THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Sciences de l'Evolution

École doctorale 231 GAIA

Unité de recherche 5554 ISE-M

Characterization of Sundaland ichthyofauna through DNA barcodes: A case study in Java island

Présentée par Hadi DAHRUDDIN

Le 12 Décembre 2019

Sous la direction de Jean-François AGNESE (Directeur de thèse)
et Nicolas HUBERT (Co-directeur)

Devant le jury composé de

Anne CHENUIL, Directrice de recherche, CNRS	Rapporteur
Bernard HUGUENY, Directeur de recherche, IRD	Rapporteur
Philippe KEITH, Professeur, MNHN	Président du jury
Sophie ARNAUD-HAOND, Chargée de Recherche, IFREMER	Examinateur
Jean-François AGNESE, Directeur de Recherche, IRD	Directeur de thèse
Nicolas HUBERT, Directeur de Recherche, IRD	Co-directeur

UNIVERSITÉ DE MONTPELLIER

Characterization of Sundaland ichthyofauna through DNA barcodes: A case study in Java island

Caractérisation de l'ichtyofaune de Sundaland par l'approche des code-barres ADN : L'exemple de l'ile de Java

Acknowledgments

The realization of this thesis was made possible thanks to funding from the Institut de Recherche pour le Développement (IRD), the French Embassy in Indonesia with the support from the Institut des Sciences de l'Evolution de Montpellier (ISE-M) and also with permission and support of the Research Center of Biology (RCB) – Indonesian Institute of Sciences (LIPI). Many thanks to Nicolas Hubert, Frederic Busson, Philippe Keith, Sopian Sauri, Aditya Hutama, Ujang Nurhaman, and Sumanta for help, support and friendship during field sampling, thanks to Bambang Dwisusilo for processing specimen images and Jean-Paul Toutain and Edmond Dounias as the successive representative of IRD Indonesia for their support. Thanks a lot to Jean Francois Agnese and Nicolas Hubert for supervising this thesis.

Many thanks to the late Renny Kurnia Hadiaty and Mohamad Rofik Sofyan for their support. Also thanks to Daisy Wowor, Rosichon Ubaidillah, Hari Sutrisno, Cahyo Rahmadi, Gono Semiadi and Wirdateti in Zoology division, RCB-LIPI.

For the first and the second year in Indonesia, thanks to members of the Ichthyology laboratory (Haryono, Gema Wahyudewantoro, Yayat Priyatna, Ilham Vemandra Utama) in Zoology division, RCB-LIPI, thank you for your help in preserving and curating the specimens. Also, thanks to the staff of Genetics laboratory for their help and support when processing samples in the laboratory and also members of Reproduction laboratory. The second and third year in Montpellier, thanks to Arni Sholihah and Erwan Delrieu-Trottin for their help and insightfull discussions.

Finally, I am very grateful to the destiny from ALLOH SWT. for this achievement because conducting a doctoral project is a purpose since high school. Thank you very much for Hj. Rosidah (mother), Kundang and Cucu (parents in law), Maryati (wife), Aditya Ramadhan D. and Halya Khairunnisa D. (son and daughter), Mohamad Sofyan, Yani Maryani, Dede Nurjaya (brothers and sister), brothers and sisters in law with nieces and nephews thanks to

suported, raising their hand (prayer), and being patient to live the life when I was in France for 6.5 months on 2018 and 7.5 months on 2019. I will never forget to say thank you very much to my father, the late H. Muslim bin H. Kosim who taught the struggle of life also gave education and support up to high school. May ALLOH SWT. bless you, my father.

Abstract

The Indonesian archipelago hosts 1218 freshwater fish species disseminated across 14,000 islands. Encompassing three majors geographic assemblages (Sundaland, Wallacea, Sahul) separated by two majors faunistic transitions (Wallace and Lyddeker lines), Indonesian islands display heterogeneous levels of species richness resulting from diverse geological and paleoecological histories. Sundaland itself hosts 68% of the total number of freshwater fish species and constitutes one of the world's most endangered fauna worldwide. By contrast with Wallacea that results from an early settlement through subduction around 40 Mya, Sundaland (Borneo, Sumatra and Java) has acquired its modern configuration during the last 5 Mya through a combination of continental fragmentation and subduction. The alarming state of Sundaland ichthyodiversity, combined with major taxonomy and distribution knowledge gaps, urges for a modern reappraisal through standardized DNA-based methods. The ichtyodiversity of Java in particular, is the most threatened and the less known of Sundaland. This dissertation aims at addressing two main questions: (1) Is DNA barcoding a suitable approach to characterize the ichthyodiversity of Java? (2) Is the geological and paeloecological history of Java a good predictor of diversity patterns and population genetic structure? The main results evidence: (1) large discrepancies between the checklist of the Java freshwater fishes based on historical records and a modern re-appraisal through DNA barcodes. Reasons invoqued are the taxonomic bias related to the interrupted inventory of Java ichthyofauna during the last 3 centuries and the rarefaction of several species targeted by artisanal fisheries. (2) A DNA-based reappraisal of species boundaries and distribution for the genera *Nemacheilus* and *Rasbora* indicated two new taxa, several cases of cryptic diversity and several cases of wrong assignement of populations to the species levels. Species range distributions appear to be much more restricted than previously thoughts and question the persistence of these species in changing landscapes. (3) A DNA-based assessment through DNA barcodes of the population genetic structure of three widespread species in Java evidences high levels of cryptic diversity and deep genetic divergences among geographically restricted and non-overlapping mitochondrial lineages. Consistent with a fragmentation related to the rise of volcanic arches in Java that prompted a long-term declines of historical effective population size, this pattern argue for the sensitive conservation status of these mitochondrial lineages. The results presented here highlights the benefits of using a standardized DNA-based approach for the fast characterization of a poorly known fauna and open new perspectives in the conservation of the ichtyofauna of Java and Bali.

Résumé en Français

Introduction- L'archipel Indonésien se situe à la pointe Sud de l'Asie du Sud-Est et constitue le plus grand archipel au monde avec près de 14,000 iles. Il est constitué de trois grands ensembles géographiques : (1) Sundaland, comprenant les îles de Java, Sumatra, Bornéo et Bali, il appartient au plateau de la Sonde, (2) Wallacea, comprenant les îles isolées de Sulawesi et des Moluques, entourées de mers profondes, (3) Sahul, qui comprend l'île de Papouasie et qui correspond au plateau de Sahul. Ces grands ensembles géographiques hébergent des faunes et flores distinct et deux principal zone de démarcation faunistiques ont été identifiées : (1) la ligne de Wallace séparant le plateau de la Sonde de Wallacea, correspondant à la transition faunistique entre Bornéo et Sulawesi et associée au détroit de Makassar ; (2) la ligne de Lydekker séparant Wallacea du plateau de Sahul, correspondant à la transition faunistique entre les Moluques et la Papouasie et associée à la mer de Seram.

L'Asie du Sud-Est héberge près de 3107 espèces valides de poissons vivant dans les rivières, estuaires et mangroves parmi lesquelles 1218 sont présentent en Indonésie. Cette diversité est distribuée de façon très hétérogène dans l'archipel. Sundaland par exemple héberge près de 75% de cette diversité avec 899 espèces contre 184 pour Wallacea et 255 pour Sahul. L'endémisme est lui aussi répartie de façon hétérogène puisque Sundaland contribue à hauteur de 68% au nombre d'espèces endémique de l'archipel contre 13% à Wallacea et 20% à Sahul. Les taux d'endémisme sont toutefois comparables entre grands ensemble puisque 48% des espèces de Sundaland sont endémiques pour 45% à Wallacea et 49% à Sahul. Le nombre de description de nouvelles espèces est en forte augmentation depuis 3 décennies indiquant que la diversité ichtyologique des eaux douces de l'archipel reste sous-estimée.

Cette forte diversité s'explique en partie par une grande diversité d'histoire géologique et paléo-écologique entre ces trois grands ensembles géographiques. La formation de l'archipel a débuté il y a 60 Ma au travers de la tectonique des plaques Asiatique et Australienne ayant résulté à une forte activité de subduction. Les iles de Wallacea ont émergées de la mer entre 40 et 25 Ma par subduction. La plateau de la sonde en revanche s'est formé beaucoup plus récemment par isolement de Bornéo du continent à partir de 20 Ma puis émergence d'ile de Sumatra par subduction vers 10 Ma puis Java vers 5 Ma. La formation du Plateau de la Sonde dans sa configuration actuelle est très récente, les iles de Bornéo, Sumatra et Java ayant été connectée entre elles et au continent jusqu'au début du Pléistocène. La faible élévation du plateau de la Sonde a aboutit par la suite à des interactions

entre géologie et paléoclimats du Pléistocène. En effet, lors des cycles glaciaires, le plateau de la Sonde s'est retrouvé régulièrement exondé, du fait de la faible profondeur de la mer de Java, lors des maximums glaciaires. Ainsi, lors des plus importants maximums glaciaires ayant entrainé des baisses de 120m du niveau de la mer, les iles du plateau de la Sonde ont fusionné avec le continent Asiatique pour former une grande masse de terres émergées. Les reconstructions paléoenvironnementalles du Pléistocène prédisent alors la formation de grandes paléorivières qui ont connectés de façon variable les bassins versants de Bornéo, Sumatra et Java. L'hétérogénéité de la distribution de la richesse ichtyologique de l'archipel est ainsi le résultat de la taille, âge et connexion au continent des différentes iles.

Objectifs de la thèse- L'Indonésie héberge une biodiversité remarquable et représente le l'un des densités d'espèce de poissons d'eau douce les plus élevées au monde après le Brésil et la République du Congo. Cette diversité est toutefois extrêmement menacée par une multitude de perturbations anthropiques liées à la conversion des paysages pour l'agriculture (déforestation) et le développement exponentiel des zones urbaines. Sundaland et Wallacea ont ainsi été labellisé 'points chauds de biodiversité' en raison du grand nombre d'espèce endémique et de l'importance des menaces d'origine anthropique. Sundaland constitue actuellement le point chaud le plus menacé au monde en raison des taux records de croissance des menaces d'origine anthropiques. Java illustre parfaitement les enjeux de conservation dans cette partie du monde. Avec une population de plus de 140 millions d'habitants, soit près de la moitié de la population de l'archipel, sur une île de 130 000 km2, l'île de Java a été la plus touchée par le développement économique rapide de l'Indonésie. Impactée par l'expansion des espèces envahissantes et la surpêche, l'ichtyofaune de Java a connu un déclin spectaculaire au cours des dernières décennies tout en attirant beaucoup moins l'attention en termes d'explorations ichtyologiques. Des études récentes ont démontré, par exemple, qu'un réexamen minutieux des limites des espèces au moyen de méthodes basées sur l'ADN permettait la détection de nouvelles espèces parmi des complexes d'espèces étroitement apparentées. Cette situation est particulièrement vraie pour les groupes d'espèces dépourvus de taxonomistes et dont la taxonomie n'est accessible qu'à quelques spécialistes dans le monde, mettant ainsi en doute la durabilité des connaissances taxonomiques et compromettant les efforts de conservation.

Cette thèse a pour objectif de combler le déficit de connaissances taxonomiques sur l'ichtyofaune de Java par la ré-évaluation de l'ichtyodiversité de l'ile par une approche standardisée basée sur l'ADN. À cet égard, les code-barres ADN (i.e. l'utilisation de 650

paire de base du gène mitochondriale de la cytochrome oxidase I comme marqueur des espèces) ouvrent de nouvelles perspectives en permettant d'examiner le statut biologique des espèces nominales à Java et en donnant accès à l'identification des espèces par tous les scientifiques, quel que soit l'état du matériel biologique à identifier. La recherche présentée dans cette thèse s'adresse aux décideurs et aux responsables gouvernementaux impliqués dans la conservation et la gestion des poissons d'eau douce de Java. Cette thèse à ainsi pour objectif de répondre à deux questions :

(1) L'approche des code-barres ADN est-elle effective pour caractériser l'ichtyodiversité de Java ? Il est question notamment d'évaluer la capacité de cette approche à capturer les limites d'espèces en vue de l'identification automatisée des espèces et revisiter les contours des espèces définis par des approches morphologiques traditionelles.

(2) Est-ce que l'histoire géologique de Java permet d'expliquer les patrons de diversité chez les poissons de Java ? L'ile de Java résulte de la fusion de deux arc volcaniques et son réseau hydrographique a été connecté de façon chronique à ceux du Sud de Bornéo lors des maximums glaciaires. Les structures populationnelles de plusieurs espèces et distributions d'espèces proches seront examinées à la lumière de ces contraintes historiques.

Dans ce contexte, les objectifs suivants ont été identifiés :

(1) Objectif 1: revisiter l'ichtyodiversité de Java au moyen de codes à barres ADN et évaluer l'utilité d'une bibliothèque de référence de codes à barres pour des identifications moléculaires automatisées plus poussées.

(2) Objectif 2: examiner la validité des espèces nominales à Java et affiner la connaissance de la répartition de leur aire de répartition pour le complexe d'espèces Rasbora spp. et Nemacheilus spp. à travers les codes barres de l'ADN

(3) Objectif 3: identifier le caractère commun de la structure de la population de multiples espèces co-distribuées à Java afin d'identifier des unités de conservation.

La librairie de code-barres ADN des poissons de Java et Bali (Article 1)- Parmi les 899 espèces de poissons d'eau douce recensées dans le point chaud de biodiversité de Sundaland, près de 50% sont endémiques. L'intégrité fonctionnelle des écosystèmes aquatiques est actuellement compromise par les activités humaines et la conversion des paysages a entraîné le déclin des populations de poissons dans plusieurs parties du Sundaland, en particulier à Java. L'inventaire de l'ichtyofaune javanaise a été discontinu et les connaissances taxonomiques sont dispersées dans la littérature. Cette étude fournit une librairie de référence

de code-barres ADN pour les poissons de l'intérieur des terres de Java et de Bali, dans le but de rationaliser l'inventaire des poissons de cette partie du Sundaland. Faute de liste de référence disponible pour l'estimation de la couverture taxonomique de cette étude, une liste de référence a été établie à partir de catalogues en ligne. Au total, 95 sites ont été visités et une bibliothèque comprenant 1046 code-barres ADN pour 159 espèces a été constituée. La distance au plus proche voisin était en moyenne 28 fois plus élevée que la distance intraspécifique maximale, un « barcoding gap » a été observée. La liste des espèces établie par les codes-barres ADN présente de grandes différences par rapport à la liste de référence compilée ici: seulement 36% (soit 77 espèces) et 60% (soit 24 espèces) des espèces connues ont été échantillonnées à Java et à Bali, respectivement. Ce résultat contraste avec le nombre élevé de nouvelles occurrences et le plafond des courbes d'accumulation pour les espèces et les genres. Ces résultats mettent en évidence la faible connaissance taxonomique de cette ichtyofaune et le décalage apparent entre les données d'occurrence actuelles et historiques doit être attribué à la disparition d'espèces, à la synonymie et aux erreurs d'identification dans les études précédentes.

Caractérisation des contours et de la distribution des espèces de Rasbora spp. et Nemacheilus spp. à Java et Bali par code-barres ADN (Article 2)- Les points chauds de biodiversité ont fourni des indicateurs géographiques utiles pour les efforts de conservation. Délimités à partir de quelques groupes d'animaux et de plantes, les points chauds de biodiversité ne reflètent pas l'état de conservation des poissons d'eau douce. Avec des centaines de nouvelles espèces décrites chaque année, les poissons constituent le groupe de vertébrés le plus mal connu. Cette situation appelle à une accélération de l'inventaire des espèces de poissons grâce à des outils moléculaires rapides et fiables tels que les code-barres de l'ADN. La présente étude porte sur la diversité des poissons d'eau douce dans le point-chaud de biodiversité de Sundaland en Asie du Sud-Est. Des études récentes ont mis en évidence d'importants écarts entre connaissances taxonomiques historiques et acutlles, ainsi que des niveaux inattendus de diversité cryptique, en particulier dans les îles de Java et de Bali. Les genres Cypriniformes *Rasbora* et *Nemacheilus* représentent la plupart des espèces endémiques de Java et de Bali, mais leur taxonomie est entachée de confusion quant à leur identité et à leur distribution. Cette étude examine le statut taxonomique des espèces de *Rasbora* et *Nemacheilus* dans les îles Java, Bali et Lombok à l'aide de codes à barres ADN, dans le but de dissiper la confusion taxonomique et d'identifier les tendances en matière de diversité génétique pouvant être utilisées ultérieurement pour des questions de conservation.

Plusieurs méthodes de délimitation des espèces basées sur des séquences d'ADN ont été utilisées et ont confirmé le statut de la plupart des espèces, mais plusieurs cas de confusion taxonomique et deux nouveaux taxons ont été détectés. Les séquences mitochondriales expliquent que la plupart des distributions d'espèces actuellement répertoriées dans la littérature sont gonflées en raison d'attributions erronées de populations au niveau de l'espèce, et mettent en évidence le statut de conservation sensible de la plupart des espèces de *Rasbora* et de *Nemacheilus* en raison de leur distribution restreinte sur les îles de Java, Bali et Lombok.

Génétique de la conservation des poissons de Java et Bali (Article 3)- La délimitation d'unités évolutives significatives à des fins de conservation est une étape cruciale de la conservation. Dans toute l'aire de répartition, les espèces présentent fréquemment une structure de population qui détermine la répartition de la diversité génétique. Ces modèles de structure et de diversité génétiques résultent d'interactions complexes entre l'histoire biogéographique et la dynamique démographique. Cependant, les connaissances biogéographiques antérieures sont rarement disponibles, une tendance particulièrement marquée sous les tropiques où l'obstacle taxonomique entrave les études biogéographiques et les efforts de conservation. Les code-barres ADN ont été initialement proposé pour favoriser les études taxonomiques grâce au développement d'un système automatisé d'identification moléculaire des espèces. Bien que son utilité pour l'identification des espèces soit de plus en plus reconnue, son utilité pour la délimitation rapide et à grande échelle d'unités évolutives significatives reste à explorer. S'ils s'avèrent utiles à cette fin, les code-barres ADN pourraient également ouvrir de nouvelles perspectives en matière de conservation en fournissant rapidement des informations préliminaires sur l'état de conservation des populations. La présente étude vise à évaluer l'utilité des code-barres d'ADN pour la délimitation de ces unités parmi les espèces de poissons d'eau douce les plus courantes de Java et de Bali, en comparant les structures génétiques des populations et les schémas de diversification de nombreuses espèces. Des niveaux substantiels de diversité cryptique sont découverts parmi les trois espèces de poissons d'eau douce largement répandues et analysées avec un total de 21 lignées mitochondriales indépendantes (BIN) observées chez *Barbodes binotatus*, *Channa gachua* et *Glyptothorax platypogon*. La distance génétique maximale pour chaque coalescent varie de 6,78 à 7,76 pourcent de divergence génétique (K2P), respectivement pour *C. gachua* et *G. platypogon*. La diversification et les analyses génétiques de population soutiennent un scénario de différenciation allopatrique. L'analyse de la distribution spatiale des BIN indique

des modèles de distribution concordants parmi les trois espèces qui permettent d'identifier 18 unités évolutives significatives. Les implications pour la conservation génétique de ces espèces sont discutées à la lumière de l'histoire de la région.

Discussion générale- Ces résultats ont mis en évidence d'important déficit des connaissances chez les poissons d'eau douce de Java et Bali. Plusieurs raisons, discutées dans les articles 1 & 2 sont avancées, sont avancées :

(1) Biais taxonomiques : L'exploration ichtyologique des eaux indonésiennes menée par une grande diversité d'ichtyologues au cours de trois grandes vagues historiques correspondant à des pratiques variées en matière de description d'espèces. La principale conséquence est la fragmentation de la littérature taxonomique et une connaissance éparpillée au fil du temps dans diverses publications scientifiques. Ainsi le manque de description détaillée des caractères morphologiques diagnostiques a souvent rendu l'identification morphologique des poissons d'eau douce indonésiens extrêmement difficile.

(2) Confusion taxonomique entre espèces proches : L'utilisation de méthodes d'inventaire basées sur l'ADN et d'algorithme de délimitation d'espèces a permis de résoudre plusieurs cas de confusion taxonomique perpétuée et de réexaminer les caractères morphologiques (article 2). Comme en témoigne *Rasbora* spp. (article 2), *R. baliensis* et *R. lateristriata* ont été largement confondus à Java et à Bali depuis la description de *R. baliensis* Hubbs 1954. Une situation similaire a également été décrite pour *Nemacheilus* spp. à Java, où les noms des espèces ont été appliqués de manière incohérente aux spécimens types. Seul un réexamen, par le biais de séquences d'ADN, comprenant des spécimens de la localité type, des limites d'espèce et de la répartition de l'aire de répartition a permis de détecter une inversion du nom de l'espèce et la réinterprétation des caractères originaux utilisés pour discriminer *N. chrysolaimos* et *N. fasciatus*.

Les résultats présentés dans cette thèse montrent que les aires de distribution des espèces sont beaucoup plus restreintes qu'on ne le pensait auparavant et que de multiples événements de fragmentation des populations conduisant à l'établissement de lignées moléculaires très divergentes sont détectés. Toutes les lignées moléculaires intra-spécifiques détectées présentent des aires de distribution limitées à un seul bassin versant ou à une poignée de rivières contigues et ne présentent aucun chevauchement (articles 2 et 3). Les schémas de répartition des aires, compatibles avec une divergence allopatrique, et des divergences

génétiques élevées, compatibles avec une fragmentation ancienne, questionnent le statut biologique des multiples lignées moléculaires détectées au sein des espèces et l'existence d'un isolement reproducteur.

Ces résultats sont à mettre au regard de l'état de conservation préoccupant de l'ichtyofaune de Java. Ainsi la répartition restreinte de la plupart des espèces et lignées moléculaires de poissons d'eau douce de Java est préoccupante du point de vue de la conservation, car leur persistance est soutenue par un ensemble géographiquement restreint de populations. Dans ce contexte, une réduction supplémentaire de la taille effective de leur population pourrait avoir des conséquences dramatiques sur leur survie, notamment en raison d'une stochasticité démographique accrue. Des programmes de restauration fondés sur l'élevage en captivité pourraient constituer une solution. Toutefois, le caractère commun de la diversité cryptique et la divergence génétique profonde observée entre les lignées moléculaires intraspécifiques s'opposent aux programmes de translocation entre bassins versants (articles 2 et 3).

Conclusions et perspectives- Cette thèse confirme l'efficacité des code-barres ADN pour capturer les contours des espèces et son utilité pour l'exploration de structures phylogéographiques intra-spécifiques. Des niveaux élevés de diversité cryptique sont détectés pour les espèces les plus largement répandues à Java et à Bali. Ces résultats indiquent que les code-barres ADN peuvent être utilisé pour l'identification d'échantillons au niveau de l'espèce et l'attribution de séquences inconnues à des taxons connus. La grande diversité de lignées cryptiques observées pour plusieurs espèces et leur ségrégation spatiale suggère également que les code-barres ADN peuvent également être appliqué pour retracer l'origine géographique des spécimens de plusieurs espèces. Ces résultats ouvrent de nouvelles perspectives pour la surveillance des ichtyodiversité de Java et de Bali. Compte tenu de l'érosion alarmante de l'ichtyodiversité de Java, les librairies de référence de codes-barres d'ADN développées pendant cette thèse constitueront probablement un outil utile à l'avenir pour le secteur universitaire et les organismes gouvernementaux chargés de la gestion des ressources ichtyologiques de Java.

List of Figures

Figure 1. Map of Indonesia including the 23 islands considered in the present review with biogeographic provinces and their boundaries. 1, Bali; 2, Bangka; 3, Batam and Bintan; 4, Belitong; 5, Buru; 6, Java; 7, Kalimantan; 8, Madura; 9 Natuna and Riau; 10, Sumatera; 11, Bacan; 12, Celebes; 13, Ceram; 14, Flores; 15, Halmahera; 16, Indonesian Timor; 17, Lombok; 18, Sumba; 19, Sumbawa; 20, Ternate; 21, Talaud; 22, Aru; 23, Indonesia New Guinea (Hubert *et al.*, 2015). 19

Figure 2. Geological reconstructions of lands and seas in the Indo-Australian Archipelago from 60 Ma to 5 Ma (Lohman *et al.*, 2011) 23

Figure 3. Epicontinental Shelf Sunda and the expected drainage river system (Hantoro, 2018) .

.....25

Figure 4. Elements of a DNA barcode records in the barcode of life data system (BOLD; Ratnasingham & Hébert, 2007).....27

Figure 5. Collection sites for the 1046 samples analysed in this study. Each point may represent several collection sites. 34

Figure 6. Distribution of genetic distances below and above species boundaries. (A) Distribution of maximum intraspecific distances (K2P). (B) Distribution of nearest neighbour distances (K2P). (C) Relationship between maximum intraspecific and nearest neighbour distances. Points above the diagonal line indicate species with a barcode gap

.....37

Figure 7. Accumulation curves and size class distributions of the 159 species analysed in this study. (A) Accumulation curves recovered from 100 iterations for species (bold curve) and genera (regular curve). (B) Distribution of size class (10 cm) of 266 species with documented maximum sizes among the 301 species of Java and Bali. (C) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 164 species with documented maximum length for euryhalin or amphidromous families. (D) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 135 species with documented maximum length for primary freshwater families..

.....42

Figure 8. Collection sites for the 241 samples analyzed in the present study following the sampling campaign detailed in Dahruddin *et al.,* 2017 and new sampling events in Lombok island. a Collection sites of *Rasbora* specimens. b Collection sites of *Nemacheilus* specimens. White dots correspond to sites where *Rasbora* or *Nemacheilus* specimens were collected. Black dots represent visited sites where no *Rasbora* or *Nemacheilus* specimens were observed. Each dot may represent several collection sites. .56

Figure 9. Bayesian maximum credibility tree of *Rasbora* DNA barcodes including 95% HPD interval for node age estimates and sequence clustering results according to the 5 species delimitation methods implemented 61

Figure 10. Bayesian Maximum Credibility Tree of *Nemacheilus* DNA barcodes including 95% HPD interval for node age estimates and sequence clustering according to the 5 species delimitation methods implemented 62

Figure 11. Selected specimen photographs of each of the 6 *Rasbora* species collected and recognized in the present study. a *Rasbora aprotaenia* (specimen BIF1501; SL = 47 mm; Ci Siih, Banten, Java). b *Rasbora argyrotaenia* (specimen BIF976; SL = 33 mm; Cilacap, Central Java). c *Rasbora baliensis* (specimen BIF2351; SL = 72 mm; Jembrana, West Bali). d *Rasbora lateristriata* (specimen BIF3619; SL = 87 mm; Kali Dauwan, Mojokerto, East Java). e *Rasbora* sp1 (specimen BIF864; SL = 61 mm; Kali Pelus, Purwokerto, Central Java). f *Rasbora* sp2 (specimen BIF155; SL = 43 mm; Ci Heulang, Sukabumi, West Java) 64

Figure 12. Species range distribution of the 6 *Rasbora* species recognized in the present study. Colored dots represent collection sites. White circles represent type localities. a *Rasbora aprotaenia.* b *Rasbora argyrotaenia*. c *Rasbora lateristriata.* d *Rasbora baliensis*. e *Rasbora* sp1. f *Rasbora* sp2 65

Figure 13. Species range distribution of the 2 *Nemacheilus* species recognized in the present study. Colored dots represent collection sites. White circles represent type localities. a *Nemacheilus chrysolaimos* b *Nemacheilus fasciatus* 66

Figure 14. Selected specimen photographs of each of the 2 *Nemacheilus* species collected and recognized in the present study. a *Nemacheilus fasciatu*s (specimen BIF495; SL = 42 mm; Ci Asem, Purwakarta, West Java). b *Nemacheilus fasciatus* (specimen BIF163; SL = 45 mm; Ci Heulang, Central Java). c *Nemacheilus chrysolaimos* (specimen BIF2032; SL = 58 mm; Ngerjo, Blitar, East Java). d *Nemacheilus chrysolaimos* (specimen BIF2074; SL = 55 mm; Bicoro, Lumajang, East Java) 67

Figure 15. Conceptual framework developed in the present study for the delineation of ESUs. Step 1a, detection of groups of populations differentiated by their haplotypes frequencies. Step 1b, detection of molecular lineages with independent evolutionary dynamics (e.g. lower connectance). Step 2, comparing population groups with molecular lineages. If population genetic structure results from ancient fragmentation of the populations, a correlation between genetic groups and BINs is expected. Conversely, the lack of correlation would indicate that population groups originated recently and either share ancient polymorphism or have been connected by gene flow in a recent past (Fu, 1999, Nielsen and Wakeley, 2001, Wakeley, 2001, Wakeley, 2003). Along the same line, several BINs may be delineated within a genetic group as a consequence of the stochastic nature of the coalescent but not disrupted gene flow (Hudson, 1982, Kingman, 1982, Tajima, 1983). Step 3, Delineation of ESUs for individual species. ESUs are defined based on either rvariation of haplotype frequencies or independent mitochondrial lineages after

comparisons with the phylogroups defined as groups of population sharing similar sets of mitochondrial lineages. 79

Figure 16. Location of the 51 collection sites for the samples analyzed in this study.....80

Figure 17. Population genetic structure of *Barbodes binotatus* (a), *Channa gachua* (b) and *Glyptothorax platypogon* (c) as inferred from the hierarchical cluster analysis and SAMOVA. 87

Figure 18. BINs diversification patterns of *Channa gachua* (a, d and g), *Glyptothorax platypogon* (b, e and h) and *Barbodes binotatus* (c, f and i) including Maximum Credibility Trees (a, b, c); Lineage Through Time plot (d, e, f) and Bayesian Generalized Skyline Plots (g, h, i). Solid black line in d, e and f is the median diversity accumulation curve, blue shaded area is the 95% highest posterior density intervals and dotted lines represent the BIN coalescent depth in Million years. The solid black line in g, h and i represent the median effective population size, blue shaded area is the 95% highest posterior density intervals and Population size scalar (in millions) = effective population size x generation time. Calibrations are derived from previously published molecular clock hypotheses applied to the maximum K2P distance foreach species - i.e. age of the MRCA (see Table 9). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).....91

Figure 19. Delineation of phylogroups and detection of phylogeographic breaks. a general hierarchical cluster based on the phylogenetic distances among sites established from the grafted chronograms of *C. gachua, G. platypogon* and *B.binotatus*, and occurrence data (see Table S2). b distribution range of phylogroups I and II, and associated phylogeographic breaks (solid black lines). c distribution range of phylogroups III, IV and V, and associated phylogeographic breaks (solid black lines). d phylogeographic breaks identified from the phylogroup geographic boundaries and further used for delineating ESUs.....92

Figure 20. Mapping of the population groups identified by SAMOVA (a, d and g), the BINs delineated by the RESL algorithm (b, e and h) and ESUs candidates (c, f and i) for *C. gachua* (a, b and c), *G. platypogon* (d, e and f) and *B.binotatus* (g, h and i). Ambiguous assignment of populations to ESUs due to trans distribution across phylogeographic breaks are highlighted by solid white circle (c, f and i). 93

Figure 21. Number of species description per decades since 1758 for the species occurring in Indonesia. A. All species, B. Endemic species (Hubert *et al.*, 2015). 110

Figure 22. *Nemacheilus chrysolaimos* Valenciennes 1846 and *Nemacheilus fasciatus* Valenciennes 1846 following Article 2 and original illustration of *Nemacheilus chrysolaimos* Valenciennes 1846. 111

List of Table

Table 1. Summary statistics of the Indonesian ichthyofauna including surface of islands, species richness, endemism and species density for the 23 major islands of the Archipelago (Hubert *et al*., 2015) 21

Table 2. Summary statistics of the genetic distances (K2P) through increasing taxonomic levels 36

Table 3. Summary statistics of the 18 species with more than a single BIN including the number of BIN, maximum intraspecific distance and BIN accession numbers 37

Table 4. Summary statistics per families of the taxonomic coverage yielded by this study including the number of species derived from online catalogues, DNA barcoding coverage and new records.....39

Table 5. Summary statistics of the 20 exotic species including the number of BIN, maximum intraspecific distance, geographic origin and BIN accession numbers.....43

Table 6. Summary statistics of *Rasbora* species genetic diversity and species delimitation schemes 62

Table 7. Summary statistics of *Nemacheilus* species genetic diversity and species delimitation schemes 63

Table 8. Partitioning of the molecular variance at various spatial scales, fixation indexes and number of population groups as inferred from SAMOVA. The percentage column indicates the amount of total variance explained by each of the hierarchical levels according to the number of groups of population. Φ-statistics estimate the correlation among haplotypes at each ofthe hierarchical levels examined and their significant departure from a random distribution of the haplotype was tested through randomization across 1000 permutations. 86

Table 9. Summary statistics of the genetic diversity including the sampling size (N), haplotype diversity (h), nucleotide diversity (π) and mean number of pairwise differences among haplotypes (mean-pairwise differences) for each of the population groups. 88

Table 10. Summary statistics of the genetic K2P distances and age estimates. The maximum and average K2P distances are provided for each BINs and the maximum K2P distances are provided for the entire coalescent trees. The age estimate of the MRCA is provided based on three alternative hypotheses of molecular clock including 0.005 genetic divergence per million years (Hardman and Lundberg, 2006; Read *et al*., 2006), 0.012 genetic divergence per million years (Bermingham *et al*., 1997) and 0.02 genetic divergence per million years (Read *et al*., 2006). Divergence estimates are derived from the Bayesian analyses based on the H2 calibration with upper and lower bounds of the prior age intervals defined by the molecular clock hypotheses H1 and H3. 89

Table of content

I. Introduction to Sundaland 19

1.1. Sundaland ichthyofauna: species richness and endemism 20

1.2. Geological history 22

1.3. Sundaland hydrography 24

II. Fundaments and objectives 25

2.1. General context 25

2.2. A DNA-based approach of the ichthyodiversity 26

2.3. Objectives and plan of the thesis 29

III. DNA Barcoding of Sundaland Freshwater Fishes: the Java library 31

3.1 Abstract 31

3.2 Introduction 32

3.3 Materials and methods 33

3.3.1 Specimen collections and identifications 33

3.3.2 Assembling a checklist of the Javanese and Balinese freshwater fishes 34

3.3.3 DNA barcode sequencing 34

3.3.4 DNA barcode analysis 35

3.4 Results 36

3.4.1 DNA barcode analyses and BIN splits 36

3.4.2 Species diversity and exotic species 39

3.5 Discussion 44

3.6 Conclusions 47

IV. Validity and distribution of endemic species of *Rasbora* spp. and *Nemacheilus* spp. in Java through DNA Barcodes 53

4.1 Abstract 53

4.2 Introduction 54

4.3 Materials and methods 56

4.3.1 Sampling and collection management 56
4.3.2 Sequencing and international repositories 57
4.3.3 Species delimitation and genetic diversity 57
4.4 Results 59
4.5 Discussion 63
4.5.1 Taxonomy of Rasbora species in Java, Bali and Lombok 63
4.5.2 Taxonomy of Nemacheilus species in Java 66
4.5.3 Species distribution and conservation genetics 67
4.6 Conclusion 69
4.7 Acknowledgements 69
V. Conservation genetics of the freshwater fishes of Java 76
5.1 Abstract 76
5.2 Introduction 77
5.3 Materials and methods 80
5.3.1 Sampling strategy and collection management 80
5.3.2 Sequencing and international repositories 81
5.3.3 Population genetic structure (Fig. 15, step 1a) 81
5.3.4 Delineating mitochondrial lineages and inferring their diversification (Fig. 15, step 1b) 82
5.3.5 Delineating ESU (Fig. 15, steps 2 & 3) 83
5.4 Results 84
5.4.1 Population genetic structure and molecular variance 84
5.4.2 Delineation and diversification of BINs 89
5.4.3 Identifying phylogroups and delineating ESUs 91
5.5 Discussion 94
5.5.1 Regional patterns and evolutionary dynamics 94
5.5.2 Implications for the conservation of aquatic ecosystems in South Sundaland 95

5.5.3 Validation of the approach and generalization 97
5.6 Conclusion 99
5.7 Acknowledgements 100
VI. General Discussion 110
6.1 State of the knowledge on the ichthyodiversity of Java 110
6.2 Local vs. regional dynamics and their influence on species range distribution and populations genetic structure 112
6.3 Implications for conservation 114
VII. Conclusions and Perspectives 116
VIII. Bibliography 116

I. Introduction to Sundaland

Located in the southern part of Southeast Asia, Indonesia is the world's largest archipelago with as many as 14,000 islands. Resulting from a complex geological history, the Indonesian archipelago is made of three major geographic units (Fig. 1): (1) Sundaland including the islands of Java, Sumatra, Borneo and Bali that belong to the shallow Sunda shelf, (2) Wallacea including isolated islands surrounded by deep seas, (3) Sahul that includes the island of Papua and corresponds to a shallow shelf (Hubert *et al*., 2015). This diversity of islands and its rich biodiversity has attracted the attention of biologist for centuries. So far, two major biotic transitions have been identified: (1) the Wallace line separating the Sunda shelf from Wallacea, corresponding to a major faunistic transition between Borneo and Sulawesi and associated to the deep waters of the Makassar straight (Fig. 1), (2) the Lydekker line separating the islands of Wallacea and the Sahul shelf, corresponding also to a major transition between the Moluccas and the island of Papua and associated to the deep waters of the Seram sea. Each of these domains results from distinct geological and paleoenvironnemental histories. Sundaland for instance, being a shallow shelf not exceeding 120m below sea level, has been repeatedly connected to the continent during glacial maxima while Wallacea islands have been physically isolated throughout the Pleistocene (Hall, 1996; Voris, 2000; Lohman *et al*., 2011).

