Vol. 46: 205–214, 2021
https://doi.org/10.3354/esr01152

ENDANGERED SPECIES RESEARCH
Endang Species Res

Published December 2

# Pleistocene paleodrainages explain the phylogeographic structure of Malaysian populations of Asian arowana better than their chromatic variation

**Norli Fauzani Mohd Abu Hassan Alshari[1,2], Sébastien Lavoué[1,*], Mohamad Asyraf Mohamad Sulaiman[3], Md Zain Khaironizam[1], Siti Azizah Mohd Nor[2], Fazimah Aziz[3]**

[1]School of Biological Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia
[2]Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus 21030, Malaysia
[3]Department of Aquatic Science, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

ABSTRACT: Little is known about the genetic structure and phylogeography of Asian arowana (*Scleropages* spp.). Natural variation in body color has led to the informal distinction of chromatic varieties, but previous studies that attempted to genetically characterize these varieties did not comprehensively cover their geographical distribution. In Malaysia, about 10 drainage-restricted populations of Asian arowana are known that are currently classified into 2 species and 3 color varieties. In this study, we used 3 molecular markers to test 2 hypotheses explaining the relationships among 9 of these populations. The first hypothesis postulates that each color variety forms a monophyletic group, whereas the second hypothesis assumes that Pleistocene paleodrainages shaped the distribution of these populations. We found that the overall genetic variability is low within Asian arowana and that the green variety is non-monophyletic, with other varieties nested within. Instead, the populations of Malaysia belong to 3 genetic lineages that are allopatrically distributed. The ages and distribution of 2 of these lineages are consistent with past connections through paleodrainages, whereas the last lineage is restricted to Central Sarawak. Overall, our results reject the first hypothesis, demonstrating that the geographic origin of specimens is a better phylogenetic indicator than their body color. This study highlights the importance of Malaysia in the conservation of Asian arowana, because it is the only country in which populations of all 3 main genetic lineages occur.

KEY WORDS: *Scleropages formosus* · *Scleropages inscriptus* · Sundaland · Phylogeography · Conservation

## 1. INTRODUCTION

*Scleropages formosus* and *S. inscriptus* (Teleostei; Osteoglossidae), collectively known as Asian arowana, are among the most endangered freshwater fish in Southeast Asia because of the inexorable destruction of their natural habitats in Sundaland and south Indochina coupled with overharvesting to supply the aquarium trade market (Ng & Tan 1997, Larson & Vidthayanon 2019). Asian arowana have already been extirpated from several regions, and the remaining small populations persist only in a

*Corresponding author: microceb@hotmail.com



Publisher: Inter-Research · www.int-res.com

few lentic or slow-moving lotic habitats scattered throughout Malaysia (Peninsular Malaysia and Sarawak), Indonesia (Sumatra and Kalimantan, i.e. the Bornean region of Indonesia), south Myanmar and south Cambodia (Fig. 1). *S. formosus* is listed as Endangered (A2cd+4cd) in the IUCN Red List of Threatened Species (Larson & Vidthayanon 2019), whereas *S. inscriptus* is only assessed as Data Deficient (Vidthayanon 2019).

As expensive aquarium fish (Voight 2016), Asian arowana have been the object of significant genetic research to characterize commercial strains and improve their farming production (reviewed by Yue et al. 2020). In comparison, little is known about the genetic structure and phylogeography of natural populations of these fish and the underlying factors that shape their extant distribution.

Asian arowana are variable in body color (discussed by van Oijen & van der Meij 2013), but the different forms, their natural variations and distributions, and their putative functions have not been thoroughly investigated. Two possible reasons explain this situation: (1) museum specimens lose most of their chromatic pigments once they are preserved, and (2) Asian arowana are now rare in their natural habitats. For commercial purposes, the aquarium trade broadly recognizes 3 main varieties within *S. formosus* (i.e. 'red', 'green' and 'gold') and 2 within *S. inscriptus* (i.e. 'green batik' and 'green nami'). One question that arises is whether the color varieties within *S. formosus* represent valid species. Pouyaud et al. (2003) claimed they are different species based on morphological and genetic evidence, leading these authors to describe 3 new species. Kottelat & Widjanarti (2005) invalidated these new species, pointing out some flaws in the study by Pouyaud et al. (2003). Van Oijen & van der Meij (2013) also commented on Pouyaud et al. (2003). More recent genetic studies on *S. formosus* attempted to characterize the red, green and gold varieties, using single-locus molecular markers (Yue et al. 2004, Mohd-Shamsudin et al. 2011, Mu et al. 2012) or by sequencing their complete genomes (Austin et al. 2015, Bian et al. 2016), but these studies neglected the possible geographical variation.

