Molecular Phylogenetics and Evolution 139 (2019) 106531

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Untangling a mess of worms: Species delimitations reveal morphological crypsis and variability in Southeast Asian semi-aquatic earthworms (Almidae, *Glyphidrilus*)

Parin Jirapatrasilp[a,b], Thierry Backeljau[c,d], Pongpun Prasankok[e], Ratmanee Chanabun[f], Somsak Panha[b,*]

[a] Biological Sciences Program, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[b] Animal Systematics Research Unit, Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[c] Royal Belgian Institute of Natural Sciences, Vautierstraat 29, B-1000 Brussels, Belgium
[d] Evolutionary Ecology Group, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium
[e] School of Biology, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand
[f] Program in Animal Science, Faculty of Agriculture Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon 47000, Thailand

ARTICLE INFO

Keywords:
Species delimitation
Morphological crypsis
Morphological variability
DNA barcodes
ITS2 secondary structure
Earthworms

ABSTRACT

Semi-aquatic freshwater earthworms in the genus *Glyphidrilus* from Southeast Asia are characterized by both an extreme morphological crypsis among divergent phylogenetic lineages and a high morphological variability within the same phylogenetic lineages. The present study provides a new taxonomic framework for this problematic genus in SE Asia by integrating DNA sequence and morphological data. When single-locus and multilocus multispecies coalescent-based (MSC) species delimitation methods were applied to DNA sequence data, they usually yielded highly incongruent results compared to morphology-based species identifications. This suggested the presence of several cryptic species and high levels of intraspecific morphological variation. Applying reciprocal monophyly to the cytochrome *c* oxidase subunit 1 (COI) gene tree allowed us to propose the existence of 33 monophyletic species. Yet, often substantially more molecular operational taxonomic units (MOTUs) were obtained when species delimitation was based on COI and 16S rRNA sequences. In contrast, the ITS1 and ITS2 sequences suggested fewer MOTUs and did not recover most of the monophyletic species from the Mekong basin. However, several of these latter taxa were better supported when MSC species delimitation methods were applied to the combined mtDNA and ITS datasets. The ITS2 secondary structure retrieved one unnamed Mekong basin species that was not uncovered by the other methods when applied to ITS2 sequences. In conclusion, based on an integrative taxonomic workflow, 26 *Glyphidrilus* candidate species were retained and two remained to be confirmed. As such, this study provides evidence to suggest nine species new to science and to synonymize 12 nominal morphospecies. It also illustrates that the uncritical use of COI as a universal DNA barcode may overestimate species diversity because COI may be unable to distinguish between divergent conspecific lineages and different candidate species.

1. Introduction

Soils contain an overwhelming diversity of often poorly known invertebrates (Fitter et al., 2005; Decaëns et al., 2006) and are, therefore, regarded as one of the most challenging frontiers in biodiversity research (André et al., 1994; Briones, 2014). Earthworms are one of the most important components of the soil macrofauna in terms of species diversity, biomass and functional roles, and act as potential ecosystem engineers (Lavelle et al., 2007; Blouin et al., 2013). They also serve as bioindicators of anthropogenic impacts and habitat changes (Fründ et al., 2011; Pérès et al., 2011), and as such, they are used as standard model organisms in several soil ecotoxicological tests (Lee et al., 2008; Römbke et al., 2016). In addition, earthworms are mass cultured as a dietary source of protein and fish bait and also can function as an organic waste management tool (Edwards et al., 2010; Lowe et al., 2014). Despite these variety of uses, the taxonomy and species diversity of earthworms in many regions of the world are still poorly known (Reynolds, 2004).

* Corresponding author.
E-mail address: somsak.pan@chula.ac.th (S. Panha).

https://doi.org/10.1016/j.ympev.2019.106531
Received 20 December 2018; Received in revised form 19 May 2019; Accepted 7 June 2019
Available online 08 June 2019

At present, approximately 6000 species of earthworms are recognized (Reynolds and Wetzel, 2017), but this figure is estimated to represent only up to 30% of the total number of existing species (Blakemore, 2009, 2016). Current earthworm taxonomy almost entirely relies on a few external and internal morphological characters, the most important of which are those from the reproductive system (Edwards and Bohlen, 1996; Jamieson, 2006). However, viewed through human eyes, earthworms are supposed to show 'morphological stasis' due to their relatively simple and stable mode of life in the soil and their structural simplicity without elaborate or specialized copulatory organs (Novo et al., 2012a; Pérez-Losada et al., 2012). This often makes earthworm taxonomy very complex, particularly when closely related species show high levels of overlapping morphological variation. Accordingly, cryptic species are easily overlooked and/or lumped (Briones et al., 2009; Erséus and Gustafsson, 2009). Hence, the overall state of earthworm taxonomy remains unstable and chaotic, and thorough taxonomic revisions are needed (Blakemore, 2016). Given this situation, it is no surprise that molecular techniques, and in particular DNA barcoding, have become a powerful tool to provide alternative lines of evidence in resolving earthworm taxonomy (Chang and James, 2011; Decaëns et al., 2013). However, the practice of combining morphological data with DNA barcodes and species delimitation methods to achieve an integrative taxonomy of earthworms is still poorly explored.

The semi-aquatic freshwater earthworm genus *Glyphidrilus* Horst, 1889 (family Almidae) is characterized by peculiar expansions of the epidermis at the clitellum, referred to as "wings" or "alae", and a quadrangular-shaped cross-section of the posterior body segments (Fig. 1c). Most species live in the transition zone between freshwater and land, such as the topsoil near the edge of freshwater habitats, including waterfalls and paddy fields with muddy to sandy sediments (Brinkhurst and Jamieson, 1971; Chanabun et al., 2013). *Glyphidrilus* is widely distributed in South and Southeast Asia, and is currently comprised of 47 nominal species, including one species and one subspecies from East Africa. This number encompasses 28 recently described species from mainland Southeast Asia, 11 of which were from the Malay Peninsula, 12 from the Mekong River basin and five from the Chao Phraya River basin in Thailand (Chanabun et al., 2013, 2017; Chanabun and Panha, 2015; Jirapatrasilp et al., 2016).

The taxonomy of *Glyphidrilus* species relies heavily on the position of the wings and clitellum, and the number and position of genital markings and spermathecae (Brinkhurst and Jamieson, 1971; Chanabun et al., 2013). However, these characters are highly variable within populations and species, and may show a substantial degree of overlap between populations and species. Chanabun et al. (2013), for example, reported that *Glyphidrilus* populations from northeastern Thailand represent only two, allegedly allopatric, species, viz. *G. chiensis* Chanabun and Panha, 2013 and *G. quadratus* Chanabun and Panha, 2013, that show extensive morphological variation within and between populations. As such, it is difficult to morphologically delimit these two species since the threshold(s) between intra- and interspecific morphological differentiation are still unclear. In contrast, using allozyme data, Jirapatrasilp et al. (2015) were able to unequivocally delimit *G. vangviengensis* Panha and Chanabun, 2011 and *G. mekongensis* Panha and Chanabun, 2012 as two distinct species that occur syntopically along the lower Mekong and show questionable morphological differences. Moreover, the same allozyme data uncovered at least two additional cryptic species from this region. These studies suggest that the actual species diversity of *Glyphidrilus* in mainland Southeast Asia is still underestimated.

In this study, we used two mitochondrial (COI and 16S rRNA) and two nuclear DNA markers (ITS1 and ITS2) to explore the species diversity of *Glyphidrilus* in mainland Southeast Asia. Several single-locus and multilocus species delimitation methods were then applied and compared under various parameter settings in order to assess the degree of congruence among the inferred species boundaries. This approach is supposed to increase the confidence in the suggested species (Carstens et al., 2013), particularly if these species can be subsequently corroborated by phenotypic (i.e. morphological) data within an integrative taxonomic framework (Dayrat, 2005; Padial et al., 2010).

2. Material and methods

2.1. Specimen collection and morphological identification

A total of 285 *Glyphidrilus* specimens were collected throughout mainland Southeast Asia (Fig. S1 and Table S1). Of these, 16 specimens were from the study of Jirapatrasilp et al. (2016). The remaining specimens were newly collected, including from 20 localities already sampled by Jirapatrasilp et al. (2015). Two specimens of *Pontoscolex corethrurus* (family Rhinodrilidae) from Thailand were used as the outgroup. Earthworms were collected by hand, cleaned and then euthanized in 30% (v/v) ethanol prior to fixation in 95% (v/v) ethanol. Adult specimens were identified following Chanabun et al. (2013) and compared with type specimens if possible. The positions of the following seven morphological characters were examined: (1) wings, (2) clitellum, (3) pre-wing median genital markings, (4) post-wing median genital markings, (5) pre-wing lateral genital markings, (6) post-wing lateral genital markings and (7) spermathecae. Juveniles and ambiguously identified specimens were putatively assigned to species by comparing their DNA sequences with those of the most closely related type specimens or, if no association with type specimens was apparent, by keeping them as unidentified molecular operational taxonomic units (MOTUs). Voucher specimens were deposited in the Museum of Zoology, Chulalongkorn University (CUMZ), Bangkok, Thailand.

2.2. Acquisition of DNA sequence data

Total genomic DNA was extracted from small pieces of posterior integument and muscular tissues using the NucleoSpin®Tissue Kit (Macherey-Nagel). Two partial mitochondrial markers (COI and 16S rDNA) and two complete nuclear ribosomal internal transcribed spacers (ITS1 and ITS2) were amplified with the primers and annealing temperatures listed in Table S2. The PCR reactions were performed as reported before (Jirapatrasilp et al., 2016) and PCR products were purified using the ExoSAP-IT® purification kit. Cycle-sequencing was performed with the BigDye® Terminator v. 1.1 cycle Sequencing kit (Applied Biosystems, Lennik, Belgium) and resolved on an ABI 3130xl capillary DNA sequencer (Applied Biosystems). In addition, published sequences of COI and 16S rDNA of 16 specimens from Jirapatrasilp et al. (2016) were included. GenBank accession numbers are given in Table S1.

2.3. Sequence phasing, alignment and phylogeny inference

The ITS sequence chromatograms showing double peaks may indicate heterozygote genotypes, whose phase reconstruction was performed in PHASE v. 2.1.1 (Stephens et al., 2001) using a threshold of 0.95. Sequences were trimmed in MEGA v. 7.0 (Kumar et al., 2016) and aligned using the MAFFT webserver (https://mafft.cbrc.jp/alignment/server/index.html) using the default settings (Katoh et al., 2017). All COI sequences were checked for substitution saturation with DAMBE v. 5.6.21 (Xia, 2013), but as no saturation was detected, all codon positions were used in subsequent analyses. General DNA sequence features and polymorphism estimates were inferred from MEGA v. 7.0 and DnaSP v. 6.10 (Rozas et al., 2017).

Three types of phylogenetic analyses of each gene and concatenated dataset were conducted: maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI). The MP analyses were run in PAUP* v. 4.0b (Swofford, 2002) using a heuristic search with 1000 replicates, and clade support was calculated by bootstrapping with 1000 replicates. Both ML and BI analyses were conducted in the CIPRES Science Gateway (Miller et al., 2010) using RAxML-HPC2 on XSEDE v.

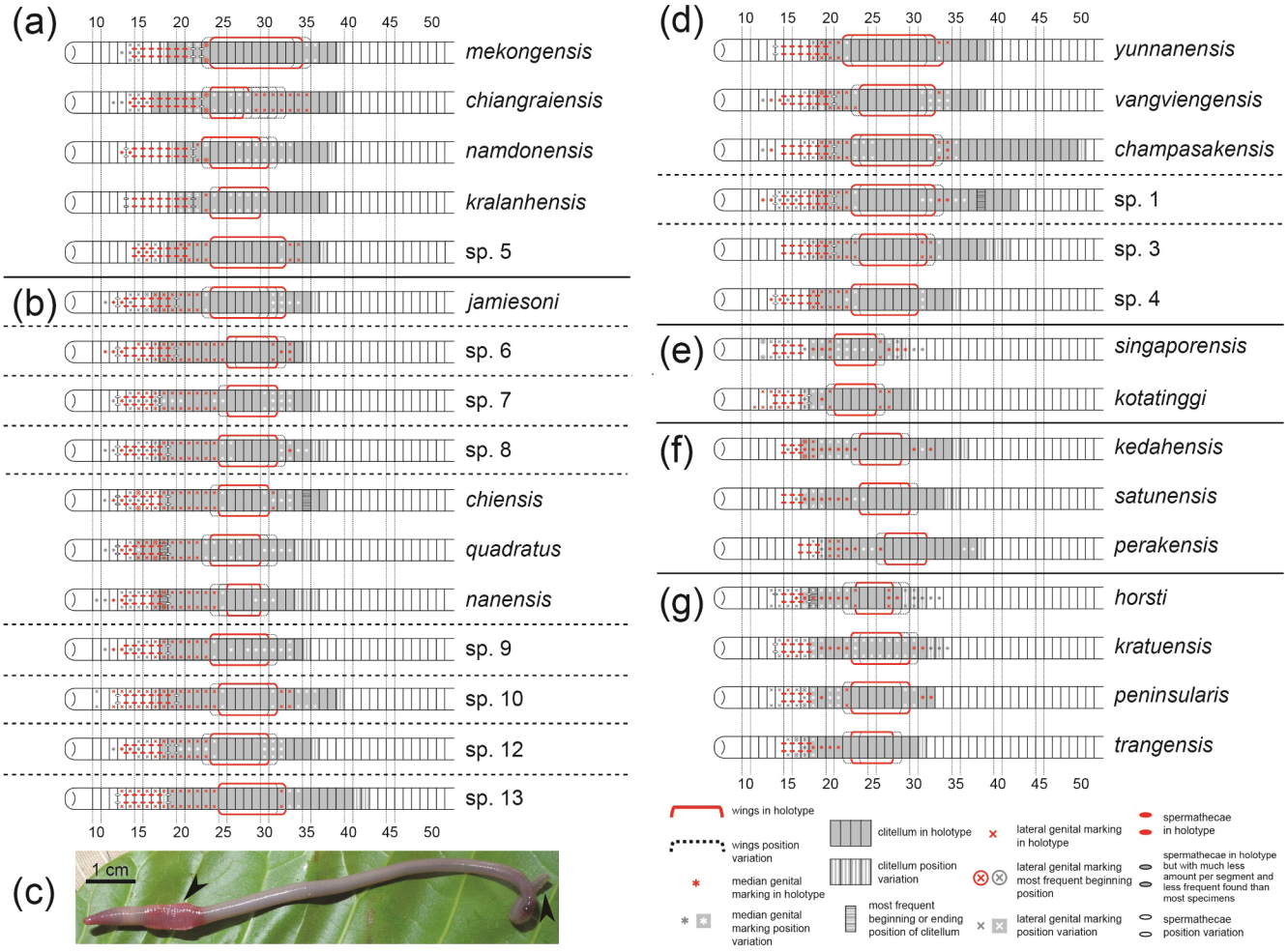

Fig. 1. Morphological comparisons among some *Glyphidrilus* spp. that show overlapping characters or belong to the same species complex/clade. (a) *G. mekongensis* species complex, (b) *Glyphidrilus* spp. with *chiensis* morph, (c) a specimen of recently preserved *G. jamiesoni*, where the arrows indicate anterior wings and a quadrangular-shaped cross-section of the posterior body segments, (d) *G. yunnanensis* species complex, (e) *G. singaporensis* clade, (f) *G. kedahensis* clade and (g) *G. horsti* clade. Those symbols indicated for characters in the holotype of the nominal species illustrate the most frequent positions in unnamed taxa instead. Spermathecae of *G. yunnanensis* were observed based on recently collected specimens.

