



Contents lists available at ScienceDirect

## Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

## Comparative molecular species delimitation in the charismatic Nawab butterflies (Nymphalidae, Charaxinae, *Polyura*)<sup>☆</sup>



Emmanuel F.A. Toussaint<sup>a,\*,1</sup>, Jérôme Morinière<sup>a</sup>, Chris J. Müller<sup>b</sup>, Krushnamegh Kunte<sup>c</sup>, Bernard Turlin<sup>d</sup>, Axel Hausmann<sup>a,e</sup>, Michael Balke<sup>a,e</sup>

<sup>a</sup> SNSB-Bavarian State Collection of Zoology, Münchhausenstraße 21, 81247 Munich, Germany

<sup>b</sup> Australian Museum, 6 College Street, Sydney, NSW 2010, Australia

<sup>c</sup> National Center for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India

<sup>d</sup> 14, Résidence du Nouveau Parc, 78570 Andrésy, France

<sup>e</sup> GeoBioCenter, Ludwig-Maximilians University, Munich, Germany

### ARTICLE INFO

#### Article history:

Received 8 December 2014

Revised 28 April 2015

Accepted 19 May 2015

Available online 30 May 2015

#### Keywords:

Australasian–Indomalayan archipelago

bGMYC

BPP

Charaxini

Cryptic diversity

Poisson Tree Processes

### ABSTRACT

The charismatic tropical *Polyura* Nawab butterflies are distributed across twelve biodiversity hotspots in the Indomalayan/Australasian archipelago. In this study, we tested an array of species delimitation methods and compared the results to existing morphology-based taxonomy. We sequenced two mitochondrial and two nuclear gene fragments to reconstruct phylogenetic relationships within *Polyura* using both Bayesian inference and maximum likelihood. Based on this phylogenetic framework, we used the recently introduced bGMYC, BPP and PTP methods to investigate species boundaries. Based on our results, we describe two new species *Polyura paulettae* Toussaint sp. n. and *Polyura smilesii* Toussaint sp. n., propose one synonym, and five populations are raised to species status. Most of the newly recognized species are single-island endemics likely resulting from the recent highly complex geological history of the Indomalayan–Australasian archipelago. Surprisingly, we also find two newly recognized species in the Indomalayan region where additional biotic or abiotic factors have fostered speciation. Species delimitation methods were largely congruent and succeeded to cross-validate most extant morphological species. PTP and BPP seem to yield more consistent and robust estimations of species boundaries with respect to morphological characters while bGMYC delivered contrasting results depending on the different gene trees considered. Our findings demonstrate the efficiency of comparative approaches using molecular species delimitation methods on empirical data. They also pave the way for the investigation of less well-known groups to unveil patterns of species richness and catalogue Earth's concealed, therefore unappreciated diversity.

Published by Elsevier Inc.

### 1. Introduction

Habitat loss and climate disruptions threaten global species diversity (Thomas et al., 2004; Brooks et al., 2006). This is especially true in the tropics, which hold the majority of the 35 biodiversity hotspots found across the planet, representing regions of extreme yet highly threatened endemism (Mittermeier et al., 2004; Williams et al., 2011). Species are disappearing at an alarming pace while a growing body of evidence underpins the dramatic impact of biodiversity loss on ecosystem functioning (Wardle et al.,

2011; Cardinale et al., 2012; Hooper et al., 2012). In this context, cataloguing Earth's biodiversity is urgent. Accelerating of species discovery and description is achievable through embracing new technology. In the past decade, some have advocated the use of molecular data instead of or in addition to morphology and/or other lines of evidence to help discover unknown diversity throughout the tree of life (e.g. Tautz et al., 2003; Padial et al., 2010; Riedel et al., 2013). In particular, the field of molecular species delimitation is of growing importance but not without contention (e.g. Bauer et al., 2011; Fujita and Leaché, 2011; Carstens et al., 2013). A wide array of new species delimitation methods has been developed recently (Fujita et al., 2012), aiming at connecting molecular variation between organisms and taxonomy using models and thresholds of different nature and level of complexity (e.g. Hebert et al., 2003; O'Meara et al., 2006; Pons et al.,

<sup>☆</sup> This paper was edited by the Associate Editor S.L. Cameron.

\* Corresponding author.

E-mail address: [Toussaint@ku.edu](mailto:Toussaint@ku.edu) (E.F.A. Toussaint).

<sup>1</sup> Present address: Department of Ecology & Evolutionary Biology & Division of Entomology, Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA.

2006; Yang and Rannala, 2010; Ence and Carstens, 2011; Reid and Carstens, 2012; Ratnasingham and Hebert, 2013; Zhang et al., 2013). Although these new methods allow the discrimination of species-level molecular entities under a certain threshold, until recently only a few studies using such approaches led to taxonomic acts (but see Jörger and Schrödl, 2013; Satler et al., 2013). Clades for which morphology-based taxonomy is relatively well-known are of prime interest to empirically test the efficiency of these methods in order to feel the way toward a more rapid assessment of diversity in focal clades.

