Biological Journal of the Linnean Society, 2020, **130**, 520–532. With 5 figures.

DNA-based species delimitation reveals cryptic and incipient species in synchronous flashing fireflies (Coleoptera: Lampyridae) of Southeast Asia

WAN F. A. JUSOH[1], LESLEY BALLANTYNE[2] and KIN ONN CHAN[1,*]

[1]*Lee Kong Chian Natural History Museum, Faculty of Science, National University of Singapore, 2 Conservatory Drive, Singapore 117377*
[2]*School of Agricultural and Wine Sciences, Charles Sturt University, Wagga Wagga 2678, Australia*

Received 26 February 2020; revised 1 May 2020; accepted for publication 5 May 2020

Synchronous flashing fireflies of the genus *Pteroptyx* are ubiquitous throughout Southeast Asia, yet fundamental knowledge about their biodiversity is lacking. Recent studies have revealed notable population-level phylogeographical structure within the *Pteroptyx tener* and *P. bearni* groups in Malaysia, suggesting that cryptic species may exist. Additionally, morphological and genetic similarities between *P. balingiana* and *P. malaccae* have raised questions about the former's validity as a distinct species. We collected samples from previously unsampled populations and assembled the most comprehensive genetic dataset for *Pteroptyx* to date, to characterize species boundaries within the *P. tener*, *P. bearni* and *P. malaccae* groups. Using a suite of species delimitation analyses, we show that *P. tener* along the west coast of Peninsular Malaysia (PM) is distinct from populations from the east coast and Borneo despite the absence of morphological differentiation. However, analyses could not conclusively differentiate *P. bearni* from Borneo and eastern PM, nor identify *P. balingiana* and *P. malaccae* as distinct species, indicating that these populations may be conspecific or represent incipient species. This study underlines the need to increase geographical, taxonomic and genetic sampling of Southeast Asian fireflies to provide a better understanding of their biodiversity.

ADDITIONAL KEYWORDS: ABGD – bGMYC – CAD – CO1 – DNA barcode – *gdi* – Lampyridae – mPTP – phylogenetics.

INTRODUCTION

Fireflies (or lightning bugs) are soft-bodied beetles of the family Lampyridae. The estimated number of species ranges from 2000 to 8000 globally (Lloyd, 2008). Approximately 1200 firefly species are known from tropical America (Faust, 2004), more than 400 species are known from Southeast Asia and the Indo-Pacific region (largely from the subfamily Luciolinae; Ballantyne *et al.*, 2015), while little information is available for tropical Africa (Lloyd, 2008).

Fireflies are best known for the ability of males to emit precise flashing patterns to attract females, although not all species are luminescent. The most spectacular of all flashing displays is the near-to-perfect synchronous flashing of fireflies from the genus *Pteroptyx* Olivier (subfamily Luciolinae), in which species are known to occur in multitudes on trees and shrubs along tidal rivers of mangrove swamps (Jusoh *et al.*, 2018). The highly precise synchronous flashing and tree-swarming behaviour have attracted scientific and public interest to explicate the mechanisms of synchronicity (Buck, 2004), elucidate congregating and courting behaviour (Lloyd *et al.*, 2006; Case, 2007), and recently, present them as icon species for sustainable ecotourism and conservation (Jusoh *et al.*, 2018). However, although fireflies have been the focus of numerous ecological, physiological and even biomedical research (Gould & Subramani, 1988; Ermentrout, 1991; Copeland & Moiseff, 2006; Fraga, 2008; Schena *et al.*, 2015), fundamental knowledge on their biodiversity and evolutionary history is still lacking (Chen *et al.*, 2019), underlining the need for more systematic and evolutionary research.

According to the latest studies (Ballantyne *et al.*, 2015, 2019; Jusoh *et al.*, 2018), *Pteroptyx* is an Oriental

*Corresponding author. E-mail: chankinonn@gmail.com

genus comprising 18 species distributed from Hong Kong, southwards through the Philippines (Ballantyne, 2001) and Southeast Asia and westward to India (Madras) (Ballantyne & McLean, 1970; Ballantyne *et al.*, 2011). Until recently, species identification and descriptions in the subfamily Luciolinae, which includes *Pteroptyx*, were based on morphology. However, without highly trained taxonomists who can examine morphological characters of the male genitalia, fireflies can be difficult to identify to species (Ballantyne, 2012). Moreover, it is almost impossible to make accurate identifications of female and larval specimens, even to genus unless the specimens were collected in association with males (Jusoh *et al.*, 2014, 2018).

Developments in molecular sequencing and analytical methods have greatly improved our ability to detect, discover and describe new species (Sites & Marshall, 2003; Wiens, 2007). Furthermore, comparing genetic divergences from informative markers has been shown to be an effective way to identify species and screen for taxonomic misidentifications and potential new species (Hebert *et al.*, 2003; Vences *et al.*, 2005; Fouquet *et al.*, 2007; Padial & De La Riva, 2007; Vieites *et al.*, 2009; Chan *et al.*, 2018; Chan & Grismer, 2019). Among the genetic markers being used, the DNA barcode based on the *CO1* mitochondrial gene is one of the most widely sequenced, especially among invertebrates. Although DNA barcoding was initially developed for species identification (Hebert *et al.*, 2003), this robust marker has also been used for species discovery and delimitation (Hebert & Gregory, 2005; Cao *et al.*, 2016; Hubert & Hanner, 2016; Sheth & Thaker, 2017; Yang & Rannala, 2017; Machado *et al.*, 2018; DeSalle & Goldstein, 2019).

DNA sequence data for fireflies, in particular the *CO1* gene, have recently been accumulating but a well-characterized profile of genetic variation and a justified threshold for species boundaries have yet to be determined. The first study that implemented DNA barcoding of several species of *Pteroptyx* in Malaysia found congruence between morphology and molecular divergences (Jusoh *et al.*, 2014). However, the molecular data also revealed considerable phylogeographical structure (Jusoh *et al.*, 2014: fig. 2), suggesting that the genus *Pteroptyx* could potentially harbour cryptic species. In particular, *P. tener* Olivier was clustered into three geographically isolated clades: along the east and west coasts of Peninsular Malaysia (PM east and PM west) and Borneo (Fig. 1). However, subsequent examination of morphological characters did not reveal significant differences

