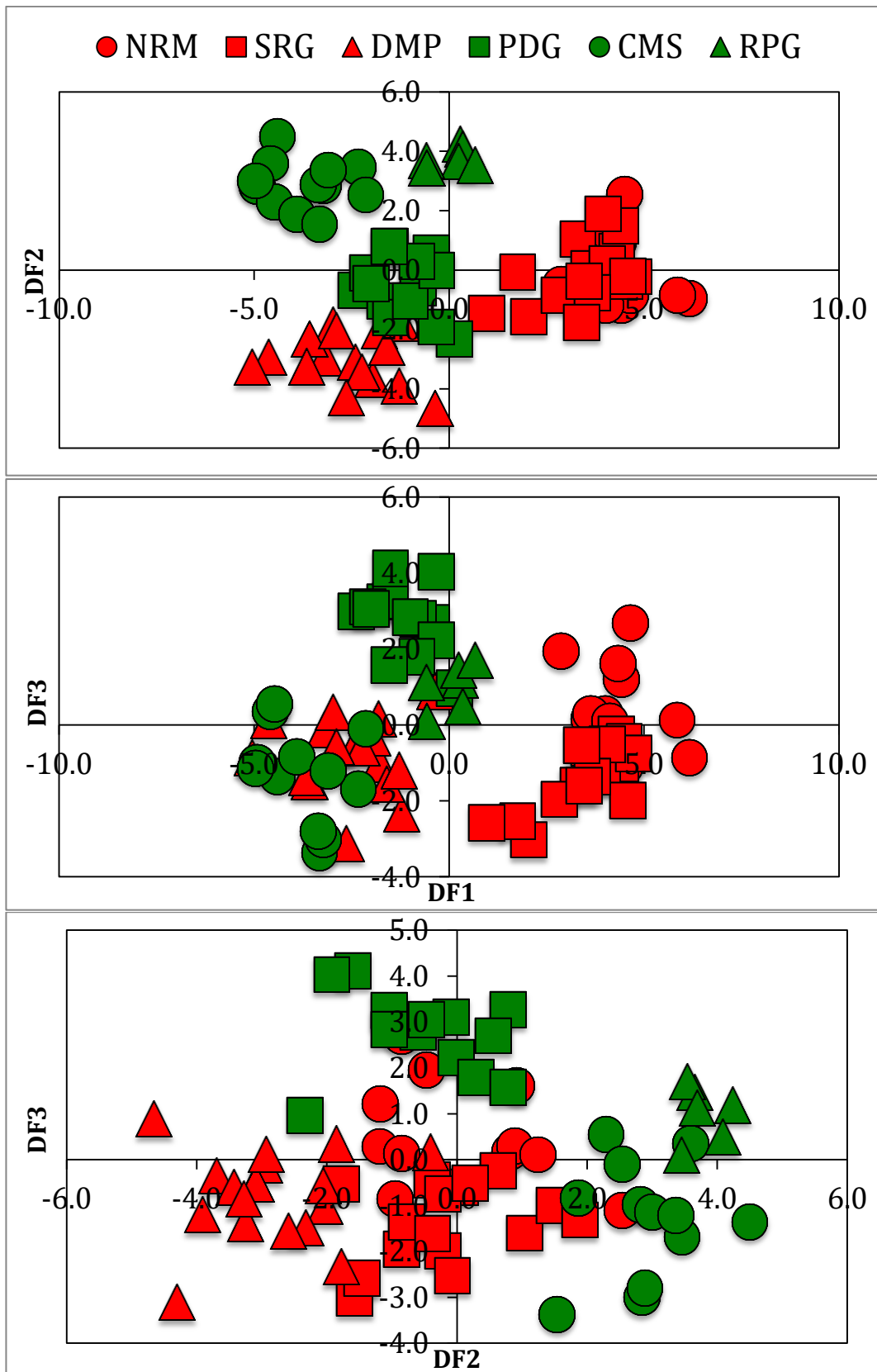


Figure 4.12. Plot of Discriminant Function scores for the first three axes for six populations of *M. albus* based on 12 ratio transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.13. Discriminant function analysis of 12 size transformed morphometric variables for six populations of *M. albus*. Variable loadings are given for the first three functions as correlations

Variables/ DFA statistics	DF1	DF2	DF3
STO	0.657	0.007	0.328
STGA	0.738	-0.306	0.291
HL	0.755	-0.089	0.499
SL	0.772	0.375	-0.077
STE	0.735	0.357	0.236
GL	0.662	-0.460	0.278
ED	0.072	-0.740	0.021
WGA	0.372	0.291	-0.491
HD	0.771	-0.545	-0.086
HW	0.535	-0.209	-0.072
GBD	0.674	-0.072	-0.581
GWB	0.235	0.205	0.110
Eigen value	5.131	4.716	2.741
% of variance	38.516	35.402	20.573
Cumulative % variance	38.516	73.918	94.491

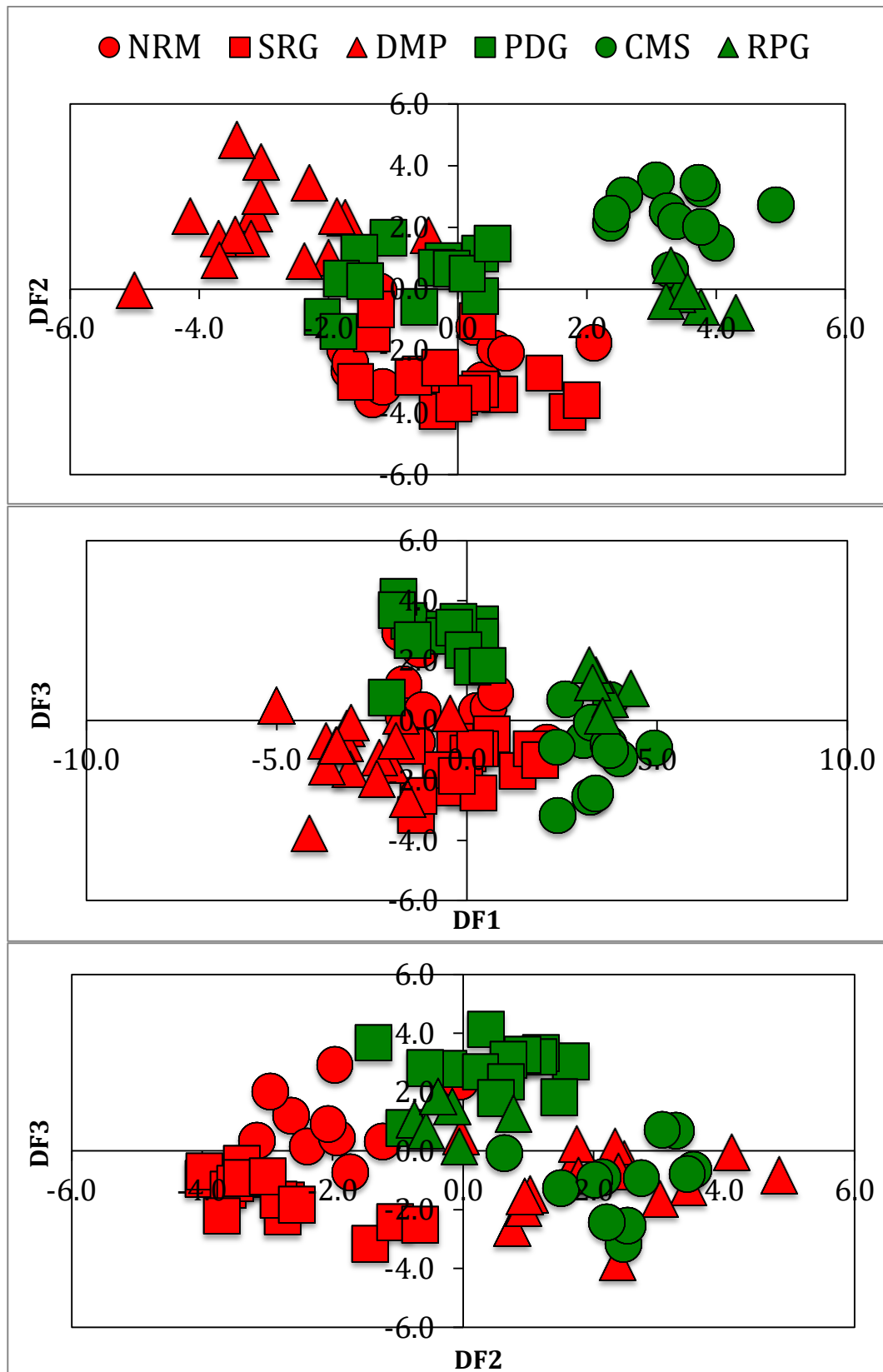


Figure 4.13. Plot of Discriminant Function scores for the first three axes for six populations of *M. albus* based on 12 size transformed morphometric variables. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.14. Two group Discriminant Function analyses of 13 log transformed and 12 ratio transformed morphometric characters for seven populations of *M. albus*. Variable loadings are given for the first three functions as correlations

Log10		Ratio	
Variables/ DFA statistics	DF1	Variables/ DFA statistics	DF1
TL	0.827		
STO	0.834	STO/TL	0.436
STGA	0.828	STGA/TL	0.366
HL	0.882	HL/TL	0.640
SL	0.822	SL/TL	0.647
STE	0.903	STE/TL	0.663
GL	0.710	GL/TL	0.060
ED	0.438	ED/TL	-0.644
WGA	0.579	WGA/TL	-0.068
HD	0.465	HD/TL	-0.156
HW	0.629	HW/TL	-0.117
GBD	0.516	GBD/TL	-0.116
GWB	0.750	GWB/TL	0.090
Eigen value	4.509	Eigen value	3.929
% of variance	100	% of variance	100

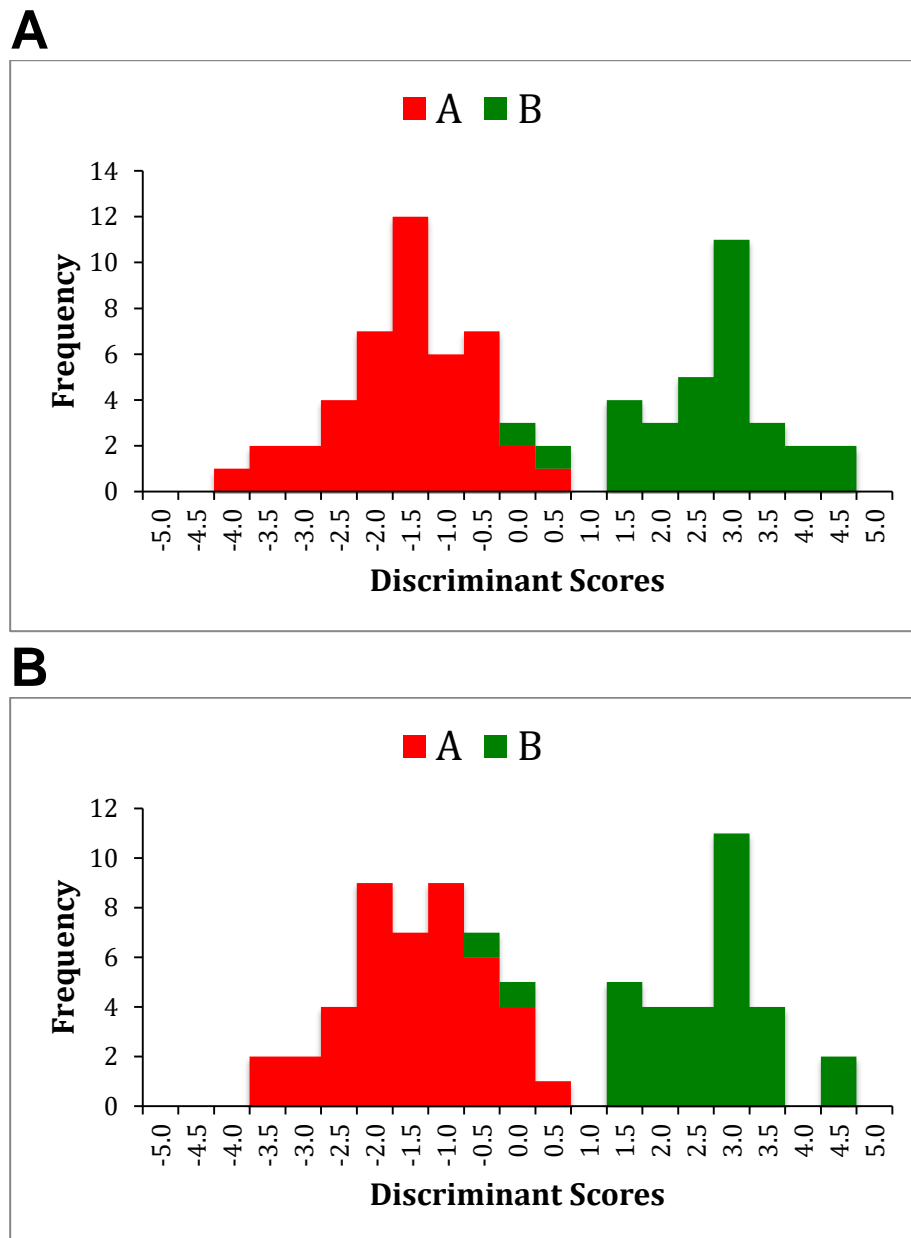


Figure 4.14. Histograms of discriminant scores from two group Discriminant Function analyses of 13 log transformed (A) and 12 ratio transformed (B) morphometric characters for seven populations of *M. albus*. Red and green colours refer to the genetically defined haplotypes A and B

Table 4.15. Classification results of discriminant function analysis using log and ratio transformed morphometric variable from seven populations of *M.albus* based on the grouping of the two haplotypes

Data set	<i>M. albus</i>	Predicted count		Total (correct)
		Haplotype A	Haplotype B	
Log10	Haplotype A	56	0	56 (100.0)
	Haplotype B	0	32	32 (100.0)
Ratio	Haplotype A	56	0	56 (100.0)
	Haplotype B	1	31	32 (96.9)

Table 4.16. The two-group Discriminant Function analyses of untransformed morphometric and meristic characters for 88 *M.albus* from seven populations of *M. albus*. The first two DFAs were based on 14 variables, including vertebrae number and the third was based on the best discriminating variables identified in the analysis of all variables in the untransformed data set. Variable loadings are given for the first three functions as correlations

Variables/ DFA statistics	DF1 Log10	DF1 Untrans- formed	DF1 Untrans - formed
TL	0.683	0.730	
STO	0.672	0.730	
STGA	0.648	0.698	
HL	0.688	0.758	
SL	0.657	0.735	
STE	0.768	0.826	0.865
G	0.638	0.659	
E	0.369	0.400	
A	0.540	0.586	
HD	0.445	0.442	
HW	0.537	0.585	
GBD	0.500	0.544	
GWB	0.605	0.668	
Vertebrae	0.903	0.904	0.947
Eigen value	5.561	6.006	3.585
% of variance	100	100	100

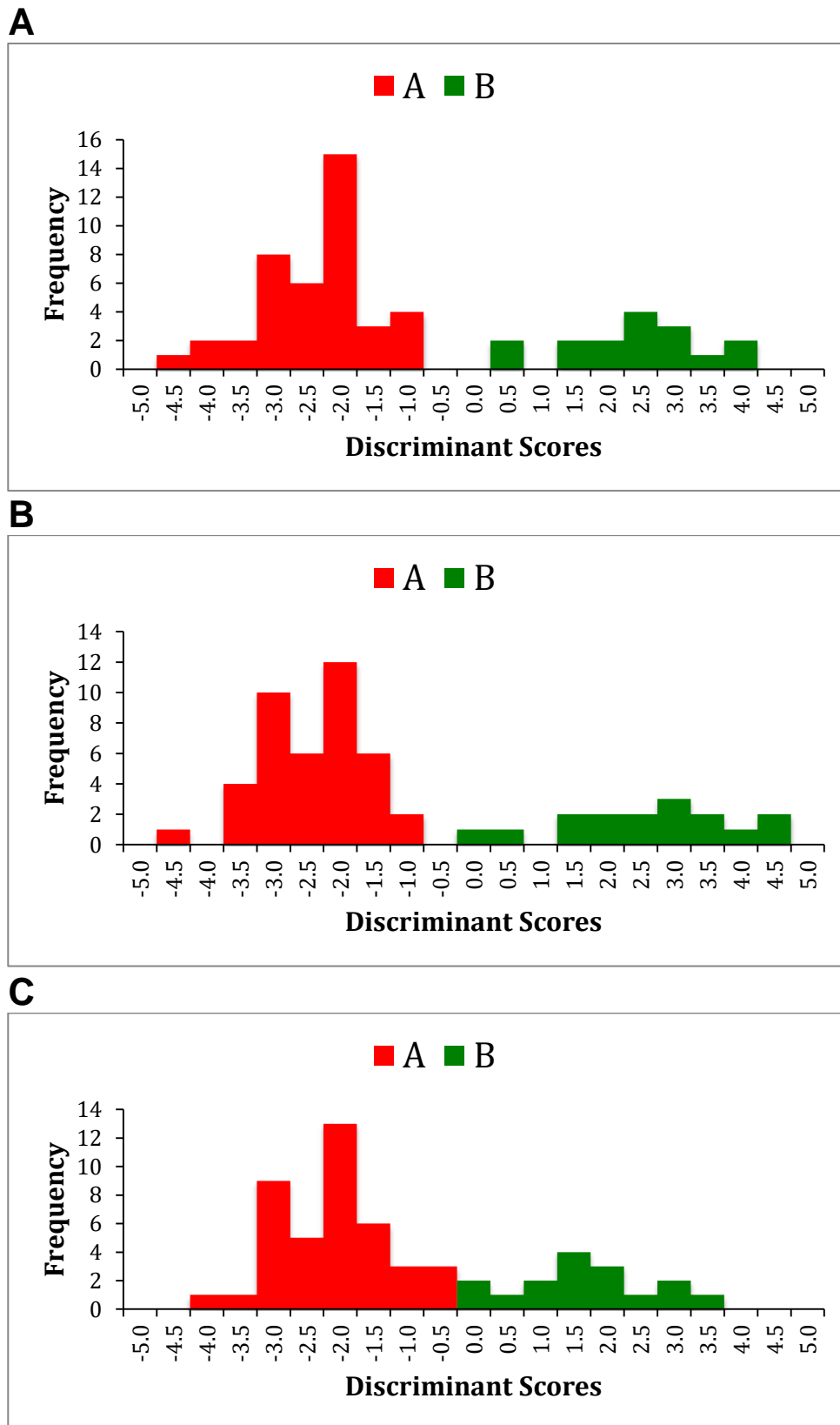


Figure 4.15. Histograms of discriminant scores from two group Discriminant Function analyses of 14 log transformed (A) and untransformed (B) morphological variables, and for the two best discriminating variables (C) for seven populations of *M. albus*. Red and green colours refer to the genetically defined haplotypes A and B