Delimiting species boundaries using DNA sequence data relies on the assumption that gene fragments can form monophyletic groups that represent species entities (see Puillandre et al., 2012 for the ABGD method that can recognize paraphyletic groups as MOTUs in certain circumstances). Paraphyly or polyphyly in certain markers is a known phenomenon in very recent species that are otherwise well-characterized morphologically and ecologically. Mapping morphologically delineated species onto gene trees can reveal such issues and molecular species delineation will likely fail to delineate nominal species in such cases. Delimiting species boundaries using DNA sequence data can be straightforward in groups where morphology is also rather divergent between putative species, but is generally more challenging where there is little morphological variation or presumably high levels of homoplasy. The fact that independently evolving lineages may not bear morphological differences due to a recent split is acknowledged by the generalized species concept (de Queiroz, 2007). However, it might still be difficult if not impossible to recognize the most suitable molecular species delimitation method to use on empirical data. As a result, the use of multiple methods has been recommended in order to avoid bias and to assess the consistency of delineated species across models (Astrin et al., 2012; Carstens et al., 2013; Satler et al., 2013).

Numerous studies have investigated cryptic diversity and species boundaries in Lepidoptera. DNA barcoding in particular has been widely used to describe new taxa using a combination of molecular clustering and other lines of evidence such as genitalia, host-plant preferences or caterpillar morphology (Burns et al., 2007, 2008, 2010; Hausmann et al., 2009; Chacón et al., 2012). However, studies using the most recent advances in the field of molecular species delimitation are scarce (but see Dincă et al., 2011; Le Ru et al., 2014, 2015; Dumas et al., 2015; Kergoat et al., in press). Yet, many lepidopteran groups are among the most taxonomically well-known groups of insects and therefore offer the opportunity to test the efficiency of such methods in an empirical framework. The tribe Charaxini (Lepidoptera, Nymphalidae) comprises the charismatic *Charaxes* (Emperors and Rajahs), *Euxanthe* (Forest Queens) and *Polyura* (Nawabs) butterflies. Passion for this group among collectors and researchers has led to a thorough assessment of morphology-based alpha-taxonomy (e.g. Smiles, 1982; Henning, 1989; Turlin, 2005, 2007a,b, 2009, 2011, 2013, 2014). About 170 species have been described from the Afrotropical region and about 50 to 60 other species spread as far as Southeast Asia, Wallacea and the Pacific Islands (Aduse-Poku et al., 2009; Müller et al., 2010). Molecular phylogenetic investigations of the group have revealed an affiliation of closely related clades within *Charaxes* despite a lack of morphological evidence (Aduse-Poku et al., 2009). Despite some taxonomic suggestions (Aduse-Poku et al., 2009), the systematics of *Charaxes* and its close relatives the genera *Euxanthe* and *Polyura* remain contentious. It is likely that *Charaxes* represents a complex paraphyletic series.

*Polyura* butterflies are restricted to the Indomalayan/Australasian archipelago (Fig. 1). This region encompasses 14 biodiversity hotspots (Mittermeier et al., 2004; Williams et al., 2011) and has a highly complex geological history (Hall, 2012, 2013),