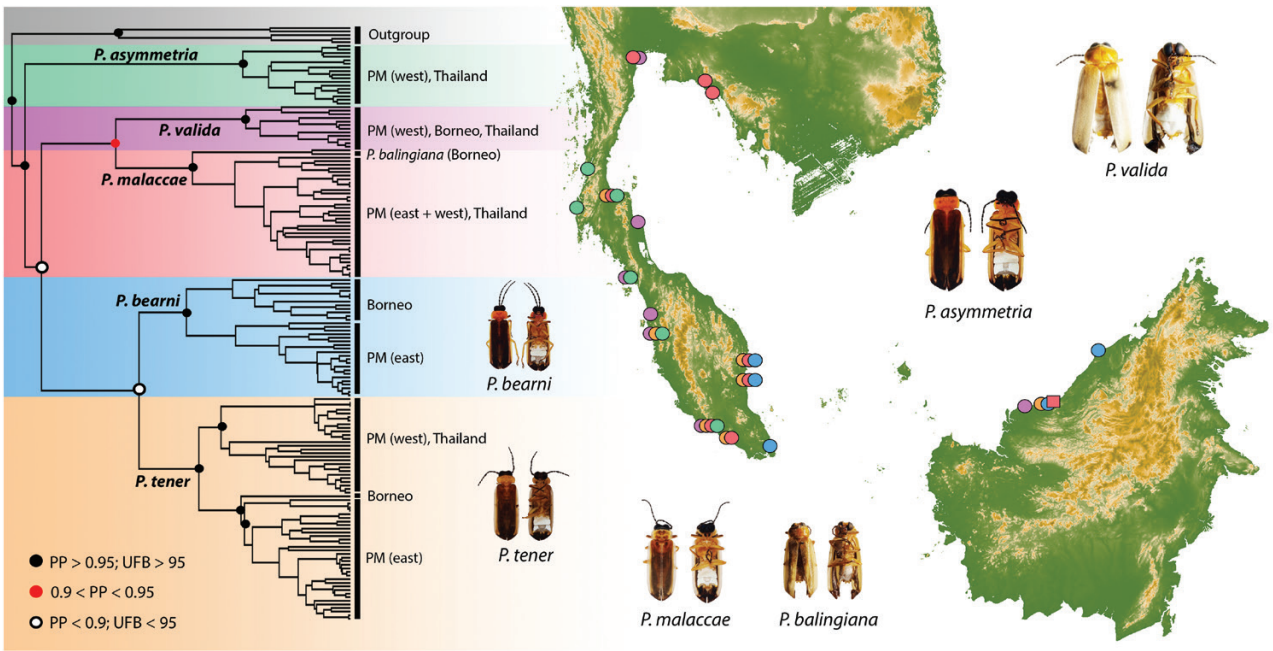

Figure 1. Geographical distribution of *Pteroptyx* samples used in this study and the BEAST phylogeny inferred from a concatenated sequence matrix consisting of 2119 bp of the *CO1* mitochondrial and *CAD* nuclear genes (maximum likelihood produced the same topology). Bayesian posterior probabilities (PP) and ultrafast bootstrap support values (UFB) for major nodes are represented by black, red and open circles (see key). Species-level clades are colour-coded to match locality points on the distribution map. Inset figures show dorsal and ventral views of males of representative species from the ingroup (not scaled to relative size).

among geographically circumscribed haplotypes (Jusoh *et al.*, 2018). In Thailand, a population of *P. tener* was discovered for the first time in 2015, which prompted the hypothesis that geographically structured polymorphisms may exist (Sriboonlert *et al.*, 2015). *Pteroptyx bearni* Olivier, on the other hand, was geographically and genetically clustered into two clades (PM east and Borneo), which also exhibited differences in coloration (Jusoh *et al.*, 2014).

Pteroptyx malaccae (Gorham) is a widely distributed species in mangroves throughout Southeast Asia (Ballantyne & McLean, 1970; Ballantyne, 1987). In Malaysia, *P. malaccae* occurs in small populations, usually in sympatry with *P. tener*, which occurs in much bigger congregations (Jusoh *et al.*, 2011, 2018). In Thailand, *P. malaccae* forms mass congregations in trees along the riverbanks in sympatry with *Pteroptyx valida* Olivier (Prasertkul, 2018). Jusoh *et al.* (2018) hypothesized that *P. malaccae* constitutes a morphologically variable species complex that requires further investigation because four morphologically distinct groups were detected. The same study also revealed that *P. malaccae* formed a sister relationship with a newly described species, *Pteroptyx balingiana* Jusoh. The authors previously argued that the latter species could be 'isolated from the *P. malaccae* by geographically structured variation in the DNA barcodes' (Jusoh *et al.*, 2014: 709), but it was uncertain whether this represented population or species level differentiation.

Recent studies have suggested that the spatio-genetic structure of populations within *P. tener*, *P. bearni* and *P. malaccae*/*P. balingiana* could represent complexes containing cryptic and undescribed species (Jusoh *et al.*, 2014, 2018), but no explicit species delimitation analyses have been performed. In this study, we collected genetic material from previously unsampled populations and assembled the most comprehensive genetic dataset of *Pteroptyx* to date, to estimate evolutionary relationships and characterize species boundaries. Specifically, we used a variety of distance-, tree- and coalescent-based species delimitation methods to (1) test whether geographically and genetically structured populations of *P. tener* and *P. bearni* represent cryptic/distinct species; and (2) determine if *P. balingiana* is sufficiently distinct from *P. malaccae* to warrant specific recognition.

MATERIALS AND METHODS

Taxon sampling and genetic data

For this study, we collected 64 new sequences from 12 unique localities across PM and Borneo (Fig. 1). Additionally, we supplemented this dataset with relevant sequences from GenBank, including published sequences from Thailand (Sartsanga *et al.*, 2018) and sequences from the first author's previous studies (Jusoh *et al.*, 2014, 2018). Apart from the focal taxa (*P. malaccae*, *P. balingiana*, *P. tener* and *P. bearni*), we also included other closely related taxa such as *Pteroptyx asymmetria* Ballantyne and *P. valida*, while outgroup samples included *Pteroptyx galbina* Jusoh, *Pteroptyx testacea* (Motschulsky), *Colophotia praeusta* (Eschscholtz) and *Colophotia brevis* Olivier. The final dataset consisted of 157 sequences of the mitochondrial cytochrome *c* oxidase subunit 1 (*CO1*) gene and 39 sequences of the nuclear protein coding gene carbamoylphosphate synthetase (*CAD*; Supporting Information). All sequences used in this study and their associated GenBank accession numbers are presented in Table S1. The *CO1* gene comprised three fragments: the Folmer region, a 665-bp fragment located at the 5′ end of the *CO1* gene (Folmer *et al.*, 1994); a ~540-bp fragment at the 3′ end (Villalba *et al.*, 2002) [trimmed to exclude any sequences of the cytochrome *c* oxidase subunit 2 (*CO2*) gene and *tRNA-Leu*]; and a long fragment (~1310 bp) that partially overlaps the first two fragments (Sartsanga *et al.*, 2018). Samples that contained more than one of these fragments were combined to form a single contig. The *CAD* gene was amplified using the primers CD439 (F) and CD688 (R) (Wild & Maddison, 2008). Both *CO1* and *CAD* sequences were checked for stop codons and aligned separately using MUSCLE with default parameters performed on MEGA-X v.10.0.5 (Kumar *et al.*, 2018). Aligned sequences are available in the Supporting Information.

Phylogenetic analyses

We used the program IQ-TREE v.1.6 (Nguyen *et al.*, 2015) to estimate a maximum likelihood (ML) phylogeny, while a Bayesian phylogeny was inferred using BEAST v.2.5 (Bouckaert *et al.*, 2014). For the ML phylogeny, sequences were partitioned by gene and the best-fit model of DNA evolution for each partition was determined using ModelFinder (Kalyaanamoorthy *et al.*, 2017). Branch support was assessed with 5000 bootstrap replicates using Ultrafast Bootstrap Approximation (UFB; Hoang *et al.*, 2018). UFB values above 95 were considered well supported. The BEAST analysis was implemented through the CIPRES portal (Miller *et al.*, 2010) and the best-fit substitution model for each partition was estimated via model averaging using the bModelTest plugin in BEAST (Bouckaert & Drummond, 2017). A relaxed log-normal and Yule model was used as the molecular clock and tree priors, respectively, while all other priors were set to default values. We executed two separate Markov chain Monte Carlo (MCMC) runs at 50 million generations each and checked for convergence using the program Tracer

v.1.6 (Rambaut *et al.*, 2018). Sampled trees from both MCMC runs were combined using *logcombiner* and a maximum clade credibility tree was constructed using *treeannotator* with a burn-in of 10%.