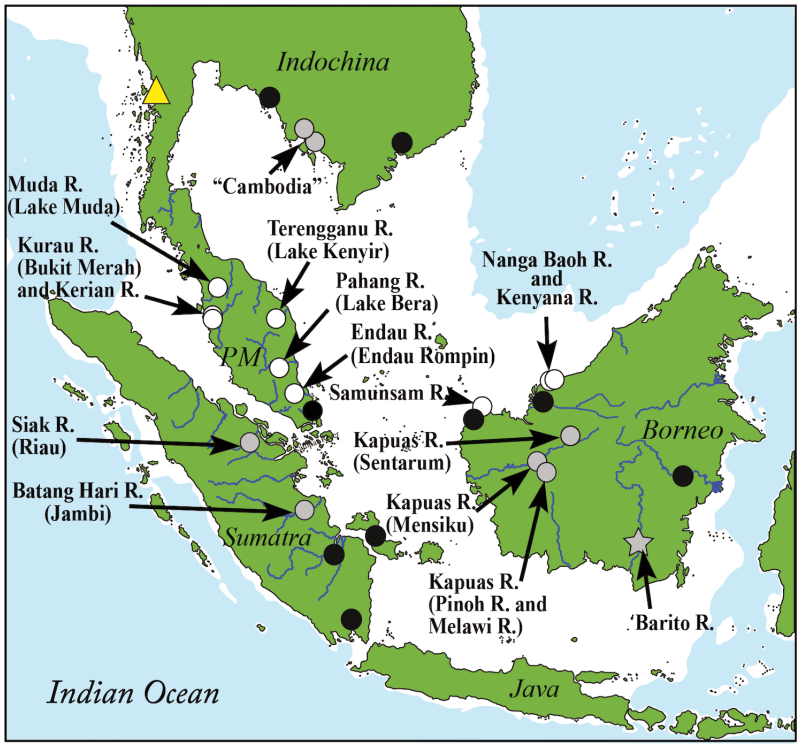


Fig. 1. Sundaland region of Southeast Asia, showing the localities where specimens of Asian arowana (*Scleropages* spp.) originated: black dots indicate historical records, white dots indicate Malaysian populations sampled for this study, and grey dots represent additional populations for which specimens were examined in this study. The grey star is the approximate geographical position of the type locality of *S. formosus* (see van Oijen & van der Meij 2013), and the yellow triangle shows the type locality of *S. inscriptus*. Additional details of the localities and examined specimens are provided in Table 1. The white area represents the current submerged part (to a maximum depth of 120 m) of the continental Sunda shelf. PM: Peninsular Malaysia

The Sundaland, where most of the natural populations of Asian arowana occur, is a biogeographical region of Southeast Asia, which includes the Malay Peninsula, the large islands of Borneo, Sumatra and Java, as well as a multitude of smaller islands scattered on the shallow immerged continental shelf (Fig. 1). Until about 400 thousand years ago (ka), the Sunda Shelf was fully exposed and the freshwater drainages of Borneo and Sumatra were connected with those of the Malay Peninsula and the Indochina mainland through 4 ancient rivers: the extended Chao Phraya River, Malacca Straits River, East Sunda River and North Sunda River (Fig. 2) (reviewed by Sathiamurthy & Voris 2006). At that time, the fish fauna of each of these ancient rivers was likely homogeneous. From 400 ka, the combined actions of Sundaland's subsidence and large marine regressions repeatedly modified the landscape of this region (Haq et al. 1987, Husson et al. 2020).

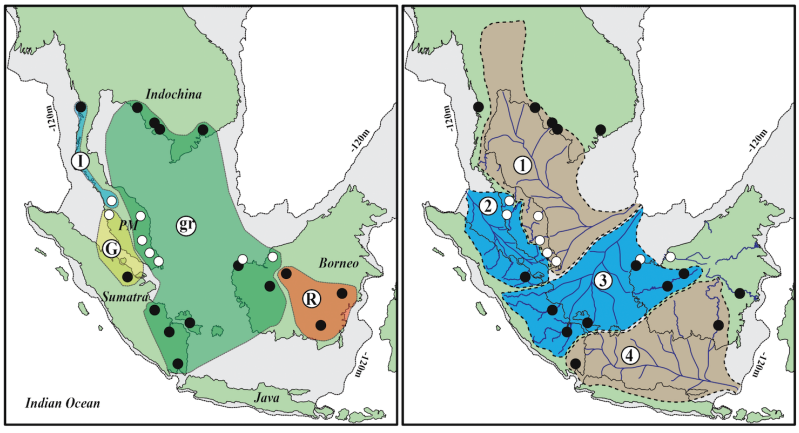


Fig. 2. The 2 hypotheses tested in this study. Black dots indicate historical records, and white dots indicate Malaysian populations sampled for this study (see Fig. 1). (A) The 'color-based species' hypothesis postulates that each of the 4 main color varieties forms a valid species of *Scleropages*: red (R), gold (G), green (gr) and green nami/batik (I) (the latter is currently recognized as *S. inscriptus*). (B) The 'paleodrainage-based' hypothesis postulates that the distribution of Asian arowana is the result of the Pleistocene sea level fluctuations which connected and disconnected different parts of the Sundaland through a network of 4 main paleodrainages (1: extended Chao Phraya River; 2: Malacca Straits River; 3: North Sunda River; 4: East Sunda River) (adapted from Sathiamurthy & Voris 2006). Assumptions of these hypotheses are discussed in Section 1

On several occasions, the Sunda shelf was completely immerged, splitting these ancient rivers into the present rivers and isolating their populations. Such a scenario of river connection–disconnection has driven the evolutionary differentiation of several species of fish through dispersal–vicariance mechanisms (de Bruyn et al. 2013, Sholihah et al. 2021), but not all species (see Beck et al. 2017).

About 10 drainage-restricted populations of Asian arowana are known from Peninsular Malaysia and Sarawak, with each population occupying only a restricted section of its drainage (Fig. 1). The populations in Sarawak and the east coast of Peninsular Malaysia are classified as green, whereas 1 of the 2 populations along the west coast of Peninsular Malaysia is classified as gold and the other one belongs to *S. inscriptus* (variety green nami). Previous genetic studies examined only some Peninsular Malaysian populations (i.e. those of east coast Peninsular Malaysia, Tasek Bera and Endau Rompin – green; and west coast Peninsular Malaysia, Bukit Merah – gold) (Mohd-Shamsudin et al. 2011), demonstrating that the green and gold varieties are genetically differentiated. As far as we know, other Malaysian populations, such as that of *S. inscriptus* and those of Sarawak, have not been studied.