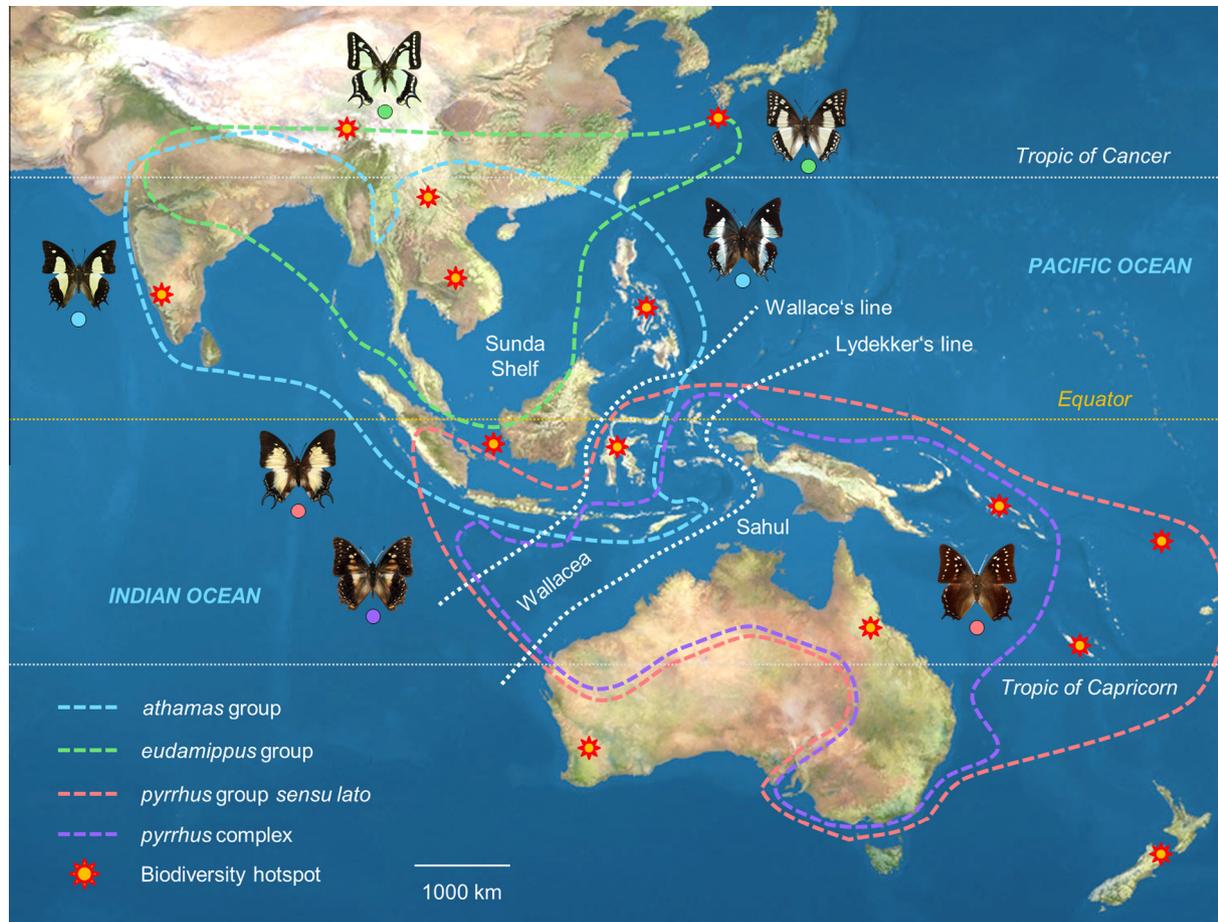
rendering it a natural laboratory to study processes of lineage diversification. *Polyura* contains 26 morphologically delineated species (*sensu* Smiles, 1982) of large, fast-flying butterflies that exhibit the typical patrolling, fighting and hill-topping behavior of the tribe. Adult *Polyura* feed on carrion and dung but also on rotten fruits and oozing sap. They are distributed from India to Fiji and from the Ryukyu Archipelago to south-eastern Australia. Numerous endemic species occur on remote islands such as Christmas Island, Fiji, New Caledonia, the Solomons and Vanuatu. Since its description (Billberg, 1820), the genus *Polyura* has been surprisingly overlooked before receiving increasing attention in the past decades with a complete revision of the group (Smiles, 1982) and some attempts to unravel phylogenetic relationships at regional scales (Wang et al., 2003, 2004; Long et al., 2006). In his comprehensive revision of *Polyura*, Robert L. Smiles noted that characters from the genitalia, larval instars or venation were of little assistance for delineating species and therefore he based his taxonomic assessment on morphological features from wing undersides that he found very informative (Smiles, 1982). Given the limitations of the traditional morphological characters used for species delimitation in butterflies, for *Polyura* species taxonomy, there is an impetus for applying novel (molecular) data to resolve the taxonomy of this group.

Here, we generated a multi-marker DNA sequence matrix comprising  $\geq 200$  specimens of all extant species of the genus *Polyura* recognized by Smiles (1982). We seek to (i) infer phylogenetic relationships between all sequenced specimens to investigate the monophyly of morphological species *sensu* Smiles (1982), (ii) delineate species boundaries using recent methods of molecular species delimitation, (iii) describe potential new species with respect to the results of species delimitation methods and the insights of geographical and morphological information derived from the literature, and (iv) compare the performance of the different species delimitation methods.

## 2. Materials and methods

### 2.1. Taxon sampling and molecular biology

We collected butterflies in India and New Guinea (permit numbers are listed in the Acknowledgments), and used museum specimens to assemble a comprehensive taxonomic sampling of the genus *Polyura*. Our dataset includes 205 specimens representing all described species except for the dubious Sulawesi endemic *P. inopinatus* which is known only from the lost holotype and may be a hybrid (Fig. 2). Specimens sequenced for this study are listed in Appendix 1. Total genomic DNA was extracted from legs and antennae tissues of dried specimens using a DNeasy kit (Qiagen, Hilden, Germany). Using PCR protocols described in Wahlberg and Wheat (2008) and Müller et al. (2010), we amplified and then sequenced the following gene fragments: *cytochrome oxidase subunit 1* (*CO1*, 471 bp), *NADH dehydrogenase subunit 5* (*ND5*, 417 bp), *ribosomal protein S5* (*Rps5*, 573 bp) and *Wingless* (396 bp). All outgroup sequences were retrieved from GenBank except *Charaxes viola* which was sequenced for the purpose of this study. We specifically sampled representatives from most *Charaxes* species groups to test the monophyly of *Polyura* (Appendix 1). The DNA sequences were edited in Geneious R6 (Biomatters, <http://www.geneious.com/>), aligned using Muscle (Edgar, 2004) and the reading frames were checked under Mesquite 2.75 (<http://mesquiteproject.org>). The different datasets used to infer phylogenetic relationships were generated under Mesquite. All sequences were deposited in GenBank (accession Nos. KT073236–KT073670 and KT073704–KT073900) and in a public dataset on BOLD (POLYU001-15–POLYU206-15).