8.2.10 (Stamatakis, 2014) and MrBayes on XSEDE v. 3.2.6 (Ronquist et al., 2012), respectively. For each gene, the best-fit models of nucleotide substitution were identified in Kakusan4 (Tanabe, 2011), with different models being applied to the COI codon partitions. The ML branch support was assessed using 1000 bootstrap replicates. Two independent BI analyses were run in parallel for 10 million generations each and sampled every 500 generations, starting with a random tree and burn-in set to 50%. Phylogenetic incongruence among the different genes was tested with the S-H test (Shimodaira and Hasegawa, 1999) in PAUP* v. 4.0b, using resampling of estimated log likelihoods (RELL) with 10,000 bootstrap replicates. As the SH test did not yield any incongruence among the four gene fragments, they were concatenated. The concatenated data were then used to infer MP trees in PAUP* v. 4.0b with the same settings as above, while inferring ML and BI trees from the concatenated data was done after partitioning the concatenated data with Kakusan4 and subsequently running ML and BI analyses as described above, except that now two independent BI analyses were ran in parallel for 50 million generations each and sampled every 500 generations. Convergence of the two runs was achieved if the average standard deviation of split frequencies were ≤ 0.01, and adequate Markov chain Monte Carlo (MCMC) samples from the posterior probability distributions were assessed if potential scale reduction factor (PSRF) values of each parameter approached 1.0 (Ronquist et al., 2012). Nodes were considered well-supported if $\geq 70\%$ for MP/ML bootstrap values (Hillis and Bull, 1993) or ≥ 0.95 for BI posterior probabilities (San Mauro and Agorreta, 2010).

Datasets for species delimitation analysis were altered to examine the taxon sampling effects. All phylogenetic trees constructed from the concatenated dataset showed that all the species from the Mekong River basin formed one well-supported clade ("Mekong clade") with *G. chaophraya* and *G. vangthongensis* as sister species (see Results). We, therefore, used the Mekong species as a separate, reduced dataset (with *G. chaophraya* and *G. vangthongensis* as the outgroup) for comparison with the full dataset.

2.4. Reciprocal monophyly as primary species proposal (PSP)

Reciprocal monophyly refers to the situation in which all extant alleles within sister taxa are completely sorted, i.e. when alleles are genealogically closer to one another within a taxon, than to any alleles in its sister taxon (Avise, 2000; Rieppel, 2010). This makes reciprocal monophyly a compulsory condition of the genealogical species concept (GSC), the evolutionary species concept (ESC) and the phylogenetic species concept (PSC) (Baum and Shaw, 1995; Hudson and Coyne, 2002; Rieppel, 2010). To apply the reciprocal monophyly criterion, all clades retrieved from the COI gene tree that showed an "interspecific" divergence of at least 13% (Jeratthitikul et al., 2017) from one another were defined as "monophyletic species" and used as PSPs. The

Table 1
Species delimitation methods used in this study.

Method		Approach[a]	Basis of method[a]	Species concept[b]	Input data	Number of loci[c]	*A priori* parameter setting	*A priori* species assignment	Statistical support	Computation time	References
Reciprocal monophyly		Monophyly-based	Reciprocal monophyly	Monophyletic species concept	Tree	One or more than one	None	No	Bootstrap support or posterior probability	*	Mishler (1985); Brower (1999)
Geneious Species Delimitation plugin (Masters et al., 2011)	Rosenberg's P_{AB}	Monophyly- and probability-based	Test of reciprocal monophyly	USC*	Tree	One	None	Yes	*p*-value	*	Rosenberg (2007)
	Rodrigo's *P (Randomly Distinct)*	Probability-based	Test of clade exclusivity	USC*	Tree	One	None	Yes	*p*-value	*	Rodrigo et al. (2008)
ABGD (Automatic Barcode Gap Discovery)		Distance threshold-based	Automatic barcoding gap detection using genetic distance	Genotypic clustering species concept (Mallet, 1995) /phenetic species concept	Sequence alignment (fasta) or distance matrix	One	P_{min} and P_{max} (minimum and maximum intraspecific distances), X (relative gap width), type of distance	No	No	*	Puillandre et al. (2012)
sGMYC (single-threshold General Mixed Yule-Coalescent)		Frequentist, phylogeny-based	Locating the transition threshold from coalescent to speciation (yule model) branching rates	GSC	Ultrametric tree	One	None	No	Likelihood ratio test and confidence interval	*	Pons et al. (2006)
bGMYC (Bayesian implementation of GMYC)		Bayesian, phylogeny-based	Bayesian implementation of GMYC integrating distributions of tree topology and branch lengths	GSC	Ultrametric trees	One	Posterior probability threshold	No	Posterior probability	***	Reid and Carstens (2012)
mPTP (multi-rate Poisson Tree Processes)		Frequentist, phylogeny-based	Difference of the number of substitutions between intraspecific coalescence and speciation applying multiple exponential distributions of substitutions	GSC	Bifurcating non-ultrametric tree	One	None	No	MCMC posterior probability/ average support values and confidence interval	*	Kapli et al. (2017)
BPP (Bayesian Phylogenetics and Phylogeography): A10 mode		Bayesian, phylogeny-based	rjMCMC sampling from the posterior distribution of multispecies coalescent model of species boundaries	BSC*, GSC	Sequence alignment (phylip), species assignment file, user-guide tree	More than one	Species divergence times (τ) and population size (θ)	Yes	Posterior probability	***	Yang and Rannala (2010); Rannala and Yang (2013)
spedeSTEM		Frequentist, phylogeny-based	Estimation of ML species trees using STEM for different scenarios of species boundaries	GSC	Ultrametric trees, species assignment file	More than one	Scaling factors and population size (θ)	Yes	Akaike Information Criterion	*	Ence and Carstens (2011)
tr2		Bayesian, phylogeny-based	Bayesian model comparison of posterior probabilities of rooted triplets' topological congruence under multispecies coalescent model	GSC	Ultrametric trees, user-guide tree	More than one	None	Yes	Posterior probability	**	Fujisawa et al. (2016)

(*continued on next page*)

Table 1 (*continued*)

Method	Approach[a]	Basis of method[a]	Species concept[b]	Input data	Number of loci[c]	*A priori* setting	*A priori* parameter	*A priori* species assignment	Statistical support	Computation time	References
ITS2 secondary structure	Character-based	Examining the folding structure and determination of compensatory base changes (CBCs) and hemi-CBCs	Typological species concept	Sequence alignment	One	None		No	No	*	Müller et al. (2007); Schultz and Wolf (2009)

[a] Expanded from Choi (2016) and Mallo and Posada (2016).

[b] BSC; biological species concept (Mayr, 1942); GSC; genealogical species concept (Baum and Shaw, 1995; within the phylogenetic species concept *sensu* de Queiroz, 2007); USC; unified species concept (de Queiroz, 2007). Asterisks indicate that the species concept was mentioned in the original publication of the method. The lists of species concepts follow de Queiroz (2007) and Zachos (2016).

[c] Number of loci the program was developed to analyze.

monophyletic species were subsequently validated by comparing them with the results from different species delimitation methods, which evaluate the PSP against the species concept(s) that underly the species delimitation methods (Sites and Crandall, 1997).

2.5. Validation of PSP using species delimitation analyses

The species delimitation methods used in this study are summarized in Table 1. Here (below) we provide some essential methodological background. The clusters of specimens delimited by these analyses were regarded as MOTUs and were compared with the monophyletic species as PSP in the following integrative taxonomic workflow (see below).

2.5.1. Geneious Species Delimitation Plugin (SDP) and Automatic Barcode Gap Discovery (ABGD)

The SDP (Masters et al., 2011) in Geneious v. 10.2.3 (Kearse et al., 2012) was applied to the ML and BI trees constructed from the full dataset of each gene to calculate *Rosenberg's* P_{AB} (Rosenberg, 2007) and *Rodrigo's P(Randomly Distinct) [Rodrigo's P(RD)]* (Rodrigo et al., 2008) for each monophyletic species. *Rosenberg's* P_{AB} tests the null hypothesis that reciprocal monophyly is due to a random coalescent process, not a speciation event, while *P(Randomly Distinct)* tests the null hypothesis that the 'distinctiveness' of a clade is a chance result of the coalescent process acting on a single, panmictic population of constant size. The smaller the probability, the less likely that the distinctiveness of a clade is the consequence of random coalescence. The sequential Bonferroni correction was applied to accommodate multiple testing bias.

The ABGD software automatically detects the threshold between intra- and interspecific divergence (barcode gap) in single locus data and uses that threshold to partition the data recursively into clusters that are interpreted as species (Puillandre et al., 2012). Thus, ABGD circumvents the drawback of establishing an *ad hoc* limit between intra- and interspecific divergence, as in other predefined distance thresholds, and the 10× rule (Hebert et al., 2004; Smith et al., 2005). To apply ABGD, sequence alignments of the full dataset of each gene were uploaded into the ABGD web server (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) and both initial and recursive partitions were reported as suggested by Puillandre et al. (2012).

Three parameters critically affect ABGD results: (1) the prior maximum divergence of intraspecific diversity (P_{max}), (2) the relative gap width (X) and (3) the type of genetic distance. Thus, we applied ABGD to each of the four genes with the following P_{max} settings: 0.01–0.13 in COI, 0.010–0.065 in 16S, 0.002–0.040 in ITS1 and 0.002–0.030 in ITS2, where the range of values coincide with the range of intraspecific distances of each gene dataset (Table S3). The default X value is 1.5 and we also applied the values 1.0 and 1.25, as lower values may yield more species (Jörger et al., 2012). All three genetic distances specified by the program were used: Jukes-Cantor 69 (JC69), Kimura-2-Parameter (K2P) and uncorrected *p*-distances. The barcode gap of each gene was retrieved as the distance value at the steepest slope position of the ranked pairwise distance graph (Puillandre et al., 2012).

2.5.2. General Mixed Yule-Coalescent (GMYC), Bayesian implementation of GMYC (bGMYC) and multi-rate Poisson Tree Processes (mPTP)

Rather than merely relying on a genetic distance threshold, the GMYC and mPTP take phylogenetic relationships into account and define the threshold between neutral coalescent branching within species and speciation branching between species. Then, GMYC applies this threshold to a time-calibrated ultrametric tree in which the branching patterns younger than this threshold involve the coalescence of the same species (Pons et al., 2006). In contrast, mPTP infers species boundaries by comparing the numbers of substitutions per site due to intraspecific coalescence and those due to speciation, without using time calibration (Kapli et al., 2017).

The GMYC can apply single (sGMYC; Pons et al., 2006) and multiple (mGMYC; Monaghan et al., 2009) time thresholds to delimit species,

and requires unique haplotypes because zero length terminal branches and polytomies affect the likelihood estimation (Fujisawa and Barraclough, 2013). Therefore, unique haplotype datasets were generated in ALTER (Glez-Peña et al., 2010), and were then used in all subsequent species delimitations. Following Hendrixson et al. (2013), we excluded the outgroup *Pontoscolex corethrurus* from the GMYC analyses of the full dataset because it was too divergent, and for consistency we then also excluded the outgroup for the analyses of the Mekong dataset. For each gene dataset, the appropriate substitution models were selected with jModelTest2 v. 2.1.6 on XSEDE (Darriba et al., 2012) in the CIPRES Science Gateway. Ultrametric trees were constructed in BEAST v. 1.8.4 (Drummond et al., 2012), as recommended by Tang et al. (2014), on XSEDE in the CIPRES Science Gateway to implement either a Yule process or a coalescent (constant size) model for a tree prior with a random starting tree, and a lognormal uncorrelated relaxed clock. The COI data were either not partitioned or partitioned into 1st, 2nd and 3rd codon positions. Two runs of 250 million generations were executed for each dataset. We sampled parameters every 1000 generations and assessed the convergence in TRACER v. 1.6 (Rambaut et al., 2014). Next, LOGCOMBINER v. 1.8.4 (Drummond and Rambaut, 2007) was used to combine the two runs and resample tree files every 5000 generations, while TREEANNOTATOR v. 1.8.4 (Drummond and Rambaut, 2007) was used to construct the maximum clade credibility trees using median heights as node heights, after discarding the first 10% of the trees as burn-in. Marginal likelihood estimation for each tree model used in each gene dataset was performed by path sampling and stepping stone sampling conducted in 100 steps. An MCMC chain length of 1,000,000 with a burn-in of 100,000 for each step was sufficient to reach stability. Bayes factors (BF) were calculated and log BF values were used to indicate the strength of support for competing models following Ritchie et al. (2017). The final ultrametric trees were entered into the R package SPLITS v. 1.0–19 (Ezard et al., 2009) and were analyzed with only sGMYC, since mGMYC often overestimates the number of delimited species (Fujisawa and Barraclough, 2013; Blair and Bryson, 2017).

We applied the bGMYC (Reid and Carstens, 2012) which can accommodate errors in the model estimates, and uncertainty of tree topologies and branch lengths that were not accounted for in GMYC. The tree outputs from each BEAST run were resampled every one million generations, following a 10% burn-in in LOGCOMBINER. The sampled output trees (237–251 trees) were analyzed with the R package bGMYC v. 1.0.2 using the default settings specified in the manual (http://nreid.github.io/assets/bGMYC_instructions_14.03.12.txt) and run for 50,000 generations, discarding the first 40,000 generations as burn-in, and using a thinning interval of 100. The function "bgmyc.point" was used to cluster individuals that were considered conspecific under the specified probability threshold. The higher the value of this probability threshold is, the larger the number of species retrieved. We specified five conspecificity probability thresholds (P_{con}): 0.05, 0.25, 0.50, 0.75 and 0.95, and constructed a box plot of those results.

The mPTP v. 0.2.3 algorithm (Kapli et al., 2017) is an improvement to PTP and bPTP (Bayesian implementation of PTP; Zhang et al., 2013) and likely gives more accurate results (Blair and Bryson, 2017). We applied mPTP to strictly bifurcating ML and BI trees for each gene. Unlike PTP and bPTP, mPTP requires strictly bifurcating trees. Yet, MrBayes returned multifurcating trees and so, we transformed the BI results into strictly bifurcating trees using the function "multi2di" in the R Package APE v. 5.0 (Paradis et al., 2004). The analyses were performed with 50 million generations, sampled every 1000 generations and discarding one million generations as burn-in. For datasets without an outgroup, mPTP will automatically root an input tree on the longest branch. Yet, for datasets with an outgroup, mPTP requires an outgroup-rooted input tree, as entering an unrooted tree and then using the tree-rooting function to root the tree with a specified taxon overestimated the species number (not shown).

2.5.3. Multilocus multispecies coalescent (MSC)-based species delimitations

The MSC methods using multilocus data incorporate the discrepancies among gene trees due to incomplete lineage sorting and do not require reciprocal monophyly of any species in focus within every gene tree (Fujita et al., 2012; Mallo and Posada, 2016). We applied three MSC methods with different underlying algorithms (see Table 1), viz. Bayesian Phylogenetics and Phylogeography (BPP; Rannala and Yang, 2003), spedeSTEM (Ence and Carstens, 2011) and the tr2 python script (Fujisawa et al., 2016). The A10 mode (species delimitation = 1, species tree = 0) of BPP v. 3.3a was selected and we entered the topology of a BI concatenated gene tree as an *a priori* guide tree of the phylogenetic relationship among the monophyletic species. Within each replicate, the BPP analyses used a total of 60,000 samples. The analyses were performed for five replicates with a burn-in of 10,000 generations.

The two most important parameters affecting the posterior probabilities of speciation events are the species divergence time (τ) and the population size parameter, which is estimated from θ, a parameter for genetic diversity (Rannala and Yang, 2003). We analyzed the dataset with four combinations of τ and θ in order to examine the effect of these parameter values on the number of delimited species. The prior combinations were: (1) moderate θ and τ [θ = G(2, 50), τ = G(2, 100)], (2) small θ and τ [θ = G(2, 200), τ = G(2, 2000)], (3) large θ and τ [θ = G(1, 10), τ = G(1, 10)], and (4) large θ and small τ [θ = G(1, 10), τ = G(2, 2000)] (Leaché and Fujita, 2010; Ortiz and Francke, 2016). Each analysis with the same algorithm was run three times to ensure consistency among runs (Yang, 2015). To decide whether a speciation event is significant, we followed McKay et al. (2013) by both assessing the 0.95 probability cutoff recommended by Leaché and Fujita (2010), and a less conservative interpretation in which we adopted a minimum cutoff value of 0.75.

SpedeSTEM requires *a priori* species assignment and an estimate of $\theta = 4N_e\mu$ for all loci in order to scale the species trees' branch lengths. Both ML and bifurcated BI gene trees of all four genes were used as different input trees in each analysis. We calculated the average Watterson's θ across all loci in DnaSP and prepared the scaling factors file as instructed in the spedeSTEM manual (https://carstenslab.osu.edu/Software_files/spedeSTEM.tutorial.pdf). The PSP of monophyletic species was used as *a priori* species assignment. The "discovery" mode of the program spedeSTEM v. 2.0 was used to generate the output of species delimitation. Applying tr2 also requires multiple gene trees as input and *a priori* species assignment via a user-specified guide tree. Similar to spedeSTEM analyses, both ML and bifurcating BI gene trees of all four genes were used as different input in each analysis. However, similar to BPP, tr2 requires *a priori* guide tree and the topology of the BI concatenated gene tree was used.

