Received: 7 October 2017 | Accepted: 18 January 2018

DOI: 10.1111/zsc.12280

ORIGINAL ARTICLE

WILEY Zoologica Scripta

Exploring species boundaries with multiple genetic loci using empirical data from non-biting midges

Xiao-Long Lin | **Elisabeth Stur** | **Torbjørn Ekrem**

Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence
Xiao-Long Lin, Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway.
Email: lin880224@gmail.com

Over the past decade, molecular approaches to species delimitation have seen rapid development. However, species delimitation based on a single locus, for example, DNA barcodes, can lead to inaccurate results in cases of recent speciation and incomplete lineage sorting. Here, we compare the performance of Automatic Barcode Gap Discovery (ABGD), Bayesian Poisson tree processes (PTP), networks, generalized mixed Yule coalescent (GMYC) and Bayesian phylogenetics and phylogeography (BPP) models to delineate cryptic species previously detected by DNA barcodes within *Tanytarsus* (Diptera: Chironomidae) non-biting midges. We compare the results from analyses of one mitochondrial (cytochrome *c* oxidase subunit I [COI]) and three nuclear (alanyl-tRNA synthetase 1 [AATS1], carbamoyl phosphate synthetase 1 [CAD1] and 6-phosphogluconate dehydrogenase [PGD]) protein-coding genes. Our results show that species delimitation based on multiple nuclear DNA markers is largely concordant with morphological variation and delimitations using a single locus, for example, the COI barcode. However, ABGD, GMYC, PTP and network models led to conflicting results based on a single locus and delineate species differently than morphology. Results from BPP analyses on multiple loci correspond best with current morphological species concept. In total, 10 lineages of the *Tanytarsus curticornis* species complex were uncovered. Excluding a Norwegian population of *Tanytarsus brundini* which might have undergone recent hybridization, this suggests six hitherto unrecognized species new to science. Five distinct species are well supported in the *Tanytarsus heusdensis* species complex, including two species new to science.

KEYWORDS

Bayesian phylogenetics, cryptic species, generalized mixed Yule coalescent, maximum likelihood, networks, species delimitation, *Tanytarsus*

1 | INTRODUCTION

Accurate assessment of species boundaries is critical for our understanding of biological diversity and speciation (Pimm et al., 2014). Moreover, limited knowledge of species' evolutionary potential makes global estimates of diversity constrained (Appeltans et al., 2012; Costello, May, & Stork, 2013; Moritz, 2002). Documentation of genetic variation between species, particular through large DNA barcoding initiatives (Hebert, Cywinska, & Ball, 2003; Hebert, Ratnasingham, & de Waard, 2003), has proven very informative for the detection and resolution of species complexes and has provided insights into the evolutionary history of species (Kress, García-Robledo, Uriarte, & Erickson, 2015).

Potentially cryptic species are detected now more frequently than ever. However, the presence of introgression (Gay et al., 2007; Martinsen, Whitham, Turek, & Keim, 2001) and incomplete lineage sorting (Ballard & Whitlock, 2004; Heckman, Mariani, Rasoloarison, & Yoder, 2007; Willyard, Cronn, & Liston, 2009) present obstacles for correct species delimitation using single genetic markers. Moreover, deep mitochondrial genetic divergence is not always accompanied by correspondingly deep nuclear differentiation. Genealogical concordance among multiple loci can provide convincing evidence for species boundaries and validate the presence of genetically distinctive but morphologically cryptic lineages. This has been explored in several insect taxa. For instance, Fossen, Ekrem, Nilsson, and Bergsten (2016) found evidence of genetically distinct lineages in closely related northern European water scavenger beetles using multiple loci, and Low et al. (2016) delineated taxonomic boundaries in the largest species complex of black flies using multiple genes, morphological and chromosomal data. Also within the Chironomidae, several potential cryptic species have been detected by DNA barcodes and subsequently confirmed by nuclear DNA markers and careful analyses of morphological characters (e.g., Anderson, Stur, & Ekrem, 2013).

There are several methods that can be used to investigate species boundaries even with low populational sampling frequency. For example, in DNA barcoding, the so-called barcode gap assumes larger interspecific than intraspecific genetic distance. Although some studies show that barcode gaps can disappear with increased sampling and geographical coverage (Bergsten et al., 2012), many investigated groups tend to retain barcodes gaps as sampling is increased (Čandek & Kuntner, 2015; Huemer, Mutanen, Sefc, & Hebert, 2014; Marín et al., 2017). This is also the case for Chironomidae (own observation in BOLD, Lin, Stur, & Ekrem, 2015). The software Automatic Barcode Gap Discovery (ABGD) recursively searches for barcode gaps in the distribution of pairwise sequence divergences and assigns input sequences into hypothetical species based on pairwise distances. It is recognized that ABGD performs well on large barcode data sets with an appropriate prior of maximum intraspecific divergence (Lin et al., 2015; Pentinsaari, Vos, & Mutanen, 2016; Puillandre, Lambert, Brouillet, & Achaz, 2012); however, it is sensitive to singleton sequences and requires knowledge of threshold values (Pentinsaari et al., 2016; Puillandre et al., 2012).

Standard phylogenetic analyses assuming dichotomous splitting of ancestral branches can also be used to recognize species as monophyletic groups in phylogenetic trees. Character-based, rigorously tested approaches include maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI). While these methods produce results that are based on similarities analysed in a strictly hierarchical framework, techniques using coalescent-based species delimitation combine population genetics and phylogenetics to objectively delineate evolutionary significant units of diversity. For single genetic loci, the generalized mixed Yule coalescent model (GMYC) (Pons et al., 2006) and the Poisson Tree Processes (PTP) (Zhang, Kapli, Pavlidis, & Stamatakis, 2013) are widely used to apply the phylogenetic species concept with assumed reciprocal monophyly in gene trees. The GMYC model combines a Yule species birth model with a neutral coalescent model of intraspecific branching (Fujisawa & Barraclough, 2013; Pons et al., 2006) and has been widely accepted for species delimitation based on single-locus data under many circumstances, including high singleton presence, taxon richness and the presence of gaps in intraspecific sampling coverage (Talavera, Dincă, & Vila, 2013). However, relative to other methods, GMYC has a tendency of oversplitting lineages resulting from errors in the reconstruction of ultrametric input trees (Paz & Crawford, 2012; Pentinsaari et al., 2016; Tänzler, Sagata, Surbakti, Balke, & Riedel, 2012). The method also shows a tendency of overlumping in cases where different lineages are results of rapid and recent divergences (Esselstyn, Evans, Sedlock, Khan, & Heaney, 2012).

The PTP model requires a rooted input tree and assumes that intra- and interspecific substitutions follow distinct Poisson processes and that intraspecific substitutions are discernibly fewer than interspecific substitution (Tang, Humphreys, Fontaneto, & Barraclough, 2014; Zhang et al., 2013). The bPTP model, an updated version of the original PTP with Bayesian support values, provides more accurate results for species delimitation (Zhang et al., 2013).

Single gene trees can be discordant and are not the same as species trees due to processes like incomplete lineage sorting. Using coalescent-based methods on multiple loci can therefore be advantageous as it uncouples gene trees and species trees, and the gene tree coalescences are allowed to be older than species tree coalescences. It is recognized that the coalescent-based method Bayesian phylogenetics and phylogeography (BPP) (Yang & Rannala, 2010) is efficient in delineating closely related species using multiple loci (Yang, 2015; Yang & Rannala, 2010, 2017). The BPP method implements a reversible jump Markov chain Monte Carlo (rjMCMC) search to estimate the posterior probability of species delimitation hypotheses. The method estimates ancestral population sizes and species population divergence times through estimated distributions of gene trees from multiple loci. The method requires sequence data and a guide species tree with defined topology as input and BPP can lead to false species delimitation when the guide tree is inaccurately specified (Rannala & Yang, 2013).

Statistical parsimony network analysis implemented in the TCS software provides a rapid and useful tool for species delimitation (Hart & Sunday, 2007), when applied to non-recombinant loci. TCS calculates the maximum number of mutational steps constituting a 95% parsimonious connection

between two haplotypes and then joins these into networks following specific algorithms (Templeton, Crandall, & Sing, 1992).

Currently, only a few studies have compared the performance of these different analytical methods for multiple genetic markers, especially for delineation of potentially cryptic species in insects.

The genus *Tanytarsus* van der Wulp, 1874 (Diptera: Chironomidae) is the most species-rich genus of the tribe Tanytarsini in subfamily Chironominae with more than 350 valid species worldwide. Larvae of *Tanytarsus* are eurytopic, occur in all types of freshwater, sometimes even in marine or terrestrial environments, and play an important role in freshwater biomonitoring. However, morphological determination of species in some *Tanytarsus* species groups can be notoriously difficult. Additionally, there are many unknown and cryptic *Tanytarsus* species where the boundaries remain uncertain. In a previous study, DNA barcodes uncovered several potential cryptic species within the *Tanytarsus curticornis* Kieffer, 1911 and *Tanytarsus heusdensis* Goetghebuer, 1923 species complexes (Lin et al., 2015).

Based on morphologically similar characteristics in the adult male, the *T. curticornis* species complex previously included *Tanytarsus brundini*, Lindeberg, 1963, *Tanytarsus congus* Lehmann, 1981, *T. curticornis* Kieffer, 1911, *Tanytarsus ikicedeus* Sasa & Suzuki, 1999, *Tanytarsus neotamaoctavus* Ree, Jeong & Nam, 2011, *Tanytarsus pseudocongus* Ekrem, 1999, *Tanytarsus salmelai* Giłka & Paasivirta, 2009, *Tanytarsus tamaoctavus* Sasa, 1980. The *T. heusdensis* species complex included four described species: *T. heusdensis* Goetghebuer, 1923, *Tanytarsus reei* Na & Bae, 2010, *Tanytarsus tamaduodecimus* Sasa, 1983, *Tanytarsus tusimatneous* Sasa & Suzuki, 1999. The similar phenotypes within the *T. curticornis* and *T. heusdensis* species complexes likely have led to misidentifications and an underestimation of species biodiversity in *Tanytarsus*. Thus, these two species complexes are well suited to explore the suitability of different molecular markers and analytical methods in the analyses of species boundaries within non-biting midges. Currently to this study, a taxonomic review of the two species complexes based on morphology and DNA barcodes was conducted and formal descriptions published (Lin, Stur, & Ekrem, 2017). In total, eight species new to science were described as follows: *Tanytarsus adustus* Lin, Stur & Ekrem, 2017, *Tanytarsus heberti* Lin, Stur & Ekrem, 2017, *Tanytarsus madeiraensis* Lin, Stur & Ekrem, 2017, *T. pseudoheusdensis* Lin, Stur & Ekrem, 2017, *Tanytarsus songi* Lin, Stur & Ekrem, 2017, *Tanytarsus thomasi* Lin, Stur & Ekrem, 2017, *Tanytarsus tongmuensis* Lin, Stur & Ekrem, 2017 and *Tanytarsus wangi* Lin, Stur & Ekrem, 2017.