**Fig. 1.** Distributional ranges of the different *Polyura* species groups in the Indomalayan–Australasian archipelago. Geographic map of the Indomalayan–Australasian archipelago featuring the distribution of the three *Polyura* species groups spread on twelve out of fourteen hotspots of biodiversity found in the region. Stars represent a rough approximation of the center of biodiversity hotspots. Names and delineations of main biogeographic barriers and geological regions are provided. Habitus of six species are presented on the edge of their distribution. From left to right: *P. bhārata* stat. rev., *P. dehanii*, *P. andrewsi*, *P. posidonius*, *P. luzonica* stat. rev., *P. weismanni* stat. rev., *P. epigenes*. Colored dots under the habitus of each species indicate the species group to which it belongs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

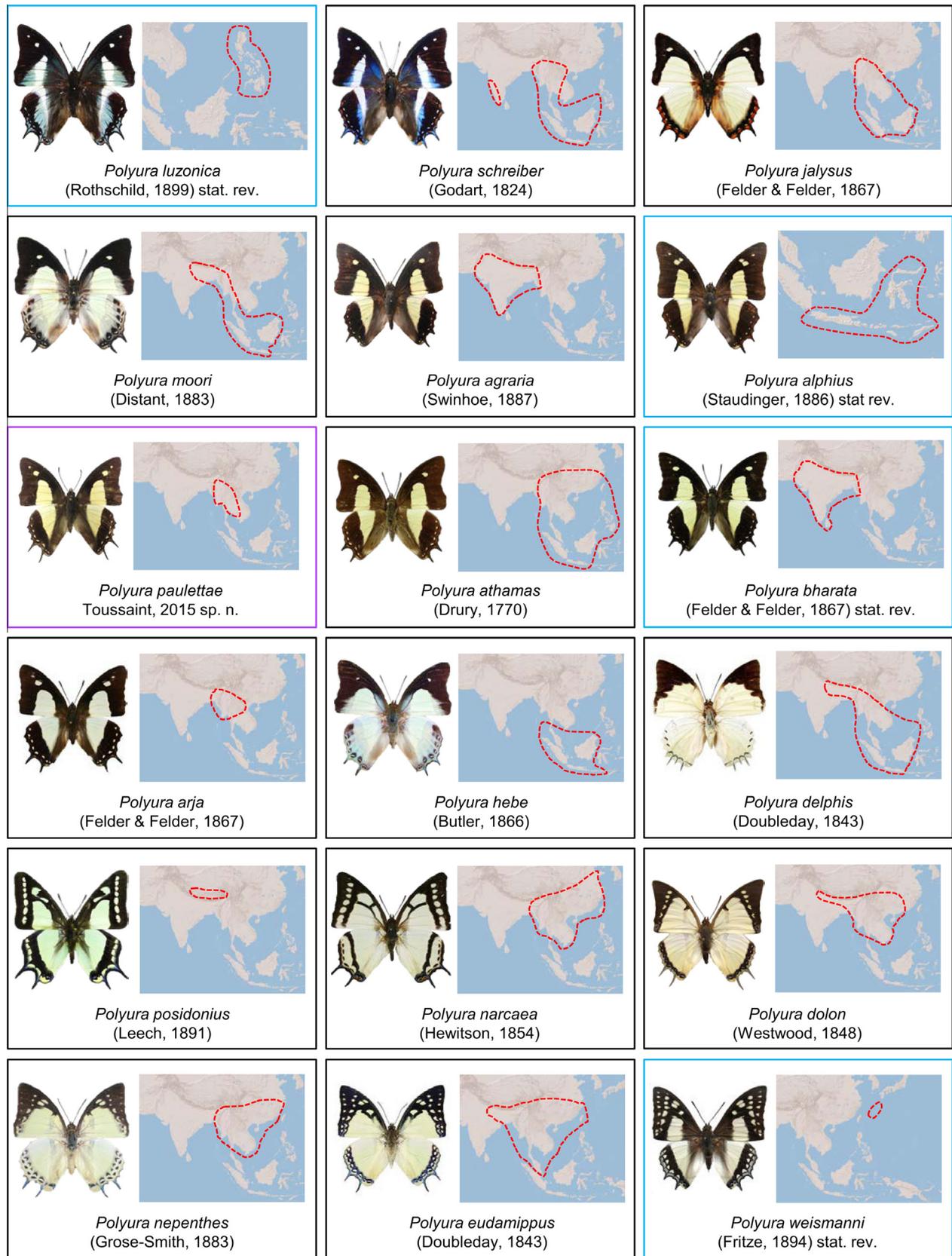
## 2.2. Molecular phylogenetics

We ran preliminary analyses in GARLI v. 0.96 (Zwickl, 2008) to reconstruct gene trees for the four markers in order to detect potential supported incongruences. Results indicated no conflict between mitochondrial gene trees and no supported incongruence between mitochondrial and nuclear gene trees. As a result we used Bayesian Inference (BI) and Maximum Likelihood (ML) to reconstruct phylogenetic relationships of all specimens sequenced using a concatenated dataset. The partitions and corresponding optimal models of substitution were searched under PartitionFinder 1.1.1 (Lanfear et al., 2012) using the greedy algorithm, and either the *mrBayes* or *raxml* set of models because MrBayes 3.2.3 (Ronquist et al., 2012) and RAXML (Stamatakis, 2006) implement different sets of substitution models. The Akaike Information Criterion corrected (AICc) was used to compare the fit of the different models. The BI analyses were performed using MrBayes 3.2.3 (Ronquist et al., 2012). Two simultaneous and independent runs consisting of eight Metropolis-coupled Markov chain Monte Carlo (MCMC, one cold and seven incrementally heated) running 80 million generations were used, with a tree sampling every 1000 generations to calculate posterior probabilities (PP). We used the partitions recovered in PartitionFinder, but instead of using the *a priori* substitution models recovered, we used reversible jump MCMC (rjMCMC) to sample the entire space of possible models