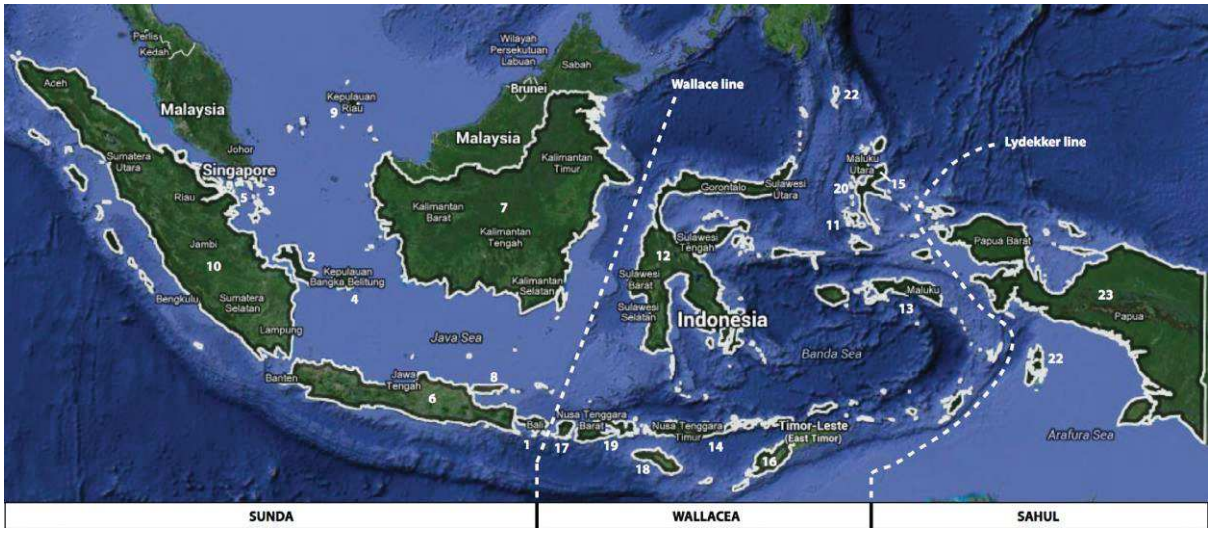


Figure 1. Map of Indonesia including the 23 islands considered in the present review (Appendix) with biogeographic provinces and their boundaries. 1, Bali; 2, Bangka; 3, Batam and Bintan; 4, Belitong; 5, Buru; 6, Java; 7, Kalimantan; 8, Madura; 9 Natuna and Riau; 10, Sumatera; 11, Bacan; 12, Celebes; 13, Ceram; 14, Flores; 15, Halmahera; 16, Indonesian Timor; 17, Lombok; 18, Sumba; 19, Sumbawa; 20, Ternate; 21, Talaud; 22, Aru; 23, Indonesia New Guinea (Hubert *et al*., 2015).

As a consequence of this diversity of geological origin and paleoenvironmental conditions across its islands, Indonesia host nearly 110,000 plants and 23,000 animals species

and as such, is considered as a mega-diverse country similarly to Brazil and the Congo Republic (Darajati *et al*., 2016). Levels of anthropogenic threats on Indonesia's biodiversity are extremely high due to landscape conversion as a consequence of deforestation for the development of agriculture and urbanism (Myers *et al*., 2000; Darajati *et al*., 2016). Thus, this exceptional diversity associated to high levels of threats led to the designation of Sundaland and Wallacea as biodiversity hotspots (Myers *et al*., 2000). To date, Sundaland constitutes one of the largest hotspot in term of species richness and endemism that ranks third after the Tropial Andes and Mesoamerica. Anthropogenic threats in Sundaland, however, are extremely high due to the high human-population density and rank it as the worlds most threatened hotspot. As such, Sundaland constitutes an absolute priority of conservation plans on a global scale.

1.1. Sundaland ichthyofauna: species richness and endemism

The world has approximately 8.7 million eukaryotic species among which 14,000 are freshwater fish species and 19,800 are marine fish species (Mora *et al*., 2011; Darajati *et al*., 2016). Southeast Asia has 3107 valid fish species in inland waters, estuaries and mangroves that encompass 707 genera and 137 families (Kottelat, 2013). Within this high diversity, Indonesia host 1218 freshwater fish species including 630 endemic species that belong to 84 families. The Cyprinidae family largely dominates with 241 species, followed by the family Gobiidae with 122 species, Osphronemidae with 81 species and Bagridae with 60 species. Sundaland itself hosts 899 species including 431 endemic species that is 74% of Indonesia freshwater fish diversity. In Sundaland, the Cyprinidae family include 231 species largely distributed in Sundaland aquatic habitats (Hubert *et al*., 2015). The diversity of Indonesia ichtyofauna is presented in Table 1.

Among the 1218 freshwater fish species of Indonesia, 1172 are native and 28 are exotic with varying status ranging from introduce without being invasive to highly invasive. Introduce not only from outside of Indonesian waters, introduction of native species between islands has also been reported. *Channa micropeltes* and *Channa striata*, for instance, are native species of Sundaland that have been introduced in Wallacea. Another example is *Puntigrus tetrazona*, an ornamental Cyprinidae species from Sumatra that has been introduced in Java waters. These introductions might be expected to have negative impacts on native species through several mechanisms such as predation, competitive exclusion and pathogens transmission.

Table 1. Summary statistics of the Indonesian ichthyofauna including surface of islands, species richness, endemism and species density for the 23 major islands of the Archipelago (Hubert *et al.*, 2015)

					All species			Endemics			
Code map	**Biogeographic domain**	**Island**	**Surface (km2)**	**N. of family**	**N. of species**	**Percent (total all species)**	**Density (Sp/1000km2)**	**N. of species**	**Percent (total endemics)**	**Percent (total all species)**	**Density (Sp/1000km2)**
1	Sunda	Bali	5561	14	38	0.3	6.8	5	0.01	13.16	0.9
2		Bangka	11330	23	35	0.9	3.1	10	0.02	28.57	0.88
3		Batam/Bintan	2280	18	26	0.1	11.4	2	0	7.69	0.88
4		Belitung	4800	14	18	0.2	3.8	4	0.01	22.22	0.83
5		Buru	9505	14	16	0.2	1.7	0	0	0	0
6		Java	126700	54	213	13.3	1.7	33	0.05	15.49	0.26
7		Kalimantan	539500	66	646	35.3	1.2	294	0.46	45.51	0.54
8		Madura	5290	11	12	0.3	2.3	1	0	8.33	0.19
9		Natuna/Riau	3420	8	19	0.1	5.6	8	0.01	42.11	2.34
10		Sumatera	425000	64	460	23.5	1.1	162	0.25	35.22	0.38
		Total	**1133386**	**73**	**899**	**74**	**0.8**	**431**	**0.68**	**47.94**	**0.38**
11	Wallacea	Bacan	1800	6	8	0.1	4.4	0	0	0	0
12		Celebes	174600	31	146	6.7	0.8	69	0.11	47.26	0.4
13		Seram	17418	25	53	1.7	3	4	0.01	7.55	0.23
14		Flores	14300	13	25	0.2	1.7	3	0	12	0.21
15		Halmahera	17780	23	51	0.4	2.9	5	0.01	9.8	0.28
16		Indonesian Timur	15770	13	23	0.3	1.5	2	0	8.7	0.13
17		Lombok	5435	10	20	0.2	3.7	4	0.01	20	0.74
18		Sumba	11153	5	9	0.2	0.8	0	0	0	0
19		Sumbawa	15448	10	18	0.1	1.2	2	0	11.11	0.13
20		Ternate	65	9	11	0.1	169.2	0	0	0	0
21		Talaud	1285	1	1	0.1	0.8	1	0	100	0.78
		Total	**275054**	**37**	**184**	**15**	**0.7**	**82**	**0.13**	**44.57**	**0.3**
22	Sahul	Aru	8563	8	22	0.5	2.6	1	0	4.55	0.12
23		Indonesian New Guinea	421981	45	249	15.4	0.6	124	0.19	49.8	0.29
		Total	**430544**	**45**	**255**	**16**	**0.6**	**125**	**0.2**	**49.02**	**0.29**
Total			**1838984**	**79**	**1172**		**0.6**	**630**			**0.35**

Some introductions, however, are of serious concern, however, due to the large size of the introduced species such as *Arapaima gigas,* the gar fish (*Lepisosteus* spp.) and alligator gar (*Atractosteus spatula*), that can grow beyond 2 meters of length, or due to their dangerousity such as piranha species (*Serrasalmus* spp.) that have been recently reported in Java (No.31/KEP-BKIPM/2017). The species identity of these introduced species, however, remains uncertain.

The diversity of Indonesian freshwater fishes might be expected to be under-estimated. In fact, the number of species described per decades has been drastically increasing during the last three decades (Hubert *et al*., 2015) to exceed 100 new species described per decade. This situation is due to the discovery of new species further described based on morphological characters but also due to the increasing use of genetic approaches that have help clarify several cases of perpetuated taxonomic confusion among closely related and morphologically similar species (Conte-Grand *et al*., 2017; Farhana *et al*., 2018, Lim *et al*., 2016). This trend is further amplifiyed by the abundance of small size species below 5 cm for which morphological characters are not easily accessible. Sundaland ichtyofauna in particular is awaiting a large-scale re-examination of species biological status through DNA-based methods. Large knowledge gap have been recently highlighted in the taxonomic knowledge of Indonesian freshwater fishes that currently bridle the development of conservation plans. In the meantime, the increasing levels of anthropogenic threats might be expected to have already impacted Sundaland ichtyofauna.

1.2. Geological history

The Indonesian archipelago results from a long geological history than span across the last 60 Mya (Lohman *et al*., 2011; Hall, 1996). The geological settlement of the Indonesian archipelago has been mainly driven by plate tectonic and the collision of the Australian and Eurasian plates (Fig. 2). At 65 Mya, Sundaland was a continental promontory at the southern end of the Eurasian plate (Hall, 1996) located at the equator with a tropical climate (Fig. 2A). This period is supposed to correspond to the thermal maximum that peaked at 56 Mya. The climate was probably weter at that time than today, but from approximately 45 Mya, a cooler and drier climate established until 23 Mya (Heaney, 1991; Morley, 2000). During the subduction of the Australian plate, several islands emerged from the sea around 40 Mya until 25 Mya that later contributed to the settlement of Wallacea islands (Figs. 2B, 2C & 2D). In the meantime, Sundaland was surrounded by subduction zones until 25 Mya (Fig. 2D).

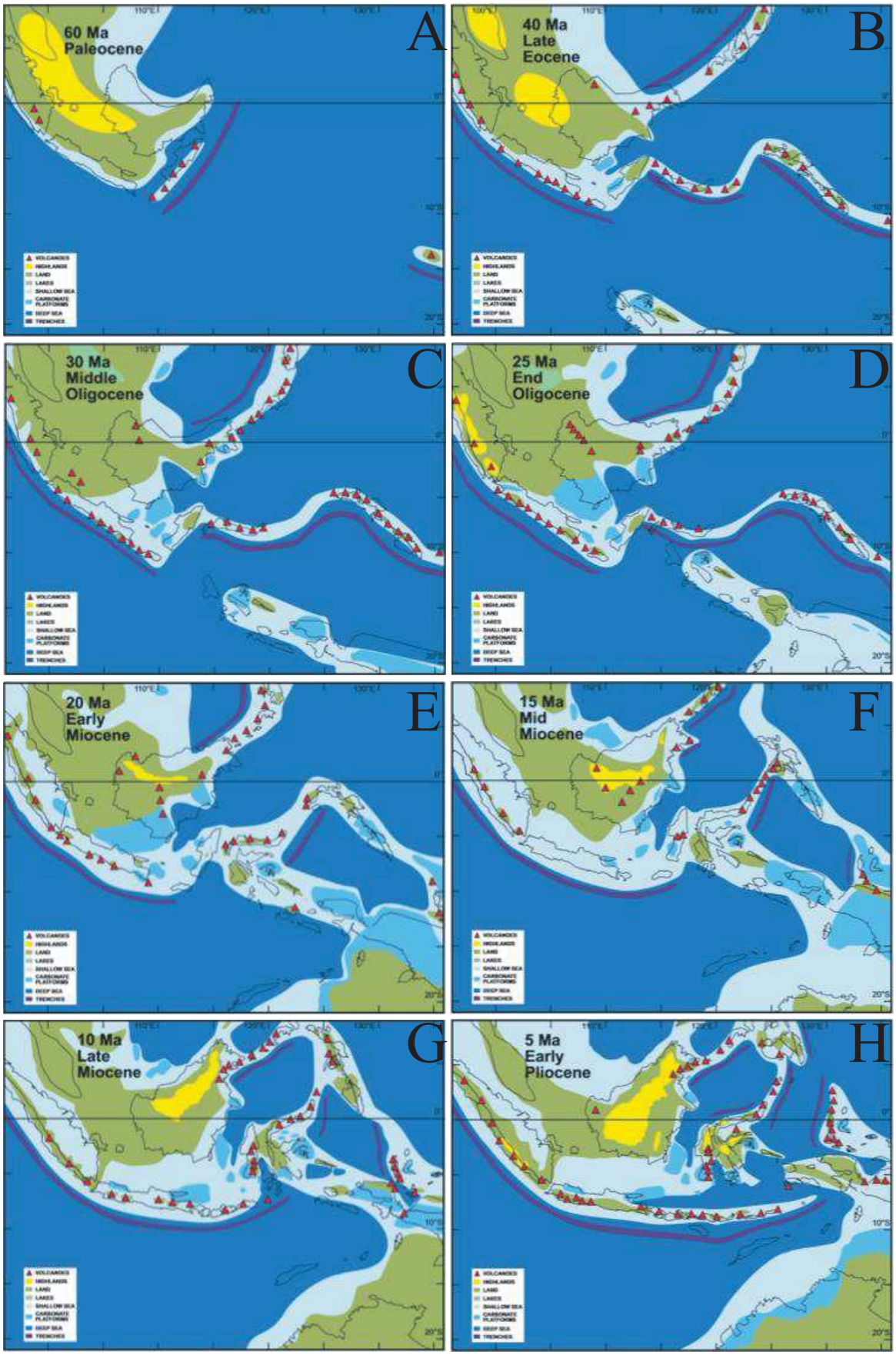


Figure 2. Geological reconstructions of lands and seas in the Indo-Australian Archipelago from 60 Ma to 5 Ma (Lohman *et al.*, 2011)

Around 23 Mya, the intensification of the subduction of the Australian plate initiated the rotation counter-clockwise of the Sunda shelf (Figs. 2F & 2G). This subduction activity prompted the settlement of Sumatra, a volcanic chains along the Sunda margin, still connected to the continent through a land bridge at the Southwestern tip of the Sunda shelf between 10 and 5 Mya (Figs. 2G & 2H). The rise of Java island is more recent and a late consequence of the Sunda shelf rotation and volcanic activity along the subduction zone that resulted in the emergence and further merging of two volcanic arches (Fig. 2G). Sundaland islands have been connected to the continent until 5 Mya (Fig. 2H) and the differentiation of the islands of Borneo, Sumatra and Java happened recently during the last 5 Mya.

The reconstruction of this geological scenario, dominated by fragmentation and the progressive settlement of isolated sets of islands, allowed the formulation of plausible vicariant mechanisms of speciation (Whitmore, 1981; Lohman *et al*., 2011) further balanced by periodic dramatic events such as the eruption of Toba that likely prompted large fires that impacted forests distribution in North Sumatra and Sunda shelf (Wilting *et al*., 2012; O'Connell *et al*., 2018). The tectonic reconstructions, however, suggest that the Southern part of Sundaland are much younger, resulting from volcanic and tectonic activity until the Plio-Pleistocene transition (De Bruyn *et al*., 2014; Giarla *et al*., 2018; Hall, 2009). Thus, islands size, age, and isolation from the continent during the geological history of the Indonesian archipelago jointly contributed to establish varying levels of species richness and endemism among islands (Heaney, 1986; MacArthur & Wilson, 1967; Rabosky & Glor, 2010; O'Connell *et al*., 2018).

1.3. Sundaland hydrography

The settlement of Sundaland hydrographic network is recent compared to other tropical systems (Lundberg, 1998) and results from the last 2-3 Mya of geological history (Fig. 2). The geological settlement of Sundaland has largely interacted with the sea-level fluctuations of the Pleistocene associated to the milankowitch cycles (Broecker and Denton 1989; Huybers, 2006; Voris, 2000; Woodruff, 2010). Sundaland in its present form is made of three major islands separated by the shallow Java sea that do not exceed 120m deep (Lohman *et al*., 2011). Sea levels dropped down to 120m lower than today (Voris, 2000; Woodruff, 2010) during glacial times. In particular, sea level was −123 m lower than today during the Last Glacial Maximum (LGM) (Hanebuth *et al*., 2009). As a consequence, the Sunda shelf was mostly emerged and Java, Sumatra and Borneo were embodied in a large

landmass connected to the continent (Fig. 3). This landmass formed a landbridge between the Malaysia peninsula, Sumatra, Java and Borneo (Bird *et al*., 2005; Hanebuth *et al*., 2011). River drainage pattern during the LGM was as follow in Sundaland (Fig. 3): (1) North Sunda paleoriver (Fig. 3, Sunda Utara) running North and draining the central part of Sumatra and western Part of Borneo; (2) East Sunda paleoriver (Fig. 3, Sunda Timur), running to the East and draining the Java island and the Southern part of Borneo (Hantoro, 2018; Metcalfe, 2009).

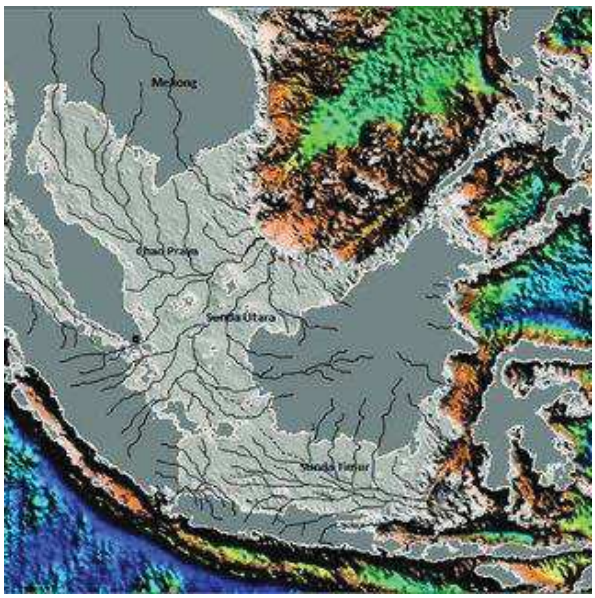


Figure 3. Epicontinental Shelf Sunda and the expected drainage river system (Hantoro 2018)

The existence of large paleorivers, varyingly connecting the islands of Sundaland, might be expected to have influenced the distribution of freshwater aquatic organisms by enabling dispersal among islands during glacial times, further balanced by fragmentation during inter-glacial times (Voris, 2000; Woodruff, 2010). For instances, recent studies have evidenced trans-islands sister-relationships among Sundaland freshwater fish species (Pouyaud & Paradis, 2009; Endra *et al*., 2016; Lim *et al*., 2016; Farhana *et al*., 2018) or population structure within species that partially match paleoriver boundaries (Beck *et al*., 2017; De Bruyn *et al*., 2013; Dodson *et al*., 1995; Nguyen *et al*., 2008). These results suggest that the Quaternary ice ages have deeply influenced Sundaland aquatic biotas at a pace similarly to what has been previously observed in temperate biomes (Bennett, 2017). The existence of large paleorivers during the glacial maxima led to the formulation of a biogeographic hypothesis, the paleoriver hypothesis, that hypothesized that the paleoriver boundaries drove allopatric speciation once the paleorivers established (Kottelat *et al*. 1993).

II. Fundaments and objectives

2.1. General context

Due to the overlays of several anthropogenic pertubations such as logging in primary forest, landscapes conversion, urbanization and associated pollution, Sundaland biomes are among the most threatened on earth and freshwater fishes are no exceptions. Because

watersheds are integer of anthropogenic perturbations at the regional scale and freshwater fishes are still a major source of animal protein in Sundaland, freshwater fishes populations have been rapidely declining during the last decades (Hubert *et al.*, 2015). Java particularly exemplifies the challenges faced by conservation plans. With a population exceeding 140 millions people, that is nearly half of the population of the archipelago, in an island of 130,000 km^2, the Java island has been the most impacted by Indonesia's fast economic development. Impacted by the expansion of invasive species and overfishing by artesanl fisheries, the Java ichtyofauna has been through a dramatic decline during the last decades while attracting much less attention in term of ichthyological explorations during the last decade (Hubert *et al*., 2015). Recents studies have demonstrated, for instance, that a carefull re-examination of species boundaries through DNA-based methods enabled the detection of new species among closely related species complex (Keith *et al*., 2015). This situation is particularly true for species groups lacking taxonomists and which taxonomy is accessible to only but a few specialists worldwide, thus questionning the sustainability of taxonomic knowledge and jeopardizing conservation efforts.

Identifying and delineating species are the primary tasks of taxonomy. Owing to the decreasing interest of the nations for taxonomy and the inventory of living beings, funds have been drastically decreasing during the last two decades for taxonomic studies (Hubert & Hanner, 2015). As a consequence, the worldwide pool of taxonomists has dramatically decreased. DNA barcoding, as an automated tool for species identification through DNA-based methods, opened new perspectives in countries were taxonomic knowledge gaps and lacks of taxonomists jeopardize conservation efforts (Dahruddin *et al*., 2017).

This dissertation aims at filling the taxonomic knowledge gap in the Java ichthyofauna through the re-exmination of the freshwater fishes diversity through a standardized DNA-based approach. On that front, DNA barcoding opens new perspective by enabling the examination of the biological status of the nominal species in Java and give access to the species identification by any scientists, whatever the state of the biological material to be identified (Ko *et al*., 2013; Hubert *et al*., 2010). The research presented in this dissertation is aiming to the decision maker and governmental officers involved in the conservation and management of freshwater fishes of Java.

2.2. A DNA-based approach of the ichthyodiversity

DNA barcoding is designed to provide accurate, and automated species identifications through the use of molecular species tags based on short, standardised gene regions. The primary goal of DNA barcoding consists in the assembly of DNA barcode reference libraries for known species in order to develop molecular tools for automated species identification. To that end, a 648-bp segment of the 5' region of mitochondrial cytochrome c oxidase I (COI) has been proposed as a universal DNA barcode for the animal kingdom (Hebert *et al.*, 2003). DNA barcode records consist essentially of seven data elements (Fig. 4) (Hubert *et al.*, 2008):

1. Species name (derived from morphological identification)
2. Voucher data (deposited in a permanent repository)
3. Collection record (geographic information)
4. Identifier of the specimen (name of the expert that performed the identification)
5. COI sequence of at least 500 bp and resulting from a sequencing forward an reverse
6. PCR primers used to generate the amplicon
7. Trace files

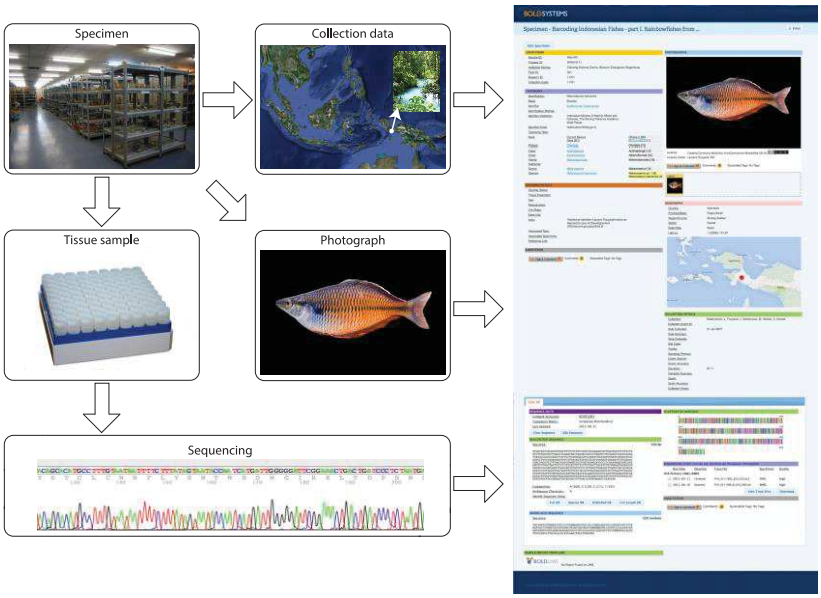


Figure 4. Elements of a DNA barcode records in the barcode of life datasystem (BOLD; Ratnasingham & Hébert, 2007)

Mitochondrial DNA presents several advantages for large scale molecular tagging. First, this genome is present in a large number of copies yielding substantial amounts of genomic DNA from a variety of extraction methods. Second, the high mutation rate and small effective population size make it often an informative genome about evolutionary patterns and processes. In most animals, however, only the mitochondrial genome from the mother is transmitted to the offspring. This pattern of mtDNA inheritance is well known as "maternal inheritance" (Sato & Sato, 2013). This constitutes the main limits when assessing species boundaries, as the sharing of DNA sequences between species cannot be determined as the origin of the paternal genome is unknown.

The sharing of genomic material between species can have different origin: (1) genes are exchanged through gene flow between species for which the reproductive isolation is not complete resulting in an introgressive hybridization of genomic material from one species to another; (2) ancestral polymorphism that is defined as variants that arose by mutation prior to the speciation event that generated the species in which they segregate (Hubert & Hanner, 2015). The presence of shared polymorphism between species may complicate the interpretation of species boundaries, particularly so if mitochondrial genome alone is examined. It is thus important to take into account the contribution of ancestral polymorphisms to variability within species and divergence between species (Nowell *et al*., 2011). There is increasing evidence that closely related species contain many polymorphisms that were present in their common ancestral species (Charlesworth *et al*., 2005). In this context, a carefull examination of morphological characters for the establishment of DNA barcode reference libraries is essential to avoid adding noise when estimating the frequency of shared polymorphisms and validated DNA barcodes records (Hubert & Hanner, 2015).

From a conceptual and statistical perspective, the coalescent theory provides new tools for the objective delimitation of species through DNA sequences (Fu & Li, 1999; Rosenberg & Nordborg, 2002). Coalescent-based methods have been developed and proposed for DNA barcoding data analysis in order to overcome the limits of a genetic distance threshold for species boundaries (Hubert & Hanner, 2015). Moreover, the publication of genetic distance trees as the only output of DNA barcoding attracted criticism to the approach. It is worth mentionning that DNA barcoding is not, in a strict sense, a phylogenetic reconstruction to solve phylogenetic issues or to classify but a species diagnostic approach. These methods should be used by users with a sound background on the phylogeny versus identification debate in the context of DNA barcoding (Casiraghi *et al*., 2010).

Matz & Nielsen (2005) proposed one of the first effort to introduce statistical formalisms in DNA barcoding data analysis. Their tree-based method takes into account phylogenetic uncertainty and uses population genetic theory to determine cut offs for species assignment in ambiguous cases. A like lihood ratio test allows to evaluate possible boundaries of intra-specific variation (for each species) on the basis of reference datasets using population genetic inferences based on coalescent theory. Since then, several methods of species delimitation through DNA sequences have been proposed. Four methods in particular are increasingly used: (1) Barcode Index Numbers (BINs) are an interim taxonomic system for animals (Ratnasingham & Hebert 2013). Based on a network approach,

BINs identify zone of haplotypes with higher connections as a signature of species boundaries, providing that ancestral haplotypes near the Most Recent Common Ancestor (MRCA) of the species show a higher number of connections in average (Milton *et al*., 2013), (2) PTP, incorporates different levels of intraspecific genetic diversity deriving from differences in either the evolutionary history or sampling of each species (Kapli *et al*., 2017), it implementes a likelihood optimization of species boundaries with two poisson distribution for intra-specific and inter-specific branching events; (3) GMYC is a likelihood method for delimiting species by fitting within- (Coalescent) and between-species (Yule diversification model) branching models to reconstructed gene trees (Fujisawa & Barraclough, 2013), (4) ABGD is an automatic procedure that sorts the sequences into hypothetical species based on the barcode gap, which can be observed whenever the divergence among organisms belonging to the same species is smaller than divergence among organisms from different species (Puillandre *et al*., 2012).

2.3. Objectives and plan of the thesis

The present dissertation is aiming at addressing two main questions, a first methological question related to the effectiveness of DNA barcoding in capturing species boundaries for the Javanese ichthyofauna, a second biological question related related to the impact of the geological history of Java and its consequences on the ichthyodiversity.

(1) DNA barcoding is an effective tool to characterize the ichthyodiversity of Java?

DNA barcoding has been proposed as a fast, easy, relatively inexpensive approach to provide alternative solutions for specimen identification. As a DNA-based method, DNA barcoding is expected to enable specimen identification to the species level, whatever the life stage under scrutiny or the biological material to be analyzed. This prediction, however, is based on the assumption that DNA barcoding aptly captures species boundaries for the biodiversity under scrutiny and that DNA barcodes constitute cluster of closely related sequences that match species boundaries as defined by morphological characters (Hubert & Hanner, 2015). Thus, the objective of the present dissertation is to explore the utility of DNA barcoding for the characterization of the Javanese ichthyofauna for: (1) specimen identification to the species level, (2) characterize the biological status of morphological species and populations to further produce recommandations for conservation plans.

(2) Is the geological history of Java a predictor of diversity patterns?

The volcanic origin of Java islands might be expected to have left a inprint on species range distribution and population structure (Lohman *et al*., 2011). Java originated from the merging of two volcanic arches, one is the West and a second in the east including East Java, that emerged between 10 and 5 Ma and further aggregated during the last 5 Million years. This scenario questionnes the dynamic of colonization of Javanese rivers by strictly freshwaters organisms and the commonness of population structure across species. The commonality of population structure, however, is key to the establishment of conservation plans.

The thesis is structured as follow:

(1) Objective 1: revisiting the ichthyodiversity of Java through DNA barcodes and assessing the utility of a DNA barcode reference library for further automated molecular identifications.

(2) Objective 2: examine the validity of nominal species in Java and refine the knowledge of their range distribution for the species complex *Rasbora* spp. and *Nemacheilus* spp. through DNA barcodes

(3) Objective 3: identify the commonality of population structure for multiple co-distributed species in Java to identify conservation units.

III. DNA Barcoding of Sundaland Freshwater Fishes: the Java library

Article 1: Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species

Hadi Dahruddin,* Aditya Hutama,† Frédéric Busson,‡ Sopian Sauri,* Robert Hanner,§ Philippe Keith,‡ Renny Hadiaty* and Nicolas Hubert¶

* LIPI, Research Center for Biology, Zoology Division, MZB, Gedung Widyasatwaloka, Jl. Raya Jakarta Bogor Km. 46, Cibinong- Bogor 16911, Indonesia.

† Faculty of Mathematics and Natural Science, Animal Bioscience, Bogor Agricultural University, Jl. Raya Darmaga, Bogor 16680, Indonesia.

‡ Muséum national d'Histoire naturelle, UMR 7208 (MNHN-CNRS-UPMC-IRD-UCBN), CP 026, 43, rue Cuvier, Paris Cedex 05 F-75231, France.

§ Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada.

¶ Institut de Recherche pour le Développement, UMR 226 ISEM (UM2-CNRS-IRD), Place Eugène Bataillon, CC 065, F-34095 Montpellier cedex 05, France

3.1 Abstract

Among the 899 species of freshwater fishes reported from Sundaland biodiversity hotspot, nearly 50% are endemics. The functional integrity of aquatic ecosystems is currently jeopardized by human activities, and landscape conversion led to the decline of fish populations in several part of Sundaland, particularly in Java. The inventory of the Javanese ichthyofauna has been discontinuous, and the taxonomic knowledge is scattered in the literature. This study provides a DNA barcode reference library for the inland fishes of Java and Bali with the aim to streamline the inventory of fishes in this part of Sundaland. Owing to the lack of available checklist for estimating the taxonomic coverage of this study, a checklist was compiled based on online catalogues. A total of 95 sites were visited, and a library including 1046 DNA barcodes for 159 species was assembled. Nearest neighbour distance was 28-fold higher than maximum intraspecific distance on average, and a DNA barcoding gap was observed. The list of species with DNA barcodes displayed large discrepancies with the checklist compiled here as only 36% (i.e. 77 species) and 60% (i.e. 24 species) of the

known species were sampled in Java and Bali, respectively. This result was contrasted by a high number of new occurrences and the ceiling of the accumulation curves for both species and genera. These results highlight the poor taxonomic knowledge of this ichthyofauna, and the apparent discrepancy between present and historical occurrence data is to be attributed to species extirpations, synonymy and misidentifications in previous studies.

Keywords: DNA barcoding, fish, fisheries management, habitat degradation, wildlife management

3.2 Introduction

Amongst the 25 biodiversity hotspots identified worldwide for their remarkable levels of endemism and anthropogenic threats, four are observed in South-East Asia (SEA), including Indo-Burma (Thailand, Cambodia, Laos, Vietnam and Myanmar), Sundaland (Malaysia, Indonesia), Wallacea (Indonesia) and the Philippines (Myers *et al*. 2000). The Sundaland and Indo-Burma hotspots exhibit the highest species richness and endemism in SEA that rank them as some of the world most speciose together with the Brazil's cerrado or West African forests (Lamoureux *et al*. 2006). These high levels of endemism and species richness may be seen as the result of the complex history of the Indo-Australian Archipelago (IAA) that has been repeatedly fragmented during its ontogenesis through tectonic events (Lohman *et al*. 2011) and eustatic fluctuations (Woodruff 2010). The Sundaland hotspot, however, is currently one of the world's most endangered and aquatic ecosystems exemplify the diversity of anthropogenic pressures faced by its biotas (Hoffman *et al*. 2010). Their functional integrity is currently jeopardized by interactions among ecological (e.g. mining, logging, land conversion, organic and inorganic contaminations) and biotic perturbations (e.g. alien species, overexploitation by inland fisheries), resulting in the destruction of foraging and spawning grounds and the decline of populations (Schilthuizen *et al*. 2005; Clements *et al*. 2006, 2008; Fraser 2006; Normile 2010; Sodhi *et al*. 2010).

Indonesia exhibits one of the world highest densities of freshwater fish species (i.e. 0.6 species per 1000 km2) ahead of Brazil (0.37 species per 1000 km2) and the Democratic Republic of Congo (0.48 species per 1000 km2) (Hubert *et al*. 2015). In Sundaland, nearly 900 species and 430 endemics have been reported, a diversity that accounts for 74% and 48% of the global and endemic diversity, respectively (Hubert *et al*. 2015). Amongst the three major islands of Sundaland, Java exhibits the highest density of species with 1.7 species per

1000 km2 (ca. 213 species) ahead of Kalimantan (1.2 species per 1000 km2 for a total of 646 species) and Sumatra (1.7 species per 1000 km2 for a total of 460 species). During the last decades, the exponential growth of the human population in Java – 130 millions of people sharing 130 000 km2 – had dramatic consequences on Javanese ecosystems and Javanese biotas are currently some of most threaten in Sundaland.

The diversity loss in Sundaland hotspot is of great concern; however, the knowledge about species taxonomy and distribution is still incomplete for many taxa, what arguably bridles the establishment of sounds conservation plans. This situation is further amplified by the complexity of accurately delineating species in megadiverse faunas due to the high number of closely related and morphologically similar species (Smith *et al*. 2005; Hubert *et al*. 2012; Jaafar *et al*. 2012), and Indonesian freshwater fishes are no exception (Kadarusman *et al*. 2012). In addition, several recent biogeographic studies have evidenced a substantial of cryptic diversity in Sundaland freshwater fishes frequently displaying intricate patterns of range distribution straddling across islands and calling for regional management plans (Nguyen *et al*. 2008; Pouyaud *et al*. 2009; De Bruyn *et al*. 2013). The objective of the present study is to provide a DNA barcode reference library of the freshwater fishes of Java and Bali islands with the aim to streamline the ongoing inventory of the Javanese and Balinese ichthyofaunas and to promote more sustainable practices for further taxonomic studies (Hubert & Hanner 2015). The checklist of the Javanese and Balinese freshwater fishes has been assembled from several online catalogues, and implications of this study on the inventory of the Java and Bali ichthyofauna are discussed.

3.3 Materials and methods

3.3.1 Specimen collections and identifications

A total of 3310 specimens, including 162 species, 110 genera and 53 families, have been collected across 95 sites in Java and Bali islands between November 2012 and May 2015 (Fig. 5). Specimens were captured using various gears including electrofishing, seine nets, cast nets and gill nets across sites encompassing the diversity of freshwater lentic and lotic habitats from outlets (i.e. sea level), floodplains, lakes and ponds to upstream tributaries (i.e. 1068 m). Specimens were identified following available monographs (Kottelat *et al*. 1993; Rachmatika 2003; Larson 2009, 2010; Keith *et al*. 2010, 2013), and species names were further validated based on several online catalogues (Froese & Pauly 2011; Eschmeyer & Fricke 2014). Specimens were photographed and individually labelled, and voucher

specimens were preserved in a 5% formalin solution. A fin clip or a muscle biopsy was taken for each specimen and fixed in a 96% ethanol solution for further genetic analyses. Both tissues and voucher specimens were deposited in the national collections at the Muzeum Zoologicum Bogoriense (MZB) in the Research Centre for Biology (RCB) from the Indonesian Institute of Sciences (LIPI).