Species delimitation

Genetic distance and ABGD: We calculated uncorrected *p*-distances within and between morpohologically determined taxa using MEGA-X (Kumar *et al.*, 2018) to obtain a profile of intra- and interspecies divergence distributions. Subsequently, the Automatic Barcoding Gap Discovery (ABGD) method was used to detect breaks between intraspecific and interspecific diversity known as the barcode gap (Puillandre *et al.*, 2012). ABGD uses pairwise distances to determine divergence between sequences and does not require a priori specification of divergence thresholds. Because ABGD analyses single-locus data and has been shown to be effective with the *CO1* gene (Puillandre *et al.*, 2012; Lin *et al.*, 2018), we performed this analysis on the *CO1* alignment. The analysis was performed through the web-server (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) using default settings and the Kimura two-parameter (K80) distance model (Ratnasingham & Hebert, 2013).

bGMYC: Compared to ABGD, the Bayesian implementation of the general mixed Yule-coalescent (bGMYC) is a phylogeny-based method that does not rely on similarity threshold parameters. It models the Yule and coalescent processes on an ultrametric tree to determine the transition between intra- and interspecific divergences. To account for potential errors in phylogenetic estimation and uncertainty in model parameters, this method integrates over uncertainty in tree topology and branch lengths via MCMC runs (Reid & Carstens, 2012). As input, we used 100 randomly selected trees that were sampled from the posterior distribution of the BEAST analysis. For each tree, we ran the MCMC sampler for 50 000 generations with a burn-in of 40 000, retaining 10 000 post-burn-in generations with a thinning interval of 100. We then assessed a range of probability thresholds ranging from conservative (0.05) to moderate (0.1) and liberal (0.25) delimitation schemes. The analysis outputs results in the form of posterior probabilities (PP) that sequences are conspecific. Hence, we considered populations with PP < 0.05 as strong support, 0.1 < PP < 0.05 as moderate support and PP > 0.1 as weak support for species divergence.

mPTP: The multi-rate Poisson tree process (mPTP) is a phylogeny-aware method that also does not rely on similarity threshold parameters. Similar to bGMYC, it models the transition between intra- and interspecific divergence determined by the coalescent and speciation parameters, where intraspecific branching events are expected to be more frequent compared to among species (Kapli *et al.*, 2017). However, mPTP differs from bGMYC in modelling speciation and coalescent events relative to numbers of substitutions rather than time (Tang *et al.*, 2014). It takes into account the evolutionary relationships of sequences and uses the number of accumulated expected substitutions between subsequent speciation events to model the branching process. We used the ML phylogeny as the input tree, and confidence of delimitation schemes were assessed using two independent MCMC chains at 5 000 000 generations each. Support values represent the fraction of sampled delimitations in which a node was part of the speciation process.

BPP: The BPP (Bayesian Phylogenetics and Phylogeography) analysis uses a Bayesian modelling approach to estimate posterior probabilities of species assignments using gene trees, while taking into account uncertainties in the coalescent process (Yang & Rannala, 2010). The A10 analysis (species delimitation using a fixed guide tree) was implemented using relationships derived from phylogenetic analyses as a guide tree. We used a diffuse prior of $\alpha = 3$ for both θ and τ priors, and the corresponding β parameter was adjusted according to the mean (m) estimate of nucleotide diversity (for θ) and node height (for τ) using the equation $m = \beta/(\alpha - 1)$, for $\alpha > 2$ (Flouri *et al.*, 2018). The mean root node height was obtained from the BEAST phylogeny. To accommodate uncertainty in the guide tree, we also performed the A11 analysis (joint species delimitation and species-tree estimation). The MCMC was set to 100 000 samples with burn-inn = 10 000 and sample frequency = 5. Convergence was assessed by comparing the consistency of posterior distributions (Yang, 2015; Flouri *et al.*, 2018; Leaché *et al.*, 2019). A PP ≥ 0.95 was considered highly supported, 0.90 ≤ PP < 0.95 was considered moderately supported and PP < 0.9 was considered weakly supported.

Heuristic gdi: because Bayesian model selection in BPP can sometimes oversplit species by detecting population splits as species divergence (Sukumaran & Knowles, 2017; Leaché *et al.*, 2019), we used the heuristic genealogical divergence index (*gdi*), which has been shown to produce more accurate results (Jackson *et al.*, 2017; Chan & Grismer, 2019; Leaché *et al.*, 2019). First, the A00 analysis in BPP was implemented to generate posterior distributions for the parameters τ and θ using the same empirically derived priors. Four separate runs were performed to ensure convergence and converged runs were combined to generate posterior distributions for the multispecies coalescent parameters that were subsequently used to

calculate the *gdi* following the equation: $gdi = 1 - e^{-2\tau/\theta}$ (Jackson *et al.*, 2017; Leaché *et al.*, 2019). Population A is distinguished from population B by using $2\tau_{AB}/\theta_A$, while $2\tau_{AB}/\theta_B$ is used to differentiate population B from population A. Populations are considered distinct species when *gdi* values are > 0.7, while low *gdi* values < 0.2 indicate that populations belong to the same species. Values of 0.2 > *gdi* < 0.7 indicate ambiguous species status (Pinho & Hey, 2010; Jackson *et al.*, 2017).

RESULTS

Phylogenetic analyses

Both ML and Bayesian phylogenies produced congruent topologies, with nominal species forming clades with high support (Fig. 1; Supporting Information, Figs S1, S2). In general, *Pteroptyx valida*, *P. asymmetria* and *P. malaccae* did not show geographically circumscribed genetic sub-structuring, with substantial mixing occurring between individuals from PM and Thailand. The exception was five individuals of *P. malaccae* from south-eastern Thailand (*P. malaccae* Thai 11–15; Figs S1, S2; Table S1) that formed a distinct clade. *Pteroptyx balingiana* was reciprocally monophyletic with *P. malaccae* with high support, while *P. bearni* formed two distinct clades (Borneo and PM). The Bornean population of *P. tener* formed a sister relationship with populations from eastern PM and this clade (Borneo + PM east) was in turn reciprocally monophyletic with populations from western PM (Fig. 1). Using these relationships as a framework for hypothetical species boundaries, we defined and compared the following geographically circumscribed and reciprocally monophyletic population pairs as candidate species for downstream species delimitation analyses: (1) *P. balingiana* vs. *P. malaccae*; (2) *P. bearni* Borneo vs. *P. bearni* PM; (3) *P. tener* Borneo vs. *P. tener* PM east; and (4) *P. tener* PM west vs. *P. tener* Borneo + PM east (Fig. 1).

Genetic divergence and species delimitation

Genetic distance: On average, the *CO1* alignment consisted of 30% missing data per sequence. Interspecies genetic divergences between the sister lineages *P. bearni*/*P. tener* and *P. valida*/*P. malaccae* were high (6.3–8.9% and 8–10.5% respectively), while divergences among populations of the same species were low (< 3%). For the candidate species, divergences were slightly higher compared to intrapopulation distributions, except for *P. tener* PM west vs. *P. tener* Borneo + *P. tener* PM east, which were higher (3.15–5.49%), but still well below interspecific divergences (Fig. 2).

ABGD: A total of seven partitions were inferred with prior maximal intraspecific distances (P) ranging from 0.001 to 0.02 (Table 1). For the initial run, the number of delimited species plateaued at ten (including outgroups). This delimitation scheme lumped *P. balingiana* with *P. malaccae*, *P. bearni* Borneo with *P. bearni* PM, and *P. tener* Borneo with *P. tener* PM east. However, *P. tener* PM west was delimited as a distinct species from *P. tener* Borneo + *P. tener* PM east (Fig. 3; Table 2). At *P* = 0.0077, the recursive run also recovered the same ten species but at *P* = 0.0046 (partition 4), *P. tener* Borneo was further split from *P. tener* PM east. Partitions 1–3 of the recursive run produced 13–26 species, which we consider as erroneous as some of the delimited groups were not monophyletic.

bGMYC: The distribution of the coalescence to Yule branching rate ratios were well above zero, indicating that the model is a good fit to the data. The results showed low support for the splitting of *P. balingiana*/*P. malaccae* (conspecificity PP = 0.29), *P. bearni* Borneo/PM (PP = 0.23) and *P. tener* Borneo/PM east (PP = 0.5). On the other hand, the splitting of *P. tener* Borneo + *P. tener* PM east/*P. tener* PM west was moderately supported (PP = 0.08; Table 2). At the most conservative threshold of 0.05, the analysis lumped *P. balingiana* with *P. malaccae*, and all populations of *P. bearni* and *P. tener*. At threshold 0.1, *P. tener* PM west was split from *P. tener* Borneo + PM east, while at the most liberal threshold of 0.25, *P. bearni* Borneo was further split from *P. bearni* PM east (Fig. 3).

mPTP: The mPTP analysis did not support the splitting of *P. balingiana* and *P. malaccae* (PP = 0.16), *P. bearni* Borneo and *P. bearni* PM (PP = 0.13), or *P. tener* Borneo and *P. tener* PM east (PP = 0.67); however, it highly supported the split between *P. tener* PM west and *P. tener* Borneo + *P. tener* PM east (PP = 1.0; Fig. 3).