The goal of this study was to investigate whether different molecular markers and analytical tools give similar results when applied to a set of morphologically similar species of Chironomidae, and whether the results are comparable to those achieved from DNA barcodes or morphological analysis alone (op. cit.).

2 | MATERIAL AND METHODS

2.1 | Taxon sampling

We used 63 specimens of the *T. curticornis* and *T. heusdensis* species complexes from Canada, China, Czech Republic, Germany, Norway and Ukraine and included additional five public COI sequences of *T. reei* from South Korea. List of all species, specimens, their individual images, georeferences, primers, sequences and other relevant laboratory data of all sequenced specimens can be seen online in the publicly accessible data sets "*T. curticornis* species complex (DS-TANYSC)," DOI: dx.doi.org/10.5883/DS-TANYSC and "*T. heusdensis* species complex (DS-HEUSDEN)," DOI: dx.doi.org/10.5883/DS-HEUSDEN in the Barcode of Life Data Systems (BOLD) (Ratnasingham & Hebert, 2007, 2013). Specimens were identified morphologically by re-examination of available type material and use of taxonomic revisions and species descriptions (Giłka & Paasivirta, 2009; Kieffer, 1911; Lindeberg, 1963; Na & Bae, 2010; Reiss & Fittkau, 1971; Sasa, 1980).

2.2 | Molecular methods and analyses

Adult specimens were preserved in 85% ethanol, immatures in 96% ethanol, and stored dark at 4°C before morphological and molecular analyses. Genomic DNA of most specimens was extracted from the thorax and head using QIAGEN® DNeasy Blood & Tissue Kit and GeneMole DNA Tissue Kit on a GeneMole® instrument (Mole Genetics, Lysaker, Norway) at the Department of Natural History, NTNU University Museum. When using QIAGEN® DNeasy Blood & Tissue Kit was used, the standard protocol of the kit was followed, except that the final elution volume was 100 µl due to small specimen size. When using GeneMole DNA Tissue Kit, the standard protocol was followed, except that 4 µl Proteinase K was mixed with 100 µl buffer for overnight lysis at 56°C and the final elution volume was 100 µl. After DNA extraction, the cleared exoskeleton was washed with 96% ethanol and mounted in Euparal on the same microscope slide as its corresponding antennae, wings, legs and abdomen following the procedure outlined by Sæther (1969). Vouchers are deposited at the Department of Natural History, NTNU University Museum, Trondheim, Norway, University Museum of Bergen, Bergen, Norway, or the College of Life Sciences, Nankai University, Tianjin, China.

Fragments of one mitochondrial protein-coding gene cytochrome *c* oxidase subunit I (COI) and three nuclear protein-coding genes (alanyl-tRNA synthetase 1

TABLE 1 Overview of gene segments and primer combinations

Gene segment	Oligo name	Oligo sequence (5′–3′)	Reference
Cytochrome *c* oxidase subunit I (COI)	LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer, Black, Hoeh, Lutz, and Vrijenhoek (1994)
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
Carbamoyl phosphate synthetase 1 (CAD1)	54F	GTNGTNTTYCARACNGGNATGGT	Moulton and Wiegmann (2004)
	405R	GCNGTRTGYTCNGGRTGRAAYTG	Moulton and Wiegmann (2004)
Alanyl-tRNA synthetase 1 (AATS1)	A1-92F	TAYCAYCAYACNTTYTTYGARATG	Regier et al. (2008)
	A1-244R	ATNCCRCARTCNATRTGYTT	Su, Narayanan Kutty, and Meier (2008)
6-phosphogluconate dehydrogenase (PGD)	PGD-2F	GATATHGARTAYGGNGAYATGCA	Regier et al. (2008)
	PGD-3R	TRTGIGCNCCRAARTARTC	B. Cassel unpublished

[AATS1], carbamoyl phosphate synthetase 1 [CAD1] and 6-phosphogluconate dehydrogenase [PGD]) were amplified. The primers used to amplify the four regions are shown in Table 1. DNA amplification of COI was carried out in 25 μl reactions using 2.5 μl 10× Takara ExTaq pcr buffer (CL), 2 μl 2.5 mM dNTP mix, 2 μl 25 mM $MgCl_2$, 0.2 μl Takara Ex Taq HS, 1 μl 10 μM of each primer, 2 μl template DNA and 14.3 μl ddH2O. Amplification cycles were performed on a Bio-Rad C1000 Thermal Cycler (Bio-Rad, California, USA) and followed a program with an initial denaturation step of 95°C for 5 min, then followed by 34 cycles of 94°C for 30 s, 51°C for 30 s, 72°C for 1 min and one final extension at 72°C for 3 min. DNA amplifications of selected three nuclear genes were carried out using 2.5 μl 10× Ex Taq Buffer, 2 μl 2.5 mM dNTP Mix, 0.1 μl Ex Taq HS (all TaKaRa Bio INC, Japan), 0.5 μl 25 mM $MgCl_2$ and 1 μl of each 10 μM primer. The amount of template DNA was adjusted according to the DNA concentration and varied between 2 and 5 μl. ddH_2O was added to make a total of 25 μl for each reaction. Amplification cycles were performed on a Bio-Rad C1000 Thermal Cycler and followed a program with an initial denaturation step of 98°C for 10 s, then 94°C for 1 min followed by five cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 2 min and seven cycles of 94°C for 30 s, 51°C for 1 min, 72°C for 2 min and 37 cycles of 94°C for 30 s, 45°C for 20 s, 72°C for 2 min 30 s and one final extension at 72°C for 3 min. PCR products were visualized on a 1% agarose gel, purified using Illustra ExoProStar 1-Step (GE Healthcare Life Sciences, Buckinghamshire, UK) and shipped to MWG Eurofins (Ebersberg, Germany) for bidirectional sequencing using BigDye 3.1 (Applied Biosystems, Foster City, CA, USA) termination. Not all individuals were successfully sequenced for all three nuclear loci (Tables S1 and S2). Sequences were assembled and edited using Sequencher 4.8 (Gene Codes Corp., Ann Arbor, Michigan, USA). The forward and reverse sequences were automatically assembled by the software, and the contig was inspected and edited manually. The appropriate International Union of Pure and Applied Chemistry (IUPAC) code was applied when the ambiguous base calls existed. Sequence information was uploaded on BOLD (www.boldsystems.org) along with an image and collateral information for each voucher specimen. The sequences names were edited using Mesquite 2.7.5 (Maddison & Maddison, 2010). Alignment of the sequences was carried out using the Muscle algorithm (Edgar, 2004) on nucleotides in MEGA 7 (Kumar, Stecher, & Tamura, 2016). Introns were detected with a reference sequence (*Chironomus tepperi*, GenBank: FJ040616) and removed from the alignment using GT-AG rule (Rogers & Wall, 1980). After removing introns, the codons were aligned. No evidence of paralogue copies was observed in any sequences.

2.2.1 | Automatic barcode gap discovery (ABGD)

Although several species had fewer than three specimens, the aligned COI barcodes of *T. curticornis* and *T. heusdensis* species complexes were sorted into hypothetical species using the ABGD method to discover the existence of the DNA barcode gaps and estimate the number of molecular OTUs. The analyses were conducted on the ABGD website (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) with a prior *p* that ranges from .005 to .1 and the K2P model, following default settings.

2.2.2 | Phylogenetic reconstructions

All nuclear genetic markers were concatenated using SequenceMatrix v1.7.8 (Vaidya, Lohman, & Meier, 2011). Phylogenetic analyses used the partition strategies and

models of sequence evolution selected based on the Bayesian information criterion (BIC) in the jModelTest 2.1.7 (Darriba, Taboada, Doallo, & Posada, 2012). We used a maximum-likelihood (ML) phylogenetic analysis on each loci and on the concatenated nuclear gene data set with RAxML8.1.2 (Stamatakis, 2006, 2014) using raxmlGUI v1.5b1 (Silvestro & Michalak, 2012), with unlinked partitions as selected by PartitionFinder (Lanfear, Calcott, Ho, & Guindon, 2012). We used 1,000 bootstrap replicates in a rapid bootstrap analysis and a thorough search for the best scoring ML tree. Results indicated no conflict between nuclear gene trees, but incongruence between mitochondrial and nuclear trees. As a result, we used a concatenated nuclear data set and mitochondrial COI data set separately to reconstruct phylogenetic relationships of all specimens sequenced. We also implemented Bayesian inference in MrBayes v3.2.6 (Ronquist et al., 2012). In the Bayesian analyses, data sets were partitioned by gene, four chains on two runs for 20 million generations, sampled every 1,000 generations with a burn-in of 0.25. Convergence of posterior probabilities in each run was monitored using Tracer v1.6 (Rambaut, Suchard, Xie, & Drummond, 2014); the first 10% of the sampled trees were discarded as burn-in.

2.2.3 | Network analyses

Ambiguous sites in AATS1, CAD1 and PGD sequences were resolved by running a PHASE algorithm (Stephens & Donnelly, 2003; Stephens, Smith, & Donnelly, 2001) under DnaSP V.5.10 (Librado & Rozas, 2009) to create haplotype pairs from these nuclear genes. A haplotype network for each gene segment was reconstructed with PopART (Leigh & Bryant, 2015) using the TCS method (Clement, Posada, & Crandall, 2000; Clement, Snell, Walker, Posada, & Crandall, 2002) with gaps and missing data excluded.

2.2.4 | GMYC

The single-threshold GMYC analyses were conducted in R v3.2.3 (R Core Team 2016) in a Linux environment, with the use of the *splits* package. The ultrametric single-locus gene tree required for the GMYC method was obtained using BEAST 1.8.2 (Drummond, Suchard, Xie, & Rambaut, 2012) on the reduced data set (identical sequences were excluded in RAxML), with 10 million MCMC generations under the Yule speciation model. A strict molecular clock was shown to be appropriate to infer the ultrametric trees through the model comparison using a Bayes factor test in Tracer 1.6 (Rambaut et al., 2014). For the *T. curticornis* species complex, the GTR + G substitution model (Tavaré, 1986) was selected for the AATS1, CAD1 and PGD genes, the HKY + G substitution model (Hasegawa, Kishino, & Yano, 1985) was selected for COI. For the *T. heusdensis* species complex, the GTR + G substitution model was selected for PGD, the HKY + G substitution model was selected for AATS1, CAD1 and COI. Effective sample sizes (ESS) and trace plots estimated with Tracer 1.6 were used as convergence diagnostics, and a burn-in of one million generations was used to avoid suboptimal trees in the final consensus tree. Ultrametric maximum clade credibility (MCC) trees were computed using the mean node heights with TreeAnnotator v1.8.2 for each locus gene.