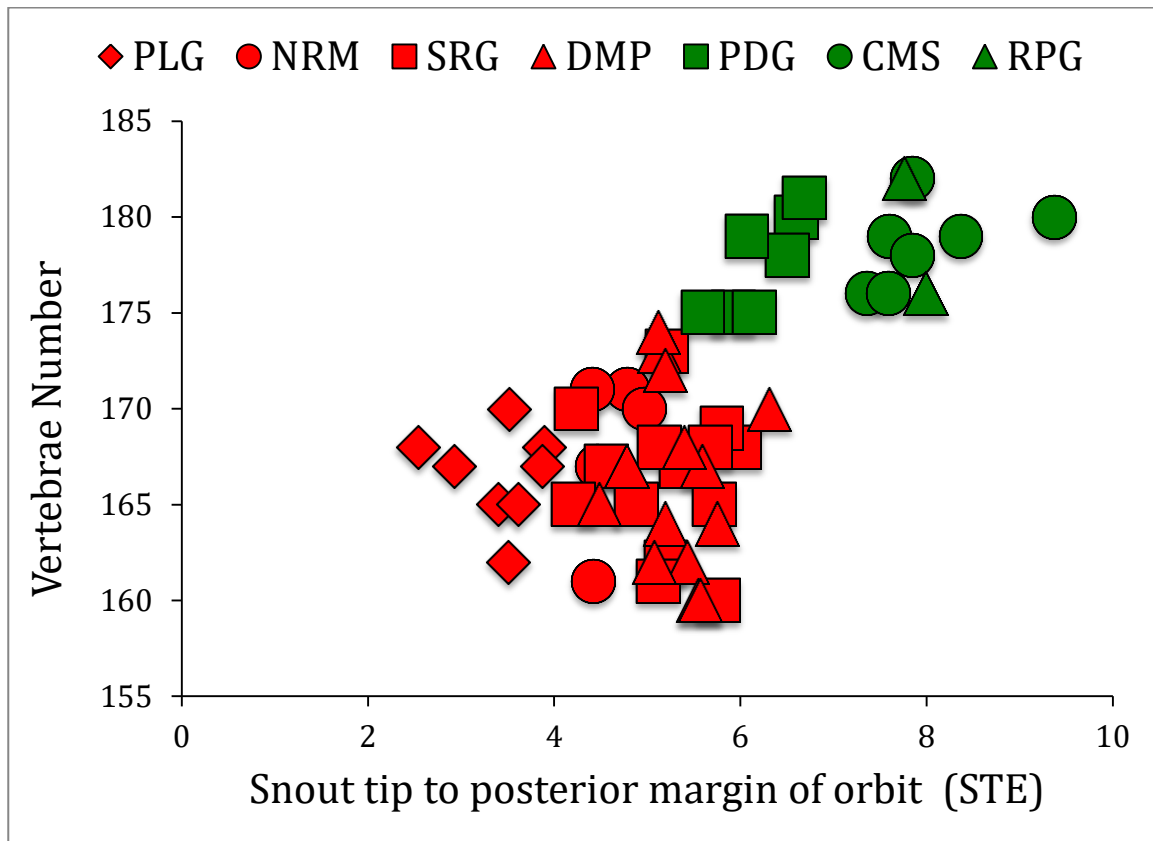


Figure 4.16. Bivariate plot of the two best discriminating variables, snout tip to posterior margin of orbit and vertebrae number identified from the DFA of 14 untransformed variables (see Table 4.16). Red and green colours refer to the genetically defined haplotypes A and B

CHAPTER 5

General Discussion

The wide distribution, economic significance, invasive potential, variable morphology, and taxonomic disputation make swamp eels an important, interesting, challenging and, at times controversial species for study (Collins *et al.*, 2002; Matsumoto *et al.*, 2010). The advancement of knowledge of the taxonomy and understanding of the distribution of genetic variation in “*M. albus*” is of utmost importance given the species significance not only with respect to food security and household incomes, but also for effective conservation management in many developing countries including Indonesia.

The series of related studies presented in this thesis, involving the acquisition of different kinds of molecular genetic information and morphological data, contributes significant new knowledge about the swamp eel, *M. albus*. Indeed this thesis represents the most comprehensive study of swamp eel molecular genetics so far undertaken, contributing to the understanding of the taxonomy and evolution of the species. This study is also likely the most comprehensive molecular genetic investigation of any Indonesian freshwater fish species so far completed.

Following from the primary aims of this study, the following section summarises and discusses the principal findings and conclusions, together with recommendations for future research.

5.1. Global taxonomy and evolution of *Monopterus albus* complex

Thirty one populations, obtained from comprehensive collections from Indonesia and two other Asian countries (Vietnam and Taiwan), have been used to reconstruct genealogical relationships using multi gene regions. Data from two different mtDNA fragments (COI and 2 fragments of 16S) and two nuclear gene regions (RAG1 and 1st intron of S7) have been utilised. Sequence data with up to 3,366 bp were analysed using specimens as detailed in Chapters 2 and 3. A remarkably consistent picture emerged from the different nucleotide sequences and genealogical analyses. Firstly, DNA barcoding using COI mtDNA has provided strong evidence that *M. albus* contains cryptic species. In this investigation, the Indonesian *M. albus* fell into two distinct groups, which are genetically distinct from each other and distinct from other forms of *M. albus* from two other Asian countries (Vietnam and Taiwan). Secondly, 16S mtDNA sequences combined with the data from previous studies supported that *M. albus* is a species complex. The results also revealed a consistent picture of Indonesian *M. albus* being composed of two distinct groups, one clade unique to Indonesia and the other clade containing *M. albus* samples from Southeast Asian countries but distinct from *M. albus sensu stricto* from Japan and two other distinct forms (Figure 3.10). Thus I am confident that there are at least five valid species within what should now be described as the *M. albus* species complex. These findings have important implications for the taxonomy of *M. albus*, which is highly confused with most authors considering *M. albus* to be one species (Rosen and Greenwood, 1976; but see Kottelat, 2013). However my findings are consistent with more recent genetics, but less comprehensive studies, which conclude that *Monopterus* is represented by multiple species (Collins *et al.*, 2002;

Matsumoto *et al.*, 2010) in western East Asia and Southeast Asia. Finding complexes of cryptic species or lineages in freshwater fish groups is becoming more common in eastern and Southeast Asian countries with the increasing number of geographically comprehensive molecular genetic and phylogenetic studies (Nguyen *et al.*, 2008; Bohlen *et al.*, 2011; Ratmuangkhwang *et al.*, 2014; Barman *et al.*, 2014).

Lastly, the use of sequences from multigene regions and additional sampling allowed the construction of robust phylogenetic hypotheses and therefore insights into the evolutionary history of the *M. albus* species complex. The overall phylogenetic relationships indicate that the highest diversity and oldest lineages within the complex are from the western part East Asia (China, Japan, Taiwan, North Vietnam) and the younger species discovered in this study to be located in Southeast Asia (Indonesia, Malaysia, and South Vietnam as detailed in Figure 3.10).

This pattern of evolution would have required migration south and dispersal across present-day sea barriers between Malaysia and Indonesia and between major Indonesian islands (e.g. Sumatra and Java). Such migration would have been possible during periods of lower sea levels (Hall, 2002) when these land masses and island were connected and creating the ancient land mass of what is referred to as Sunderland (Figure 5.1). Dispersal within and between present-day countries would have been facilitated by the presence of large ancient rivers (Voris, 2000) as illustrated in Figure 5.2. The Sundaland land mass would have appeared several times during the last ice age (110,000 to 12,000 years ago) when sea levels were lower and would have also been present during earlier periods extending in the Miocene (Hall, 2012). The presence of Sundaland has been used to explain the otherwise unusual and disjunct distribution of freshwater fish species in Southeast Asia including

arowana (*Scleropages formosus*) (Tang *et al.*, 2004), *Barbodes gonionotus* (McConnell, 2004), mahseer (*Tor* species) (Nguyen *et al.*, 2008) and loaches (*Pangio*) (Bohlen *et al.*, 2011).

5.2. Genetic variation of the Indonesian *Monopterus albus* complex

The swamp eel is one of the most popular fish in Indonesia due to its reputation as valuable food. The fish is a common inhabitant of rice fields. However, they have not been cultured intensively yet. Therefore, it is important that management of stocks and genetic improvement programs utilize the best genetic information available to ensure benefits accrue to small scale farmers who depends on this fish species for their livelihoods and food security, and to ensure the wild conservation of the species indigenous to Indonesia occurring on the islands of Java and several adjacent islands to the east.

While swamp eels have been studied using a variety of molecular markers in several countries, detailed knowledge of genetic variation of Indonesian swamp eels using microsatellite markers have not previously been conducted. The previous studies mostly focused on Chinese swamp eel populations from a relatively restricted geographic distribution (Cai *et al.*, 2008; Cai *et al.*, 2012). This thesis presents a comprehensive study on swamp eel genetic diversity in Indonesia, using microsatellite markers that have been previously used for swamp eels (Li *et al.*, 2007; Lei *et al.*, 2011) and other markers, including direct DNA sequencing of mitochondrial and nuclear genes and the DGGE procedures that have not been applied to this species before (Chapters 2 and 3).

In general, the mitochondrial DNA and nuclear DNA are consistent and revealed Indonesian swamp eel populations fall into two genetically distinct groups, being distinct from other forms of swamp eel from several Asian countries. In addition, these studies provided evidence of mixing of stocks of the two Indonesian species. An important finding was the haplotypes and alleles characterising the two forms of Indonesian swamp eel were correlated in the unmixed and mixed populations providing strong evidence meeting the requirement of the biological species concept (Mayr, 2000). The findings of a small number of hybrid individual in the mixed populations needs further research relating to reproductive behaviour of each species and their ecology including niche specialisation.

The examination of variation in mitochondrial and nuclear markers documented high level of genetic diversity in both forms of Indonesian swamp eels. High level of genetic diversity of the fish is useful for aquaculture in the future. Therefore, there is a clear need to preserve and take advantage of genetic variation of the fish to improve breeding programs and hatchery productions which can potentially give beneficial impacts for economic improvement of small-scale fish farmers and integrated agriculture in Indonesia.

Finding swamp eels on either side of the Wallacea line, and considering their documented introduction to the United States on a minimum of three occasions (Collins *et al.*, 2002), together with their known hardiness and adaptability and commercial importance, means that the species could become a serious pest and a significant invasive species (Nico *et al.*, 2011). In addition to being a possible vector for parasites, translocated populations could displace local species, effect local freshwater communities through increased predation rates and complicate the interpretation of geographic patterns of genetic variability.

5.3. Morphological variation and taxonomy

Integrative taxonomy which includes molecular, morphological, ecological and geographical data has recently developed as one of the most promising approaches to species delimitation in taxonomically difficult groups (Baco and Cairns, 2012; Novo *et al.*, 2012; Silva *et al.* 2013). The *Monopterus albus* complex represents an interesting taxonomical study case, as this group of fish is defined by a lack of informative morphological characters. This research tested the effectiveness of an integrative taxonomy approach through delimiting two species within the “*M. albus*” species complex in Indonesia

While identifying clear morphometric differences among the populations of the two Indonesian populations was challenging, univariate, bivariate and multivariate morphological analyses did provide support for phenotypic divergence between the species. Morphometric studies of variation among fish populations and species are confronted by difficulties in dealing with variation in size, phenotypic plasticity and local adaptation, which are amplified in freshwater species with fragmented distributions. Working with swamp eels is further complicated by their conservative morphology, the absence of scales (i.e. the body lacks rigidity compared with scaled fish), the absence of fins, and a long tail that is readily injured or damaged.

Despite these problems the morphological analyses were effective in providing support for the genetically defined groups. Further, characters potentially useful for taxonomic identification were identified and the two most useful were highlighted (vertebrae number and snout tip to posterior margin of orbit). In addition, the multiple discriminant function analysis was successful in

distinguishing individuals from the two genetic forms and the classification function could be useful for the taxonomic identification of unknown individuals. One of the interesting outcomes of the morphological analyses was the lack of effectiveness of the transformations used to reduce the effects of size variation (Atchley and Anderson, 1978; Bookstein, 1982). The most useful approach was using the untransformed or log transformed data and PCA and DFA, which provided the best separation of the two species and were effective in partitioning size and shape related variation (Thorpe, 1976; Albrecht, 1980).

5.4. Further studies

Even though this study has significantly enhanced the understanding of the taxonomic and phylogenetic relationships and genetic diversity of the *M. albus* species complex, especially in relation to Indonesian stocks, there are still many areas that need further research. These include aspects relating to the geographic extension of this study, and details of the life history and ecology of swamp eel species that can now be conducted more effectively as it is clear that the *M. albus* complex contains several biological species.

The finding of a new species endemic to Indonesia will likely have important implications. Local fishermen and farmers already recognise that two forms exist, with one growing larger (i.e. haplotype B), and hence being a common target as broodstock for ponds especially in western Sumatra, and the smaller form (i.e. haplotype A) having a favoured more savory taste, meaning it is often sold in preference in markets (especially eastern Java). This study provides a robust scientific basis for further exploring biological attributes of the two

species in the region relating to commercial exploitation, cultural use and conservation.

In general, proper genetic, phylogenetic, biogeographical and ecological studies can not be conducted without a stable species level taxonomy and tools that allow identification of species. It is thus extremely important that extensive morphological and taxonomic studies similar to those carried out by Tancioni *et al.* (2013) in leuciscine fish and Silva *et al.* (2013) in species of *Tetragonopterus* be conducted. The limitation of this study is that morphological analyses only included a limited number of populations from Indonesia representing only two of a least five species of the complex. Any new studies need to use molecular tools to ensure a priori identification of specimens of swamp eels subject to morphological and taxonomic analysis. Such studies need to relate findings to existing taxonomic treatments, thus relevant type specimens need to be examined (if they have not been lost) and fish from type localities (where known) should be collected where possible. This is required to ensure the rules of nomenclature are followed and the appropriate and available species names are applied (Blackwelder, 1967; Schuh, 2000), especially considering the extensive list of synonyms for *M. albus*. Where no valid species name is available or appropriate then the taxon should be described as new, which will be likely most be the case for species corresponding to haplotype A from Indonesia. In addition, further molecular studies should be conducted with more comprehensive sampling, as *M. albus* has an extremely wide distribution (Figure 3.10) and this and earlier studies have still yet to fully sample the likely molecular and genetic variation with the species complex. Priority countries for additional sampling of *Monopterus* include Thailand, Myanmar, Cambodia, Laos, Philippines and some parts of Malaysia (Sarawak) and central China. Only with further studies

encompassing these countries will we have a full understanding of the diversity and evolution of the swamp eel.

New genetic and taxonomic studies of *M.albus* can take advantage of the DGGE or RFLP technique used in this study for rapid and relatively inexpensive assignment of individuals to putative species. In addition, an extension of the microsatellite investigation would be useful for taxonomy and populations genetics and also for constructing a genetic linkage map for *M. albus* and determine the potential for the identification of Quantitative Trait Loci (QTL) and Marker Assisted Selection (MAS) as means enhancing aquaculture productivity of this species (Baranski *et al.*, 2006; Baranski *et al.*, 2008; Ozaki *et al.*, 2013; Yue, 2013).

The finding of admixture of two forms of Indonesian *M. albus* at several locations and the the presence of only a limited number of hybrids indicates that a study of the reproductive behaviour of the fish would be of significant interest to determine how the species maintains reproductive isolation. In addition, genomic studies have identified several genes that contribute to sex-determination and sexual development in *M. albus* (Liu and Zhou, 2001; Zhou *et al.*, 2002; Lu *et al.*, 2003; Zhou *et al.*, 2003; Yu *et al.*, 2003; Wang *et al.*, 2003; Xia *et al.*, 2004; Huang *et al.*, 2005; Ye *et al.*, 2007; Yu *et al.*, 2008; Liu *et al.*, 2009) and these studies should be extended to Indonesian *M. albus* populations. *Monopterus albus* could become an increasingly important model organism for a range of fundamental research investigations as it is hardy, grows quickly and is readily maintained in captivity. This again emphasises the need to extend this study to develop a comprehensive knowledge base on the diversity of this important and interesting group of fishes through the distribution, underpinned by a stable and reliable taxonomy.