2.5.4. ITS2 secondary structure

The secondary structure of the nuclear ribosomal ITSs has long been used as another type of molecular character for delimiting species (reviewed in Schultz and Wolf, 2009). The ITS2, 5.8S and 28S flanking regions were annotated in the ITS2-Annotation module (Keller et al., 2009) at http://its2.bioapps.biozentrum.uni-wuerzburg.de/ (Koetschan et al., 2010). The most frequent ITS2 haplotype from each monophyletic species was entered into the RNA folding module of the mfold webserver (Zuker, 2003), using the default settings. The pairing between 5.8S and 28S flanking regions was constrained as this improved the accuracy of the folding algorithms (Morgan and Blair, 1998). The program yielded the optimum folding result with the lowest minimum free energy (ΔG). Secondary structures were checked for congruent patterns and used to align ITS2 sequences in 4SALE v. 1.7 (Seibel et al., 2006). Both compensatory base changes (CBCs: the two nucleotides of a paired site change, while maintaining the pairing) and hemi-CBCs (only a single nucleotide of a paired site changes) were compared between different species by aligning ITS2 secondary structures using fixed references of single-stranded regions between consecutive helices (Caisová et al., 2013). The CBC and hemi-CBC comparison matrices

were calculated using CBCAnalyzer v. 2.0.0 (Wolf et al., 2005).

2.6. Workflow of integrative taxonomy

We applied the candidate species approach following Vieites et al. (2009) and Padial et al. (2010) as a work protocol for integrative taxonomy to decide on the species diversity of Southeast Asian *Glyphidrilus*. All 33 monophyletic species identified as PSPs were applied to this work protocol by comparison with MOTUs delimited by the previous species delimitation methods, and were classified into different categories of species hypotheses, viz. confirmed candidate species (CCS), unconfirmed candidate species (UCS) and deep conspecific lineage (DCL) (Vieites et al., 2009; Padial et al., 2010). The CCSs were identified if at least three out of four of the following criteria were met: (1) reciprocal monophyly and species delimitation results of ITS1 and ITS2, (2) distinct ITS2 secondary structure, (3) species delimitation results of multilocus MSC methods and (4) consistent morphological differences. The species delimitation results of the COI and 16S sequences were not included in these criteria as most of the results were similar to the monophyletic species. Monophyletic species which meet only the criteria of distinct ITS2 secondary structure and/or species delimitation results of multilocus MSC methods are regarded as UCSs, while monophyletic species that fail to meet any criterion are regarded as DCLs (Fig. 2).

3. Results

DNA sequence features and polymorphism estimates for each of the four genes are reported in Table S4.

3.1. Morphological identification, phylogenies and reciprocal monophyly as PSPs

Using the reciprocal monophyly criterion, all 285 specimens were clustered into 33 monophyletic species as PSP from the COI gene tree (Fig. S2). Of the 186 out of the 285 *Glyphidrilus* specimens that were assigned to 31 nominal morphospecies, they were clustered into only 19 monophyletic species. The remaining specimens could not be assigned to any existing morphospecies and were clustered into 14 monophyletic species. The morphological characters of the specimens in this study are provided in Table S1. Some *Glyphidrilus* spp. which showed overlapping characters, or which represented the variation within a species complex, are illustrated in Fig. 1.

The 33 monophyletic species also appeared in the 16S and the concatenated gene trees, except that *Glyphidrilus* sp. 6 was paraphyletic and divided into two clades (Figs. 3 and S3). The Mekong clade was supported by all analyses of the concatenated gene tree (Fig. S4a–c), BI analysis of ITS1 gene tree (Fig. S5c) and all analyses of ITS2 gene tree (Fig. S5d–f). All remaining species from the Chao Phraya River, northern and western Thailand, Myanmar and the Malay Peninsula, are referred to as the non-Mekong clades (Fig. 3). Seventeen and 16 monophyletic species were found to belong to the Mekong and non-Mekong clades, respectively. Topological variations in the Mekong clade and support values among the MP/ML/BI concatenated gene trees are shown in Fig. S4a–c. Within the Mekong clade, we distinguished four species complexes, all of which were supported by ML and BI support values of the concatenated gene tree (Fig. S4b and c), viz. *G. jamiesoni*, *G. yunnanensis*, *G. mekongensis* and *G. chiensis* (Table 2). *Glyphidrilus sekongensis* and *Glyphidrilus* sp. 8 were kept separate in the Mekong clade because they showed no supported affiliations with any species complex (Figs. 3 and 4). Topological variations in the Mekong clade and support values from all tree construction methods for the COI and 16S gene trees are shown in Fig. S4d and e.

Five clades in the concatenated, COI and 16S gene trees were comprised of specimens of different morphospecies and within each clade some specimens from different morphospecies shared the exact mitochondrial haplotypes (Figs. 3, S2 and S3). The morphology of the different morphospecies within those five clades along with those within each species complex was re-examined and are shown in Fig. 1. Seven unnamed *Glyphidrilus* species (spp. 6–10, 12 and 13) were morphologically similar to *G. chiensis/quadratus/nanensis* in the positions of wings, genital markings and spermathecae (Fig. 1b).

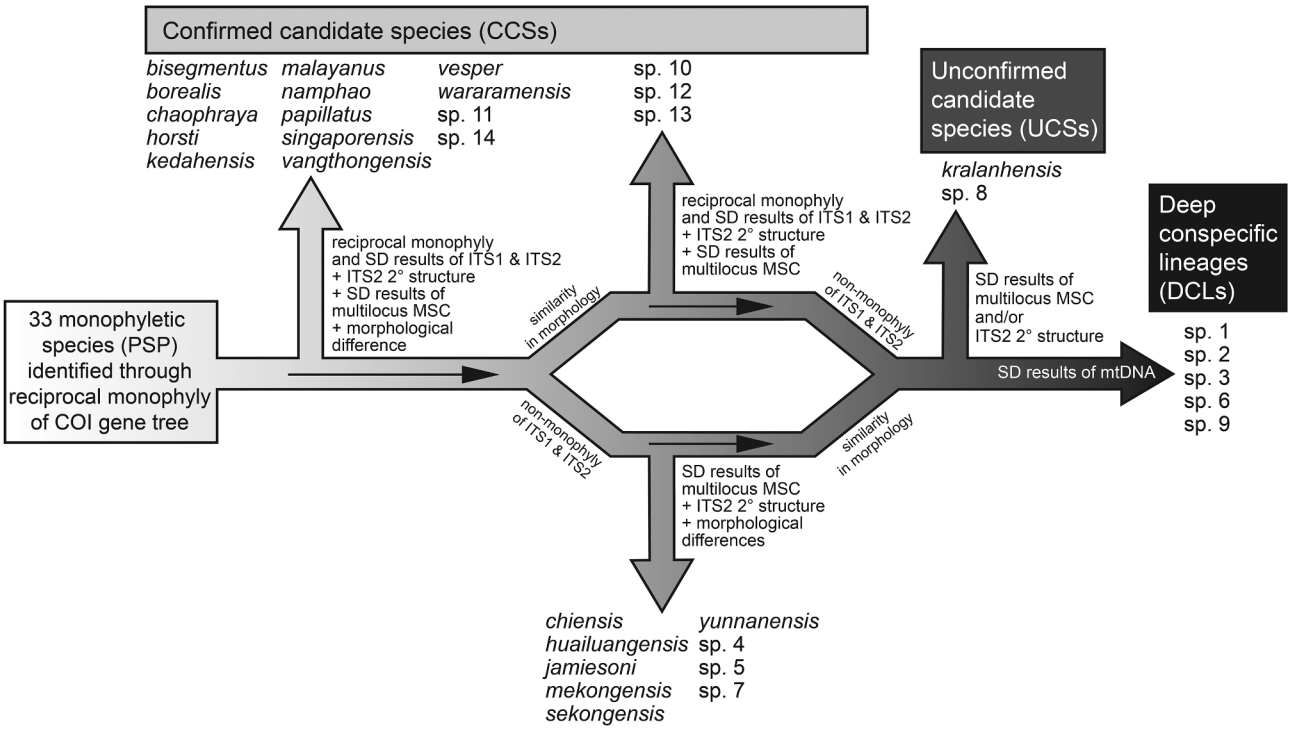

Fig. 2. Schematic workflow in integrative taxonomy of Southeast Asian *Glyphidrilus* in this study modified from the candidate species approach following Vieites et al. (2009) and Padial et al. (2010). SD: species delimitation, MSC: multispecies coalescent-based methods.

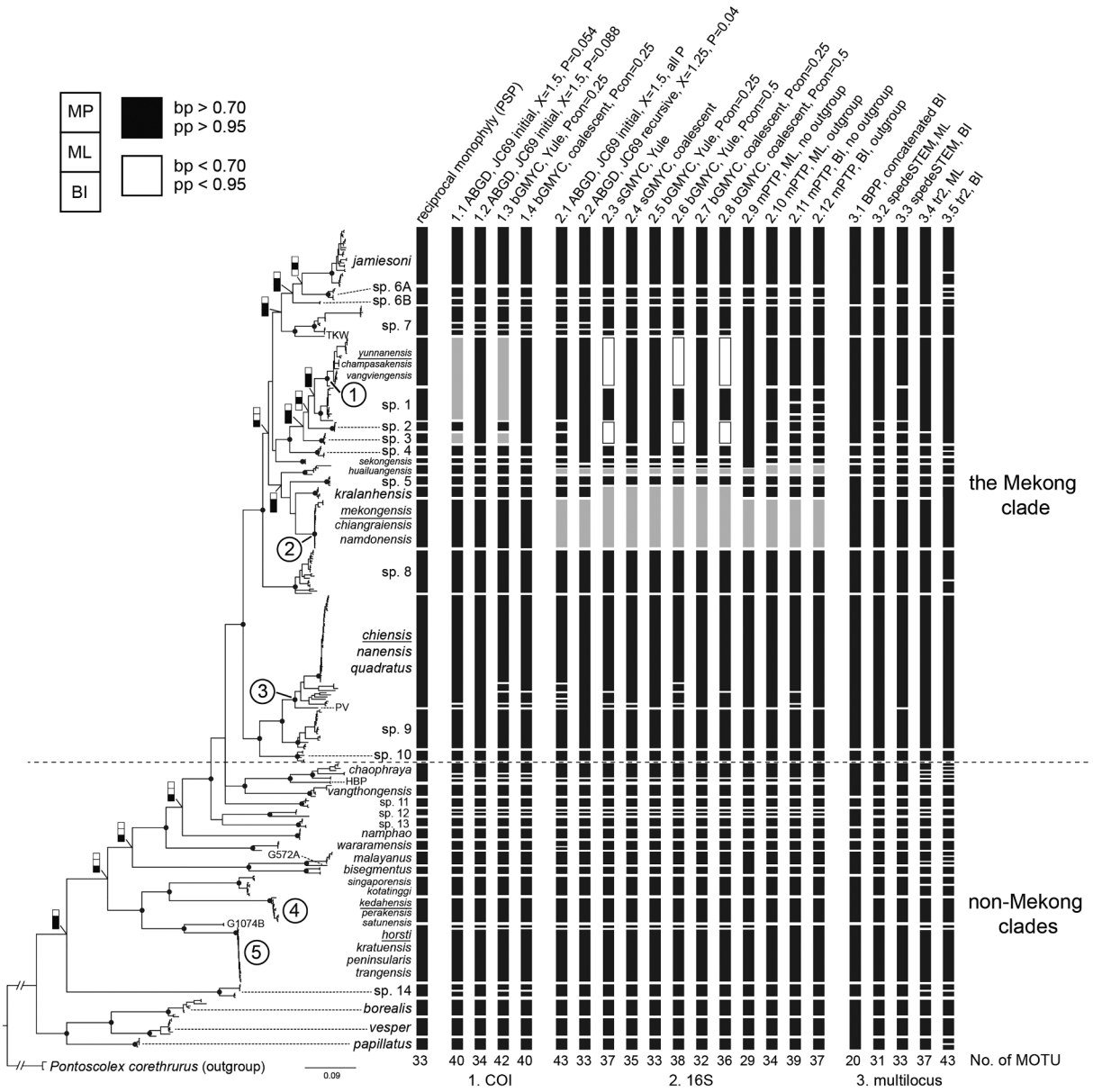

Fig. 3. BI concatenated tree of the full dataset and species delimitation results of 1. COI, 2. 16S rDNA and 3. multilocus methods. The nodes with MP and ML bootstraps > 0.70 and BI posterior probability of > 0.95 are indicated with black circles. The support values of the remaining nodes are shown as a three-tile square positioned next to each node. Grey or white bars within the same columns belong to the same clustering. Column number is located on the top of each column, indicates the result from the analysis mentioned in the text. MOTU number obtained from each analysis is located under each column. Numbers in the circle indicate the clades comprising specimens of different morphospecies, some of which share the same mitochondrial haplotype. The oldest available species names in those clades are underlined. P: prior maximum divergence of intraspecific diversity; X: relative gap width; Pcon: conspecificity probability.

Comparing the PSP to both the ITS1 and ITS2 gene trees confirmed all the monophyletic species in the non-Mekong clades. However, in the Mekong clade only six monophyletic species (*G. huailuangensis*, *G. sekongensis*, *Glyphidrilus* sp. 5, sp. 7, sp. 8 and sp. 10) were supported in the ITS1 gene tree, while three monophyletic species (*Glyphidrilus* sp. 5, sp. 8 and sp. 10) were supported in the ITS2 gene tree (Fig. 4). Topological differences in the *Glyphidrilus* monophyletic species and support values from all methods between ITS1 and ITS2 gene trees are shown in Fig. S5.

3.2. Results of the SDP, ABGD and barcode gap analyses

All the SDP results are reported in Table S5. *Rosenberg's* P_{AB} and *Rodrigo's P(RD)* could neither be calculated for *G. horsti* and *Glyphidrilus* sp. 12, nor for *G. huailuangensis* and *G. mekongensis* as they were not monophyletic in the COI and 16S trees, respectively. The significances of *Rosenberg's* P_{AB} and *Rodrigo's P(RD)* are shown in Table S5, and the results of species delimited before and after applying a sequential Bonferroni correction in each analysis are shown in Table S6. Monophyletic species with high *Intra/Inter* ratios are recognizable in Fig. 5, where they show up close to the threshold lines at which the maximum intraspecific distance and the minimum distance to the nearest neighbor are equal. Monophyletic species near these threshold lines in the COI and 16S plots were likely to be split in subsequent species delimitation analyses (Fig. 5a and b). In the ITS1 + ITS2 plot, the maximum intraspecific distances of the Mekong species were higher than the minimum distances to their nearest neighbors (Fig. 5c).

The ABGD relative gap widths of $X = 1$ and 1.25 yielded one MOTU in some analyses of COI and 16S (Table S7). The JC69 distances in COI yielded the smallest numbers of MOTUs, while *p*-distances produced the

largest numbers. Similarly, JC69 yielded the smallest numbers of MOTUs in 16S, while both K2P and *p*-distances produced larger numbers of MOTUs. The effect of the distances was less pronounced with ITS1 and ITS2, except for *p*-distances in ITS2, which produced markedly smaller numbers of MOTUs.

Most ABGD results with K2P and *p*-distances for COI and 16S yielded two to three times as many MOTUs as the number of monophyletic species. Most results based on the JC69 distances of COI and 16S yielded 40 and 43 MOTUs, respectively (Fig. S6a and b). These delimitations corresponded to most monophyletic species, but ABGD with the COI and 16S datasets compared with the monophyletic species did not always delimit the species in the same way (Fig. 3, columns 1.1, 1.2, 2.1 and 2.2). In contrast, ABGD analyses applied to ITS1 and ITS2 usually yielded 20 and 17 MOTUs, respectively, (Fig. S6c and d), which was fewer than the number of monophyletic species. This is due to the Mekong clade which is regarded by ABGD as only one or two MOTUs (Fig. 4, columns 4.1 and 5.1). In addition, recursive partitions within ITS1 and ITS2 did not provide evidence for a further subdivision of the Mekong clade (Fig. 4, columns 4.2 and 5.2).