Here we examined the phylogeography of almost all known populations of Asian arowana in Malaysia using 3 molecular markers (the mitochondrial cytochrome *c* oxidase I [*COI*] and cytochrome *b* [*cytb*] genes and the nuclear recombination activating 1 [*rag1*] gene). The objective was to test 2 hypotheses explaining the phylogeographic structure of Asian arowana (Fig. 2). The 'color-based' hypothesis postulates that each main color variety forms a species regardless of its distribution. The alternative 'paleodrainage-based' hypothesis assumes that Pleistocene paleodrainages shaped the distribution of Asian arowana. Predictions can be made under each of these 2 hypotheses. Under the first hypothesis, it is predicted that each color variety forms a monophyletic group and that these groups should genetically diverge from each other by at least 2% using *COI*. Under the second hypothesis, populations will form monophyletic groups according to the paleoriver network to which they belong, and the ages of each of these paleodrainage-based monophyletic groups should be younger than about 400 ka, which was when the Sunda shelf was periodically exposed allowing freshwater organisms to disperse. If these monophyletic groups are older than 400 ka, it means that another paleo-mechanism (pre-dating the periodic exposure of the Sunda shelf through eustatic sea level variation) shaped the distribution and differentiation of Asian arowana.

## 2. MATERIALS AND METHODS

### 2.1. Taxonomic and character sampling

All molecular studies aiming to examine the phylogeography of Asian arowana, including the present investigation, face difficulties in ensuring sound taxonomic sampling because these fish are rare in the field and the geographical origins of commercial or captive specimens are difficult to trace. In this study, as much as it was possible, we reused published data from known-origin Malaysian specimens, including 1 specimen from Tasek Bera (Pahang River, east coast Peninsular Malaysia [PM]) and 1 specimen from Bukit Merah (Kurau River, west coast PM) (Mohd-Shamsudin et al. 2011, Rahman et al. 2010). To complete our taxonomic sampling of Malaysian populations, we collected specimens from Lake Muda

(Muda River, west coast PM; 1 specimen), Lake Kenyir (Terengganu River, east coast PM; 1 specimen), Endau Rompin (Endau River, east coast PM; 1 specimen) and 1 cultivated specimen assumed to have descended from wild parents captured in the Kerian River (west coast PM). In addition, we examined 2 specimens from the Samunsam River (Southwest Sarawak) and 4 specimens from Central Sarawak, i.e. 2 from the Nanga Baoh River and 2 from Kenyana River (Fig. 1, Table 1). From each specimen, we clipped a piece of the caudal fin that we immediately preserved in 96% ethanol. Most of the specimens were then released. To broaden the comparative analysis of this study, we included the *cytb* dataset of Pouyaud et al. (2003) used by these authors to describe their new species along with *cytb* and *rag1* sequences available in GenBank from 6 specimens: 4 from Indonesia, 1 from Cambodia and 1 from PM (Table 1).

We analyzed 3 molecular markers, the barcode fragment of the *COI* gene (651 bp), the complete *cytb* gene (1141 bp) and the partial nuclear *rag1* gene (1230 bp).

### 2.2. Genomic DNA extraction, PCR and sequencing

Detailed molecular methods are described in Text S1 in the Supplement at www.int-res.com/articles/suppl/n046p205_supp.pdf. Briefly, for each sample, genomic DNA was first extracted from ethanol-fixed

Table 1. List of 27 specimens of Asian arowana examined in this study along with their corresponding GenBank (GB) accession numbers for the mitochondrial cytochrome *c* oxidase I (*COI*) and cytochrome *b* (*cytb*) genes, and the nuclear recombination activating 1 gene (*rag1*). Eleven specimens from Malaysia were newly sequenced for this study (accession numbers are highlighted in **bold**). Eight Indonesian specimens with wild localities were selected from Pouyaud et al. (2003) (specimen code 'Pyd'), along with 2 Malaysian specimens, one from Rahman et al. (2010) and the other from Mohd-Shamsudin et al. (2011). To complete our taxonomic sampling, unpublished sequences of cytochrome *b* and recombination activating 1 from 6 specimens were mined from GB: 4 from Indonesia, 1 from Peninsular Malaysia (PM) and 1 from Cambodia. Imprecise or tentative specimen localities are emphasized with asterisks. R.: river; Kalim.: Kalimantan (Indonesian part of Borneo). (–) No corresponding sequence available

| Specimen code | Origin | Color var. | *COI* | *cytb* | *rag1* |
|---|---|---|---|---|---|
| S5 | Samunsam R., Sarawak | Green | **MW590648** | **MW596605** | **MW596616** |
| S6 | Samunsam R., Sarawak | Green | **MW590649** | **MW596606** | **MW596617** |
| N2 | Nanga Baoh R., Sarawak | Green | **MW590650** | **MW596607** | **MW596618** |
| N3 | Nanga Baoh R., Sarawak | Green | **MW590651** | **MW596608** | **MW596619** |
| K1 | Kenyana R., Sarawak | Green | **MW590652** | **MW596609** | **MW596620** |
| K3 | Kenyana R., Sarawak | Green | **MW590653** | **MW596610** | **MW596621** |
| Sc059 | Endau Rompin, East PM | Green | **MW590654** | **MW596611** | **MW596622** |
| KEN01 | Lake Kenyir, East PM | Green | **MW590655** | **MW596612** | **MW596623** |
| 2sc | Kerian R., West PM* | Gold | **MW590656** | **MW596613** | **MW596624** |
| 7sc 'inscriptus' | Lake Muda, West PM | Green nami | **MW590657** | **MW596614** | **MW596625** |
| CAM02 | Cambodia* | Green | **MW590658** | **MW596615** | **MW596626** |
| TB79 | Tasik Bera, East PM | Green | JF946666[1] | DQ864663[1] | – |
| BlueGold3 | Bukit Merah, West PM | Gold | HM156429[2] | HM156454[2] | – |
| Pyd1 'legendrei' | West Kalimantan* | Red | – | Not in GB[3] | – |
| Pyd2 'aureus' | Siak R., Sumatra* | Gold | – | Not in GB[3] | – |
| Pyd3 'macrocephalus' | Melawi R., West Kalim. | Red | – | Not in GB[3] | – |
| Pyd4 'macrocephalus' | Barito R., Central Kalim. | Red | – | Not in GB[3] | – |
| Pyd6 | Sumatra* | Green | – | Not in GB[3] | – |
| Pyd7 | Melawi R., West Kalim. | Green | – | Not in GB[3] | – |
| Pyd8 | Pinoh R., West Kalim. | Green | – | Not in GB[3] | – |
| Pyd9 | Central Kalim.* | Green | – | Not in GB[3] | – |
| MAL01 | East PM* | Green | – | FJ887091[4] | FJ887608[4] |
| JAM01 | Jambi, Sumatra | Green | – | FJ887071[4] | FJ887588[4] |
| MJK01 | Mensiku, West Kalim. | Green | – | FJ887110[4] | FJ887627[4] |
| KAG01 | Kalimantan* | Green | – | FJ887081[4] | FJ887598[4] |
| IG01 | Sumatra* | Gold | – | FJ887061[4] | FJ887578[4] |
| CAM01 | Cambodia* | Green | – | FJ887053[4] | FJ887570[4] |