BPP and gdi: All candidate populations were delimited as distinct species with high support (PP = 1.0) in the BPP A10 and A11 analyses (Fig. 4; Table 2). The heuristic *gdi* analysis provided high support for the splitting of *P. balingiana*/*P. malaccae* and *P. tener* Borneo + PM east/PM west, but uncertain support for *P. bearni* Borneo/PM and *P. tener* Borneo/PM east (Figs 3, 4; Table 2).

DISCUSSION

Phylogenetic relationships

Previous phylogenetic relationships within the subfamily Luciolinae were based on 438 morphological characters (Ballantyne *et al.*, 2013, 2015) until

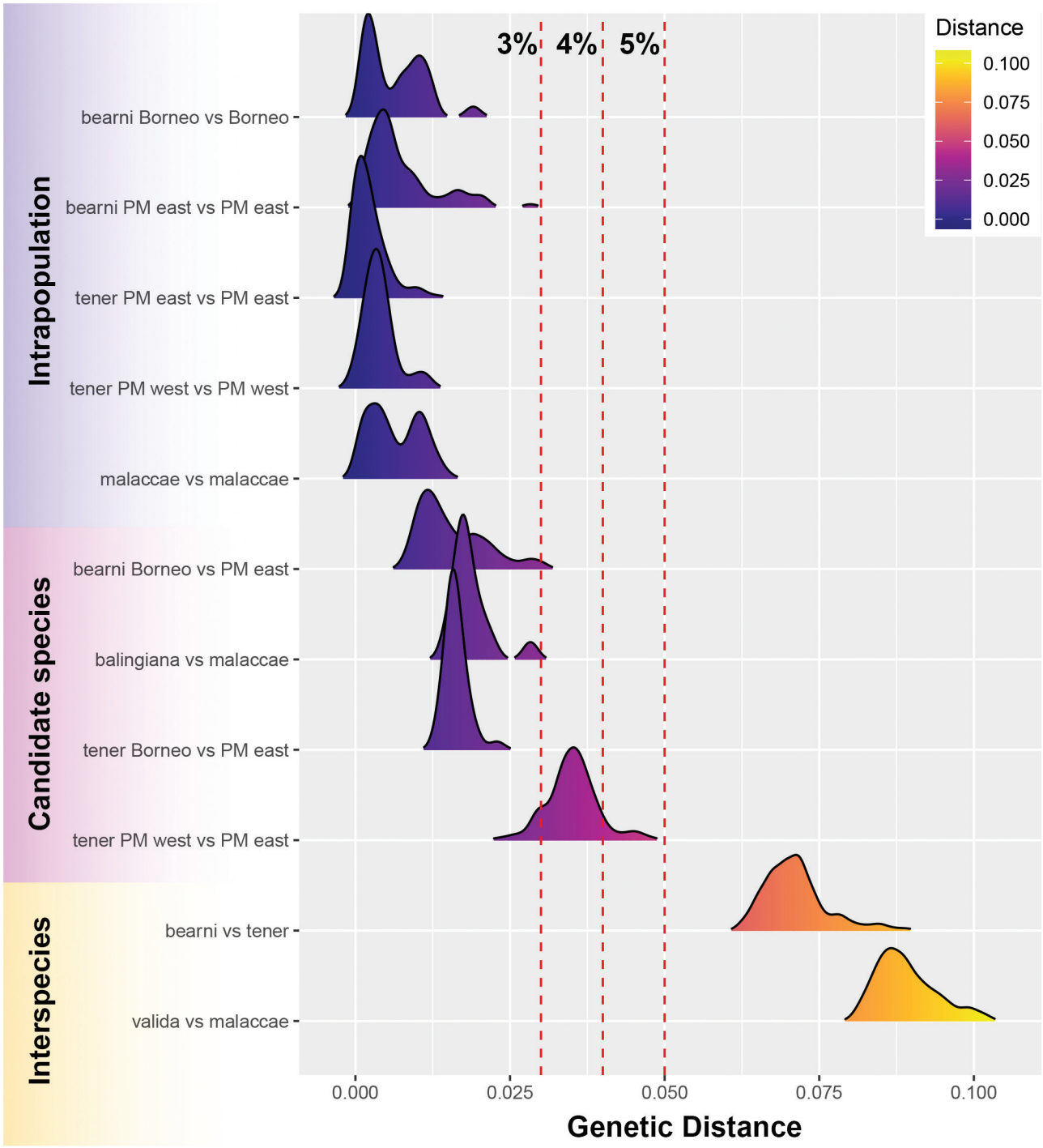

Figure 2. Distribution of genetic distances based on the *CO1* mitochondrial gene. Candidate species are represented by reciprocally monophyletic and geographically circumscribed populations that were subjected to downstream species delimitation analyses. Dotted red lines represent the 3, 4 and 5% genetic distance thresholds that are commonly used to represent species-level divergence.

Jusoh *et al.* (2018) made the first attempt to combine and jointly analyse morphological data with a small subset of molecular data derived from 25 taxa (total = 158 taxa) using Bayesian inference (BI) and maximum parsimony (MP). Although molecular taxon sampling was limited, the data produced a similar topology to that with the larger morphological matrix, but the 'total evidence' morphological and molecular

dataset analysis produced different relationships. Within *Pteroptyx*, MP analysis (Jusoh *et al.*, 2018: fig. 2, part 2) revealed that *P. balingiana* formed a clade with *Pteroptyx macdermotti* McLean and *P. gelasina*, and this clade was reciprocally monophyletic with *P. malaccae* (the BI analysis was unresolved). This does not conflict with the topology from the present study as no molecular samples of *P. macdermotti* and *P. gelasina* were available for analyses. The phylogenetic positions of *P. valida* and *P. asymmetria* from Jusoh *et al.* (2018) were not concordant with the results from this study. Our results showed *P. asymmetria* as the basal lineage with regard to *P. valida*, *P. malaccae*, *P. bearni* and *P. tener*, but with low support. In contrast, Jusoh *et al.* (2018) recovered *P. asymmetria* within the *P. bearni* + *P. tener* subclade, which was in turn sister to the *P. valida* clade with high support. Based on morphology, *P. asymmetria* has stronger affinity to *P. tener* and *P. bearni*, and the phylogeny given by Jusoh *et al.* (2018) was based on a combination of morphological characters and molecular sequences. As such, we consider the latter topology to be more robust than ours with regard to those clades. Unfortunately, with the current data, we are still unable to resolve the species relationships of Malaysian *Pteroptyx* with high certainty and more data in the form of additional molecular loci and taxon sampling will be needed to provide more robust inferences.

Table 1. Partitions, number of species (initial run followed by recursive run in parentheses), and corresponding prior maximal distance from the ABGD analysis using the Kimura (K80) distance model and a relative gap with of $X = 1.5$

	Number of species	Prior max. distance
Partition 1	10 (26)	0.001
Partition 2	10 (14)	0.001668
Partition 3	10 (13)	0.002783
Partition 4	10 (11)	0.004642
Partition 5	10 (10)	0.007743
Partition 6	9 (10)	0.012915
Partition 7	1 (1)	0.021544