The COI analysis consistently yielded the highest K2P interspecific distance values, followed by 16S, while ITS1 and ITS2 showed much lower, but mutually similar values (Fig. S7). The barcode gap from ABGD analyses of COI K2P distances is approximately 0.09–0.10, less than both the maximum intraspecific and minimum interspecific distances (Fig. S8a and b). The K2P barcode gap of 16S was 0.04–0.07, more than the minimum interspecific distances, while the maximum intraspecific distance was within range of this barcode gap (Fig. S8c and d). In contrast, the steepest slopes of the ranked pairwise distance graphs of ITS1 and ITS2 were not prominent, so the barcode gaps of those genes were not inferred (Fig. S8e–h).

3.3. Results of the GMYC, bGMYC and mPTP analyses

Most GMYC analyses with the full dataset, especially those involving COI, yielded more MOTUs than monophyletic species (Fig. 6a). The BFs calculated between different analyses of competing tree models of each gene dataset are shown in Table S8, indicating that for all gene datasets the use of a coalescent tree prior provided a better fit, except for ITS1 and ITS2 in the Mekong dataset. In addition, a coalescent tree prior with codon partitioned was the best fit for the COI dataset. The sGMYC using a BEAST tree with a Yule prior generated more MOTUs than a coalescent prior did in the partitioned COI, ITS1 and ITS2, while

Table 2
Conclusive list of mainland SE Asian *Glyphidrilus* confirmed candidate species (CCSs) and unconfirmed candidate species (UCSs) delimited under integrative taxonomy in this study. The nominal (morpho)species and monophyletic species which are regarded as deep conspecific lineages (DCLs) and belong to the same clade are indicated with an asterisk.

Affiliation	Species delimited in this study	Type locality	Distribution range
Mekong clade	**Confirmed candidate species**		
	***G. jamiesoni* species complex**		
	1. *G. jamiesoni* Jirapatrasilp, Chanabun and Panha in Jirapatrasilp et al., 2016	Stream near Praduk Temple, Siem Riep, Cambodia	Northeastern Thailand, Northern Lao P.D.R. and Siem Riep, Cambodia
	**Glyphidrilus* sp. 6		
	2. *Glyphidrilus* sp. 7	–	Eastern Thailand
	***G. yunnanensis* species complex**		
	3. *G. yunnanensis* Chen and Xu, 1977	Luosuo River, Menglun, Yunnan, P.R. China	Mekong River, its tributaries in Yunnan, Thailand and Lao P.D.R., and Ping River, Thailand
	**G. vangviengensis* Panha and Chanabun in Chanabun et al., 2011		
	**G. champasakensis* Chanabun and Panha in Chanabun et al., 2017		
	**Glyphidrilus* sp. 1		
	**Glyphidrilus* sp. 2		
	**Glyphidrilus* sp. 3		
	4. *Glyphidrilus* sp. 4	–	Songkhram River, Bueng Kan, Thailand
	***G. mekongensis* species complex**		
	5. *G. mekongensis* Panha and Chanabun in Chanabun et al., 2012a	Toh Sang Hotel, Mekong River, Ubon Ratchathani, Thailand	Mekong River and its tributaries in Lao P.D.R.
	**G. chiangraiensis* Chanabun and Panha in Chanabun et al., 2017		
	**G. namdonensis* Chanabun and Panha in Chanabun et al., 2017		
	6. *G. huailuangensis* Chanabun and Panha in Chanabun et al., 2013	Huailuang Waterfall, Ubon Ratchathani, Thailand	Only from the type locality
	7. *Glyphidrilus* sp. 5	–	Mekong and Mun Rivers, Ubon Ratchathani, Thailand
	***G. chiensis* species complex**		
	8. *G. chiensis* Chanabun and Panha in Chanabun et al., 2013	Rice field near Chi River at Ban Thatoom, Maha Sarakham, Thailand	Central and Northeastern Thailand, and Salavan, Lao P.D.R.
	**G. quadratus* Chanabun and Panha in Chanabun et al., 2013		
	**G. nanensis* Chanabun and Panha in Chanabun et al., 2017		
	**Glyphidrilus* sp. 9		
	9. *Glyphidrilus* sp. 10	–	Eastern Thailand
	Not in any species complex		
	10. *G. sekongensis* Chanabun and Panha in Chanabun et al., 2017	Stream at Ban Kiangkong, Salavan, Lao P.D.R.	Only from the type locality
	Unconfirmed candidate species		
	1. *G. kralanhensis* Jirapatrasilp, Chanabun and Panha in Jirapatrasilp et al., 2016 (within *G. mekongensis* species complex)	Tapan River, Kralanh, Siem Riep, Cambodia	The type locality and Wang River, Tak, Thailand
	2. *Glyphidrilus* sp. 8 (not in any species complex)	–	Northeastern Thailand, Lao P.D.R., and Siem Riep, Cambodia

(*continued on next page*)

Table 2 (*continued*)

Affiliation	Species delimited in this study	Type locality	Distribution range
Non-Mekong clades	**Confirmed candidate species**		
	1. *G. chaophraya* Chanabun and Panha in Chanabun et al., 2013	Chao Phraya River, Nakhon Sawan, Thailand	Central Thailand
	2. *G. vangthongensis* Chanabun and Panha in Chanabun et al., 2013	Sakunothayan Waterfall, Phitsanulok, Thailand	Khaek River and its tributaries in Phitsanulok, Central Thailand
	3. *Glyphidrilus* sp. 11	–	Mekong River, Champasak, Lao P.D.R.
	4. *Glyphidrilus* sp. 12	–	Chainat and Saraburi, Central Thailand
	5. *Glyphidrilus* sp. 13	–	Khao Nang Panthurat Forest Park, Phetchaburi, Thailand
	6. *G. namphao* Chanabun and Panha in Chanabun et al., 2017	Phao River, Bolikhamsai, Lao P.D.R.	Only from the type locality
	7. *G. wararamensis* Chanabun and Panha in Chanabun et al., 2013	Sok Stream near Tham Wararam Temple, Surat Thani, Thailand	Surat Thani, Thailand
	8. *G. malayanus* Michaelsen, 1902	Lubock Paku, Pahang River	Malay Peninsula
	9. *G. bisegmentus* Chanabun and Panha in Chanabun et al., 2012b	Air Banun Pandig, Perak, Malaysia	Phuket, Southern Thailand and Perak, Malaysia
	10. *G. singaporensis* Shen and Yeo, 2005	Bukit Timah Nature Reserve, Singapore	Johor, Malaysia and Singapore
	**G. kotatinggi* Chanabun and Panha in Chanabun et al., 2012b		
	11. *G. kedahensis* Chanabun and Panha, 2015	Sedim River, Kedah, Malaysia	Satun, Southern Thailand, Kedah and Perak, Malaysia
	**G. perakensis* Chanabun and Panha, 2015		
	**G. satunensis* Chanabun and Panha in Chanabun et al., 2017		
	12. *G. horsti* Stephenson, 1930	Pulau Berhala, Straits of Malacca	Malay Peninsula
	**G. peninsularis* Chanabun and Panha in Chanabun et al., 2012b		
	**G. kratuensis* Chanabun and Panha in Chanabun et al., 2013		
	**G. trangensis* Chanabun and Panha in Chanabun et al., 2013		
	13. *Glyphidrilus* sp. 14	–	Mandalay and Magway, Myanmar
	14. *G. borealis* Chanabun and Panha in Chanabun et al., 2013	Mae Klang Waterfall, Chiang Mai, Thailand	Northern Thailand and Northern Lao P.D.R.
	15. *G. vesper* Chanabun and Panha in Chanabun et al., 2013	Thi Lo Su Waterfall, Tak, Thailand	Western Thailand
	16. *G. papillatus* (Rosa, 1890)	Cobapo, Burma	Myanmar

more similar results were obtained within the non-partitioned COI and 16S dataset. Compared to COI, the 16S results with both tree priors were more in line with the monophyletic species (Fig. 3, columns 2.3 and 2.4). For ITS1 and ITS2, only the results with coalescent priors were in line with the clades in the ITS1 and ITS2 gene trees (Fig. 4, columns 4.3 and 5.3). In the Mekong dataset of COI and 16S, all GMYC analyses yielded many more MOTUs (sometimes up to 3x) than the number of the monophyletic species (Fig. 6b). Although the GMYC analyses of ITS yielded the same (ITS2) or similar (ITS1) numbers of MOTUs as monophyletic species (Fig. 6b), the compositions of the MOTUs differed (not shown). Some analyses with the Mekong dataset received a lower statistical significance of likelihood ratio tests of the results compared to the full dataset (Fig. 6b).

The bGMYC analyses of the full dataset, using a coalescent prior, yielded a narrower range of MOTU numbers (P_{con} = 0.05–0.95) than with a Yule prior, except for with COI, which showed a small difference between both priors (Fig. 7a). Even though some bGMYC analyses of COI yielded similar MOTU numbers (40–42) as with ABGD, the two results differed in the composition of MOTUs in the Mekong clade (Fig. 3, columns 1.1, 1.3 and 1.4). The 16S analyses with P_{con} = 0.25–0.5 using both priors yielded similar MOTU numbers (32–38 species) and a similar MOTU composition to those of the monophyletic species (Fig. 3, columns 2.5–2.8). Most analyses of ITS1 and ITS2 returned fewer MOTUs than monophyletic species due to the lumping of some Mekong monophyletic species (Fig. 7a and 4, columns 4.4–4.6 and 5.4–5.6). For the Mekong dataset, the analyses with $P_{con} \geq 0.25$ in COI and 16S always yielded more MOTUs than monophyletic species and retrieved most of the monophyletic species in the Mekong clade (Fig. 7b and S9, columns 6.3 and 7.1). Conversely, most results from ITS1 and ITS2 yielded fewer MOTUs than monophyletic species (Fig. 7b and S9, columns 8.1 and 9.1).

The mPTP analyses of the full dataset of the ML and BI COI trees always yielded more MOTUs, while those of ITS1 and ITS2 always yielded fewer MOTUs than monophyletic species (Fig. 6c). In addition, the analysis of 16S trees yielded more or less MOTUs similar to the number of monophyletic species, yet the composition of MOTUs varied slightly (Fig. 3, columns 2.9–2.12). Comparing the mPTP analyses with and without the outgroup showed 1 or 2 MOTUs difference in the COI, 16S and ITS1 results (except in the 16S BI tree). For ITS2, both the analysis with and without the outgroup showed the same MOTU number but the composition of the MOTUs differed (Fig. 4, columns 4.8, 4.9, 5.8 and 5.9). For the COI and the 16S, the use of bifurcating BI trees returned more MOTUs than the ML (Fig. 6c), while with ITS1 and ITS2 the BI and ML results were more similar to each other (e.g. Fig. 4, columns 4.7 and 4.8). For the Mekong dataset, the mPTP analyses of 16S ML with the outgroup yielded 17 MOTUs, the composition of which was closest to that of the monophyletic species (Fig. S9, column 7.2). Yet, the analyses of the 16S ML without the outgroup and both BI with and without the outgroup yielded more MOTUs (Fig. S9, columns 7.3–7.5). In contrast, the analyses with other markers either produced more (COI) or far fewer (ITS1, ITS2) MOTUs (Fig. 6d).

3.4. Multilocus MSC species delimitation

Among the multilocus MSC methods, BPP delimited the smallest number (20) of MOTUs (Fig. 3, column 3.1), where BPP with the four combinations of θ and τ always returned the highest posterior probabilities (> 0.95) at all deep nodes. However, more recent nodes received lower posterior probabilities (0.11–0.88) for all combinations of θ and τ, especially the nodes of the more recently diverging sister species. There was no obvious trend in which combinations of θ and τ yielded higher posterior probabilities (Fig. 8). The spedeSTEM result from the BI gene trees yielded 33 MOTUs, which is the exact number of monophyletic species and showed the same species composition, while

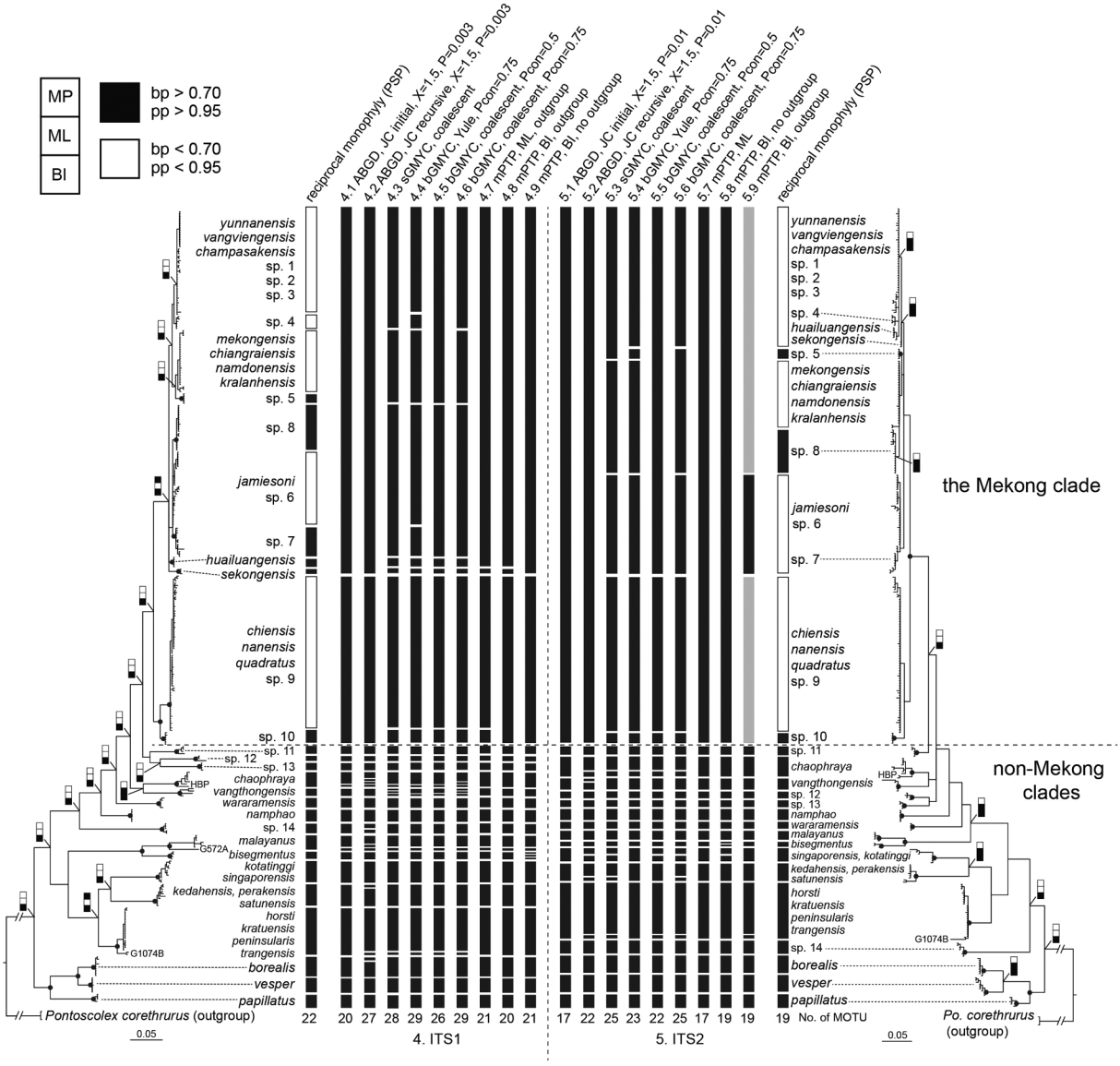

Fig. 4. Bayesian inference ITS1 (left) and ITS2 (right) gene trees of the full dataset and species delimitation results of both ITSs. The legend follows Fig. 3.

the result from ML yielded slightly fewer (31) MOTUs (Table S9; Fig. 3, columns 3.2 and 3.3). In contrast, using ML and BI gene trees in the tr2 analysis yielded 37 and 43 MOTUs, respectively, which were more than the number of monophyletic species (Fig. 3, columns 3.4 and 3.5).

3.5. ITS2 secondary structure

The common pattern of the secondary structure of *Glyphidrilus* ITS2 consisted of three helices (IB, II and III) radiating from a central ring, with helix III branching into two stems (IIIA and IIIB; Fig. 9a). This pattern was present in all the Mekong monophyletic species. However, some species in the non-Mekong clades exhibited additional helices, such as helix IA and helix IV, with helix IV mostly found in the Malay Peninsula species. Five monophyletic species showed an unbranched helix III, none of which were phylogenetically closely related. One invariable motif 5′-CGGCMGGAGCGRRGYGGCGAG-3′ occurred at the 5′ side near the tip of helix III (or the tip of helix IIIB in the case of a branched helix III; Fig. 9a).