(Huelsenbeck et al., 2004). In order to investigate the convergence of the runs we investigated the split frequencies and Effective Sample Size (ESS) of all the parameters, and plotted the log-likelihood of the samples against the number of generations in Tracer 1.5 (<http://BEAST.bio.ed.ac.uk/Tracer>). A value of ESS > 200 was acknowledged as a good indicator of convergence. All the trees that predated the time needed to reach a log-likelihood plateau were discarded as burn-in, and the remaining samples were used to generate a 50% majority rule consensus tree. The ML analyses were conducted with the best partitioning scheme selected in PartitionFinder 1.1.1 (Lanfear et al., 2012) using RAXML (Stamatakis, 2006). We performed 1000 Bootstrap replicates (BS) to investigate the level of support at each node. A calculated PP  $\geq 0.95$  or a BS  $\geq 70$  was considered to indicate strong support for a given clade (Hillis and Bull, 1993; Erixon et al., 2003).

## 2.3. Molecular species delimitation

First we used the Poisson Tree Processes (PTP) model (Zhang et al., 2013) to infer molecular clades based on our inferred molecular phylogeny. The PTP method estimates the mean expected number of substitutions per site between two branching events using the branch length information of a phylogeny and then implements two independent classes of Poisson processes (intra and inter-specific branching events) before clustering the



**Fig. 2.** Distributional range maps of all extant *Polyura* species in the Indomalayan–Australasian archipelago following the taxonomic ranking of this study. Species ordering follows the phylogenetic affinities as depicted in Fig. 3. A habitus of each species is presented along with a map highlighting in red the distributional range according to the literature, field notes and examination of multiple museum specimen labels. Species delineated by a blue, violet or brown rectangle are respectively subspecies having been raised to species level, new species to science or a new species combination. *P. inopinatus* is highlighted in a red rectangle to indicate its likely extinction *in natura*. A drawing of this species is presented since the monotype was destroyed during World War II. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