2.2.5 | PTP

A rooted input tree for each gene was generated with RAxML using rapid Bootstrap with 1,000 replicates and the GTR + G + I substitution model. The PTP and bPTP analyses for each gene were run on a web server (http://species.h-its.org/ptp/) with 500,000 MCMC generations, excluding outgroup, following the remaining default settings.

2.2.6 | BPP

We combined the data sets of the *T. curticornis* and *T. heusdensis* species complexes for the BPP analyses because the statistical power of BPP can be increased when closely related outgroups are included (Rannala & Yang, 2013). The multilocus Bayesian species delimitation method in BPP X1.2.2 (Yang, 2015; Yang & Rannala, 2010) was used with two concatenated data sets (three nuclear loci [AATS1, CAD1 and PGD] and all loci [AATS1, COI, CAD1 and PGD]).

Two start guide species trees were estimated in *BEAST v1.8.2 (Drummond et al., 2012) on the above concatenated data sets and run with 40 million MCMC generations under the Yule Process speciation model. The HKY + G substitution model was selected for AATS1, COI, CAD1 and PGD genes for the *T. curticornis* and *T. heusdensis* species complexes. Effective sample sizes (ESS) and trace plots were examined in Tracer 1.6 and used as convergence diagnostics. A burn-in of one million generations was used to avoid suboptimal trees in the final consensus tree. Ultrametric maximum clade credibility (MCC) trees were computed using the mean node heights with TreeAnnotator v1.8.2. Trees were visualized using FigTree 1.4.3 (available at http://tree.bio.ed.ac.uk/software/figtree).

We used algorithm 0 with a default fine-tuning parameter $\varepsilon = 2$ and species model prior to 1 as uniform rooted trees. The estimation of the marginal posterior probability of speciation associated with each node in the guide tree is performed by summarizing the probabilities for all models that support a particular speciation event with probability values of ≥95% (Leaché & Fujita, 2010). The posterior probabilities for models can be mainly affected by the prior distributions on the ancestral population size (θ) and root age (τ), with large values for θ and small values for τ favouring conservative models containing fewer species (Yang & Rannala, 2010). As no empirical data were available for the

studied species, we ran the species delimitation analyses by the following combinations of gamma distributions: 1. Θ: G (2: 1,000), τ: G (2: 200); 2. Θ: G (2: 1,000), τ: G (2: 2,000); 3. Θ: G (2: 100), τ: G (2: 200); 4. Θ: G (2: 100), τ: G (2: 2,000); 5. Θ: G (2: 100), τ: G (2: 500). All BPP analyses were run for 500,000 generations with sampling every five generations, after discarding an initial burn-in of 20,000 generations. Heredity scalars were set to 1.0 for AATS1, COI, CAD1 and PGD, while algorithm was set to "0." Every analysis was run twice to check for convergence between runs and agreement on the posterior probability of the species delimitation models.

3 | RESULTS

3.1 | Sequencing results

The aligned length (bp) for the four loci used in the full analysis was as follows: AATS1 (408), CAD1 (909), COI (658), PGD (747). The number of variable and parsimony informative sites as well as the average nucleotide composition in each genetic marker is shown in Tables S3 and S4. The COI sequences were heavily AT-biased, especially in third position (>82%), in both species complexes.

3.2 | ABGD

In the *T. curticornis* species complex, the COI sequences were sorted into 10 molecular OTUs, but no definite "barcode gap" was observed in the pairwise K2P distances as some morphospecies showed high intraspecific divergence (Figure S1a). For the *T. heusdensis* species complex, two gaps were observed (Figure S1b), and the COI sequences were sorted into five molecular OTUs when the threshold was placed at 9% according to the second gap in the distribution of pairwise nucleotide distances (Figure S1b).

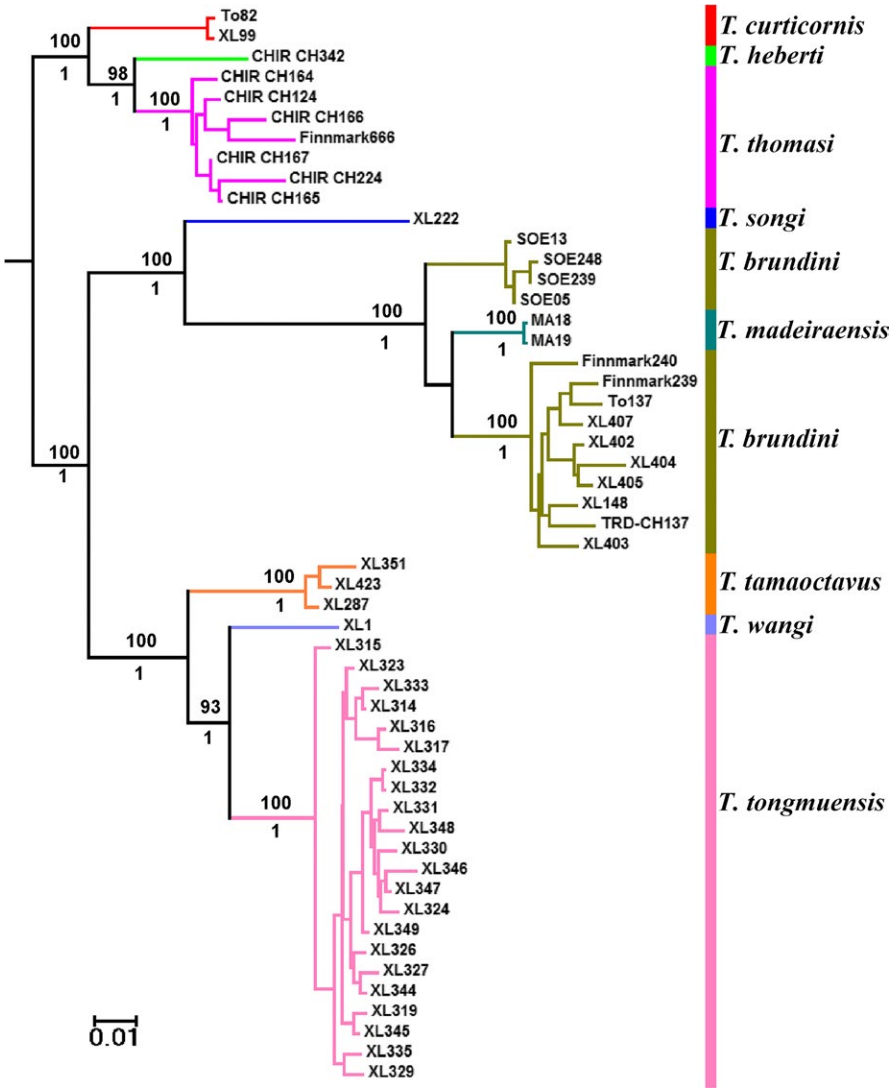

FIGURE 1 Maximum-likelihood tree based on the concatenated nuclear data set of the *Tanytarsus curticornis* species complex. Bootstrap support (1,000 replicates) and posterior probabilities of nodes are indicated above and below the branches, respectively. Only nodes with BS > 70% and/or BP > 0.95 are labelled

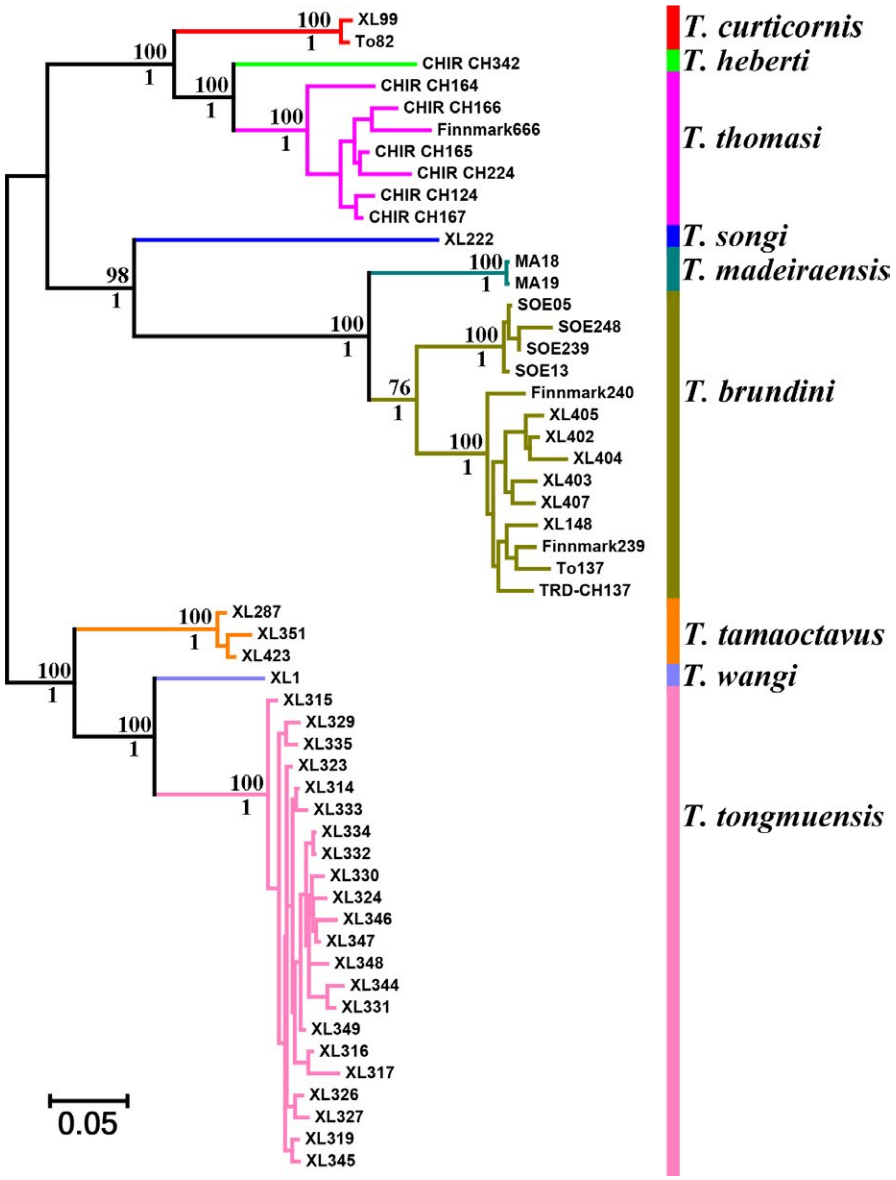

FIGURE 2 Maximum-likelihood tree based on the concatenated mitochondrial and nuclear data set of the *Tanytarsus curticornis* species complex. Bootstrap support (1,000 replicates) and posterior probabilities of nodes are indicated above and below the branches, respectively. Only nodes with BS > 70% and/or BP > 0.95 are labelled