[1]From Rahman et al. (2010); [2]From Mohd-Shamsudin et al. (2011); [3]Partial, 310 bp long *cytb* sequences extracted from Table 2 in Pouyaud et al. (2003); [4]Unpublished sequences mined from GB

fin clips, then the 3 nucleotide markers were amplified by PCR with specific primer pairs (listed in Table S1), and PCR products were purified and sequenced by First BASE Laboratories Sdn. Bhd. (Selangor, Malaysia) using Sanger Sequencing. Chromatograms were edited using MEGA X (Stecher et al. 2020). All new sequences obtained in this study have been submitted to GenBank (accession numbers provided in Table 1).

### 2.3. Analytical comparison

Alignment of each gene was done separately by eye. This resulted in alignments of 651 (for 13 individuals), 1141 (for 27 individuals) and 1230 (for 17 individuals) nucleotide positions for *COI*, *cytb* and *rag1*, respectively. Missing data were coded with '–' and treated as uninformative characters. We conducted maximum-likelihood (ML) phylogenetic analyses using 3 datasets which allowed us to compare the quantity and quality of the phylogenetic signal and the resulting topology: (1) the mitochondrial (mt) dataset which combined *COI* and *cytb* genes (i.e. 1792 nucleotide positions) for all 27 specimens regardless of their completeness for *COI*; (2) the nuclear (nuc) *rag1* dataset comprising 17 specimens; and (3) the complete dataset combining all genes (i.e. 3022 nucleotide positions) and all 27 specimens (with or without *COI* and *rag1* sequences).

To reconstruct the ML trees for each dataset, we used the phylogenetic reconstruction program RAxML-NG (Kozlov et al. 2019) as implemented in the graphical interface raxmlGUI 2.0 (Edler et al. 2021). Datasets were partitioned according to codon position and, for the complete dataset, gene origin (i.e. mt versus nuc). The best model of sequence evolution for each partition was selected by using ModelTest-NG (Darriba et al. 2020) implemented in raxmlGUI and using the Bayesian information criterion. To evaluate the robustness of the internal branches of the ML tree, standard non-parametric bootstrap proportions (1000 replicates) were calculated for each data set. Each phylogenetic inference was rooted at its midpoint because the sister group of Asian arowana (i.e. the 2 Australian species of *Scleropages*) is too distantly related to provide a reliable root position as the groups diverged from each other more than 60 million years ago (Lavoué 2015). Uncorrected pairwise genetic distances (i.e. *p*-distances) among and within main lineages revealed by the phylogenetic analyses were calculated with MEGA X (Stecher et al. 2020).

We then estimated a time-calibrated coalescent-based lineage-tree using *BEAST v.2.6.5 (Star-BEAST) (Heled & Drummond 2010) as implemented in the package BEAST v.2.6.5 (Bouckaert et al. 2019). For that, specimens for which *COI*, *cytb* and *rag1* sequences were available were assigned to the 3 lineages identified in the ML inference. Substitution and clock models were unlinked across all loci. We linked the mitochondrial *cytb* and *COI* tree models. A lognormal relaxed-clock model was applied to each gene under a Yule tree prior and a linear with constant root population size model. To calibrate the dating analyses in Star-BEAST, the rate of nucleotide substitution (= molecular clock rate) was set for *COI* to 0.012 substitutions per site per million years, corresponding to the mean estimation from 19 trans-Panamanian Isthmus pairs of marine fishes (Bermingham et al. 1997). Two BEAST analyses of the Star-BEAST dataset were independently run for 100 000 000 generations each (parameters and trees sampled every 10 000 generations). We used Tracer v.1.7 (Rambaut et al. 2018) to check for convergence of all analytical parameters, ensuring that all effective sample sizes were >200, and to select burn-in cycles (25%). The 2 separate runs were combined using Logcombiner v2.6.5, and a maximum clade credibility time tree was generated in TreeAnnotator v.2.6.5. Figtree v1.4.4 (https://github.com/rambaut/figtree/releases) and Densitree v.2.2.7 (Bouckaert & Heled preprint doi:10.1101/012401) were used to visualize the trees.

## 3. RESULTS

### 3.1. Gene trees and time-calibrated lineage tree reconstruction

We determined new *COI*, *cytb* and *rag1* sequences of 10 specimens of *Scleropages formosus* and 1 specimen of *S. inscriptus* (Table 1). Our mitochondrial (mt) dataset, combining *COI* and *cytb* genes, comprises 27 specimens and 1792 bp (about 25% of missing data). Out of a total of 1792 (651 + 1141) positions, 83 positions are variable and 63 are phylogenetically informative.