Examples of variation in helix II are shown in Figs. 9 and S10. Overall the secondary structure of ITS2 did not differentiate all of the Mekong monophyletic species, but the stem pattern of the helix II in the Mekong clade distinguished three species groups (Fig. 9b–d). The loop section of helix II showed a greater level of variation, while the stem section revealed a few species-specific hemi-CBCs. There is one CBC in helix II between *G. yunnanensis* and *Glyphidrilus* sp. 4, both of which belong to the *G. yunnanensis* species complex. In contrast, helix II of the non-Mekong monophyletic species showed more differences in terms of stem length and pattern, with the longest stem being found in *G. bisegmentus* (Fig. S10). Two and one hemi-CBC(s) distinguished *G. chaophraya* HBP and *G. horsti* 1074B from the other *G. chaophraya* and *G. horsti* specimens, respectively. Although most species had their own unique ITS2 secondary structure, some monophyletic species pairs had identical helix II patterns, such as *G. huailuangensis* and *Glyphidrilus* sp. 3, and *G. borealis* and *G. vesper*. More CBCs and hemi-CBCs were found in comparisons among the non-Mekong monophyletic species (Fig. S10), where species could differ by up to five CBCs (Table S10). Although *G. borealis* and *G. vesper* had an identical helix II, they differed by two CBCs at helix III (not shown).

3.6. Summary of the species delimitation results and the integrative taxonomic workflow

All species delimitation results are summarized in Table S11 and comparisons between the morphological identification and molecular

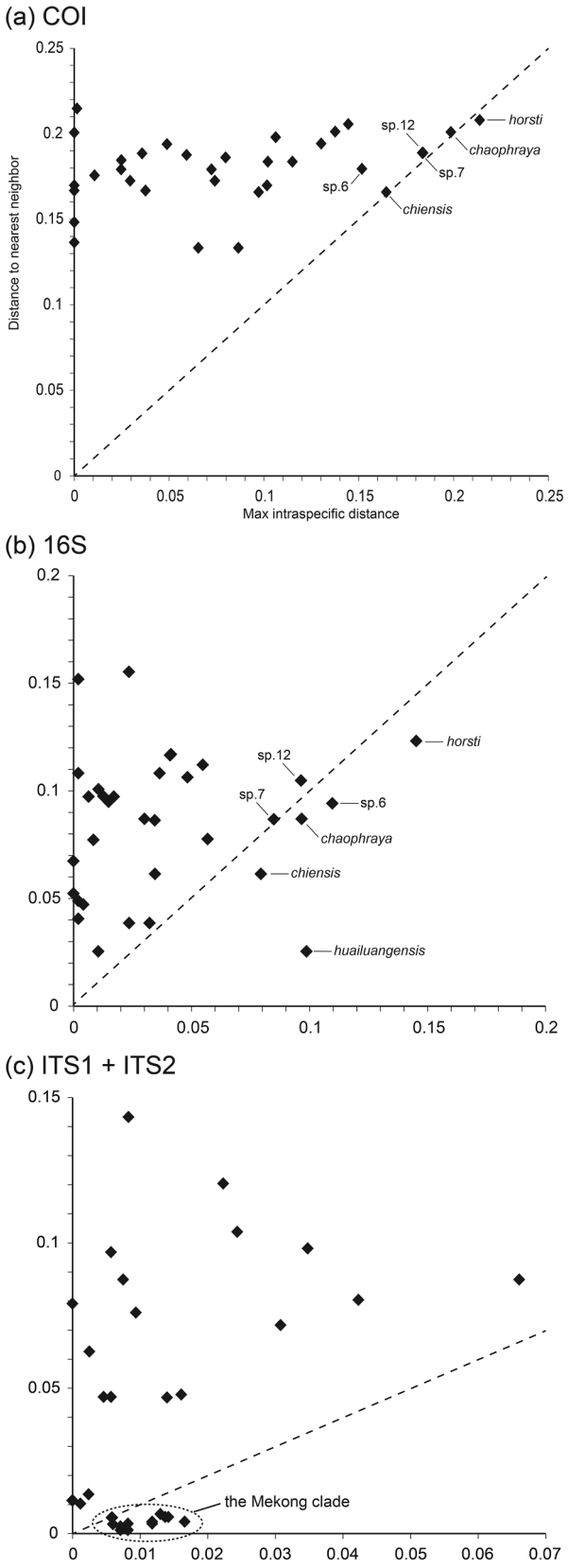

Fig. 5. Plot comparisons of distances estimated from the best-fit model between maximum intraspecific and minimum interspecific to the nearest neighbor in (a) COI, (b) 16S rDNA and (c) ITS1 + ITS2. Dashed lines indicate the threshold where the maximum intraspecific distances and the minimum interspecific distances to the nearest neighbor are equal. Monophyletic species with high *Intra/Inter* ratios are labeled in the graphs.

species delimitation are summarized in Table S12. Overall, 13 nominal (morpho)species were not recognized but were nested within clades with older species names (Table 2). The results from the integrative taxonomic workflow are illustrated in Fig. 2. Fourteen out of 16 non-Mekong monophyletic species met all four criteria set out in section 2.6. Nine out of 17 Mekong monophyletic species met three criteria without reciprocal monophyly and species delimitation results of ITS1 and ITS2, while two non-Mekong (sp. 12 and sp. 13) and one Mekong species (sp. 10) met three criteria without any morphological differences. Hence, this results in a total of 26 CCSs. Two species, *G. kralanhensis* and *Glyphidrilus* sp. 8 met only two out of four criteria and are regarded as UCSs. Five unnamed *Glyphidrilus* species (spp. 1–3, 6 and 9) were only recognized and supported by species delimitation using mtDNA data but failed to meet any criterion in the workflow and are thus regarded as DCLs (Fig. 2).

4. Discussion

4.1. Comparison of different genetic markers

Even though DNA sequence data are commonly used in earthworm phylogenetics, only a few studies have applied them to delimit species (Jeratthitikul et al., 2017; Taheri et al., 2018). Our interspecific COI K2P distances were high compared to those in other earthworm studies (Table S3). Compared to the proposed thresholds (< 9% intraspecific and > 15% interspecific), the average COI interspecific distances among *Glyphidrilus* monophyletic (22.26%) and candidate species (20.74%) were > 15% and the average intraspecific distances among monophyletic (4.2%) and candidate species (3.59%) were < 9% (Table S3; Chang and James, 2011). However, the lowest interspecific (13.33%) and highest intraspecific distances of monophyletic (12.21%) and candidate species (11.45%) fell within the 9–15% range, so that the taxonomic interpretation of sequence divergences in this range may be difficult. In addition, if applying the COI barcode gap retrieved from ABGD (9–10%), which is lower than the average interspecific distance as a threshold to delimit species, this may overestimate the species number. Different COI barcode thresholds have been suggested, viz. 13% for earthworms in general (Jeratthitikul et al., 2017) and 9% for French lumbricids (Porco et al., 2018), yet these thresholds were clade-specific, not general (Novo et al., 2012b). Thus, it seems ill-advised to apply a single threshold over a highly diverse dataset, such as the one presented in this study.

Our 16S data yielded considerably higher inter- and intraspecific genetic distances than other earthworm studies (Table S3). Compared to COI, our 16S data showed a lower ratio of maximum intraspecific distances to minimum interspecific distances to the nearest neighbor, so that 16S often yielded fewer MOTUs, which were more in line with the monophyletic species. So while 16S appears to be a good species delimitation marker, it still does not yet achieve the same level of efficiency as COI, because (1) there are fewer 16S reference sequences of earthworms in databases like GenBank and (2) there is still no BOLD reference database for 16S (Klarica et al., 2012).

The plot between maximum intraspecific distances and minimum interspecific distances to the nearest neighbor helps to avoid the subjectivity of choosing an arbitrary threshold to differentiate different species. Most MOTUs shown in this plot are in line with the monophyletic species, except for some MOTUs in which the maximum intraspecific distances were larger than the minimum distances to the nearest neighbor (Fig. 5). High degrees of intraspecific COI sequence divergence within earthworm species have been reported previously to suggest the presence of cryptic species (Novo et al., 2010; Shekhovtsov et al., 2017). However, the analysis of mtDNA data alone could lead to inconclusive taxonomic decisions. For example, in the absence of nuclear DNA data, the COI results of Jeratthitikul et al. (2017) would not allow the distinction between the presence of cryptic species or conspecific deep diverging lineages. Alternatively, Taheri et al. (2018)

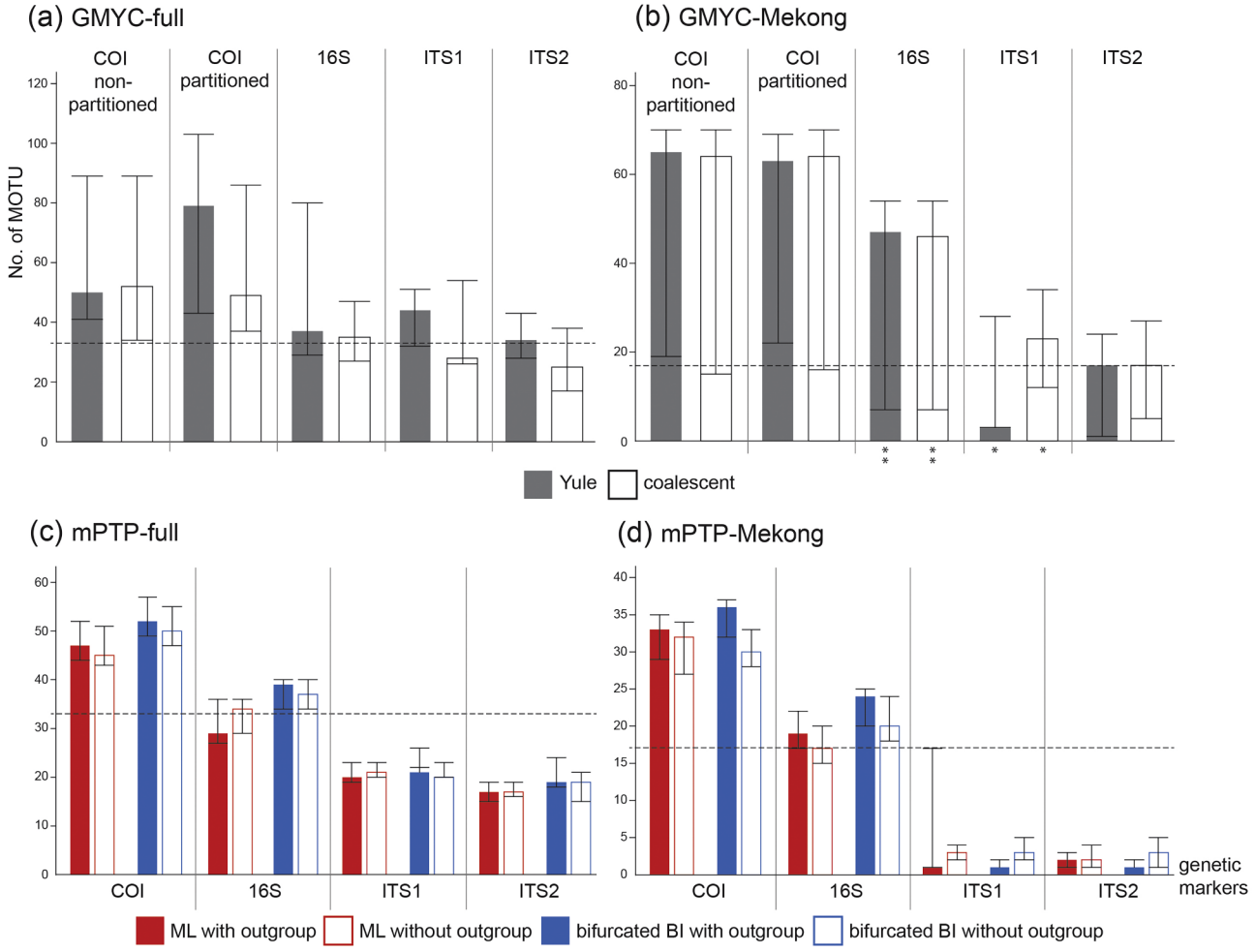

Fig. 6. General Mixed Yule-Coalescent (GMYC) and multi-rate Poisson Tree Processes (mPTP) results of (a, c) full and (b, d) the Mekong dataset. Dashed lines indicate the number of monophyletic species as PSP. All GMYC results were supported by statistical significance of likelihood ratio tests as specified from the program at $p < 0.001$ except ones with asterisks along the X-axis: $^{*}p < 0.05$, $^{**}p < 0.01$.

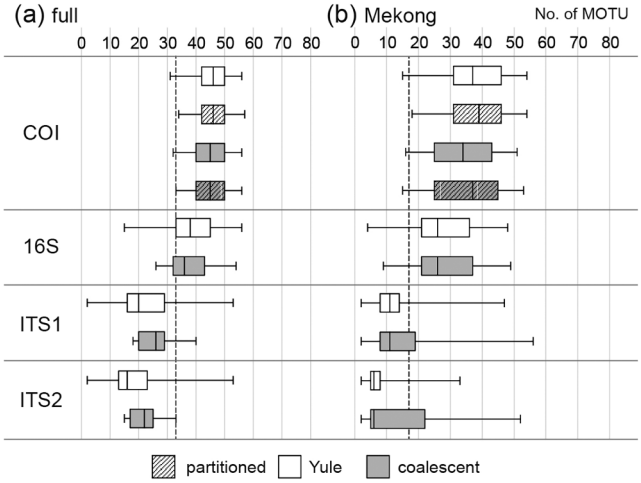

Fig. 7. Box plots of Bayesian implementation of GMYC results from (a) full and (b) the Mekong dataset. Dashed lines indicate the number of monophyletic species as PSP.

validated their COI results by conducting BPP with two nuclear loci. Thus, the results from mtDNA need confirmation from nuclear gene fragment(s).

The interspecific ITS distances among the Mekong species were on average lower than those among *Glyphidrilus* species from elsewhere (Table S3; Fig. 5). The only other earthworm study using ITS sequence data (Shekhovtsov et al., 2013) reported average interspecific distances for ITS among *Eisenia* species that were approximately two-fold lower than among *Glyphidrilus* species, while the average intraspecific distances in *Eisenia* were nearly similar to those in *Glyphidrilus* (Table S3). Neither of the ITS fragments showed well-defined barcode gaps in *Glyphidrilus*, which is in line with studies in other organisms (e.g. Wang et al., 2015; Ortiz and Francke, 2016) but see Klinth et al. (2017). Moreover, both ITS datasets provided limited resolution to recognize the Mekong monophyletic species.