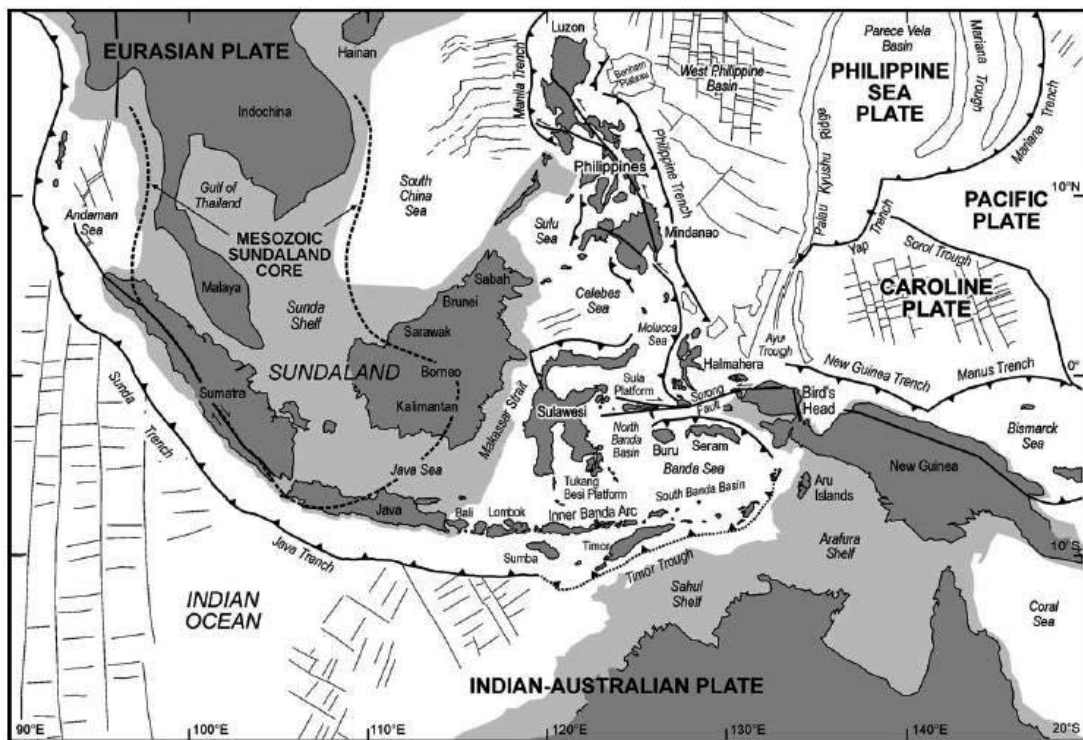


Figure 5.1. Principal geographic features of the region covered in Southeast Asia reconstruction. The light shade areas are the continental shelves of Eurasia and Australia drawn at the 200 m isobath (Hall, 2002)



Figure 5.2. Map of Pleistocene sea levels in Southeast Asia and Austral-Asia. Light grey shading indicates -75 m sea level contour, dark lines show freshwater catchments at this time (Voris, 2000)

Appendices

Appendix 1. Allele frequency of *M. albus* haplotype A (18 populations)

Locus	Allele	NRM	TWG	PKY	PLG	GSK	NGT	SRG	DMP	WLK	SKP	KYR	PJK	CPK	LMJ	NGR-A	PLB-A	GMP-A	BRS-A
MaI01	214	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.025	0.150	-
	217	0.375	0.375	0.200	0.150	0.200	0.250	0.200	0.350	0.380	0.325	0.300	-	0.300	0.250	0.207	0.350	0.150	0.167
	219	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.050	-	-
	221	0.525	0.518	0.350	0.800	0.800	0.750	0.800	0.650	0.620	0.675	0.700	0.725	0.700	0.500	0.732	0.425	0.525	0.429
	223	0.100	0.107	0.450	0.050	-	-	-	-	-	-	-	0.275	-	0.250	0.061	0.150	0.175	0.357
	225	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MaI07	184	0.350	0.179	0.300	-	-	-	-	-	-	-	-	-	-	-	0.513	0.225	0.312	-
	186	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.025	-
	188	0.275	0.411	0.500	0.625	0.650	0.550	0.500	1.000	0.300	0.300	0.425	0.400	0.100	0.525	0.244	0.400	-	0.381
	190	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.244	0.100	0.075	0.143
	192	0.300	0.161	-	-	-	-	0.175	-	0.320	0.450	0.325	0.400	-	-	0.378	0.175	0.413	0.190
	194	-	0.303	-	-	0.100	0.175	0.150	-	0.160	-	0.125	-	-	-	-	-	0.175	-
	196	0.075	0.107	0.200	0.375	0.250	0.275	0.175	-	0.220	0.250	0.125	0.200	-	0.475	0.061	0.100	-	0.286
MaI007	331	-	0.429	0.475	0.500	0.650	0.600	0.600	0.600	0.500	0.450	0.525	0.450	0.525	0.550	0.390	0.375	0.225	0.381
	333	-	-	-	-	-	-	-	-	-	0.225	0.150	0.175	0.175	-	-	-	0.087	-
	335	0.275	0.286	-	0.500	0.350	0.400	0.400	-	0.500	0.325	0.325	0.375	0.300	0.450	0.427	0.475	0.312	0.381
	337	0.250	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.188	-
	341	-	-	0.525	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	343	0.275	0.196	-	-	-	-	-	0.400	-	-	-	-	-	-	-	-	0.188	-
	349	0.125	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	351	0.075	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	353	-	0.089	-	-	-	-	-	-	-	-	-	-	-	-	0.122	0.100	-	0.238
	357	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.050	-	-
359	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.061	-	-	-	

Appendix 1. Continued

Locus	NRM	TWG	PKY	PLG	GSK	NGT	SRG	DMP	WLK	SKP	KYR	PJK	CPK	LMJ	NGR-A	PLB-A	GMP-A	BRS-A
Mal008	336	0.025	-	-	-	-	-	-	-	0.075	0.100	0.250	0.225	-	-	-	-	-
	338	0.300	-	-	-	-	0.150	0.325	0.220	0.200	0.175	-	-	0.200	0.219	-	0.250	0.357
	340	0.075	-	-	-	-	-	0.500	0.300	0.450	0.500	0.400	-	0.175	-	-	0.050	-
	342	-	-	0.125	0.100	0.100	0.100	-	-	-	-	-	-	0.050	-	0.175	-	-
	344	-	0.304	-	-	-	0.100	0.175	-	-	-	-	-	0.175	-	-	-	-
	346	0.050	0.393	0.300	-	-	-	-	-	0.175	-	-	-	0.100	0.183	-	0.200	0.238
	348	-	-	0.475	0.050	0.100	-	-	-	-	-	0.150	0.450	-	0.171	-	0.075	0.214
	350	0.125	0.303	0.225	0.300	0.050	0.200	-	0.180	-	0.225	-	-	-	-	-	0.012	0.191
	352	-	-	0.250	0.300	0.275	0.200	-	0.180	0.100	-	0.200	0.325	0.225	0.110	0.275	0.063	-
	354	-	-	-	0.175	0.275	0.250	-	0.180	-	-	-	-	-	-	0.275	0.050	-
	356	-	-	-	0.050	-	-	-	-	-	-	-	-	-	-	-	-	-
	358	0.075	-	-	0.150	0.025	-	-	-	-	-	-	-	0.075	-	-	0.063	-
	360	0.300	-	-	0.175	0.175	-	-	-	-	-	-	-	-	0.195	-	0.112	-
	362	0.050	-	-	-	-	-	-	0.060	-	-	-	-	-	0.122	-	0.125	-
	364	-	-	-	-	-	-	-	0.060	-	-	-	-	-	-	-	-	-
Mal013	212	0.275	0.339	0.325	0.300	0.400	0.350	0.425	0.320	-	0.225	0.175	0.350	0.350	0.378	0.475	0.300	0.119
	216	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.075	0.143
	222	0.725	0.536	0.525	0.600	0.600	0.650	0.575	0.520	-	0.575	0.625	0.650	0.450	0.427	0.525	0.462	0.738
	224	-	0.125	0.150	-	-	-	-	0.160	-	0.200	0.200	-	0.200	0.195	-	0.163	-
	228	-	-	-	0.100	-	-	-	-	-	-	-	-	-	-	-	-	-

The population codes are given in Table 2.1.

Appendix 2. Allele frequency of *M. albus* haplotype B (6 populations)

Locus	Allele	KMR	RPG	NGR-B	PLB-B	GMP-B	BRS-B
<i>Mal01</i>	204	0.080	-	0.300	0.210	-	-
	214	0.780	0.432	0.575	0.790	0.608	0.408
	217	-	-	-	-	0.270	0.066
	219	0.140	0.25	-	-	0.027	0.171
	221	-	-	0.075	-	0.068	0.355
	223	-	-	0.050	-	0.027	-
	225	-	0.318	-	-	-	-
<i>Mal07</i>	184	-	-	-	-	0.013	-
	186	0.140	-	0.400	-	0.013	0.197
	190	0.860	0.100	0.575	0.952	0.974	0.750
	192	-	-	0.025	0.048	-	0.053
<i>Mal007</i>	327	-	0.409	-	-	-	-
	331	-	-	-	-	-	0.053
	335	-	-	-	-	-	0.079
	341	-	-	0.025	-	-	-
	345	0.680	0.318	0.525	0.452	0.338	0.592
	349	-	-	0.050	-	-	0.118
	353	-	-	0.050	-	-	-
	355	-	-	0.350	-	0.162	-
	357	0.160	-	-	0.161	0.311	-
	359	0.160	0.182	-	0.242	-	-
	361	-	-	-	-	-	0.158
	363	-	0.091	-	0.145	0.189	-
<i>Mal008</i>	320	0.180	-	0.400	0.242	0.446	0.224
	324	0.660	-	0.325	0.419	0.311	0.395
	328	0.120	0.100	0.275	0.339	0.189	0.224
	330	0.040	-	-	-	0.054	-
	342	-	-	-	-	-	0.039
	344	-	-	-	-	-	0.066
	348	-	-	-	-	-	0.052
<i>Mal13</i>	207	0.360	0.091	0.400	0.436	0.446	0.355
	210	-	0.273	-	-	-	-
	212	-	-	-	0.032	-	-
	214	0.080	-	-	-	0.081	-
	216	0.560	0.204	0.450	0.387	0.392	0.500
	222	-	-	0.150	-	0.081	0.145
	226	-	0.432	-	0.145	-	-

The population codes are given in Table 2.1.

Appendix 3. Allele frequency of *M. albus* admixture populations (4 populations)

Locus	Allele	NGR	PLB	GMP	BRS
<i>Mal01</i>	204	0.098	0.127	-	-
	214	0.189	0.490	0.370	0.263
	217	0.139	0.137	0.208	0.102
	219	-	0.020	0.013	0.110
	221	0.517	0.167	0.305	0.381
	223	0.057	0.059	0.104	0.127
	225	-	-	-	0.017
<i>Mal07</i>	184	0.345	0.088	0.169	-
	186	0.131	-	0.019	0.127
	188	0.016	0.157	-	0.135
	190	0.205	0.618	0.507	0.534
	192	0.262	0.098	0.214	0.102
	196	0.041	0.039	0.091	0.102
<i>Mal007</i>	331	0.262	0.147	0.117	0.170
	333	-	-	0.046	-
	335	0.287	0.186	0.162	0.186
	337	-	-	0.097	-
	341	0.008	-	-	-
	343	-	-	0.097	-
	345	0.172	0.275	0.162	0.381
	349	0.017	-	-	0.076
	353	0.098	0.039	-	0.085
	355	0.115	-	0.078	-
	357	-	0.118	0.150	-
	359	0.041	0.147	-	-
	361	-	-	-	0.102
	363	-	0.088	0.091	-
<i>Mal008</i>	320	0.131	0.147	0.214	0.144
	324	0.107	0.255	0.149	0.254
	328	0.090	0.206	0.091	0.144
	330	-	-	0.026	-
	338	0.147	-	0.130	0.127
	340	-	-	0.026	-
	342	-	0.068	-	0.026
	344	-	-	0.104	0.127
	346	0.123	-	0.039	0.076
	348	0.115	-	0.007	0.102
	350	-	0.108	-	-
	352	-	0.108	0.032	-
	354	0.074	0.108	0.026	-
	358	-	-	0.032	-
	360	0.131	-	0.058	-
	362	0.082	-	0.065	-
<i>Mal13</i>	207	0.131	0.265	0.214	0.229
	212	0.254	0.206	0.156	0.042
	214	-	-	0.040	-
	216	0.148	0.235	0.227	0.373
	222	0.287	0.206	0.279	0.356
	224	0.131	-	0.084	-
	226	0.049	0.088	-	-

The population codes are given in Table 2.1.

Appendix 4. Sequence of 16S mitochondrial gene (primer 16Sar and 16Sbr) of *Ophisternon gutturale* and *Ophisternon sp*

>Seq1 [organism=*Ophisternon gutturale*] *Ophisternon gutturale*
voucher RCK 16S rRNA gene, partial sequence; mitochondrial
CCTGACTATACGTTCAACGGCCACGGTATCCTAACCGTGCAAAGGTAGCG
CAATCACTTGTCTTTTAAATTGGAGACCTGTATGAATGGTCTAACGAGAGC
TTGACTGTCTCCTCATTAAGTCAATAAAATTGATCTTCCCGTGCAGAAG
CGGGAATAAAAACATAAGACGAGAAGACCCTGTGGAGCTTTAGACACTAA
AGCAGCTCACAAACCTATAACAAGTTAACCTAACGATTCCTGCCCTAATGT
CTTCGGTTGGGGCGACCAAGGGGAATTAACAACCCCCATGCGGACCAGG
AAAACCTTTCCCAAACCTAAGAGCCACAACCTCTAACAAATAGAACTTCTA
ACCATTTACATAGACCCGGCACTGCCGATCTACGAACCAAGTTACCCAG
GGATAACAGCGCAATCCCCTTTTCAGAGCCCTTATCGACAAGGGGGTTTAC
GACCTCGATGTTGGATCAGGACACCCCAATGGTGCAGCCGCTA

>Seq2 [organism=*Ophisternon sp. n. 1*] *Ophisternon sp. n. 1* TA-2015
voucher AGR1 16S rRNA gene, partial sequence; mitochondrial
ACTGACTATACGTTCAACGGCCACGGTATCCTAACCATGCGAAGGTAGCG
CAATCACTTGTCTTTTAAATTGAAGACCTGTATGAATGGCCGAACGAGAGC
TTAACTGTCTCCTCTTTAAAGTCAATAAAATTGATCTTCCCGTGCAGAAG
CGGGAATAAAAACATAAGACGAGAAGACCCTGTGGAGCTTTAGACACTAA
AGCAGCTCACACAAACCTATAAAAGTATTTCCCTATGACCCCTGCCCTAA
TGTCTTTGGTTGGGGCGACCAAGGGGAATTAACAACCCCCATGTGGATC
AGGAACACCTTTCCCAAACCTAAGAGCCACAACCTCTAACAAATAGAACCT
CTAACCATTTAACAAGACCCGGCAACGCCGATCTACGAACCAAGTTACCC
CAGGGATAACAGCGCAATCCCCTTTTCAGAGCCCCTATCGACAAGGGGGTT
TACGACCTCGATGTTGGATCAGGACACCCCAATGGTGCAGCCGCTA

>Seq3 [organism=*Ophisternon sp. n. 1*] *Ophisternon sp. n. 1* TA-2015
voucher AGR2 16S rRNA gene, partial sequence; mitochondrial
ACTGACTATACGTTCAACGGCCACGGTATCCTAACCATGCGAAGGTAGCG
CAATCACTTGTCTTTTAAATTGAAGACCTGTATGAATGGCCGAACGAGAGC
TTAACTGTCTCCTCTTTAAAGTCAATAAAATTGATCTTCCCGTGCAGAAG
CGGGAATAAAAACATAAGACGAGAAGACCCTGTGGAGCTTTAGACACTAA
AGCAGCTCACACAAACCTATAAAAGTATTTCCCTATGACCCCTGCCCTAA
TGTCTTTGGTTGGGGCGACCAAGGGGAATTAACAACCCCCATGTGGATC
AGGAACACCTTTCCCAAACCTAAGAGCCACAACCTCTAACAAATAGAACCT
CTAACCATTTAACAAGACCCGGCAACGCCGATCTACGAACCAAGTTACCC
CAGGGATAACAGCGCAATCCCCTTTTCAGAGCCCCTATCGACAAGGGGGTT
TACGACCTCGATGTTGGATCAGGACACCCCAATGGTGCAGCCGCTA