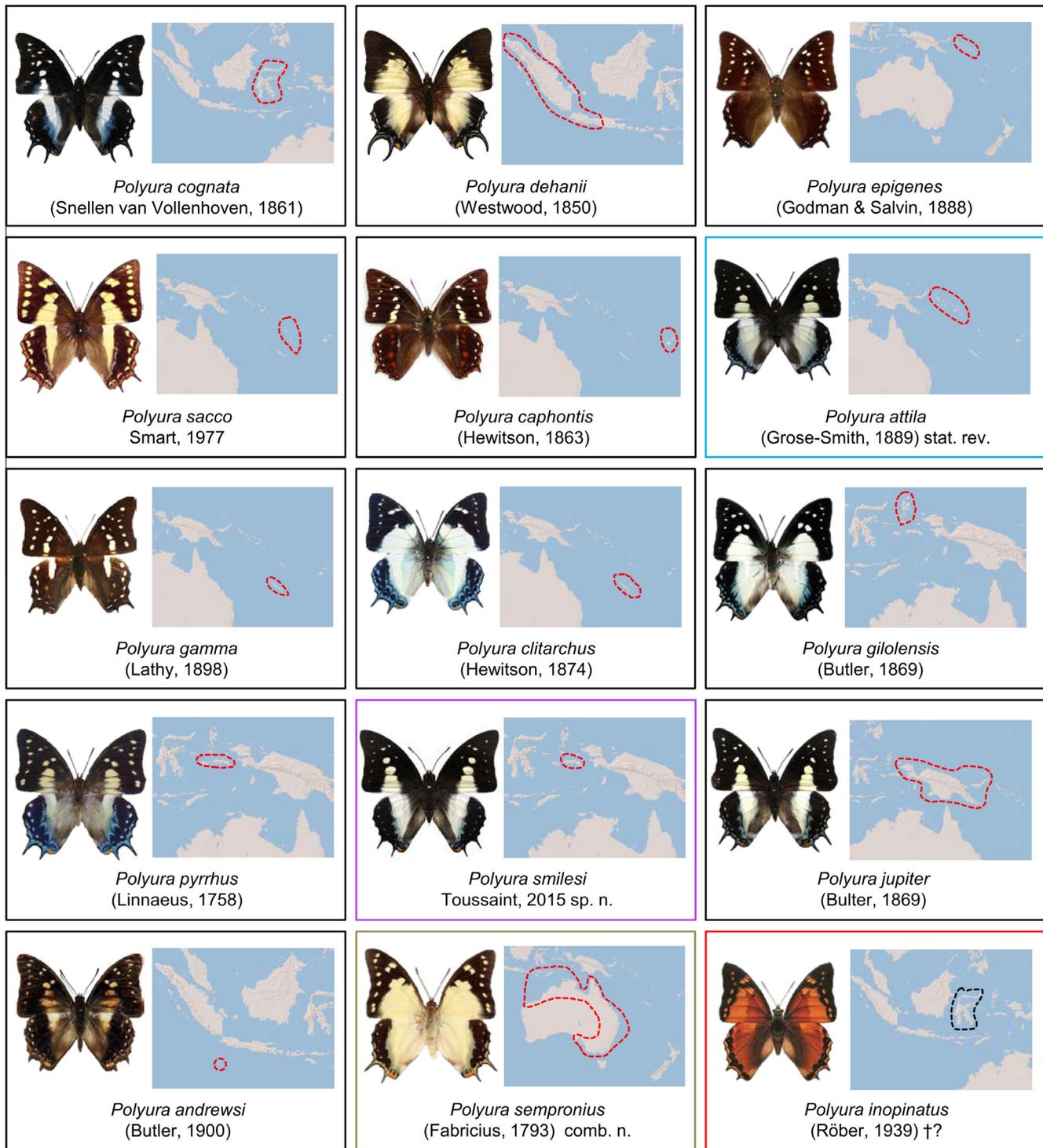


Fig. 2 (continued)

phylogenetic tree according to the results. The analyses were conducted on the web server for PTP (available at <http://species.h-its.org/ptp/>) using the RAxML topology as advocated for this method (Zhang et al., 2013; Tang et al., 2014).

Second, we used bGMYC (Reid and Carstens, 2012), a Bayesian implementation of the GMYC approach (Pons et al., 2006). The GMYC model searches in an ultrametric gene tree the threshold at which branching patterns represent coalescent events or speciation events (Pons et al., 2006). As a result, the phylogenetic uncertainty, taxon sampling and the ultrametrization of the tree have a

substantial impact on the calculation of this threshold. The bGMYC implementation allows such shortcomings to be alleviated by providing the means to use posterior distributions of trees as an input instead of a single tree (Reid and Carstens, 2012). We therefore conducted the bGMYC approach using ultrametric gene trees inferred in the BEAST 1.8.0 (Drummond et al., 2012) without outgroups under a strict clock model and a *Speciation: Yule Process Tree Model*. The runs consisted of 10 million generations sampled every 1000 cycles. Convergence was assessed by ESS values. A conservative burn-in of 10% was performed after checking the

log-likelihood curves in Tracer 1.5. As recommended by Reid and Carstens (2012), 100 trees sampled at intervals from the posterior distribution of trees using LogCombiner 1.8.0 (Drummond et al., 2012) were used to perform the bGMYC analyses. Species delimitation analyses were conducted in R using the package 'bGMYC'. The analyses consisted for each of the 100 trees selected of 250,000 generations with a burnin of 25,000 and a thinning parameter of 100.

Third, we used Bayesian species delimitation as implemented in Bayesian Phylogenetics and Phylogeography (BPP) 2.2 (Rannala and Yang, 2003; Yang and Rannala, 2010). This method accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism. A gamma prior  $G\theta s(\alpha, \beta)$ , with mean  $\alpha/\beta$ , is used on the population size parameters ( $\theta s$ ). The age of the root in the species tree ( $\tau_0$ ) is assigned the gamma prior  $G\tau_0(\alpha, \beta)$ , whereas the other divergence time parameters are assigned the Dirichlet prior (Yang and Rannala, 2010: Eq. (2)). The morphological species recognized by Smiles (1982), geographic clades as well as species recovered by the PTP and bGMYC analyses (with PP  $\geq 0.95$ ) were used as putative species, yielding a total of 38 taxa to test (Fig. 3). We used \*BEAST 1.8.0 (Heled and Drummond, 2010) to estimate the species tree using the four alignments and assigned each specimen to its corresponding putative species. Since some species were not successfully sequenced for all the genes, we generated artificial uninformative sequences comprising ambiguities for the few specimens lacking, that we included afterwards in the alignment files. For the four different partitions, we specified an uncorrelated lognormal prior for the clock, a Yule Process model as Species Tree Prior and a Piecewise constant Population Size Model. The analysis consisted of 50 million generations with a sampling interval of 5000 and a conservative burnin of 25%. As advocated by Leaché and Fujita (2010), we conducted three different sets of analyses with different values of  $\alpha$  and  $\beta$  allowing  $\theta s$  and  $\tau_0$  to account for (i) large ancestral population sizes and deep divergence between species using  $G\theta s(1, 10)$  and  $G\tau_0(1, 10)$ , (ii) small ancestral population sizes and shallow divergence between species using  $G\theta s(2, 2000)$  and  $G\tau_0(2, 2000)$ , and finally (iii) large ancestral population sizes and shallow divergence between species using  $G\theta s(1, 10)$  and  $G\tau_0(2, 2000)$ . The analyses were performed with the following settings: *speciesdelimitation* = 1, *algorithm* = 0, *finetune* *e* = 2, *usedata* = 1 and *cleandata* = 0. The reversible-jump MCMC analyses consisted of 50,000 generations (sampling interval of 2) with 25,000 samples being discarded as burn-in. Each analysis was run twice using different starting seeds to confirm consistency between runs.