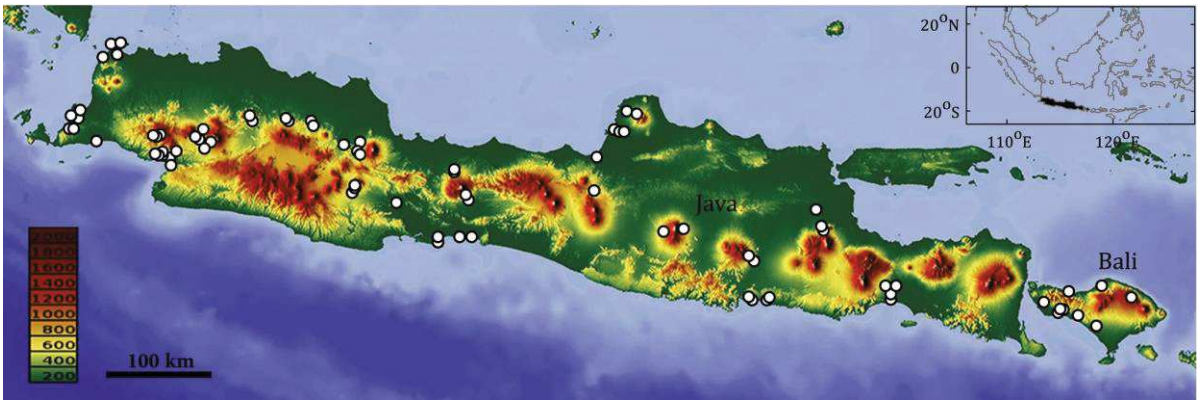


Figure 5. Collection sites for the 1046 samples analysed in this study. Each point may represent several collection sites.

3.3.2 Assembling a checklist of the Javanese and Balinese freshwater fishes

The checklist of the Javanese and Balinese freshwater fishes has been assembled from available online catalogues including FISHBASE (Froese & Pauly 2011) and Eschmeyer catalogue of fishes (Eschmeyer & Fricke 2014) as detailed in Hubert and colleagues (Hubert *et al*. 2015). Occurrences were further refined based on available monographs (Kottelat *et al*. 1993; Kottelat 2013). This checklist was used to estimate the taxonomic coverage of the present DNA barcoding campaign and flag potential new occurrences. Additional information was included as follows: (i) authors of the original description, (ii) maximum observed length, (iii) type of length measurement, (iv) status of the species namely native or introduced, (v) source references for the distribution, (vi) status of the distribution including endemic, occurring in other countries or original distribution range for introduced species, (vii) occurrence in Java, (viii) potential new occurrence in Java, (ix) occurrence in Bali and (x) potential new occurrence in Bali. The database is provided as an online supplementary material (Table S1, Supporting information).

3.3.3 DNA barcode sequencing

A total of 1403 specimens were selected for sequencing in order to cover as much as possible of (i) the intraspecific genetic diversity by selecting specimens throughout the occurrence range, and (ii) the ontogenetic stages for each species by including juveniles, sub

adults and adults. Genomic DNA was extracted by a Qiagen DNeasy 96 tissue extraction kit following the manufacturer's specifications. The standard 652-bp segment from the 5' region of the cytochrome oxidase I gene (COI) was amplified using primer cocktail C_FishF1t1/C_FishR1t1 including a M13 tails (Ivanova *et al*. 2007). PCR amplifications were performed on a Veriti 96-well Fast thermocycler (ABI – Applied Biosystems) with a final volume of 10.0 lL containing 5.0 lL buffer 2X, 3.3 lL ultrapure water, 1.0 lL each primer (10 lM), 0.2 lL enzyme Phire! Hot Start II DNA polymerase (5U) and 0.5 lL of DNA template (~50 ng). Amplifications were conducted as follows: initial denaturation at 98 °C for 5 min followed by 30 cycles of denaturation at 98 °C for 5 s, annealing at 56 °C for 20 s and extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. The PCR products were purified with ExoSap-IT! (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. Sequencing reactions were performed using the BigDye TERMINATOR v3.1 Cycle Sequencing Ready Reaction', and sequencing was performed on the automatic sequencer ABI 3130 DNA Analyzer (Applied Biosystems).

DNA barcodes, photographs, sequences and collection data were deposited on the Barcode of Life Datasystem (BOLD) in the projects 'Barcoding Indonesian Fishes – part II. Inland fishes of Java and Bali [BIFB]', 'Barcoding Indonesian Fishes – part III. Sicydiinae of Sundaland [BIFC]', 'Barcoding Indonesian Fishes – part VIb. Widespread primary freshwater fishes of Java and Bali [BIFGA]', 'Barcoding of Indonesian Fishes – part VIIb Rasbora spp [BIFHB]' in the container 'Barcoding Indonesian Fishes' of the 'Barcoding Fish (FishBOL)' campaign.

3.3.4 DNA barcode analysis

DNA sequence divergence was calculated using the Kimura 2-parameter (K2P) model (Kimura 1980). The mid-point-rooted neighbour-joining (NJ) tree of K2P distances was constructed to provide a graphic representation of the species divergence as implemented in the Sequence Analysis module of BOLD (Ratnasingham & Hebert 2007). Sequence divergence was considered below and above species boundaries by calculating the maximum intraspecific distance and the distance to the closest phylogenetic neighbour in the data set. The distribution of both distances was examined through sequence divergence class of 1% in order to check for a potential overlap between intraspecific and interspecific sequence divergence. We further checked for a DNA barcoding gap in our data (Meyer & Paulay 2005). Instead of considering potential overlap in the distribution of sequence divergences,

we examined the relationships between the maximum intraspecific distance and the distance to the nearest neighbour as it displays potential overlap on an individual basis considering potential overlap for each species (Blagoev *et al*. 2016). Sequence divergence within species, and the potential occurrence of cryptic diversity, was further explored through the Refined Single Linkage (RESL) algorithm to reach decision on the number of operational taxonomic units (OTUs) referenced as Barcode Index Numbers (BIN) in BOLD (Ratnasingham & Hebert 2013).

The taxonomic coverage of the present DNA barcoding campaign was estimated according to the checklist assembled here, and its completeness was further explored through the accumulation curve analysis implemented in the Sequence Analysis module in BOLD (Ratnasingham & Hebert 2007). Accumulation curves were established independently for species and genera across 100 iterations. The taxonomic coverage was further explored through the distribution of the percentage of sampled and nonsampled species across 10-cm size class for euryhaline and amphidromous or primary freshwater families independently.

3.4 Results

3.4.1 DNA barcode analyses and BIN splits

A total of 1046 sequences belonging to 159 species, 107 genera and 50 families were successfully obtained (Table S2, Supporting information). All the sequences were above 500 bp of length and no codon stops were detected, suggesting that the DNA barcodes collected represent functional coding regions. Although the cocktail of primers failed to amplify 357 specimens (i.e. 25%), they were effective in amplifying and sequencing 159 of the 162 species sampled (i.e. 98%) as no DNA barcodes were recovered for only three species, including *Lethrinus fulviflamma*, *Megalops cyprinoides* and an unidentified South American Loricariid assigned to the genus *Ancistrus*.

Table 2. Summary statistics of the genetic distances (K2P) through increasing taxonomic levels

	N(Taxa/ Specimens)	Minimum distance (K2P)	Mean distance (K2P)	Maximum distance (K2P)
Within species	125/1012	0	0.57	12.81
Within genus	36/698	1.56	10.21	23.7
Within family	16/813	6.35	19.35	33.02

Intraspecific distances ranged from 0 to 12.81% and averaged 0.57%, while interspecific distance within genus was 18-fold higher, ranging from 1.56% to 23.7% and averaging 10.21% (Table 1). Although the distribution of the maximum intraspecific distance and the distance to the nearest neighbour overlapped (Fig. 6A,B), nearest neighbour distances were 28-fold higher on average than maximum intraspecific distances (Table S3, Supporting information). Their relationship indicated that maximum intraspecific distances only exceeded nearest neighbour distance in a single species pair and a barcoding gap was generally observed (Fig. 6C). Several species exhibited high maximum intraspecific distances, and BIN counts were higher than species count in 18 cases (Table 2). The highest BIN divergences were observed in *Eleotris fusca* with 12.81% between two BINs and *Giuris margaritacea* with 12.56% between two BINs. The lowest BIN divergences were observed in *Acentrogobius caninus* (two BINS, 1.58%) and *Trichopodus trichopterus* (two BINS, 1.2%). The highest number of BIN was observed in *Channa gachua* with five BINS with a maximum divergence of 5.63% (Table 2).

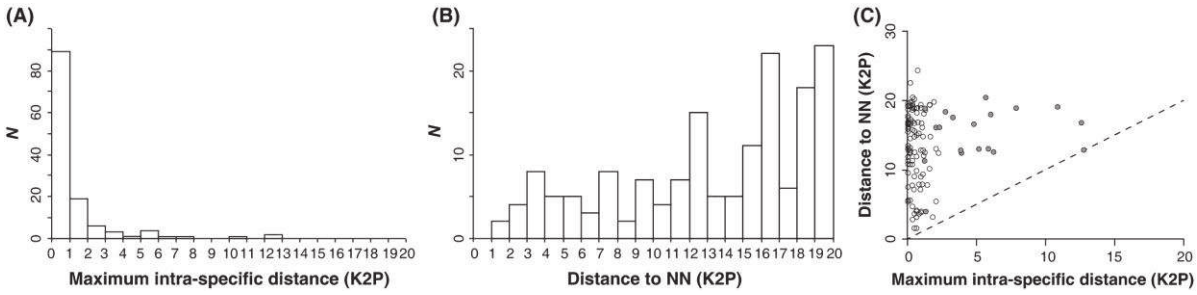


Figure 6. Distribution of genetic distances below and above species boundaries. (A) Distribution of maximum intraspecific distances (K2P). (B) Distribution of nearest neighbour distances (K2P). (C) Relationship between maximum intraspecific and nearest neighbour distances. Points above the diagonal line indicate species with a barcode gap.

Table 3. Summary statistics of the 18 species with more than a single BIN including the number of BIN, maximum intraspecific distance and BIN accession numbers

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Sympatric	BIN
1.	Gobiidae	*Acentrogobius caninus*	2	1.58		BOLD:ACE4188 BOLD:ACT1041
2.	Cyprinidae	*Barbodes binotatus*	3	5.13		BOLD:ACP5712 BOLD:ACP6025 BOLD:ACP6290
3.	Eleotridae	*Belobranchus*	2	3.82		BOLD:ACQ5484

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Sympatric	BIN
		belobranchus				BOLD:ACQ5485
4.	Channidae	*Channa gachua*	5	5.63		BOLD:ACQ0290
						BOLD:ACQ0291
						BOLD:ACQ0292
						BOLD:ACQ6939
						BOLD:ACQ6940
5.	Clariidae	*Clarias gariepinus*	2	2.03		BOLD:AAB2256
						BOLD:ACF4787
6.	Hemiramphidae	*Dermogenys pusilla*	3	2.28		BOLD:ACH7708
						BOLD:ACH7709
						BOLD:ACT1438
7.	Eleotridae	*Eleotris fusca*	2	12.81		BOLD:AAF0108
						BOLD:ACQ5280
8.	Eleotridae	*Eleotris melanosoma*	2	6.2		BOLD:AAF0109
						BOLD:AAK9481
9.	Gerreidae	*Gerres filamentosus*	2	10.83		BOLD:AAC0381
						BOLD:AAY1477
10.	Eleotridae	*Giuris margaritacea*	2	12.56		BOLD:AAV6427
						BOLD:ACP9929
11.	Sisoridae	*Glyptothorax platypogon*	5	5.99		BOLD:AAY1028
						BOLD:ACP5850
						BOLD:ACP5898
						BOLD:ACP6223
						BOLD:ACP6224
12.	Cobitidae	*Lepidocephalichthys hasselti*	2	3.25		BOLD:ACT2693
						BOLD:ACT6514
13.	Mastacembelidae	*Macrognathus maculatus*	4	3.87		BOLD:ACT1648
						BOLD:ACT1649
						BOLD:ACT1650
						BOLD:ACT1890
14.	Synbranchidae	*Monopterus albus*	2	7.83		BOLD:AAF8880
						BOLD:ACT5080

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Sympatric	BIN
15.	Adrianichthyidae	*Oryzias javanicus*	2	2.71		BOLD:ACT2454 BOLD:ACT6896
16.	Gobiidae	*Periophthalmus argentilineatus*	2	5.82		BOLD:AAY1920 BOLD:ACQ9240
17.	Cyprinidae	*Rasbora lateristriata*	2	1.79		BOLD:ACQ7159 BOLD:ACQ7160
18.	Osphronemidae	*Trichopodus trichopterus*	2	1.2		BOLD:AAE8555 BOLD:AAW0021

3.4.2 Species diversity and exotic species

The checklist of the freshwater fishes of Java and Bali yielded a total of 227 species belonging to 181 genera and 66 families (Table S1, Supporting information) among which 216 and 40 have been reported from Java and Bali, respectively (Table 3). The species list obtained from the present DNA barcoding campaign, however, poorly matched the checklist compiled from online catalogues. Amongst the 159 species of the reference library, 85 species have been previously reported from Java and Bali, while 74 species correspond to new records for both islands. In Java, DNA barcodes of only 77 species were recovered of the 227 species previously reported (i.e. 36%), while 75 species with DNA barcodes correspond to new records, increasing the species richness of Java by 130%. A similar trend is observed for Bali as amongst the 40 species previously reported, DNA barcodes were recovered for 24 species (i.e. 60%), while 34 species with DNA barcodes correspond to new records, increasing the diversity by 185%.

Table 4. Summary statistics per families of the taxonomic coverage yielded by this study including the number of species derived from online catalogues, DNA barcoding coverage and new records

Family	*N*species from checklist	*N* species in Java			*N* species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	*N* species	Percentage of overage
Adrianichthyidae	2	2	1 (50)	0	1	1 (100)	0	2	50
Akysidae	3	3	0 (0)	0	0	N/A	0	3	0
Ambassidae	3	3	0 (0)	3	0	N/A	2	7	57

Family	*N*species from checklist	*N* species in Java			*N* species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	*N* species	Percentage of overage
Anabantidae	0	0	N/A	1	0	N/A	0	1	100
Anguillidae	3	3	2 (67)	0	1	0 (0)	1	3	67
Antennariidae	1	1	0 (0)	0	0	N/A	0	1	0
Aplocheilidae	1	1	1 (100)	0	1	1 (100)	0	1	100
Apogonidae	2	2	0 (0)	0	1	0 (0)	0	2	0
Ariidae	3	3	1 (33)	0	0	N/A	0	3	33
Bagridae	9	9	4 (44)	0	0	N/A	0	9	44
Balitoridae	4	4	0 (0)	0	0	N/A	0	4	0
Belonidae	0	0	N/A	2	0	N/A	0	2	100
Blenniidae	1	1	0 (0)	0	0	N/A	0	1	0
Carangidae	0	0	N/A	1	0	N/A	0	1	100
Chanidae	0	0	N/A	1	0	N/A	0	1	100
Channidae	4	4	2 (50)	0	3	2 (67)	0	4	50
Cichlidae	0	0	N/A	8	0	N/A	2	8	100
Clariidae	3	3	1 (33)	1	0	N/A	0	4	50
Cobitidae	8	8	2 (25)	1	0	N/A	0	9	44
Cyprinidae	44	43	14 (33)	6	3	3 (100)	0	49	41
Dasyatidae	1	1	0 (0)	0	0	N/A	0	1	0
Eleotridae	11	10	6 (60)	4	4	4 (100)	5	15	73
Engraulidae	2	2	0 (0)	2	1	0 (0)	0	4	50
Gerreidae	1	1	1 (100)	1	1	1 (100)	0	2	100
Gobiidae	38	30	17 (57)	18	19	10 (53)	14	54	70
Haemulidae	1	1	0 (0)	1	0	N/A	0	2	50
Helostomatidae	0	0	N/A	1	0	N/A	0	1	100
Kuhliidae	2	2	1 (50)	0	0	N/A	0	2	50
Latidae	0	0	N/A	1	0	N/A	0	1	100
Leiognathidae	1	1	0 (0)	1	0	N/A	0	2	50
Loricariidae	0	0	N/A	1	0	N/A	0	1	100
Lutjanidae	1	1	1 (100)	0	0	N/A	0	1	100

Family	*N*species from checklist	*N* species in Java			*N* species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	*N* species	Percentage of overage
Mastacembelidae	3	3	2 (67)	1	0	N/A	0	4	75
Megalopidae	1	1	0 (0)	0	0	N/A	0	1	0
Monodactylidae	1	1	0 (0)	0	0	N/A	0	1	0
Mugilidae	6	6	0 (0)	0	0	N/A	0	6	0
Muraenidae	2	2	0 (0)	0	0	N/A	0	2	0
Nandidae	1	1	0 (0)	0	0	N/A	0	1	0
Nemacheilidae	2	2	2 (100)	0	0	N/A	0	2	100
Notopteridae	3	3	1 (33)	0	0	N/A	0	3	33
Ophichthidae	1	1	0 (0)	2	0	N/A	2	4	75
Osphronemidae	4	4	3 (75)	1	0	N/A	1	5	80
Pangasiidae	2	2	0 (0)	2	0	N/A	0	4	50
Platycephalidae	0	0	N/A	1	0	N/A	0	1	100
Plotosidae	1	1	1 (100)	1	0	N/A	0	2	100
Poeciliidae	0	0	N/A	4	0	N/A	4	5	100
Pristidae	2	2	0 (0)	0	0	N/A	0	2	0
Pristigasteridae	2	2	0 (0)	0	0	N/A	0	2	0
Pristolepididae	1	1	0 (0)	0	0	N/A	0	1	0
Rhyacichthyidae	1	1	1 (100)	0	1	1 (100)	0	1	100
Scatophagidae	1	1	1 (100)	0	0	N/A	0	1	100
Schilbeidae	1	1	1 (100)	0	0	N/A	0	1	100
Sciaenidae	4	4	2 (50)	0	0	N/A	0	4	50
Serrasalmidae	0	0	N/A	1	0	N/A	0	1	100
Siganidae	0	0	N/A	2	0	N/A	0	2	100
Sillaginidae	2	2	0 (0)	2	1	0 (0)	0	4	50
Siluridae	11	11	0 (0)	0	0	N/A	0	11	0
Sisoridae	2	2	1 (50)	0	0	N/A	0	2	50
Sparidae	1	1	0 (0)	0	0	N/A	0	1	0
Synbranchidae	1	1	1 (100)	0	0	N/A	1	1	100

Family	*N*species from checklist	*N* species in Java			*N* species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	*N* species	Percentage of overage
Syngnathidae	8	7	3 (43)	2	1	1 (100)	2	9	56
Terapontidae	3	3	1 (33)	0	0	N/A	0	3	33
Tetraodontidae	4	4	0 (0)	1	0	N/A	0	5	20
Tetrarogidae	3	3	1 (33)	0	2	0 (0)	0	3	33
Toxotidae	1	1	1 (100)	0	0	N/A	0	1	100
Zenarchopteridae	3	3	1 (33)	1	0	N/A	0	4	50
Total	227	216	77 (36)	75	40	24 (60)	34	301	

The accumulation curves displayed an asymptotic trend, indicating that the present sampling was representative of the Java and Bali ichthyofauna, whatever the taxonomic level considered (Fig. 7A).

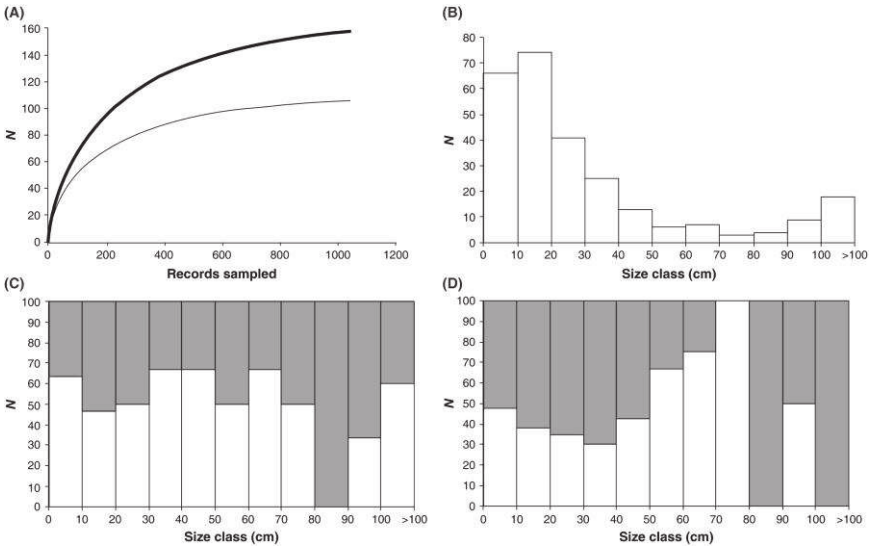


Figure 7. Accumulation curves and size class distributions of the 159 species analysed in this study. (A) Accumulation curves recovered from 100 iterations for species (bold curve) and genera (regular curve). (B) Distribution of size class (10 cm) of 266 species with documented maximum sizes among the 301 species of Java and Bali. (C) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 164 species with documented maximum length for euryhalin or amphidromous families. (D) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 135 species with documented maximum length for primary freshwater families.

We further examined the distribution of the taxonomic gap in the present library through size class of 10 cm. The Javanese and Balinese ichthyofauna are dominated by species smaller than 20 cm, representing nearly 70% of the overall species pool (Fig. 7B). The distribution of the taxonomic gap through size class varied between euryhaline or amphidromous and primary freshwater families (Fig. 7C,D). A poor taxonomic coverage was recovered at the largest size classes for primary freshwater family (Fig. 7D), as 12 of the 19 species exceeding

70 cm of maximum length were not sampled (i.e. 73%). Although nearly 45% of the largest species (i.e. >70 cm) for euryhaline and amphidromous families were not sampled, the taxonomic coverage was more evenly distributed across size classes, excepting the 80- to 90-cm size class with a single species that was not sampled (Fig. 7C). The influence of species size on taxonomic coverage, however, was not significant for either euryhaline and amphidromous families (v2 = 6.22; P = 0.79) or primary freshwater families (v2 = 11.38; P = 0.33), despite that more heterogeneity was observed for the later.

Amongst the 74 new records for both Java and Bali, a total of 20 introduced species belonging to 16 genera and eight families were newly recorded during the present campaign (Table 4). All of them displayed a single BIN with very low maximum intraspecific distances, excepting for *Clarias gariepinus* with two BINs diverging by 2.03% and *Poecilia reticulata* with two BINs diverging by 4.77%. The families Cichlidae and Poeciliidae were the most speciose with eight and five species, respectively. Amongst the 20 introduced species, four species originate from South America, nine from North and Central America, three from Africa and four from Asia.

Table 5. Summary statistics of the 20 exotic species including the number of BIN, maximum intraspecific distance, geographic origin and BIN accession numbers

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Origin	BIN
1.	Cichlidae	*Amphilophus citrinellus*	1	–	Central America (Nicaragua, Costa Rica)	BOLD:AAA7015
2.	Cichlidae	*Andinoacara rivulatus*	1	0	South America (Ecuador, Peru)	BOLD:AAJ5733
3.	Cichlidae	*Archocentrus nigrofasciatus*	1	0	Central America (Panama)	BOLD:AAD6407
4.	Cichlidae	*Oreochromis mossambicus*	1	0	Africa (Mozambique, Malawi)	BOLD:AAA8511
5.	Cichlidae	*Oreochromis niloticus*	1	–	Africa (widely distributed)	BOLD:ACR5811
6.	Cichlidae	*Parachromis managuensis*	1	–	Central America (Honduras, Costa Rica)	BOLD:AAB8227
7.	Cichlidae	*Paraneetroplus fenestratus*	1	–	North America (Mexico)	BOLD:AAD2571
8.	Cichlidae	*Paraneetroplus maculicauda*	1	–	Central America (Guatemala, Panama)	BOLD:AAB9907
9.	Clariidae	*Clarias gariepinus*	2	2.03	Africa (widely	BOLD:AAB2256

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Origin	BIN
					distributed)	BOLD:ACF4787
10.	Cyprinidae	*Ctenopharyngodon idella*	1	0.16	Asia (China, Russia)	BOLD:ACL1923
11.	Cyprinidae	*Puntigrus tetrazona*	1	0	Asia (Indonesia: Sumatra, Borneo)	BOLD:AAD9761
12.	Loricariidae	*Liposarcus pardalis*	1	0	South America (Brazil, Peru, Bolivia)	BOLD:ACK1995
13.	Osphronemidae	*Trichopodus pectoralis*	1	0	Asia (Vietnam, Thailand, Cambodia)	BOLD:AAE8555
14.	Pangasiidae	*Pangasianodon hypophthalmus*	1	0	Asia (Vietnam, Thailand, Cambodia)	BOLD:AAE3237
15.	Poeciliidae	*Gambusia affinis*	1	0.72	North and Central America (USA, Mexico)	BOLD:AAC2756
16.	Poeciliidae	*Poecilia latipinna*	1	0	North America (USA, Mexico)	BOLD:ACE4147
17.	Poeciliidae	*Poecilia reticulata*	2	4.77	South America (widely distributed)	BOLD:ACC0443 BOLD:ACE3484
18.	Poeciliidae	*Xiphophorus hellerii*	1	0.16	North and Central America (Mexico to Honduras)	BOLD:AAB8020
19.	Poeciliidae	*Xiphophorus maculatus*	1	0	North and Central America (USA, Mexico)	BOLD:AAB7239
20.	Serrasalmidae	*Colossoma macropomum*	1	0.16	South America (widely distributed)	BOLD:AAD6423

3.5 Discussion

The present study provides a DNA barcode reference library consisting of 159 species, including 139 native and 20 exotic species. Although maximum intraspecific distances exceeded nearest neighbour distances in a species pair (i.e. Eleotris fusca and E. melanosoma), all the species examined consisted of a diagnostic set of DNA barcodes arranged into monophyletic units. This study revealed that nearest neighbour distances were 28-fold higher on average than intraspecific maximum distances, suggesting that the present library is useful for the identification of Javanese and Balinese freshwater fishes. The average distance among congeneric, however, was 18- fold higher than the average intraspecific

distance (0.57% vs. 10.21%), a range lower than that observed for maximum intraspecific and nearest neighbour distances. This difference was not unexpected considering the geographic focus of the present study. Nearest neighbour distances were estimated based on a geographically focused sampling. Considering that the Java and Bali islands include 227 species of the 900 species in Sundaland, the closest phylogenetic relatives have been probably sampled in only a few cases, thereby inflating estimates of nearest neighbour distances. This bias is not unexpected considering that the Pleistocene eustatic fluctuations repeatedly connected Sundaland major islands and offered temporary opportunities of dispersal during times of low sea levels and subsequent fragmentation during high sea level periods (Voris 2000; Woodruff 2010). As a consequence, patterns of allopatric speciation and trans-islands sister-species relationships have been frequently described (Dodson *et al*. 1995; Nguyen *et al*. 2008; Pouyaud *et al*. 2009; De Bruyn *et al*. 2013).

Although the barcoding gap was large in many species, expanding the spatial coverage of the present library will yield an increasing amount of closely related species, calling for its reassessment based on a more comprehensive spatial and taxonomic sampling. Along the same line, expanding the spatial coverage may be expected to reassess upward maximum intraspecific distances as previously demonstrated (Bergsten *et al*. 2012), particularly in fragmented landscapes (Geiger *et al*. 2014). Considering that very large maximum intraspecific distances were observed at the island level, even in numerous species consisting of a single BIN, the bias towards underestimated intraspecific distances may be high when incorporating a more comprehensive coverage of the distribution range for species straddling across the islands of Sundaland. The extent of this bias, however, is tightly linked to the dynamic that generated these high levels of intraspecific genetic diversity. If intraspecific divergence results from admixtures after secondary contacts of ancient populations previously isolated in allopatry, the bias may be limited. If intraspecific divergence results from the isolation of populations throughout fragmented landscapes, this bias may be expected to be high. Considering the intricate history of merging and isolation of populations, as a consequence of the Pleistocene eustatic fluctuations (Voris 2000; Woodruff 2010), both dynamics are likely to be involved and warrant further assessment of the barcoding gap at larger spatial scales in Sundaland. This hypothesis happens to be likely considering that the present study revealed 18 species with several BINs displaying a large extent of maximum divergence (1.2–12.81%).

The present study provided surprising results if considering the taxonomic coverage. Amongst the 227 species previously reported from Java and Bali, only 36% and 60% were

sampled, respectively. This observation is dramatically contrasted by the 74 new records, among which 54 likely correspond to native species. Considering that most native species in Java and Bali correspond to old descriptions (Table S1, Supporting information) and that no comprehensive inventories of the Java and Bali ichthyofauna have been conducted during the last decade, new records were not unexpected, particularly for exotic species.

The international trade in ornamental fishes has grown rapidly over the last decades, as well as the aquaculture trade, resulting in a global homogenization of the world ichthyofauna (Leprieur *et al*. 2007; Blanchet *et al*. 2010). The exotic species observed during the present campaign are among the most common of the ornamental (e.g. *Xiphophorus* spp., *Poecilia* spp., *Paraneetroplus* spp. *Archocentrus nigrofasciatus*, *Puntigrus tetrazona*) or aquaculture trade (e.g. *Oreochromis* spp., *Pangasianodon hypophthalmus*, *Clarias gariepinus*). Considering that Indonesia rank as one of the major contributors of the international ornamental fish trade (Ling & Lim 2005), this result highlights that avoiding escaping from fish farms or releasing in nature by fish hobbyist is challenging, particularly for species with high adaptive abilities in terms of breeding such as *Oreochromis* spp., *Xiphophorus* spp. or *Poecilia* spp. This situation is further amplified by the lack of knowledge by local populations of the geographic origin of most ornamental fish species traded in Indonesia and the impact of artificial introduction of exotic species. During the present survey, restocking with exotic species for recreational purposes was frequently observed in both Java (e.g. Lake Rawa Pening) and Bali (e.g. Lake Batur).

The observation of 54 new occurrences of species with no obvious economic interest for the ornamental trade is more challenging in terms of basic knowledge about the taxonomic composition of the Java and Bali ichthyofauna. This is particularly evident considering that this high number of new occurrences is drastically contrasted by the poor taxonomic coverage of the present campaign for the previously reported species according to online catalogues. Identifications were performed independently by several of the co-authors based on available monographs and later cross-validated, including based on DNA barcode data available in BOLD. Misidentifications are unlikely to account alone for this apparent discrepancy, an assertion further supported by the asymptotic trend of the accumulation curve for both species and genera. Among the 16 species endemic of Java, for instance, only three were sampled, including *Nemacheilus chrysolaimos*, *Rasbora aprotaenia* and *Sicyopterus parvei*. The status of these endemic species should be revised, however, as *S. parvei* have been observed also in Bali. By contrast, the two endemic species of Bali were sampled (i.e. *Lentipes whittenorum* and *Rasbora baliensis*), as well as the three species endemic of Java

and Bali (i.e. *Lentipes ikeae*, *Sycopus rubicundus* and *Stiphodon aureofuscus*). The validity of several endemic species of Java has been previously discussed as a substantial proportion of them correspond to old descriptions based on a single specimen such as *Barbodes platysoma*, *Mystus abbreviatus*, *Ompok javanesis*, *Puntius aphya* and *Puntius bramoides* (Kottelat 2013; Eschmeyer & Fricke 2014). The contrasted coverage of the campaigns conducted in Java and Bali urther questions their validity as sampling has been conducted nearby the type locality for several of them. Nomenclatural issues, however, are not sufficient per se to account for the large gap in the present taxonomic coverage, and extirpations are also likely considering the exponential increase of human populations, and associated anthropogenic perturbations, in Java. Despite that the influence of species maximum size on taxonomic coverage was not significant, for either euryhalin and amphidromous or primary freshwater families, many large-sized and emblematic primary freshwater species are missing from Java, as for instance *Pangasius djambal* (90 cm of SL), *Chitala chitala* (122 cm of SL), *Chitala lopis* (150 cm of SL), *Bagarius bagarius* (200 cm of TL), *Wallago attu* (240 cm of TL), *Tor tambra* (100 cm of TL) and *Tor soro* (100 cm of TL). Considering that several inland fisheries have been visited several times during the course of the study, including the Mojokerto fish market aggregating fisheries of the largest Javanese river (i.e. Brantas River), extirpations as a consequence of overexploitation cannot be discarded. Worth mentioning, the presence of most of these large species in Java is derived from the early ichthyological exploration of Javanese inland waters several decades or centuries ago (Roberts 1993).

3.6 Conclusions

The present study highlights the difficulty to develop accurate DNA barcode reference libraries in Sundaland. Despite two centuries of ichthyological exploration of the inland waters of Java, the build-up of the taxonomic knowledge has been scattered in the literature, and as a consequence, the taxonomy of Javanese fishes has been accessible to only but a few specialists (Kottelat 2013). Considering the large discrepancies between the checklist established here and those available on online catalogues, the inventory of Javanese fishes is still far from comprehensive and is currently plagued by uncertainties in the validity and occurrence of many species. The development of this DNA barcode library will allow more researchers to explore the ichthyodiversity in this part of Sundaland, what will without doubt help refine the checklist of Javanese fishes and probably shed a new light on the validity of the endemic species of Java. The Sundaland ichthyofauna is one of the world's most

endangered and establishing accurate checklists for its major islands is an absolute priority. This library is a primer to that end.

Acknowledgements

The authors wish to thank Dr. Siti Nuramaliati Prijono, Dr. Witjaksono, Mohammad Irham M.Sc., Dr. Marlina Adriyani, Dr. Rosichon Ubaidillah, Dr. Sri Sulandari, at Research Centre for Biology (RCB-LIPI), Dr. Jean-Paul Toutain, Dr. Jean-François Agnèse and Dr. Domenico Caruso from the 'Institut de Recherche pour le Développement', Dr. Joel LeBail at the French embassy in Jakarta, Dr. Bambang Suryobroto and Dr. Achmad Farajallah at the Bogor Agronomy University for their support as well as Dr. Lukas Rüber and Dr. Fabian Herder for constructive discussions. We are thankful to Dr. Daisy Wowor and Dr. Daniel Lumbantobing at RCB-LIPI, Sumanta and Bambang Dwisusilo at IRD Jakarta and Dr. Helen Larson from the Museum and Art Gallery of the Northern Territory for their help during either the field sampling or the identifications of the specimens collected in Java and Bali. Part of the present study was funded by the Institut de Recherche pour le Développement (UMR226 ISE-M and IRD headquarter through incentive funds), the MNHN (UMR BOREA), the RCB-LIPI, the French Ichthyological Society (SFI), the Foundation de France and the French embassy in Jakarta. The Indonesian Ministry of Research approved this study, and field sampling was conducted according to the research permits 097/SIP/FRP/SM/IV/2014 for Philippe Keith, 60/EXT/SIP/FRP/SM/XI/2014 for Frédéric Busson and 41/EXT/SIP/FRP/SM/VIII/2014 for Nicolas Hubert. Sequence analysis was aided by funding from the Government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project. This publication has ISEM number 2015-204.

References

Bergsten, J, Bilton, DT, Fujisawa, T *et al*. (2012) The effect of geographical scale of sampling on DNA barcoding. Systematic Biology, 61, 851– 869.

Blagoev, GA, deWaard, JR, Ratnasingham, S *et al*. (2016) Untangling taxonomy: a DNA barcode reference library for Canadian spiders. Molecular Ecology Resources, 16, 325–341.

Blanchet, S, Grenouillet, G, Beauchard, O *et al*. (2010) Non-native species disrupt the worldwide patterns of freshwater fish body size: implications for Bergmann's rule. Ecology Letters, 13, 421– 431.

Clements, R, Sodhi, NS, Schilthuizen, M, Ng, PKL (2006) Limestone karsts of Southeast Asia: imperiled arks of biodiversity. BioScience, 56, 733– 742.

Clements, R, Ng, PKL, Lu, XX *et al*. (2008) Using biogeographical patterns of endemic land snails to improve conservation planning for limestone karsts. Biological Conservation, 141, 2751– 2764.

De Bruyn, M, Rüber, L, Nylinder, S *et al*. (2013) Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast asian biodiversity hotspots. Systematic Biology, 62, 398– 410.

Dodson, JJ, Colombani, F, Ng, PKL (1995) Phylogeographic structure in mitochondrial DNA of a South-east Asian freshwater fish, Hemibagrus nemurus (Sluroidei; Bagridae) and Pleistocene sea-level changes on the Sunda shelf. Molecular Ecology, 4, 331– 346.

Eschmeyer, WN, Fricke, R (2014) Catalog of fishes electronic version (updated 6 October 2014. San Francisco.

Fraser, S (2006) Threats to biodiversity. In: The Ecology of Indonesia (eds AJ Marshall, BM Beehler), pp. 1199– 1229. Periplus Editions, Singapore.