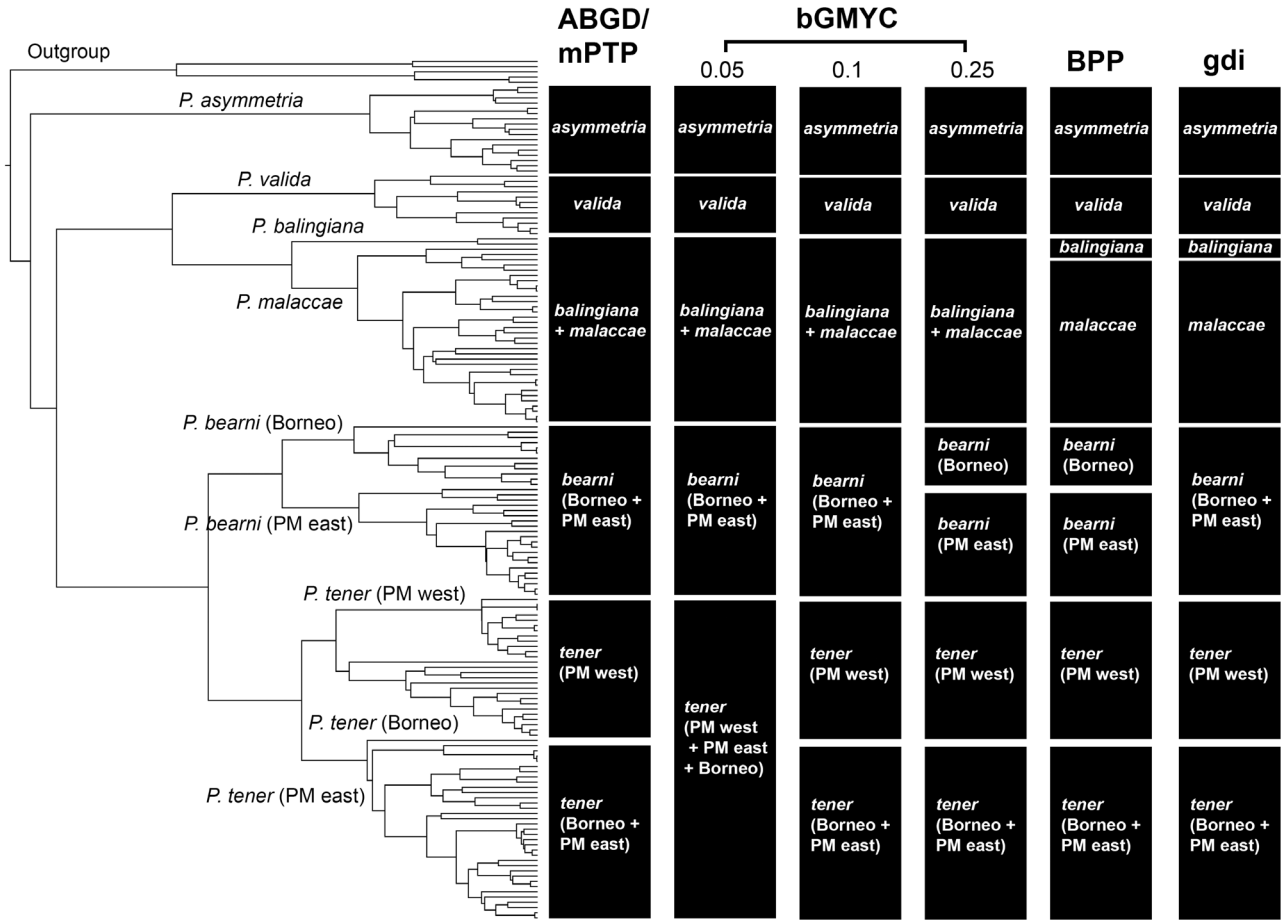

Figure 3. Summary of putative species boundaries resulting from the various species delimitation analysis.

Table 2. Summary of the species delimitation results on the candidate species

	P. balingiana vs. *P. malaccae*	*P. bearni* Borneo vs. *P. bearni* PM	*P. tener* Borneo vs. *P. tener* PM east	*P. tener* Borneo+PM east vs. *P. tener* PM west
P-distance (%)	1.48–2.92	0.84–5.00	1.39–2.29	3.15–5.49
ABGD*	Lump/split	Lump/split	Lump/split	Split
mPTP†	Low (0.16)	Low (0.13)	Low (0.67)	High (1.0)
bGMYC‡	Low (0.29)	Low (0.23)	Low (0.23)	Moderately high (0.08)
BPP (A10)†	High (1.0)	High (1.0)	High (1.0)	High (1.0)
BPP (A11)†	High (1.0)	High (1.0)	High (1.0)	High (1.0)
gdi§	High	Uncertain	Uncertain	High

*Combined results from partitions 4–6 that delimited nine to ten species. See Results for more details.
†Support values denote posterior probabilities that lineages are separate species. PP > 0.95 = High; 0.9 < PP < 0.95 = Moderate; PP < 0.9 = Low.
‡Support values denote posterior probabilities that lineages are conspecific. Hence, lineages are highly supported as separate species if PP < 0.05, while 0.1 > PP > 0.05 = moderate support and PP > 1.0 = low support.
§Average *gdi* > 0.7 highly supports lineages as separate species, 0.2 < *gdi* < 0.7 denotes uncertain species status, while *gdi* < 0.2 denotes lineages are the same species.

Species boundaries and systematics

All previous species delimitation studies of Southeast Asian fireflies have been predicated on traditional morphology, which is ineffective for cryptic groups. Our study is the first to implement robust species delimitation analysis using molecular data to elucidate cryptic species boundaries of Southeast Asia fireflies. Using a suite of species delimitation analyses that are based on different models and assumptions (Fig. 3), we showed that *Pteroptyx* from Southeast Asia runs the gamut of the speciation continuum and exhibits strong phylogeographical patterns that are in some cases incongruent with morphological differentiation. Populations of *P. tener* from Borneo + PM east showed high genetic divergence and were highly supported as distinct from the PM west population across all analyses. Furthermore, each of these populations was represented by multiple individuals. As such, the distinction between these lineages can be posited with relatively high confidence. However, there are no diagnostic morphological characters that distinguish these lineages (see Fig. 5A; Jusoh *et al.*, 2018), indicating that they could constitute cryptic species (Bickford *et al.*, 2007).

Pteroptyx balingiana was described as a distinct species from *P. malaccae* on the basis of morphological and molecular differentiation (Jusoh *et al.*, 2018). Morphologically, *P. balingiana* can be distinguished from *P. malaccae* by structural differences in the light-emitting organ (Fig. 5B, C). More in-depth analyses performed here have demonstrated that genetic divergence between the two taxa was moderate and their status as distinct species was not unequivocally supported by our species delimitation analyses. Nevertheless, the data also did not strongly support the lumping of these taxa. As such, we retain the current taxonomic status of *P. balingiana* and *P. malaccae* as distinct species.

Similarly, species delimitation results for populations of *P. bearni* from Borneo/PM and populations of *P. tener* from Borneo/PM east were also largely inconclusive. According to the morphological examination of freshly preserved specimens by Jusoh *et al.* (2018), individuals of *P. bearni* from PM and Borneo can be distinguished by the coloration on the pronotum and head (Fig. 5D). However, these differences were not apparent in specimens preserved for > 1 year. No morphological differences were detected between populations of *P. tener* from Borneo and eastern PM (Fig. 5A), but this could be due to the limited number of Bornean samples. Based on the relatively shallow bipartitions of these lineages and the absence of notable differences in morphology, we interpret them as being in the early stages of divergence. The distribution of genetic divergences showed slight but clear shifts away from population-level variation, but they have yet to attain the levels of divergence observed among species (Fig. 2). We therefore consider these lineages as incipient species that have begun to diverge, but remain in the grey zone of the speciation continuum (Feder *et al.*, 2012; Nosil & Feder, 2012; Roux *et al.*, 2016). Because evidence was not compelling, we refrain from splitting these populations into separate species pending further investigation.