3.3 | Phylogenetic analyses

The phylogenetic analyses under ML and Bayesian inference produced identical trees in the *T. curticornis* and *T. heusdensis* species complexes for the concatenated nuclear genes data. In the *T. curticornis* species complex, the concatenated nuclear genes data yielded 10 well-supported monophyletic groups (Figure 1). A population of *T. brundini* from Sølendet, Norway, was separated from other populations of *T. brundini* in all three nuclear markers (Figure 1) and in the trees resulting from analyses of a concatenated mitochondrial and nuclear data set (Figure 2). All *T. brundini* sequences clustered together in the trees based on COI barcodes (Figure 3). In the *T. heusdensis* species complex, the data set based on concatenated nuclear genes as well as the data set based on COI barcodes yielded five well-supported monophyletic groups (Figure 4a,b).

3.4 | Patterns of haplotype diversity

For the *T. curticornis* species complex, generally the networks based on mitochondrial and nuclear genes showed 10 haplotype groups (Figures 5 and 6). However, sequences of *T. thomasi* were sorted into one haplotype group in AATS1 gene network, two haplotype groups in COI and two haplotype groups in CAD1. The PGD marker gave three haplotype groups. Furthermore, sequences of *T. brundini* were arranged into one haplotype group in the COI network and two groups in all networks based on nuclear markers (Figures 5b and 6).

For the *T. heusdensis* species complex, the TCS network of mitochondrial haplotypes showed six groups where the sequences of *T. reei* split into two haplotype groups (Figure 7a), resulting from the high intraspecific divergence in COI sequences for this species. As expected, the TCS networks

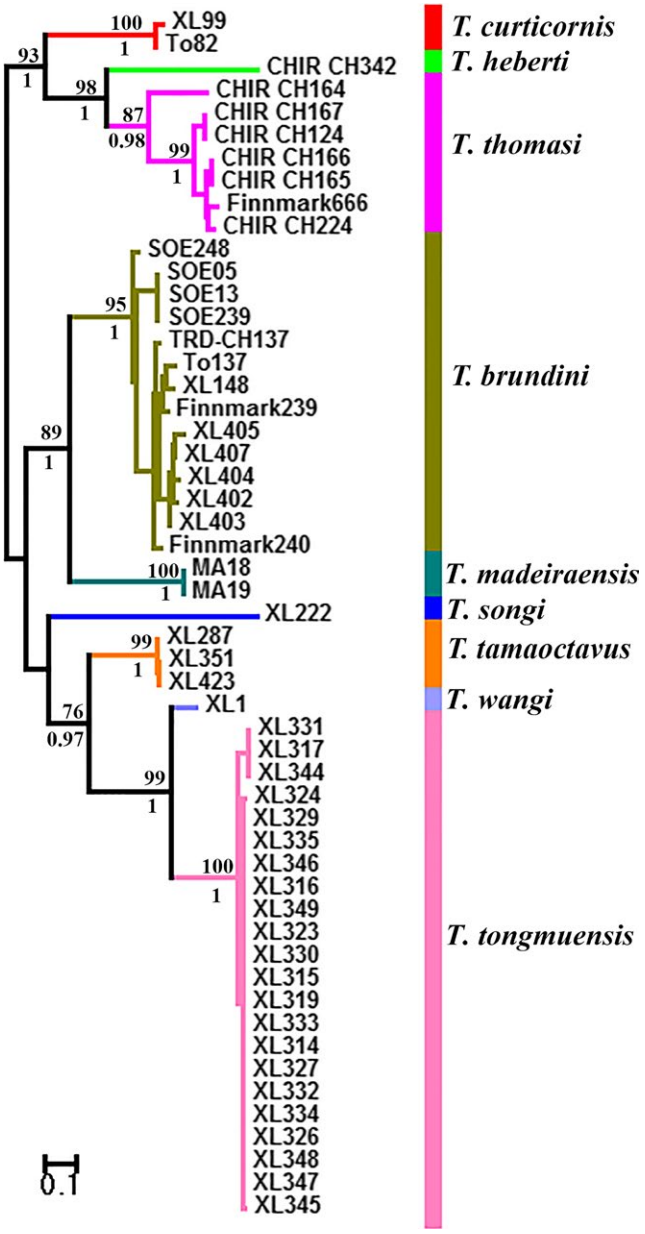

FIGURE 3 Maximum-likelihood tree based on the cytochrome *c* oxidase subunit I barcodes of the *Tanytarsus curticornis* species complex. Bootstrap support (1,000 replicates) and posterior probabilities of nodes are indicated above and below the branches, respectively. Only nodes with BS > 70% and/or BP > 0.95 are labelled

based on all three nuclear alleles confirmed the results obtained in the phylogenetic trees and retrieved five genetic groups (Figure 7b–d).

3.5 | Species delimitation

3.5.1 | Species delimitation with GMYC

In the *T. curticornis* species complex, the GMYC model resulted in a slight oversplitting in COI- (Figure S2), CAD1- (Figure S3) and PGD-data (Figure S4) varying between 10 and 14 OTUs (Table 2). Surprisingly, the GMYC analysis of AATS1 using an ultrametric tree with 44 terminals returned a result where only three OTUs were distinguished (Figure 8). This might be a result of insufficient sampling, low intraspecific divergences and recent speciation of the *T. curticornis* species complex (Timothy Barraclough pers. comm.). We excluded a few sequences with little divergence and generated a new ultrametric tree with 33 individuals under the same settings in BEAST. This AATS1 data set yielded seven OTUs (Figure 9). We also ran the smaller data set for AATS1 using different ultrametric input trees, but still got a lower number of distinguished clusters (3–7 OTUs) compared to the other markers. Recent and rapid divergences can result in uncertainty in the GMYC model and lead to a certain tendency of overlumping (Esselstyn et al., 2012; Reid & Carstens, 2012). However, the variation observed for AATS1 in our data is comparable to that of the other markers. It is difficult to explain why the results are so different with the AATS1 data set and we speculate that the observed pattern might be caused by a systematic error with the GMYC model.

In the *T. heusdensis* species complex, the GMYC analyses delimited five species with a single threshold. The most likely solution showed concordant results between all nuclear markers and corresponded well to defined morphospecies. The analyses of CAD1 (Figure S5) and PGD (Figure S6) both distinguished five molecular OTUs, but as we failed to amplify the AATS1 segment for *T. adustus*, this marker resulted in four distinct molecular OTUs for the *T. heusdensis* species complex (Figure S7). For COI, GMYC analysis resulted in six distinguished clusters as geographically divergent populations of *T. reei* from Germany and Eastern Asia formed two separate OTUs (Figure S8). The lack of similar pattern in the nuclear sequence data sets is difficult to explain, but could be due to higher evolutionary rate of the COI barcodes. Thus, in the *T. heusdensis* species complex, species delimitations based on the AATS1, CAD1 and PGD nuclear markers appear more reliable than those using the mitochondrial COI gene.

3.5.2 | Species delimitation with PTP and bPTP

The bPTP analysis of COI for the *T. curticornis* species complex failed to reach convergence by 500,000 MCMC generations, which is the upper limit of the web server. Disregarding this, the PTP and bPTP analyses yielded similar result with the GMYC analysis of the COI, CAD1 and PGD data sets delineating 10–14 OTUs (Table 3). For the marker AATS1, the PTP and bPTP analyses resulted in 14

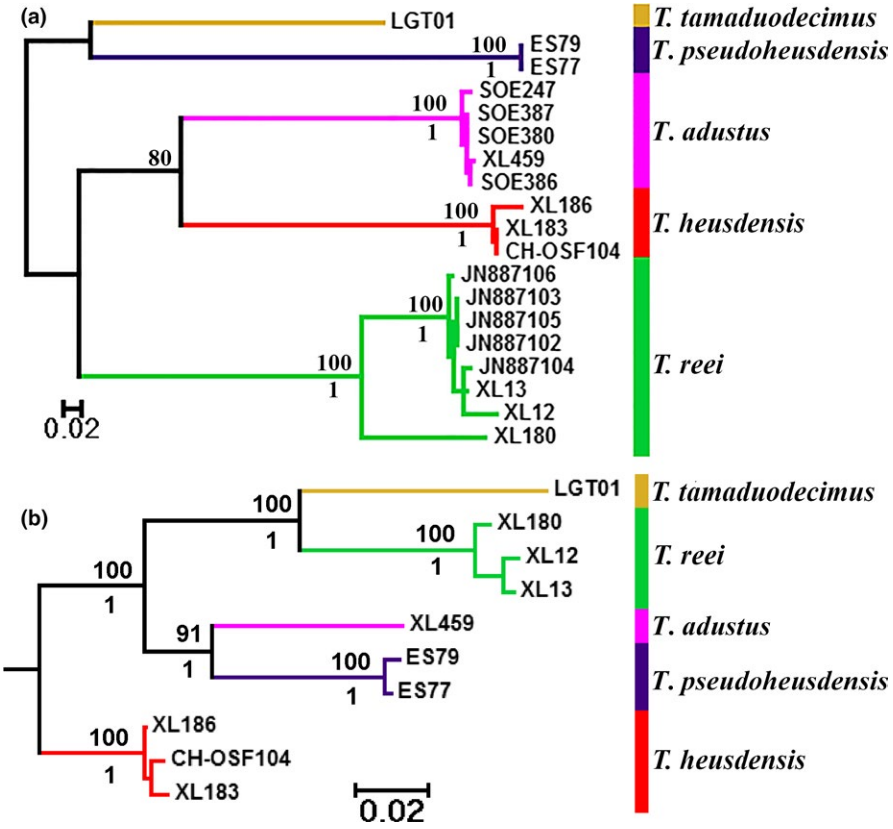

FIGURE 4 Maximum-likelihood tree based on the cytochrome *c* oxidase subunit I (a) and the concatenated nuclear (b) data sets of the *Tanytarsus heusdensis* species complex. Bootstrap support (1,000 replicates) and posterior probabilities of nodes are indicated above and below the branches, respectively. Only nodes with BS > 70% and/or BP > 0.95 are labelled