Froese, R, Pauly, D (2011) Fishbase. http://www.fishbase.org

Geiger, MF, Herder, F, Monaghan, MT *et al*. (2014) Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes. Molecular Ecology Resources, 14, 1210– 1221.

Hoffman, M, Hilton-Taylor, C, Angulo, A *et al*. (2010) The impact of conservation on the status of the world's vertebrates. Science, 330, 1503– 1509.

Hubert, N, Hanner, R (2015) DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes, 3, 44– 58.

Hubert, N, Meyer, C, Bruggemann, JH *et al*. (2012) Cryptic diversity in Indo-Pacific coral reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. PLoS ONE, 7, e28987.

Hubert, N, Kadarusman, , Wibowo, A *et al*. (2015) DNA barcoding Indonesian freshwater fishes: challenges and prospects. DNA Barcodes, 3, 144– 169.

Ivanova, NV, Zemlak, TS, Hanner, RH, Hébert, PDN (2007) Universal primers cocktails for fish DNA barcoding. Molecular Ecology Notes, 7, 544– 548.

Jaafar, TNAM, Taylor, MI, Mohd Nor, SA, De Bruyn, M, Carvalho, GR (2012) DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay carangidae (Teleosteii: Perciformes). PLoS ONE, 7, e49623.

Kadarusman, , Hubert, N, Hadiaty, RK *et al*. (2012) Cryptic diversity in Indo-Australian rainbowfishes revealed by DNA Barcoding: implications for conservation in a biodiversity hotspot candidate. PLoS ONE, 7, e40627.

Keith, P, Marquet, G, Lord, C, Kalfatak, D, Vigneux, E (2010) Poissons et Crustacés D'eau Douce du Vanuatu. Société Française d'Ichtyologie, Paris, France.

Keith, P, Marquet, G, Gerbeaux, P, Vigneux, E, Lord, C (2013) Poissons et Crustacés D'eau Douce de Polynésie. Société Française d'Ichthyologie, Paris, France.

Kimura, M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 15, 111– 120.

Kottelat, M (2013) The fishes of the inland waters of Southeast Asia: a catalog and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. The Raffles Bulletin of Zoology, 27(Supplement), 1– 663.

Kottelat, M, Whitten, AJ, Kartikasari, SR, Wirjoatmodjo, S (1993) Freshwater Fishes of Western Indonesia and Sulawesi. Periplus Editions, Singapore.

Lamoureux, JF, Morrison, JC, Ricketts, TH *et al*. (2006) Global tests of biodiversity concordance and the importance of endemism. Nature, 440, 212– 214.

Larson, HK (2009) Review of the gobiid fish genera Eugnathogobius and Pseudogobiopsis (Gobioidei: Gobiidae: Gobionellinae), with descriptions of three new species. The Raffles Bulletin of Zoology, 57, 127– 181.

Larson, HK (2010) A review of the gobiid fish genus Redigobius (Teleostei: Gobionellinae); with descriptions of two new species. Ichthyological Exploration of Freshwaters, 21, 123– 191.

Leprieur, F, Beauchard, O, Blanchet, S, Oberdorff, T, Brosse, S (2007) Fish invasions in the world's river systems: when natural processes are blurred by human activities. Plos Biology, 6, 404– 410.

Ling, KH, Lim, LY (2005) The status of ornamental fish industry in Singapore. Singapore Journal of Primary Industries, 32, 59– 69.

Lohman, K, De Bruyn, M, Page, T *et al*. (2011) Biogeography of the Indo-Australian archipelago. Annual Review of Ecology, Evolution and Systematics, 42, 205– 226.

Meyer, C, Paulay, G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology, 3, 2229– 2238.

Myers, N, Mittermeier, RA, Mittermeier, CG, Fonseca, GAB, Kent, J (2000) Biodiversity hotspots for conservation priorities. Nature, 403, 853– 858.

Nguyen, TTT, 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.

Normile, D (2010) Saving forests to save biodiversity. Science, 329, 1278– 1280.

Pouyaud, L, Sudarto, , Paradis, E (2009) The phylogenetic structure of habitat shift and morphological convergence in Asian Clarias (Teleostei, Siluriformes: Clariidae). Journal of Zoological Systematics and Evolutionary Research, 47, 344– 356.

Rachmatika, I (2003) Fish fauna of the Gunung Halimun national park, West Java. Biodiversity Conservation Project LIPI-JICA-PHKA, Jakarta, Indonesia.

Ratnasingham, S, Hebert, PDN (2007) Bold: the barcode of life data system (http://www.barcodinglife.org). Molecular Ecology Notes, 7, 355– 364.

Ratnasingham, S, Hebert, PDN (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE, 8, e66213.

Roberts, TR (1993) The freshwater fishes of Java, as observed by Kuhl and van Hasselt in 1820-23. Zoologische Verhandelingen, 285, 3– 94.

Schilthuizen, M, Liew, TS, Bin Elehan, B, Lackman-Ancrenaz, I (2005) Effects of karst forest degradation on pulmonate and prosobranch land snail communities in Sabah, Malaysian Borneo. Conservation Biology, 19, 949– 954.

Smith, AM, Fisher, BL, Hebert, PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society, Serie B, 360, 1825– 1834.

Sodhi, NS, Koh, LP, Clements, R *et al*. (2010) Conserving southeast asian forest biodiversity in human-modified landscapes. Biological Conservation, 143, 2375– 2384.

Voris, HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. Journal of Biogeography, 27, 1153– 1167.

Woodruff, DS (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugium-phase biodiversity. Biodiversity and Conservation, 19, 919– 941.

Data accessibility

All collecting and sequence data are available on the Barcode of Life Datasystem (BOLD) in the projects 'Barcoding Indonesian Fishes – part II. Inland fishes of Java and Bali [BIFB]', 'Barcoding Indonesian Fishes – part III. Sicydiinae of Sundaland [BIFC]', 'Barcoding Indonesian Fishes – part VIb. Widespread primary freshwater fishes of Java and Bali [BIFGA]', 'Barcoding of Indonesian Fishes – part VIIb Rasbora spp [BIFHB]' in the container 'Barcoding Indonesian Fishes' of the 'Barcoding Fish (FishBOL)' campaign. The sequence alignment and neighbour-joining tree (as both PDF and Newick files) have all been uploaded to DRYAD (doi: 10.5061/dryad.tk5rj). The sequences are also available on GenBank (see Table S2, Supporting information for accession numbers).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Midpoint rooted Neighbor-joining tree of the 1046 DNA barcodes collected from the 159 species analyzed in this study.

Table S1 Checklist of the freshwater fishes of Java and Bali including the authors and date of the original description, maximum length, type of length measurement, status of the species namely native or introduced, source references for the distribution, status of the distribution including endemic, occurring in other countries or original distribution range for introduced species, occurrence in Java, potential new occurrence in Java, occurrence in Bali, potential new occurrence in Bali.

Table S2 Collecting data and sequence information.

Table S3 Barcoding gap in the species analyzed in the present study including mean intra-specific, maximum intra-specific and nearest neighbor distances for the 159 species analyzed in the present study.

IV. Validity and distribution of endemic species of *Rasbora* spp. and *Nemacheilus* spp. in Java through DNA Barcodes

Article 2: Revisiting species boundaries and distribution ranges of *Nemacheilus* spp. (Cypriniformes: Nemacheilidae) and *Rasbora* spp. (Cypriniformes: Cyprinidae) in Java, Bali and Lombok through DNA barcodes: implications for conservation in a biodiversity hotspot.

Nicolas Hubert[1], Daniel Lumbantobing[2], Arni Sholihah[1,3], Hadi Dahruddin[2], Erwan Delrieu-Trottin[1,4], Frédéric Busson[1,5], Sopian Sauri[2], Renny Hadiaty[2], Philippe Keith[5]

[1] Institut de Recherche pour le Développement, UMR 226 ISEM (UM, CNRS, IRD, EPHE), Université de Montpellier, Montpellier cedex 05, France
[2] Division of Zoology, Research Center for Biology, Indonesian Institute of Sciences (LIPI), Cibinong, Indonesia
[3] Instut Teknologi Bandung, School of Life Science and Technology, Bandung, Indonesia
[4] Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Berlin, Germany
[5] UMR 7208 BOREA (MNHN-CNRS-UPMC-IRD-UCBN), Muséum National d'Histoire Naturelle, Paris cedex 05, France

4.1 Abstract

Biodiversity hotspots have provided useful geographic proxies for conservation efforts. Delineated from a few groups of animals and plants, biodiversity hotspots do not reflect the conservation status of freshwater fishes. With hundreds of new species described on a yearly basis, fishes constitute the most poorly known group of vertebrates. This situation urges for an acceleration of the fish species inventory through fast and reliable molecular tools such as DNA barcoding. The present study focuses on the freshwater fishes diversity in the Sundaland biodiversity hotspot in Southeast Asia. Recent studies evidenced large taxonomic gaps as well as unexpectedly high levels of cryptic diversity, particularly so in the islands of Java and Bali. The Cypriniformes genera *Rasbora* and *Nemacheilus* account for most of the endemic species in Java and Bali, however their taxonomy is plagued by confusion about species identity and distribution. This study examines the taxonomic status of the *Rasbora* and *Nemacheilus* species in Java, Bali and Lombok islands through DNA barcodes, with the objective to resolve taxonomic confusion and identify trends in genetic diversity that can be

further used for conservation matters. Several species delimitation methods based on DNA sequences were used and confirmed the status of most species, however several cases of taxonomic confusion and two new taxa are detected. Mitochondrial sequences argue that most species range distributions currently reported in the literature are inflated due to erroneous population assignments to the species level, and further highlight the sensitive conservation status of most *Rasbora* and *Nemacheilus* species on the islands of Java, Bali and Lombok.

4.2 Introduction

Biodiversity hotspots are characterized by high proportions of endemic species and high levels of anthropogenic threats (Myers *et al*. 2000). Identified to maximize conservation efforts in a world with finite human and funding resources for conservation matters, biodiversity hotspots have provided useful geographic proxies for conservation efforts. While those biodiversity hotspots have been delineated based on a limited set of well-known vertebrate taxa such as mammals, birds, amphibians and reptiles, the diversity and status of the world's most diverse vertebrate group, that is fishes, is still largely unknown (Myers *et al*. 2000; Lamoreux *et al*. 2006; Hoffman *et al*. 2010). With hundreds of new species described on a yearly basis, freshwater fishes suffer from an important taxonomic knowledge gaps that, combined with the taxonomic impediment (i.e. the rarefaction of taxonomists worldwide), currently plagues conservation efforts in most biodiversity hotspots (Winemiller *et al*. 2016; Garnett and Christidis 2017). This situation arguably accounts for their exclusion from most of the large-scale meta-analyses conducted so far on global diversity patterns (Myers *et al*. 2000; Lamoreux *et al*. 2006; Hoffman *et al*. 2010). In insular South-East Asia (SEA), the Sundaland hotspot exemplifies the stakes faced by conservation stakeholders due to antagonistic interests in the use of biological resources. Including the islands of Java, Sumatra and Borneo, Sundaland is currently among the largest hotspots in terms of number of species and endemics (Myers *et al*. 2000). Recent threat analyses, however rank it as one of the most threatened (Lamoreux *et al*. 2006; Hoffman *et al*. 2010). With nearly 900 species and 430 endemics, Sundaland accounts respectively for 74% and 48% of the total and endemic diversity of the approximately 1200 fish species cited from rivers of the Indonesian archipelago (Hubert *et al*. 2015). Within Sundaland, Java exhibits one the highest fish species density with 1.7 species/1000 km2 (213 species) together with Sumatra (460 species) and ahead of Kalimantan (Indonesian Borneo; 1.2 species per 1000 km2 and 646 species).

Hosting 130 million of people sharing 130,000 km2, Javanese aquatic ecosystems have faced a dramatic increase of anthropogenic threats during the last decade.The recent molecular inventory of the Javanese ichthyofauna evidenced large discrepancies between the checklist of Java freshwater fishes established from historical records (Hubert *et al*. 2015) and a modern reappraisal based on DNA sequences (Dahruddin *et al*. 2017), hence highlighting major gaps in the taxonomic knowledge of this ichthyofauna. Along the same line Hutama *et al*. (2017) evidenced high levels of cryptic diversity (i.e. morphologically unnoticed diversity) in widespread fish species of Java deriving from a late Pleistocene fragmentation of the populations associated with population bottlenecks. Considering that Sundaland is currently in a refugial state and that its emerged lands represent only a small fraction of its average surface during the Pleistocene (Woodruff 2010; Lohman *et al*. 2011), the state of Sundaland ichthyofauna urges for an acceleration of the ichthyological exploration of its freshwaters.

Initially designed to circumvent the taxonomic impediment by proposing a standard molecular framework for species identification through the use of the mitochondrial cytochrome oxidase I gene as an internal species tag, DNA barcoding opened new perspectives in the inventory of freshwater fishes (Hubert *et al*. 2008; Ward *et al*. 2009; Steinke and Hanner 2011). While large scale fish DNA barcoding campaigns have been tackled during the last decade (April *et al*. 2011; Hubert *et al*. 2012, 2018; Pereira *et al*. 2013; Geiger *et al*. 2014; Knebelsberger *et al*. 2015; Dahruddin *et al*. 2017; Durand *et al*. 2017; Machado *et al*. 2018), it becomes more and more evident that the pace of species description is surpassed by the astonishing underestimation of species diversity, often referring to cryptic diversity, and the complexity of fish biodiversity (Hubert *et al*. 2012; Jaafar *et al*. 2012; Kadarusman *et al*. 2012; Geiger *et al*. 2014; Winterbottom *et al*. 2014). We focus in the present study on the diversity and range distribution in South Sundaland of two Cypriniformes genera, namely *Rasbora* (Cyprinidae) and *Nemacheilus* (Nemacheilidae) that constitute emblematic endemic lineages in Java and Lesser Sunda Islands (Bali, Lombok) due to their occurrence in a large array of aquatic ecosystems and their high levels of endemism compared to other genera occurring in Java. Mostly described during the eighteenth and nineteenth centuries, *Rasbora* and *Nemacheilus* taxonomy and distribution is confusing in Java due to the lack of traceability of the taxonomic information often associated with old descriptions. Type localities are available for most of these species (Kottelat 2013), however range distribution are currently unknown (Froese and Pauly 2014; Hubert *et al*. 2015; Eschmeyer *et al*. 2018), most Rasbora and Nemacheilus species being reported in Java and/or

Bali without further details. With the aim to re-examine *Rasbora* and *Nemacheilus* diversity on the islands of Java, Bali and Lombok, we produced a DNA barcode reference library with the following objectives: (1) exploration of species biological boundaries through DNA-based species delimitation methods, (2) validation of species identity and taxonomy and precise range distribution by producing DNA barcodes from type localities or neighboring watersheds, (3) estimation of species genetic diversity and production of recommendations for conservation genetics purposes.

4.3 Materials and methods

4.3.1 Sampling and collection management

The authors previously conducted a large-scale DNA barcoding campaign across 95 sites in Java and Bali Island between November 2012 and May 2015 (Dahruddin *et al*. 2017). During this initial inventory, a total of 3310 specimens, including 162 species belonging to 110 genera and 53 families were collected. This was complemented by an additional campaign in Lombok island on March 2015 resulting in the sampling of an additional set of 367 specimens belonging to 54 species and 44 genera sampled across 12 sites. With the objective to produce a DNA barcode reference library for the Java and Bali ichthyofauna, a total of 24 specimens for 4 species of Rasbora and 15 specimens for 2 species of Nemacheilus were previously sequenced (Dahruddin *et al*. 2017).

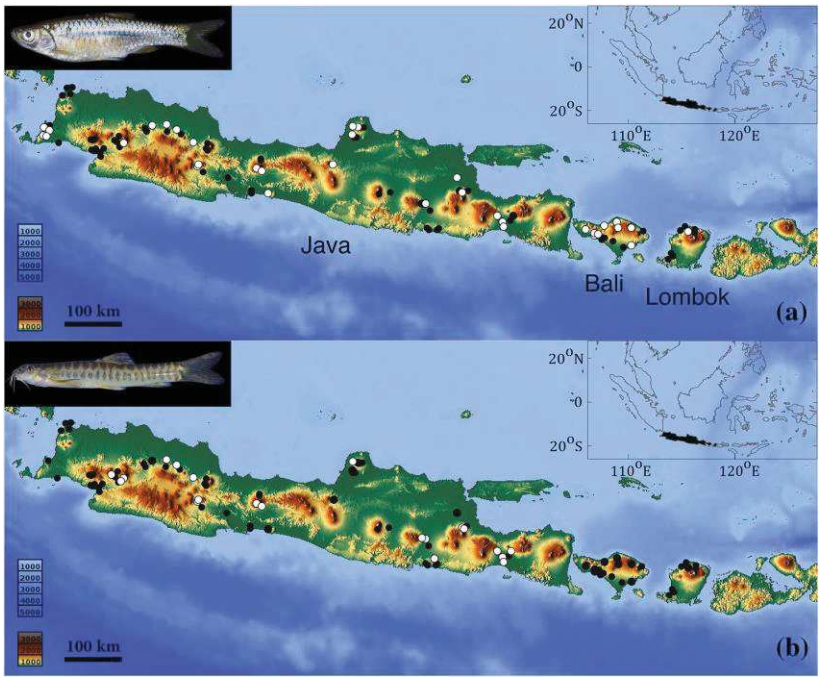


Figure 8. Collection sites for the 241 samples analyzed in the present study following the sampling campaign detailed in Dahruddin *et al*. 2017 and new sampling events in Lombok island. a Collection sites of *Rasbora* specimens. b Collection sites of *Nemacheilus* specimens. White dots correspond to sites where *Rasbora* or *Nemacheilus* specimens were collected. Black dots represent visited sites where no *Rasbora* or *Nemacheilus* specimens were observed. Each dot may represent several collection sites

Considering the objectives of the present study, an additional set of 84 specimens of *Nemacheilus* and 118 specimens of *Rasbora* were selected at all the sites these genera were sampled during the initial campaign

for further sequencing (Fig. 8). Thus, a total of 99 specimens belonging to 2 species of *Nemacheilus* and 142 specimens belonging to 4 species of *Rasbora* were analyzed in the present study (Table S1).

Specimens were captured using various gears including electrofishing, seine nets, cast nets and gill nets across sites encompassing the diversity of freshwater lentic and lotic habitats. Specimens were identified following available monographs (Kottelat *et al*. 1993), and species names were further validated based on several online catalogues (Froese and Pauly 2014; Eschmeyer *et al*. 2018). Specimens were photographed and individually labeled, and voucher specimens were preserved in a 5% formalin solution. A fin clip or a muscle biopsy was taken for each specimen and fixed in a 96% ethanol solution for genetic analyses. Both tissues and voucher specimens were deposited in the national collections at the Muzeum Zoologicum Bogoriense (MZB) in the Research Centre for Biology (RCB) from the Indonesian Institute of Sciences (LIPI).

4.3.2 Sequencing and international repositories

Genomic DNA was extracted using a Qiagen DNeasy 96 tissue extraction kit following the manufacturer's specifications. A 651-bp segment from the 5′ region of the cytochrome oxidase I gene (COI) was amplified using primer cocktails C_FishF1t1/C_FishR1t1 including M13 tails (Ivanova *et al*. 2007). PCR amplifications were done on a Veriti 96-well Fast (ABI-AppliedBiosystems) thermocycler with a final volume of 10.0 µl containing 5.0 µl Buffer 2×, 3.3 µl ultrapure water, 1.0 µl each primer (10 µM), 0.2 µl enzyme Phire® Hot Start II DNA polymerase (5 U) and 0.5 µl of DNA template (~ 50 ng). Amplifications were conducted as follow: initial denaturation at 98 °C for 5 min followed by 30 cycles denaturation at 98 °C for 5 s, annealing at 56 °C for 20 s and extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. The PCR products were purified with ExoSap-IT® (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. Sequencing reactions were performed using the "BigDye® Terminator v3.1 Cycle Sequencing Ready Reaction" and sequencing was performed on the automatic sequencer ABI 3130 DNA Analyzer (Applied Biosystems). The sequences and collateral information have been deposited in BOLD (Ratnasingham and Hebert 2007) and are available in the projects BIFH, BIFHB, BIFI and BIFB. DNA sequences were submitted to GenBank (accession numbers are accessible directly at the individual records in BOLD).

4.3.3 Species delimitation and genetic diversity

A maximum likelihood (ML) tree was first reconstructed using phyml 3.0.1 (Guindon and Gascuel 2003) based on the most likely substitution model selected by JMODELTEST 2.1.7 (Darriba *et al*. 2012). An ultrametric and fully resolved tree was reconstructed using the Bayesian approach implemented in BEAST 2.4.8 (Bouckaert *et al*. 2014). Two markov chain of 50 million each were ran independently using Yule pure birth model tree prior and an uncorrelated relaxed lognormal clock model for both Rasbora and Nemacheilus data sets. The ML tree was converted into an ultrametric tree using a relaxed clock model of the chromos function in the R package ape 4.1 (Paradis 2004) implemented in R (R Core Team 2018) and further used to initiate tree searches for the Bayesian analyses. Calibrations of ML and Bayesian analyses were established following Hutama *et al*. (2017). Age intervals for the Most Recent Common Ancestor (MRCA) of *Rasbora* spp. and *Nemacheilus* spp. were estimated based on the canonical 1.2% (+/− 0.5%) of genetic distance per million years for the fish COI gene (Bermingham *et al*. 1997). The average genetic distances between species pairs involving a direct ancestry with the MRCAs of *Rasbora* and *Nemacheilus* were calculated using MEGA 6 (Tamura *et al*. 2013) and used to estimate the age interval of the MRCAs. An additional calibration was added in the *Rasbora* tree including *Rasbora baliensis, R. lateristriata* and *R. aprotaenia* and also in *Nemacheilus* tree for the MRCA of *N. chrysolaimos* haplotypes following the same methodology. Trees were sampled every 10,000 states after an initial burning period of 10 million and both runs were combined using LogCombiner 2.4.8 (Bouckaert *et al*. 2014). The maximum credibility tree was constructed using TreeAnnotator 2.4.7 (Bouckaert *et al*. 2014). Several alternative methods have been proposed for delimitating molecular lineages (Pons *et al*. 2006; Puillandre *et al*. 2012; Ratnasingham and Hebert 2013; Zhang *et al*. 2013; Hubert and Hanner 2015). These methods rely on different approaches and assumptions but they all have in common the detection of transitions between mutation/drift (within species) and speciation/extinction (between species) dynamics (Hubert and Hanner 2015). Each of these methods is prone to pitfalls, particularly regarding singletons (i.e. delimitated lineages represented by a single sequence) and combining different approaches is increasingly used to circumvent potential pitfalls arising from, for instance, uneven sampling among species (Kekkonen and Hebert 2014; Kekkonen *et al*. 2015; Blair and Bryson 2017). Here, four sequence-based methods of species delimitation were used to delimitate species, and a final delimitation scheme was established based on a 50% consensus among methods in order to produce a robust delimitation scheme. For the sake of clarity, species identified based on morphological characters are referred to as species while species delimitated by DNA sequences are referred to as Operational

Taxonomic Units (OTU), defined as diagnosable molecular lineages (Avise 1989; Moritz 1994; Vogler and DeSalle 1994; Hutama *et al*. 2017). OTUs were delimitated using the following algorithms: (1) Refined single linkage (RESL) as implemented in BOLD and used to produce Barcode Index Numbers (BIN) (Ratnasingham and Hebert 2013), (2) Automatic barcode gap discovery (ABGD) (Puillandre *et al*. 2012), (3) Poisson tree process (PTP) in its multiple rates version (mPTP) as implemented in the standalone software mptp_0.2.3 (Zhang *et al*. 2013; Kapli *et al*. 2017), and (4) General mixed yule-coalescent (GMYC) in its single rate version (sGMYC) as implemented in the R package splits 1.0-19 (Ezard *et al*. 2009; Fujisawa and Barraclough 2013). RESL and ABGD used the DNA alignments as inputs while the ML tree was used for mPTP. Two delimitation schemes were collected for sGMYC: (1) a scheme based on the maximum credibility tree from the Bayesian analysis as input (sGMYC), (2) a consensus scheme with OTUs selected if present in more than 50% of the 10 replicates of sGMYC based on 10 Bayesian trees sampled along the Markov chain (sGMYC*).

We quantified the match among methods and their relative power using the match ratio, the Relative Taxonomic Index of Congruence index (Rtax) and the Taxonomic Index of Congruence (Ctax) following Blair and Bryson (2017). The match ratio is a measure of concordance among methods and is defined as twice the number of matches divided by the sum of the number of delimitated OTUs and the number of morphological species (Arhens *et al*. 2016). The Rtax index quantifies the relative power of a method to infer all estimated speciation events and is defined as the number of speciation events identified by a method divided by the total number of speciation events identified by the different methods (Miralles and Vences 2013). The Ctax index is a measure of congruence in species assignments between two methods and is calculated by dividing the number of speciation events inferred jointly by the two methods by the total number of speciation events inferred. Considering the number of comparisons involved, an average Ctax index was calculated for each method.

For each species, Kimura 2-parameter (K2P) pairwisem genetic distances were calculated using the R package ape 4.1 (Paradis 2004). Maximum intraspecific and nearest neighbor genetic distances were calculated from the matrice of pairwise K2P genetic distances using the R package SPIDER 1.5 (Brown *et al*. 2012). Haplotype diversity (h) and nucleotide diversity (π) were calculated for each species using the R package pegas 0.1 (Paradis 2010).

4.4 Results

99 and 142 sequences were successfully obtained for *Nemacheilus* and *Rasbora* respectively. All the sequences were above 500 bp of length and no stop codons were detected, suggesting that the sequences collected represent functional coding regions. The maximum credibility tree of *Rasbora* spp. identified a group of closely related species including *R. aprotaenia*, *R. lateristriata* and *R. baliensis* as well as two unknown taxa labeled as *R.* sp1 and *R.* sp2 (Fig. 9). The age of the MRCA of this clade of closely related species is inferred to trace back around 3 million years ago (Ma), and the split between *Rasbora argyrotaenia* and the remaining *Rasbora* is inferred to happen around 11 Ma. The age of *Rasbora* species MRCAs ranged between 0.5 Ma for R. baliensis and 1 Ma for R. sp1. The maximum credibility tree clearly separated the two *Nemacheilus* species genealogy into two distinct clades with a MRCA dated around 10 Ma (Fig. 10). The MRCA of *N. chrysolaimos* and N. *fasciatus* genealogies are dated around 1.5 and 0.5 Ma respectively. Delimitation methods largely converged in identifying 8 OTUs within the 6 species of *Rasbora* recognized here. Of the two partitioning schemes obtained with sGMYC, only the consensus partitioning scheme derived from 10 replicates (sGMYC*) is consistent with other methods in *Rasbora* (Fig. 9), with a number of OTUs ranging from 7 to 9 across sampled trees. Applying sGMYC to the Bayesian maximum credibility tree resulted in an inflated number of OTUs as 51 lineages were delineated (Table 5). Two OTUs were detected within *R. lateristriata* and *R*. sp1. The match ratio was similar among methods excepting sGMYC and the highest resolution power was observed for sGMYC with a Rtax of 1 (Table 5). The highest taxonomy congruence was observed for BIN, ABGD, mPTP and sGMYC* with a Ctax of 0.784 (Table 5). Delimitation methods produced concordant delimitation schemes within *Nemacheilus*, as all methods, excepting sGMYC, delineated one OTU for each of the two species (Fig. 10). As observed for *Rasbora*, sGMYC inflated the number of OTUs with 4 OTUs delimitated within *N. chrysolaimos* (Table 6). The match ratio and the taxonomic concordance (Ctax) were the highest for all methods excepting sGMYC (Table 6). The resolution power was estimated to be the highest for sGMYC as revealed by Rtax (Table 6).

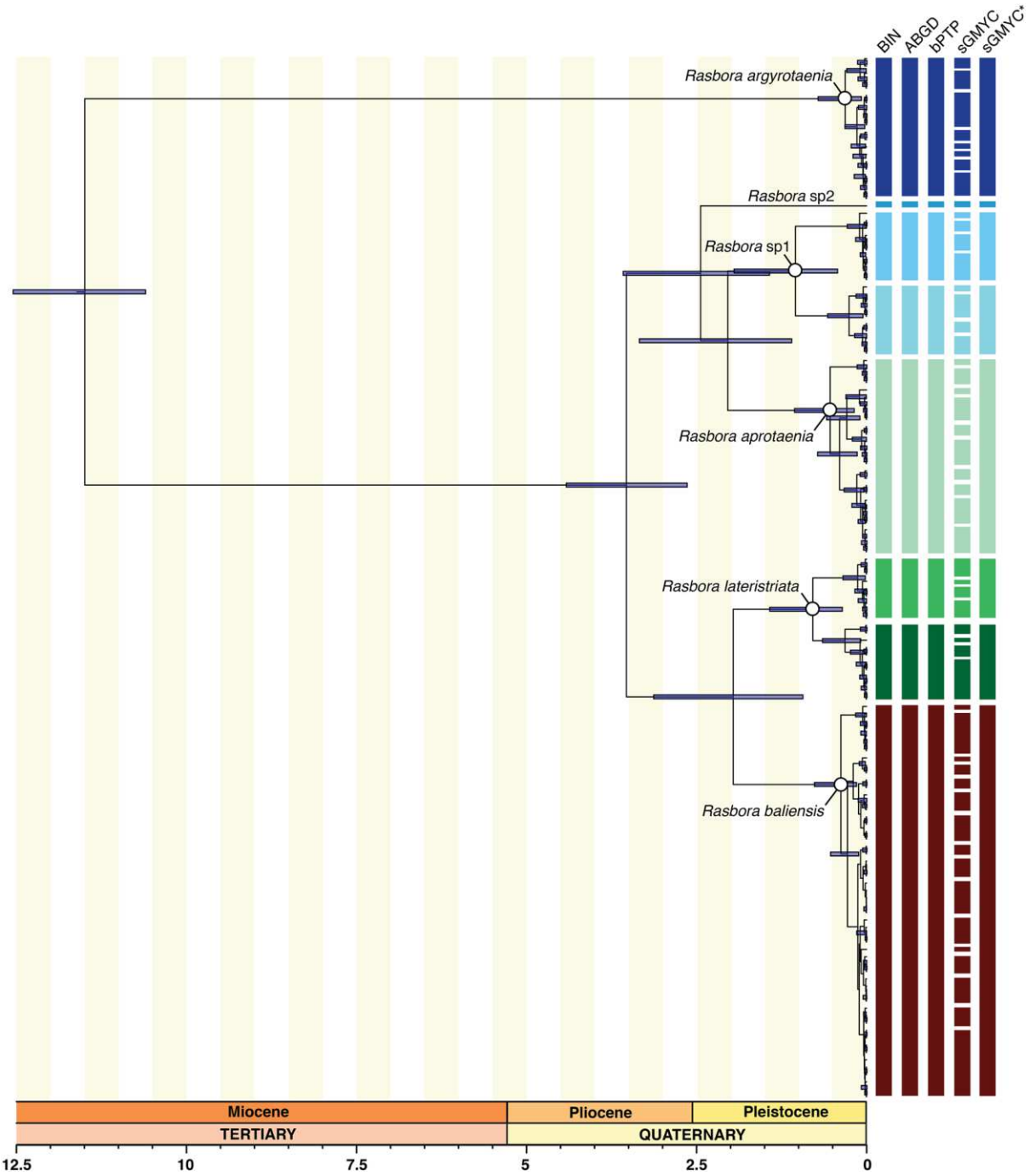


Figure 9. Bayesian maximum credibility tree of *Rasbora* DNA barcodes including 95% HPD interval for node age estimates and sequence clustering results according to the 5 species delimitation methods implemented

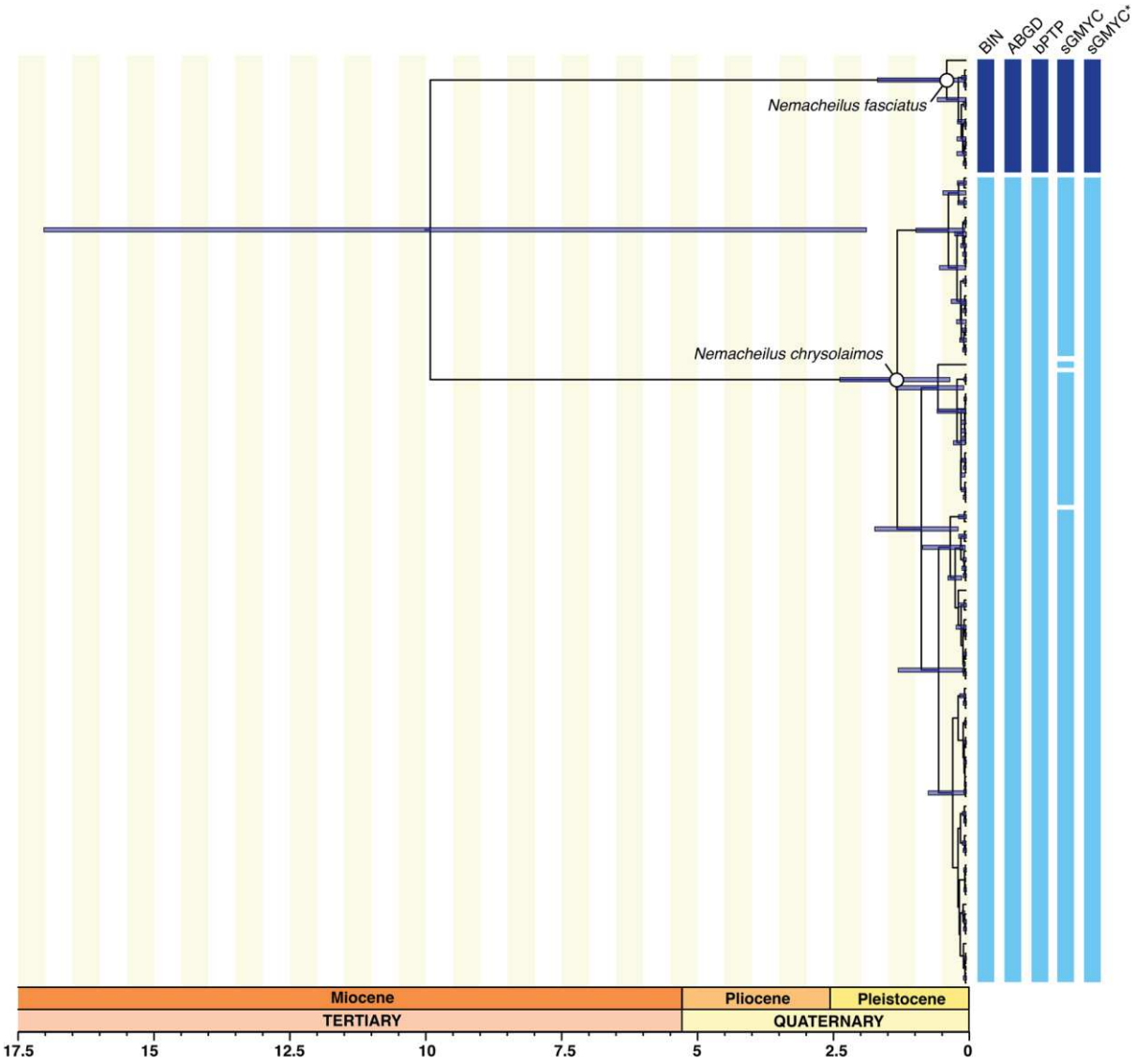


Figure 10. Bayesian Maximum Credibility Tree of *Nemacheilus* DNA barcodes including 95% HPD interval for node age estimates and sequence clustering according to the 5 species delimitation methods implemented

Table 6. Summary statistics of *Rasbora* species genetic diversity and species delimitation schemes

Taxon	n	Max. K2P distance	Min. nearest neighbor K2P distance	h	π	BIN	ABGD	bPTP	sGMYC	sGMYC*
R. argyrotaenia	20	0.62	9.67	0.742	0.001	1	1	1	8	1
R. aprotaenia	27	0.93	2.68	0.795	0.004	1	1	1	10	1
R. lateristriata	20	2.08	3.04	0.753	0.009	2	2	2	8	2
R. baliensis	54	0.77	3.04	0.505	0.002	1	1	1	16	1
***R.* sp1**	20	2.51	3.33	0.658	0.012	2	2	2	8	2
***R.* sp2**	1	NA	NA	NA	NA	1	1	1	1	1
Total	142	–	–	–	–	8	8	8	51	8
Match ratio	–	–	–	–	–	0.571	0.571	0.571	0.035	0.571
R_{tax}	–	–	–	–	–	0.16	0.16	0.16	1	0.16
Mean C_{tax}	–	–	–	–	–	0.784	0.784	0.784	0.04	0.784

Table 7. Summary statistics of *Nemacheilus* species genetic diversity and species delimitation schemes

Taxon	n	Max. K2P distance	Min. nearest neighbour K2P distance	h	π	BIN	ABGD	bPTP	sGMYC	sGMYC*
N. fasciatus	14	0.47	14.78	0.168	0.001	1	1	1	1	1
N. chrysolaimos	85	1.72	14.78	0.828	0.007	1	1	1	4	1
Total	99	–	–	–	–	2	2	2	5	2
Match ratio	–	–	–	–	–	1	1	1	0.286	1
R_{tax}	–	–	–	–	–	0.25	0.25	0.25	1	0.25
Mean C_{tax}	–	–	–	–	–	0.813	0.813	0.813	0.25	0.813

Maximum intraspecific K2P distances ranged between 0.62 in *R. argyrotaenia* and 2.51 in *R*. sp1 (Table 5), and the minimum nearest neighbor K2P distance ranged between 2.68 in *R. aprotaenia* and 9.67 in *R. argyrotaenia*. Haplotype diversity was globally high for all species with a haplotype diversity ranging from 0.795 in *R. aprotaenia* to 0.505 in *R. baliensis*. Nucleotide diversity was low for all species, with the lowest values ranging from 0.001 in *R. argyrotaenia* to 0.004 in *R. aprotaenia*, excepting for *R. lateristriata* and *R*. sp1, the only two species with two OTUs (Table 5). Maximum intraspecific K2P distances were 0.47 and 1.72 for *N. fasciatus* and *N. chrysolaimos*, respectively and the K2P distance between the two species was 14.78 (Table 6). Genetic diversity is markedly different between the two *Nemacheilus* species with *N. fasciatus* exhibiting a low genetic diversity with a haplotype diversity of 0.168 and a nucleotide diversity of 0.001 while *N. chrysolaimos* has a haplotype diversity of 0.828 and a nucleotide diversity of 0.007 (Table 6).