4.2. Arbitrariness of data and parameter selection in species delimitation

Many sophisticated species delimitation methods have been developed recently, all of which are based on different methodological approaches and species concepts and designed for different types of input data and number of loci (Table 1). However, this study indicated that those methods are not without any problems, one of which is the arbitrariness of the parameter selection. In ABGD, lower values of P_{max} and X always yielded more MOTUs, (Khedkar et al., 2014; Schwarzfeld and Sperling, 2015), while the impact of genetic distances has been reported previously (Kekkonen et al., 2015; Cao et al., 2016). In our analyses, the use of *p*-distances always yielded more MOTUs than with K2P or JC69, and JC69 distances yielded MOTU numbers that were mostly in line with the monophyletic species. Thus, limiting DNA

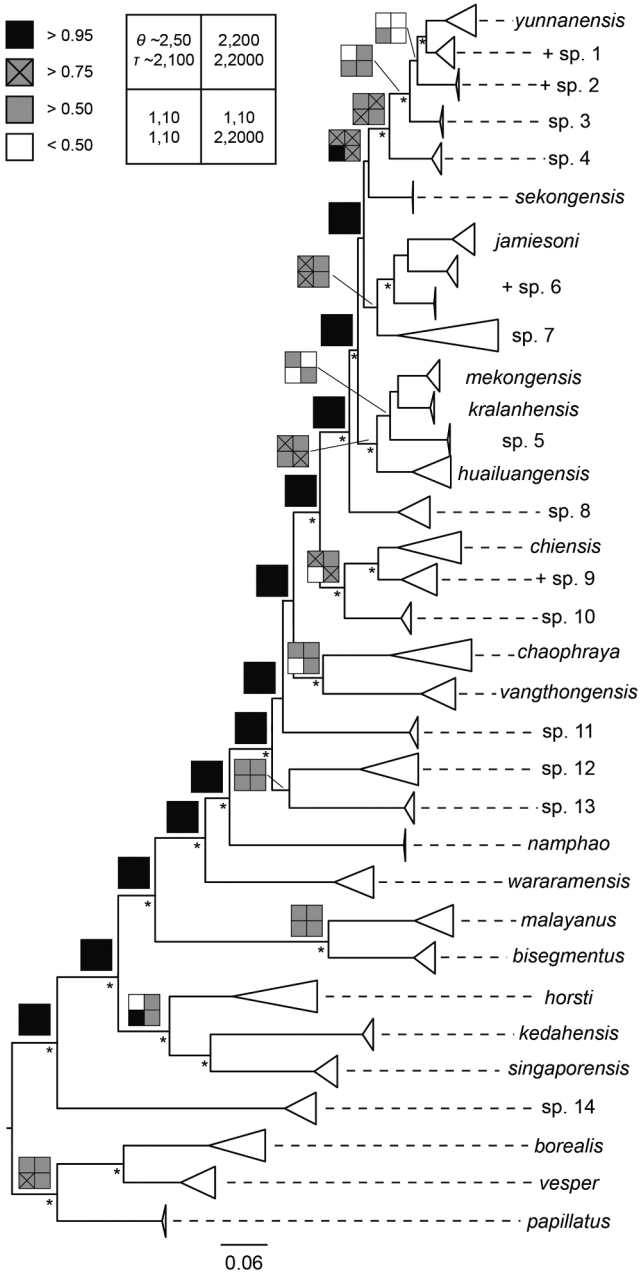

Fig. 8. BPP results of posterior probabilities from each θ and τ recombination shown as a four-tile square overlaid on each node of the BI concatenated tree. Asterisks indicate the posterior probability nodal support of > 0.95 from MrBayes.

barcoding studies to K2P distances, as in common usage, may yield biased results (Collins et al., 2012; Srivathsan and Meier, 2012).

In the GMYC analyses, it is unclear how the choice of tree model priors in BEAST affects the species delimitation results. In some studies, the results from a coalescent prior were similar to those with a Yule prior (Schwarzfeld and Sperling, 2015; da Cruz and Weksler, 2018). In contrast, our results showed that a Yule prior was less adequate and often produced large confidence intervals, as in some other previous studies (Ortiz and Francke, 2016; Ritchie et al., 2017). The lower adequacy of a Yule prior is possibly because GMYC uses coalescence as a null model (Monaghan et al., 2009). Anyway, both the coalescent and Yule priors have their limitations: the Yule model assumes a constant rate of speciation and no extinction, while the coalescent model assumes panmixia and constant population size. These assumptions are likely to be violated when dealing with real data (including *Glyphidrilus*), so that the threshold between intraspecific coalescence and speciation in GMYC may be ill-defined and so also the inferred species delimitations (Esselstyn et al., 2012; Ritchie et al., 2016). The calculation of BFs between competing tree models was adopted herein to indicate which model is more suitable for the construction of the ultrametric tree (Ritchie et al., 2017) and could mitigate the error in species estimation.

In the bGMYC, there is no consensus as to how the conspecificity posterior probabilities (P_{con}) should be chosen; $P_{con} > 0.95$ indicates the highest support of MOTUs but also the most stringent conspecificity, leading to the highest estimates of species numbers (Reid and Carstens, 2012). Thus, diverse values of P_{con} have been adopted in previous studies, often with subjective preferences (Gélin et al., 2017; Musher and Cracraft, 2018). In our study, we could not decide about *a priori* values of P_{con} so we generated a boxplot of the results between P_{con} = 0.05–0.95 (Fig. 7). The results from applying the median value of P_{con} (0.5) yielded more congruent, though not identical, MOTUs to the monophyletic species (except in COI which yielded markedly more MOTUs), and much more MOTUs were retrieved if higher P_{con} values were chosen.

In our mPTP results, different effects between using BI and ML trees were associated with the different markers analyzed and showed no obvious trend. Nevertheless, PTP results will be robust as long as a robust gene tree is provided (Tang et al., 2014). In BPP, adopting four combinations of θ and τ values yielded consistent posterior probabilities for most nodes, which is in line with previous reports (Rato et al., 2016; Toussaint et al., 2016). Although it has been suggested that higher values of θ would favor fewer MOTUs (Leaché and Fujita, 2010; McKay et al., 2013), this was not always the case in *Glyphidrilus*, since larger values of θ sometimes yielded recent nodes with posterior probabilities in the range of 0.76–1.00 that supported more MOTUs (Fig. 8).

Another problem is that tree-based species delimitation methods in particular (e.g. GMYC) are susceptible to taxon sampling effects (Fujisawa and Barraclough, 2013; Hamilton et al., 2014). As such, GMYC and PTP yield more species for parts of a larger dataset if they are applied to only that part of the larger dataset (Schwarzfeld and Sperling, 2015; Ahrens et al., 2016). Although the rate heterogeneity and the proportion of unresolved nodes may increase, which may affect the species number estimates (Talavera et al., 2013; Tang et al., 2014), including all subclades probably yields more accurate parameter estimates for the species delimitations and may improve their local performance on a tree (Ahrens et al., 2016). In this regard, one should include as many different relevant taxa as possible and sample from a wide geographical area so that both deeply and recently diverged lineages are represented in order to avoid the poorer performance from limited taxon sampling (Talavera et al., 2013; Ahrens et al., 2016).

4.3. Secondary ITS2 structure and its application in species delimitation

The secondary structure of ITS can provide relevant evidence for species delimitation (Ruhl et al., 2010; Wolf et al., 2013), but this has not yet been explored in earthworms. The secondary structure of ITS2 in *Glyphidrilus* (Fig. 9) complied with the eukaryotic common core structure (Coleman, 2003, 2007, 2015; Schultz et al., 2005). However, 23 monophyletic species, including all Mekong species, had no helix IV, as in many other animals (Ruhl et al., 2010; Salvi et al., 2010), but contained a double-branched helix III, such as in nematodes (Ma et al., 2008). Some *Glyphidrilus* species exhibited an additional helix IA upstream from helix IB, as has been reported in, for example, anthozoans (Aguilar and Reimer, 2010).

In general, CBCs are suggested to occur between different, reproductively isolated species (Müller et al., 2007). As such, CBC-based species delimitation has been adopted to distinguish closely related but morphologically similar eukaryotic species (Ahvenniemi et al., 2009; Shazib et al., 2016). However, CBC analysis has rarely been applied to

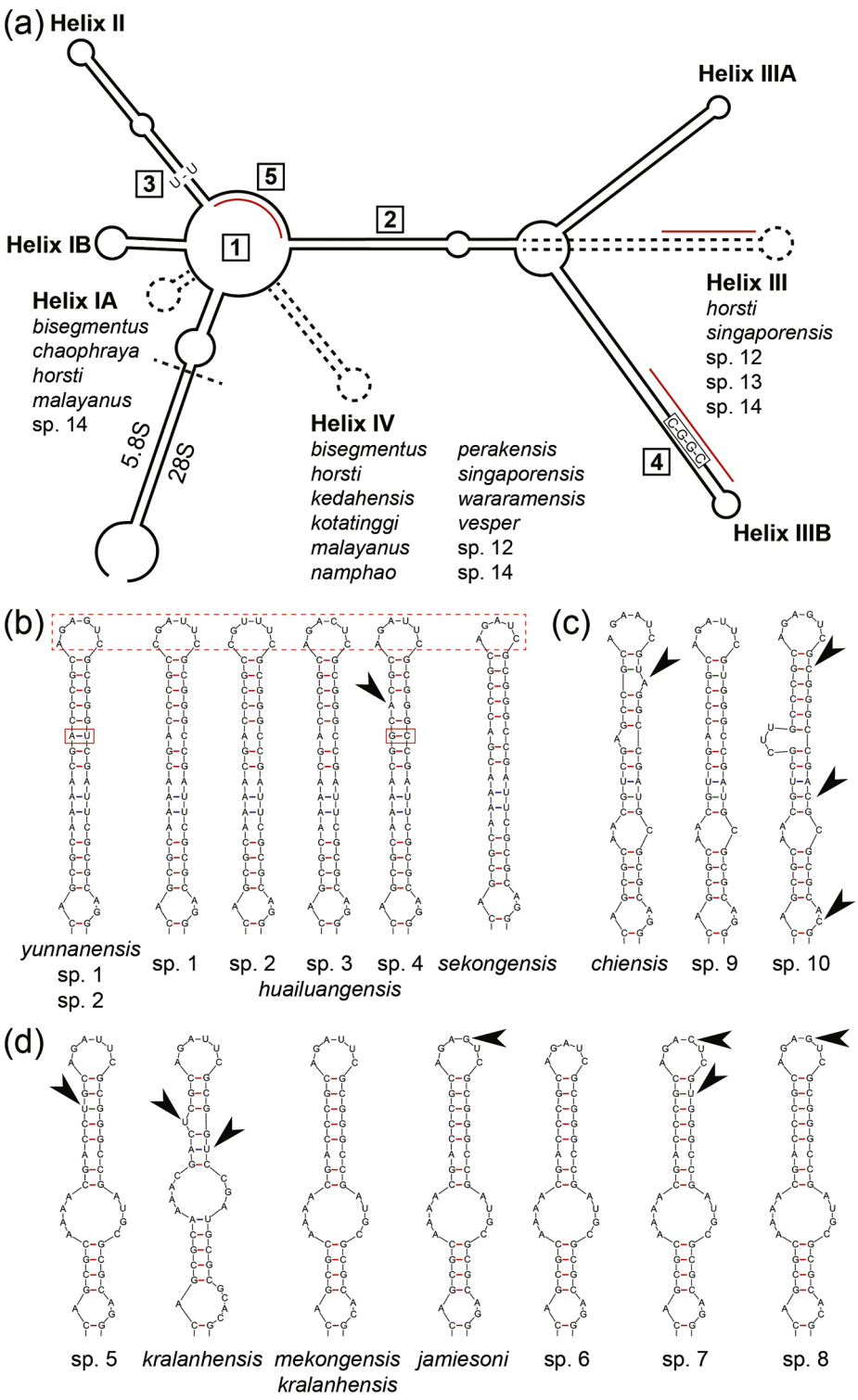

Fig. 9. Secondary structure of ITS2 showing (a) the common core with details on its helices and all hallmarks, and variable part of helix II of each species within the Mekong clade comprising (b) *G. yunnanensis* (plus *sekongensis* and *huailuangensis*), (c) *G. chiensis* and (d) *G. jamiesoni* and *G. mekongensis* (plus sp. 8) species complexes. All hallmarks are as follows: (1) four helices, (2) helix III is the longest, (3) helix II shows a U-U mismatch, (4) the 5′ side of helix III shows a YGGY motif (in this case CGGC), and (5) there is a conserved single-stranded purine-rich region between helix II and III. Red lines indicate the position of conserved regions. Arrows indicate the positions of some hemi-CBCs and different nucleotides in the loop regions. A red dashed box illustrates the difference of loop region among *G. yunnanensis* (plus *sekongensis* and *huailuangensis*) species complex. Red boxes indicate the positions of CBCs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

animals (Coleman, 2003). We observed up to five CBCs in pairwise comparisons among *Glyphidrilus* species, which is more than in other animal studies (e.g. Young and Coleman, 2004; Di Capua et al., 2017). Although the multiple occurrence of hemi-CBCs is sometimes also used to recognize different biological species (Shazib et al., 2016), we refrained from doing so as we observed hemi-CBCs between ITS2 sequences within the same morphospecies (not shown).

4.4. Taxonomic implications

Five unnamed species (*Glyphidrilus* spp. 1–3, 6 and 9) did not meet the criteria of the integrative taxonomic workflow to be interpreted as candidate species and were regarded as DCLs. Their spermathecae have the same positions within each clade, whereas their different wing positions might reflect intraspecific morphological variation (Fig. 1). In contrast, *Glyphidrilus* sp. 7 and sp. 10 are retained as CCSs, as these two species met at least three out of the four criteria, and *Glyphidrilus* sp. 7 also has a different position of the spermathecae. In addition, both

species' distribution ranges were restricted to eastern Thailand (Fig. 1b and S1). We also retained *Glyphidrilus* sp. 12 and sp. 13 as CCSs because they were supported by reciprocal monophyly and species delimitation results in both ITSs and multilocus MSC methods. In contrast, *Glyphidrilus* sp. 8 is regarded as an UCS because it is only supported by multilocus MSC methods (Fig. 2). However, the similar morphology and overlapping distributions of *Glyphidrilus* sp. 8, sp. 12 and sp. 13 with the *G. chiensis* complex need further investigation (Fig. 1b and S1). Furthermore, the problem of distinguishing between cryptic species and diverging DCLs in *Glyphidrilus* is further aggravated by the fact that some key morphological characters, like the position of male pores and anterior blood vessels, are not always included in species descriptions (Horst, 1893; Brinkhurst and Jamieson, 1971).

Several *Glyphidrilus* clades morphologically coincide with *G. chiensis* and *G. quadratus*, while some specimens of both these species *sensu* Chanabun et al. (2013) are now assigned to different species (Table S1 and Fig. 1b). For example, some northeastern Thai specimens that were previously identified as *G. chiensis* or *G. quadratus* are now considered to belong to *G. jamiesoni* (Table S1). In this regard, *G. jamiesoni* specimens tend to have longer wings and a more posterior-ward position of the spermathecae than *G. chiensis* delimited in this study (Fig. 1b).

There are three cases where a remarkable morphological differentiation was not reflected in the DNA analyses, so that corresponding morphospecies were assigned to the same clades. The first case involves *G. champasakensis* with its clitellum that is 10 segments longer than in *G. yunnanensis* and *G. vangviengensis* (Fig. 1d). However, intraspecific clitellum length variation has been reported in other species (Gates, 1972; Terhivuo and Saura, 1993) and so that is how this difference is putatively interpreted here. The second case involves the different wing lengths and positions of the genital markings among, for example, *G. mekongensis*, *G. chiangraiensis* and *G. namdonensis*. These differences are also regarded as intraspecific variation, because those nominal species have similar numbers and positions of the spermathecae within the same clade (Fig. 1a). Although *G. kralanhensis* was monophyletic in mtDNA trees, most species delimitation methods lumped it with *G. mekongensis*. This is consistent with the positions of the wings and spermathecae in *G. kralanhensis* which fall within the range of *G. chiangraiensis* and *G. namdonensis*. For the time being, we thus retain *G. kralanhensis* as an UCS. The third case involves the posterior shift of the wings and spermathecae by 2–3 segments in *G. perakensis* compared to *G. kedahensis* and *G. satunensis* (Fig. 1f). Anterior and posterior displacements of genital openings and reproductive organs have often been reported in earthworms (Morgan, 1895; Gates, 1958, 1972) and could be caused by amputation of anterior body segments resulting in either hypomeric (segment loss) or hypermeric (segment gain) regeneration (Gates, 1927, 1941). Thus, the posterior displacement of some organs is believed to be mainly caused by hypermeric regeneration, but aberrant ontogenetic development was also proposed to be an alternative cause (Gates, 1951, 1957). Anyway, *G. perakensis* individuals might involve *G. kedahensis* with posterior displacement of all important diagnostic characters.

4.5. Cryptic species or diverging conspecific lineages in an integrative taxonomic framework

Morphological crypsis in earthworms is more the rule than the exception (Erséus and Gustafsson, 2009). Deeply diverging mitochondrial lineages were discovered within well-established morphospecies in several families, such as the Lumbricidae (Martinsson and Erséus, 2017; Porco et al., 2018), Megascolecidae (Buckley et al., 2011; Jeratthitikul et al., 2017) and *Glyphidrilus* (Almidae), in which some of these deeply diverging mitochondrial lineages may be morphologically cryptic species. However, it is well known that species delimitations tend to overestimate species numbers due to the effect of geographical structuring (Lohse, 2009). For instance, ABGD can be prone to over-splitting if the barcode gap is inferred from deeply diverging conspecific populations (Hamilton et al., 2014; Ortiz and Francke, 2016).