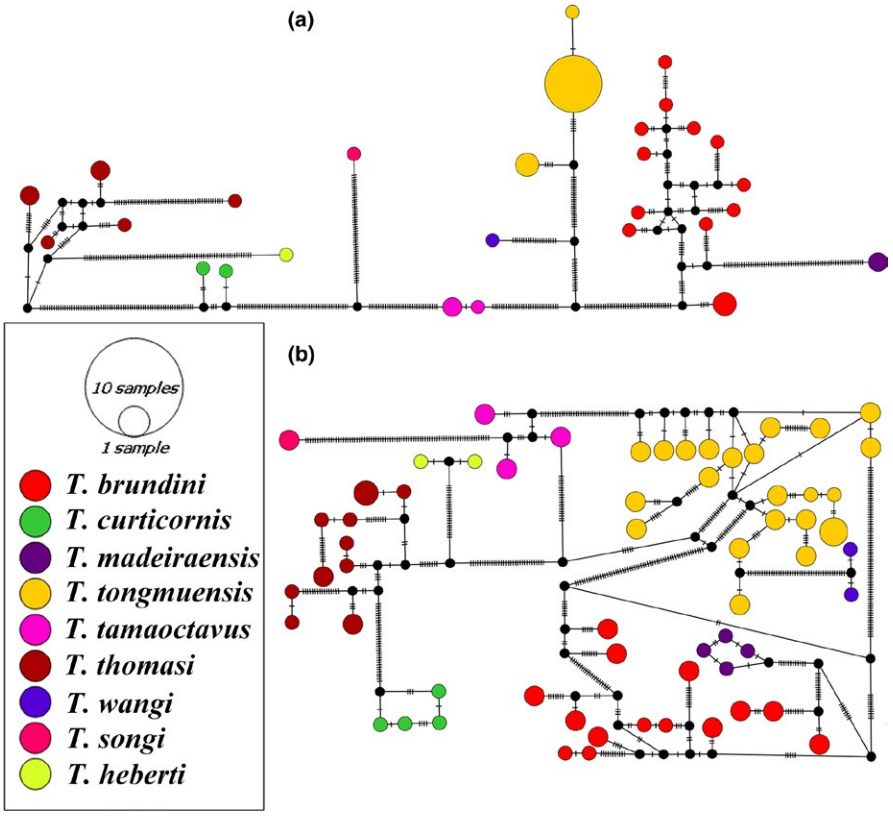

FIGURE 5 TCS haplotype networks based on the mitochondrial cytochrome *c* oxidase subunit I (a) and the nuclear carbamoyl phosphate synthetase 1 (b) data sets of the *Tanytarsus curticornis* species complex. Different colours correspond to the different putative species. Mutations are shown as lines on the branches

and 16 OTUs, respectively, considerably higher than the results of the GMYC analyses as well as more than the expected nine morphospecies. The PTP and bPTP analyses of the *T. heusdensis* species complex yielded same results as the GMYC model, delineating 4–6 OTUs for each marker (Table 3).

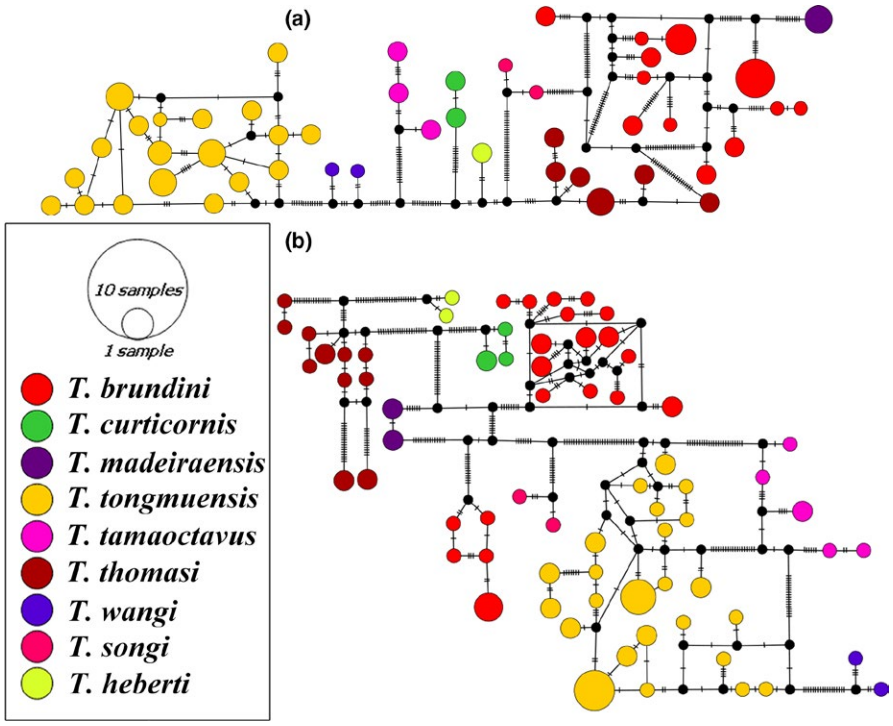

FIGURE 6 TCS haplotype networks based on the nuclear alanyl-tRNA synthetase 1 (a) and 6-phosphogluconate dehydrogenase (b) data sets of the *Tanytarsus curticornis* species complex. Different colours correspond to the different putative species. Mutations are shown as lines on the branches

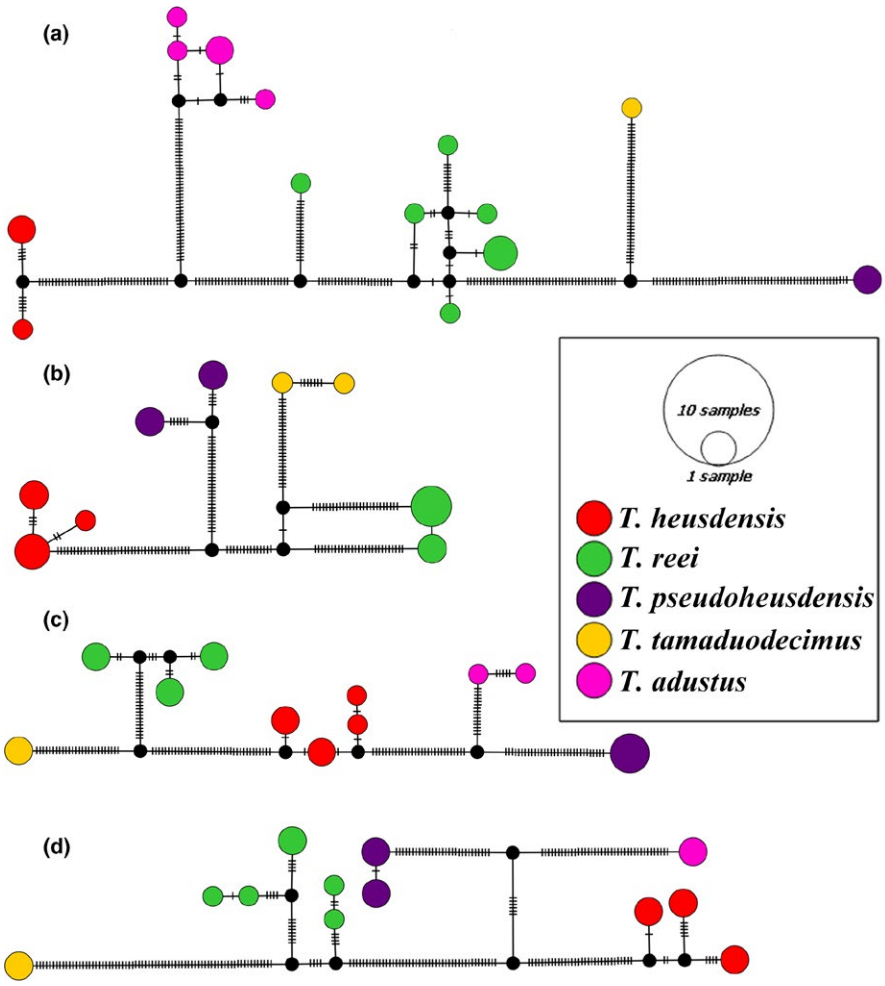

FIGURE 7 TCS haplotype networks based on the mitochondrial cytochrome *c* oxidase subunit I (a) and the nuclear alanyl-tRNA synthetase 1 (b), 6-phosphogluconate dehydrogenase (c), Carbamoyl phosphate synthetase 1 (d) data sets of the *Tanytarsus heusdensis* species complex. Different colours correspond to the different putative species. Mutations are shown as lines on the branches

TABLE 2 Results of species delimitation with the generalized mixed Yule coalescent (GMYC) model

Gene	OTUs	Likelihood of null model	Maximum likelihood of GMYC model	Likelihood ratio
Tanytarsus curticornis species complex				
COI	10	196.7564	203.6784	13.84397
AATS1	3–7	210.2633–299.4416	211.2459–300.2484	1.613635–1.965244
CAD1	12	365.6418	367.1871	3.090735
PGD	14	326.1291	328.7436	5.229101
Tanytarsus heusdensis species complex				
COI	6	62.55684	69.92735	14.74101
AATS1	4	30.13732	34.00053	7.726431
CAD1	5	40.20682	42.42386	4.434082
PGD	5	33.71124	37.10057	6.778666

AATS1, alanyl-tRNA synthetase 1; CAD1, carbamoyl phosphate synthetase 1; COI, cytochrome *c* oxidase subunit I; PGD, 6-phosphogluconate dehydrogenase.