4.5 Discussion

4.5.1 Taxonomy of Rasbora species in Java, Bali and Lombok

Species boundaries of the four *Rasbora* species were successfully recovered by all, except sGMYC, species delimitation methods. This result has several implications regarding the taxonomy of the *Rasbora* genus in Java. The morphological characters described by Kottelat (1993) were not all operational as different and non-standardized meristic features were described for each species. Coloration patterns were used to propose an initial set of morphological identifications that resulted in the acknowledgement of four species based on the following key: (1) *R. aprotaenia* is distinguished by two dark spots along the lateral line being connected by a thin and diffuse dark line, the first below the origin of the dorsal fin, the second on the caudal peduncle and a dark spot along the proximal margin of the anal fin (Fig. 11). (2) *R. argyrotaenia* can be further separated in having a continuous dark line that cover

nearly entirely the lateral line and a dark line that underlines the entire proximal margin of the anal fin. (3) *R. laterstriata* and *R. baliensis* are further distinguished by the number of scale on the lateral line with 26–28 in *R. baliensis* and 29–33 in *R. lateristriata* (Kottelat *et al*. 1993). The concordance between the DNA-based and coloration- and meristic-based delimitation schemes for the four known Rasbora species confirms their biological species status. In addition, the examination of range distributions and known type localities further confirm this concordance with type localities being contained within the observed distribution range for *R. argyrotaenia*, *R. aprotenia* and *R. baliensis* (Fig. 12 a, b, d). *R. lateristriata* has been initially described based on a series of specimens collected throughout the Western part of Java (Kottelat 2013) at type localities that are not included in the range distribution observed here (Fig. 12c). Most of the type localities belong to the watershed draining the largest urban areas in Java (i.e. Jakarta, Bogor, Bandung), which are highly impacted by anthropogenic activities. Previous observations highlighted that some localities in this part of Java were dominated by invasive species including *Xiphophorus* spp., *Oreochromis niloticus*, *Clarias gariepinus* and *Poecilia* spp. (Dahruddin *et al*. 2017), and very few of the sites visited in the western part of Java resulted in the capture of *Rasbora* specimens (Jawa Barat, Table S1).

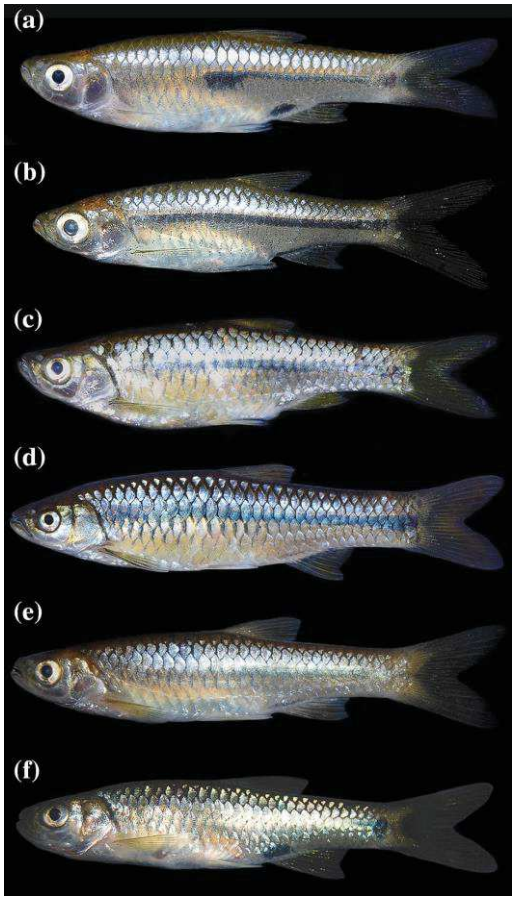


Figure 11. Selected specimen photographs of each of the 6 *Rasbora* species collected and recognized in the present study. a *Rasbora aprotaenia* (specimen BIF1501; SL = 47 mm; Ci Siih, Banten, Java). b *Rasbora argyrotaenia* (specimen BIF976; SL = 33 mm; Cilacap, Central Java). c *Rasbora baliensis* (specimen BIF2351; SL = 72 mm; Jembrana, West Bali). d *Rasbora lateristriata* (specimen BIF3619; SL = 87 mm; Kali Dauwan, Mojokerto, East Java). e *Rasbora* sp1 (specimen BIF864; SL = 61 mm; Kali Pelus, Purwokerto, Central Java). f *Rasbora* sp2 (specimen BIF155; SL = 43 mm; Ci Heulang, Sukabumi, West Java)

Also it is possible that we failed to capture *R. lateristriata* while the species was present, this result suggests that it has become rare, at least in this part of Java. The concordance between morphology-based and sequence-based species delimitation, however seems to confirm its validity. *Rasbora baliensis* was reported as a species endemic to the crater lakes in Bali while *R. lateristriata* replaced it in the rivers of Bali (Kottelat *et al*. 1993; Kottelat 2013). Our study shows that *R. baliensis* has a range

distribution surprisingly wider than expected as its presence is detected until East Java and Lombok (Fig. 12d). The confusion that reigns over the know range distribution of *R. baliensis* and *R. lateristriata* is surprising considering that they show non-overlapping numbers of scale at the lateral line, a character that has been previously overlooked. The otherwise morphological similarity between the two species is likely to account for the reported occurrence of *R. lateristriata* in most Lesser Sunda Islands including Bali, Lombok and Sumbawa (Kottelat *et al*. 1993). The disjunctive range distribution of both species, as well as the reciprocal monophyly that was captured by all DNA-based delimitation methods, argue that *R. baliensis* is a valid taxon. Furthermore, the presence of *R. baliensis* in east Java and Lombok is reported here for the first time while we show that *R. lateristriata* is restricted to central Java (Fig. 12d).

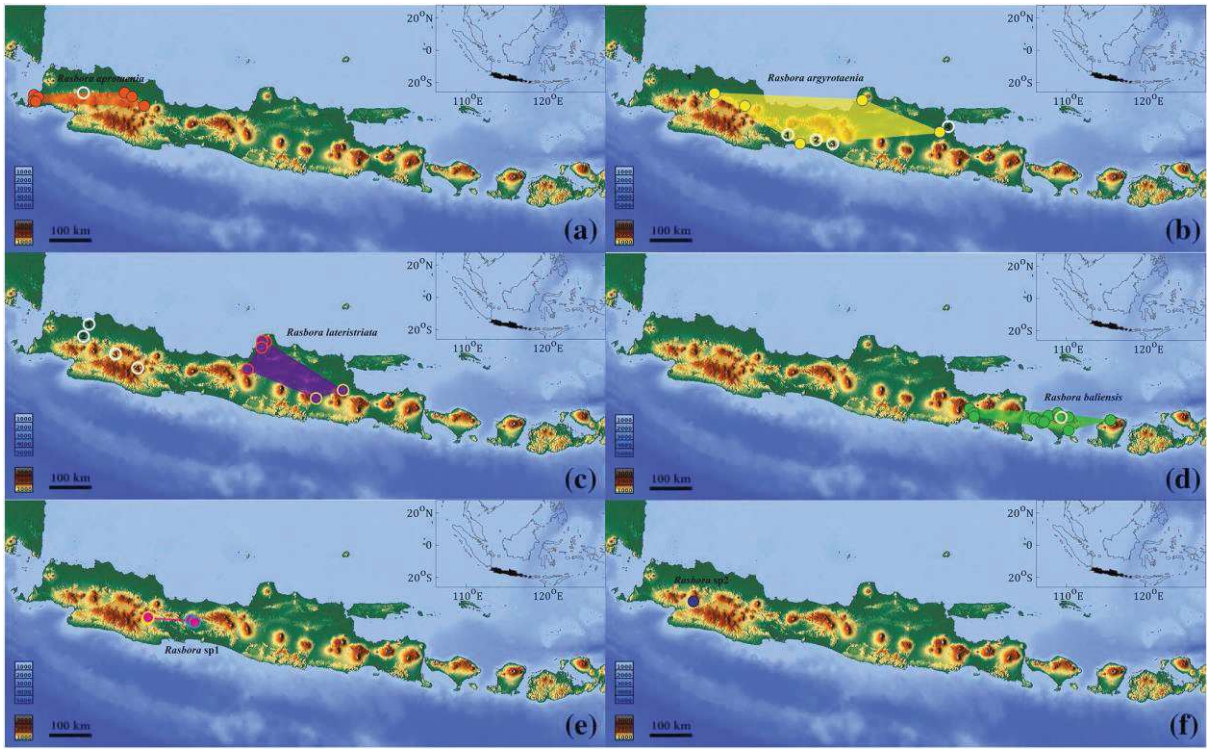


Figure 12. Species range distribution of the 6 *Rasbora* species recognized in the present study. Colored dots represent collection sites. White circles represent type localities. a *Rasbora aprotaenia*. b *Rasbora argyrotaenia*. c *Rasbora lateristriata*. d *Rasbora baliensis*. e *Rasbora* sp1. f *Rasbora* sp2

Finally, two new taxa are discovered here, *Rasbora* sp1 and *Rasbora* sp2 (Figs. 2, 4). The COI tree suggests that both taxa are closely related to *R. aprotaenia*. The coloration pattern of R. sp1 is markedly different from *R. aprotaenia*, however as R. sp1 exhibit a thin dark line, instead of a dark spot in *R. aprotaenia*, at the proximal margin of the anal fin and a thin and a diffuse dark line along the lateral line (Fig. 11e). *Rasbora* sp2 is much more

similar to *R. aprotaenia* in terms of coloration pattern, and the fact that a single specimen was captured warrants further studies to enable its formal description. The diversification of this clade of closely related species is inferred to happen during the last 2.5 Ma in a restricted area of the western part of Java, a pattern previously described for several species largely distributed in Java such as *Barbodes binotatus, Channa gachua* and *Glyptothorax platypogon* (Hutama *et al*. 2017).

4.5.2 Taxonomy of Nemacheilus species in Java

The DNA-based delimitation schemes are consistent with the presence of two species in Java, namely *Nemacheilus fasciatus*, reported to occur in Java and Sumatra, and *Nemacheilus chrysolaimos*, an endemic species of Java (Kottelat *et al*. 1993; Kottelat 2013). The examination of the type locality of *N. fasciatus* provided a conflicting result with the initial identification performed based on the diagnostic morphological characters described in Kottelat (1993).

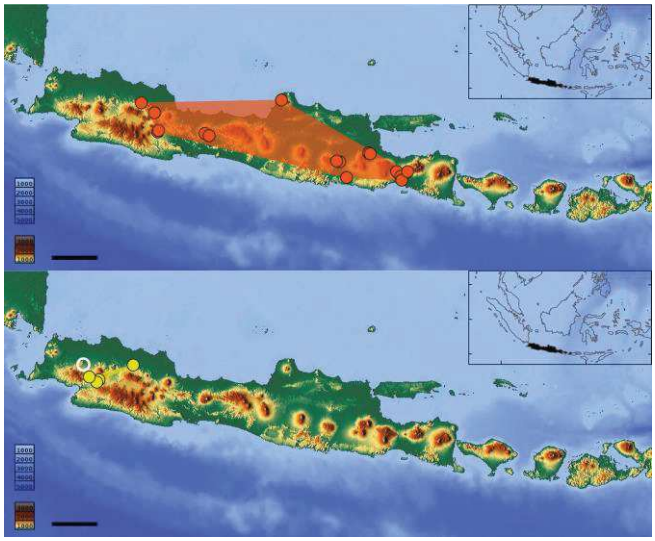


Figure 13. Species range distribution of the 2 *Nemacheilus* species recognized in the present study. Colored dots represent collection sites. White circles represent type localities. a *Nemacheilus chrysolaimos* b *Nemacheilus fasciatus*

The type locality of *N. chrysolaimos* is unknown and the type locality of *N. fasciatus* is located in the western part of Java (Fig. 13). Following Kottelat's monograph (1993), the Western species is *Nemacheilus chrysolaimos*. According to our study, however the *Nemacheilus* species of western Java is *N. fasciatus*, *N. chrysolaimos* being largely distributed in Central and Eastern Java (Fig. 13). A reexamination of the diagnostic morphological characters proposed to differentiate both *Nemacheilus* species in Java at the light of the present results indicates that *N. fasciatus* is the species that is distinguished in having a winged flap on the margin of the anterior naris (Fig. 14c, d), contrasting with an anterior naris valve pierced at the tip of a tube in *N. chrysolaimos* (Fig. 14a, b). The coloration patterns of each species are also distinct with *N. chrysolaimos* presenting 14–18 dark blotches along lateral line with 11–12 dark saddles across the back contrasting with *N. fasciatus* exhibiting 9–18 dark bars of irregular shape (Kottelat *et al*. 1993). The identity of both species was further confirmed by the examination of the type

specimens of both species (*N. chrysolaimos*, B2972, B3961; *N. fasciatus*, B2798) deposited at the MNHN Paris, as well as the original illustration of *N. chrysolaimos* depicting a coloration pattern consisting of irregular saddles, consistent with the coloration pattern of the specimens assigned to *N. chrysolaimos* in the present study.

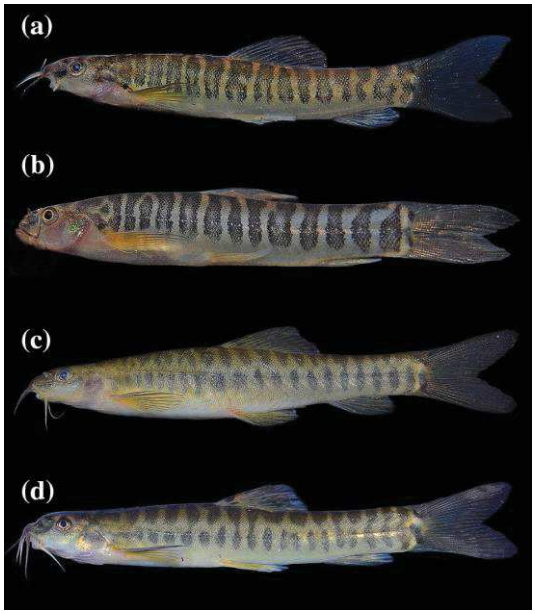


Figure 14. Selected specimen photographs of each of the 2 *Nemacheilus* species collected and recognized in the present study. a *Nemacheilus fasciatus* (specimen BIF495; SL = 42 mm; Ci Asem, Purwakarta, West Java). b *Nemacheilus fasciatus* (specimen BIF163; SL = 45 mm; Ci Heulang, Central Java). c *Nemacheilus chrysolaimos* (specimen BIF2032; SL = 58 mm; Ngerjo, Blitar, East Java). d *Nemacheilus chrysolaimos* (specimen BIF2074; SL = 55 mm; Bicoro, Lumajang, East Java)

4.5.3 Species distribution and conservation genetics

All the species under study here were previously reported as endemic species of the islands of Java, Bali and Lombok. The present study highlights that most of the species examined here have range distribution much more restricted than previously acknowledged. Excepting *N. chrysolaimos* and *R. argyrotaenia* detected in east and central Java, most species are confined to a restricted set of watersheds. A clear separation is observed between the eastern versus the central and east Java lineages, a pattern previously reported among cryptic lineages within widespread species in Java such as *Barbodes binotatus*, *Glyptothorax platypogon* and *Channa gachua* (Dahruddin *et al*. 2017; Hutama *et al*. 2017). This biogeographic transition between East and West watersheds in Java is related to the ontogeny of Java, resulting from the merging of two volcanic arches, one is the west and a second in the east (including East Java and Bali), that emerged between 10 and 5 Ma and further aggregated during the last 5 million years (Lohman *et al*. 2011). This scenario is reflected in Rasbora with the range distribution of *R. baliensis* encompassing the eastern part of Java, Bali and Lombok. A similar trans-distribution across the Bali strait was previously observed in *Barbodes binotatus* and *Channa gachua* (Hutama *et al*. 2017). Divergence levels between the western most species and the central/east species are surprisingly higher than previously observed in Java with a divergence between *N. fasciatus* and *N. chrysolaimus* inferred at around 10 Ma (1.01–17.90 Ma, 95% HPD) and the divergence of *R. argyrotaenia* from other Rasbora inferred at

around 11.5 Ma (10.61–12.55 Ma, 95% HPD). These age estimates are much older than reported for other widespread species in Java with species MRCA estimated to occur around 3 Ma for several species (Hutama *et al*. 2017). The age of the MRCA of sampled species of *Nemacheilus* and *Rasbora* also trace back beyond the ontogeny of Java, suggesting that the evolutionary history of these Javanese species started prior to the rise of this island, during the initial settlement of Sundaland (Lohman *et al*. 2011). These old splits particularly contrast in *Rasbora* with divergence age estimates for other *Rasbora* species that do not exceed 3.5 Ma, an age consistent with previous age estimates of the MRCA of the coalescent trees of several widespread species in Java (Hutama *et al*. 2017). Interestingly, *N. fasciatus* and *R. argyrotaenia* are both cited from Sumatra suggesting a recent colonization of Java from Sumatra for those species. This scenario was recently suggested for *Rasbora* based on a phylogenetic assessment of Java species based on mitochondrial genomes (Kusuma *et al*. 2016), and highlights the different origins of the western and eastern species of *Nemacheilus* and *Rasbora*. The *Rasbora* tree topology, supporting close relationships between central and east Java species, and age estimates, below 3.5 Ma, suggest that most *Rasbora* species diversified through insitu speciation in Java, a hypothesis previously suggested (Kusuma *et al*. 2016). The general trend of highly restricted range distribution, related to a hypothesized origin though in-situ diversification in Java and Bali for Rasbora, highlights the sensitive conservation status of most endemic species in Java. The distribution patterns observed here are highly fragmented with little overlap between range distributions of sister-species, a pattern that was expected considering the volcanic origin of Java, Bali and Lombok islands during the Quaternary (Lohman *et al*. 2011). Genetic diversity patterns are consistent with these observations as nucleotidic diversity is generally low for *Rasbora* and *Nemacheilus* species, a molecular signature also consistent with a recent origin through in-situ diversification. Cryptic diversity is also reported here for *R. lateristriata* and *R*. sp1, with cryptic lineages diverging around 1 Ma for both species, further accentuating the fragmented status of the populations of these species. As such, most *Rasbora* species should be considered as complexes of small populations, highly sensitive to further reductions in population size that may eventually lead to a vortex of extinction (Gilpin and Soulé 1986; Fagan and Holmes 2006). Java is the most densely populated island in Sundaland and anthropogenic activities have an alarming impact on the rivers of this region (Spracklen *et al*. 2015; Breckwoldt *et al*. 2016; Hayati *et al*. 2017; Garg *et al*. 2018). Severe reductions of aquatic habitat size during the last two decades have been reported with significant impacts of water pollution generated by industrial wastes and agricultural run-offs, but also the

introduction of exotic species (Eidman 1989; Dahruddin *et al*. 2017). Considering the very narrow range distributions observed here for most endemic species and the concerning environmental context of Java, this study stresses that conservation efforts even at a small spatial scale could have significant impacts in the conservation of these endemic species. Translocations of livestock are common in Indonesia and are often used as a management measure (Dahruddin *et al*. 2017). The presence of cryptic diversity, newly discovered, yet undescribed, taxa and their close morphological affinities make accurate species identifications difficult. In this context, the introduction of multiple species through translocation programs due to species misidentifications are likely if not guided by DNA barcoding. Considering the young age of most of the species examined here and their allopatric distributions, the effectiveness of reproductive isolation mechanisms after secondary contact is questionable and the occurrence of hybridization cannot by discarded. Both genera have become scarce in the Western part of Java. Combined with the low genetic diversity at the nucleotide level, restoration program through genetic rescue (i.e. increasing population fitness through the repletion of genetic diversity by immigration) are probably required (Whiteley *et al*. 2014). If such translocation programs were to be implemented in the future, we advocate for short spatial scale translocations in order to avoid secondary contact among closely related species and identify immigrants, to the species level using DNA barcodes.

4.6 Conclusion

As in previous molecular studies in the area, the present study highlights gaps of knowledge in the taxonomy and distribution of the freshwater fishes of Java and Bali, and further highlights the complexity of diversity patterns in this part of Sundaland. This assessment of *Rasbora* and *Nemacheilus* diversity through DNA barcodes sheds light on the species biological status and distribution in South Sundaland and warrants further studies. In particular, the two new taxa discovered here need additional sampling efforts to be accurately described. The present study highlights the sensitive status of most species owing to their restricted range and low genetic diversity.

4.7 Acknowledgements

The authors wish to thank Siti Nuramaliati Prijono, Bambang Sunarko, Witjaksono, Mohammad Irham, Marlina Adriyani, Ruliyana Susanti, Rosichon Ubaidillah, Hari Sutrisno

and Muhamad Syamsul Arifin Zein at Research Centre for Biology (RCBLIPI); Jean-Paul Toutain, Robert Arfi, Valérie Verdier and Jean-François Agnèse from the 'Institut de Recherche pour le Développement'; Joel Le Bail and Nicolas Gascoin at the French embassy in Jakarta for their continuous support. We are thankful Sumanta at IRD Jakarta for his help during the field sampling. Part of the present study was funded by the Institut de Recherche pour le Développement (UMR226 ISE-M and IRD through incentive funds), the MNHN (UMR BOREA), the RCB-LIPI, the French Ichthyological Society (SFI), the Foundation de France and the French embassy in Jakarta. The Indonesian Ministry of Research and Technology approved this study and field sampling was conducted according to the research permits 097/SIP/FRP/SM/IV/2014 for Philippe Keith, 60/EXT/SIP/FRP/SM/XI/2014 for Frédéric Busson and 41/EXT/SIP/FRP/SM/VIII/2014 for Nicolas Hubert. Sequence analysis was aided by funding from the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project. We thank Paul Hebert, Robert Hanner and Evgeny Zakharov as well as BOLD and CCDB staffs at the University of Guelph for their valuable support. Finally, we thank Anti Vasemägi and the three anonymous reviewers for providing constructive comments of earlier versions of the manuscript. This publication has ISEM Number 2018-279-SUD.

References

April J, Mayden R, Hanner L, Bernatchez RH L (2011) Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci USA 108:10602–10607

Arhens D, Fujisawa T, Krammer HJ, Eberle J, Fabrizi S, Vogler AP (2016) Rarity and incomplete sampling in DNA-based species delimitation. Syst Biol 65:478–494

Avise JC (1989) Molecular markers, natural history and evolution. Chapman and Hall, New York

Bermingham E, McCafferty S, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian isthmus. In: Kocher TD, Stepien CA (eds) Molecular systematics of fishes. CA Academic Press, San Diego, pp 113–128

Blair C, Bryson JRW (2017) Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: Phrynosoma). Mol Ecol Resour 17:1168–1182

Bouckaert RR, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537

Breckwoldt A, Dsikowitzky L, Baum G, Ferse SCA, van der Wulp S, Kusumanti I, Ramadhan A, Adrianto L (2016) A review of stressors, uses and management perspectives for the larger Jakarta Bay Area, Indonesia. Mar Pollut Bull 110:790–794

Brown SDJ, Collins RA, Boyer S, Lefort C, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) SPIDER: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12:562–565

Dahruddin H, Hutama A, Busson F, Sauri S, Hanner R, Keith P, Hadiaty RK, Hubert N (2017) Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. Mol Ecol Resour 17:288–299

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

Durand JD, Hubert N, Shen KN, Borsa P (2017) DNA barcoding grey mullets. Rev Fish Biol Fisheries 27:233–243

Eidman HM (1989) Exotic aquatic species introduction into Indonesia, vol 3. Exotic aquatic organisms in Asia Asian Fisheries Society Special Publication, Bethesda, pp 57–62

Eschmeyer WN, Fricke R, van der Laan R (2018) Catalog of fishes electronic version. Accessed March 2018

Ezard T, Fujisawa T, Barraclough TG (2009) Splits: SPecies' LImits by Threshold Statistics. R package version 1.0-18/r45. Available from http://R-Forge.R-project.org/projects/splits/

Fagan WF, Holmes EE (2006) Quantifying the extinction vortex. Ecol Lett 9:51–60

Froese R, Pauly D (2014) FishBase. World Wide Web electronic publication. http://www.fishbase.org

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol, 62:707–24

Garg T, Hamilton SE, Hochard JP, Kresch EP, Talbot J (2018) (Not so) gently down the stream: River pollution and health in Indonesia. J Environ Econ Manag 92:35–53

Garnett ST, Christidis L (2017) Taxonomy anarchy hampers conservation. Nature 546:25–27

Geiger MF, Herder F, Monaghan MT, Almada V, Barbieri R, Bariche M, Berrebi, Bohlen P, Casal-Lopez M, Delmastro GB (2014) Spatial heterogeneity in the mediterranean biodiversity hotspot affects barcoding accuracy of its freshwater fishes. Mol Ecol Resour 14:1210–1221

Gilpin E, Soulé M (1986) Minimum viable populations: processes of species extinction. In: Soulé ME (ed) Conservation biology: the science of scarcity and diversity. Sinauer, Sunderland. pp 19–34

Guindon S, Gascuel O (2003) A simple, fast and Accurate algorithm to estimate large phylogenies by Maximum Likelihood. Syst Biol 52:696–704

Hayati A, Tiantono N, Mirza MF, Putra IDS, Abdizen MM, Seta AR, Solikha BM, Fu'adil MH, Putranto TWC, Affandi M, Rosmaninda (2017) Water quality and fish diversity in the Brantas river, East Java, Indonesia. J Biol Res 22:43–49

Hoffman M, Hilton-Taylor C, Angulo A, Böhm M, Brooks TM, Butchart SHM, Carpenter KE, Chanson J, Collen B, Cox NA *et al* (2010) The impact of conservation on the status of the world's vertebrates. Science 330:1503–1509

Hubert N, Hanner R (2015) DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3:44–58

Hubert N, Hanner RH, Holm E, Mandrak NE, Taylor EB, Burridge M, Watkinson DA, Dumont P, Curry A, Bentzen P, Zhang J, April J, Bernatchez L (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE, 3:e2490

Hubert N, Meyer C, Bruggemann JH, Guérin F, Komeno RJL, Espiau B, Causse R, Williams JT, Planes S (2012) Cryptic diversity in Indo-Pacific coral reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. PLoS ONE, 7:e28987

Hubert N, Wibowo A, Busson F, Caruso D, Sulandari S, Nafiqoh N, Rüber L, Pouyaud L, Avarre JC, Herder F, Hanner R, Keith P, Hadiaty RK (2015) DNA barcoding Indonesian freshwater fishes: challenges and prospects. DNA Barcodes 3:144–169

Hubert N, Dettai A, Pruvost P, Cruaud C, Kulbicki M, Myers RF, Borsa P (2018) Geography and life history traits account for the accumulation of cryptic diversity among Indo-West Pacific coral reef fishes. Mar Ecol Prog Ser 583:179–193

Hutama A, Dahruddin H, Busson F, Sauri S, Keith P, Hadiaty RK, Hanner R, Suryobroto B, Hubert N (2017) Identifying spatially concordant evolutionary significant units across multiple species through DNA barcodes: application to the conservation genetics of the freshwater fishes of Java and Bali. Global Ecol Conserv 12:170–187

Ivanova NV, Zemlak TS, Hanner RH, Hébert PDN (2007) Universal primers cocktails for fish DNA barcoding. Mol Ecol Notes 7:544–548

Jaafar TNAM, Taylor MI, Mohd Nor SA, De Bruyn M, Carvalho GR (2012) DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay carangidae (Teleosteii: Perciformes). PLoS ONE, 7:e49623

Kadarusman HN, Hadiaty RK, Paradis E, Pouyaud L (2012) Cryptic diversity in Indo-Australian rainbowfishes revealed by DNA Barcoding: implications for conservation in a biodiversity hotspot candidate. Plos ONE 7:e40627

Kapli P, Zhang SL, Kobert J, Pavlidis K, Stamatakis P, Flouri A T (2017) Multi-rate Poisson Tree Processes for single-locus species delimitation under Maximum Likelihood and Markov Chain Monte Carlo. Bioinformatics 33:1630–1638

Kekkonen M, Hebert PDN (2014) DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. Mol Ecol Resour 14:706–715

Kekkonen M, Mutanen M, Kaila L, Nieminen M, Hebert PDN (2015) Delineating species with DNA barcodes: a case of taxon dependent method performance in moths. PLoS ONE 10:e0122481

Knebelsberger T, Dunz AR, Neumann D, Geiger MF (2015) Molecular diversity of Germany's freshwater fishes and lampreys assessed by DNA barcoding. Mol Ecol Resour 15:562–572

Kottelat M (2013) The fishes of the inland waters of Southeast Asia: a catalog and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. Raffles Bull Zool Suppl 27:1–663

Kottelat M, Whitten AJ, Kartikasari SR, Wirjoatmodjo S (1993) Freshwater fishes of western indonesia and sulawesi. Periplus editions, Singapore

Kusuma WE, Ratmuangkhwang S, Kumazawa Y (2016) Molecular phylogeny and historical biogeography of the Indonesian freshwater fish Rasbora lateristriata species complex (Actinopterygii: Cyprinidae): cryptic species and west-to-east divergences. Mol Phylogenet Evol 105:212–223

Lamoreux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight M, Shugart HH (2006) Global tests of biodiversity concordance and the importance of endemism. Nature 440:212–214

Lohman K, De Bruyn M, Page T, Von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR, Von Rintelen T (2011) Biogeography of the Indo-Australian archipelago. Annu Rev Ecol Evol Syst 42:205–226

Machado VN, Collins RA, Ota RP, Andrade MC, Farias IP, Hrbek T (2018) One thousand DNA barcodes of piranhas and pacus reveal geographic structure and unrecognised diversity in the Amazon. Sci Rep 8:8387

Miralles A, Vences M (2013) New metrics for comparison of taxonomies eveal striking discrepancies among species delimitation methods in Madascincus lizards. PLoS ONE 8:e68242

Moritz C (1994) Defining 'Evolutionary Significant Units' for conservation. Trends Ecol Evol 9:373–375

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858

Paradis E (2010) pegas: an R package for poluation genetics with an integrated modular approach. Bioinformatics 26:419–420

Paradis E, Claude J, Strimmer K (2004) ape: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

Pereira LHG, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna ? BMC Genet 14:20

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol 55:595–606

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol 21:1864–1877

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system. Mol Ecol Notes 7:355–364

Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE 8:e66213

Spracklen DV, Reddington CL, Gaveau DLA (2015) Industrial concessions, fires and air pollution in Equatorial Asia. Environ Res Lett 10:091001

Steinke D, Hanner R (2011) The FISH-BOL collaborators' protocol. Mitochondrial DNA 22:10–14

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

Vogler AP, DeSalle R (1994) Diagnosing units of conservation management. Conserv Biol 6:170–178

Ward RD, Hanner RH, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74:329–356

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2014) Genetic rescue to the rescue. Trends Ecol Evol 30:42–49

Winemiller KO, McIntyre PB, Castello L, Fluet-Chouinard E, Giarrizzo T, Nam S, Baird IG, Darwall W, Lujan NK, Harrison I, Stiassny MLJ, Silvano RAM, Fitzgerald DB, Pelicice FM, Agostinho AA, Gomes LC, Albert JS, Baran E, Petrere M, Zarfl C, Mulligan M, Sullivan JP, Arantes CC, Sousa LM, Koning AA, Hoeinghaus DJ, Sabaj M, Lundberg JG, Armbruster J, Thieme ML, Petry P, Zuanon J, Vilara GT, Snoeks J, Ou C, Rainboth W, Pavanelli CS, Akama A, Soesbergen AV, Sáenz L (2016) Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. Science 351:128–129

Winterbottom R, Hanner R, Burridge M, Zur M (2014) A cornucopia of cryptic species—a DNA barcode analysis of the gobiid genus Trimma (Percomorpha, Gobiiformes). Zookeys 381:79–111

Woodruff DS (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugium-phase biodiversity. Biodivers Conserv 19:919–941

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876

V. Conservation genetics of the freshwater fishes of Java

Article 3: Identifying spatially concordant evolutionary significant units across multiple species through DNA barcodes: Application to the conservation genetics of the freshwater fishes of Java and Bali

Aditya Hutama[a], Hadi Dahruddin[b], Frédéric Busson[c], Sopian Sauri[b], Philippe Keith[c], Renny Kurnia Hadiaty[b], Robert Hanner[d], Bambang Suryobroto[a], NicolasHubert[e]

[a] Bogor Agricultural University, Faculty of Mathematics and Natural Science, Animal Bioscience, Jl. Raya Darmaga, Bogor 16680, Indonesia
[b] LIPI, Research Center for Biology, Zoology Division, MZB, Gedung Widyasatwaloka, Jl. Raya Jakarta Bogor Km. 46, Cibinong- Bogor 16911, Indonesia
[c] Muséum national d'Histoire naturelle, UMR 7208 (MNHN-CNRS-UPMC-IRD-UCBN), CP 026, 43, rue Cuvier, F-75231 Paris Cedex 05, France
[d] Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada
[e] Institut de Recherche pour le Développement, UMR 226 ISEM (UM2-CNRS-IRD), Place Eugène Bataillon, CC 065, F-34095 Montpellier cedex 05, France

5.1 Abstract

Delineating Evolutionary Significant Units for conservation purposes is a crucial step in conservation. Across a distribution range, species frequently display population structure that drives the distribution of genetic diversity. These patterns of genetic structure and diversity result from intricate interactions between biogeographic history and demographic dynamics. Prior biogeographic knowledge, however, is scarcely available, a trend particularly pronounced in the tropics where the taxonomic impediment is hampering biogeographic studies and conservation efforts. DNA barcoding has been initially proposed to foster taxonomic studies through the development of an automated molecular system of species identification. While its utility for species identification is increasingly acknowledged, its usefulness for fast and large-scale delineation of ESU remains to be explored. If proved to be useful for that purpose, DNA barcoding may also open new perspectives in conservation by quickly providing preliminary information about population conservation status. The present study aims at assessing the utility of DNA barcoding for the delineation of ESUs among the

most common freshwater fish species of Java and Bali through the comparison of population genetic structures and diversification patterns across multiple species. Substantial levels of cryptic diversity are discovered among the three widely distributed freshwater fish species analyzed with a total of 21 evolutionary independent mitochondrial lineages (BINs) observed in *Barbodes binotatus*, *Channa gachua* and *Glyptothorax platypogon*. The maximum genetic distance for each coalescent tree ranges from 6.78 to 7.76 K2P genetic distances for *C. gachua* and *G. platypogon*, respectively. Diversification and population genetic analyses support a scenario of allopatric differentiation. The analysis of the BINs spatial distribution indicates concordant distribution patterns among the three species that allow identifying 18 ESUs. Implications for the conservation genetics of these species are discussed at the light of the history of the region.

5.2 Introduction

Conservation aims at preserving species evolutionary potential to sustain their adaptive abilities in fluctuating environments and fuel evolution on a long-term perspective. The nature of the biological units to be targeted for achieving this goal, however, has been a contentious issue since the 1990s (Crandall *et al*., 2000). With the objective to get around taxonomic confusion, Ryder (1986) proposed the concept of Evolutionary Significant Units (ESU) that he defined as 'subset of the more inclusive entity species, which possess genetic attributes significant for the present and future generations of the species'. He proposed to delineate ESUs based on concordant evidences from ecological, physiological and genetic perspectives (Ryder, 1986). Following this initial formulation, several recognition criteria were proposed with varying emphasis on reproductive isolation (Waples, 1991, Waples, 1995), historical trends in population structure (Avise, 1989, Moritz, 1994), shared character states (Vogler and DeSalle, 1994) and genetic or ecological exchangeability (Crandall *et al*., 2000). From a genetic perspective, most of the criteria focus on detecting the imprint of disrupted gene flow in mitochondrial and nuclear genomes, either from a short-term perspective when ESUs are delineated based on differences in allele frequencies (Waples, 1991, Waples, 1995, Dizon *et al*., 1992, Waples, 1995) or long-term perspective when ESUs are based on reciprocal monophyly of their genes genealogies (Avise, 1989, Moritz, 1994).