Appendix 5. Sequence of 16S mitochondrial gene (primer L1567 and H2196) of *Ophisternon gutturale* and *Ophisternon* sp.

```
>Seq1 [organism=Ophisternon gutturale] Ophisternon gutturale
voucher RCK 16S rRNA gene, partial sequence; mitochondrial
ACAGATAAAGCACCTCACTTACACCGAGGTAAATACCAGTGCAAACCCGG
TCGCTTTGATACCCAACAGCTAGCACAAATATTCCAACAACAACAGTCAC
CATAAATAAAACATAAECTCACCTAAAACCTATATAACAAAAACATTTACC
TTCCAAGTATGGGAGACAGAAAAGAACTATGCGCGATAGAGTAAGTACC
GTAAGGGAACGCTGAAAAAGAACTGAAATAAATCAGTCAAGTTAAAAAAA
GCAGAGACTAATACTCGTACCTTTTGCATCATGATTTAGCTAGTCCACCC
AAGCAAAGCGTACTTTAGTTTGTATACCCCGAAAATAAGGGAGCTACTCCA
AGACAGCTTACAAATAGAGCCTACCCGTCTCTGTAGCAAAAAGAGTGGGGA
GAACTTCGAGTAGAGGGCATAAACAACCGAACTTAGTTATAGCTGGTTG
CCTGAGAACCGAATATGAGTTCAACCTCCCGATTTCTCCATACACCTTAT
TCTTCTCCCCTCCCTCCTATAAGCTAACAAAAGAATGTCGAGAGAGCTAG
TCAAAGGGGGGAACAACCTCTCTGACACAAGATACAACCTTTTACCAGAAG
ATAAGGATCATAACACCAAAGGTATTTT
```

```
>Seq2 [organism=Ophisternon sp. n. 1] Ophisternon sp. n. 1 TA-2015
voucher AGR1 16S rRNA gene, partial sequence; mitochondrial
ATAGATAAAGCACCTCACTTACACCGAGGCAAATACCAGTGCAAACCCGG
TCGCTTTGATACCCAACAGCTAGCACATCACTCCAACAACAACAATCAC
TATAAATAAAACACAACCCCGCCTAAAACCTATACAACAAAAACATTTTACC
TTCAAAGTATGGGAGACAGAAAAGAACTATGCGCTATAGAATAAGTACC
GTAAGGGAACGCTGAAAAAGAAATTGAAATAAATCAGTTAAGTTAAAAAAA
GCAGAGACTAACACTCGTACCTTTTGCATCATGATTTAGCTAGTCTACCC
AAGCAAAGCGCACTTTAGTTTGGTGCCCCGAAAATAAGGGAGCTACTCCA
AGACAGCTTATATAAATAGAGCCTACCCATCTCTGTGGCAAAAAGAGTGGC
GAGAGCTTTGAGTAGAGGTGACAAACCAATCGAACTTAGTTATAGCTGGT
TGCCTGAGAACCGAATATGAGTTCAACCTCCCGCTTCTTCATACACTAT
CTTTAACCTTACCCCCACCCACACAGGCTAACAAAAGAACACCGAGAG
AGCTAGTCAAAGGGGGGAACAACCTCCTTTGACACAAGACACAACCTTTACCA
GAAGGGTAAGAATCACAACACCTAAGGTACTTTT
```

```
>Seq3 [organism=Ophisternon sp. n. 1] Ophisternon sp. n. 1 TA-2015
voucher AGR2 16S rRNA gene, partial sequence; mitochondrial
ATAGATAAAGCACCTCACTTACACCGAGGCAAATACCAGTGCAAACCCGG
TCGCTTTGATACCCAACAGCTAGCACATCACTCCAACAACAACAATCAC
TATAAATAAAACACAACCCCGCCTAAAACCTATACAACAAAAACATTTTACC
TTCAAAGTATGGGAGACAGAAAAGAACTATGCGCTATAGAATAAGTACC
GTAAGGGAACGCTGAAAAAGAAATTGAAATAAATCAGTTAAGTTAAAAAAA
GCAGAGACTAACACTCGTACCTTTTGCATCATGATTTAGCTAGTCTACCC
AAGCAAAGCGCACTTTAGTTTGGTGCCCCGAAAATAAGGGAGCTACTCCA
AGACAGCTTATATAAATAGAGCCTACCCATCTCTGTGGCAAAAAGAGTGGC
GAGAGCTTTGAGTAGAGGTGACAAACCAATCGAACTTAGTTATAGCTGGT
TGCCTGAGAACCGAATATGAGTTCAACCTCCCGCTTCTTCATACACTAT
CTTTAACCTTACCCCCACCCACACAGGCTAACAAAAGAACACCGAGAG
AGCTAGTCAAAGGGGGGAACAACCTCCTTTGACACAAGACACAACCTTTACCA
GAAGGGTAAGAATCACAACACCTAAGGTACTTTT
```

Appendix 6. Sequence of COI mitochondrial gene of *Ophisternon gutturale* and *Ophisternon* sp.

>Seq1 [organism=*Ophisternon gutturale*] *Ophisternon gutturale*
voucher RCK cytochrome oxydase I (COI) gene, partial sequence;
mitochondrial

CCAGCAGAATTATGTCAACCAGGCTCTCTTATGGGCGACGACCAAATCTA
TAATGTTATCGTTACAGCACATGCCTTTGTAATAATTTCTTTATAGTCA
TACCAATCATAATCGGGGGCTTCGGAAACTGATTAGTGCCCTCATAATC
GGCGCTCCAGATATAGCATTCCCCGAATAAATAATATAAGCTTCTGACT
CCTCCCTCCTTCATTTCTACTTCTGCTAGCCTCTTCTGGCGTAGAAGCTG
GAGCAGGCACCGGCTGGACAGTTTACCCCCCTTTAGCTGGTAATCTAGCT
CACGCCGGGGCCTCAGTAGACTTAACAATCTTTTCTTACATTTAGCAGG
TGTCTCCTCAATTTTAGGGCTATTAATTTTCATCACTACTATTATTAACA
TAAACCCCCAGCTATCTCTCAATATCAAACACCCCTTTTGTGTTGATCC
GTAATAATTACTGCTGTCTACTCCTCCTCCTACCAGTATTGGCAGC
AGGAATCACAATATTATTAACAGACC

>Seq2 [organism=*Ophisternon* sp. n. 1] *Ophisternon* sp. n. 1 TA-2015
voucher AGR1 cytochrome oxydase I (COI) gene, partial sequence;
mitochondrial

CGAGCAGAGTTATGTCAACCAGGCTCTCTTATAGGCGATGACCAAATCTA
TAATGTTATCGTTACAGCACATGCATTCGTAATAATCTTCTTTATAGTAA
TACCAATTATAATTGGGGGTTTCGGAAACTGATTAGTACCACTAATAATT
GGCGCCCCAGACATAGCCTTCCCTCGAATAAATAATATAAGCTTTTACT
CCTTCCCCCTCATTTCTACTTTTATTAGCTTCTTCTGGTGTAGAAGCTG
GAGCAGGCACCGGCTGAACAGTTTACCCCCCTTCTGCTGGTAATTTAGCC
CACGCCGGAGCCTCAGTAGACTTGACAATCTTTTCTTACACTTAGCAGG
CATCTCCTCAATTCTAGGGCTATCAACTTCATCACTACTATCATTAAATA
TAAACCTCCAACCATCTCCCAATACCAGACACCTCTCTTTGTGTTGATCC
GTAATAATTACTGCCATCTGCTCCTACTTTCCCTCCCAGTATTGGCAGC
AGGAATTACAATATTATTAACAGACC

>Seq3 [organism=*Ophisternon* sp. n. 1] *Ophisternon* sp. n. 1 TA-2015
voucher AGR2 cytochrome oxydase I (COI) gene, partial sequence;
mitochondrial

CGAGCAGAGTTATGTCAACCAGGCTCTCTTATAGGCGATGACCAAATCTA
TAATGTTATCGTTACAGCACATGCATTCGTAATAATCTTCTTTATAGTAA
TACCAATTATAATTGGGGGTTTCGGAAACTGATTAGTACCACTAATAATT
GGCGCCCCAGACATAGCCTTCCCTCGAATAAATAATATAAGCTTTTACT
CCTTCCCCCTCATTTCTACTTTTATTAGCTTCTTCTGGTGTAGAAGCTG
GAGCAGGCACCGGCTGAACAGTTTACCCCCCTTCTGCTGGTAATTTAGCC
CACGCCGGAGCCTCAGTAGACTTGACAATCTTTTCTTACACTTAGCAGG
CATCTCCTCAATTCTAGGGCTATCAACTTCATCACTACTATCATTAAATA
TAAACCTCCAACCATCTCCCAATACCAGACACCTCTCTTTGTGTTGATCC
GTAATAATTACTGCCATCTGCTCCTACTTTCCCTCCCAGTATTGGCAGC
AGGAATTACAATATTATTAACAGACC

Appendix 7. Sequence of RAG1 nuclear gene of *Ophisternon gutturale* and *Ophisternon* sp.

```
>Seq1 [organism=Ophisternon gutturale] Ophisternon gutturale
voucher RCK recombination activating gene 1 (RAG1) mRNA gene,
partial sequence; nuclear
CGAGCTGCTGAGAAGGAGCTTATCCCTGGCTTTCACCAGTTTGAATGGCAGCCCGCTCTC
AAGAATGTGTCTCCATCCTGCAATGTTGGCATTATTAATGGGCTCTCTGGATGGGCTTCC
TCAGTGGATGACTCCCCAGCTGATACCATCAGCCGGCGGTTTCGCTACGATGTGGCACTG
GTGTCAGCGTTAAAGGATCTGGAGGAGGACATCATGGAGGGGCTGAGAGAAAGTGGGATG
GAAGACAGCGCTTGCACCTCAGGCTTTAACGTCATGATCAAGGAATGTTGTGATGGCATG
GGTGACGTCAGCGAGAAGCACGGCGGAGGACCAGTTGTTCCCTGAGAAAGCTGTACGTTTC
TCTTTCACTGTTATGTCTGTCTGTCTGGGCGGATGATAGGAAGGAGGAGGTTACCATT
TTCAGTGAAGCCAAAGCCAAACTCAGAAGTGTCTGTAAGCCCCTTTCCTAATGTTTGTG
GATGAGTCAGACCATGAGACACTCACAGCTGTCTGGGGCCCGTAGTTGCAGAGCGTAAC
GCAATGAAAGAGAGCAGGCTCATCTAGCTATTTGGCGGACTGCCTCGCTCCATCCGCTTC
CACTTCAGAGGCACGGGATACGATGAGAAGATGGTGCGAGAGATGGAGGGCCTGGAGGCC
TCTGGGTCTACATACATCTGCACTCTGTGTGACTCAAGTCGGGCAGAGGCCTCTCAAAC
ATGGTCTTACTCCATCACC CGCAGTCATGAAGAGAACCTAGAAGTTATGAAATATGG
AGAACCAACCCCTTCTCTGAGTCTGTGGATGAGCTGCGAGACAGAGTCAAAGGGGTGTG
GCCAAGCCCTTCATGGATACCCAGCCACACTAGATGCATTACACTGTGACATTGGCAAT
GCCACTGAGTTCTATAAAATCTTCCAGGATGAGATTGGGGAGGTGTATCAAAGGTCAAT
CCCAGCCGGGAGGAACGGCGTAGCTGGAGGGCAGCCCTAGATAAACAGCTGAGGAAGAAG
ATGAAGCTTAAACCGGTAATGAGGATGAATGGGAACATGCCCGCAGGCTAATGACCTTA
GAGGCTGTGGAGGTGGTGTGTGAGCTGGTGGCCCTCAGAGGAGAGAAGGGAGGCCCTAAGG
GAGCTTATGCGACTCTACCTCCAAATGAAGCCTGTGTGGCGTGCCACCTGCCAGCCAAG
GAGTGCCCTGACCAGCTGTGCCGCTACAGCTTTAACTCCCAGCGTTTTGCCGACATCCTC
TCCTCTACATTCAAATATAGGTACAATGGAAAGATAACCAATTACCTGCACAAGACCCTG
GCCCATGTGCCTGAAATCATAGAGAGAGATGGATCCATAGGAGCCTGGGCCAGCGAGGGT
AATG
```

Appendix 7. continued

>Seq2 [organism=Ophisternon sp. n. 1] Ophisternon sp. n. 1 TA-2015
voucher AGR1 recombination activating gene 1 (RAG1) mRNA gene,
partial sequence; nuclear

```
CGAGCTGCTGAGAAGGAGCTTATCCCTGGCTTTCACCAGTTTGAATGGCAGCCCCGCTCTC
AAGAATGTGTCTCCATCCTGCAATGTTGGCATTATTAATGGGCTCTCTGGATGGGCTTCC
TCAGTGGATGACTCCCCAGCTGATACCATCAGCCGGCGGTTTCGCTACGATGTGGCACTG
GTGTCAGCGTTAAAGGATCTGGAGGAGGACATCATGGAGGGGCTGAGAGAAAGTGGGATG
GAAGACAGCGCTTGCACCTCAGGCTTTAACGTCATGATCAAGGAATGTTGTGATGGCATG
GGTGACGTCAGCGAGAAGCACGGCGGAGGACCAGTTGTTCTGAGAAAGCTGTACGTTTC
TCTTTCACTGTTATGTCTGTCTGTCTGGGCGGATGATAGGAAGGAGGAGGTTACCATT
TTCACTGAGCCAAAGCCAAACTCAGAACTGTCTGTAAGCCCCTTTGCCTAATGTTTGTG
GATGAGTCAGACCATGAGACACTCACAGCTGTCTGGGGCCCGTAGTTGCAGAGCGTAAC
GCAATGAAAGAGAGCAGGCTCATTCTAGCTATTGGCGGACTGCCTCGCTCCATCCGCTTC
CACTTCAGAGGCACGGGATACGATGAGAAGATGGTGCAGAGATGGAGGGCCTGGAGGCC
TCTGGGTCTACATACATCTGCACTCTGTGTGACTCAAGTCGGGCAGAGGCCCTCAAAC
ATGGTTCTACACTCCATCACCCGAGTCATGAAGAGAACCCTAGAACGTTATGAAATATGG
AGAACCAACCCCTTCTCTGAGTCTGTGGATGAGCTGCGAGACAGAGTCAAAGGGGTGTCG
GCCAAGCCCTTCATGGATACCCAGCCCACGCTAGATGCATTACACTGTGACATTGGCAAT
GCCACTGAGTTCTATAAAATCTTCCAGGATGAGATTGGGGAGGTGTATCAAAGGTCAAT
CCCAGCCGGGAGGAACGGCGTAGCTGGAGGGCAGCCCTAGATAAACAGCTGAGGAAGAAG
ATGAAGCTTAAACCGTAATGAGGATGAATGGGAACTATGCTCGCAGGCTAATGACCTTA
GAGACTGTGGAGGTGGTGTGTGAGCTGGTGCCTCAGAGGAGAGAAGGGAGGCCCTAAGG
GAGCTTATGCGACTCTACCTCCAAATGAAGCCTGTGTGGCGTGCCACTTGCCCAGCCAAG
GAGTGCCCTGACCAGCTGTGCCGCTACAGCTTTAACTCCCAGCGTTTTGCCGACATCCTC
TCCTCTACATTCAAATATAGGTACAATGGAAAGATAACCAATTACCTGCACAAGACCCTG
GCCCATGTGCCTGAAATCATAGAGAGAGATGGATCCATAGGAGCCTGGGCCAGCGAGGGT
AATG
```

Appendix 7. continued

>Seq3 [organism=Ophisternon sp.nov1] Ophisternon sp.nov1 voucher
AGR2 recombination activating gene 1 (RAG1) mRNA gene, partial
sequence; nuclear

```
CGAGCTGCTGAGAAGGAGCTTATCCCTGGCTTTCACCAGTTTGAATGGCAGCCCCGCTCTC
AAGAATGTGTCTCCATCCTGCAATGTTGGCATTATTAATGGGCTCTCTGGATGGGCTTCC
TCAGTGGATGACTCCCCAGCTGATACCATCAGCCGGCGGTTTCGCTACGATGTGGCACTG
GTGTCAGCGTTAAAGGATCTGGAGGAGGACATCATGGAGGGGCTGAGAGAAAGTGGGATG
GAAGACAGCGCTTGCACCTCAGGCTTTAACGTCATGATCAAGGAATGTTGTGATGGCATG
GGTGACGTCAGCGAGAAGCACGGCGGAGGACCAGTTGTTCTGAGAAAGCTGTACGTTTC
TCTTTCACTGTTATGTCTGTCTGTCTGGGCGGATGATAGGAAGGAGGAGGTTACCATT
TTCACTGAGCCAAAGCCAAACTCAGAACTGTCTGTAAAGCCCCTTTGCCTAATGTTTGTG
GATGAGTCAGACCATGAGACACTCACAGCTGTCTGGGGCCCGTAGTTGCAGAGCGTAAC
GCAATGAAAGAGAGCAGGCTCATTCTAGCTATTGGCGGACTGCCTCGCTCCATCCGCTTC
CACTTCAGAGGCACGGGATACGATGAGAAGATGGTGCAGAGATGGAGGGCCTGGAGGCC
TCTGGGTCTACATACATCTGCACTCTGTGTGACTCAAGTCGGGCAGAGGCCCTCAAAC
ATGGTTCTACACTCCATCACCCGAGTCATGAAGAGAACC TAGAACGTTATGAAATATGG
AGAACCAACCCCTTCTCTGAGTCTGTGGATGAGCTGCGAGACAGAGTCAAAGGGGTGTCG
GCCAAGCCCTTCATGGATACCCAGCCCACGCTAGATGCATTACACTGTGACATTGGCAAT
GCCACTGAGTTCTATAAAATCTTCCAGGATGAGATTGGGGAGGTGTATCAAAGGTCAAT
CCCAGCCGGGAGGAACGGCGTAGCTGGAGGGCAGCCCTAGATAAACAGCTGAGGAAGAAG
ATGAAGCTTAAACCGGTAATGAGGATGAATGGGAACTATGCTCGCAGGCTAATGACCTTA
GAGACTGTGGAGGTGGTGTGTGAGCTGGTGCCTCAGAGGAGAGAAGGGAGGCCCTAAGG
GAGCTTATGCGACTCTACCTCCAAATGAAGCCTGTGTGGCGTGCCACTTGCCCAGCCAAG
GAGTGCCCTGACCAGCTGTGCCGCTACAGCTTTAACTCCCAGCGTTTTGCCGACATCCTC
TCCTCTACATTCAAATATAGGTACAATGGAAAGATAACCAATTACCTGCACAAGACCCTG
GCCCATGTGCCTGAAATCATAGAGAGAGATGGATCCATAGGAGCCTGGGCCAGCGAGGGT
AATG
```

Appendix 8. Sequence of 1st intron of S7 nuclear gene of *Ophisternon gutturale* and *Ophisternon* sp.