CAVEATS AND LIMITATIONS

Both mPTP and bGMYC methods seek to identify the transition point between the speciation and coalescent process, and hence adequate taxon representation is needed to provide robust estimates. Consequently, populations represented by singletons or doubletons can affect the accuracy of the analysis (Reid & Carstens, 2012; Kapli *et al.*, 2017). While BPP has

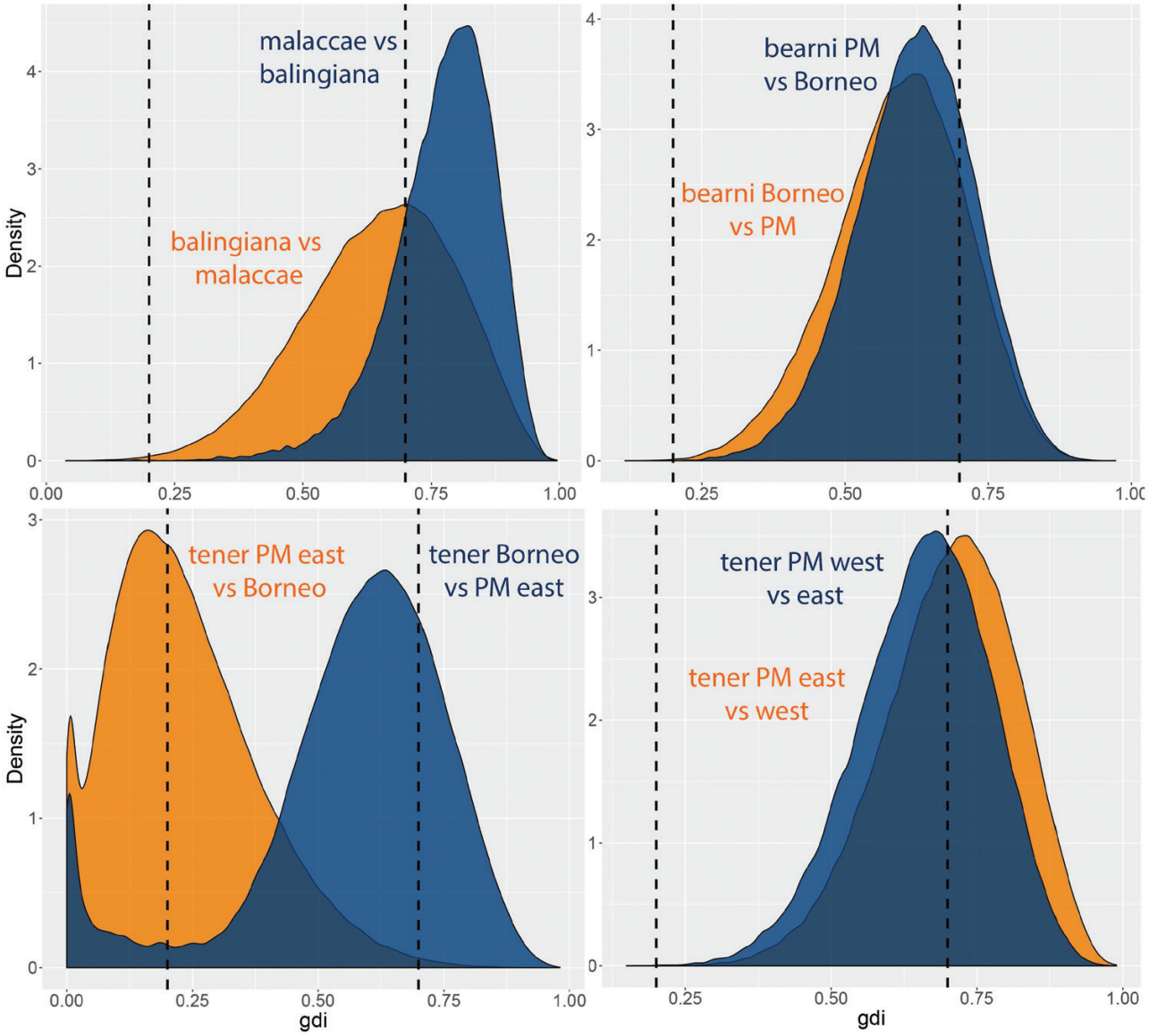

Figure 4. Density plots of *gdi* values with vertical dotted lines representing the 0.2 and 0.7 thresholds. Populations are considered distinct species if *gdi* > 0.7, while *gdi* < 0.2 indicates that populations belong to the same species. Values of 0.2 > *gdi* < 0.7 indicate ambiguous species status.

been shown to be less sensitive to taxon rarity or low numbers of loci (Yang & Rannala, 2017), this analysis is also known to oversplit species due to its inability to differentiate between population structure and species divergence under a protracted speciation model (Sukumaran & Knowles, 2017; Chan & Grismer, 2019; Leaché *et al.*, 2019). This could explain the unrealistically high support for splitting every one of the focal populations, which is in conflict with results from the other species delimitation analyses. Therefore, the BPP analyses and results pertaining to populations with singletons or doubletons should be treated with caution.

The limited number of genes that were available for analysis could also potentially affect the results. Unfortunately, there is a severe lack of nuclear gene representation for fireflies in public databases. Furthermore, additional studies are required to identify informative genes that are more useful for phylogenetics and species delimitation in fireflies, as this has yet to be determined. There is also a dearth of taxonomic representation in molecular samples as only eight out of 18 species of *Pteroptyx* have ever been sequenced, rendering the phylogenetic relationships of this genus uncertain. Morphologically, the genus *Pteroptyx* has three subdivisions: Group I [deflexed

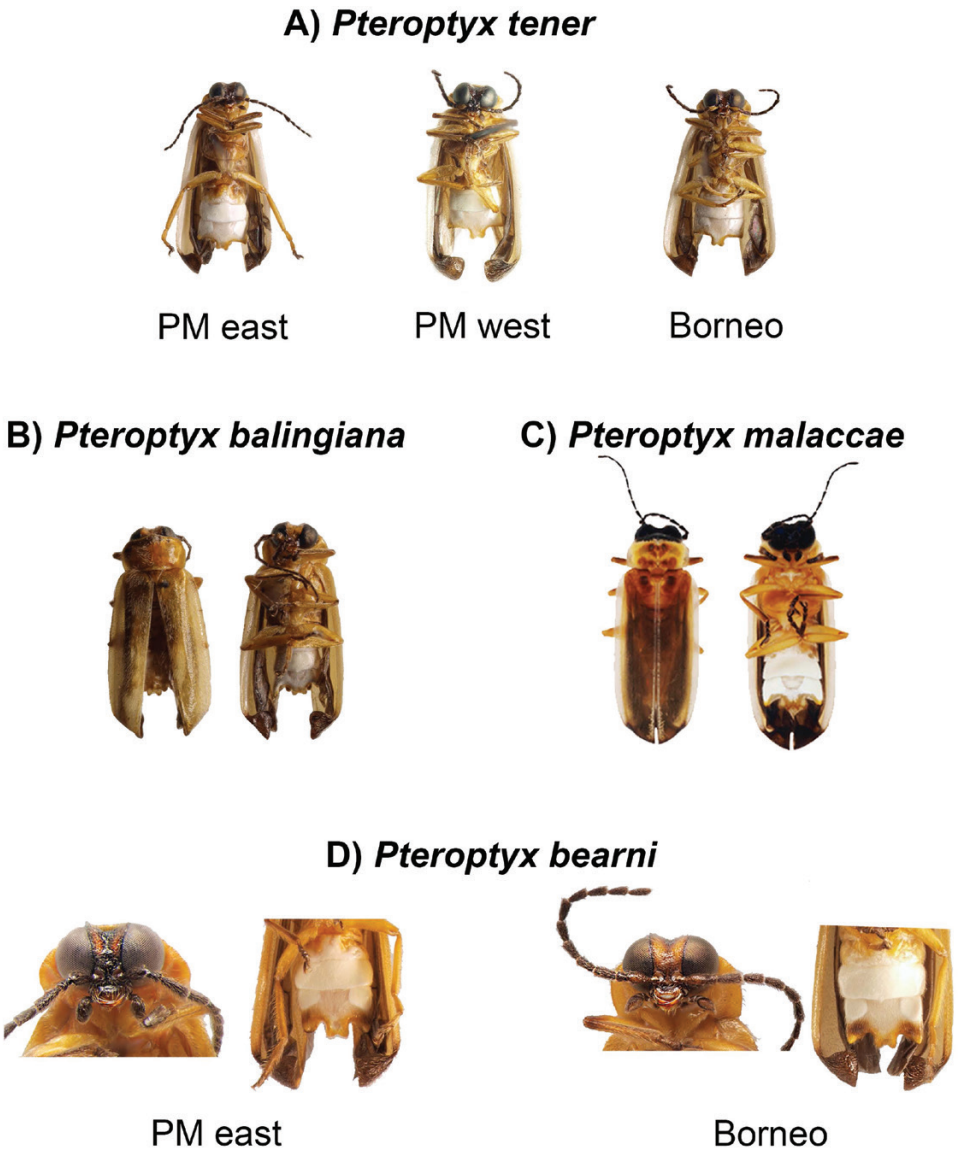

Figure 5. A, ventral view of *Pteroptyx tener* from PM east, PM west and Borneo; B and C, dorsal and ventral views of *P. balingiana* and *P. malaccae* (note the difference in the shape of the light organ); D, *P. bearni* from PM east and Borneo can be differentiated by the colour of the head but both populations have similar light organ structures.