Fraser and Bernatchez (2001) highlighted that the procedures for delineating ESUs are based on criteria, not mandatory properties, which applicability depends on biogeographical and ecological context. Since then, major improvements happened in terms of genome

analysis and statistical tools to link population genetic data with landscape ecology (Holderegger and Wagner, 2006, Anderson *et al*., 2010, Sork and Waits, 2010) or landscape history (Beheregaray, 2008). These methodological developments opened new perspectives in conservation, in particular, landscape genetics and niche modeling shed a new light at the way we look at dispersal and gene flow (Endo *et al*., 2014, Gutiérrez-Tapia and Palma, 2016). Usually implemented at small spatial scales, this process-based approach requires a prior knowledge of population structure at the regional scale and identifying ESUs is still a preliminary step in conservation (Fraser and Bernatchez, 2001, Pearse and Crandall, 2004).

Delineating ESU is context dependent as Earth biotas originated from diverse biogeographical histories resulting in species of varying age, distribution range and population structure (Wiens and Donoghue, 2004, Mittelbach *et al*., 2007, Weir and Schluter, 2007, Hua and Wiens, 2010). Thus, prior biogeographic knowledge should provide a useful framework to guide the delineation of ESUs (Avise, 2000, Fraser and Bernatchez, 2001). This knowledge, however, is not sufficiently detailed for many regions of the world, a situation frequently observed in the tropics where high species richness has largely amplified the problem (Beheregaray, 2008, Hubert and Hanner, 2015). DNA barcoding, the use of the cytochrome oxidase I gene as an internal species tag for molecular identifications, has opened new perspectives on collecting genomic resources across multiple species (Hebert *et al*., 2004, Hebert and Gregory, 2005, Janzen *et al*., 2005, Smith *et al*., 2005, Steinke *et al*., 2009, Hubert *et al*., 2012). A quick scan of population structure across multiple species through DNA barcode may provide a preliminary and insightful approach in conservation, prior to a more comprehensive assessment using both mitochondrial and nuclear markers, by (i) delineating mitochondrial lineages that depart from population dynamics and display independent mutation/drift dynamics (Hajibabaei *et al*., 2007, Vernooy *et al*., 2010, Ratnasingham and Hebert, 2013, Kekkonen and Hebert, 2014), (ii) providing genomic resources for quick species and ESUs molecular identification (Hajibabaei *et al*., 2011, Gibson *et al*., 2014), (iii) easing the characterization of ESUs range distribution and their spatial match across multiple species to identify conservation priorities.

Such strategy may be particularly relevant in areas facing massive anthropogenic threats and where conservation strategies are impeded by the lack of appropriate knowledge on species evolutionary dynamics and taxonomic confusion, a trend further amplified by the recent rarefaction of taxonomists worldwide (i.e. taxonomic impediment). This situation is currently observed in South-East Asia where the four biodiversity hotspots identified are among the most threatened to date (Myers *et al*., 2000, Lamoreux *et al*., 2006, Hoffman *et al*.,

2010). This is particularly evident for the insular hotspots of the Indonesian archipelago (i.e. Sundaland and Wallacea), where the impact of anthropogenic activities is amplified by their refugial state, particularly so in the Sundaland hotspot (Kottelat, 1989, Woodruff, 2010, Lohman *et al*., 2011).

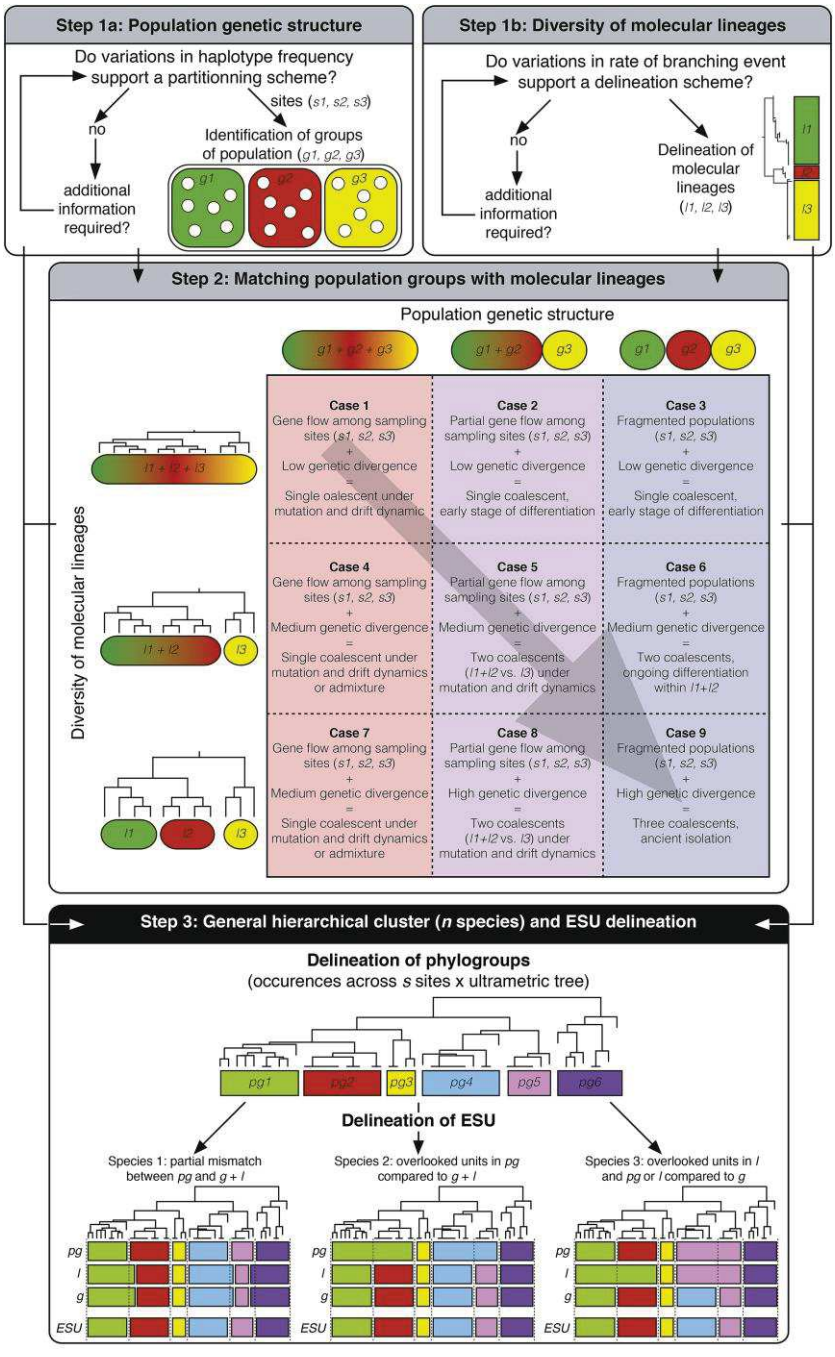


Figure 15. Conceptual framework developed in the present study for the delineation of ESUs. Step 1a, detection of groups of populations differentiated by their haplotypes frequencies. Step 1b, detection of molecular lineages with independent evolutionary dynamics (e.g. lower connectance). Step 2, comparing population groups with molecular lineages. If population genetic structure results from ancient fragmentation of the populations, a correlation between genetic groups and BINs is expected. Conversely, the lack of correlation would indicate that population groups originated recently and either share ancient polymorphism or have been connected by gene flow in a recent past (Fu, 1999, Nielsen and Wakeley, 2001, Wakeley, 2001, Wakeley, 2003). Along the same line, several BINs may be delineated within a genetic group as a consequence of the stochastic nature of the coalescent but not disrupted gene flow (Hudson, 1982, Kingman, 1982, Tajima, 1983). Step 3, Delineation of ESUs for individual species. ESUs are defined based on either variation of haplotype frequencies or independent mitochondrial lineages after comparisons with the phylogroups defined as groups of population sharing similar sets of mitochondrial lineages.

Considering that Indonesia increased its Gross Domestic Product (GDP) and carbon emissions by 1000% and 400% during the last two decades (World Bank), respectively, with a population reaching 260 Millions people, it becomes evident that anthropogenic threats have increased severely. The present study focuses on delineating ESUs and their spatial concordance through the development of a DNA barcode reference library for the widespread

freshwater fishes of the islands of Java and Bali, two of the less explored islands of the Sundaland hotspot (Hubert *et al*., 2015b, Dahruddin *et al*., 2017). Our objective is to provide a 3-step general framework for the delineation of ESUs through the analysis of the spatial and temporal population structures among multiple species based on mitochondrial coalescent trees at the COI gene (Fig. 15). Finally, this framework is used to provide recommendations for evidence-based conservation strategies in the area.

5.3 Materials and methods

5.3.1 Sampling strategy and collection management

A large-scale DNA barcoding campaign was conducted by the authors between November 2012 and May 2015 across 95 sites in Java and Bali islands (Dahruddin *et al*., 2017). A total of 3310 specimens, including 162 species, 110 genera and 53 families were collected, providing a comprehensive assessment of the Javanese and Balinese ichthyofauna. Among those 162 species, three species (*Barbodes binotatus*, *Glyptothorax platypogon* and *Channa gachua*) were sampled in more than 50% of the sites visited, displayed unusually high maximum within-species genetic distances based on a previous assessment using a restricted set of individuals (Dahruddin *et al*., 2017), and as such represented suitable candidates considering the objective of the present studies. Each of the three species belong to different orders displaying varied ecological preferences (Froese and Pauly, 2011), and as such, they ensure that the concordance of their ESUs spatial patterns is unlikely to results from the history of a restricted set of aquatic habitats but originated through common evolutionary dynamics of broad impact on aquatic ecosystems (Avise, 2000, Avise *et al*., 2016).

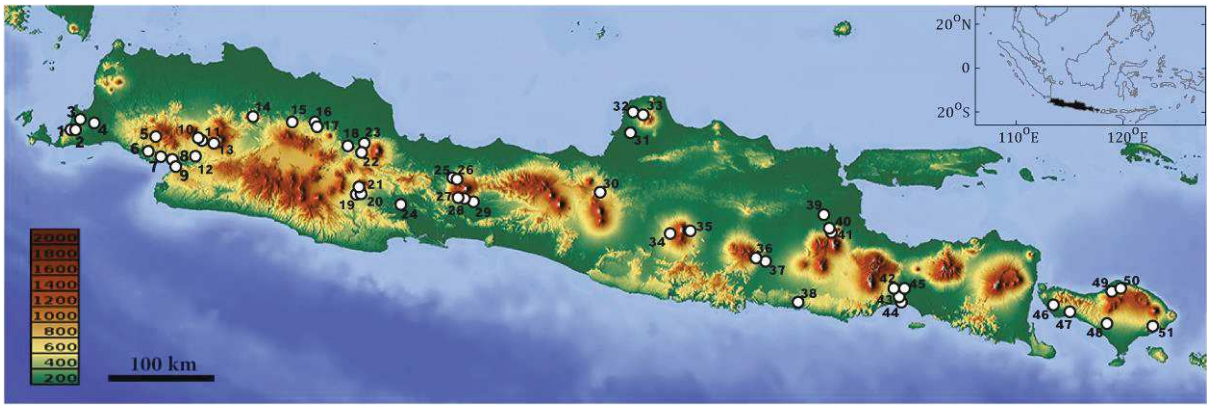


Figure 16. Location of the 51 collection sites for the samples analyzed in this study.

The three species were collected across 51 sites distributed across the islands of Java and Bali (Fig. 16, Supplementary Table S1). Specimens were captured using various gears including electrofishing, seine nets, cast nets and gill nets. Specimens were photographed, individually labeled and voucher specimens were preserved in a 5% formalin solution. A fin clip or a muscle biopsy was taken for each specimen and fixed in a 96% ethanol solution for further genetic analyses. Both tissues and voucher specimens were deposited at the national collections at the Museum Zoologicum Bogoriense (MZB) in the Research Centre for Biology (RCB) from the Indonesian Institute of Sciences (LIPI).

5.3.2 Sequencing and international repositories

Genomic DNA was extracted using a Qiagen DNeasy 96 tissue extraction kit following the manufacturer's specifications. A 651-bp segment from the 5′ region of the cytochrome oxidase I gene (COI) was amplified using primers cocktails C_FishF1t1/C_FishR1t1 including a M13 tails (Ivanova *et al*., 2007). PCR amplifications were done on a Veriti 96-well Fast (ABI-Applied Biosystems) thermocycler with a final volume of 10.0 µl containing 5.0 µl Buffer 2X, 3.3 µl ultrapure water, 1.0 µl each primer (10 µM), 0.2 µl enzyme Phire® Hot Start II DNA polymerase (5U) and 0.5 µl of DNA template (~50 ng). Amplifications were conducted as follow: initial denaturation at 98 °C for 5 min followed by 30 cycles denaturation at 98 °C for 5s, annealing at 56 °C for 20s and extension at 72 °C for 30s, followed by a final extension step at 72 °C for 5 min. The PCR products were purified with ExoSap-IT® (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. Sequencing reactions were performed using the "BigDye® Terminator v3.1 Cycle Sequencing Ready Reaction" and sequencing was performed on the automatic sequencer ABI 3130 DNA Analyzer (Applied Biosystems). The sequences and collateral information have been deposited in BOLD (Ratnasingham and Hebert, 2007) in the projects BIFGA and BIFG in the container 'Barcoding Indonesian Fishes' of the 'Barcoding Fish (FishBOL)' campaign and DNA sequences were submitted to GenBank (accession numbers are accessible directly at the individual records in BOLD).

5.3.3 Population genetic structure (Fig. 15, step 1a)

We examined the distribution of molecular variance through an additive partitioning across increasing spatial scales as implemented in AMOVA (Excoffier *et al*., 1992, Excoffier and Smouse, 1994). We opted for the SAMOVA version that both defines groups of

populations without a priori partitioning scheme and estimates molecular variance within populations, among populations within groups and among groups, based on a simulated annealing approach (Dupanloup *et al*., 2002). SAMOVA is not a decision-based method pointing to the optimal number of groups, we hence applied an empirical threshold and stopped increasing the number of groups once the partitioning scheme produced at least one group consisting of a single sampling site. The SAMOVA were performed for each species using the software SAMOVA 2.0 (Dupanloup *et al*., 2002). SAMOVA partitioning schemes were further compared to the results of a hierarchical clustering based on genetic distances. Mean genetic K2P distances were computed among sampling sites using MEGA 6 (Tamura *et al*., 2013) and used to produce hierarchical clusters derived from the complete linkage algorithm as implemented in hclust function of the R Stats ver. 3.1.2 package (R_Core_Team, 2014). Finally, traditional parameters of population genetic diversity including the number of haplotypes, nucleotidic diversity and mean pairwise K2P distance for each groups were computed using ARLEQUIN 3.5 (Excoffier *et al*., 2005).

5.3.4 Delineating mitochondrial lineages and inferring their diversification (Fig. 15, step 1b)

Several alternative methods have been proposed for delineating molecular lineages (Schloss and Handelsman, 2005, Pons *et al*., 2006, Puillandre *et al*., 2012). They have all in common to detect transition zones in branching patterns resulting from different segments of the gene genealogies that originated from phylogenetic diversification (speciation and extinction) or coalescent dynamics (mutation and drift). We opted for the Refined Single Linkage (RESL) algorithm that considers the number of connections of each sequences in a network estimated through the silhouette index (Rousseuw, 1987) as implemented in BOLD. Sequence connectivity is explored through random walks and optimal partitioning schemes are identified through Markov clustering. At the end, each cluster of sequence is assigned to a Barcode Index Number (BIN) in BOLD (Ratnasingham and Hebert, 2013).

Once BINs were delineated (Table S2), their timing of diversification was explored through the Bayesian approach implemented in BEAST 1.8.1 (Drummond *et al*., 2012). In order to establish robust prior, the best-fit substitution model was selected through the Bayesian Information Criterion (BIC) as implemented in JMODELTEST 2.1.7 (Darriba *et al*., 2012) and further used as a prior for the joint reconstruction of tree topology and divergence times. The initial tree topology was obtained with an UPGMA starting tree and a Coalescent model was used as a tree prior (Kingman, 1982). We used the canonical fish

substitution rate of 1.2% of genetic divergence per Millions years for mitochondrial protein coding gene (Bermingham *et al*., 1997) and applied it to the maximum K2P distance within each species to estimate the age of the Most Recent Common Ancestor (MRCA) to be used as a prior for Bayesian analyses. The prior upper and lower bounds for the MRCAs age intervals were derived from the known highest and smallest substitution rates for fish mitochondrial genomes (Hardman and Lundberg, 2006, Read *et al*., 2006). We ran one MCMC of 10 × 106 step long, sampled every 1000 states with a burn-in period of 10 000. The maximum credibility tree was obtained with TreeAnnotator 1.8.1 after an additional burn-in period of 10 000. Median node ages and 95% highest posterior density (HPD) intervals were plotted in the chronogram. Duplicated sequences were removed for these analyses. We further explored the properties of BINs diversification through Lineage Through Time (LTT) plots (Harvey *et al*., 1994) and the Generalized Skyline Plots (GSP) (Strimmer and Pybus, 2001) methods. The Bayesian LTT was conducted through similar MCMC parameters as of the tree analyses and duplicated sequences were removed. The Bayesian GSP analyses were conducted on the entire data set, including duplicated sequences, a HKY substitution model and MCMC chains of 50 × 106 steps long sampled as described for LTT and tree reconstruction analyses.

5.3.5 Delineating ESU (Fig. 15, steps 2 & 3)

We examined the relationship between the genetic groups delineated by SAMOVA and the BINs (Fig. 15, Step 2) by testing the independence of each partitioning scheme using the Pearson chi-square test of independence as implemented in the R Stats ver. 3.1.2 package (R_Core_Team, 2014). The objective was to examine to what extent the SAMOVA groups of populations were determined by the BINs delineated by the RESL algorithm (Fig. 15, Step 2). We further examined the concordance of the BINs spatial distribution among species by performing a general hierarchical cluster analysis based on the average phylogenetic distance of the BINs among sites as implemented in hclust function of the R Stats ver. 3.1.2 package (R_Core_Team, 2014). The matrix of phylogenetic distance among sites was computed based on the composite chronogram of the three grafted maximum credibility trees of *Barbodes binotatus*, *Channa gachua* and *Glyptothorax platypogon* and BINs occurrence data using the R package PICANTE (Kembel *et al*., 2010). The composite chronogram was constructed by adding internal branches set to 0.001 Millions years between (i) the MRCA of *B. binotatus* and the MRCA of *G. platypogon*, (ii) the MRCA of both *B. binotatus* and *G. platypogon* and

the MRCA of the three species. The internal branch between the MRCA of C. gachua and the MRCA of the three species was further adjusted to produce an ultrametric tree. BINs were considered as absent from sites out of their distribution range but BINs absence was coded as missing data at sites within their distribution range in order to account for sampling uncertainty i.e. a BIN present but not sampled (Table S3). We finally produced distribution maps of the population groups (Fig. 15, Step 1a), the BINs (Fig. 15, Step 1b) and the general hierarchical cluster (Fig. 15, Step 3). The general hierarchical cluster was furthered used to define phylogroups i.e. groups of sites hosting phylogenetically related BINs. The distribution of the phylogroups was used as a template to spot phylogeographic breaks, as exemplified by the phylogroup geographic boundaries, and compared to the distribution of SAMOVA groups and BINS to propose ESUs (Fig. 15, Step 3).

5.4 Results

A total of 317 new sequences were successfully generated for the three species including 108 sequences of *Barbodes binotatus*, 99 sequences of *Channa gachua* and 110 sequences of *Glyptothorax platypogon*. Together with the 37 sequences previously published for these species (Dahruddin *et al*., 2017), a total of 354 sequences were analyzed including 122 sequences of *Barbodes binotatus*, 109 sequences of *Channa gachua* and 123 sequences of *Glyptothorax platypogon* (Table S1). All the sequences were above 500 bp of length and no codon stops were detected, suggesting that the sequences collected represent functional coding regions. The 354 sequences were collected from 39 sites for *B. binotatus*, 28 sites for *C. gachua* and 25 sites for *G. platypogon*.

5.4.1 Population genetic structure and molecular variance

A total of 5, 7 and 6 groups of populations were delineated for *Channa gachua*, *Glyptothorax platypogon* and *Barbodes binotatus* respectively (Table 7, Table S1). For the three species, most of the molecular variance is explained by differences among groups of populations with a percentage of the total variance ranging from 69.9 percent in *B. binotatus* to 78.75 percent in *G. platypogon*. 'Among populations within groups' is the second level in terms of the total variance explained with 24.69 percent for C. gachua and 11.34 percent for *G. platypogon*. *B. binotatus* differs from the two other species in having a slightly higher proportion of the total variance explained by differences within populations, with 19.05 percent, than among populations within groups, with 10.75 percent. The fixation indexes are

significant at all spatial scales supporting spatially structured populations in the three species. This result was further confirmed by the hierarchical cluster analysis of the populations that indicates a substantial genetic differentiation of each group with K2P genetic distances among groups above 0.04, 0.02 and 0.01 for *C. gachua*, *G. platypogon* and *B. binotatus*, respectively (Fig. 17). Within groups of populations, haplotype diversity is high on average with h above 0.5 excepting for population groups II, V and VII of *G. platypogon* with h below 0.2 (Table 8). This high haplotype diversity is opposed to the low nucleotide diversity and the low mean-pairwise differences within groups for the three species.

Table 8. Partitioning of the molecular variance at various spatial scales, fixation indexes and number of population groups as inferred from SAMOVA. The percentage column indicates the amount of total variance explained by each of the hierarchical levels according to the number of groups of population. Φ-statistics estimate the correlation among haplotypes at each of the hierarchical levels examined and their significant departure from a random distribution of the haplotype was tested through randomization across 1000 permutations.

Number of Group	***C. gachua***				***G. platypogon***				***B. binotatus***			
	5				**7**				**6**			
Variance component	**Variance**	**% of total**	***P***	**Φ-statistics**	**Variance**	**% of total**	***P***	**Φ-statistics**	**Variance**	**% of total**	***P***	**Φ-statistics**
Among groups	8.788	70.56	<0.001	Φ*CT*=0.706	8.042	78.75	<0.001	Φ*CT*=0.788	6.958	69.90	<0.001	Φ*CT*=0.702
Among populations within groups	3.071	24.69	<0.001	Φ*SC*=0.839	1.516	11.34	<0.001	Φ*SC*=0.534	2.094	10.75	<0.001	Φ*SC*=0.361
Within populations	0.591	4.75	<0.001	Φ*ST*=0.953	1.015	9.91	<0.001	ΦST=0.901				

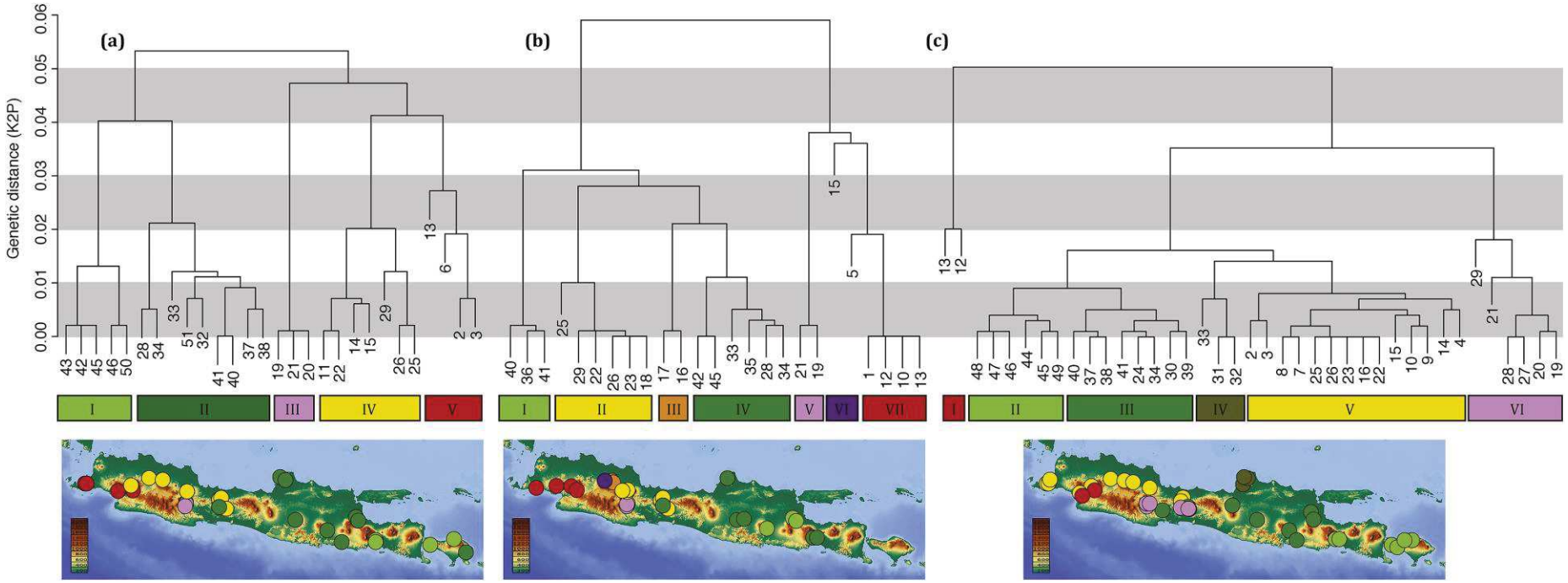


Figure 17. Population genetic structure of *Barbodes binotatus* (a), *Channa gachua* (b) and *Glyptothorax platypogon* (c) as inferred from the hierarchical cluster analysis and SAMOVA.

Table 9. Summary statistics of the genetic diversity including the sampling size (N), haplotype diversity (h), nucleotide diversity (π) and mean number of pairwise differences among haplotypes (mean-pairwise differences) for each of the population groups.

N	***h***	π	**mean-pairwise differences**	
Barbodes binotatus				
I	4	1	0.019	12.17
II	22	0.87	0.002	1.41
III	33	0.83	0.002	1.37
IV	12	0.55	0.005	2.73
V	36	0.82	0.004	2.35
VI	15	0.89	0.011	6.88
Channa gachua				
I	14	0.88	0.006	3.91
II	34	0.93	0.010	6.07
III	11	0.51	0.008	0.51
IV	36	0.81	0.009	5.85
V	14	0.58	0.011	6.97
Glyptothorax platypogon				
I	16	0.58	0.001	0.75
II	34	0.28	0.004	2.29
III	6	0.60	0.001	0.60
IV	39	0.75	0.006	3.62
V	10	0.20	0.002	1.20
VI	7	0.91	0.032	20.80
VII	11	0.18	0.001	0.18

5.4.2 Delineation and diversification of BINs

A total of 9, 7 and 3 molecular lineages (i.e. BINs) were delineated by the RESL algorithm on BOLD for *Channa gachua*, *Glyptothorax platypogon* and *Barbodes binotatus*, respectively (Table S1). Both maximum and average K2P distances are low on average within BINs and contrast with the high maximum K2P distances observed for each of the three coalescent trees (Table 9). Among the 19 BINs delineated, four were represented by singletons in *C. gachua* (ACQ3951, ACQ6941, ACQ6939, ACQ6940). The Bayesian analyses yielded three maximum credibility trees (Fig. 18a, b, c) that confirmed the contrast between the deep divergence among BINs, ranging from 0.47 to 2.75 Million years ago (Ma), and the shallow coalescent depth of the BINs ranging from 0.08 to 0.89 Ma (Table 9). The age of the MRCA was very similar among the three species ranging from 2.71 Ma for *B. binotatus* to 3.14 Ma for *G. platypogon* (Table 9). Plotting the BINs coalescent depth on the LTT curves further confirmed that the molecular diversity within BINs accumulated very recently within the three species (Fig. 18d, e, f). The Bayesian GSP yielded very similar demographic trajectories for the three species with a global trend of constant population size over the last 2 millions years and a steep decline of population size during the last 100.000 years (Fig. 18g, h, i).

Table 10. Summary statistics of the genetic K2P distances and age estimates. The maximum and average K2P distances are provided for each BINs and the maximum K2P distances are provided for the entire coalescent trees. The age estimate of the MRCA is provided based on three alternative hypotheses of molecular clock including 0.005 genetic divergence per million years (Hardman and Lundberg, 2006, Read *et al.*, 2006), 0.012 genetic divergence per million years (Bermingham *et al.*, 1997) and 0.02 genetic divergence per million years (Read *et al.*, 2006). Divergence estimates are derived from the Bayesian analyses based on the H2 calibration with upper and lower bounds of the prior age intervals defined by the molecular clock hypotheses H1 and H3.

K2P distance (%)		MRCA H1[a](Myr)	MRCA H2[b](Myr)	MRCA H3[c](Myr)	Divergence estimates (Myr)	Coalescent depth (Myr)	
Maximum	**Average**						
Channa gachua							
Coalescent	6.78		6.78	2.83	1.70	2.75 (6.02-1.70)	
ACQ0292	2.1	0.87				2.04 (4.56-0.82)	0.66 (1.60-0.26)
ACQ0290	1.4	0.61				2.04 (4.56-0.82)	0.61 (1.49-0.16)
ACQ0291	0.15	0.08				2.75 (6.02-1.70)	0.08 (0.31-0.01)
ACQ3951	-	-				1.23 (2.87-0.41)	-
ACQ6941	-	-				0.73 (1.75-0.24)	-
ACQ6940	-	-				0.47 (1.17-0.24)	-

K2P distance (%)		MRCA H1[a](Myr)	MRCA H2[b](Myr)	MRCA H3[c](Myr)	Divergence estimates (Myr)	Coalescent depth (Myr)	
Maximum	**Average**						
ACQ6939	0.00	0.00				0.47 (1.17-0.24)	-
ACQ3952	1.4	0.31				0.87 (2.08-0.31)	0.14 (1.40-0.14)
ACQ3950	0.93	0.32				0.87 (2.08-0.31)	0.33 (0.84-0.08)
Glyptothorax platypogon							
Coalescent	7.76		7.76	3.24	1.94	3.14 (6.74-1.94)	
ACQ5850	-	-				1.05 (2.68-0.21)	0.13 (0.45-0.01)
AAY1028	0.00	0.00				1.05 (2.68-0.21)	0.08 (0.33-0.01)
ACQ5898	0.93	0.23				1.75 (4.22-0.51)	0.39 (1.13-0.05)
ACP6225	0.31	0.12				1.98 (4.71-0.73)	0.23 (0.72-0.03)
ACQ6223	1.24	0.55				1.09 (2.63-0.32)	0.50 (1.35-0.14)
ACP6117	0.15	0.09				0.69 (1.75-0.18)	0.08 (0.34-0.01)
ACQ6224	0.46	0.05				0.69 (1.75-0.18)	0.21 (0.56-0.05)
Barbodes binotatus							
Coalescent	6.80		6.8	2.84	1.70	2.71 (5.96-1.70)	
ACP6290	0.01	0.00				2.71 (5.96-1.70)	0.12 (0.41-0.01)
ACP6025	0.57	1.09				1.74 (4.26-0.48)	0.38 (1.12-0.08)
ACP5712	1.94	0.73				1.74 (4.26-0.48)	0.89 (2.33-0.48)

a
0.005.

b
0.012.

c
0.02.

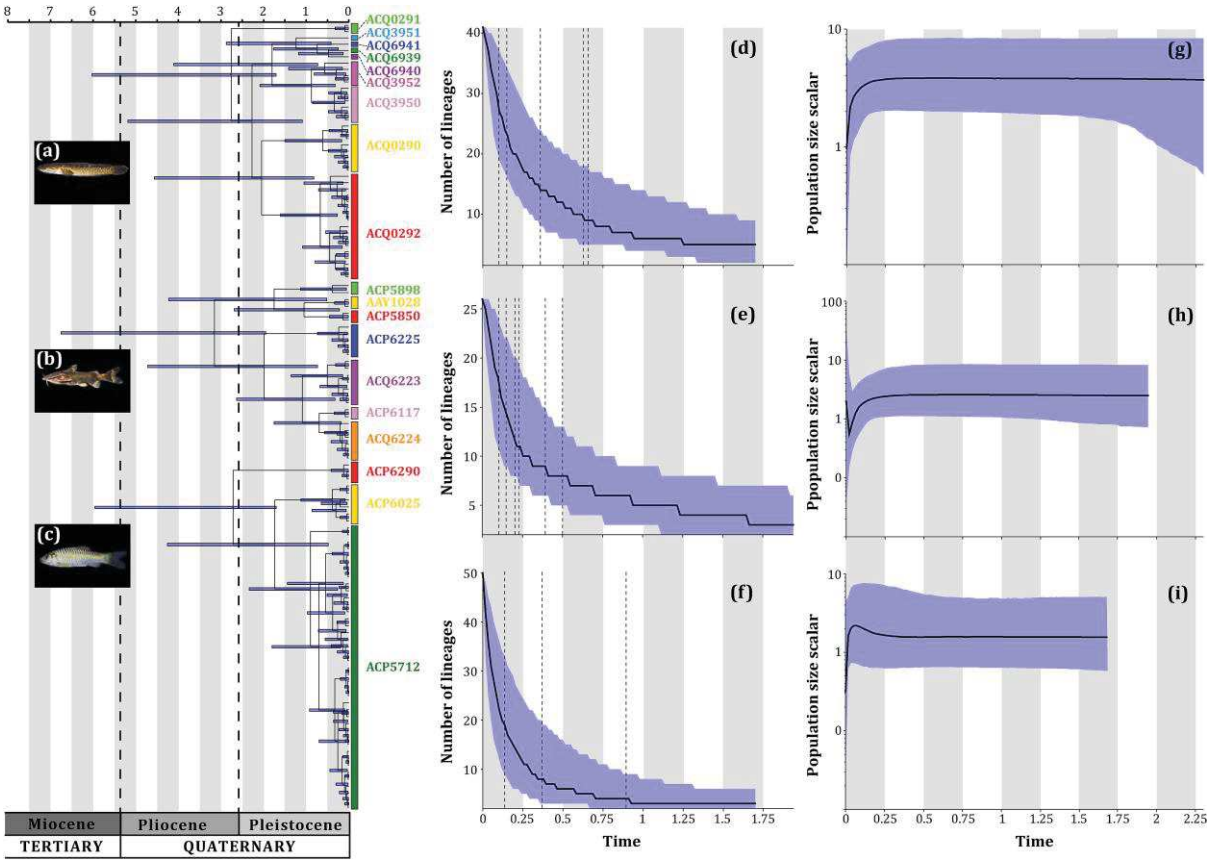


Figure 18. BINs diversification patterns of *Channa gachua* (a, d and g), *Glyptothorax platypogon* (b, e and h) and *Barbodes binotatus* (c, f and i) including Maximum Credibility Trees (a, b, c); Lineage Through Time plot (d, e, f) and Bayesian Generalized Skyline Plots (g, h, i). Solid black line in d, e and f is the median diversity accumulation curve, blue shaded area is the 95% highest posterior density intervals and dotted lines represent the BIN coalescent depth in Million years. The solid black line in g, h and i represent the median effective population size, blue shaded area is the 95% highest posterior density intervals and Population size scalar (in millions) = effective population size x generation time. Calibrations are derived from previously published molecular clock hypotheses applied to the maximum K2P distance for each species - i.e. age of the MRCA (see Table 9). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4.3 Identifying phylogroups and delineating ESUs

The null hypothesis of independence between population groups delineated by SAMOVA and BINs was rejected by the Pearson chi-square test ($X2 = 4486.7$, p-value<0.001) suggesting that the delineation of population groups is largely supported by alternative distribution of BINs among populations. The general hierarchical cluster constructed using the phylogenetic distance among sites resulted in the delineation of five major phylogroups, diverging by more than 5 Million years on average, and ordered into two main clusters (Fig. 19). The geographic range of phylogroups I and II showed no overlap and allowed identifying three phylogeographic breaks (Fig. 19b). Phylogroup I is restricted to the western

part of Java while phylogroup II shows an alternative distribution in eastern Java and Bali. The first phylogeographic break identified is located in the Java-Bali straight based on the segregation of phylogroups I and II on each side of the straight (Fig. 19, break 1). The second phylogeographic break identified is located in central Java, in between the western and central volcanic arches (Fig. 19, break 2). The third phylogeographic break is located on the western most slope of the western volcanic arch based on the western limit of phylogroup II (Fig. 19, break 3). The distribution ranges of phylogroups III, IV and V are more intricated but show some similarities with range distribution of phylogroups I and II. The Java-Bali straight is also identified as a phylogeographic break segregating phylogroups IV and V, each being alternatively distributed in Java or Bali (Fig. 19, break 1). The central Java phylogeographic break is also associated with the segregation of phylogroups III and V (Fig. 19, break 2) but two additional phylogeographic breaks are identified in central java, in between the western and central volcanic arches, that are associated with the segregation of phylogroup III from the phylogroup V (Fig. 19, break 4) and the segregation of an isolated site of phylogroup V from the eastern sites (Fig. 19, break 5). The western most phylogeographic break is also associated with the segregation of phylogroups IV and V in their western range but with some overlaps (Fig. 19, break 3).