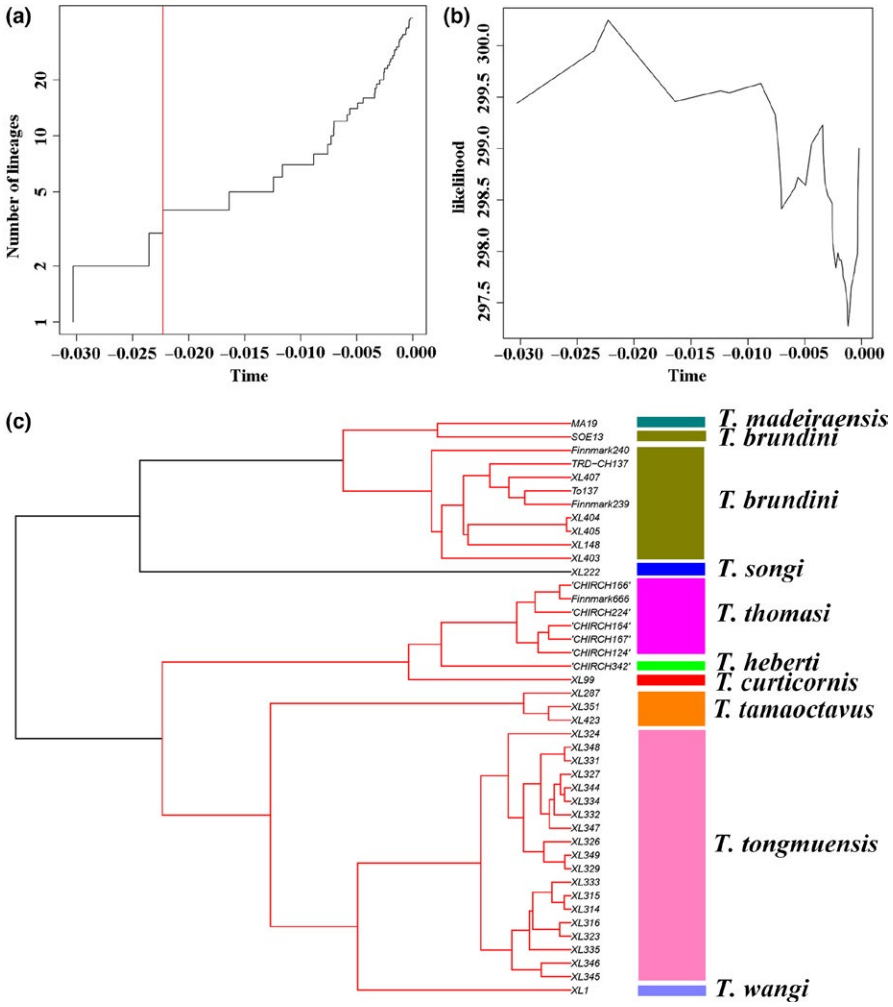

FIGURE 8 Results of the species delimitation analysis for the *Tanytarsus curticornis* species complex according to the generalized mixed Yule coalescent (GMYC) single-threshold model on the alanyl-tRNA synthetase 1 (AATS1) data set with 44 individuals. (a) Lineage-through-time plot based on the ultrametric tree obtained from AATS1 sequences. The sharp increase in branching rate, corresponding to the transition from interspecific to intraspecific branching events, is indicated by a red vertical line. The *x*-axes (both in panels a and b) show substitutions per nucleotide site; (b) likelihood function produced by GMYC to estimate the peak of transition between cladogenesis (interspecific diversification) and allele intraspecific coalescence along the branches; (c) ultrametric tree with 44 individuals obtained in BEAST setting coalescent prior and strict clock model. Red clusters and black lines (singletons) indicate putative species calculated by the model

3.5.3 | Species delimitation with BPP

Initial runs showed errors in the RJ fine-tune variable (≤0) when the parameters were set as follows: Θ: G (2: 100), τ0: G (2: 2,000) and Θ: G (2: 100), τ0: G (2: 1,000). Thus, we used the parameters as Θ: G (2: 100), τ0: G (2: 500) in the final runs. The results from BPP analyses on the concatenated data sets of both nuclear genes and all genes (including COI) showed that 15 candidate species were well supported (posterior probabilities 0.99–1.00; Table 4).

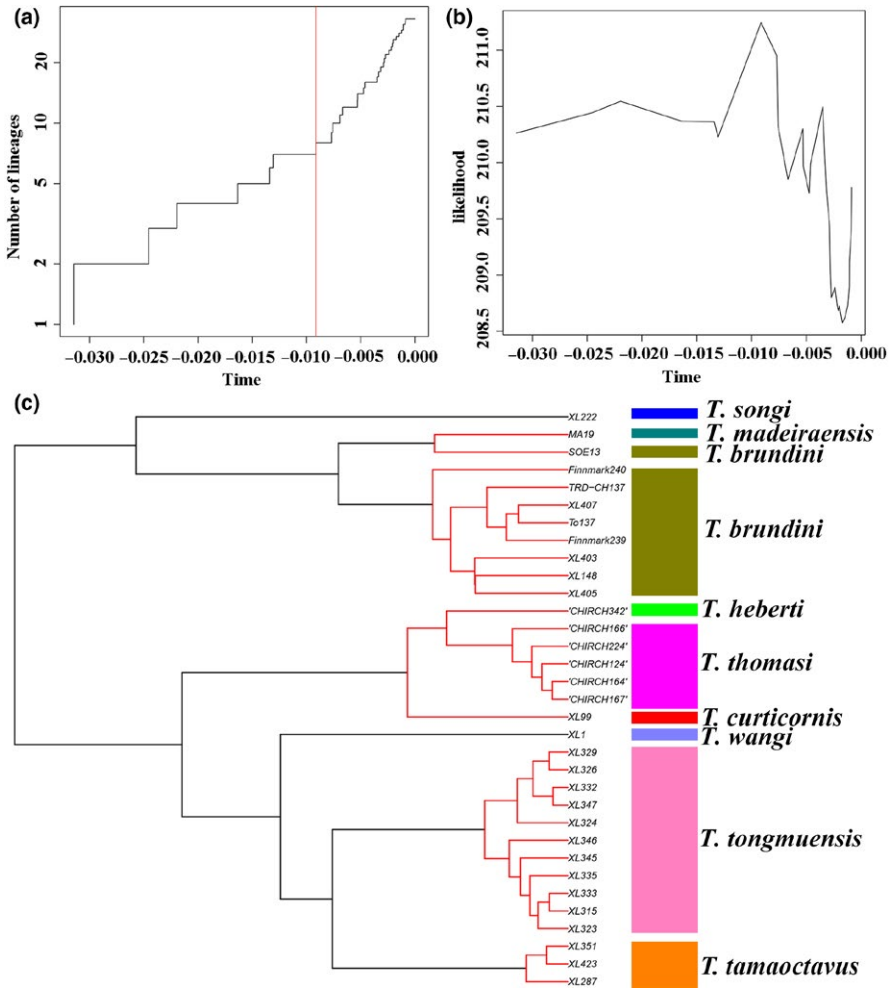

FIGURE 9 Results of the species delimitation analysis for the *Tanytarsus curticornis* species complex according to the generalized mixed Yule coalescent single-threshold model on the alanyl-tRNA synthetase 1 (AATS1) data set with 33 individuals. (a) Lineage-through-time plot based on the ultrametric tree obtained from AATS1 sequences. The sharp increase in branching rate, corresponding to the transition from interspecific to intraspecific branching events, is indicated by a red vertical line. The *x*-axes (both in panels a and b) show substitutions per nucleotide site; (b) likelihood function produced by generalized mixed Yule coalescent to estimate the peak of transition between cladogenesis (interspecific diversification) and allele intraspecific coalescence along the branches; (c) ultrametric tree with 33 individuals obtained in BEAST setting coalescent prior and strict clock model. Red clusters and black lines (singletons) indicate putative species calculated by the model

TABLE 3 Results of species delimitation with poisson tree processes (PTP) and Bayesian Poisson tree processes (bPTP) models

	Tanytarsus curticornis species complex			*Tanytarsus heusdensis* species complex		
	PTP	**bPTP**	**Acceptance rate**	**PTP**	**bPTP**	**Acceptance rate**
COI	10 OTUs	–	0.52	6 OTUs	6 OTUs	0.39
AATS1	14 OTUs	16 OTUs	0.67	4 OTUs	4 OTUs	0.34
CAD1	11 OTUs	11 OTUs	0.38	5 OTUs	5 OTUs	0.37
PGD	13 OTUs	14 OTUs	0.56	5 OTUs	5 OTUs	0.39

TABLE 4 Posterior probabilities for the number of delimited species using different priors for model parameters in Bayesian phylogenetics and phylogeography on concatenated data sets of nuclear markers and all genetic markers

Prior	**Posterior probability for the number of delimited species (all nuclear genes)**	**Posterior probability for the number of delimited species (all genes)**
Θ: G (2: 1,000), τ: G (2: 200)	$p_{15} = 1.000$	$p_{13} = 1.000$
Θ: G (2: 1,000), τ: G (2: 2,000)	$p_{15} = 1.000$	$p_{15} = 1.000$
Θ: G (2: 100), τ: G (2: 200)	$p_{15} = .995$, $p_{14} = .005$	$p_{15} = .999$, $p_{14} = .001$
Θ: G (2: 100), τ: G (2: 500)	$p_{15} = .995$, $p_{14} = .005$	$p_{15} = .999$, $p_{14} = .001$

The *T. curticornis* species complex was divided into 10 species where the Sølendet population of *T. brundini* was isolated as a separate species. For the *T. heusdensis* species complex, both data sets isolated five species in the BPP analyses.

4 | DISCUSSION

Several previous studies have evaluated the performance of the species delimitation approaches used here on single-locus data. GMYC appears to have a tendency to oversplit and sometimes overlump lineages due to sampling bias, differences in population size and speciation rates (Dellicour & Flot, 2015; Esselstyn et al., 2012; Fujisawa & Barraclough, 2013; Pentinsaari et al., 2016; Reid & Carstens, 2012; Talavera et al., 2013). PTP generates more robust results or results that are highly congruent with GMYC (Pentinsaari et al., 2016; Tang et al., 2014). While ABGD and parsimony networks appear to perform well when speciation rates are low and interspecific divergence is high (Dellicour & Flot, 2015). When sampling is comprehensive within species and effective population sizes are small, these species delimitation methods using single-locus data sets generally yield the same results.

However, when intraspecific divergence is high and interspecific divergence is low, species delimitation models using single-locus data set often are unable to separate species properly. Moreover, non-monophyletic species caused by infrequent horizontal gene flow and incomplete lineage sorting may also lead to inaccurate species delimitation results when using single-locus data set (Camargo, Morando, Avila, & Sites, 2012; Fontaneto, Flot, & Tang, 2015; Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012). To overcome these problems, species delimitation using multiple loci can be used. The BPP species delimitation method is perhaps the most popular method using multiple loci and has been proven efficient in species separation (Fehlauer-Ale et al., 2014; Leaché et al., 2017; Yang & Rannala, 2010). Using various species delimitation approaches with different criteria and searching a consensus from different outcomes may increase our confidence regarding species boundaries of target groups.