#### 2.4. Nuclear DNA haplotype network

We also inferred a haplotype network based on a matrix of both nuclear genes concatenated for the *athamas* species complex. We selected one specimen per putative species (Fig. 3) with the most complete sequences for both Rps5 and Wingless yielding a matrix of 10 individuals. We used SplitsTree v. 4.13.1 (Huson and Bryant, 2006) with calculated uncorrected *p*-distances and the NeighborNet algorithm to reconstruct the haplotype network. We then run 1000 bootstrap replicates to test the robustness of the inference.

### 3. Results

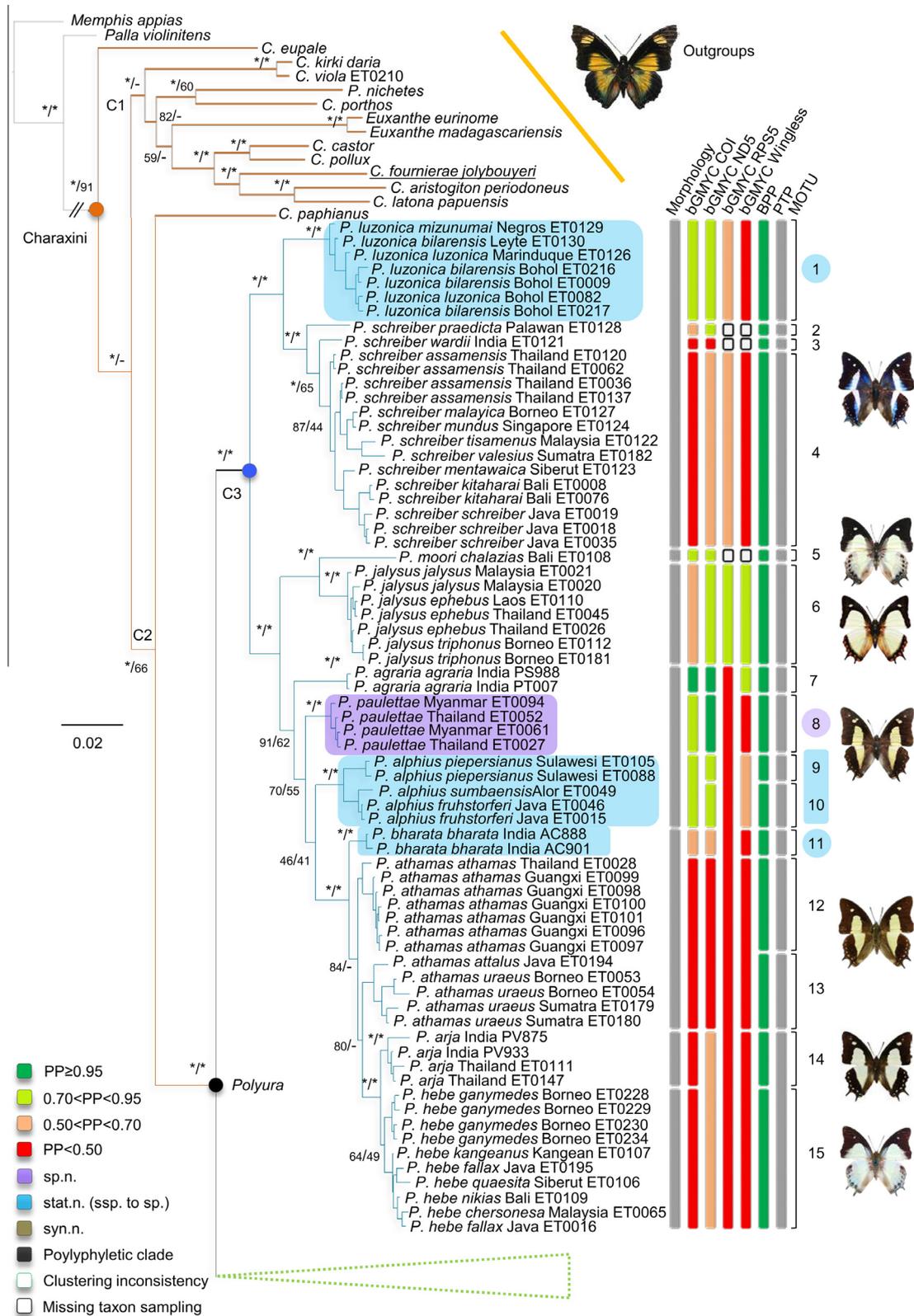
#### 3.1. Phylogenetic relationships

All information relative to the sequencing results and data quality is provided in Table 1. Results from the phylogenetic analyses conducted with the concatenated dataset are presented