Moreover, GMYC and PTP may misinterpret isolated genetic clusters and (deeply) diverged intraspecific populations as different species (Papadopoulou et al., 2008, 2009; Fujisawa and Barraclough, 2013). This holds true in previous studies, where the presence of a mitochondrial barcode gap or deep lineages in earthworms does not always mean the presence of cryptic species (Giska et al., 2015; Martinsson et al., 2017). Highly divergent mtDNA lineages in earthworms could be explained by several factors, such as an admixture among previously geographically isolated lineages, large effective population size, balancing selection and introgression (Giska et al., 2015). Those are likely the causes in *Glyphidrilus*, especially the population admixture along the rivers and the dynamics of river pattern changes in mainland Southeast Asia (Chanabun et al., 2013; Jirapatrasilp et al., 2015, 2016). Nevertheless, measures of intraspecific gene flow or introgression between interspecific sympatric populations using genomic data are recommended.

Nuclear loci commonly do not show the reciprocal monophyly of mtDNA clades (Rato et al., 2016), as mtDNA often shows a stronger phylogenetic signal because of its reduced effective population size, lower level of recombination and higher substitution rates (Ballard and Whitlock, 2004; Rubinoff et al., 2005, 2006). As such, multispecies coalescent-based methods, which do not require that each species is reciprocally monophyletic for each gene (Fujita et al., 2012), are able to delimit species recognized as mtDNA clades, but that are not monophyletic with nuclear loci (Rato et al., 2016). This is shown in our Mekong monophyletic species. Nevertheless, multispecies coalescent methods may still overestimate species numbers, especially when there are intraspecific population structures (McKay et al., 2013; Ortiz and Francke, 2016), which can be confounded with species boundaries (Sukumaran and Knowles, 2017; Leaché et al., 2019). Of course, this caution applies to all species delimitation methods that are susceptible to the paucity of genetic markers used, different substitution rates among loci and mitonuclear discordance.

Our integrative taxonomic framework incorporates a wide array of species delimitation results of both mitochondrial and nuclear markers, and also takes morphology and ITS2 secondary structure into account. We are well-aware that different species concepts underlie different species delimitation methods in our study (Table 1). The reconciliation of alternative species concepts thus complies with the unified species concept (de Queiroz, 2007), which could accommodate species delimitation in an integrative taxonomic framework. In addition, our framework is "iterative" in the sense that our PSPs (species boundaries) were further refined by different additional datasets (Yeates et al., 2011). Thus, an increase in the rigor and robustness of the delimited species is expected (Schlick-Steiner et al., 2010). However, an integrative taxonomic approach is not without any argument, as a subjective decision is still needed to resolve incongruences among different delimitation results (Reydon and Kunz, 2019). By applying the candidate species approach, the monophyletic species were evaluated to what extent those species as PSPs comply with different levels of reliability in species boundary by classifying them into the three categories of CCS, UCS and DCL (Vieites et al., 2009). With this approach we aimed to mitigate our subjective judgment on species entities and also avoid the risk of contributing to a potentially ill-founded taxonomic inflation (Jörger et al., 2012; Carstens et al., 2013) in cases where species delimitation was not unequivocally supported by an integrative approach. In addition, additional external data (e.g. biochemical, physiology, behavior, etc.) are crucial to underpin the interpretation of any species delimitation results as eventual testable species hypotheses (Bernardo, 2011).

5. Conclusion

This integrative taxonomic study of Southeast Asian *Glyphidrilus* yielded 26 CCSs and two UCSs, nine of which are probably new to

science and need formal description. Our study also provided evidence to synonymize 12 nominal morphospecies names (Table 2). The integrative taxonomic framework, here incorporating the ITS2 secondary structure of earthworms for the first time, is beneficial in tackling this particular case of morphological crypsis among divergent lineages and morphological variability within the same phylogenetic lineages. Some cryptic species obtained from species delimitation of only mtDNA datasets may in fact be deep intraspecific lineages, as those did not pass the criteria in our integrative taxonomic framework. Thus, the exclusive use of COI as the universal DNA barcode may overestimate the species diversity and will complicate the application in ecology and conservation. Genome-wide approaches (such as RADseq and genotyping by sequencing) will hereby be promising to further elucidate the true extent of *Glyphidrilus* cryptic species.

Acknowledgements

This study was supported by a graduate studentship to PJ through the Royal Golden Jubilee PhD Program (PHD/0113/2556), Thailand Research Fund, Thailand. The main funding was provided by Center of Excellence on Biodiversity (BDC-PG2-160012), Thailand. We thank all members of the Animal Systematics Research Unit (ASRU) at Chulalongkorn University, Thailand for assistance in field sampling. Special thanks go to K. Inkhavilay, B. Ng, C. Sutcharit and E. Jeratthitikul for collecting and providing specimens from Lao P.D.R., Malaysia and P.R. China. Special thanks also go to Prof. Dr. J.J. Jaeger, Dr. Y. Chaimanee, French-Myanmar colleagues, and Fauna and Flora International for their kindness to invite ASRU members to join the field expeditions in Myanmar in 2013–2014 and 2016. Sincere thanks are extended to Robert Butcher and the journal's referees for their valuable and helpful comments that substantially improved the manuscript.

Data accessibility

DNA sequences: GenBank Accession numbers: KX238575 – KX238779, MG920572 – MG920779, MG921689 – MG922030, MG922112 – MG922457, MG923003 – MG923209.

See Table S1 for specific Accession numbers.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106531.

References

Aguilar, C., Reimer, J.D., 2010. Molecular phylogenetic hypotheses of *Zoanthus* species (Anthozoa: Hexacorallia) using RNA secondary structure of the internal transcribed spacer 2 (ITS2). Mar. Biodiv. 40, 195–204.
Ahrens, D., Fujisawa, T., Krammer, H.-J., Eberle, J., Fabrizi, S., Vogler, A.P., 2016. Rarity and incomplete sampling in DNA-based species delimitation. Syst. Biol. 65, 478–494.
Ahvenniemi, P., Wolf, M., Lehtonen, M.J., Wilson, P., German-Kinnari, M., Valkonen, J.P.T., 2009. Evolutionary diversification indicated by compensatory base changes in ITS2 secondary structures in a complex fungal species, *Rhizoctonia solani*. J. Mol. Evol. 69, 150–163.
André, H.M., Noti, M.-I., Lebrun, P., 1994. The soil fauna: the other last biotic frontier. Biodiv. Conserv. 3, 45–56.
Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.
Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744.
Baum, D.A., Shaw, K.L., 1995. Genealogical perspectives on the species problem. In: Hoch, P.C., Stephenson, A.G. (Eds.), Experimental and Molecular Approaches to Plant Biosystematics. Missouri Botanical Garden, St. Louis, pp. 289–303.
Bernardo, J., 2011. A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. In: Hodkinson, T.R., Jones, M.B., Waldren, S., Parnell, J.A.N. (Eds.), Climate Change. Cambridge University Press, Ecology and Systematics.
Blair, C., Bryson, R.W., 2017. Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: *Phrynosoma*). Mol. Ecol. Resour. 17, 1168–1182.
Blakemore, R.J., 2009. Cosmopolitan Earthworms — A Global and Historical Perspective. In: Shain, D.H. (Ed.), Annelids in Modern Biology. John Wiley and Sons, Inc., pp. 257–283.
Blakemore, R.J., 2016. Cosmopolitan Earthworms – an Eco-Taxonomic Guide to the Peregrine Species of the World, sixth ed. VermEcology, Yokohama, Japan.
Blouin, M., Hodson, M.E., Delgado, E.A., Baker, G., Brussaard, L., Butt, K.R., Dai, J., Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D., Brun, J.J., 2013. A review of earthworm impact on soil function and ecosystem services. Eur. J. Soil Sci. 64, 161–182.
Brinkhurst, R.O., Jamieson, B.G.M., 1971. Aquatic Oligochaeta of the World. University of Toronto Press, Edinburgh.
Briones, M.J.I., 2014. Soil fauna and soil functions: a jigsaw puzzle. Front. Environ. Sci. 2.
Briones, M.J.I., Morán, P., Posada, D., 2009. Are the sexual, somatic and genetic characters enough to solve nomenclatural problems in lumbricid taxonomy? Soil Biol. Biochem. 41, 2257–2271.
Brower, A.V.Z., 1999. Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. Syst. Biol. 48, 199–213.
Buckley, T.R., James, S., Allwood, J., Bartlam, S., Howitt, R., Prada, D., 2011. Phylogenetic analysis of New Zealand earthworms (Oligochaeta: Megascolecidae) reveals ancient clades and cryptic taxonomic diversity. Mol. Phylogen. Evol. 58, 85–96.
Caisová, L., Marin, B., Melkonian, M., 2013. A consensus secondary structure of ITS2 in the Chlorophyta identified by phylogenetic reconstruction. Protist 164, 482–496.
Cao, X., Liu, J., Chen, J., Zheng, G., Kuntner, M., Agnarsson, I., 2016. Rapid dissemination of taxonomic discoveries based on DNA barcoding and morphology. Sci. Rep. 6, 37066.
Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.
Chanabun, R., Bantaowong, U., Sutcharit, C., Tongkerd, P., Inkavilay, K., James, S.W., Panha, S., 2011. A new species of semi-aquatic freshwater earthworm of the genus *Glyphidrilus* Horst, 1889 from Laos (Oligochaeta: Almidae). Trop. Nat. Hist. 11, 213–222.
Chanabun, R., Bantaowong, U., Sutcharit, C., Tongkerd, P., James, S.W., Panha, S., 2012a. A new species of semi-aquatic freshwater earthworm of the genus *Glyphidrilus* Horst, 1889 from the Mekong River (Oligochaeta: Almidae). Raffles Bull. Zool. 60, 265–277.
Chanabun, R., Inkavilay, K., Panha, S., 2017. New species of semi-aquatic freshwater earthworm genus *Glyphidrilus* Horst, 1889 from Thailand and Laos (Oligochaeta, Almidae). ZooKeys 672, 1–34.
Chanabun, R., Panha, S., 2015. Two new species of semi-aquatic earthworms genus *Glyphidrilus* Horst, 1889 from Malaysia (Oligochaeta: Almidae). Trop. Nat. Hist. 15, 179–189.
Chanabun, R., Sutcharit, C., Tongkerd, P., Panha, S., 2013. The semi-aquatic freshwater earthworms of the genus *Glyphidrilus* Horst, 1889 from Thailand (Oligochaeta, Almidae) with re-descriptions of several species. ZooKeys 265, 1–76.
Chanabun, R., Sutcharit, C., Tongkerd, P., Tan, S.-H.A., Panha, S., 2012b. Three new species of semi-aquatic freshwater earthworms of the genus *Glyphidrilus* Horst, 1889 from Malaysia (Clitellata: Oligochaeta: Almidae). Zootaxa 3458, 120–132.
Chang, C.-H., James, S.W., 2011. A critique of earthworm molecular phylogenetics. Pedobiologia 54 (Supplement), S3–S9.
Chen, Y., Xu, Z.F., 1977. On some new earthworms from China II. Acta Zool. Sinica 23, 175–181.
Choi, S.C., 2016. Methods for delimiting species via population genetics and phylogenetics using genotype data. Genes Genom. 38, 905–915.
Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 19, 370–375.
Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic Acids Res. 10.
Coleman, A.W., 2015. Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. Trends Genet. 31, 157–163.
Collins, R.A., Boykin, L.M., Cruickshank, R.H., Armstrong, K.F., 2012. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. Meth. Ecol. Evol. 3, 457–465.
da Cruz, M.d.O.R., Weksler, M., 2018. Impact of tree priors in species delimitation and phylogenetics of the genus *Oligoryzomys* (Rodentia: Cricetidae). Mol. Phylogen. Evol. 119, 1–12.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Meth. 9, 772.
Dayrat, B., 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85, 407–415.
de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.
Decaëns, T., Jiménez, J.J., Gioia, C., Measey, G.J., Lavelle, P., 2006. The values of soil animals for conservation biology. Eur. J. Soil Biol. 42, S23–S38.
Decaëns, T., Porco, D., Rougerie, R., Brown, G.G., James, S.W., 2013. Potential of DNA barcoding for earthworm research in taxonomy and ecology. Appl. Soil Ecol. 65, 35–42.
Di Capua, I., Maffucci, F., Pannone, R., Mazzocchi, M.G., Biffali, E., Amato, A., 2017. Molecular phylogeny of Oncaeidae (Copepoda) using nuclear ribosomal internal transcribed spacer (ITS rDNA). PLoS One 12, e0175662.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
Edwards, C.A., Arancon, N.Q., Sherman, R.L., 2010. Vermiculture Technology: Earthworms, Organic Wastes, and Environmental Management. CRC Press.
Edwards, C.A., Bohlen, P.J., 1996. Biology and Ecology of Earthworms. Chapman and Hall, London.
Ence, D.D., Carstens, B.C., 2011. SpedeSTEM: a rapid and accurate method for species