>Seq1 [organism=*Ophisternon gutturale*] *Ophisternon gutturale*
voucher RCK S7 ribosomal protein gene, intron, partial sequence;
nuclear

AAGTTTACGTACATTCAAGCATCTTAAGTCTCTGATACATTAGATTATGT
GGTTTGCAGCGGCGGATGGCAGTTTTATAACACAAAATAAAGCGTAGTAG
AGTAGCTCCGGTGAATGGGTTCTGGTTGTAGGGAGTAGTTAGCAGGCCGT
GCATGTGTTTCAGACAGCAGCTAGAGGAGCGACTGGAGAAGGTGGCAGACT
GAAGATATTGTGCTCTGTTTATGGTTTCGTTTTATGTAAAATTAATCAAGT
TACTGCTGGTTGTCACATAAACTCTCCAGCCAAATTCTAGCAGAAATAA
GACAACCGGCTAAGTTATATTAGCT

>Seq2 [organism=*Ophisternon* sp. n. 1] *Ophisternon* sp. n. 1 TA-2015
voucher AGR1 S7 ribosomal protein gene, intron, partial sequence;
nuclear

AAGTTTACGTACATTAAAGCATCTTGAGTCTCTGATACATTAGATTATGT
GGTTTGCAGTGGCGGATGGCAGTTTTATAACACAAAATAAAGCGTAGTAG
AGTAGCTCCTGTGAATGGGTTCTGGTTGTAGGGAGTAGTTAGCAGGCCGT
GCATGTGTTTCAGATAGCAGCTAGAGGAGCGACTGGAGAAGGTGGCAGACT
GAAGATATTGTGCTCTGTTTATGGTTTGTGTTTTATGTAAAATTAATCAAGT
TACTGCTGGTTGTCACATAAACTCTCCAGCCAAATTCTAGCAGAAATAA
GACAACCGGCTAAGTTATATTAGCT

>Seq3 [organism=*Ophisternon* sp. n. 1] *Ophisternon* sp. n. 1 TA-2015
voucher AGR2 S7 ribosomal protein gene, intron, partial sequence;
nuclear

AAGTTTACGTACATTAAAGCATCTTGAGTCTCTGATACATTAGATTATGT
GGTTTGCAGTGGCGGATGGCAGTTTTATAACACAAAATAAAGCGTAGTAG
AGTAGCTCCTGTGAATGGGTTCTGGTTGTAGGGAGTAGTTAGCAGGCCGT
GCATGTGTTTCAGATAGCAGCTAGAGGAGCGACTGGAGAAGGTGGCAGACT
GAAGATATTGTGCTCTGTTTATGGTTTGTGTTTTATGTAAAATTAATCAAGT
TACTGCTGGTTGTCACATAAACTCTCCAGCCAAATTCTAGCAGAAATAA
GACAACCGGCTAAGTTATATTAGCT

Bibliography

- Abbott, L.A., Bisby, F.A., Rogers, D.J. 1985. *Taxonomic Analysis in Biology*. Columbia University Press, New York, USA.
- Abdurahman, M., Sabirhazi, G., Liu, B., Yin, L., Pan, B. 2012. Comparison of five *Calligonum* species in Tarim basin based on morphological and molecular data. *EXCLI Journal*, **11**, 776-782.
- Agapow, P.M. 2004. The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*, **79**(2), 161-179.
- Agnarsson, I., Kuntner, M. 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. *Systematic Biology*, **56**(3), 531-539.
- Agorreta, A., Dominguez-Dominguez, O., Reina, R.G., Miranda, R., Bermingham, E., Doadrio, I. 2013. Phylogenetic relationships and biogeography of *Pseudoxiphophorus* (Teleostei: Poeciliidae) based on mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, **66**(1), 80-90.
- Agrawal, A.A. 2001. Phenotypic plasticity in the interactions and evolution of species. *Science*, **294** (5541), 321-326.
- Albrecht, G.H. 1980. Multivariate analysis and the study of form, with special reference to canonical variate analysis. *American Zoologist*, **20**, 679-693.
- Alda, F., Reina, R.G., Doadrio, I., Bermingham, E. 2013. Phylogeny and biogeography of the *Poecilia sphenops* species complex (Actinopterygii, Poeciliidae) in Central America. *Molecular Phylogenetics and Evolution*, **66**(3), 1011-1026.
- Allen, G.R., Erdmann, M.V. 2012. *Reef Fishes of the East Indies*. Tropical Reef Research, Perth, Australia.
- Allen, G.R., Midgley, S.H., Allen, M. 2002. *Freshwater Fishes of Australia*. Western Australian Museum, Perth, Australia.
- Allendorf, F.W., Ryman, N., Utter, F. 1987. Genetics and Fishery Management: Past, Present and Future. In: *Population Genetics and Fishery Management*, (Eds.) N. Ryman, F. Utter, University of Washington Press. Seattle (USA), pp. 1-19.
- Anker, A., Baeza, J.A. 2012. Molecular and morphological phylogeny of hooded shrimps, genera *Betaeus* and *Betaeopsis* (Decapoda, Alpheidae): testing the center of origin biogeographic model and evolution of life history traits. *Molecular Phylogenetics and Evolution*, **64**(3), 401-415.
- Arif, I.A., Khan, H.A. 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal Biodiversity and Conservation*, **32**(1), 9-17.

- Arribas, P., Andújar, C., Sánchez-Fernández, D., Abellán, P., Millán, A. 2013. Integrative taxonomy and conservation of cryptic beetles in the Mediterranean region (Hydrophilidae). *Zoologica Scripta*, **42**(2), 182-200.
- Arthofer, W., Rauch, H., Thaler-Knoflach, B., Moder, K., Muster, C., Schlick-Steiner, B.C., Steiner, F.M. 2013. How diverse is *Mitopus morio*? Integrative taxonomy detects cryptic species in a small-scale sample of a widespread harvestman. *Molecular Ecology*, **22**(14), 3850-3863.
- Asgharian, H., Sahafi, H.H., Ardalan, A.A., Shekarriz, S., Elahi, E. 2011. Cytochrome c oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf and analysis using meta-data flag several cryptic species. *Molecular Ecology Resources*, **11**(3), 461-472.
- Atchley, W.R., Anderson, D. 1978. Ratios and the statistical analysis of biological data. *Systematic Zoology*, **27**, 71-78.
- Aung, O., Nguyen, T.T.T., Poopuang, S., Kamonrat, W. 2010. Microsatellite DNA markers revealed genetic population structure among captive stocks and wild populations of mrigal, *Cirrhinus cirrhosus* in Myanmar. *Aquaculture*, **299**(1-4), 37-43.
- Austin, C.M., Knott, B. 1996. Systematics of the freshwater crayfish genus *Cherax* Erichson (Decapoda: Parastacidae) in South-western Australia: Electrophoretic, morphological and habitat variation. *Australian Journal of Zoology*, **44**, 223-258.
- Austin, C.M., Ryan, S.G. 2002. Allozyme evidence for a new species of freshwater crayfish of the genus *Cherax* Erichson (Decapoda: Parastacidae) from the south-west of Western Australia. *Invertebrate Systematics*, **16**, 357-367.
- Avice, J.C. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, USA.
- Baco, A.R., Cairns, S.D. 2012. Comparing molecular variation to morphological species designations in the deep-sea coral *Narella* reveals new insights into seamount coral ranges. *PLoS One*, **7**(9), 1-15 (e45555).
- Bailey, R.M., Gans, C. 1998. Two new synbranchid fishes-*Monopterus roseni* from Peninsular India and *M.desilvai* from Sri Langka. *Occasional Papers of the Museum of Zoology The University of Michigan*, **726**, 1-18.
- Baillie, V.L., Bouwer, G. 2011. Development of highly sensitive assays for detection of genetic variation in key *Helicoverpa armigera* nucleopolyhedrovirus genes. *Journal of Virological Methods*, **178**(1-2), 179-185.
- Balakhrihnan, R. 2005. Species concepts, species boundaries and species identification: A view from the tropics. *Systematic Biology*, **54**, 689-693.
- Balaraj, S., Basheer, M. 2012. Genetic diversity among *Puntius sophore* complex using restriction fragment length polymorphism. *Journal of Medical and Allied Sciences*, **2**(2), 49-53.

- Baranski, M., Loughnan, S., Austin, C.M., Robinson, N. 2006. A microsatellite linkage map for the blacklip abalone *Haliotis rubra*. *Animal Genetics*, **37**, 563-570.
- Baranski, M., Rourke, M., Loughnan, S., Hayes, B., Austin, C.M., Robinson, N. 2008. Detection of QTL for growth rate in the blacklip abalone (*Haliotis rubra* Leach) using selective DNA pooling. *Animal Genetics*, **39**, 606-614.
- Barman, A.S., Singh, M., Singh, R.K., Sarkar, T., Lal, K.K. 2014. Molecular identification and phylogeny of *Channa* species from Indo-Myanmar biodiversity hotspots using mitochondrial COI gene sequences. *Biochemical Systematics and Ecology*, **57**, 367-373.
- Barkley, T.M., DePriest, P., Funk, V., Kiger, R.W., Kress, W.J., Moore, G. 2004. Linnaean nomenclature in the 21st century - a report from a workshop on integrating traditional nomenclature and phylogenetic classification. *Taxon*, **53**(1), 153-158.
- Barreta, J., Gutierrez-Gil, B., Iniguez, V., Saavedra, V., Chiri, R., Latorre, E., Arranz, J.J. 2013. Analysis of mitochondrial DNA in Bolivian llama, alpaca and vicuna populations: a contribution to the phylogeny of the South American camelids. *Animal Genetics*, **44**(2), 158-168.
- Bektas, Y., Belduz, A.O. 2009. Morphological variation among Atlantic Horse Mackerel *Trachurus trachurus* populations from Turkish Coastal Waters. *Journal of Animal and Veterinary Advances*, **8**(3), 511-517.
- Beheregaray, L.B., Caccione, A. 2007. Cryptic biodiversity in a changing world. *Journal of Biology*, **6**(9), 1-5.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. 2004. *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Berra, T.M. 2007. *Freshwater Fish Distribution*. The University of Chicago Press, Chicago.
- Berry, P.E. 2002. Biological inventories and the Phylocode. *Taxon*, **51**(1), 27-29.
- Bertorelle, G., Bruford, M.W., Hauffe, H.C., Rizzoli, A., Vernesi, C. 2009. *Population Genetics for Animal Conservation*. Cambridge University Press, Cambridge, UK.
- Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Ingram, K.K., Das, I. 2006. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**(3), 148-155.
- Blackwelder, R.E. 1967. *Taxonomy: A Text and Reference Book*. Wiley, New York.

- Blanton, E.B., Page, L.M., Hilber, S.A. 2013. Timing of clade divergence and discordant estimates of genetic and morphological diversity in the Slender Madtom, *Nocturus exilis* (Ictaluridae). *Molecular Phylogenetics and Evolution*, **66**, 679-693.
- Bloch, M.E. 1795. *Naturgeschichte der ausländischen Fische*, Berlin (Germany).
- Bloom, D.D., Lovejoy, N.R. 2012. Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *Journal of Evolutionary Biology*, **25**(4), 701-715.
- Bohlen, J., Slechtova, V., Tan, H.H., Britz, R. 2011. Phylogeny of the Southeast Asian freshwater fish genus *Pangio* (Cypriniformes; Cobitidae). *Molecular Phylogenetics and Evolution*, **61**(3), 854-865.
- Bookstein, F.L. 1982. Foundations of morphometrics. *Annual Review of Ecology and Systematics*, **13**, 451-470.
- Bortolus, A. 2008. Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. *AMBIO: A journal of Human Environment*, **37**(2), 114-118.
- Bristol, R.M., Fabre, P.H., Irestedt, M., Jonsson, K.A., Shah, N.J., Tatayah, V., Warren, B.H., Groombridge, J.J. 2013. Molecular phylogeny of the Indian Ocean Terpsiphone paradise flycatchers: undetected evolutionary diversity revealed amongst island populations. *Molecular Phylogenetics and Evolution*, **67**(2), 336-347.
- Britz, R., Lalremsanga, H.T., Lalrotluanga, Lalramliana. 2011. *Monopterus ichthyophoides*, a new species of scaled swamp eel (Teleostei: Synbranchiformes: Synbranchidae) from Mizoram, India. *Zootaxa*, **2936**, 51-58.
- Buhay, E. 2009. 'Coi-like' sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology*, **29**(1), 96-110.
- Cai, X., Gou, X., Zeng, F., Zhang, T., Jiang, L., Fan, D., Pu, D., Zeng, X. 2008. Mitochondrial DNA diversity of *Monopterus albus* from the Sichuan Basin of China. *Biochemical Genetics*, **46**(9-10), 583-589.
- Cai, X., Yu, S., Mipam, T., Zhang, X., Yue, B. 2012. Phylogenetic lineages of *Monopterus albus* (Synbranchiformes: Synbranchidae) in China inferred from mitochondrial control region. *Journal of Zoological Systematics and Evolutionary Research*, **51**(1), 38-44.
- Cerutti-Pereyra, F., Meekan, M.G., Wei, N.W., O'Shea, O., Bradshaw, C.J., Austin, C.M. 2012. Identification of rays through DNA barcoding: an application for ecologists. *PLoS One*, **7**(6), 1-10 (e36479).
- Chakrabarty, P. 2010. Genotypes: a concept to help integrate molecular phylogenetics and taxonomy. *Zootaxa*, **2632**, 67-68.

- Chan, C.X., Ragan, M.A. 2013. Next-generation phylogenomics. *Biology Direct*, **8**(3), 1-6.
- Chan, C.X., Bernard, G., Poirion, O., Hogan, J.M., Ragan, M.A. 2014. Inferring phylogenies of evolving sequences without multiple sequence alignment. *Scientific Reports*, **4**, 1-9.
- Chauduri, B.L. 1913. Zoological result of the Abor expedition, 1911-12.XVII. Fish. *Records of the Indian Museum*, **8**(3), 243-257.
- Chauhan, T., Rajiv, K. 2010. Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology*, **1**(04), 281-291.
- Chen, W., 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. *Molecular Phylogenetics and Evolution*, **26**, 262-288.
- Cheng, Q., Lu, D. 2005. PCR-RFLP analysis of cytochrome b gene does not support *Coilia ectenes taihuensis* being a subspecies of *Coilia ectenes*. *Journal of Genetics* **84**(3), 307-310.
- Cheng, Q. 2010. Morphological variation of *Coilia mystus* (Clupeiformes: Engraulidae) in three Chinese estuaries. *Journal of Life Sciences*, **4**(6), 29-34.
- Chow, S., Hazama, K. 1998. Universal primers of S7 ribosomal protein gene introns in fish. *Molecular Ecology*, **7**, 1247-1263.
- Cognato, A.I. 2006. Standard percent DNA sequence difference for insects does not predict species boundaries. *Journal of Economic Entomology*, **99**, 1037-1045.
- Collins, T.M., Trexler, J.C., Nico, L.G., Rawlings, T.A. 2002. Genetic diversity in a morphologically conservative invasive taxon: multiple introductions of swamp eels to the Southeastern United States. *Conservation Biology*, **16**, 1024-1035.
- Cook, L.G., Edward, R.D., Crisp, M.D., Hardy, N.B. 2010. Need morphology always be required for new species descriptions. *Invertebrate Systematics*, **24**, 322-326.
- Cowman, P.F., Bellwood, D.R. 2011. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary Biology*, **24**(12), 2543-2562.
- Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of Linnean Society*, **85**, 407-415.
- Dayrat, B., Cantino, P.D., Clarke, J.A., de Queiroz, K. 2008. Species Names in the PhyloCode: the approach adopted by the International Society for Phylogenetic Nomenclature. *Systematic Biology*, **57**(3), 507-514.
- De Hoog, G.S. 1981. Methodology of taxonomy. *Taxon*, **30**(4), 779-783.

DeSalle, R., Egan, M.G., Siddall, M. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society London B Biological Sciences*, **360**, 1905-1916.

Doberty, D., McCarthy, T.K. 2004. Morphometric and meristic characteristics analyses of two western irish populations of arctic char, *Salvelinus alpinus* (L.). *Proceedings of The Royal Irish Academy*, **104B**(1), 75-85.

Dooms, S., Papakostas, S., Hoffman, S., Delbare, D., Dierckens, K., Triantafyllidis, A., De Wolf, T., Vadstein, O., Abatzopoulos, T.J., Sorgeloos, P., Bossier, P. 2007. Denaturing Gradient Gel Electrophoresis (DGGE) as a tool for the characterisation of *Brachionus* sp. strains. *Aquaculture*, **262**(1), 29-40.

Dornburg, A., Moore, J.A., Webster, R., Warren, D.L., Brandley, M.C., Iglesias, T.L., Wainwright, P.C., Near, T.J. 2012. Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. *Molecular Phylogenetics and Evolution*, **65**(2), 727-738.

Duda, R., Hart, P., Stork, D. 2001. *Pattern Classification. 2nd edition*. John Wiley and Sons, New York, USA.

Dudu, A., Georgescu, S., Costache, M. 2011. PCR-RFLP method to identify Salmonid species of economic importance. *Animal Science and Biotechnologies*, **44**(1), 193-196.

Dunham, R.A. 2004. *Aquaculture and Fisheries Biotechnology-Genetic Approaches*. CABI Publishing, USA.