elytra + meta femoral comb (MFC) + bipartite light organ (BLO)] (see Fig. 5); Group II [non-deflexed elytra + MFC + entire light organ (ELO)]; and Group III (non-deflexed elytra + no MFC + BLO; see Jusoh *et al.*, 2018). The present study only dealt with taxa from Group I, commonly known as the bent-winged fireflies, with single representatives from the other groups (Group II = *P. galbina*, Group III = *P. testacea*). Future phylogenetic studies should, ideally, also include key taxa that are closely related to *Pteroptyx* such as *Medeopteryx* Ballantyne and *Pyrophanes* Olivier. Another species that is morphologically similar to *P. malaccae* and *P. balingiana* is *P. gelasina*, which is only known from Sabah, Borneo. Unfortunately, no molecular data are available for this species. Therefore, besides increasing the genetic sampling of *P. balingiana*, acquiring molecular data for *P. gelasina* is crucial to determine the species boundaries and phylogeographical structure of this species complex. While missing taxa could affect the topology of the overall phylogeny, we do not anticipate that it will significantly disrupt our species delimitation results as those were performed at the population level (i.e. populations within the *P. malaccae*, *P. bearni* and *P. tener* groups), which should not be affected by uncertainty in relationships at the species level.

CONCLUSIONS

Despite the charismatic attraction that fireflies hold in cultures worldwide, this study has revealed that fundamental knowledge on Southeast Asian firefly biodiversity and evolutionary history is still lacking. More disconcertingly, much of their habitat, specifically forests and mangroves, are suffering severe degradation and loss from pollution and land conversion (Jusoh & Hashim, 2012; Prasertkul, 2018), setting the stage for an arms race to elucidate the true extent of firefly diversity before it is lost. Our data showed that *Pteroptyx* fireflies in Malaysia contain potential undiscovered cryptic and incipient species that are phylogeographically structured, thereby demonstrating the dynamic and complicated interplay between genes and the environment that drives, maintains and distributes firefly biodiversity. Overall, this study contributes to the growing body of genetic work on fireflies and underlines the urgent need to increase the breadth and depth of geographical and genetic sampling to better understand the evolutionary history of fireflies.

ACKNOWLEDGEMENTS

We thank Chris L. Lambkin of Queensland Museum, John James Wilson of World Museum, Liverpool, and two anonymous reviewers for their helpful comments.

REFERENCES

Ballantyne L. 1987. Further revisional studies on the firefly genus *Pteroptyx* Olivier (Coleoptera: Lampyridae: Luciolinae: Luciolini). *Transactions of the American Entomological Society* **113:** 117–170.

Ballantyne L. 2001. The bent winged fireflies of Cambodia, Indonesia, Malaysia, Philippines and Thailand (Coleoptera: Lampyridae: Luciolinae: Luciolini), *Pteroptyx* spp. of the Polunin collection. *Serangga* **6:** 51–95.

Ballantyne L. 2012. Taxonomy – help or hindrance in south east asia? *Lampyrid* **2:** 1–12.

Ballantyne L, **Fu XH**, **Shih CH**, **Cheng CY**, **Yiu V. 2011.** *Pteroptyx maipo* Ballantyne, a new species of bent-winged firefly (Coleoptera: Lampyridae) from Hong Kong, and its relevance to firefly biology and conservation. *Zootaxa* **34:** 8–34.

Ballantyne L, **Lambkin CL**, **Boontop Y**, **Jusoh WFA. 2015.** Revisional studies on the Luciolinae fireflies of Asia (Coleoptera: Lampyridae): 1. The genus *Pyrophanes* Olivier with two new species. 2. Four new species of *Pteroptyx* Olivier and 3. A new genus *Inflata* Boontop, with redescription of *Luciola indica* (Motsc.). *Zootaxa* **3959:** 1–84.

Ballantyne LA, **Lambkin CL. 2013.** Systematics and phylogenetics of Indo-Pacific Luciolinae fireflies (Coleoptera: Lampyridae) and the description of new genera. *Zootaxa* **3653:** 1–162.

Ballantyne LA, **Lambkin CL**, **Ho JZ**, **Jusoh WFA**, **Nada B**, **Nak-Eiam S**, **Thancharoen A**, **Wattanachaiyingcharoen W**, **Yiu V. 2019.** The Luciolinae of S. E. Asia and the Australopacific region: a revisionary checklist (Coleoptera: Lampyridae) including description of three new genera and 13 new species. *Zootaxa* **4687:** 1–174.

Ballantyne LA, **McLean MR. 1970.** Revisional studies on the firefly genus *Pteroptyx* Olivier (Coleoptera: Lampyridae: Luciolinae: Luciolini). *Transactions of the American Entomological Society* **96:** 223–305.

Bickford D, **Lohman DJ**, **Sodhi NS**, **Ng PKL**, **Meier R**, **Winker K**, **Ingram KK**, **Das I. 2007.** Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22:** 148–155.

Bouckaert R, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** 1–6.

Bouckaert RR, **Drummond AJ. 2017.** bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* **17:** 42.

Buck J. 2004. Synchronous rhythmic flashing of fireflies. II. *The Quarterly Review of Biology* **63:** 265–289.

Cao X, **Liu J**, **Chen J**, **Zheng G**, **Kuntner M**, **Agnarsson I. 2016.** Rapid dissemination of taxonomic discoveries based on DNA barcoding and morphology. *Scientific Reports* **6:** 1–13.

Case JF. 2007. Courting behavior in a synchronously flashing, aggregative firefly, *Pteroptyx tener*. *The Biological Bulletin* **159:** 613–625.

Chan KO, **Abraham RK**, **Grismer JL**, **Grismer LL. 2018.** Elevational size variation and two new species of torrent frogs from Peninsular Malaysia (Anura: Ranidae: Amolops Cope). *Zootaxa* **4434:** 250–264.

Chan KO, **Grismer LL. 2019.** To split or not to split? Multilocus phylogeny and molecular species delimitation of Southeast Asian toads (Family: Bufonidae). *BMC Evolutionary Biology* **19:** 95.

Chan KO, **Grismer LL**, **Brown RM. 2018.** Comprehensive multi-locus phylogeny of Old World tree frogs (Anura: Rhacophoridae) reveals taxonomic uncertainties and potential cases of over- and underestimation of species diversity. *Molecular Phylogenetics and Evolution* **127:** 1010–1019.

Chen X, **Dong Z**, **Liu G**, **He J**, **Zhao R**, **Wang W**, **Peng Y**, **Li X. 2019.** Phylogenetic analysis provides insights into the evolution of Asian fireflies and adult bioluminescence. *Molecular Phylogenetics and Evolution* **140:** 106600.

Copeland J, **Moiseff A. 2006.** Flash precision at the start of synchrony in *Photuris frontalis*. *Integrative and Comparative Biology* **44:** 259–263.

DeSalle R, **Goldstein P. 2019.** Review and interpretation of trends in DNA barcoding. *Frontiers in Ecology and Evolution* **7:** 302.

Ermentrout B. 1991. An adaptive model for synchrony in the firefly *Pteroptyx malaccae*. *Journal of Mathematical Biology* **29:** 571–585.

Faust LF. 2004. Fireflies as a catalyst for science education. *Integrative and Comparative Biology* **44:** 264–265.

Feder JL, **Egan SP**, **Nosil P. 2012.** The genomics of speciation-with-gene-flow. *Trends in Genetics* **28:** 342–350.

Flouri T, **Jiao X**, **Rannala B**, **Yang Z. 2018.** Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution* **35:** 2585–2593.