delimitation. Mol. Ecol. Resour. 11, 473–480.
Erséus, C., Gustafsson, D., 2009. Cryptic speciation in clitellate model organisms. In: Shain, D.H. (Ed.), Annelids in Modern Biology. John Wiley and Sons Inc, pp. 31–46.
Esselstyn, J.A., Evans, B.J., Sedlock, J.L., Anwarali Khan, F.A., Heaney, L.R., 2012. Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. Proc. R. Soc. B 279, 3678–3686.
Ezard, T., Fujisawa, T., Barraclough, T. 2009. SPLITS: SPecies' Limits by Threshold Statistics. R Packag. version 1.0-14/r31.
Fitter, A.H., Gilligan, C.A., Hollingworth, K., Kleczkowski, A., Twyman, R.M., Pitchford, J.W., 2005. Biodiversity and ecosystem function in soil. Func. Ecol. 19, 369–377.
Fründ, H.-C., Graefe, U., Tischer, S., 2011. Earthworms as bioindicators of soil quality. In: Karaca, A. (Ed.), Biology of Earthworms. Springer, Berlin, Heidelberg, pp. 261–278.
Fujisawa, T., Aswad, A., Barraclough, T.G., 2016. A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. Syst. Biol. 65, 759–771.
Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on simulated data sets. Syst. Biol. 62, 707–724.
Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27, 480–488.
Gates, G.E., 1927. Regeneration in a tropical earthworm *Perionyx excavatus* E. Perr. Biol. Bull. 53, 351–364.
Gates, G.E., 1941. Further notes on regeneration in a tropical earthworm, *Perionyx excavatus* E. Perrier 1872. J. Exp. Zool. 88, 161–185.
Gates, G.E., 1951. An abnormal earthworm. Am. Midl. Nat. 45, 471–473.
Gates, G.E., 1957. On the origin of the homoeosis in an aberrant earthworm of the lumbricid species *Allolobophora tuberculata* Eisen, 1874. Ann. Mag. Nat. Hist. 10, 204–208.
Gates, G.E., 1958. On homoeosis, as well as other aberrations, and their origin in an earthworm species, *Eisenia foetida* (Savigny, 1826), along with some deductions as to morphogenesis in the Lumbricidae. Am. Midl. Nat. 59, 452–464.
Gates, G.E., 1972. Burmese earthworms: an introduction to the systematics and biology of megadrile oligochaetes with special reference to Southeast Asia. Trans. Am. Phil. Soc. 62, 1–326.
Gélin, P., Postaire, B., Fauvelot, C., Magalon, H., 2017. Reevaluating species number, distribution and endemism of the coral genus *Pocillopora* Lamarck, 1816 using species delimitation methods and microsatellites. Mol. Phylogen. Evol. 109, 430–446.
Giska, I., Sechi, P., Babik, W., 2015. Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. BMC Evol. Biol. 15, 217.
Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., Posada, D., 2010. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Res. 38, W14–W18.
Hamilton, C.A., Hendrixson, B.E., Brewer, M.S., Bond, J.E., 2014. An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). Mol. Phylogen. Evol. 71, 79–93.
Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2, e312.
Hendrixson, B.E., DeRussy, B.M., Hamilton, C.A., Bond, J.E., 2013. An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, *Aphonopelma*). Mol. Phylogen. Evol. 66, 327–340.
Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Horst, R., 1893. Earthworms from the Malay Archipelago. Zoologische Ergebnisse einer Reise in Niederländisch Ost-Indien 3, 28–83.
Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. Evolution 56, 1557–1565.
Jamieson, B.G.M., 2006. Chapter 8: Non-leech Clitellata. In: Rouse, G., Pleijel, F. (Eds.), Reproductive Biology and Phylogeny of Annelida. Science Publishers, Enfield, pp. 235–392.
Jeratthitikul, E., Bantaowong, U., Panha, S., 2017. DNA barcoding of the Thai species of terrestrial earthworms in the genera *Amynthas* and *Metaphire* (Haplotaxida: Megascolecidae). Eur. J. Soil Biol. 81, 39–47.
Jirapatrasilp, P., Prasankok, P., Chanabun, R., Panha, S., 2015. Allozyme data reveal genetic diversity and isolation by distance in sympatric *Glyphidrilus* Horst, 1889 (Oligochaeta: Almidae) of the Lower Mekong River Basin. Biochem. Syst. Ecol. 61, 35–43.
Jirapatrasilp, P., Prasankok, P., Sutcharit, C., Chanabun, R., Panha, S., 2016. Two new Cambodian semi-aquatic earthworms in the genus *Glyphidrilus* Horst, 1889 (Oligochaeta, Almidae), based on morphological and molecular data. Zootaxa 4189, 543–558.
Jörger, K.M., Norenburg, J.L., Wilson, N.G., Schrödl, M., 2012. Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. BMC Evol. Biol. 12, 245.
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.
Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. bbx108.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kekkonen, M., Mutanen, M., Kaila, L., Nieminen, M., Hebert, P.D.N., 2015. Delineating species with DNA barcodes: a case of taxon dependent method performance in moths. PLoS One 10, e0122481.
Keller, A., Schleicher, T., Schultz, J., Müller, T., Dandekar, T., Wolf, M., 2009. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene 430, 50–57.
Khedkar, G.D., Jamdade, R., Naik, S., David, L., Haymer, D., 2014. DNA barcodes for the fishes of the Narmada, one of India's longest rivers. PLoS One 9, e101460.
Klarica, J., Kloss-Brandstätter, A., Traugott, M., Juen, A., 2012. Comparing four mitochondrial genes in earthworms – implications for identification, phylogenetics, and discovery of cryptic species. Soil Biol. Biochem. 45, 23–30.
Klinth, M.J., Martinsson, S., Erséus, C., 2017. Phylogeny and species delimitation of North European *Lumbricillus* (Clitellata, Enchytraeidae). Zool. Scr. 46, 96–110.
Koetschan, C., Förster, F., Keller, A., Schleicher, T., Ruderisch, B., Schwarz, R., Müller, T., Wolf, M., Schultz, J., 2010. The ITS2 Database III—sequences and structures for phylogeny. Nucleic Acids Res. 38, D275–D279.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Lavelle, P., Barot, S., Blouin, M., Decaëns, T., Jimenez, J.J., Jouquet, P., 2007. Earthworms as key actors in self-organized soil systems. In: Cuddington, K., Byers, J., Wilson, W., Hastings, A. (Eds.), Ecosystem Engineers Plants to Protists. Academic Press, pp. 77–107.
Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. B 277, 3071–3077.
Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2019. The spectre of too many species. Syst. Biol. 68, 168–181.
Lee, B.-T., Shin, K.-H., Kim, J.-Y., Kim, K.-W., 2008. Progress in Earthworm Ecotoxicology. In: Kim, Y.J., Platt, U. (Eds.), Advanced Environmental Monitoring. Springer, Netherlands, Dordrecht, pp. 248–258.
Lohse, K., 2009. Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). Syst. Biol. 58, 439–442.
Lowe, C.N., Butt, K.R., Sherman, R.L., 2014. Current and Potential Benefits of Mass Earthworm Culture. In: Morales-Ramos, J.A., Rojas, M.G., Shapiro-Ilan, D.I. (Eds.), Mass Production of Beneficial Organisms. Academic Press, San Diego, pp. 683–709.
Ma, H., Overstreet, R.M., Subbotin, S.A., 2008. ITS2 secondary structure and phylogeny of cyst-forming nematodes of the genus *Heterodera* (Tylenchida: Heteroderidae). Org. Divers. Evol. 8, 182–193.
Mallet, J., 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10, 294–299.
Mallo, D., Posada, D., 2016. Multilocus inference of species trees and DNA barcoding. Phil. Trans. R. Soc. B 371, 20150335.
Martinsson, S., Erséus, C., 2017. Cryptic speciation and limited hybridization within *Lumbricus* earthworms (Clitellata: Lumbricidae). Mol. Phylogen. Evol. 106, 18–27.
Martinsson, S., Rhodén, C., Erséus, C., 2017. Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae). Mitochondrial DNA A 28, 147–155.
Masters, B.C., Fan, V., Ross, H.A., 2011. Species delimitation – a geneious plugin for the exploration of species boundaries. Mol. Ecol. Resour. 11, 154–157.
Mayr, E., 1942. Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press, New York.
McKay, B.D., Mays, H.L., Wu, Y., Li, H., Yao, C.-T., Nishiumi, I., Zou, F., 2013. An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. Mol. Ecol. 22, 4943–4957.
Michaelsen, W., 1902. Neue Oligochaeten und neue Fundorte alt-bekannter. Jahrbuch der Hamburgischen Wissenschaftlichen Anstalten 19, 1–54.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, pp. 1–8.
Mishler, B.D., 1985. The morphological, developmental, and phylogenetic basis of species concepts in bryophytes. The Bryologist 88, 207–214.
Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J.G., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G., Vogler, A.P., 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Syst. Biol. 58, 298–311.
Morgan, J.A.T., Blair, D., 1998. Trematode and Monogenean rRNA ITS2 secondary structures support a four-domain model. J. Mol. Evol. 47, 406–419.
Morgan, T.H., 1895. A study of metamerism. Q. J. Microsc. Sci. 148, 395–476.
Müller, T., Philippi, N., Dandekar, T., Schultz, J., Wolf, M., 2007. Distinguishing species. RNA 13, 1469–1472.
Musher, L.J., Cracraft, J., 2018. Phylogenomics and species delimitation of a complex radiation of Neotropical suboscine birds (*Pachyramphus*). Mol. Phylogen. Evol. 118, 204–221.
Novo, M., Almodóvar, A., Fernández, R., Trigo, D., Díaz-Cosín, D.J., Giribet, G., 2012a. Appearances can be deceptive: different diversification patterns within a group of Mediterranean earthworms (Oligochaeta, Hormogastridae). Mol. Ecol. 21, 3776–3793.
Novo, M., Almodóvar, A., Fernández, R., Trigo, D., Díaz Cosín, D.J., 2010. Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. Mol. Phylogen. Evol. 56, 507–512.
Novo, M., Fernández, R., Fernández Marchán, D., Gutiérrez, M., Diaz Cosin, D., 2012b. Compilation of morphological and molecular data, a necessity for taxonomy: The case of *Hormogaster abbatissae* sp. n. (Annelida, Clitellata, Hormogastridae). ZooKeys 242, 1–17.
Ortiz, D., Francke, O.F., 2016. Two DNA barcodes and morphology for multi-method species delimitation in *Bonnetina* tarantulas (Araneae: Theraphosidae). Mol. Phylogen. Evol. 101, 176–193.
Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of

taxonomy. Front. Zool. 7, 16.
Papadopoulou, A., Bergsten, J., Fujisawa, T., Monaghan, M.T., Barraclough, T.G., Vogler, A.P., 2008. Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. Phil. Trans. R. Soc. B 363, 2987–2996.
Papadopoulou, A., Monaghan, M.T., Barraclough, T.G., Vogler, A.P., 2009. Sampling error does not invalidate the Yule-coalescent model for species delimitation. A response to Lohse (2009). Syst. Biol. 58, 442–444.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Pérès, G., Vandenbulcke, F., Guernion, M., Hedde, M., Beguiristain, T., Douay, F., Houot, S., Piron, D., Richard, A., Bispo, A., Grand, C., Galsomies, L., Cluzeau, D., 2011. Earthworm indicators as tools for soil monitoring, characterization and risk assessment. An example from the national Bioindicator programme (France). Pedobiologia 54, S77–S87.
Pérez-Losada, M., Bloch, R., Breinholt, J.W., Pfenninger, M., Domínguez, J., 2012. Taxonomic assessment of Lumbricidae (Oligochaeta) earthworm genera using DNA barcodes. Eur. J. Soil Biol. 48, 41–47.
Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., Hedin, M., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609.
Porco, D., Chang, C.-H., Dupont, L., James, S., Richard, B., Decaëns, T., 2018. A reference library of DNA barcodes for the earthworms from Upper Normandy: biodiversity assessment, new records, potential cases of cryptic diversity and ongoing speciation. Appl. Soil Ecol. 124, 362–371.
Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21, 1864–1877.
Rambaut, A., Suchard, M. A., Xie, D., Drummond, A. J. 2014. Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/.
Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.
Rannala, B., Yang, Z., 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194, 245–253.
Rato, C., Harris, D.J., Carranza, S., Machado, L., Perera, A., 2016. The taxonomy of the *Tarentola mauritanica* species complex (Gekkota: Phyllodactylidae): Bayesian species delimitation supports six candidate species. Mol. Phylogen. Evol. 94, 271–278.
Reid, N.M., Carstens, B.C., 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. 12, 196.
Reydon, T.A.C., Kunz, W., 2019. Species as natural entities, instrumental units and ranked taxa: new perspectives on the grouping and ranking problems. Biol. J. Linn. Soc. 126, 623–636.
Reynolds, J.W., 2004. The status of earthworm biogeography, diversity, and taxonomy in North America revisited with glimpses into the future. Earthworm Ecol. 63–74 CRC Press.
Reynolds, J.W., Wetzel, M.J., 2017. Nomenclatura Oligochaetologica – A Catalogue of Names, Descriptions and Type Specimens. Editio Secunda.
Rieppel, O., 2010. Species monophyly. J. Zool. Syst. Evol. Res. 48, 1–8.
Ritchie, A.M., Lo, N., Ho, S.Y.W., 2016. Examining the sensitivity of molecular species delimitations to the choice of mitochondrial marker. Org. Divers. Evol. 16, 467–480.
Ritchie, A.M., Lo, N., Ho, S.Y.W., 2017. The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. Syst. Biol. 66, 413–425.
Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H., Tsai, P., 2008. The perils of plenty: what are we going to do with all these genes? Phil. Trans. R. Soc. B 363, 3893.
Römbke, J., Aira, M., Backeljau, T., Breugelmans, K., Domínguez, J., Funke, E., Graf, N., Hajibabaei, M., Pérez-Losada, M., Porto, P.G., Schmelz, R.M., Vierna, J., Vizcaíno, A., Pfenninger, M., 2016. DNA barcoding of earthworms (*Eisenia fetida/andrei* complex) from 28 ecotoxicological test laboratories. Appl. Soil Ecol. 104, 3–11.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Rosa, D., 1890. Viaggio di Leonardo Fea in Birmania e Regioni vicine, xxv, Moniligastridi, Geoscolecidi. Ann. Mus. Civ. Stor. Nat. Genova 9, 386–400.
Rosenberg, N.A., 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61, 317–323.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302.
Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97, 581–594.
Rubinoff, D., Holland, B.S., Savolainen, V., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54, 952–961.
Ruhl, M.W., Wolf, M., Jenkins, T.M., 2010. Compensatory base changes illuminate morphologically difficult taxonomy. Mol. Phylogen. Evol. 54, 664–669.
Salvi, D., Bellavia, G., Cervelli, M., Mariottini, P., 2010. The analysis of rRNA sequence-structure in phylogenetics: an application to the family Pectinidae (Mollusca: Bivalvia). Mol. Phylogen. Evol. 56, 1059–1067.
San Mauro, D., Agorreta, A., 2010. Molecular systematics: a synthesis of the common methods and the state of knowledge. Cell. Mol. Biol. Lett. 15, 311–341.
Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55, 421–438.
Schultz, J., Maisel, S., Gerlach, D., Müller, T., Wolf, M., 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11, 361–364.
Schultz, J., Wolf, M., 2009. ITS2 sequence–structure analysis in phylogenetics: a how-to manual for molecular systematics. Mol. Phylogen. Evol. 52, 520–523.
Schwarzfeld, M.D., Sperling, F.A.H., 2015. Comparison of five methods for delimitating species in *Ophion* Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae). Mol. Phylogen. Evol. 93, 234–248.
Seibel, P.N., Müller, T., Dandekar, T., Schultz, J., Wolf, M., 2006. 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinf. 7, 498.
Shazib, S.U.A., Vďačný, P., Kim, J.H., Jang, S.W., Shin, M.K., 2016. Molecular phylogeny and species delimitation within the ciliate genus *Spirostomum* (Ciliophora, Postciliodesmatophora, Heterotrichea), using the internal transcribed spacer region. Mol. Phylogen. Evol. 102, 128–144.
Shekhovtsov, S.V., Berman, D.I., Bulakhova, N.A., Vinokurov, N.N., Peltek, S.E., 2017. Phylogeography of *Eisenia nordenskioldi nordenskioldi* (Lumbricidae, Oligochaeta) from the north of Asia. Polar Biol. 41, 237–247.
Shekhovtsov, S.V., Golovanova, E.V., Peltek, S.E., 2013. Cryptic diversity within the Nordenskiold's earthworm, *Eisenia nordenskioldi* subsp. *nordenskioldi* (Lumbricidae, Annelida). Eur. J. Soil Biol. 58, 13–18.
Shen, H.-P., Yeo, D., 2005. Terrestrial earthworms (Oligochaeta) from Singapore. Raffles Bull. Zool. 53, 13–33.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114.
Sites, J.W.J., Crandall, K.A., 1997. Testing species boundaries in biodiversity studies. Conserv. Biol. 11, 1289–1297.
Smith, M.A., Fisher, B.L., Hebert, P.D.N., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Phil. Trans. R. Soc. B 360, 1825.
Srivathsan, A., Meier, R., 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28, 190–194.
Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.
Stephenson, J., 1930. On some Oligochaeta from Berhala Island in the Straits of Malacca. Misc. Zool. Sumatrana 48, 1–5.
Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. PNAS 114, 1607–1612.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods), Version 4. Sinauer Associates, Sunderland, MA.
Taheri, S., James, S., Roy, V., Decaëns, T., Williams, B.W., Anderson, F., Rougerie, R., Chang, C.H., Brown, G., Cunha, L., Stanton, D.W.G., Da Silva, E., Chen, J.H., Lemmon, A.R., Moriarty Lemmon, E., Bartz, M., Baretta, D., Barois, I., Lapied, E., Coulis, M., Dupont, L., 2018. Complex taxonomy of the 'brush tail' peregrine earthworm *Pontoscolex corethrurus*. Mol. Phylogen. Evol. 124, 60–70.
Talavera, G., Dincă, V., Vila, R., 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. Meth. Ecol. Evol. 4, 1101–1110.
Tanabe, A.S., 2011. Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. Mol. Ecol. Resour. 11, 914–921.
Tang, C.Q., Humphreys, A.M., Fontaneto, D., Barraclough, T.G., 2014. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. Meth. Ecol. Evol. 5, 1086–1094.
Terhivuo, J., Saura, A., 1993. Genic and morphological variation of the parthenogenetic earthworm *Aporrectodea rosea* in southern Finland (Oligochaeta, Lumbricidae). Ann. Zool. Fenn. 30, 215–224.
Toussaint, E.F.A., Morinière, J., Lam, A., Turlin, B., Fls, M.B., 2016. Bayesian Poisson tree processes and multispecies coalescent models shed new light on the diversification of Nawab butterflies in the Solomon Islands (Nymphalidae, Charaxinae, *Polyura*). Zool. J. Linn. Soc. 178, 241–256.
Vieites, D.R., Wollenberg, K.C., Andreone, F., Köhler, J., Glaw, F., Vences, M., 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. PNAS 106, 8267.
Wang, X.-C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J.-H., Cai, D., Li, J.-Q., 2015. ITS1: a DNA barcode better than ITS2 in eukaryotes? Mol. Ecol. Resour. 15, 573–586.
Wolf, M., Chen, S., Song, J., Ankenbrand, M., Müller, T., 2013. Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences – A proof of concept. PLoS ONE 8, e66726.
Wolf, M., Friedrich, J., Dandekar, T., Müller, T., 2005. CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. Silico Biol. 5, 291–294.
Xia, X., 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30, 1720–1728.
Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61, 854–865.
Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence

data. PNAS 107, 9264–9269.
Yeates, D.K., Seago, A., Nelson, L., Cameron, S.L., Joseph, L., Trueman, J.W.H., 2011. Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36, 209–217.
Young, I., Coleman, A.W., 2004. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. Mol. Phylogen. Evol. 30, 236–242.
Zachos, F.E., 2016. Species Concepts in Biology. Springer International Publishing, Switzerland.
Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.