Fig. 3A shows the ML tree inferred from the mt dataset which is rooted at its mid-point. In this tree, 3 main lineages are identifiable, each supported by a bootstrap proportion >89%. The first and largest lineage (Lineage 1) comprises all specimens from west PM, southwest Sarawak and Indonesia (Kalimantan and Sumatra). Lineage 2 comprises only the specimens from central Sarawak (i.e. from Nanga Baoh

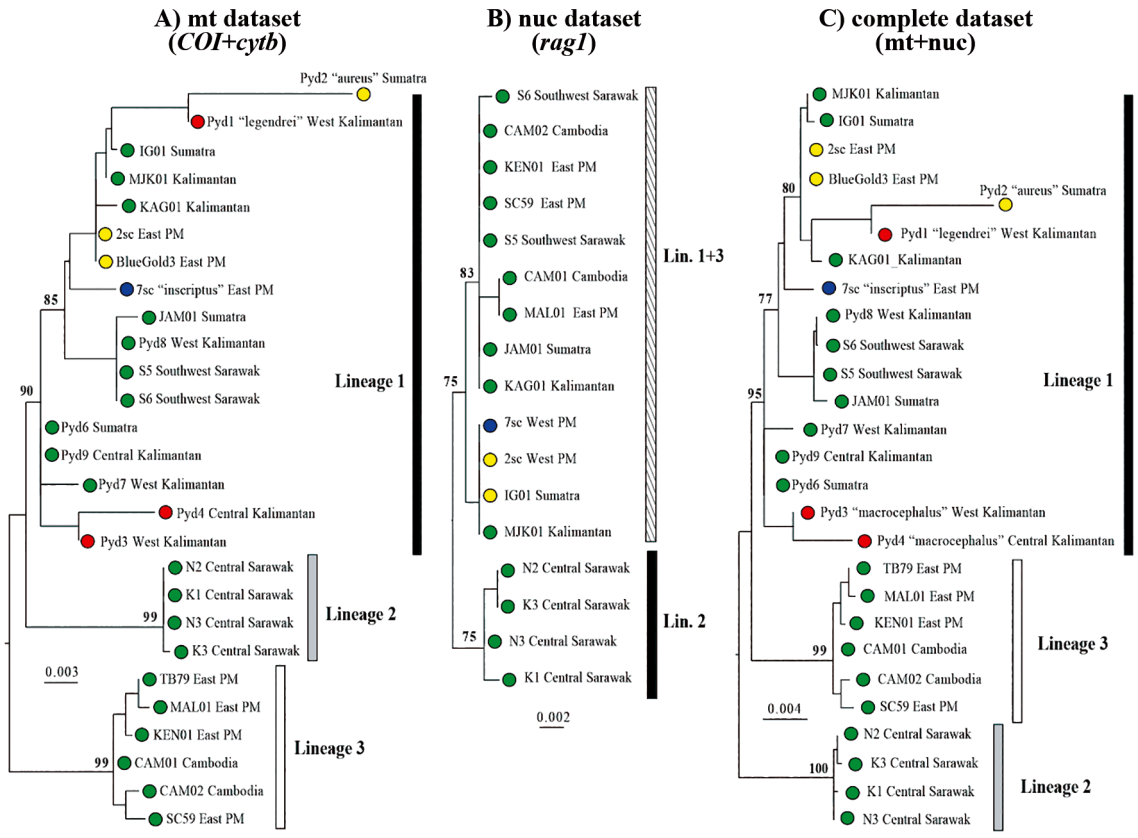


Fig. 3. Maximum likelihood (ML) phylogenetic trees based on the analyses of: (A) the mitochondrial (mt) dataset (combining *COI* and *cytb* genes; 27 specimens, 1792 characters), (B) the nuclear (nuc) dataset (*rag1* gene; 17 specimens, 1230 characters) and (C) the complete dataset combining all 3 genes (*COI*, *cytb* and *rag1* genes; 27 specimens, 3022 characters). ML trees are rooted at their midpoint, branch lengths are proportional to the number of substitutions, and numbers at branches indicate bootstrap proportions (if >75%). Yellow, green, red and blue dots indicate the specimen color varieties: gold, green, red and green nami, respectively

River and Kenyana River; Fig. 1), while Lineage 3 comprises all specimens from east PM and Cambodia. Lineage 1 comprises sequences of all species of Asian arowana described in the 21st century (i.e. *S. aureus*, *S. legendrei*, *S. macrocephalus* and *S. inscriptus*) (Pouyaud et al. 2003, Roberts 2012). Because our taxonomic sampling is quantitatively limited, we verified that it covers the known genetic variability by comparing the *COI* and *cytb* sequences used in our study against all sequences available in GenBank from cultivated specimens of unknown origin (see Figs. S1 & S2).

Out of a total of 1230 positions in the nuclear (nuc) *rag1* dataset, 9 are variable and only 7 are phylogenetically informative. The topology of the ML tree (Fig. 3B) was only partially resolved because of this lack of variability. However, the specimens from Central Sarawak (Lineage 2) form a monophyletic group (bootstrap proportion [BP] = 75%), distinct from the rest of the specimens (Lineages 1 and 3).

The ML phylogenetic tree using the complete dataset (Fig. 3C) is fully resolved and topologically congruent with that of the *mt* dataset in recovering the same 3 main lineages with similar bootstrap proportions.

The time-calibrated Bayesian lineage tree reconstruction is shown in Fig. 4 for a dataset including all 11 specimens for which *COI*, *cytb* and *rag1* sequences were available. In this tree, Lineage 1 is the sister group of Lineage 3 (as in the nuc and mt+nuc ML trees; Fig. 3B,C) but with a posterior probability support of only 0.59. The alternative hypothesis suggests a sister relationship between Lineages 1 and 2 (as in the mt ML tree). Using the mitochondrial *COI* substitution rate of 0.012 substitutions per site per million years, as estimated by Bermingham et al. (1997), the age of the crown group of Asian arowana is estimated to be 420 ka (95% credibility interval [CI] = 691–167 ka) whereas the divergence time between Lineages 1 and 3 is estimated to be 283 ka [CI = 518–70 ka].