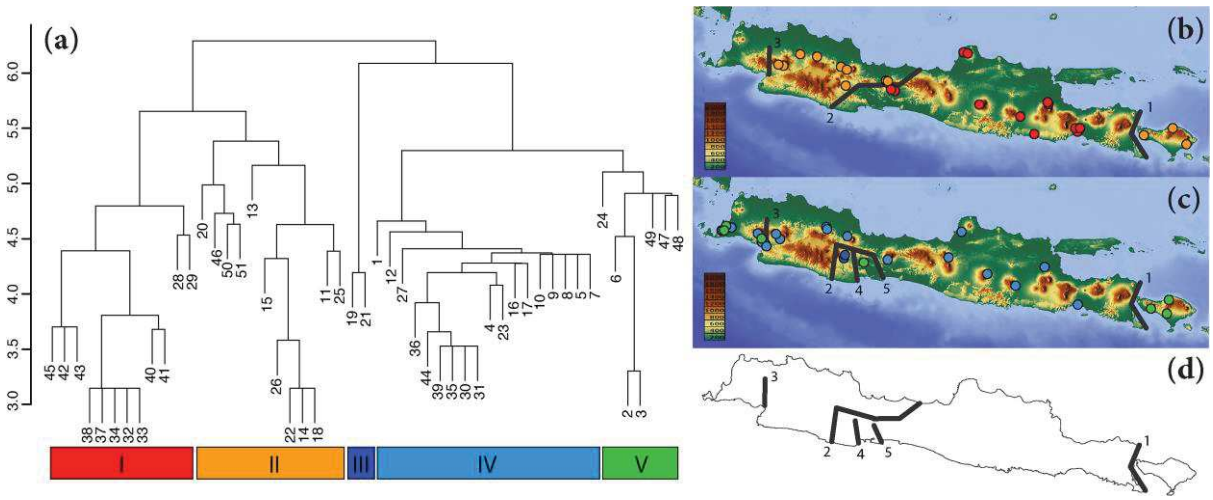


Figure 19. Delineation of phylogroups and detection of phylogeographic breaks. a general hierarchical cluster based on the phylogenetic distances among sites established from the grafted chronograms of *C. gachua, G. platypogon* and *B. binotatus*, and occurrence data (see Table S2). b distribution range of phylogroups I and II, and associated phylogeographic breaks (solid black lines). c distribution range of phylogroups III, IV and V, and associated phylogeographic breaks (solid black lines). d phylogeographic breaks identified from the phylogroup geographic boundaries and further used for delineating ESUs.

The boundaries of population groups and BINs range distribution are generally associated with the five phylogeographic breaks for the three species with some exceptions (Fig. 20). Three cases of trans distribution across phylogeographic breaks are detected in *C. gachua* (Fig. 20a and b). The first case is located at the phylogeographic break 3 with the population group V (Fig. 20a, red). The population group V, however, is associated with four independent BINs. The second case is associated with the phylogeographic break 2 with the population group IV (Fig. 20a, yellow). Finally, the population groups I and II are distributed on each side of the phylogeographic break 1 (Fig. 20a, green groups). The two cases of trans distribution across phylogeographic breaks 1 and 2 are also observed for *G. platypogon* with population group VII overlapping break 1 (Fig. 20d, red) and population group II overlapping break 2 (Fig. 20d, yellow). In addition, two groups of population corresponding to two distinct BINS in Eastern Java are detected that are not associated with a phylogeographic break (Fig. 20d, green groups; Fig. 20e, purple and blue). Trans distribution across the phylogeographic breaks 1 and 3 are also observed in *B. binotatus* for population groups II and III (Fig. 20g, green groups) and population group V (Fig. 20g, yellow), respectively. Other trans distributions are also observed for the population group VI (Fig. 20g, pink) across the phylogeographic breaks 4 and 5, and the population group III (Fig. 20g, dark green) across the phylogeographic breaks 2 and 5. Similarly with *G. platypogon*, additional population groups not associated with previously identified phylogeographic breaks are detected for B. binotatus (Fig. 20g, green groups). These populations groups, however, do not match BINs boundaries as population groups II, III and IV belong to the same BIN.

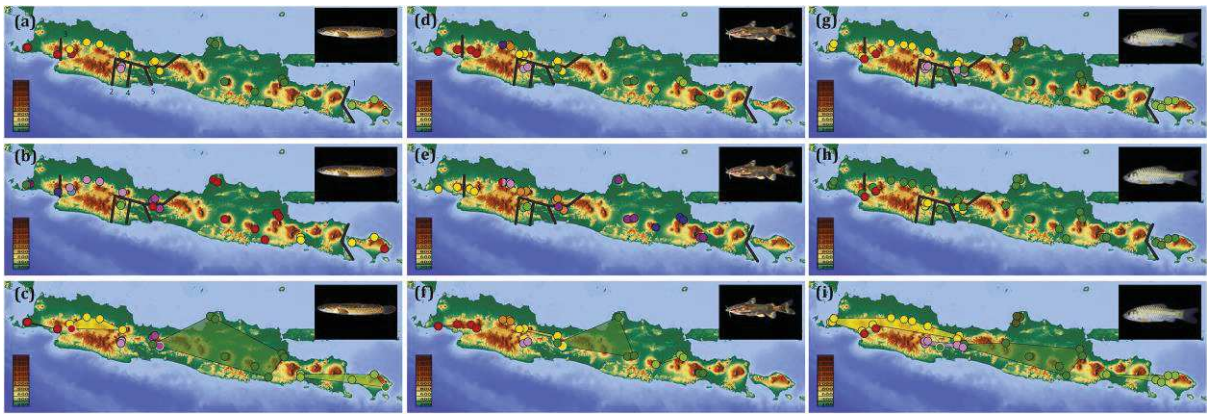


Figure 20. Mapping of the population groups identified by SAMOVA (a, d and g), the BINs delineated by the RESL algorithm (b, e and h) and ESUs candidates (c, f and i) for *C. gachua* (a, b and c), *G. platypogon* (d, e and f) and *B. binotatus* (g, h and i). Ambiguous assignment of populations to ESUs due to trans distribution across phylogeographic breaks are highlighted by solid white circle (c, f and i).

Following the identified match between phylogeographic breaks, population groups and BINs, 6 ESUs are proposed for *C. gachua* (Fig. 20c), *G. platypogon* (Fig. 20f) and *B. binotatus* (Fig. 20i) that show similar range distribution. In the three species, most population groups and BINs show great overlap in their distribution. Assigning sites to ESUs was conflicting in only a few cases associated with the cases of trans distribution above mentioned (Fig. 20c, f and i; solid white circle). Following the theoretical framework of the present study (Fig. 15), populations associated with trans distribution were assigned to ESUs according to their relative position to phylogeographic breaks. Along the same line, ESUs were delineated based on either population groups or BINs whenever each grouping scheme provided conflicting sorting e.g. two population groups within the same BIN or several BINs within a population group. A single exception was done to this general principle for the BINs in *C. gachua* represented by singletons (Fig. 20b; purple, dark blue, light blue, dark green) and associated to the same population group (Fig. 20a; red) that were grouped into a single ESU.

5.5 Discussion

5.5.1 Regional patterns and evolutionary dynamics

Our study emphasizes how the joint use of population clustering and molecular lineage delineation methods enhances our understanding of population evolutionary dynamics and resulting patterns of genetic structure. Population dynamics are ruled by stochastic and deterministic mechanisms acting at varying scales. While deterministic mechanisms, such as fragmentation through geological dynamics, are usually predominant in shaping population structure by limiting gene flow (Hanski, 1991, Hastings and Harrison, 1994, Moilanen and Hanski, 1998), stochastic dynamics predominate at local scale within populations or tightly connected populations where they determine genetic diversity through genetic drift (Vellend, 2005, Vellend and Geber, 2005, Hubert *et al*., 2015a). Landscapes are dynamic, particularly so in insular systems (Warren *et al*., 2014), and deterministic mechanisms may predominate during major landscape changes while stochastic mechanisms may take over during equilibrium dynamics (Alonso *et al*., 2006, Hubert *et al*., 2015a).

The present study illustrates the benefits of comparing patterns of population structure across multiple species to disentangle the relative contribution of deterministic and stochastic mechanisms on population genetic structure, particularly so in the absence of a priori biogeographic information. The global concordance in the population genetic structure of *B.*

binotatus, *C. gachua* and *G. platypogon* argue for a predominant influence of deterministic processes in shaping population structure. This trend could be expected considering the volcanic origin of Java and Bali islands and the stringent constraint of geological history on species range evolution (Lohman *et al*., 2011). For instance, Java originated from the merging of two volcanic arches, one is the West and a second in the east including East Java and Bali, that emerged between 10 and 5 Ma and further aggregated during the last 5 Million years. This scenario was reflected in the closer genetic affinities between the central-eastern Java and Bali population groups than between the central-eastern and the western population groups in Java of C. gachua and B. binotatus. This pattern supported the identification of the phylogeographic break 2 located in between the central and western volcanic arches of Java.

The orogeny of Java started at 10 million years and landscapes likely achieved their modern configuration during the last 5 million years (Lohman *et al*., 2011). The ancient history of Java landscapes is reflected in the lack of independence between the groups of population and the delineated molecular lineages, stressing that the population structures detected have been established for a long time. This trend was further reflected in the large difference between genetic distances and age estimates within and among population groups. The spatial structures depicted here and inferred timing of population settlement further highlight the limited dispersal opportunities for aquatic biotas during the geological history of Java and Bali. This trend suggest that the colonization of Java and Bali rivers by primary freshwater fishes happened during the earliest geological stage, as exemplified by the deep divergence among BINs, with a subsequent fragmentation of the populations rather than multiple colonization through time. This assumption is further supported by the similar age estimates of the BINs in between 750.000 and 150.000 years ago, suggesting a synchronous fragmentation of the populations. Along the same line, multiple colonization events through range expansion are frequently associated with a signature of past population growth (Slatkin and Hudson, 1991, Rogers and Harpending, 1992, Harpending *et al*., 1998). The inferred demographic trajectories, and their remarkable similarity for the three species, do not support a scenario of colonization through range expansion as constant population sizes are inferred during the last 2 Million years.

5.5.2 Implications for the conservation of aquatic ecosystems in South Sundaland

The settlement of Sundaland results from intricate interactions between plate tectonics and eustatic changes (Kottelat, 1989, Kottelat *et al*., 1993, Woodruff, 2010, Lohman *et al*., 2011).

The emergence of Sundaland resulted from the subduction of the Australian plate beneath Sundaland during the last 10 Ma (Lohman *et al*., 2011, Hall, 2013), a process initiated with the isolation of Borneo and continued with the emergence of Sumatra and Java from 5 Ma onward. Once the maximum depth of the Java sea reached nearly 120 m, however, sea levels fluctuations associated with Milankovitch cycle (Hays *et al*., 1976) started to interfere with the orogeny of Sundaland islands (Kottelat *et al*., 1993, Woodruff, 2010). During sea level low-stands, Sundaland islands were connected to each others, prompting the settlement of four large palaeodrainages straddling across islands (Kottelat, 1989, Kottelat *et al*., 1993). Such extended watershed surface happened repeatedly during the late Pleistocene and as a consequence, Sundaland ecosystems are currently in a refugial state, occupying only 50–75% of their maximal Pleistocene extent (Woodruff, 2010).

The demographic inferences produced here, as well as population genetic diversity estimates, are consistent with a scenario of biome contraction during the Pleistocene. Steep declines in population size are detected in *C. gachua*, *G. platypogon* and *B. binotatus* and their inferred time-frame are consistent with a late Pleistocene fragmentation of aquatic biotas in Sundaland. The shallow coalescent trees of the population groups and low nucleotide diversity further support that populations experienced bottleneck in the past. The high haplotype diversity and observed reciprocal monophyly among most population groups, however, argue that population size reduction is historical and populations are recovering. This observation is suggested by the demographic inference of *G. platypogon* where a recent population growth is detected. This trend, however, is not observed in *B. binotatus* and *C. gachua* and considering the notorious difficulties to disentangle multiple demographic events in structured populations (Markovtsova *et al*., 2000, Heller *et al*., 2013), the recovery state of the populations remain to be explicitly addressed.

The present study highlight several properties of the ESUs defined here that provides some guidelines for future conservation plans of the Javanese ichthyofauna. First, each ESU display substantial levels of genetic divergence, with divergence age estimates beyond 1 Million years. The population genetic structure inferred argue for a differentiation in allopatry and considering that these BINs have been consistently assigned to the same nominal species by taxonomists (Kottelat, 2013, Ng and Kottelat, 2016), they show conserved morphological attributes. This trend could be expected when differentiation happens in allopatry across homogeneous environmental and biotic conditions (Hubert *et al*., 2015a). These large levels of genetic divergence, however, question the reproductive status of those ESUs as genomic incompatibilities may accumulate through time, despite conserved

eco-morphological characteristics, and produce post-zygotic isolation (Orr and Turelli, 2001, Brideau *et al*., 2006). Second, these ESUs show remarkably restricted range distribution, particularly so in Western Java, and low genetic diversity at the nucleotide level. Considering the demographic trend of population contraction observed in the three species, these ESUs should be treated as small populations, particularly sensitive to further size reduction that may result from anthropogenic perturbations. Nevertheless, Java is currently the most densely populated island in Sundaland and the reduction of aquatic habitats has been severe during the last two decades. This situation is clearly of concern as it is now widely acknowledged that important shifts in life history traits (LHT) and reduction of genetic diversity are precursor of extinction vortex - reduced population size and increased demographic variance induce either a spatial fragmentation or a decrease of population adaptive potential (Gilpin and Soulé, 1986, Fagan and Holmes, 2006). These two characteristics of the 18 ESUs call for a specific assessment of their genetic exchangeability if breeding programs were to be considered. In particular, the success of restoration program through genetic rescue, aiming at increasing population fitness through the repletion of genetic diversity by immigration, is tightly dependent on the exchangeability of the immigrants (Whiteley *et al*., 2014). From an ecological perspective, the 18 ESUs show no pattern of altitudinal zonation (Fig. S1) that usually results from adaptive changes to heterogeneous thermal regimes and biotic interactions (Angilletta *et al*., 2006). This was expected considering the commonness of these three species during the field sampling. From a genetic perspective, however, the high level of genetic divergence among BINs calls for a conservative approach in breeding program that focus on maintaining the genetic integrity of the ESUs identified here. A few tenth of generation might be sufficient to accumulate adaptive combinations of alleles for selected genes and disrupting those allelic combinations may result in significant decreases of fitness (Whiteley *et al*., 2014). Before further evidences are collected from ecological and population genetic perspectives, the present study argue against any translocation program at large spatial scale for those species that, if not accounting for the deep population structure observed, may have dramatic consequences on populations fitness.

5.5.3 Validation of the approach and generalization

Defining biological units for conservation purposes is challenging due to the multiple interactions that drive population dynamics in nature (Vellend and Geber, 2005, Urban *et al*., 2008, Vellend, 2010, Hubert *et al*., 2015a). As a consequence, the concept of ESUs has been

subject of much debate since its earliest development (Crandall *et al*., 2000). Preserving populations genetic diversity is a mandatory step for successful conservation or restoration programs as genetic diversity has direct consequences on ecological characteristics (Hughes *et al*., 2008) and preserving genetic diversity usually implies accounting for population genetic structure in delineating ESUs that comprehensively cover species genetic diversity (Waples, 1991, Waples, 1995, Dizon *et al*., 1992, Moritz, 1994, Waples, 1995, Fraser and Bernatchez, 2001). The present study proposes a heuristic framework that enables the delineation of ESUs in the absence of a priori phylogeographic knowledge through the joint use of population clustering and molecular delineation methods across multiple species.

The three species present marked population genetic structure, a trend that may constitute a particularity of the Java and Bali aquatic systems. The concordance between the groups of population defined by SAMOVA and the BINs delineated by the RESL algorithm argue that most population groups achieved reciprocal monophyly. In this context, population groups may be treated as independent cryptic lineages instead of groups of population, a trend that has been previously observed for several lineages of freshwater fishes in Sundaland (Nguyen *et al*., 2008, Pouyaud *et al*., 2009, De Bruyn *et al*., 2013). Such system, however, exemplify the potential of a community-level assessment of population genetic structure through a fast method of molecular screening such as DNA barcoding. First, the concordance between the gene genealogies of *C. gachua*, *B. binotatus* and *G. platypogon* helped identify common trends in population grouping and associated phylogeographic breaks in the absence of a priori phylogeographic knowledge. This general trend further enabled the delineation of ESUs based on a newly established biogeographic background that helped refine population assignments in case of conflicting grouping among species. Second, this approach allowed to identify populations that may have a specific status, such as populations located in secondary contact zone and having experienced hybridization and introgression. Several cases were identified here with populations that have been assigned to a different ESU based on their trans distribution across phylogeographic breaks. Further addressing the status of these populations may gather important information about their reproductive status. For instance, none of the populations analyzed in the present study harbor more than one BIN, excepting one population of *G. platypogon*, despite the mosaic distribution across several phylogeographic breaks suggesting a dynamic of secondary contact. Third, the present study is based on a fast method of molecular screening based on mitochondrial DNA. The ease of amplification of COI sequences for fishes (Ward *et al*., 2009, April *et al*., 2011, Hubert *et al*., 2012, Pereira *et al*., 2013) offers a fast and effective heuristic approach to the exploration of

population genetic structure and ESUs delineation (Hajibabaei *et al*., 2007). This approach is thus particularly welcomed when phylogeographic and taxonomic knowledge is scarce and the extent of anthropogenic threat on biodiversity urge for an objective assessment of conservation priorities.

The effectiveness of our approach for delineating ESUs in the present system is due in part to the deep population genetic structure of the species analyzed. As such, the usefulness of the method remains to be explored for less fragmented and open systems. Insular biotas are frequently built upon in situ diversification instead of immigration, as observed in continental systems, and evolutionary diversification happen at smaller spatial scales (Emerson and Gillespie, 2008, Warren *et al*., 2014, Hubert *et al*., 2015a). The usefulness of the present method depends on the biogeographic context that dictates the spatial scale of diversification and population genetic structure (Holt, 1993, Ricklefs and Schluter, 1993). In opened systems, shared patterns of population genetic structure may not be a common trend and ESU should be delineated based on an individual basis as each species may have experienced distinct evolutionary history (Bowen *et al*., 2016). Along the same line, shallow population genetic structure may be determined by subtle variations in haplotype frequencies and substitution rates at COI might not be sufficient to reliably detect groups of population. In that case, alternative approaches should be favored based on the assessment of molecular markers displaying higher levels of polymorphism, such as those resulting from length polymorphism (Selkoe and Toonen, 2006). In addition, the maternal inheritance of mitochondrial DNA is a further limit for assessing gene flow, particularly so when subtle variations in haplotypes frequencies are observed (Nielsen and Wakeley, 2001). Here, both groups of population and molecular lineages were determined based on the same data set. The present framework, however, is flexible and both population grouping and molecular lineages may be delineated based on independent data sets. In addition, the general hierarchical cluster was established based on phylogenetic relationships of the BINs as the groups of population were largely delineated based on BINs boundaries. The general cluster may be established based on population groups instead of BINs providing that both delineating schemes are independent and shallow population genetic structure are observed. In this case, variations in allelic frequencies might be more successful at identifying common phylogeographic breaks, particularly so if population genetic structures established recently.

5.6 Conclusion

While we are facing a major biodiversity crisis, fast, cheap and universal methods for delineating ESUs are needed. The inventory of earth diversity and our understanding of its origin through space and time are still fragmentary, particularly in the tropics where the taxonomic impediment dramatically hampers conservation (Garnett and Christidis, 2017). DNA barcoding has already established as a new reference for species identification and its utility for the inventory of tropical biotas is increasingly acknowledged (Hebert *et al*., 2004, Smith *et al*., 2005, Smith *et al*., 2007, Monaghan *et al*., 2009, Hubert *et al*., 2012, Tänzler *et al*., 2012, Riedel *et al*., 2013). Its application for the delineation of ESUs, however, has been much less explored. The present study suggests that DNA barcoding may be successfully used for this purpose, providing that the limits of the mitochondrial genome, due to its maternal inherence, are balanced by the comparison of population genetic structure across multiple species. In the context of Java and Bali aquatic biotas, our approach helped identify ESUs in the absence of prior phylogeographic knowledge and produce recommendations for the conservation of the Java and Bali freshwater fishes.

5.7 Acknowledgements

The authors wish to thank Siti Nuramaliati Prijono, Witjaksono, Mohammad Irham, Marlina Adriyani, Ruliyana Susanti, Hari Sutrisno and Sri Sulandari, at Research Centre for Biology (RCB-LIPI); Jean-Paul Toutain, Robert Arfi and Jean-François Agnèse from the 'Institut de Recherche pour le Développement'; Joel Le Bail and Nicolas Gascoin at the French embassy in Jakarta for their continuous support. We are thankful to Daisy Wowor and Ujang Nurhaman at RCB-LIPI, Sumanta and Bambang Dwisusilo at IRD Jakarta for their help during the field sampling in Java and Bali. Part of the present study was funded by the Institut de Recherche pour le Développement (UMR226 ISE-M and IRD through incentive funds), the MNHN (UMR BOREA), the RCB-LIPI, the French Ichthyological Society (SFI), the Foundation de France and the French embassy in Jakarta. The Indonesian Ministry of Research approved this study and field sampling was conducted according to the research permits 097/SIP/FRP/SM/IV/2014 for Philippe Keith, 60/EXT/SIP/FRP/SM/XI/2014 for Frédéric Busson and 41/EXT/SIP/FRP/SM/VIII/2014 for Nicolas Hubert. Sequence analysis was aided by funding from the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project. We thank Paul Hébert and Evgeny Zakharov as well as BOLD and CCDB staffs at the University of Guelph for their valuable support. This publication has ISEM number 2017-278 SUD.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.gecco.2017.11.005.

References

Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. Trends Ecol. Evol. 21, 451-457.

Anderson, C.D., Epperson, B.K., Fortin, M.-J., Holderegger, R., James, P., Rosenberg, M.S., Scribner, K.T., Spear, S., 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol. Ecol. 19, 3565-3575.

Angilletta, M.J.J., Bennett, A.F., Guderley, H., Navas, C.A., Seebacher, F., Wilson, R.S., 2006. Coadaptation: a unifying principle in evolutionary thermal biology. Physiological Biochem. Zool. 79, 282-294.

April, J., Mayden, R., Hanner, L.,R.H., Bernatchez, L., 2011. Genetic calibration of species diversity among North America's freshwater fishes. Proc. Natl. Acad. Sci. U. S. A. 108, 10602-10607.

Avise, J.C., 1989. Molecular Markers, Natural History and Evolution. New York.

Avise, J.C., 2000. Phylogeography: the History and Formation of Species. Harvard University Press.

Avise, J.C., Bowen, B.W., Ayala, F., 2016. In the light of evolution X: comparative phylogeography. Proc. Natl. Acad. Sci. U. S. A. 113, 7957-7961.

Beheregaray, L.B., 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern hemisphere. Mol. Ecol. 17, 3754-3774.

Bermingham, E., McCafferty, S., Martin, A.P., 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian isthmus. In: Kocher, T.D., Stepien, C.A. (Eds.), Molecular Systematics of Fishes. CA Academic Press, San Diego, pp. 113-128.

Bowen, B.W., Gaither, M.R., DiBattista, J.D., Iacchei, M., Andrews, K.R., Grant, W.S., Toonen, R.J., Briggs, J.C., 2016. Comparative phylogeography of the ocean planet. Proc. Natl. Acad. Sci. U. S. A. 113, 7962-7969.

Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X., Barbash, D.A., 2006. Two dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science 314, 1292-1295.

Crandall, K.A., Bininda-Emonds, O.R.P., MAce, G.M., Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15, 290-295.

Dahruddin, H., Hutama, A., Busson, F., Sauri, S., Hanner, R., Keith, P., Hadiaty, R.K., Hubert, N., 2017. Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. Mol. Ecol. Resour. 17, 288-299.

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

De Bruyn, M., Rüber, L., Nylinder, S., Stelbrink, B., Lovejoy, N.R., Lavoue, S., Heok Hui, T., Nugroho, E., Wowor, D., Ng, P.K.L., Siti Azizah, M.N., Von Rintelen, T., Hall, R., Carvalho, G.R., 2013. Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast asian biodiversity hotspots. Syst. Biol. 62, 398-410.

Dizon, A.E., Lickyer, C., Perrin, W.F., *et al*., 1992. Rethinking th stock concept: a phylogeographical approach. Conserv. Biol. 6, 24-36.

Drummond, A., Suchard, M., Xie, D., Rambault, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29.

Dupanloup, I., Schneider, S., Excoffier, L., 2002. A simulated annealing approach to define the genetic structure of populations. Mol. Ecol. 11, 2571-2581.

Emerson, B.C., Gillespie, R.G., 2008. Phylogenetic analysis of community assembly and structure over space and time. Trends Ecol. Evol. 23, 519-530.

Endo, Y., Nash, M., Hoffmann, A.A., Slatyer, R., Miller, A.D., 2014. Comparative phylogeography of alpine invertebrates indicates deep lineage diversification and hisgorical refugie in the Australian Alps. J. Biogeogr. 42, 89-102.

Excoffier, L., Smouse, P., 1994. Using allele frequencies and geographic subdivision to reconstruct gene genealogies within a species. Molecular variance parsimony. genetics 136, 343-359.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol. Bioinforma. online 1, 47-50.

Excoffier, L., Smouse, P., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. genetics 131, 479-491.

Fagan, W.F., Holmes, E.E., 2006. Quantifying the extinction vortex. Ecol. Lett. 9, 51-60.

Fraser, D.J., Bernatchez, L., 2001. Adaptative evolutionary conservation: toward a unified concept for defining conservation units. Mol. Ecol. 10, 2741-2752.

Froese, R., Pauly, D., 2011. Fishbase. Worldwide web electronic publication. http://www.fishbase.org. version (06/2011).

Fu, Y.-X., 1999. Coalescing into the 21st Century: an overview and prospects of coalescent theory. Theor. Popul. Biol. 56, 1-10.

Garnett, S.T., Christidis, L., 2017. Taxonomy anarchy hampers conservation. Nature 546, 25-27.

Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W., Hajibabaei, M., 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proc. Natl. Acad. Sci. 111, 8007-8012.

Gilpin, E., Soule, M., 1986. Minimum viable populations: processes of species extinction. In: Soule, M.E. (Ed.), Conservation Biology: the Science of Scarcity and Diversity. Sinauer, Sunderland, Mass, pp. 19-34.

Gutierrez-Tapia, P., Palma, R.E., 2016. Integrating phylogeography and species distribution models: cryptic distributional reponses to past climate change in an endemic rodent from central Chile hotspot. Divers. Distrib. 22, 638-650.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6, e17497.

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

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

Hanski, I., 1991. Single-species metapopulation dynamics: concepts, models and observations. Biol. J. Linn. Soc. 42, 17-38.

Hardman, M., Lundberg, J.G., 2006. Molecular phylogeny and chronology of diversification for "phractocephaline" catfishes (Siluriformes: Pimelodidae) based on mitochondrial DNA and DNA recombination activating gene 2 sequences. Mol. Phylogenet. Evol. 40, 410-418.

Harpending, H., Batzer, M., Geurven, M., Jorde, L., Rogers, A., Sherry, S., 1998. Genetic traces of ancient demography. Proc. Natl. Acad. Sci. 95.

Harvey, P.H., May, R.M., Nee, S., 1994. Phylogenies without fossils. Evolution 48, 523-529.

Hastings, A., Harrison, S., 1994. Metapopulation dynamics and genetics. Annu. Rev. Ecol. Syst. 25, 167-188.

Hays, J.D., Imbrie, J., Shackleton, N.J., 1976. Variations in the earth's orbit: pacemaker of the ice ages. Science 194, 1121-1132.

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

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. U. S. A. 101, 14812-14817.

Heller, R., Chikhi, L., Siegismund, H.R., 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. Plos One 8, e62992.

Hoffman, M., Hilton-Taylor, C., Angulo, A., B€ohm, M., Brooks, T.M., Butchart, S.H.M., Carpenter, K.E., Chanson, J., Collen, B., Cox, N.A., *et al*., 2010. The impact of Conservation on the status of the world's vertebrates. Science 330, 1503-1509.

Holderegger, R., Wagner, H.H., 2006. A brief guide to landscape genetics. Landsc. Ecol. 21, 793-796.

Holt, R., 1993. Ecology at the mesoscale: the influence of regional processes on local communities. In: Ricklefs, R., Schluter, D. (Eds.), Species Diversity in Ecological Communities: Historical and Geographical Perspectives. University of Chicago Press, Chicago.

Hua, X., Wiens, J.J., 2010. Latitudinal variation in speciation mechanisms in frogs. Evolution 64, 429-443.

Hubert, N., Calcagno, V., Etienne, R.S., Mouquet, N., 2015a. Metacommunity speciation models and their implication for diversification theory. Ecol. Lett. 18, 864-881.

Hubert, N., Hanner, R., 2015. DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3, 44-58.

Hubert, N., Kadarusman, A., Wibowo, F., Busson, Caruso, D., Sulandari, S., Nafiqoh, N., Rüber, L., Pouyaud, L., Avarre, J.C., Herder, F., Hanner, R., Keith, P., Hadiaty, R.K., 2015b. DNA barcoding Indonesian freshwater fishes: challenges and prospects. DNA Barcodes 3, 144-169.

Hubert, N., Meyer, C., Bruggemann, J.H., Guerin, F., Komeno, R.J.L., Espiau, B., Causse, R., Williams, J.T., Planes, S., 2012. Cryptic diversity in Indo-Pacific coral reef fishes

revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. Plos One 7, e28987.

Hudson, R., 1982. Testing the constant rate neutral allele model with protein sequence data. Evolution 37, 203-217.

Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N., Vellend, M., 2008. Ecological consequences of genetic diversity. Ecol. Lett. 11, 609-623.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H., Hebert, P.D.N., 2007. Universal primers cocktails for fish DNA barcoding. Mol. Ecol. notes 7, 544-548.

Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs,W., Remigio, E., Hebert, P.D.N., 2005.Wedding biodiversity inventory of a large and complex lepidoptera fauna with DNA barcoding. Phil. Trans. R. Soc. Serie B 360, 1835-1845.

Kekkonen, M., Hebert, P.D.N., 2014. DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. Mol. Ecol. Resour. 14, 706-715.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463-1464.

Kingman, J.F.C., 1982. The coalescent. Stoch. Process Appl. 13, 245-248.

Kottelat, M., 1989. Zoogeography of the fishes from Indochinese inland water with annotated checklist. Bull. Zool. Mus. Univ. Van Amst. 12, 1-55.

Kottelat, M., 2013. The fishes of the inland waters of Southeast Asia: a catalog and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. Raffles Bull. Zool. Suppl. 27, 1-663.

Kottelat, M., Whitten, A.J., Kartikasari, S.R., Wirjoatmodjo, S., 1993. Freshwater Fishes of Western Indonesia and Sulawesi. Periplus editions, Singapore.

Lamoreux, J.F., Morrison, J.C., Ricketts, T.H., Olson, D.M., Dinerstein, E., McKnight, M., Shugart, H.H., 2006. Global tests of biodiversity concordance and the importance of endemism. Nature 440, 212-214.

Lohman, K., De Bruyn, M., Page, T., Von Rintelen, K., Hall, R., Ng, P.K.L., Shih, H.-T., Carvalho, G.R., Von Rintelen, T., 2011. Biogeography of the Indo-Australian archipelago. Annual Review Ecolo. Evol. Syst. 42, 205-226.

Markovtsova, L., Marjoram, P., Tavare, S., 2000. The effects of rate variation on ancestral inference in the coalescent. genetics 156, 1427-1436.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune,

A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M., Turelli, M., 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol. Lett. 10, 315-331.

Moilanen, A., Hanski, I., 1998. Metapopulation dynamics: effects of habitat quality and landscape structure. Ecology 79, 2503-2515.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G., Vogler, A.P., 2009. Accelerated secies inventories on Madagascar using coalescent-based models of species delineation. Syst. Biol. 58, 298-311.

Moritz, C., 1994. Defining 'evolutionary significant units' for conservation. Trends Ecol. Evol. 9, 373-375.

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

Ng, H.H., Kottelat, M., 2016. The Glyptothorax of Sundaland: a revisionary study (Teleostei: Sisoridae). Zootaxa 4188, 001-092.

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. Mol. Phylogenet. Evol. 48, 1223-1231.

Nielsen, R., Wakeley, J., 2001. Distinguishing migration from isolation: a markov chain approach. genetics 158, 885-896.

Orr, H.A., Turelli, M., 2001. The evolution of post-zygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55, 1085-1094.

Pearse, D.E., Crandall, K.A., 2004. Beyond FST: analysis of population genetic data for conservation. Conserv. Genet. 5, 585-602.

Pereira, L.H.G., Hanner, R., Foresti, F., Oliveira, C., 2013. Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna. BMC Genet. 14, 20.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595-606.

Pouyaud, L., Sudarto, Paradis, E., 2009. The phylogenetic structure of habitat shift and morphological convergence in Asian Clarias (Teleostei, Siluriformes: Clariidae). J. Zool. Syst. Evol. Res. 47, 344-356.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, automatic barcode gap discovery for primary species delimitation. Mol. Ecol. 21, 1864-1877.

R_Core_Team, 2014. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Ratnasingham, S., Hebert, P.D.N., 2007. Bold: the barcode of life data system. Mol. Ecol. notes 7, 355-364. http://www.barcodinglife.org.

Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS One 8, e66213.

Read, C.I., Bellwood, D.R., Herwerden van, L., 2006. Ancient origins of Indo-Pacific coral reef fish biodiversity: a case study of the Leopard wrasses (Labridae: Macropharyngodon). Mol. Phylogenet. Evol. 38, 808-819.

Ricklefs, R.E., Schluter, D., 1993. Species Diversity in Ecological Communities: Historical and Geographical Perspectives. University of Chicago Press, Chicago.

Riedel, A., Sagata, K., Suhardjono, Y.R., T€anzler, R., Balke, M., 2013. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. Front. Zool. 10, 15.

Rogers, A.R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552-569.

Rousseuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53-65.

Ryder, O.A., 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1, 9-10.

Schloss, P.D., Handelsman, J., 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71, 1501-1506.

Selkoe, K.A., Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9, 615-629.

Slatkin, M., Hudson, R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponantially growing populations. genetics 129, 555-562.

Smith, A.M., Fisher, B.L., Hebert, P.D.N., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Phil. Trans. R. Soc. Serie B 360, 1825-1834.

Smith, M.A., Wood, D.M., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl. Acad. Sci. U. S. A. 104, 4967-4972.

Sork, V.L., Waits, L., 2010. Contributions of landscape geneticseapproaches, insights, and future potential. Mol. Ecol. 19, 3489-3495.

Steinke, D., Zemlak, T.S., Hebert, P.D.N., 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. Plos One 4, e6300.

Strimmer, K., Pybus, O.G., 2001. Exploring the demographic history of DNA sequences using the generalised skyline plot. Mol. Biol. Evol. 18, 2298-2305.

Tajima, F., 1983. Evolutionary relationships of DNA sequences in finite populations. genetics 105, 437-460.

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

T€anzler, R., Sagata, K., Surbakti, S., Balke, M., Riedel, A., 2012. DNA barcoding for community ecology - how to tackle a hyperdiverse, mostly undescribed melanesian fauna. PLoS One 7, e28832.

Urban, M.C., LEibold, M.A., Amarasekare, P., De Meester, L., Gomulkiewicz, R., Hochberg, M.E., Klausmeier, C.A., Loeuille, N., De Mazancourt, C., Norberg, J., Pantel, J.H., Strauss, S.Y., Vellend, M., Wade, M.J., 2008. The evolutionary ecology of metacommunities. Trends Ecol. Evol. 23, 311-317.

Vellend, M., 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. Am. Nat. 166, 199-215.

Vellend, M., 2010. Conceptual synthesis in community ecology. Q. Rev. Biol. 85, 183-206.

Vellend, M., Geber, M.A., 2005. Connections between species diversity and genetic diversity. Ecol. Lett. 8, 767-781.

Vernooy, R., Haribabu, E., Muller, M.R., Vogel, J.H., Hebert, P.D.N., Schindel, D.E., Shimura, J., Singer, G.A.C., 2010. Barcoding life to conserve biological diversity: beyond the taxonomic imperative. PLoS Biol. 8, e1000417.

Vogler, A.P., DeSalle, R., 1994. Diagnosing units of conservation management. Conserv. Biol. 6, 170-178.

Wakeley, J., 2001. The coalescent in an island model of population subdivision with variation among demes. Theor. Popul. Biol. 59, 133-144.

Wakeley, J., 2003. Polymorphism and divergence for island-model species. genetics 163, 411-420.

Waples, R.S., 1991. Pacific salmon, Oncorhynchus spp. & the definition of 'species' under the endagered species act. Mar. Fish. Rev. 53, 11-22.

Waples, R.S., 1995. Evolutionary significant units and the conservation of biological diversity under the endangered species act. In: Nielsen, J.L., Power, G.A. (Eds.), Evolution and the Aquatic Ecosystem: Defining Unique Units in Population Conservation. American Fisheries Society, Bethesda, Maryland, pp. 8-27.

Ward, R.D., Hanner, R.H., Hebert, P.D.N., 2009. The campaign to DNA barcode all fishes, FISH-BOL. J. Fish Biol. 74, 329-356.