Eapen, K.C. 1963. A new species of *Monopterus* from South India. *Bulletin of the Department of Marine Biology & Oceanography*, **1**, 130-132.

Ebach, M.C., Williams, D.M. 2004. Classification. *Taxon*, **53**(3), 791-794.

Ebach, M.C., Holdrege, C. 2005. DNA barcoding is no substitute for taxonomy. *Nature*, **434**(7034), 697.

El-Hanafy, A.A., Salem, H.H. 2009. PCR-RFLP of IGFBP-3 gene in some Egyptian sheep breeds. *American-Eurasian Journal of Agricultural and Environmental Sciences*, **5**(1), 82-85.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**(6), 435-445.

Epifanio, J., Nielsen, J. 2001. The role of hybridization in the distribution, conservation and management of aquatic species. *Reviews in Fish Biology and Fisheries*, **10**, 245-251.

Eschmeyer, W. N. (ed). Catalog Of Fishes: Genera, Species, References. (<http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>). Electronic version accessed 20 November 2015. [This version was edited by Bill Eschmeyer.]

- Excoffier, L., Laval, G., Schneider, S. 2006. *ARLEQUIN Ver 3.1. An Intergrated Software Package for Population Genetics*. Computational and Molecular Genetics Lab (CMPG), Institute of Zoology, University of Berne, Switzerland.
- Fassatoui, C., Mdelgi, E., Romdhane, M.S. 2009. A preliminary investigation allozyme genetic variation and population structure in common pandora (*Pagellus erythrinus*, Sparidae) from Tunisian and Libyan coasts. *Ichthyology Research*, **56**, 301-307.
- Felsenstein, J. 2003. *Theoretical Evolutionary Genetics*. Department of Genome Sciences, University of Washington, Washington, USA.
- Ferguson, J.W.H. 2002. On the use of genetic divergence for identifying species. *Biological Journal of Linnean Society*, **75**, 509-516.
- Ferreira, J.M., Martins, F., Ditchfield, A., Morgante, J.S. 2005. The use of PCR-RFLP as an identification tool for two closely related species of bats of genus *Platyrrhinus*. *Genetics and Molecular Biology*, **28**(1), 120-122.
- Fischer, S.G., Lerman, L.S. 1980. Separation of random fragments of DNA according to properties of their sequences. *Proceedings of the National Academic of Sciences U.S.A.*, **77**, 4420–4424.
- Fischer, S.G., Lerman, L.S. 1983. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academic of Sciences U.S.A.*, **80**, 1579–1583.
- Forey, P.L. 2002. PhyloCode: Pain, no gain. *Taxon*, **51**(1), 43-45.
- Forey, P.L., Fortey, R.A., Kenrick, P., Smith, A.B. 2004. Taxonomy and fossils: a critical appraisal. *Philosophical Transactions of the Royal Society London Series B Biological Sciences*, **359**(1444), 639-653.
- Frezal, L., Leblois, R. 2008. Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution*, **8**(5), 727-736.
- Fu, Y. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.
- Fujita, M.K., Leache, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**(9), 480-488.
- Gardner, S.N., Hall, B.G. 2013. When whole-genome alignments just won't work : kSNPv2 software fro alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLOS One*, **8**(12), e81760.
- Gebiola, M., Gomez-Zurita, J., Monti, M.M., Navone, P., Bernardo, U. 2012. Integration of molecular, ecological, morphological and endosymbiont data for species delimitation within the *Pnigalio soemius* complex (Hymenoptera: Eulophidae). *Molecular Ecology*, **21**(5), 1190-1208.

Gong, J., Forster, R.J., Hai, Y. 2002. Diversity and phylogenetic analysis of bacteria in the mucosa of chicken ceca and comparison with bacteria in the cecal lumen. *FEMS Microbiology Letters*, **208**(1), 1-7.

Gopi, K.C. 2002. A new synbranchid fish, *Monopterus digressus* from Kerala, Peninsular India. *Records of the Zoological Survey of India*, **100**, 137-143.

Gordon, A.D. 1987. A review of hierarchical classification. *Journal of the Royal Statistical Society. Series A (General)*, **150**(2), 119-137.

Goude, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goude (1995), Vol. 2013.

Graham, J.B. 1977. *Air-Breathing Fishes*. Academic Press, San Diego.

Grant, V. 2003. Incongruence between cladistic and taxonomic systems. *American Journal of Botany*, **90**(9), 1263-1270.

Greuter, W., McNeill, J., Barrie, F.R., Burdet, H.M., Demoulin, V., Filgueiras, T.S., Nicolson, D.H., Silva, P.C., Skog, J.E., Treharne, P., Turland, N.J., Hawksworth, D.L. 2000. *International Code of Botanical Nomenclature (St Louis Code)*. *Regnum Vegetabile 138*. Koeltz Scientific Books, Königstein.

Greuter, W. 2004. Recent developments in international biological nomenclature. *Turkish Journal of Botany*, **28**, 17-26.

Greve, C., Hutterer, R., Groh, K., Haase, M., Misof, B. 2010. Evolutionary diversification of the genus *Theba* (Gastropoda: Helicidae) in space and time: a land snail conquering islands and continents. *Molecular Phylogenetics Evolution*, **57**(2), 572-584.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., Petit, R.J. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources*, **11**(4), 591-611.

Gum, B., Gross, R., Geist, J. 2009. Conservation genetics and management implications for European grayling, *Thymallus thymallus*: synthesis of phylogeography and population genetics. *Fisheries Management and Ecology*, **16**(1), 37-51.

Hajibabaei, M., Singer, G.A., Hebert, P.D., Hickey, D.A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**(4), 167-172.

Hall, B.G. 2011. *Phylogenetic trees made easy: a how-to manual 4th ed.* Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Hall, R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, **20**, 353-431.

- Hall, R. 2012. Sundaland and Wallacea: Geology, Plate Tectonics and Palaeogeography. in: *Biotic Evolution and Environmental Change in Southeast Asia*, (Eds.) D.J. Gower, J.E. Richardson, B.R. Rosen, L. Rüber, S.T. Williams, Cambridge University Press. UK, pp. 32-78.
- Hamilton, F. 1822. *An account of the fishes found in the river Ganges and its branches*, Edinburgh and London.
- Hammer, M.P., Adams, M., Hughes, J.M. 2013. Evolutionary Processes and Biodiversity. in: *Ecology of Australian Freshwater Fishes*, (Eds.) P. Humphries, K.F. Walker, CSIRO Publishing. Collingwood, Victoria, pp. 49-79.
- Harris, R. 2005. Attacks on taxonomy: a contentious biological classification system may make cataloging biodiversity tricky. *American Scientist*, **93**(4), 311-312.
- Hassanien, H.A., Kamel, E.A., Salem, M.A., Dorgham, A.S. 2011. Multivariate analysis of morphometric parameters in wild and cultured Nile Tilapia *Oreochromis niloticus*. *Journal of the Arabian Aquaculture Society*, **6**(2), 237-250.
- Hausdorf, B., Hennig, C. 2010. Species delimitation using dominant and codominant multilocus markers. *Systematic Biology*, **59**, 491-503.
- Hausdorf, B. 2011. Progress toward a general species concept. *Evolution*, **65**(4), 923-931.
- Hebert, P.D., Cywinska, A., Ball, S.L., deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of Biological Society B: Biological Sciences*, **270**(1512), 313-321.
- Hebert, P.D., Ratnasingham, S., deWaard, J.R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B Biological Sciences*, **270 Suppl 1**, S96-99.
- Hebert, P.D.N., Stoeckle, M.Y., Zemplak, T.S., Francis, C.M. 2004. Identification of birds through DNA barcodes. *PLoS Biology*, **2**(10), 1657-1663.
- Hellberg, M.E. 2009. Gene Flow and Isolation among Populations of Marine Animals. *Annual Review of Ecology, Evolution, and Systematics*, **40**(1), 291-310.
- Hermida, M., Fernandez, J.C., Amaro, R., Miguel, E.S. 2005. Morphometric and meristic variation in Galician threespine stickleback populations, northwest Spain. *Environmental Biological Fisheries*, **73**(2), 189-200.
- Hickerson, M.J., Meyer, C.P., Moritz, C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**(5), 729-739.
- Hill, J.E., Watson, C.A. 2007. Diet of the Nonindigenous Asian Swamp Eel in Tropical Ornamental Aquaculture Ponds in West-Central Florida. *North American Journal of Aquaculture*, **69**(2), 139-146.

Hill, R.V. 2005. Integration of morphological data sets for phylogenetic analysis of Amniota: The importance of integumentary characters and increased taxonomic sampling. *Systematic Biology*, **54**(4), 530-547.

Hillis, D.M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology*, **47**(1), 3-8.

Hillis, D.M., Moritz, C., Mable, B.K. 1996. *Molecular Systematics*. 2nd ed. Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Holmes, B.H., Steinke, D., Ward, R.D. 2009. Identification of shark and ray fins using DNA barcoding. *Fisheries Research*, **95**(2-3), 280-288.

Hörandl, E., Stuessy, T.F. 2010. Paraphyletic groups as natural units of biological classification. *Taxon*, **59**(6), 1641-1653.

Huang, X., Guo, Y., Shui, Y., Gao, S., Yu, H., Cheng, H., Zhou, R. 2005. Multiple alternative splicing and differential expression of *dmrt1* during gonad transformation of the rice field eel. *Biology of Reproduction*, **73**(5), 1017-1024.

Hurtado-Burillo, M., Ruiz, C., May-Itza, W., Quezada-Euan, J.J.G., De La Rúa, P. 2013. Barcoding stingless bees: genetic diversity of the economically important genus *Scaptotrigona* in Mesoamerica. *Apidologie*, **44**, 1-10.

Huson, D.H., Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**(2), 254-267.

Ihssen, P.E., Booke, H.E., Casselman, J.M., McGlade, J.M., Payne, N.R., Utter, F.M. 1981. Stock identification: materials and methods. *Canadian Journal of Fisheries and Aquatic Science*, **38**(12), 1838-1855.

Jalalizand, A.R., Karimi, A., Modaresi, M., Mahmoodi, E. 2012. Determining morphological traits and genetic diversity of rose aphids using RAPD and RFLP PCR molecular markers. in: *International Conference on Applied Life Sciences Turkey*, pp. 446-452.

Jensen, R.J. 2009. Phenetics: Revolution, reform or natural consequence. *Taxon*, **58**(1), 50-60.

Jo, H., Gim, J., Jeong, K., Kim, H., Joo, G. 2014. Application of DNA barcoding for identification of freshwater carnivorous fish diets: Is number of prey items dependent on size class for *Micropterus salmoides*. *Ecology and Evolution*, **4**(2), 219-229.

John, B.A., Prasannakumar, C., Lyla, P.S., Ajmal Khan, S., Jalal, K.C.A. 2010. DNA barcoding of *Lates calcarifer* (Bloch, 1970). *Research Journal of Biological Sciences*, **5**(6), 414-419.

Jin, G., Nakhleh, L., Shir, S., Tuller, T. 2006. Inferring phylogenetic networks by the maximum parsimony criterion: a case study. *Molecular Biology and Evolution*, **24**(1), 324-337.

- Kalinowski, S.T., Doornik, D.M., Kozfkay, C.C., Waples, R.S. 2012. Genetic diversity in the Snake River sockeye salmon captive broodstock program as estimated from broodstock records. *Conservation Genetics*, **13**(5), 1183-1193.
- Kanno, Y., Vokoun, J.C., Letcher, B.H. 2010. Sibship reconstruction for inferring mating systems, dispersal and effective population size in headwater brook trout (*Salvelinus fontinalis*) populations. *Conservation Genetics*, **12**(3), 619-628.
- Karaiskou, N., Lappa, M., Kalomoiris, S., Oikonomidis, G., Psaltopoulou, C., Abatzopoulos, T.J., Triantaphyllidis, C., Triantafyllidis, A. 2011. Genetic monitoring and effects of stocking practices on small *Cyprinus carpio* populations. *Conservation Genetics*, **12**(5), 1299-1311.
- Keskin, E., Atar, H.H. 2011. Genetic divergence of *Octopus vulgaris* species in the eastern Mediterranean. *Biochemical Systematics and Ecology*, **39**(4-6), 277-282.
- Khan, M.A., Miyan, K., Khan, S., Patel, D.K., Ansari, N.G. 2012. Studies on the elemental profile of otoliths and truss network analysis for stock discrimination of the threatened stinging catfish *Heteropneustes fossilis* (Bloch 1794) from the Ganga River and its tributaries. *Zoological Studies*, **51**(7), 1195-1206.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111-120.
- Kitanishi, S., Yamamoto, T., Edo, K., Higashi, S. 2012. Influences of habitat fragmentation by damming on the genetic structure of masu salmon populations in Hokkaido, Japan. *Conservation Genetics*, **13**(4), 1017-1026.
- Knapp, L.A. 2009. Single nucleotide polymorphism screening with denaturing gradient gel electrophoresis. *Methods in Molecular Biology*, **578**, 137-151.
- Knowlton, N. 2000. Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia*, **420**, 73-90.
- Kotsakiozi, P., Parmakelis, A., Giokas, S., Papanikolaou, I., Valakos, E.D. 2012. Mitochondrial phylogeny and biogeographic history of the Greek endemic land-snail genus *Codringtonia* Kobelt 1898 (Gastropoda, Pulmonata, Helicidae). *Molecular Phylogenetics and Evolution*, **62**(2), 681-692.
- Kottelat, M., Whittorn, A.J., Kartikasari, S.N., Wirjoatmojo, S. 1993. *Freshwater Fishes of Western Indonesia and Sulawesi*. Periplus ed. Limited, Indonesia.
- Kottelat, M. 2013. The fishes of the inland waters of Southeast Asia: A Catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *The Raffles Bulletin of Zoology*, **Supplement No. 27**, 1-663.
- Kress, W.J., DePriest, P.T. 2002. Linnaeus visits the 21st century. *Taxon*, **51**(1), 5-6.

- Kriwy, P., Uthicke, S. 2011. Microbial diversity in marine biofilms along a water quality gradient on the Great Barrier Reef. *Systematic and Applied Microbiology*, **34**(2), 116-126.
- Kuntner, M., Agnarsson, I. 2006. Are the Linnaean & phylogenetic nomenclatural systems combinable? Recommendations for biological nomenclature. *Systematic Biology*, **55**(5), 774-784.
- Lacépède, B.G.E. 1800. *Histoire naturelle des poissons, par le citoyen La Cèpede*. membre de l'Institut National, et Professeur du Museum d'Histoire Naturelle, Paris. ixiv+632 pp
- Ladner, J.T., Palumbi, S.R. 2012. Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Molecular Ecology*, **21**(9), 2224-2238.
- Lance, R.F., Kennedy, M.L., Leberg, P.L. 2000. Classification bias in discriminant function analyses used to evaluate putatively different taxa. *Journal of Mammalogy*, **81**(1), 245-249.
- Laurin, M., de Queiroz, K., Cantino, P.D., Cellinese, N., Olmstead, R. 2005. The PhyloCode, types, ranks and monophyly. *Cladistics*, **21**, 605-607.
- Le, M., Reid, B.N., McCord, W.P., Naro-Maciel, E., Raxworthy, C.J., Amato, G., Georges, A. 2013. Resolving the phylogenetic history of the short-necked turtles, genera *Elseya* and *Myuchelys* (Testudines: Chelidae) from Australia and New Guinea. *Molecular Phylogenetics and Evolution*, **68**, 251-258.
- Lee, M.S.Y. 2004. The molecularisation of taxonomy. *Invertebrate Systematics*, **18**, 1-6.
- Lei, L., Feng, L., Jian, T.R., Yue, G.H. 2011. Characterization and multiplex genotyping of novel microsatellites from Asian swamp eel, *Monopterus albus*. *Conservation Genetics Resources*, **4**(2), 363-365.
- Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J., Tedesco, P.A. 2007. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, **595**(1), 545-567.
- Li, W.T., Liao, X.L., Yu, X.M., Cheng, L., Tong, J. 2007. Isolation and characterization of polymorphic microsatellites in a sex-reversal fish, rice field eel (*Monopterus albus*). *Molecular Ecology Notes*, **7**(4), 705-707.
- Librado, P., Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**(11), 1451-1452.
- Liem, K.F. 1963. Sex reversal as a natural process in the Synbranchiform fish *Monopterus albus*. *Copeia*, **1963**(2), 303-312.
- Lim, G.S., Balke, M., Meier, R. 2012. Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology*, **61**(1), 165-169.