In this study, species delimitation analyses based on single loci using ABGD, parsimony networks, GMYC and PTP give the same results for the *T. heusdensis* species complex, but different results for the *T. curticornis* species complex. The BPP species delimitation model is based on multiple loci and provides a result that better reflects the observations on morphological divergence.

The above results show that while the COI marker divides the *T. curticornis* species complex into nine lineages, the three nuclear genes identify 10 lineages. The conflicting result between mitochondrial and nuclear genes is caused by a Norwegian population of *T. brundini* which has COI sequences similar to other populations of *T. brundini*, while all nuclear markers show deep divergence. We are not able to detect morphological differences between the specimens of this particular population and other populations of *T. brundini* at present, but have only compared adult males. It is widely recognized that the discordance among gene trees can be caused by the stochastic process of lineage sorting (Maddison, 1997; Pamilo & Nei, 1988) and numerous examples exist in literature. For instance, the phylogenetic incongruence in the *Drosophila melanogaster* species complex is caused by incomplete lineage sorting (Pollard, Iyer, Moses, & Eisen, 2006). Thus, incomplete lineage sorting in the nuclear markers is a possible explanation for the discordance between mitochondrial and nuclear gene trees. Another explanation can be horizontal gene transfer by infrequent hybridization between two cryptic species, resulting in equal mitochondrial genotypes while keeping divergent nuclear genomes. This pattern is previously documented for crickets (Shaw, 2002) and water fleas (Taylor, Sprenger, & Ishida, 2005) and would fit well with our observations.

Based on the observed genetic divergence, we searched and found fine but consistent morphological differences that separate six of the distinct clusters of the *T. curticornis* species complex from previously described species (*T. heberti*, *T. madeiraensis*, *T. songi*, *T. thomasi*, *T. tongmuensis* and *T. wangi*). These taxa are diagnosed and described elsewhere (Lin et al., 2017).

The five mitochondrial DNA lineages we previously identified in the *T. heusdensis* species complex (Lin et al., 2015) are also recognized in the analyses based on nuclear markers. The divergence among the *T. heusdensis* sensu lato lineages is on par with that between other recognized *Tanytarsus* species, suggesting that the complex as of now comprises five distinct species. Two are recognized as new to science based on morphology (*T. adustus* and *T. pseudoheusdensis*) and are described by Lin et al. (2017).

Overall, DNA barcodes are very effective in distinguishing chironomid species and provide novel insight into the taxonomy of some groups. However, DNA barcodes occasionally fail to separate genetically distinct species and can give inaccurate results due to deep intraspecific divergence (Meier, Shiyang, Vaidya, & Ng, 2006; Zhou, Adamowicz, Jacobus, DeWalt, & Hebert, 2009). Multiple reasons why gene trees and species trees are often not the same exist (Maddison, 1997; Nichols, 2001; Rosenberg, 2002) and incomplete lineage sorting, insufficient taxon sampling, horizontal gene flow or recent speciation can be difficult to distinguish regardless of analytic method implemented. Thus, species delimitation analyses based on multiple loci with coalescent models are widely accepted as it improves the discovery, resolution, consistency and stability of our understanding of species (Fujita et al., 2012; Leaché & Fujita,

2010). Our findings are consistent with those of Dupuis, Roe, and Sperling (2012) who found that one marker is not enough for species delimitation in closely related animals and fungi. Also in insects, multiloci-based species delimitation has proved to be more suitable as it is not equally susceptible to introgression (Boykin et al., 2014; Dincă, Lukhtanov, Talavera, & Vila, 2011; Hsieh, Ko, Chung, & Wang, 2014; Malausa et al., 2011; Schutze et al., 2015; Song & Ahn, 2014). Our results are in agreement with this and demonstrate that species delimitation analyses based on multiple loci give a more credible result than a single locus.

The discovery, description and naming of cryptic species obviously are important for both biological conservation and estimates of species richness (e.g., Delić, Trontelj, Rendoš, & Fišer, 2017; Hebert, Penton, Burns, Janzen, & Hallwachs, 2004). But it also can be of significance for environmental management if the cryptic lineages have different have preferences or react differently to environment stressors (Feckler et al., 2014). We do not have sufficient information to evaluate the potential ecological differences between the species in the *T. curticornis* and *T. heusdensis* complexes, but acknowledge the possibility for such comparative studies now that molecular characterization of these taxa exists.

5 | CONCLUSION

In our study, species delimitations based on the AATS1, CAD1 and PGD nuclear DNA markers were largely consistent with delimitations using the mitochondrial COI gene alone. The results were only conflicting for the species *T. brundini*, in which nuclear markers separated a Norwegian population as a distinct species. Bayesian species delimitation based on multiple loci gives a more reliable result than single locus-based species delimitation methods. In total, 15 species of the *T. curticornis* and *T. heusdensis* species complexes were differentiated genetically. Subsequent detection of morphological characters that support these species boundaries led to the integrative discovery of eight species new to science.

ACKNOWLEDGEMENTS

This article is part of the first author's thesis for the partial fulfilment of a PhD degree of the Norwegian University of Science and Technology, Norway, entitled "Systematics and evolutionary history of *Tanytarsus* van der Wulp, 1874 (Diptera: Chironomidae)." Many thanks to Xin-Hua Wang and Chao Song (College of Life Sciences, Nankai University, China), Viktor Baranov (Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany), Qiang Wang (Shanghai Entry-Exit Inspection and Quarantine Bureau, Shanghai, China) for collecting and sending material and Koichiro Kawai (Graduate School of Biosphere Science, Hiroshima University, Hiroshima, Japan) for sharing DNA barcode data and morphological observations of Japanese specimens.

ORCID

Xiao-Long Lin http://orcid.org/0000-0001-6544-6204

REFERENCES

Anderson, A. M., Stur, E., & Ekrem, T. (2013). Molecular and morphological methods reveal cryptic diversity and three new species of Nearctic *Micropsectra* (Diptera: Chironomidae). *Freshwater Science*, *32*, 892–921. https://doi.org/10.1899/12-026.1

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., … Błażewicz-Paszkowycz, M. (2012). The magnitude of global marine species diversity. *Current Biology*, *22*, 2189–2202. https://doi.org/10.1016/j.cub.2012.09.036

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., … Ribera, I. (2012). The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, *61*, 851–869. https://doi.org/10.1093/sysbio/sys037

Boykin, L. M., Schutze, M. K., Krosch, M. N., Chomič, A., Chapman, T. A., Englezou, A., … Cameron, S. L. (2014). Multi-gene phylogenetic analysis of south-east Asian pest members of the *Bactrocera dorsalis* species complex (Diptera: Tephritidae) does not support current taxonomy. *Journal of Applied Entomology*, *138*, 235–253. https://doi.org/10.1111/jen.12047

Camargo, A., Morando, M., Avila, L., & Sites, J. Jr (2012). Species delimitation with ABC and other coalescent-based methods in lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution*, *66*, 2834–2849. https://doi.org/10.1111/j.1558-5646.2012.01640.x

Čandek, K., & Kuntner, M. (2015). DNA barcoding gap: Reliable species identification over morphological and geographical scales. *Molecular Ecology Resources*, *15*, 268–277. https://doi.org/10.1111/1755-0998.12304

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Clement, M., Snell, Q., Walker, P., Posada, D., & Crandall, K. (2002). *TCS: Estimating gene genealogies* (p. 184). *Proceeding 16th International Parallel Distributed Processing Symposium.*

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct? *Science*, *339*, 413–416. https://doi.org/10.1126/science.1230318

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*, 772. https://doi.org/10.1038/nmeth.2109

Delić, T., Trontelj, P., Rendoš, M., & Fišer, C. (2017). The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Scientific Reports*, *7*, 3391. https://doi.org/10.1038/s41598-017-02938-z

Dellicour, S., & Flot, J. F. (2015). Delimiting species-poor datasets using single molecular markers: A study of barcode gaps, haplowebs and

GMYC. *Systematic Biology*, *64*, 900–908. https://doi.org/10.1093/sysbio/syu130

Dincă, V., Lukhtanov, V. A., Talavera, G., & Vila, R. (2011). Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nature Communications*, *2*, 324. https://doi.org/10.1038/ncomms1329

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. https://doi.org/10.1093/molbev/mss075

Dupuis, J. R., Roe, A. D., & Sperling, F. A. (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology*, *21*, 4422–4436. https://doi.org/10.1111/j.1365-294X.2012.05642.x

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Esselstyn, J. A., Evans, B. J., Sedlock, J. L., Khan, F. A. 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. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 3678–3686. https://doi.org/10.1098/rspb.2012.0705

Feckler, A., Zubrod, J. P., Thielsch, A., Schwenk, K., Schulz, R., & Bundschuh, M. (2014). Cryptic species diversity: An overlooked factor in environmental management? *Journal of Applied Ecology*, *51*, 958–967. https://doi.org/10.1111/1365-2664.12246

Fehlauer-Ale, K. H., Mackie, J. A., Lim-Fong, G. E., Ale, E., Pie, M. R., & Waeschenbach, A. (2014). Cryptic species in the cosmopolitan *Bugula neritina* complex (Bryozoa, Cheilostomata). *Zoologica Scripta*, *43*, 193–205. https://doi.org/10.1111/zsc.12042

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

Fontaneto, D., Flot, J. F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity*, *45*, 433–451. https://doi.org/10.1007/s12526-015-0319-7

Fossen, E. I., Ekrem, T., Nilsson, A. N., & Bergsten, J. (2016). Species delimitation in northern European water scavenger beetles of the genus *Hydrobius* (Coleoptera, Hydrophilidae). *ZooKeys*, *564*, 71–120. https://doi.org/10.3897/zookeys.564.6558

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. *Systematic Biology*, *62*, 707–724. https://doi.org/10.1093/sysbio/syt033

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, *27*, 480–488. https://doi.org/10.1016/j.tree.2012.04.012

Gay, L., Neubauer, G., Zagalska-Neubauer, M., Debain, C., Pons, J. M., David, P., & Crochet, P. A. (2007). Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact. *Molecular Ecology*, *16*, 3215–3227. https://doi.org/10.1111/j.1365-294X.2007.03363.x

Giłka, W., & Paasivirta, L. (2009). Evaluation of diagnostic characters of the *Tanytarsus chinyensis* group (Diptera: Chironomidae), with description of a new species from Lapland. *Zootaxa*, *2197*, 31–42. https://doi.org/10.5281/zenodo.189527

Hart, M. W., & Sunday, J. (2007). Things fall apart: Biological species form unconnected parsimony networks. *Biology Letters*, *3*, 509–512. https://doi.org/10.1098/rsbl.2007.0307

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*, 160–174. https://doi.org/10.1007/BF02101694

Hebert, P. D. N., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*, 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 14812–14817. https://doi.org/10.1073/pnas.0406166101

Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*, S96–S99. https://doi.org/10.1098/rsbl.2003.0025

Heckman, K. L., Mariani, C. L., Rasoloarison, R., & Yoder, A. D. (2007). Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Molecular Phylogenetics and Evolution*, *43*, 353–367. https://doi.org/10.1016/j.ympev.2007.03.005

Hsieh, C. H., Ko, C. C., Chung, C. H., & Wang, H. Y. (2014). Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Phylogenetics and Evolution*, *76*, 172–180. https://doi.org/10.1016/j.ympev.2014.03.021

Huemer, P., Mutanen, M., Sefc, K. M., & Hebert, P. D. (2014). Testing DNA barcode performance in 1000 species of European Lepidoptera: Large geographic distances have small genetic impacts. *PLoS ONE*, *9*, e115774. https://doi.org/10.1371/journal.pone.0115774

Kieffer, J. J. (1911). Nouvelles descriptions de chironomides obtenus d'éclosion. *Bulletin de la Société d'Histoire naturelle de Metz*, *27*, 1–60.