Folmer O, **Black M**, **Hoeh W**, **Lutz R**, **Vrijenhoek R. 1994.** DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3:** 294–299.

Fouquet A, **Gilles A**, **Vences M**, **Marty C**, **Blanc M**, **Gemmell NJ. 2007.** Underestimation of species richness in neotropical frogs revealed by mtDNA analyses. *PLoS ONE* **2:** e1109.

Fraga H. 2008. Firefly luminescence: a historical perspective and recent developments. *Photochemical and Photobiological Sciences* **7:** 146–158.

Gould SJ, **Subramani S. 1988.** Firefly luciferase as a tool in molecular and cell biology. *Analytical Biochemistry* **175:** 5–13.

Hebert PDN, **Cywinska A**, **Ball SL**, **DeWaard JR. 2003.** Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270:** 313–321.

Hebert PDN, **Gregory TR. 2005.** The promise of DNA barcoding for taxonomy. *Systematic Biology* **54:** 852–859.

Hoang DT, **Chernomor O**, **Von Haeseler A**, **Minh BQ**, **Vinh LS. 2018.** UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35:** 518–522.

Hubert N, **Hanner R. 2016.** DNA barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes* **3:** 44–58.

Jackson ND, **Carstens BC**, **Morales AE**, **O'Meara BC. 2017.** Species delimitation with gene flow. *Systematic Biology* **66:** 799–812.

Jusoh WFA, **Ballantyne L**, **Lambkin CL**, **Hashim NR**, **Wahlberg N. 2018.** The firefly genus *Pteroptyx* Olivier revisited (Coleoptera: Lampyridae: Luciolinae). *Zootaxa* **4456:** 1–71.

Jusoh WFA, **Hashim NR**, **Sääksjärvi IE**, **Adam NA**, **Wahlberg N. 2014.** Species delineation of Malaysian mangrove fireflies (Coleoptera: Lampyridae) using DNA barcodes. *The Coleopterists Bulletin* **68:** 703–711.

Jusoh WFAW, **Hashim NR. 2012.** The effect of habitat modification on firefly populations at the Rembau-Linggi estuary, Peninsular Malaysia. *Lampyrid* **2:** 149–155.

Jusoh WFAW, **Wong CH**, **Hashim NR. 2011.** Zonation of firefly species and their display trees along Kerteh River, Terengganu. *Serangga* **16:** 59–66.

Kalyaanamoorthy S, **Minh BQ**, **Wong TKF**, **Von Haeseler A**, **Jermiin LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14:** 587–589.

Kapli P, **Lutteropp S**, **Zhang J**, **Kobert K**, **Pavlidis P**, **Stamatakis A**, **Flouri T. 2017.** Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* **33:** 1630–1638.

Kumar S, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K. 2018.** MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35:** 1547–1549.

Leaché AD, **Zhu T**, **Rannala B**, **Yang Z. 2019.** The spectre of too many species. *Systematic Biology* **68:** 168–181.

Lin XL, **Stur E**, **Ekrem T. 2018.** Exploring species boundaries with multiple genetic loci using empirical data from non-biting midges. *Zoologica Scripta* **47:** 325–341.

Lloyd JE. 2008. Fireflies (Coleoptera: Lampyridae). In: Capinera JL, ed. *Encyclopedia of entomology*. Dordrecht: Springer, 1429–1452.

Lloyd JE, **Wing SR**, **Hongtrakul T. 2006.** Flash behavior and ecology of Thai *Luciola* fireflies (Coleoptera: Lampyridae). *The Florida Entomologist* **72:** 80.

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. *Scientific Reports* **8:** 8387.

Miller MA, **Pfeiffer W**, **Schwartz T. 2010.** Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop GCE* **2010:** 1–8.

Nguyen LT, **Schmidt HA**, **Von Haeseler A**, **Minh BQ. 2015.** IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32:** 268–274.

Nosil P, **Feder JL. 2012.** Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367:** 332–342.

Padial JM, **De La Riva I. 2007.** Integrative taxonomists should use and produce DNA barcodes. *Zootaxa* **1586:** 67–68.

Pinho C, **Hey J. 2010.** Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics* **41:** 215–230.

Prasertkul T. 2018. Characteristics of *Pteroptyx* firefly congregations in a human dominated habitat. *Journal of Insect Behavior* **31:** 436–457.

Puillandre N, **Lambert A**, **Brouillet S**, **Achaz G. 2012.** ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology* **21:** 1864–1877.

Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67:** 901–904.

Ratnasingham S, **Hebert PDN. 2013.** A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE* **8:** e66213.

Reid NM, **Carstens BC. 2012.** Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology* **12:** 196.

Roux C, **Fraïsse C**, **Romiguier J**, **Anciaux Y**, **Galtier N**, **Bierne N. 2016.** Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology* **14:** 1–22.

Sartsanga C, **Swatdipong A**, **Sriboonlert A. 2018.** Distribution of the firefly genus *Pteroptyx* Olivier and a

new record of *Pteroptyx asymmetria* Ballantyne (Coleoptera: Lampyridae: Luciolinae) in Thailand. *The Coleopterists Bulletin* **72:** 171–183.

Schena A, **Griss R**, **Johnsson K. 2015.** Modulating protein activity using tethered ligands with mutually exclusive binding sites. *Nature Communications* **6:** 7830.

Sheth BP, **Thaker VS. 2017.** DNA barcoding and traditional taxonomy: an integrated approach for biodiversity conservation. *Genome* **60:** 618–628.

Sites JWJ, **Marshall JC. 2003.** Delimiting species: a renaissance issue in systematic biology. *Trends in Ecology and Evolution* **18:** 462–470.

Sriboonlert A, **Swatdipong A**, **Wonnapinij P**, **E-Kobon T**, **Thancharoen A. 2015.** New record of *Pteroptyx tener* Olivier (Coleoptera: Lampyridae: Luciolinae) in Thailand. *The Coleopterists Bulletin* **69:** 332–336.

Sukumaran J, **Knowles LL. 2017.** Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences USA* **114:** 1607–1612.

Tang CQ, **Humphreys AM**, **Fontaneto D**, **Barraclough TG. 2014.** Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution* **5:** 1086–1094.

Vences M, **Thomas M**, **Bonett RM**, **Vieites DR. 2005.** Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360:** 1859–1868.

Vences M, **Thomas M**, **Meijden AVD**, **Chiari Y**, **Vieites DR. 2005.** Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in Zoology* **12:** 1–12.

Vieites DR, **Wollenberg KC**, **Andreone F**, **Kohler J**, **Glaw F**, **Vences M. 2009.** Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences USA* **106:** 8267–8272.

Villalba S, **Lobo JM**, **Martin-Piera F**, **Zardoya R. 2002.** Phylogenetic relationships of Iberian Dung Beetles (Coleoptera: Scarabaeinae): insights on the evolution of nesting behavior. *Journal of Molecular Evolution* **55:** 116–126.

Wiens JJ. 2007. Species delimitation: new approaches for discovering diversity. *Systematic Biology* **56:** 875–878.

Wild AL, **Maddison DR. 2008.** Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics & Evolution* **48:** 877–891.

Yang Z, **Rannala B. 2010.** Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences USA* **107:** 9264–9269.

Yang Z, **Rannala B. 2017.** Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology* **26:** 3028–3036.

Yang Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology* **61:** 854–865.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Maximum likelihood phylogeny estimated using a partitioned analysis on a concatenated sequence matrix consisting of 2119 bp of the *CO1* mitochondrial and *CAD* nuclear genes. Node values denote ultrafast bootstrap support.

Figure S2. Bayesian phylogeny inferred from a concatenated sequence matrix consisting of 2119 bp of the *CO1* mitochondrial and *CAD* nuclear genes. Node values denote posterior probabilities.

Table S1. List of genetic samples used in this study and their corresponding GenBank accession numbers. New sequences are marked with an asterisk (*).

Supporting information 1. Aligned sequences for the *CO1* mitochondrial gene.

Supporting information 2. Aligned sequences for the *CAD* nuclear gene.