- Lin, Y., Tzeng, C., Hwang, J. 2005. Reassessment of morphological characteristics in freshwater eels (genus *Anguilla*, Anguillidae) shows congruence with molecular phylogeny estimates. *Zoologica Scripta*, **34**, 225-234.
- Linnaeus, C. 1753. *Species Plantarum*. Laurentii Salvii, Stockholm.
- Linnaeus, C. 1758. *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata. Tomus I*. Laurentii Salvii, Holmiae.
- Liu, J.F., Guiguen, Y., Liu, S.J. 2009. Aromatase (P450arom) and 11beta-hydroxylase (P45011beta) genes are differentially expressed during the sex change process of the protogynous rice field eel, *Monopterus albus*. *Fish Physiology and Biochemistry*, **35**(3), 511-518.
- Liu, L., Wang, W., Zeng, B., Luo, Y., Han, Q. 2005. RAPD analysis of *Monopterus albus* (Zuiew) populations with three colors. *Fisheries Science*, **24**, 22-25.
- Liu, L., Zhou, R. 2001. Isolation of Sox11a, Sox11b and Sox 19 genes from rice field eel (*Monopterus albus*) using degenerate primers and nested PCR. *Aquatic Science*, **63**, 191-195.
- Lu, H., Cheng, H., Yiqing, G.U.O., Zhou, R. 2003. Two alleles of the Sox9a2 in the rice field eel. *Journal of Experimental Zoology*, **299B**, 36-40.
- Lu, S.-Q., Liu, S.-J., Liu, H.J., Liu, Z., Liu, Y. 2005. Screening of microsatellite primer and its application to conservation genetics of *Monopterus albus*. *Journal of Fisheries of China*, **29**, 612-618.
- Lumley, L.M., Sperling, F.A. 2011. Utility of microsatellites and mitochondrial DNA for species delimitation in the spruce budworm (*Choristoneura fumiferana*) species complex (Lepidoptera: Tortricidae). *Molecular Phylogenetics and Evolution*, **58**(2), 232-243.
- Ma, H., Ma, C., Ma, L. 2012. Molecular identification of genus *Scylla* (Decapoda: Portunidae) based on DNA barcoding and polymerase chain reaction. *Biochemical Systematics and Ecology*, **41**, 41-47.
- Maan, M.E., Seehausen, O. 2011. Ecology, sexual selection and speciation. *Ecology Letters*, **14**(6), 591-602.
- Maciel, N.M., Collevatti, R.G., Colli, G.R., Schwartz, E.F. 2010. Late Miocene diversification and phylogenetic relationships of the huge toads in the *Rhinella marina* (Linnaeus, 1758) species group (Anura: Bufonidae). *Molecular Phylogenetics and Evolution*, **57**(2), 787-797.
- Maddison, W.P., Maddison, D.R. 2010. Mesquite: A modular system for evolutionary analysis version 2.74. (<http://www.mesquiteproject.org>).
- Makowsky, R., Cox, C.L., Roelke, C.E., Chippindale, P.T. 2013. The relative utility of sequence divergence and phylogenetic informativeness profiling in phylogenetic study design. *Molecular Phylogenetics and Evolution*, **66**(1), 437.

- Mallet, J. 1995. The species definition for the modern synthesis. *Trends in Ecology and Evolution*, **10**(7), 294-299.
- Matsumoto, S., Fujimoto, H., Takehara, K., Nishida, M., Kohda, M. 2007. Ecology and morph traits of the swamp eel *Monopterus albus* (Synbranchiformes: Synbranchidae) on the Ryukyu Islands, Japan. *Kansai Organ Conservation Nature*, **29**, 5-18.
- Matsumoto, S., Iwata, K. 1997. Paternal egg guarding and mouthbrooding in the bubble nest in the swamp-eel, *Monopterus albus*. *Japan Journal of Ichthyology*, **44**, 35-41.
- Matsumoto, S., Kon, T., Yamaguchi, M., Takeshima, H., Yamazaki, Y., Mukai, T., Kuriwa, K., Kohda, M., Nishida, M. 2010. Cryptic diversification of the swamp eel *Monopterus albus* in East and Southeast Asia, with special reference to the Ryukyuan populations. *Ichthyological Research*, **57**(1), 71-77.
- Mayden, R.L. 1997. A Hierarchy of Species Concepts: The Denouement in The Saga of The Species Problem. in: *Species: The Unit of Biodiversity*, (Eds.) M.F. Claridge, H.A. Dawah, M.R. Wilson, Chapman and Hall. London, pp. 381-424.
- Mayr, E. 2000. The Biological Species Concept. in: *Species Concept and Phylogenetic Theory: A Debate*, (Eds.) Q.D. Wheeler, R. Meier, Columbia University Press. New York, pp. 17-29.
- Mayr, E., Bock, W.J. 2002. Classification and other ordering systems. *Journal of Zoological Systematics and Evolutionary Research*, **40**, 169-194.
- McAuliffe, L., Ellis, R.J., Lawes, J.R., Ayling, R.D., Nicholas, R.A. 2005. 16S rDNA PCR and denaturing gradient gel electrophoresis; a single generic test for detecting and differentiating Mycoplasma species. *Journal of Medical Microbiology*, **54**(Pt 8), 731-739.
- McConnell, S.K.J. 2004. Mapping aquatic faunal exchanges across the Sunda shelf, South-east Asia, using distributional and genetic data sets from the cyprinid fish *Barbodes gonionotus* (Bleeker, 1858). *Journal of Natural History*, **38**, 651-670.
- McCormack, J.E., Huang, H., Knowles, L.L. 2009. Maximum likelihood estimates of species tree: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology*, **58**, 501-508.
- McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**(2), 526-538.
- McNeill, J. 2000. Naming the groups: Developing a stable and efficient nomenclature. *Taxon*, **49**(4), 705-720.

McNeill, J., Barrie, F.R., Burdet, H.M., Demoulin, V., Hawksworth, D.L., Marhold, K., Nicolson, D.H., Prado, J., Silva, P.C., Skog, J.E., Wiersema, J.H., Turland, N.J. 2006. *International Code of Botanical Nomenclature (Vienna Code)*. *Regnum Vegetabile* 146. Gantner, Ruggel.

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L. 2006. DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**(5), 715-728.

Menon, A.G.K. 1999. Freshwater fishes of India. *Records of the Zoological Survey of India, Occasional Paper*(175), i-xxviii + 1-366.

Minos, G., Imsiridou, A., Katselis, G. 2012. Use of morphological differences for the identification of two picarel species *Spicara flexuosa* and *Spicara maena* (Pisces: Centracanthidae). *Mediterranean Marine Science*, **14**(3), 26-31.

Mishler, B.D. 2009. Three centuries of paradigm changes in biological classification: is the end in sight? *Taxon*, **58**(1), 61-67.

Mishler, B.D., Theriot, E.C. 2000. The Phylogenetic Species Concept (sensu Mishler and Theriot): Monophyly, Apomorphy, and Phylogenetic Species Concept. in: *Species Concept and Phylogenetic Theory: A Debate*, (Eds.) Q.D. Wheeler, R. Meier, Columbia University Press. New York, pp. 44-54.

Miya, M., Kawaguchi, A., Nishida, M. 2001. Mitogenomic exploration of higher teleostean phylogenies: A case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, **18**(11), 1993-2009.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N. B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S. M., Nishida, M. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 121-138.

Modica, M.V., Bouchet, P., Cruaud, C., Utge, J., Oliverio, M. 2011. Molecular phylogeny of the nutmeg shells (Neogastropoda, Cancellariidae). *Molecular Phylogenetics and Evolution*, **59**(3), 685-697.

Morin, P.A., Martien, K.K., Archer, F.I., Cipriano, F., Steel, D., Jackson, J., Taylor, B.L. 2010. Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *Journal of Heredity*, **101**(1), 1-10.

Moritz, C., Hillis, D.M. 1996. Molecular Systematics: Context and Controversies. in: *Molecular Systematics*. 2nd. ed., (Eds.) D.M. Hillis, C. Moritz, B.K. Mable, Sinauer Associates, Inc. Sunderland, Massachusetts, USA, pp. 385-394.

Moritz, C., Cicero, C. 2004. DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**(10), 1529-1531.

- Muchlisin, Z.A. 2013. Morphometric variations of *Rasbora* group (Pisces: Cyprinidae) in Lake Laut Tawar, Aceh Province, Indonesia, based on Truss Character Analysis. *HAYATI*, **20**(3), 138-143.
- Murta, A.G. 2000. Morphological variation of horse mackerel (*Trachurus trachurus*) in the Iberian and North Africa Atlantic: implications for stock identification. *ICES Journal of Marine Science*, **57**(4), 1240-1248.
- Myers, R.M., Fischer, S.G., Lerman, L.S., Maniatis, T. 1985. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Research* **13**, 3131-3145.
- Nayar, K.K. 1952. Some burrowing fishes from Travancore. *Proceeding Indian Academy of Science (B)*, **34**, 311-328.
- Neto, R.V.R., De Freitas, R.T.F., Serafini, M.A., Costa, A.C., Freato, T.A., Rosa, P.V., Allaman, I.B. 2012. Interrelationships between morphometric variables and rounded fish body yields evaluated by path analysis. *Revista Brasileira de Zootecnia*, **41**(7), 1576-1582.
- Nguyen, V.H. 2005. Ca Nuoc Ngot Viet Nam. Tap III **Vol. 3**, 1-759.
- Nguyen, T.T.T., Na-Nakorn, U., Sukmanomon, S., ZiMing, C. 2008. A study on phylogeny and biogeography of mahseer species (Pisces: Cyprinidae) using sequences of three mitochondrial DNA gene regions. *Molecular Phylogenetics and Evolution*, **48**, 1223-1231.
- Ni, L., Li, Q., Kong, L., Huang, S., Li, L. 2012. DNA barcoding and phylogeny in the family Mactridae (Bivalvia: Heterodonta): Evidence for cryptic species. *Biochemical Systematics and Ecology*, **44**, 164-172.
- Nichols, J.T. 1943. *The Freshwater of China. Natural History of Central Asia*. The American Museum of Natural History, New York.
- Nico, L.G., Sharp, P., Collins, T.M. 2011. Imported Asian swamp eels (Synbranchidae-*Monopterus*) in North American live food markets: Potential vectors of non-native parasites. *Aquatic Invasions*, **6**(1), 69-76.
- Niklas, K.J., De Queiroz, K., Donoghue, M., Gest, H., Favinger, J., Abbey, D.M. 2001. Taxing debate for taxonomists. *Science*, **292**(5525), 2249-2250.
- Nosil, P., Feder, J.L. 2012. Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society London B Biological Sciences*, **367**(1587), 332-342.
- Novo, M., Fernandez, R., Marchan, D.F., Monica, G., Cosin, D.J. 2012. Compilation of morphological and molecular data, a necessity for taxonomy: The case of *Hormogaster abbatissae* sp. n. (Annelida, Clitellata, Hormogastridae). *Zookeys* **242**, 1-16.

Nunes, V.L., Mendes, R., Marabuto, E., Novais, B.M., Hertach, T., Quartau, J.A., Seabra, S.G., Paulo, O.S., Simoes, P.C. 2014. Conflicting patterns of DNA barcoding and taxonomy in the cicada genus *Tettigetta* from southern Europe (Hemiptera: Cicadidae). *Molecular Ecology Resources*, **14**, 27-38.

Oliver, L.A., Rittmeyer, E.N., Kraus, F., Richards, S.J., Austin, C.C. 2013. Phylogeny and phylogeography of *Mantophryne* (Anura: Microhylidae) reveals cryptic diversity in New Guinea. *Molecular Phylogenetics and Evolution*, **67**(3), 600-607.

Ozaki, A., Yoshida, K., Fuji, K., Kubota, S., Kai, W., Aoki, J., Kawabata, Y., Suzuki, J., Akita, K., Koyama, T., Nakagawa, M., Hotta, T., Tsuzaki, T., Okamoto, N., Araki, K., Sakamoto, T. 2013. Quantitative trait loci (QTL) associated with resistance to a Monogenean parasite (*Benedenia seriolae*) in yellowtail (*Seriola quinqueradiata*) through genome wide analysis. *PLoS One*, **8**(6), 1-14 (e64987).

Packert, M., Martens, J., Hering, J., Kvist, L., Illera, J.C. 2013. Return flight to the Canary Islands-the key role of peripheral populations of Afrocanarian blue tits (Aves: Cyanistes teneriffae) in multi-gene reconstructions of colonization pathways. *Molecular Phylogenetics and Evolution*, **67**(2), 458-467.

Padial, J.M., Miralles, A., De la Riva, I., Vences, M. 2010. The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.

Padian, K. 1999. Charles Darwin's views of classification in theory and practice. *Systematic Biology*, **48**(2), 352-364.

Paetkau, D., Slade, R., Burden, M., Estoup, A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55-65.

Page, R.D.M., Holmes, E.C. 2003. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science Ltd., USA.

Palumbi, S.R. 1996. Nucleic acids II: The Polymerase Chain Reaction. In: *Molecular Systematics*, (Eds.) D.M. Hillis, C. Moritz, B.K. Mable, Sinauer Associates, Sunderland. Massachusetts, pp. 205-247.

Papakostas, S., Doms, S., Triantafyllidis, A., Deloof, D., Kappas, I., Dierckens, K., Wolf, T.D., Bossier, P., Vadstein, O., Kui, S., Sorgeloos, P., Abatzopoulos, T.J. 2006. Evaluation of DNA methodologies in identifying *Brachionus* species used in European hatcheries. *Aquaculture*, **255**(1-4), 557-564.

Patterson, D.J., Remsen, D., Marino, W.A., Norton, C. 2006. Taxonomic indexing-extending the role of taxonomy. *Systematic Biology*, **55**(3), 367-373.

Peakall, R., Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**(19), 2537-2539.

Pellegrin, J. 1922. Sur un nouveau poisson aveugle des eaux douces de l'Afrique occidentale. *Comptes Rendus de Académie des Sciences*, **174**, 884-885.

- Pereira, F., Carneiro, J., Amorim, A. 2008. Identification of species with DNA-based technology: Current progress and challenges. *Recent Patents on DNA and Gene Sequences*, **2**, 187-200.
- Pérez-Losada, M., Bloch, R., Breinholt, J.W., Pfenninger, M., Domínguez, J. 2012. Taxonomic assessment of Lumbricidae (Oligochaeta) earthworm genera using DNA barcodes. *European Journal of Soil Biology*, **48**, 41-47.
- Perdices, A., Doadrio, I., Bermingham, E. 2005. Evolutionary history of the synbranchid eels (Teleostei: Synbranchidae) in Central America and the Caribbean islands inferred from their molecular phylogeny. *Molecular Phylogenetics and Evolution*, **37**(2), 460-73.
- Pickett, K.M. 2005. The new and improved PhyloCode, now with types, ranks, and even polyphyly: a conference report from the First International Phylogenetic Nomenclature Meeting. *Cladistics*, **21**, 79-82.
- Pinheiro, A., Teixeira, C.M., Rego, A.L., Marques, J.F., Cabral, H.N. 2005. Genetic and morphological variation of *Solea lascaris* (Risso, 1810) along the Portuguese coast. *Fishery Research*, **73**(1-2), 67-78.
- Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., Estoup, A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**(6), 536-539.
- Pisani, D., Benton, M.J., Wilkinson, M. 2007. Congruence of morphological and molecular phylogenies. *Acta Biotheoretica*, **55**(3), 269-281.
- Podani, J. 2009. Taxonomy versus evolution. *Taxon*, **58**(4), 1049-1053.
- Posada, D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**(7), 1253-1256.
- Posada, D., Buckley, T.R. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**(5), 793-808.
- Price, T.D., Qvarnstrom, A., Irwin, D.E. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B Biological Sciences*, **270**(1523), 1433-1440.
- Pritchard, J.K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- Puillandre, N., Lambert, A., Brouille, S., Achaz, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**(8), 1864-1877.
- Pujolar, J.M., Bevacqua, D., Capoccioni, F., Ciccotti, E., Leo, G.A., Zane, L. 2011. No apparent genetic bottleneck in the demographically declining European eel using molecular genetics and forward-time simulations. *Conservation Genetics*, **12**(3), 813-825.

- Pusey, B., Kennard, M., Arthington, A. 2004. *Freshwater Fishes of North-Eastern Australia*. Centre for Riverine Landscapes, Griffith University, Queensland, Australia.
- Queiroz, K. 2006. The PhyloCode and the distinction between taxonomy and nomenclature. *Systematic Biology*, **55**(1), 160-162.
- Queiroz, K., Gauthier, J. 1992. Phylogenetic taxonomy. *Annual Review of Ecology and Systematics*, **23**, 449-480.
- Quenouille, B., Bermingham, E., Planes, S. 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **31**(1), 66-88.
- Raadik, T.A. 2014. Fifteen from one: a revision of the *Galaxias olidus* Günther, 1866 complex (Teleostei, Galaxiidae) in south-eastern Australia recognises three previously described taxa and describes 12 new species. *Zootaxa*, **3898**(1), 1-198.
- Rambaut, A. 2012. FigTree v 1.4.0. (<http://tree.bio.ed.ac.uk/software/figtree/>).
- Rasmussen, R.S., Morrissey, M.T., Hebert, P.D. 2009. DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. *Journal of Agricultural and Food Chemistry*, **57**(18), 8379-8385.
- Ratmuangkhwang, S., Musikasinthorn, P., Kumazawa, Y. 2014. Molecular phylogeny and biogeography of air sac catfishes of the *Heteropneustes fossilis* species complex (Siluriformes: Heteropneustidae). *Molecular Phylogenetics and Evolution*, **79**, 82-91.
- Regan, C.T. 1912. The anatomy and classification of the Symbranchoid eels. *The Annals and Magazine of Natural History*, **9**, 387-390.
- Reid, G.M. 2009. Carolus Linnaeus (1707-1778): His life, philosophy and science and its relationship to modern biology and medicine. *Taxon*, **58**(1), 18-31.
- Renaud, A.K., Savage, J., Adamowicz, S.J. 2012. DNA barcoding of Northern Neractic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. *BMC Ecology*, **12**, 1-15.
- Rhee, S., Kitchener, D., Brown, T., Merrill, R., Dilts, R., Tighe, S. 2004. Report on Biodiversity and Tropical Forest in Indonesia USAID.
- Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**(1), 223-225.
- Richardson, B.J., Baverstock, P.R., Adam, M. 1986. *Allozyme Electrophoresis*. Academic Press, Australia.
- Ride, W.D.L., Cogger, H.G., Dupuis, C., Kraus, O., Minelli, A., Thompson, F.C., Tubbs, P.K. 1999. *International Code of Zoological Nomenclature. 4th edition*. The International Trust for Zoological Nomenclature, London.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**(3), 539-542.