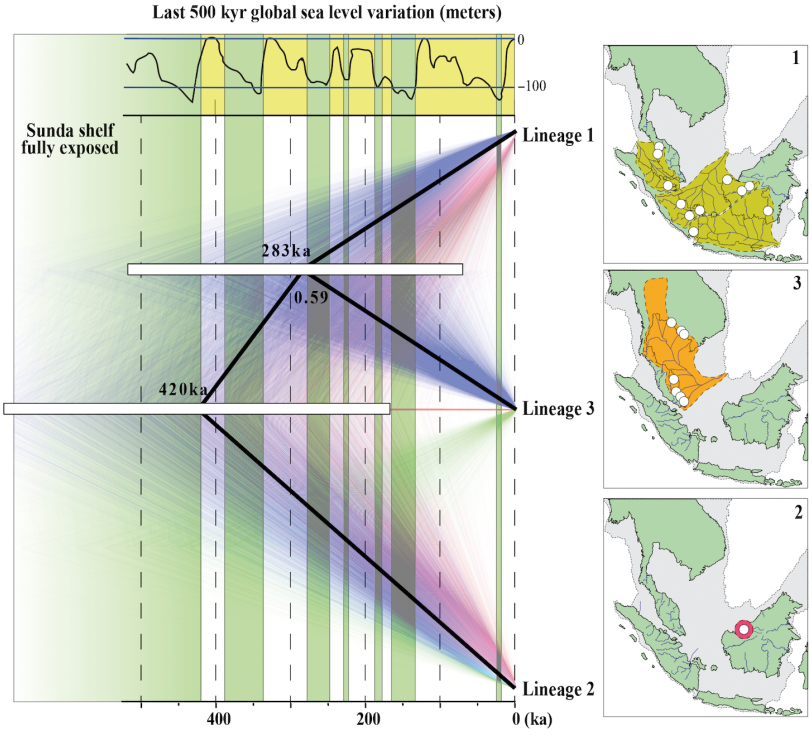


Fig. 4. Time-calibrated Bayesian lineage tree of Asian arowana (genus *Scleropages*) using Star-BEAST in BEAST v.2.6.5 and presented as a maximum clade credibility tree (shown with black thick branches). Node bars give the 95% credibility interval (CI) limits of the node heights. The 18 000 sampled trees are shown superimposed using DensiTree. Numbers above nodes are their mean divergence time estimates. The number below the node supporting Lineage 1 + Lineage 3 is its posterior probability (0.59). Horizontal timescale is in thousands of years ago (ka). Estimation of the global sea level variation of the last 500 thousand years (kyr) modified from Spratt & Lisiecki (2016) and Berends et al. (2021). Periods of full exposure of the Sunda shelf as hypothesized by Husson et al. (2020) are indicated with green columns. Right inset maps indicate the distribution of each lineage of Asian arowana (olive, orange and red for Lineages 1, 3 and 2, respectively); each white dot represents a collection locality

### 3.2. Genetic distances

Mean uncorrected pairwise genetic distances (*p*-distances) were calculated separately for *COI*, *cytb* and *rag1*, among and within the 3 lineages previously evidenced. Among these 3 lineages, mean *COI*, *cytb* and *rag1* genetic distances are between 1.3 and 1.5%, 2.1 and 2.8%, and 0.1 and 0.3%, respectively (Table 2). Within each of these lineages, genetic distances are always below 1% (Table 2). Lineage 1 exhibits the largest within-lineage genetic variability (mean *COI* and *cytb* *p*-distances of 0.6 and 0.7%, respectively); it is also the most widely distributed (Kalimantan, Sumatra, southwest Sarawak, West PM and likely Myanmar). Furthermore, this lineage contains all red, gold and green nami specimens, along with some green specimens.

Table 2. Mean uncorrected pairwise distances (*p*-distances, in %) between (regular text) and within (*italics*) the 3 lineages of Asian arowana for each gene (*COI*, *cytb* and *rag1*; see Table 1) as estimated in MEGA X

| | Lineage 1 | Lineage 2 | Lineage 3 |
|---|---|---|---|
| Lineage 1 | | | |
| *COI* | *0.6* | | |
| *cytb* | *0.7* | | |
| *rag1* | *0.0* | | |
| Lineage 2 | | | |
| *COI* | 1.5 | *0.1* | |
| *cytb* | 2.1 | *0.0* | |
| *rag1* | 0.3 | *0.0* | |
| Lineage 3 | | | |
| *COI* | 1.5 | 1.3 | *0.2* |
| *cytb* | 2.2 | 2.8 | *0.5* |
| *rag1* | 0.1 | 0.2 | *0.0* |

### 3.3. Hypotheses testing results

Our phylogenetic results reject the 'color-based' species hypothesis because neither the green nor the gold variety is monophyletic. The red and gold specimens, along with the specimen of *S. inscriptus*, are deeply nested within the green specimens (Fig. 3C). Furthermore, a gold specimen is more closely related to a red specimen than to the other gold specimens. On the other hand, our results show a better fit with the 'paleodrainage-based' hypothesis for explaining the distribution of Malaysian populations of Asian arowana in recovering an east coast PM–Cambodia monophyletic group and a clade comprising specimens from the west coast PM, Sumatra, Kalimantan and the extreme southwest part of Sarawak. The relationships within this latter clade are mostly unresolved, as no finer structure is distinguishable. The inferred divergence times among lineages (<500 ka) are also compatible with the occurrence of the late Pleistocene ancient river network (Fig. 2).

## 4. DISCUSSION

Our study on the phylogenetic structure of Malaysian populations of *Scle-*

*ropages formosus* and *S. inscriptus* reveal 3 main lineages which do not follow the color-based delineation. These results have implications for taxonomy, phylogeography and conservation.

### 4.1. Genetic diversity and taxonomy

To date, 5 species of Asian arowana have been formally described (Müller & Schlegel 1840–1845, Pouyaud et al. 2003, Roberts 2012). Only 2 of these species were considered valid prior to the present study, namely *S. formosus* and *S. inscriptus* (Kottelat 2013, Fricke et al. 2021).

In contrast to recent molecular studies which focused on the distinctiveness of the main color varieties of *S. formosus* but without looking at possible geographical variability within each color variety (Mohd-Shamsudin et al. 2011), we considered geographical variability by including specimens originating from as many localities as possible covering most of the natural distribution of *S. formosus*, and by including *S. inscriptus*.