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, *30*, 25–35. https://doi.org/10.1016/j.tree.2014.10.008

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701. https://doi.org/10.1093/molbev/mss020

Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological Sciences*, *277*, 3071–3077. https://doi.org/10.1098/rspb.2010.0662

Leaché, A. D., Grummer, J. A., Miller, M., Krishnan, S., Fujita, M. K., Böhme, W., … Ofori-Boateng, C. (2017). Bayesian inference of species diffusion in the West African *Agama agama* species group

(Reptilia, Agamidae). *Systematics and Biodiversity*, *15*, 192–203. https://doi.org/10.1080/14772000.2016.1238018

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116. https://doi.org/10.1111/2041-210X.12410

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lin, X. L., Stur, E., & Ekrem, T. (2015). Exploring genetic divergence in a species-rich insect genus using 2790 DNA Barcodes. *PLoS ONE*, *10*, e0138993. https://doi.org/10.1371/journal.pone.0138993

Lin, X. L., Stur, E., & Ekrem, T. (2017). DNA barcodes and morphology reveal unrecognized species of Chironomidae (Diptera). *Insect Systematics and Evolution*. https://doi.org/10.1163/18763 12X-00002172

Lindeberg, B. (1963). Taxonomy, biology and biometry of *Tanytarsus curticornis* Kieff. and *T. brundini* n. sp. (Dipt., Chironomidae). *Annales Entomologici Fennici*, *29*, 118–130.

Low, V. L., Takaoka, H., Pramual, P., Adler, P. H., Ya'cob, Z., Huang, Y. T., … Sofian-Azirun, M. (2016). Delineating taxonomic boundaries in the largest species complex of black flies (Simuliidae) in the Oriental Region. *Scientific Reports*, *6*, 20346. https://doi.org/10.1038/srep20346

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523–536. https://doi.org/10.1093/sysbio/46.3.523

Maddison, W. P., & Maddison, D. R. (2010). *Mesquite: A modular system for evolutionary analysis*. Version 2.75. Retrieved from mesquiteproject.org/mesquite/download/download.html.

Malausa, T., Fenis, A., Warot, S., Germain, J. F., Ris, N., Prado, E., … Couloux, A. (2011). DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae). *Journal of Applied Entomology*, *135*, 142–155. https://doi.org/10.1111/j.1439-0418.2009.01495.x

Marín, M., Cadavid, I., Valdés, L., Álvarez, C., Uribe, S., Vila, R., & Pyrcz, T. W. (2017). DNA barcoding of an assembly of montane Andean butterflies (Satyrinae): Geographical Scale and Identification Performance. *Neotropical Entomology*, *46*, 1–10.

Martinsen, G. D., Whitham, T. G., Turek, R. J., & Keim, P. (2001). Hybrid populations selectively filter gene introgression between species. *Evolution*, *55*, 1325–1335. https://doi.org/10.1111/j.0014-3820.2001.tb00655.x

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*, 715–728. https://doi.org/10.1080/10635150600969864

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, *51*, 238–254. https://doi.org/10.1080/10635150252899752

Moulton, J. K., & Wiegmann, B. M. (2004). Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Molecular Phylogenetics and Evolution*, *31*, 363–378. https://doi.org/10.1016/S1055-7903(03)00284-7

Na, K. B., & Bae, Y. J. (2010). New species of *Stictochironomus*, *Tanytarsus* and *Conchapelopia* (Diptera: Chironomidae) from Korea. *Entomological Research Bulletin*, *26*, 33–39.

Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology and Evolution*, *16*, 358–364. https://doi.org/10.1016/S0169-5347(01)02203-0

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, *5*, 568–583.

Paz, A., & Crawford, A. J. (2012). Molecular-based rapid inventories of sympatric diversity: A comparison of DNA barcode clustering methods applied to geography-based vs clade-based sampling of amphibians. *Journal of Biosciences*, *37*, 887–896. https://doi.org/10.1007/s12038-012-9255-x

Pentinsaari, M., Vos, R., & Mutanen, M. (2016). Algorithmic single-locus species delimitation: Effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular Ecology Resources*, *17*, 393–404. https://doi.org/10.1111/1755-0998.12557

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., … Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752. https://doi.org/10.1126/science.1246752

Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006). Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. *PLoS Genetics*, *2*, e173. https://doi.org/10.1371/journal.pgen.0020173

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., … Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*, 595–609. https://doi.org/10.1080/10635150600852011

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*, 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*. Retrieved from http://beast.bio.ed.ac.uk/Tracer

Rannala, B., & Yang, Z. (2013). Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, *194*, 245–253. https://doi.org/10.1534/genetics.112.149039

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, *7*, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: The barcode index number (BIN) system. *PLoS ONE*, *8*, e66213. https://doi.org/10.1371/journal.pone.0066213

Regier, J. C., Shultz, J. W., Ganley, A. R., Hussey, A., Shi, D., Ball, B., … Cunningham, C. W. (2008). Resolving arthropod phylogeny: Exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, *57*, 920–938. https://doi.org/10.1080/10635150802570791

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 Evolutionary Biology*, *12*, 196. https://doi.org/10.1186/1471-2148-12-196

Reiss, F., & Fittkau, E. J. (1971). Taxonomie und Ökologie europäisch verbreiteter *Tanytarsus*-Arten (Chironomidae, Diptera). *Archiv für Hydrobiologie, Supplement*, *40*, 75–200.

Rogers, J., & Wall, R. (1980). A mechanism for RNA splicing. *Proceedings of the National Academy of Sciences of the United States of America*, *77*, 1877–1879. https://doi.org/10.1073/pnas.77.4.1877

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenberg, N. A. (2002). The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, *61*, 225–247. https://doi.org/10.1006/tpbi.2001.1568

Sæther, O. A. (1969). Some Nearctic Podonominae, Diamesinae, and Orthocladiinae (Diptera: Chironomidae). *Bulletin of the Fisheries Research Board of Canada*, *170*, 1–154.

Sasa, M. (1980). Studies on chironomid midges of the Tama River. Part 2. Description of 20 species of Chironominae recovered from a tributary. *Research Report from the National Institute for Environmental Studies, Japan*, *13*, 9–107.

Schutze, M. K., Mahmood, K., Pavasovic, A., Bo, W., Newman, J., Clarke, A. R., … Cameron, S. L. (2015). One and the same: Integrative taxonomic evidence that *Bactrocera invadens* (Diptera: Tephritidae) is the same species as the Oriental fruit fly *Bactrocera dorsalis*. *Systematic Entomology*, *40*, 472–486. https://doi.org/10.1111/syen.12114

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 16122–16127. https://doi.org/10.1073/pnas.242585899

Silvestro, D., & Michalak, I. (2012). raxmlGUI: A graphical front-end for RAxML. *Organisms Diversity and Evolution*, *12*, 335–337. https://doi.org/10.1007/s13127-011-0056-0

Song, J. H., & Ahn, K. J. (2014). Species delimitation in the *Aleochara fucicola* species complex (Coleoptera: Staphylinidae: Aleocharinae) and its phylogenetic relationships. *Zoologica Scripta*, *43*, 629–640. https://doi.org/10.1111/zsc.12077

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics*, *73*, 1162–1169. https://doi.org/10.1086/379378

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*, 978–989. https://doi.org/10.1086/319501

Su, K. F. Y., Narayanan Kutty, S., & Meier, R. (2008). Morphology versus molecules: The phylogenetic relationships of Sepsidae (Diptera: Cyclorrhapha) based on morphology and DNA sequence data from ten genes. *Cladistics*, *24*, 902–916. https://doi.org/10.1111/j.1096-0031.2008.00222.x

Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*, 1101–1110. https://doi.org/10.1111/2041-210X.12107

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. *Methods in Ecology and Evolution*, *5*, 1086–1094. https://doi.org/10.1111/2041-210X.12246

Tänzler, R., Sagata, K., Surbakti, S., Balke, M., & Riedel, A. (2012). DNA barcoding for community ecology-how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS ONE*, *7*, e28832. https://doi.org/10.1371/journal.pone.0028832

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Taylor, D. J., Sprenger, H. L., & Ishida, S. (2005). Geographic and phylogenetic evidence for dispersed nuclear introgression in a daphniid with sexual propagules. *Molecular Ecology*, *14*, 525–537. https://doi.org/10.1111/j.1365-294X.2005.02415.x

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, *132*, 619–633.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*, 171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Willyard, A., Cronn, R., & Liston, A. (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution*, *52*, 498–511. https://doi.org/10.1016/j.ympev.2009.02.011

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*, 854–865. https://doi.org/10.1093/czoolo/61.5.854

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9264–9269. https://doi.org/10.1073/pnas.0913022107

Yang, Z., & Rannala, B. (2017). Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, *26*, 3028–3036. https://doi.org/10.1111/mec.14093

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*, 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Zhou, X., Adamowicz, S. J., Jacobus, L. M., DeWalt, R. E., & Hebert, P. D. N. (2009). Towards a comprehensive barcode library for arctic life-Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology*, *6*, 30. https://doi.org/10.1186/1742-9994-6-30

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lin X-L, Stur E, Ekrem T. Exploring species boundaries with multiple genetic loci using empirical data from non-biting midges. *Zool Scr*. 2018;00:1–17. https://doi.org/10.1111/zsc.12280