Rosen, D.E. 1975. A vicariance model of Caribbean biogeography. *Systematic Zoology*, **24**(4), 431-464.

Rosen, D.E., Rumney, A. 1972. Evidence of a second species of *Synbranchus* (Pisces, Teleostei) in South America. *The American Museum of Natural History*, **2497**, 1-45.

Rosen, E.R., Greenwood, P.H. 1976. A fourth neotropical species of synbranchid eel and the phylogeny and systematics of synbranchiform fishes. *Bulletin of the American Museum of Natural History*, **157**(1), 1-70.

Roth, W.M. 2005. Making classifications (at) work: Ordering practices in science. *Social Studies of Science*, **35**(4), 581-621.

Roy, R. 2013. *Budidaya Belut*. P.T. AgroMedia Pustaka, Jakarta, Indonesia.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**(1), 103-106.

Rubinoff, D., Cameron, S., Will, K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity*, **97**(6), 581-594

Sales, J.B., Shaw, P.W., Haimovici, M., Markaida, U., Cunha, D.B., Ready, J., Figueiredo-Ready, W.M., Schneider, H., Sampaio, I. 2013. New molecular phylogeny of the squids of the family Loliginidae with emphasis on the genus *Doryteuthis* Naef, 1912: Mitochondrial and nuclear sequences indicate the presence of cryptic species in the southern Atlantic Ocean. *Molecular Phylogenetics and Evolution*, **68**, 293-299.

Saltzgiver, M.J., Heist, E.J., Hedrick, P.W. 2012. Genetic evaluation of the initiation of a captive population: the general approach and a case study in the endangered pallid sturgeon (*Scaphirhynchus albus*). *Conservation Genetics*, **13**(5), 1381-1391.

Samaradivakara, S.P., Hirimuthugoda, N.Y., Gunawardana, R.H.A.N.M., Illeperuma, R.J., Fernandopulle, N.D., De Silva, A.D., Alexander, P.A.B.D. 2012. Morphological variation of four *Tilapia* population in selected reservoirs in Sri Lanka. *Tropical Agricultural Research*, **23**(2), 105-116.

Santini, F., Sorenson, L., Marcroft, T., Dornburg, A., Alfaro, M.E. 2013. A multilocus molecular phylogeny of boxfishes (Araucanidae, Ostraciidae; Tetraodontiformes). *Molecular Phylogenetics and Evolution*, **66**(1), 153-160.

Saparinto, C. 2009. *Panduan Lengkap Belut*. Penebar Swadaya, Jakarta, Indonesia.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, **55**, 421-438.

Schonhuth, S., Hillis, D.M., Neely, D.A., Lozano-Vilano, L., Perdices, A., Mayden, R.L. 2012. Phylogeny, diversity, and species delimitation of the North American Round-Nosed Minnows (Teleostei: Dionda), as inferred from mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **62**(1), 427-446.

Schuh, R.T. 2000. *Biological Systematics*. Cornell University Press, New York.

Schuh, R.T. 2003. The Linnaean system and its 250 year persistence. *Botanical Review*, **69**(1), 59-78.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233-234.

Scribner, K.T., Page, K.S., Bartron, M.L. 2001. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, **10**, 293-323.

Sekino, M., Saitoh, K., Shimizu, D., Wada, T., Kamiyama, K., Gambe, S., Chen, S., Aritaki, M. 2010. Genetic structure in species with shallow evolutionary lineages: a case study of the rare flatfish *Verasper variegatus*. *Conservation Genetics*, **12**(1), 139-159.

Sembiring, A., Pertiwi, N.P.D., Mahardini, A., Wulandari, R., Kurniasih, E.M., Kuncoro, A.W., Cahyani, N.K.D., Anggoro, A.W., Ulfa, M., Maddupa, H., Carpenter, K.E., Barber, P.H., Mahardika, G.N. 2015. DNA barcoding reveals targeted fisheries for endangered sharks in Indonesia. *Fisheries Research*, **164**, 130-134.

Senar, J.C., Leonart, J., Metcalfe, N.B. 1994. Wing-shape variation between resident and transient wintering Siskins *Carduelis spinus* *Journal of Avian Biology*, **25**(1), 50-54.

Sereno, P.C. 2005. The logical basis of phylogenetic taxonomy. *Systematic Biology*, **54**(4), 595-619.

Serrao, N.R., Steinke, D., Hanner, R.H. 2014. Calibrating snakehead diversity with DNA barcodes: Expanding taxonomic coverage to enable identification of potential and established invasive species. *PLoS One*, **9**(6), 1-13 (e99546).

Shafland, P.L., Gestring, K.B., Stanford, M.S. 2010. An assessment of the Asian swamp eel (*Monopterus albus*) in Florida. *Reviews in Fisheries Science*, **18**(1), 25-39.

Shaffer, H.B., Thomson, R.C. 2007. Delimiting species in recent radiations. *Systematic Biology*, **56**(6), 896-906.

Sheffield, V.C., Cox, D.R., Lerman, L.S., Myers, R.M. 1989. Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academic of Sciences U.S.A.*, **86**, 232–236.

Shlens, J. 2009. *A Tutorial on Principal Component Analysis*. Center for Neural Science, New York University, New York, USA.

Silas, E.G., Dawson, E. 1961. *Amphipnous indicus*, a new synbranchoid eel from India, with a redefinition of the genus and a synopsis to the species of *Amphipnous* Muller. *The Journal of Bombay Natural History Society*, **58**(2), 366-378.

Silva, A. 2003. Morphometric variation among sardine (*Sardina pilchardus*) populations from the northeastern Atlantic and the western Mediterranean. *ICES Journal of Marine Science*, **60**, 1352-1360.

Silva, G.S., Melo, B.F., Oliveira, C., Benine, R.C. 2013. Morphological and molecular evidence for two new species of *Tetragonopterus* (Characiformes: Characidae) from central Brazil. *Journal of Fish Biology*, **82**(5), 1613-1631.

Simon, K.D., Bakar, Y., Temple, S.E., Mazlan, A.G. 2010. Morphometric and meristic variation in two congeneric archer fishes *Toxocates chatareus* (Hamilton 1822) and *Toxotes jaculatrix* (Pallas 1767) inhabiting Malaysian coastal waters. *Journal of Zhejiang University-SCIENCE B (Biomedicine and Biotechnology)*, **11**(11), 871-879.

Singh, R.K., Lal, K.K., Mohindra, V., Punia, P., Sah, R.S., Mishra, A.K., Kumar, R., Mishra, B.N., Lakra, W.S. 2010. Assessing genetic differentiation in geographic populations of *Labeo calbasu* using allozyme markers. *Biochemical Genetics*, **48**(9-10), 760-778.

Sneath, P.H.A. 1995. Thirty years of numerical taxonomy. *Systematic Biology*, **44**(3), 281-298.

Sneath, P.H.A., Sokal, R.R. 1973. *Numerical Taxonomy*. W.H. Freeman, San Fransisco (USA).

So, N., Maes, G.E., Volckaert, F.A. 2006. High genetic diversity in cryptic populations of the migratory sutchi catfish *Pangasianodon hypophthalmus* in the Mekong River. *Heredity (Edinb)*, **96**(2), 166-174.

Stevens, P.F. 2002. Why do we name organisms-some reminders from the past. *Taxon*, **51**(1), 11-26.

Stuessy, T.F. 2009. Paradigms in biological classification (1707-2007): Has anything really changed? *Taxon*, **58**(1), 68-76.

Swainson. 1838. *The Natural History and Classification of Fishes, Amphibians, and Reptiles, or Monocardian Animals*, London.

Swofford, D.L. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts, USA.

Takahashi, T., Hori, M. 2012. Genetic and morphological evidence implies existence of two sympatric species in *Cyathopharynx furcifer* (Teleostei: Cichlidae) from Lake Tanganyika. *International Journal of Evolutionary Biology*, **2012**, 1-7.

Talwar, P.K., Jhingran, A.G. 1992. *Inland Fishes of India and Adjacent Countries*. A.A. Balkema, Rotterdam.

Tajima, F. 1989. The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597-601.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**(10), 2731-2739.

Tancioni, L., Russo, T., Cataudella, S., Malana, V., Hett, A.K., Corsi, E., Rossi, A.R. 2013. Testing species delimitations in four Italian sympatric leuciscine fishes in the Tiber River-A combined morphological and molecular approach. *PLoS One*, **8**(4), 1-10 (e60392).

Tang, P.Y., Sivananthan, J., Pillay, S.O., Muniandy, S. 2004. Genetic structure and biogeography of Asian arowana (*Scleropages formosus*) determined by microsatellite and mitochondrial DNA analysis. *Asian Fisheries Science*, **17**, 81-92.

Thai, B.T., Pham, A.T., Austin, C.M. 2006. Genetic diversity of common carp in Vietnam using direct sequencing and SSCP analysis of the mitochondrial DNA control region. *Aquaculture* **258**, 228-240.

Thaulow, J., Borgstrøm, R., Heun, M. 2012. Brown trout population structure highly affected by multiple stocking and river diversion in a high mountain national park. *Conservation Genetics*, **14**(1), 145-158.

Thornhill, D.J., Kemp, D.W., Sampayo, E.M., Schmidt, G.W. 2009. Comparative analyses of amplicon migration behavior in differing denaturing gradient gel electrophoresis (DGGE) systems. *Coral Reefs*, **29**(1), 83-91.

Thorpe, R.S. 1976. Biometric analysis of geographic variation and racial affinities. *Biological Reviews of the Cambridge Philosophical Society*, **51**, 407-452.

Torres, R.A. 2000. O genero *Synbranchus*: interrelacoes citotaxonomicas, evolutivas e natureza da variabilidade cariotipica, Vol. Doctoral Universidade Estadual Paulista, Instituto Biociencias, Campus de Botucatu, SP. Brazil.

Valdez-Moreno, M., Ivanova, N.V., Elias-Gutierrez, M., Contreras-Balderas, S., Hebert, P.D. 2009. Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. *Journal of Fish Biology*, **74**(2), 377-402.

- Van de Peer, Y. 2009. Phylogenetic Inference Based on Distance Methods. in: *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, (Eds.) P. Lemey, M. Salemi, A.-M. Vandamme, Cambridge University Press. UK, pp. 142-180.
- Vanhaecke, D., Garcia de Leaniz, C., Gajardo, G., Young, K., Sanzana, J., Orellana, G., Fowler, D., Howes, P., Monzon-Arguello, C., Consuegra, S. 2012. DNA barcoding and microsatellites help species delimitation and hybrid identification in endangered galaxiid fishes. *PLoS One*, **7**(3), 1-10 (e32939).
- Von Rintelen, K., Page, T.J., Cai, Y., Roe, K., Stelbrink, B., Kuhajda, B.R., Iliffe, T.M., Hughes, J., von Rintelen, T. 2012. Drawn to the dark side: a molecular phylogeny of freshwater shrimps (Crustacea: Decapoda: Caridea: Atyidae) reveals frequent cave invasions and challenges current taxonomic hypotheses. *Molecular Phylogenetics and Evolution*, **63**(1), 82-96.
- Voris, H.K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153-1167.
- Wallis, Y. 2002. Denaturing gradient gel electrophoresis. *Methods in Molecular Biology*, **187**, 125-135.
- Wang, R., Cheng, H., Xia, L., Guo, Y., Huang, X., Zhou, R. 2003. Molecular cloning and expression of Sox17 in gonads during sex reversal in the rice field eel, a teleost fish with a characteristic of natural sex transformation. *Biochemical and Biophysical Research Communications*, **303**(2), 452-457.
- Ward, R.D., Grewe, P.M. 1994. Appraisal of Molecular Genetic Techniques in Fisheries. in: *Molecular Genetics in Fisheries*, (Ed.) T.J. Pitcher, Chapman and Hall. UK, pp. 29-54.
- Ward, R.D., Zemplak, T.S., Innes, B.H., Last, P.R., Hebert, P.D. 2005. DNA barcoding Australia's fish species. *Philosophical Transaction of the Royal Society B Biological Sciences*, **360**(1462), 1847-1857.
- Warheit, K.I. 1992. The role of morphometrics and cladistics in the taxonomy of fossils-a paleornithological example. *Systematic Biology*, **41**(3), 345-369.
- Wayne, L.G. 1981. Numerical taxonomy and cooperative studies: roles and limits. *Reviews of Infectious Diseases*, **3**(5), 822-828.
- Weissing, F.J., Edelaar, P., van Doorn, G.S. 2011. Adaptive speciation theory: a conceptual review. *Behavioral Ecology and Sociobiology*, **65**(3), 461-480.
- Wiens, J.J., Penkrot, T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic Biology*, **51**(1), 69-91.
- Will, K.W., Mishler, B.D., Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**(5), 844-851.

- Wilson, E.O. 2004. Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society London B Biological Sciences*, **359**(1444), 739.
- Winsor, M.P. 2009. Taxonomy was the foundation of Darwin's evolution. *Taxon*, **58**(1), 43-49.
- White, W.T., Last, P.R. 2012. A review of the taxonomy of chondrichthyan fishes: a modern perspective. *Journal of Fish Biology*, **80**(5), 901-917.
- Whitworth, T.L., Dawson, R.D., Magalon, H., Baudry, E. 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of The Royal Society B Biological Science*, **274**, 1731-1739.
- Wheeler, Q.D. 2008. Introductory: Toward the New Taxonomy. in: *The New Taxonomy, Systematics Association Special Volume Series* (Ed.) Q.D. Wheeler, CRC Press, Boca Raton, FL, in press.
- Wheeler, W.C., Lucaroni, N., Hong, L., Crowley, L.M., Varon, A. 2015. POY version 5: phylogenetics analysis using dynamic homologies under multiple optimality criteria. *Cladistics*, **31**, 189-196.
- Wu, C. 2001^a. The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**(6), 851-865.
- Wu, C. 2001^b. Genes and speciation. *Journal of Evolutionary Biology*, **14**(6), 889-891.
- Xia, L., Cheng, H., Yu, H., Guo, Y., Zhou, R. 2004. Molecular cloning and expression of the osteoclast-stimulating-factor-like gene from the rice field eel. *Journal Experimental Zoology Part B Molecular and Developmental Evolution*, **302**(2), 174-181.
- Yamaguchi, M., Miya, M., Okiyama, M., Nishida, M. 2000. Molecular phylogeny and larval morphological diversity of the lanternfish genus *Hypogomum* (Teleostei: Myctophidae). *Molecular Phylogenetics and Evolution*, **15**, 103-114.
- Yang, F., Zhou, Q., Zhang, Y., Li, X. 2005. Variation of esterase isoenzyme among three kinds of bodily color *Monopterus albus* in Poyang Lake region. *Journal of Economic Animal*, **9**, 110-113.
- Yang, Z., Rannala, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceeding of the National Academy of Sciences of the USA*, **107**, 9264-9269.
- Ye, D., Lv, D., Song, P., Peng, M., Chen, Y., Guo, M., Yang, Q., Hu, Y. 2007. Cloning and characterization of a rice field eel vasa-like gene cDNA and its expression in gonads during natural sex transformation. *Biochemical Genetics*, **45**(3-4), 211-224.
- Yilmaz, C., Turkozan, O., Bardakci, F. 2011. Genetic structure of loggerhead turtle (*Caretta caretta*) populations in Turkey. *Biochemical Systematics and Ecology*, **39**(4-6), 266-276.

- Yin, S., Zhou, G., Liu, Y. 2005. Population genetic structure of rice field eel (*Monopterus albus*) with RAPD markers. *Chinese Journal of Applied and Environmental Biology*, **11**, 328-332.
- Yu, H., Cheng, H., Guo, Y., Xia, L., Zhou, R. 2003. Alternative splicing and differential expression of P450c17 (CYP17) in gonads during sex transformation in the rice field eel. *Biochemical and Biophysical Research Communications*, **307**(1), 165-171.
- Yu, J.H., Tang, Y.K., Li, J.L. 2008. Cloning, structure, and expression pattern of the P-450 aromatase gene in rice field eel (*Monopterus albus*). *Biochemical Genetics*, **46**(5-6), 267-280.
- Yue, G.H. 2013. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries*, **15**(3), 376-396.
- Zemlak, T.S., Ward, R.D., Connell, A.D., Holmes, B.H., Hebert, P.D. 2009. DNA barcoding reveals overlooked marine fishes. *Molecular Ecology Resources*, **9 Suppl s1**, 237-242.
- Zhang, C., Zhang, D.-X., Zhu, T., Yang, Z. 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747-761.
- Zhou, R., Cheng, H., Zhang, Q., Guo, Y., Cooper, R.K., Tiersch, T.R. 2002. SRY-related genes in the genome of the rice field eel (*Monopterus albus*). *Genetics Selection Evolution*, **34**(1), 129-137.
- Zhou, R., Liu, L., Guo, Y., Yu, H., Cheng, H., Huang, X., Tiersch, T.R., Berta, P. 2003. Similar gene structure of two Sox9a genes and their expression patterns during gonadal differentiation in a teleost fish, rice field eel (*Monopterus albus*). *Molecular Reproduction and Development* **66**(3), 211-217.
- Zhu, S., Fu, J., Wang, Q., Li, J. 2013. Identification of *Channa* species using the partial cytochrome c oxidase subunit I (COI) gene as a DNA barcoding marker. *Biochemical Systematics and Ecology*, **51**, 117-122.
- Zuiew, B. 1753. Biga muraenarum, novae species descriptae. *Nova Acta Academiae Scientiarum Petro-politanae*, **7**, 296-301.