Our genetic results show a low, but spatially structured, level of diversity within Asian arowana with the recognition of 3 lineages. The inter-lineage genetic distances are below the standard inter-specific threshold. These results support the hypothesis that adult color-body variation is a poor taxonomic criterion and refute the conclusions of 2 previous studies recognizing more than 1 species within *S. formosus* based on such color patterns (Pouyaud et al. 2003, Roberts 2012). All specimens of *S. aureus*, *S. macrocephalus*, *S. legendrei* and *S. inscriptus* belong to only 1 lineage along with specimens of *S. formosus* from or near its type locality, and they diverge from each other by less than 1% (*COI*-based). The current evidence led to the conclusion that these species are not valid and are all junior synonyms of *S. formosus* (Kottelat & Widjanarti 2005). Further research should investigate morphological variation among the 3 main genetic lineages evidenced herein, taking into consideration that Asian arowana are Critically Endangered, and any imprudent taxonomical change will negatively impact their conservation. Our results also show that the 5 complete whole genomes already determined for this species (Austin et al. 2015, Bian et al. 2016) only capture a fraction of the total genetic diversity, as all 5 specimens belong to Lineage 1.

### 4.2. Phylogeography

Past overseas connections among the currently separated land masses that are the Malay Peninsula, Sumatra, Borneo and Java, through large ancient river systems exposed during Pleistocene sea level regressions, have congruently shaped the distribution of many freshwater organisms (e.g. Woodruff 2010, de Bruyn et al. 2013, Sholihah et al. 2021). As a result, paleodrainage system-based predictions can be made to explain the distribution of populations. We found that the distributions of Malaysian populations of Asian arowana are generally congruent with the configuration of these paleodrainages. The extended Chao Phraya paleodrainage explains the close relationship between Cambodian and east coast PM populations (Lineage 3) ,whereas the Malacca and North Sunda drainages could explain the close relationships among the populations from West PM, Sumatra and Borneo (including the Southwest Sarawak populations but excluding the Central Sarawak population). These populations (along with those of Central Kalimantan) form a monophyletic group (Lineage 1). However, we note that our taxonomic sampling of Indonesian populations is too incomplete to precisely distinguish individual drainage patterns within this clade (Lineage 1). More variable molecular markers along with a better taxonomic coverage of the populations from Indonesia are needed to disentangle the complete phylogeography of Lineage 1.

### 4.3. Conservation

Asian arowana are now very rare in the field as their natural habitats are inexorably shrinking. In addition, the continuous demand for wild specimens to supply the aquarium trade has contributed to the extinction of some populations and put others on the edge of extinction, despite the fact that this species is locally protected (although not in PM) and its international trade is strictly regulated.

Our results have important consequences for the conservation of Asian arowana in Malaysia, as they (1) identify the main evolutionary lineages needing protection, (2) inform about possible restocking plans and (3) provide molecular resources to track the origin of traded specimens. Our results identify Malaysia as the country harboring the largest genetic diversity of this species in having populations from each of 3 main genetic lineages. Each of these lineages faces different threats in Malaysia. Populations on the east side of PM (Lineage 3) are likely the least threatened because (1) they are the most distributed, (2) some of them are found within protected areas, and (3) they belong to the commercially less valuable variety (i.e. green) and are therefore less exploited

compared to the other varieties. On the other hand, populations of Lineages 1 and 2 are highly threatened. Populations in Central Sarawak (Lineage 2) occupy only a small area where intense habitat conversions occur. Populations of Lineage 1 include those on the west coast of PM (i.e. Lake Muda and Bukit Merah) along with the population in the extreme southwest of Sarawak (i.e. Samunsam region). The *in situ* population of Bukit Merah (Kurau River) has likely been extinct since the early 21st century, and the situation of the population of Lake Muda is alarming, with only very scarce recent observations. The populations of Samunsam may be better protected, as they are located within a protected area. To preserve the genetic diversity of Asian arowana in Malaysia, conservation actions targeting each of these 3 lineages must be immediately implemented or reinforced.

## 5. CONCLUSIONS

The diversification of Asian arowana occurred at the end of the Pleistocene (likely within the last 500 kyr), yet the inter-population genetic structure is strong with the recognition of 3 main lineages which do not correspond to color varieties. Color-based variability appears to be a poor taxonomic criterion, and of the 5 species of Asian arowana previously described, only 1 is valid, i.e. *S. formosus*, which comprises all 3 lineages. The late Pleistocene paleodrainages hypothesis better explains the distribution of Asian arowana than color-based variation, but more variable genomic markers and a better taxonomic coverage of the Indonesian populations are needed to fully understand to which extent this hypothesis explains its distribution. Malaysia comprises a large proportion of the genetic diversity of Asian arowana, and additional conservation actions are immediately needed to preserve this diversity. To this end, this study provides molecular resources useful for tracing the origin of Malaysian specimens of Asian arowana and developing non-invasive molecular methods for monitoring natural populations.

*Acknowledgements*. Research permits to collect specimens of *Scleropages formosus* in Sarawak were granted by the Sarawak Forestry Department under number NCCD.907.4.4 (JLD.13)-279. This work was funded by the Malaysian Ministry of Higher Education (FRGS/1/2020/STG03/UMT/01/1 to S.A.M.N.), the Universiti Sains Malaysia (short term 304/PBIOLOGI/6315400 to S.L.) and UNIMAS (Internal Grant RAGS E14099 F07 69 1320/2015 to F.A.). We thank the following persons for their support and help throughout this project: Lincoln Chua, Chawal Supasansanee, Amirrudin Ahmad, Jamsari Amirul Firdaus Jamaluddin, Siti Zafirah binti Ghazali and Noor Zahidah Ahmad Sobri. This study is part of the PhD thesis of N.F.M.A.H.A.