Warren, B.H., Simberloff, D., Ricklefs, R.E., Aguilee, R., Condamine, F.L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernandez-Palacios, J.M., Hengl, T., Norder, S.J., Rijsdijk, F., Sanmartin, I., Strasberg, D., Triantis, K.A., Valente, L.M., Whittaker, R.J., Gillespie, R., Emerson, B.C., Thebaud, C., 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecol. Lett. 18 (2), 200-217.

Weir, J.T., Schluter, D., 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315, 1574-1576.

Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C., Tallmon, D.A., 2014. Genetic rescue to the rescue. Trends Ecol. Evol. 30, 42-49.

Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. Trends Ecol. Evol. 19, 639-644.

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

VI. General Discussion

6.1 State of the knowledge on the ichthyodiversity of Java

The results presented here highlight major knowledge gaps in the diversity of the freshwater fishes in Java and Bali. The synthesis on the Java and Bali diversity (Article 1) indicates that 216 and 40 species have been reported in the inland waters of Java and Bali, respectively (Froese & Pauly, 2011; Kottelat *et al*., 1993; Kottelat, 2013). This result contrasts with the 159 species sampled (Article 1). Several reasons have been presented in Articles 1 & 2 that may explain such large discrepancies between historical species records and the result of a modern reappraisal based on a standardized DNA-based inventory. The following trends have been identified that may account for such discrepancies:

(1) Taxonomic biases: the inland fishes reported in Java and Bali correspond to original description that span across two centuries of ichthyological research in Southeast Asia (Fig. 21). The taxonomic knowledge of inland fishes in Indonesia mainly results from three main historical waves of original descriptions that correspond to different political periods of Indonesia. As such, the ichthyological exploration of Indonesian waters has been discontinued through time and conducted by a large diversity of ichtyologists with varying practices of species description. The main consequence is the fragmented taxonomic literature and a knowledge that has been scattered through time in varying scientific publications. Despite the effort made by Kottelat and colleagues (1993) to gather this taxonomic knowledge, the lack of proper description of diagnostic morphological characters in many cases has made the morphological identification of Indonesian freshwater fishes a periously task. The results presented in Article 1 demonstrate that Java and Bali are no exceptions. Non-reproducible species identifications are likely to have led to an inflated number of species reported in Java.

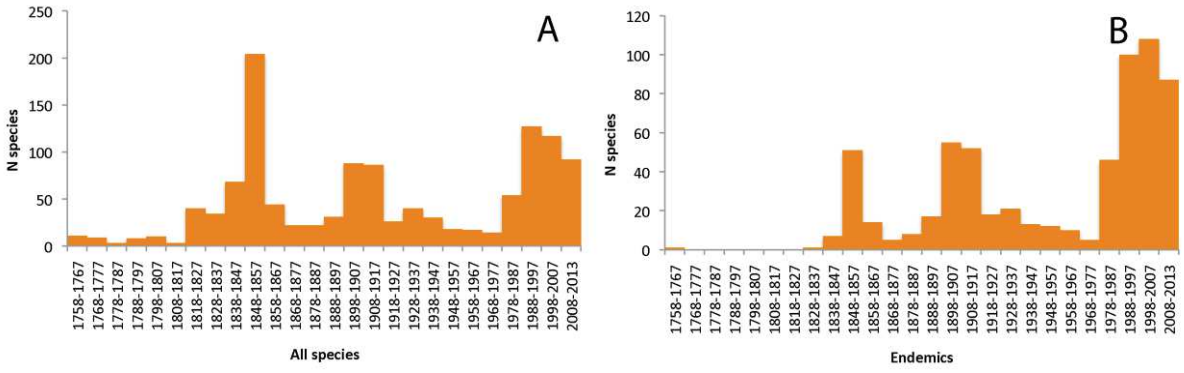


Figure 21. Number of species description per decades since 1758 for the species occurring in Indonesia. A. All species, B. Endemic species (Hubert *et al*., 2015).

Along the same line, the status of several endemic species of Java is questionable considering that among the 16 species endemic of Java, only 3 have been observed (Article 1) while 5 of the 13 unobserved species have been described by a single specimen and correspond to old descriptions (e.g. *Barbonymus platysoma* (Bleeker 1855), *Mystus abbreviates* (Valenciennes 1840), *Ompok javanensis* Hardenberg 1938, *Puntius aphya* (Bleeker 1851) and *Barbonymus bramoides* Valenciennes 1842).

(2) Taxonomic confusion among closely related species: The use of DNA-based methods of inventory and species delimitation algorithm helped resolved several cases of perpetuated taxonomic confusion and re-examine morphological characters (article 2). As evidenced for *Rasbora* spp. (article 2), *R. baliensis* and *R. lateristriata* have been largely confused in Java and Bali since the description of *R. baliensis* Hubbs 1954. A similar situation was also depicted for *Nemacheilus* spp. in Java where species names have been inconsistently applied to the original type specimens. Only a re-examination through DNA sequences, including specimens from the type locality, of the species boundaries and range distribution allowed to detect an inversion of species name and the re-interpretation of the original characters used to discriminate *N. chrysolaimos* and *N. fasciatus* (Fig. 22).

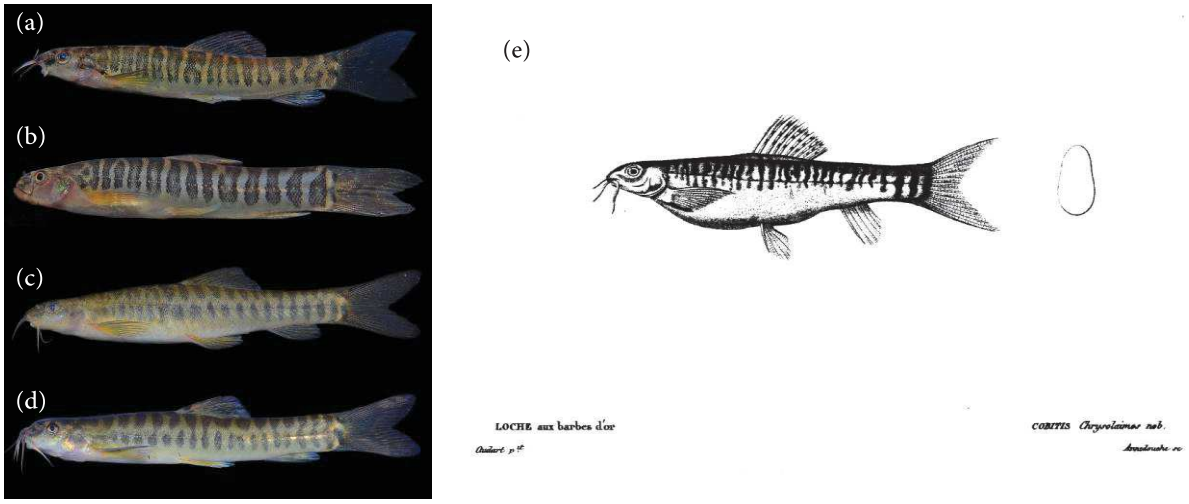


Figure 22. *Nemacheilus chrysolaimos* Valenciennes 1846 and *Nemacheilus fasciatus* Valenciennes 1846 following Article 2 and original illustration of *Nemacheilus chrysolaimos* Valenciennes 1846.

Another major result of the DNA-based inventories depicted in Articles 1, 2 and 3 is the commonness of morphologically similar molecular lineages within species with unsually high genetic distances among them (i.e. cryptic diversity). This pattern questions the actual species richness of the Java ichthyofauna considering that many cryptic lineages diverge by more than 2% of genetic divergence, a threshold that is frequently observed between fish

morphological species (Ward, 2009). This result suggests that several species encompassing multiple highly divergent lineages have been overlooked due to the occurrence of phenotypic bias toward stable morphological attributes. As stated by Hillis *et al.* (1996) morphological characters might be inadequate for species delimitation, particularly when species diverged in allopatry through geographic isolation across similar biotic and abiotic conditions stabilizing eco-morphological features. Along the same line, Bickford *et al.* (2006) stated that one of the most general assumptions about cryptic species is that cryptic diversity results from recent speciation events and a timespan since speciation too short to have left an inprint on morphological characters. The situation, however, is different in Java several cryptic lineages display genetic divergences above 3%, while other species with diagnostic morphological characters display genetic divergences below 3%. Genetic divergence can happen because of a geographic isolation inhibiting gene flow and fostering genetic drift but morphological divergence requires diversifying selection and adaptation, conditions that might not establish when geographic isolation happen in homogeneous landscapes (Bickford *et al*., 2006).

6.2 Local vs. regional dynamics and their influence on species range distribution and populations genetic structure

The results presented in this dissertation (Articles 2 & 3) evidence that species range distribution are more restricted than previously thought and that multiple events of population fragmentation leading to the establishment of highly divergent molecular lineages are detected. For instance, multiple lineages were detected in *Barbodes binotatus*, *Glyptothorax platypogon*, and *Channa gachua* with a MRCA of each species dated in between 2.5 Mya and 0.5 Mya. A similar trend was observed in *Rasbora* with *R*. sp1 and *R. lateristriata* displaying lineages with a divergence dated around 0.5 Mya. All the molecular lineages detected show very restricted range distribution limited to a single watershed or a set of geographically contiguous watersheds with no overlap detected (Articles 2 & 3). This pattern is consistent with the geological history of Java island that isolated from Sumatra and Borneo during the Pleistocene and has been home of an intense volcanic activity due to its position at the forefront of the subduction zone (Lohman *et al*., 2011). By contrast with Borneo that originated through the isolation of a pience of the Asian plate, Java established through the rise of two major volcanic arches, one in the west and one the east. The emergence of the Java island through the merging of two volcanic arches is reflected in the distribution of the molecular lineages detected in *B. binotatus*, *G. platypogon* and *C. gachua* (Article 3) and also in the range distribution of *Rasbora* spp. (Article 2) and associated to the detection of a major

phylogeographic break. This geological context gave multiple opportunities for allopatric divergence across rugged mountainous landscapes that produced many isolated watersheds (De Bruyn *et al*., 2012; April *et al*., 2013). Most of the species associated to high altitude freshwater habitats display higher diversities of molecular lineages within species (e.g. *C. gachua*, *G. platypogon*) than lowland species (e.g. *B. binotatus*) and more restricted range distribution. A result further confirmed by the restricted range distribution observed for *Rasbora* spp. and *Nemacheilus* spp., two genera typical of the hilly freshwater habitats in Java. This pattern of geographic divergence questions the effectiveness of allopatric differentiation in generating reproductive isolation. At first, reproduction isolation appears as consequence of geographic isolation among populations can still interbreed (Nowell *et al*., 2011). Then, the two populations accumulate genetic divergence to the point of establishing post-zygotic isolation and prevent gene flow in case of secondary contact (Rull *et al*., 2013; Seehausen *et al*., 2014; Rajkov *et al*., 2018). The patterns of range distribution, consistent with an allopatric divergence, and genetic divergence, consistent with an ancient fragmentation, thus question the biological status of the multiple molecular lineages detected within species here and their reproductive compatibility.

Bridle *et al*. (2009) stated that complex topography and diversity of habitat foster diversification and the accumulation of species, a statement that seems to be corroborated by the ichthyodiversity in Java Island. Another factor that can explaine the high ichtyodiversity in Java is the sea level fluctuations during the Plestocene (Voris, 2000; Woodruff, 2010). Glacial cycles during the Pleistocene drove sea level fluctuations that resulted in the connection of the Sundaland islands (Gorog *et al*., 2004; Bennett, 2017) and connection of watersheds among island through the establishment of Paleorivers. These large ancient rivers offered opportunities for dispersal among islands of aquatic organisms and likely contributed to the accumulation of species in Java island (Heaney, 1991; Hantoro, 2018). Several cases have been reported in the literature that corroborate this scenario of dispersal. *Barbonymus balleroides*, for instance, is present in Java (Dahruddin *et al*., 2017; Haryono *et al*., 2014), and South Borneo (Mote *et al*., 2014), a pettern matching the boundaries of the East Sunda Paleoriver (Kottelat *et al*., 1993; Roberts, 1993; Weber & Beaufort, 1916). From another line of evidence, the phylogeographic structure of the catfish *Hemibagrus nemurus*, observed here in Java (article 1), is consistant with a structure driven by paleorivers boundaries in Sundaland (Dodson *et al*., 1995; McConnell, 2004). A similar pattern was observed for the cyprinid species *Barbonymus gonionotus*, present in Java (Kottelat & Widjanarti, 2005) but also Sumatra and Borneo, which population structure is also matching the boundaries of the

East Sunda and North Sunda paleorivers (McConnell, 2004). Along the same line, *Tor tambroides*, observed in Java (article 1), Sumatra and Borneo (Nguyen *et al*., 2008), is widely distributed but display low level of genetic variability its range distribution suggesting a recent expansion, a scenario compatible with the time frame of the sea level fluctuations of the Pleistocene (Nguyen *et al*., 2008). The above species seems to confirm that during glacial time, Java, Borneo and Sumatra where connected to the mainland leading to faunal exchanges (McConnell, 2004; Voris, 2000).

Also paleorivers offered opportunities of dispersal among Borneo, Sumatra and Java, and Java geological history offered further opportunities of speciation, Java hosts only a few endemic species compared to Sumatra and Borneo (Table 1) with an endemism around 5% in Java (*vs*. 25% in Sumatra, 46% in Borneo). This results suggests that in-situ diversification happened only in a few cases, mainly for the species associated to hilly habitats as reflected in *Rasbora*, *B. binotatus* and *C. gachua* (Articles 2 & 3). This result suggest that species immigration from elsewhere in Sundaland mainly contributed to the accumulation of species in Java. Likewise, the recent age of Java might explain the low rate of in-situ diversification compared to Borneo, a trend that seems to be corroborated by the young age of the molecular lineages inferred for *B. binotatus*, *C. gachua* and *G. platypogon*, as well as among *Rasbora* species in Java.

6.3 Implications for conservation

Fishes diversity in Indonesia is one of highest in the world and is facing similar threats as other mega-diverse country such as Brasil and Congo Republic due to human population growth and fast development of urban areas and associated infrastructures. This context of habitat desctruction is further amplified by the importance of inland fisheries as a source of protein for local populations, particularly so in Java and Sumatra. The threats are extremely high in Java island due to the high population density, the highest in Sundaland, that induced the emergence of multiple stressors on freshwater habitats including the conversion of landscapes for agriculture (e.g. culture of rice in terraces), the construction of dams for energy supply and the lack of appropriate treatment of wasted waters from urban areas. Its has been evidenced for instance that large Dams disrupted gene flow for some species such as *Barbonymus balleroides* in Serayu river (Bahiyah *et al*., 2013). Anthropogenic perturbations, however, had more dramatic consequences as exemplified by the DNA barcoding survey conducted in Java (Article 1), where several large freshwater fish species targeted by artisanal fisheries haven't been observed. This trend is further confirmed

by previous ichthyological surveys conducted in the Ciliwung river, the watershed encompassing the largest urban areas in Java (Jakarta). In 1910, 187 native species were reported from the Ciliwung river but only 20 species were observed in 2010 (Hadiaty, 2011). Along the same line, 135 species were observed in the Cisadane river (Bogor) in 1910 and only 33 species, including 6 introduced species, were observed in 2009 (Hadiaty, 2011). Compared to 1910 records, the loss of diversity is estimated to reach 92.5% and 75.6%, respectively in Ciliwung and Cisadane river (Hadiaty, 2011). Introduction of exotic species is also a concern. Based the national database of Museum Zoologicum Bogoriense, Research Center for Biology, Indonesian Institute of Siences, 1 exotic species was observed in the Ciliwung and Cisadane rivers in 1970 but 5 species in the Ciliwung river and 6 species in Cisadane river were observed in 1990. This result was confirmed by the DNA barcoding survey (Article 1) that led to the detection of 20 exotic species across Java and originating from varying continent such as Africa, South and Central America.

The conservation status of the Java ichthyofauna is of particular concern if considering the small range distribution observed here for several species and molecular lineages (Articles 2 & 3). DNA barcoding evidenced several cases of range distribution that are actually much more restricted than previously thoughts (Article 2). *Nemacheilus chrysolaimos* for instances has been previously reported as occurring in Java at large (Kottelat *et al*., 1993) while the results of Article 2 show a much more restricted range distribution limited to a few watershed in the western tip of Java. The same situation has been observed for *Rasbora aprotaenia*, another endemic species restricted to the western tip of Java (Article 3). The case of *Rasbora lateristriata* is more intriguing as its type localities are located in west Java but the species was not observed throughout the western half of Java. Considering that only a few specimens of *Rasbora* were captures in Western Java and given the worrying environmental conditions of the watersheds in this part of Java (i.e. Cisadane and Ciliwung rivers), extirpations of *R. lateristriata* in the western part of Java cannot be discarded. A similar trend is observed for the intra-specific molecular lineages observed here in *Barbodes binotatus*, *Glyptothorax platypogon* and *Channa gachua* that show extremely restricted range distribution compared to the global range distribution of each species. As discussed in Articles 2 & 3, the restricted range distribution of most of the Java freshwater fish species and lineages is of concern from a conservation perspective because their persistence is supported by small and geographically restricted set of populations. In this context, the further reduction of their effective population size might have dramatic consequences on their survival due to increased demographic stochasticity for instance.

Restoration programs based on captive breeding might be a solution, however, the commonness of cryptic diversity and the deep genetic divergence observed among intra-specific molecular lineages argue against translocation programs among watersheds (Articles 2 & 3).

VII. Conclusions and Perspectives

The present dissertation highlights the effectiveness of DNA barcoding in capturing species boundaries and also its usefulness for the exploration of intra-specific phylogeographic patterns. High levels of cryptic diversity are detected through DNA barcoding for the most widely distributed species in Java and Bali. These results indicate that DNA barcoding can be used for specimen identifications to the species level and the assignement of unknown sequences to known taxa. The high diversity of cryptic lineages observed for several species and their spatial segregation further suggests that DNA barcoding can also be applied to trace the geographic origin of specimens for several species. Theses results open new perspectives for the monitoring of the Java and Bali ichthyodiversity by enabling automated molecular-based identifications and further applications such as meta-barcoding and/or environmental DNA (eDNA). Considering the alarming erosion of the Java ichthyodiversity, the present DNA barcode reference libraries will likely constitute a usefull tool in the future with applications as diverse as the detection and tracing of biological invasions or the temporal monitoring of species assemblages through eDNA. These libraries also enable specimen identifications to the species level by non-specialists, a benefit for a large community of the academic sector and governmental agencies in charge of the management of the Java ichthyological ressources.

From a basic perspective, the high diversity of cryptic lineages detected here questions the biological status of several species and call for a thorough assessment of their diversity through more comprehensive genomic approaches. Due to its maternal inheritance, mitochondrial DNA provide incomplete information about gene flow and the genomic consequences of fragmentation for instance. Several potential cases of secondary contacts have been highlighted that question the impact of geographic isolation on the reproductive compatibility of allopatric lineages and zones of secondary contact constitute excellent models to explore the genomic consequences of genetic drift on reproductive isolation.

VIII. Bibliography

April, J., Hanner, R.H., Dion-Cote, A-M., Bernatchez, L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology* 22: 409–422

Bahiyah., D.D.Solihin., R. Affandi. 2013. Variasi Genetik Populasi Ikan Brek (Barbonymus balleroides Val. 1842) Sebagai Dampak Fragmentasi Habitat di Sungai Serayu. *Jurnal Iktiologi Indonesia*. 13 (2): 175-186.

Beck, SV, GR Carvalho, Axel Barlow, Lukas Ru, Heok Hui Tan, A Muchlisin, and Mark De Bruyn. 2017. "Plio-Pleistocene Phylogeography of the Southeast Asian Blue Panchax Killifish, Aplocheilus Panchax," 1–17. https://doi.org/10.5061/dryad.5j501

Bennett, K D. 2017. "Milankovitch Cycles and Their Effects on Species in Ecological and Evolutionary Time" 16 (1): 11–21.

Bickford, David, David J Lohman, Navjot S Sodhi, Peter K L Ng, Rudolf Meier, Kevin Winker, Krista K Ingram, and Indraneil Das. 2006. "Cryptic Species as a Window on Diversity and Conservation" 22 (3). https://doi.org/10.1016/j.tree.2006.11.004

Bird, Michael I, David Taylor, and Chris Hunt. 2005. "Palaeoenvironments of Insular Southeast Asia during the Last Glacial Period : A Savanna Corridor in Sundaland ?" 24: 2228–42. https://doi.org/10.1016/j.quascirev.2005.04.004.

Bridle JR, Butlin RK, Schulter D. 2009. *Speciation and Patterns of Diversity*. Cambridge (GB). Cambridge University Press.

Broecker, Wallace S, and George H Denton. 1989. "What Drives Glacial Cycles ?"

Casiraghi, Maurizio, Massimo Labra, Emanuele Ferri, Andrea Galimberti, and Fabrizio De Mattia. 2010. "DNA Barcoding : A Six-Question Tour to Improve Users ' Awareness about the Method" 11 (4): 440–53. https://doi.org/10.1093/bib/bbq003.

Charlesworth, Brian, Carolina Bartolomé, and Véronique NoëL. 2005. "The Detection of Shared and Ancestral Polymorphisms." *Genetical Research* 86 (2): 149–57. https://doi.org/10.1017/S0016672305007743.

Conte-Grand, C., R. Britz, N. Dahanukar, R. Raghavan, R. Pethiyagoda, H. H. Tan, R. K. Hadiaty, N. S. Yaakob, and L. Rüber. 2017. Barcoding snakeheads (Teleostei, Channidae) revisited: Discovering greater species diversity and resolving perpetuated taxonomic confusions. PLoS One 12:e0184017. Public Library of Science.

Dahruddin, Hadi, Aditya Hutama, Frédéric Busson, Sopian Sauri, Robert Hanner, Philippe Keith, Renny Hadiaty, and Nicolas Hubert. 2017. "Revisiting the Ichthyodiversity of Java and Bali through DNA Barcodes: Taxonomic Coverage, Identification Accuracy, Cryptic Diversity and Identification of Exotic Species." *Molecular Ecology Resources*.

https://doi.org/10.1111/1755-0998.12528.

Darajati W., S. Pratiwi, E. Herwinda, A.D. Radiansyah, V.S. Nalang, B. Nooryanto, J.S. Rahajoe, R. Ubaidillah, I. Maryanto, R. Kurniawan *et al*. 2016. *Indonesian Biodiversity Strategy and Action Plan 2015-2020*. The National Development Planning Agency, Jakarta.

De Bruyn, M., B. Stelbrink, R.J. Morley, R. Hall, G.R. Carvalho, C.H. Channon, G. Van Den Bergh, *et al*. 2014. "Borneo and Indochina Are Major Evolutionary Hotspots for Southeast Asian Biodiversity" 63 (6): 879–901. https://doi.org/10.1093/sysbio/syu047.

De Bruyn, M., L. Rüber, S. Nylinder, B. Stelbrink, N. R. Lovejoy, S. Lavoué, T. Heok Hui, E. Nugroho, D. Wowor, P. K. L. Ng, M. N. Siti Azizah, T. Von Rintelen, R. Hall, and G. R. Carvalho. 2013. Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast asian biodiversity hotspots. Syst. Biol. 62:398–410.

De Bruyn, M., Von Rintelen, T., Von Rintelen, K., Mather, P.B., Carvalho, G.R. 2012. Molecular biogeography and phylogeography of the freshwater fauna of the Indo-Australian Archipelago in *Biotic Evolution and Environmental Change in Southeast Asia*, eds D. J. Gower *et al*. Cambridge University Press.

Dodson, JJ, Colombani, F, Ng, PKL. 1995. Phylogeographic structure in mitochondrial DNA of a South-east Asian freshwater fish, *Hemibagrus nemurus* (Sluroidei; Bagridae) and Pleistocene sea-level changes on the Sunda shelf. *Molecular Ecology*, 4, 331– 346.

Endra, Wahyu, Sahat Ratmuangkhwang, and Yoshinori Kumazawa. 2016. "Molecular Phylogenetics and Evolution Molecular Phylogeny and Historical Biogeography of the Indonesian Freshwater Fish Rasbora Lateristriata Species Complex (Actinopterygii : Cyprinidae): Cryptic Species and West-to-East Divergences." *Molecular Phylogenetics and Evolution* 105: 212–23. https://doi.org/10.1016/j.ympev.2016.08.014.

Farhana, Samsudin Nurul, Zainal Abidin Muchlisin, Thuy Yen Duong, and Suwat Tanyaros. 2018. "Exploring Hidden Diversity in Southeast Asia ' s *Dermogenys* Spp. (Beloniformes : Zenarchopteridae) through DNA Barcoding." *Scientific Reports*, no. January: 1–11. https://doi.org/10.1038/s41598-018-29049-7.

Froese R, Pauly D. 2011. FishBase. World Wide Web electronic publication. http://www.fishb ase.org

Fu, YX., Li, WH. 1999. Coalescing into the 21st Century: An Overview and Prospects of Coalescent Theory. *Theoretical Population Biology* 56(1): 1-10

Fujisawa, T., Barraclough, TG. 2013. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on

simulated data sets. *Syst Biol* 62:707–24

Giarla TC, Maher SP, Achmadi AS, Moore MK, Swanson MT, Rowe KC, Esselstyn JA. 2018. Isolation by marine barriers and climate explain areas of endemism in an island rodent. *J Biogeogr*. 45(9): 2053-206

Gorog, Antonia J, Martua H Sinaga, and Mark D Engstrom. 2004. "Vicariance or Dispersal ? Historical Biogeography of Three Sunda Shelf Murine Rodents (*Maxomys Surifer* , *Leopoldamys Sabanus* and *Maxomys Whiteheadi*)," 91–109.

Hadiaty, RK. 2011. Diversitas dan hilangnya jenis-jenis ikan di sungai Ciliwung dan Sungai Cisadane. *Berita* Biologi 10(4):491-504.

Hall, R. 2009. "Southeast Asia's Changing Palaeogeography." *Blumea: Journal of Plant Taxonomy and Plant Geography* 54 (1–3): 148–61. https://doi.org/10.3767/000651909X475941.

Hall, R. 1996. "Reconstructing Cenozoic SE Asia." *Geological Society, London, Special Publications* 106 (1): 153–84. https://doi.org/10.1144/GSL.SP.1996.106.01.11.

Hanebuth, T J J, K Stattegger, and A Bojanowski. 2009. "Termination of the Last Glacial Maximum Sea-Level Lowstand : The Sunda-Shelf Data Revisited." *Global and Planetary Change* 66 (1–2): 76–84. https://doi.org/10.1016/j.gloplacha.2008.03.011.

Hanebuth, Till J.J., Harold K. Voris, Yusuke Yokoyama, Yoshiki Saito, and Jun'ichi Okuno. 2011. "Formation and Fate of Sedimentary Depocentres on Southeast Asia's Sunda Shelf over the Past Sea-Level Cycle and Biogeographic Implications." *Earth-Science Reviews* 104 (1–3): 92–110. https://doi.org/10.1016/j.earscirev.2010.09.006.

Hantoro, Wahyoe Soepri. 2018. "Sunda Epicontinental Shelf and Quaternary Glacial-Interglacial Sea Level Variation and Their Implications to the Regional and Global Environmental Change." *IOP Conference Series: Earth and Environmental Science* 118 (1). https://doi.org/10.1088/1755-1315/118/1/012053.

Haryono, MF. Rahardjo, Mulyadi & R. Affandi. 2014. Komunitas ikan di perairan Sungai Serayu yang terfragmentasi waduk di wilayah Kabupaten Banjarnegara. *Zoo Indonesia*. 23(1): 35-43

Heaney, LR. 1986. Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. *Biological Journal of the Linnean Society* 28: 127–165

Heaney, LR. 1991. "A Synopsis of Climatic and Vegetational Change in Southeast Asia." *Climatic Change* 19 (1–2): 53–61. https://doi.org/10.1007/BF00142213.

Hebert, PDN, Alina Cywinska, Shelley L Ball, and R Jeremy. 2003. "Biological Identifications through DNA Barcodes," no. September 2002: 313–21.

https://doi.org/10.1098/rspb.2002.2218.

Hillis DM., Moris C., Mable. 1996. *Molecular Systematics*. Sinauer Associetes Inc. Senderland, Massachusets. USA.

Hubert, Nicolas, Erwan Delrieu-trottin, Jean-olivier Irisson, Christopher Meyer, and Serge Planes. 2010. "Molecular Phylogenetics and Evolution Identifying Coral Reef Fish Larvae through DNA Barcoding : A Test Case with the Families Acanthuridae and Holocentridae" 55: 1195–1203. https://doi.org/10.1016/j.ympev.2010.02.023.

Hubert, N., and R. Hanner. 2015. DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3:44–58.

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

Hubert, Nicolas, Arif Wibowo, Frédéric Busson, Domenico Caruso, Sri Sulandari, Nuna Nafiqoh, Laurent Pouyaud, *et al*. 2015. "DNA Barcoding Indonesian Freshwater Fishes : Challenges and Prospects," 144–69. https://doi.org/10.1515/dna-2015-0018.

Huybers, Peter. 2006. "Early Pleistocene Glacial Cycles and the Integrated Summer Insolation Forcing." *Science* 313.

Kapli, P., S. Lutteropp, J. Zhang, K. Kobert, P. Pavlidis, A. Stamatakis, and T. Flouri. 2017. "Multi-Rate Poisson Tree Processes for Single-Locus Species Delimitation under Maximum Likelihood and Markov Chain Monte Carlo." *Bioinformatics* 33 (11): 1630–38. https://doi.org/10.1093/bioinformatics/btx025.

Keith, P., C. Lord, and K. Maeda. 2015. Indo-Pacific Sicydiinae Gobies - Biodiversity, life traits and conservation. Société Française d'Ichtyologie.

Kottelat M. 2013. The fishes of the inland waters of Southeast Asia: a catalog and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries, *Raffles Bull. Zool. Supplement* 27, 1-663.

Ko, H.L., Y.T. Wang, T.S. Chiu, M.A. Lee, M.Y. Leus, K.J. Chang, W.Y. Chen & K.T. Shao 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *Plos One* 8 (1): e53451. doi:10.1371/journal.pone.0053451.

Kottelat, M., A. J. Whitten, S. R. Kartikasari, and S. Wirjoatmodjo. 1993. *Freshwater fishes of western indonesia and sulawesi*. Periplus editions, Singapore.

Kottelat, M. & E. Widjanarti. 2005. The fishes of Danau Sentarum National Park and the Kapuas lakes area, Kalimantan Barat, Indonesia. *The Raffles Bulletin of Zoology Supplement*. 13: 139-173.

Lim, Hongchiun, Muchlisin Zainal Abidin, and Chaidir Parlindungan Pulungan. 2016. "DNA Barcoding Reveals High Cryptic Diversity of the Freshwater Halfbeak Genus *Hemirhamphodon* from Sundaland," 1–17. https://doi.org/10.1371/journal.pone.0163596.

Lohman K, De Bruyn M, Page T, Von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR, Von Rintelen T. 2011. Biogeography of the Indo-Australian archipelago. Annu Rev Ecol Evol Syst 42:205–226

Lundberg, JG. 1998. The Temporal Context for the Diversification of Neotropical Fishes. P. in L. R. Malabarba Reis, R. E., Vari, R. P., Lucena, Z. M., Lucena, C. A. S., ed. Phylogeny and Classification of Neotropical Fishes. Edipucrs, Porto Alegre.

MacArthur, R.H., Wilson, E.O. 1967. *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Matz, M.V., Nielsen, R., 2005. A likelihood ratio test for species membership based on DNA sequence data. *Philos. Trans. R. Soc. B* 360: 1969–1974.

McConnell, SKJ., 2004. Mapping faunal exchanges across the Sunda shelf, South-East Asia, using distributional and genetic data sets from the cyprinid fish *Barbodes gonionotus* (Bleeker, 1850). *Journal of Natural History*, 38: 651-670.

Metcalfe, I. 2009. "Late Palaeozoic and Mesozoic Tectonic and Palaeogeographical Evolution of SE Asia." *Geological Society, London, Special Publications* 315 (1): 7–23. https://doi.org/10.1144/SP315.2.

Milton M, Pierossi P, Ratnasingham S. 2013. *Barcode of Life Data Systems Handbook*. Ontario (CND): Biodiversity Institute of Ontario.

Mora, C., Derek P Tittensor, Sina Adl, Alastair G B Simpson, and Boris Worm. 2011. "How Many Species Are There on Earth and in the Ocean ?" 9 (8): 1–8. https://doi.org/10.1371/journal.pbio.1001127.

Mote, N., R. Affandi dan Haryono. 2014. Biologi reproduksi ikan brek (Barbonymus balleroides Cuvier & Val. 1842) di Sungai Serayu zona atas dan bawah Waduk Panglima Besar Soedirman, Jawa Tengah. *Jurnal Iktiologi Indonesia*. Vol 14 (2): 111-122.

Morley, CG. 2000. "Origin and Evolution of Tropical Rain Forests." *Geologicql Evolution of Tropical Rainforests* 15 (6): 362. https://doi.org/10.2307/3515620.

Myers, N., Russell A Mittermeier, Cristina G Mittermeier, Gustavo A B Fonseca, and Jennifer Kent. 2000. "Biodiversity Hotspots for Conservation Priorities" 403 (February): 853–58.

Nguyen, TTT, Uthairat Na-nakorn, Srijanya Sukmanomon, and Chen Ziming. 2008. “Short Communication A Study on Phylogeny and Biogeography of Mahseer Species (Pisces : Cyprinidae) Using Sequences of Three Mitochondrial DNA Gene Regions” 48: 1223–31. https://doi.org/10.1016/j.ympev.2008.01.006.

Nowell, Reuben W., Brian Charlesworth, and Penelope R. Haddrill. 2011. “Ancestral Polymorphisms in Drosophila Pseudoobscura and Drosophila Miranda.” *Genetics Research* 93 (4): 255–63. https://doi.org/10.1017/S0016672311000206.

O’Connell, Kyle A., Eric N. Smith, Kyle J. Shaney, Umilaela Arifin, Nia Kurniawan, Irvan Sidik, and Matthew K. Fujita. 2018. “Coalescent Species Delimitation of a Sumatran Parachuting Frog.” *Zoologica Scripta* 47 (1): 33–43. https://doi.org/10.1111/zsc.12248.

Pouyaud, L, and E Paradis. 2009. “The Phylogenetic Structure of Habitat Shift and Morphological Convergence in Asian Clarias (Teleostei , Siluriformes : Clariidae)” 1 (August 2008): 344–56. https://doi.org/10.1111/j.1439-0469.2008.00507.x.

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz. 2012. “ABGD, Automatic Barcode Gap Discovery for Primary Species Delimitation.” *Molecular Ecology* 21 (8): 1864–77. https://doi.org/10.1111/j.1365-294X.2011.05239.x.

Rabosky, D. L., Glor, R. E. 2010. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proceedings of the National Academy of Sciences of the United States of America* 107: 22178– 22183

Rajkov, J., Weber A.A-T., Salzburger, W., Egger, B. 2018. Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes. *Evolution* 72(11): 2553–2564

Ratnasingham, S., and P. D. N. Hebert. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoSONE 8:e66213.

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

Rosenberg, N. A., Nordborg, M. 2002. Genealogical trees, coalescent theory, and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3: 380-390

Rull, J., Abraham, S., Kovaleski, A., Segura, D.F., Mendoza, M., Liendo, C.M., Vera, M.T. 2013. Evolution of pre-zygotic and post-zygotic barriers to gene flow among three cryptic species within the Anastrepha fraterculus complex. *Entomologia Experimentalis et Applicata* 148: 213–222

Sato, Miyuki, and Ken Sato. 2013. “Maternal Inheritance of Mitochondrial DNA by Diverse Mechanisms to Eliminate Paternal Mitochondrial DNA.” *Biochimica et Biophysica Acta*

- *Molecular Cell Research* 1833 (8): 1979–84. https://doi.org/10.1016/j.bbamcr.2013.03.010.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, L.C., Saetre, G-P. et al. 2014. Genomics and the origin of species. *Nature Reviews Genetics* 15(3): 176–192.

Voris, HK. 2000. Maps of Pleistocene Sea Levels in Southeast Asia : Shorelines , River Systems and Time Durations. *Journal of Biogeography* 27: 1153–1167.

Ward, RD. 2009. "DNA Barcode Divergence among Species and Genera of Birds and Fishes," 1077–85. https://doi.org/10.1111/j.1755-0998.2009.02541.x.

Weber, M. & L. F. de Beaufort, 1916. The Fishes of the Indo- Australian Archipelago. III. Ostariophysi: II Cyprinoidea, Apodes, Synbranchi. E. J. Brill, Leiden. xv+455 pp.

Wilting, Andreas, Rahel Sollmann, Erik Meijaard, Kristofer M. Helgen, and Jörns Fickel. 2012. "Mentawai's Endemic, Relictual Fauna: Is It Evidence for Pleistocene Extinctions on Sumatra?" *Journal of Biogeography* 39 (9): 1608–20. https://doi.org/10.1111/j.1365-2699.2012.02717.x.

Whitmore TC. 1981. *Wallace's Line and Plate Tectonics*. Oxford, UK: Clarendon Press. 200 pp.

Woodruff, David S. 2010. "Biogeography and Conservation in Southeast Asia: How 2.7 Million Years of Repeated Environmental Fluctuations Affect Today's Patterns and the Future of the Remaining Refugial-Phase Biodiversity." *Biodiversity and Conservation* 19 (4): 919–41. https://doi.org/10.1007/s10531-010-9783-3.