## LITERATURE CITED

Austin CM, Tan MH, Croft LJ, Hammer MP, Gan HM (2015) Whole genome sequencing of the Asian arowana (*Scleropages formosus*) provides insights into the evolution of ray-finned fishes. Genome Biol Evol 7:2885–2895

Beck SV, Carvalho GR, Barlow A, Rüber L and others (2017) Plio-Pleistocene phylogeography of the Southeast Asian blue panchax killifish, *Aplocheilus panchax*. PLOS ONE 12:e0179557

Berends CJ, de Boer B, van de Wal RSW (2021) Reconstructing the evolution of ice sheets, sea level, and atmospheric $CO_2$ during the past 3.6 million years. Clim Past 17:361–377

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. Academic Press, San Diego, CA, p 113–128

Bian C, Hu Y, Ravi V, Kuznetsova IS and others (2016) The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. Sci Rep 6:24501

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S and others (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLOS Comput Biol 15:e1006650

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2020) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol 37:291–294

de Bruyn M, Rüber L, Nylinder S, Stelbrink B and others (2013) Paleo-drainage basin connectivity predicts evolutionary relationships across three southeast Asian biodiversity hotspots. Syst Biol 62:398–410

Edler D, Klein J, Antonelli A, Silvestro D (2021) raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol Evol 12:373–377

Fricke R, Eschmeyer WN, Van der Laan R (2021) Eschmeyer's catalog of fishes. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (accessed 3 January 2021)

Haq BU, Hardenbol J, Vail PR (1987) Fluctuating sea levels since the Triassic. Science 235:1156–1167

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570–580

Husson L, Boucher FC, Sarr AC, Sepulchre P, Cahyarini SY (2020) Evidence of Sundaland's subsidence requires revisiting its biogeography. J Biogeogr 47:843–853

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

Kottelat M, Widjanarti E (2005) The fishes of Danau Sentarum National Park and the Kapuas Lakes area, Kalimantan Barat, Indonesia. Raffles Bull Zool Suppl 13: 139–173

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAxML-NG: a fast, scalable and user-friendly tool

for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455

Larson H, Vidthayanon C (2019) *Scleropages formosus*. The IUCN Red List of Threatened Species 2019: e.T152320185A89797267. doi.org/10.2305/IUCN.UK.2019-3.RLTS.T152320185A89797267.en

Lavoué S (2015) Testing a time hypothesis in the biogeography of the arowana genus *Scleropages* (Osteoglossidae). J Biogeogr 42:2427–2439

Mohd-Shamsudin MI, Fard MZ, Mather PB, Suleiman Z, Hassan R, Othman RY, Bhassu S (2011) Molecular characterization of relatedness among colour variants of Asian arowana (*Scleropages formosus*). Gene 490:47–53

Mu X, Song H, Wang X, Yang Y and others (2012) Genetic variability of the Asian arowana, *Scleropages formosus*, based on mitochondrial DNA genes. Biochem Syst Ecol 44:141–148

Müller S, Schlegel H (1840–1845) Beschrijving van een nieuwen zoetwatervisch van Borneo, *Osteoglossum formosum*. Verhandelingen over de natuurlijke geschiedenis der Nederlandsche overzeesche bezittingen, Leiden

Ng PKL, Tan HH (1997) Freshwater fishes of Southeast Asia: potential for the aquarium fish trade and conservation issues. Aquar Sci Conserv 1:79–90

Pouyaud L, Sudarto, Teugels GG (2003) The different colour varieties of the Asian arowana *Scleropages formosus* (Osteoglossidae) are distinct species: morphologic and genetic evidences. Cybium 27:287–305

Rahman S, Zakaria-Ismail M, Tang PY, Rizman-Idid M, Muniandy S (2010) Genetic variability and estimation of effective population sizes of the natural populations of green arowana, *Scleropages formosus* in Peninsular Malaysia. Adv Environ Biol 4:14–23

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904

Roberts TR (2012) *Scleropages inscriptus*, a new fish species from the Tananthayi or Tenasserim River basin, Malay Peninsula of Myanmar (Osteoglossidae: Osteoglossiformes). Aqua Int J Ichthyol 18:113–118

Sathiamurthy E, Voris HK (2006) Maps of Holocene sea level transgression and submerged lakes on the Sunda Shelf. Nat Hist J Chulalongkorn Univ Suppl 2:1–43

Sholihah A, Delrieu-Trottin E, Condamine FL, Wowor D and others (2021) Impact of Pleistocene eustatic fluctuations on evolutionary dynamics in Southeast Asian biodiversity hotspots. Syst Biol 70:940–960

Spratt RM, Lisiecki LE (2016) A late Pleistocene sea level stack. Clim Past 12:1079–1092

Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol 37:1237–1239

van Oijen MJP, van der Meij SET (2013) The types of *Osteoglossum formosum* Müller & Schlegel, 1840 (Teleostei, Osteoglossidae). Zootaxa 3722:361–371

Vidthayanon C (2019) Myanmar arowana. *Scleropages inscriptus*. The IUCN Red List of Threatened Species 2019: e.T90330470A152320337. doi.org/10.2305/IUCN.UK.2019-3.RLTS.T90330470A152320337.en

Voight E (2016) The dragon behind the glass: a true story of power, obsession, and the world's most coveted fish. Scribner, New York, NY

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 refugial-phase biodiversity. Biodivers Conserv 19:919–941

Yue GH, Li Y, Lim LC, Orban L (2004) Monitoring the genetic diversity of three Asian arowana (*Scleropages formosus*) captive stocks using AFLP and microsatellites. Aquaculture 237:89–102

Yue GH, Chang A, Yuzer A, Suwanto A (2020) Current knowledge on the biology and aquaculture of the Endangered Asian arowana. Rev Fish Sci Aquacult 28:193–210

*Editorial responsibility: Constance M. O'Connor, Toronto, Ontario, Canada*
*Reviewed by: 2 anonymous referees*

*Submitted: February 13, 2021*
*Accepted: August 11, 2021*
*Proofs received from author(s): November 19, 2021*