in Fig. 3. The first clade (C1) contains all sampled *Charaxes* species from Africa (including the genus *Euxanthe*) and the Indomalayan/Australasian archipelago except *C. paphianus*. In a second clade (C2) the genus *Polyura* is recovered as monophyletic with strong support (1.0/100) with *C. paphianus* as sister taxon (0.95/66) rendering *Charaxes* paraphyletic. Within *Polyura*, three main clades are recovered (Fig. 3). Clade C3 (1.0/99) contained all species of the *P. athamas* group with *P. schreiber sensu Smiles (1982, MOTUs 1 to 4)* in a sister position to the remaining species. *Polyura schreiber*, *P. jalysus* (MOTU 6), *P. arja* (MOTU 14) and *P. hebe* (MOTU 15) are recovered as monophyletic with strong support except the last presenting lower support (0.64/49), whereas *P. agraria* (MOTUs 7–10) and *P. athamas* (MOTUs 11–13) were paraphyletic. In *P. schreiber sensu Smiles (1982)*, specimens from the Philippines form a well-delineated clade sister to all other specimens distributed across the rest of the distributional range. *Polyura agraria sensu Smiles (1982)* is divided into three strongly supported subclades representing different geographic areas; (i) Malaysian peninsula, (ii) Sunda (Borneo, Java, Sumatra), the Lesser Sunda Islands and Sulawesi, and (iii) India. In the second subclade, specimens from Sulawesi on one hand and from Sunda and the Lesser Sunda Islands on the other hand are also clearly separated with strong support. *Polyura athamas sensu Smiles (1982)* is equally split in three subclades roughly matching the same geographic areas as in *P. agraria sensu Smiles (1982)*.

The second clade C4 (0.54/61) contains all representatives of the *P. eudamippus*-group with *P. delphis* (MOTU 16) as sister to the rest of the species. In this clade all species are recovered as monophyletic with strong support. The main incongruence between BI and ML topologies is the placement of *P. posidonius* (MOTU 18) from Tibet which is recovered as sister to *P. narcaea* (MOTU 17) in ML and in a more derived position in BI (Fig. 2). Specimens of *P. eudamippus weismanni* (MOTU 21) from Okinawa Island form a well delineated clade sister to the rest of the *P. eudamippus* specimens (MOTU 22).

The clade C5 (1.0/99), below referred to as the *P. pyrrhus* group *sensu lato*, comprises *P. cognata* (MOTU 26), *P. dehanii* (MOTU 23), *P. epigenes* (MOTUs 24 and 25) and all representatives of the *pyrrhus* group *sensu stricto* (MOTUs 27–38). In BI two weakly supported subclades are recovered for this division whereas the ML topology only recovers a succession of increasingly derived clades. The first BI subclade contains *P. dehanii* (MOTU 23) and *P. epigenes* (MOTUs 24 and 25) whereas in ML *P. epigenes* (MOTUs 24 and 25) is found to be sister to the rest of the species of the subclade. All species recognized by Smiles (1982) are recovered as monophyletic with strong support except in the last clade C6 here referred to as the *P. pyrrhus* complex (MOTUs 31–38). Within the latter, *P. pyrrhus* (MOTU 32) and *P. gilolensis* (MOTU 34) are recovered as monophyletic. *P. jupiter* (MOTUs 31, 33 and 35) is recovered as polyphyletic in three subclades one of which comprising specimens from Solomons (1.0/100) is found sister to the rest of the *pyrrhus* complex and is clearly delineated from the rest of the other species. A second subclade from Seram is found sister to the *P. pyrrhus* complex except *P. pyrrhus* (MOTU 32) whereas the third one from New Britain, New Guinea and New Ireland is sister to *P. andrewsi* (MOTU 36), *P. sempronius* (MOTU 37) and *P. galaxia* (MOTU 38). Overall the phylogenetic reconstructions recover mostly congruent and highly supported clades at the inter- and intra-specific levels. The \*BEAST cloudogram (Fig. 4) from which was derived the input tree used to run the BPP analyses is broadly congruent with the topology presented in Fig. 3 and does not showcase a signature of gene inconsistency in the latest clades. Deeper nodes on the other hand show a more contrasted congruence signal as should be expected with the combined use of mitochondrial and nuclear data.



**Fig. 3.** *Polyura* molecular phylogenetic relationships and species boundaries. Bayesian molecular phylogeny of *CO1*, *ND5*, *Rps5* and *Wingless* gene fragments recovered under MrBayes. Posterior probabilities and bootstrap values from the RAXML analysis are presented for the most important nodes (asterisks indicate  $PP \geq 0.95$  or  $BS \geq 70$ ; - indicate that the node was not recovered in the RAXML topology). Double bar at the root indicates that gray branches have been reduced in length and are not proportional to the scale. Branches within the genera *Charaxes* and *Euxanthe* are shown in orange and branches for *Polyura* are respectively shown in blue, green and red for the *athamas*, *eudamippus* and *pyrrhus* group (*sensu lato*). Pictures of habitus are presented for *C. fourmierae* and all morphological species from *Smiles* (1982) as indicated by the delineation of the gray "Morphology" bars. Within the *pyrrhus* complex, the habitus presented from top to bottom are: *P. pyrrhus*, *P. jupiter*, *P. gilolensis*, *P. andrewsi*, *P. sempronius* and *P. galaxia sensu Smiles* (1982). Rectangles in the 6 other columns at the right present the results of the different species delimitation methods. Numbers on the right correspond to the 38 putative MOTUs delineated using bGMYC and PTP and used in the \*BEAST and BPP analyses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

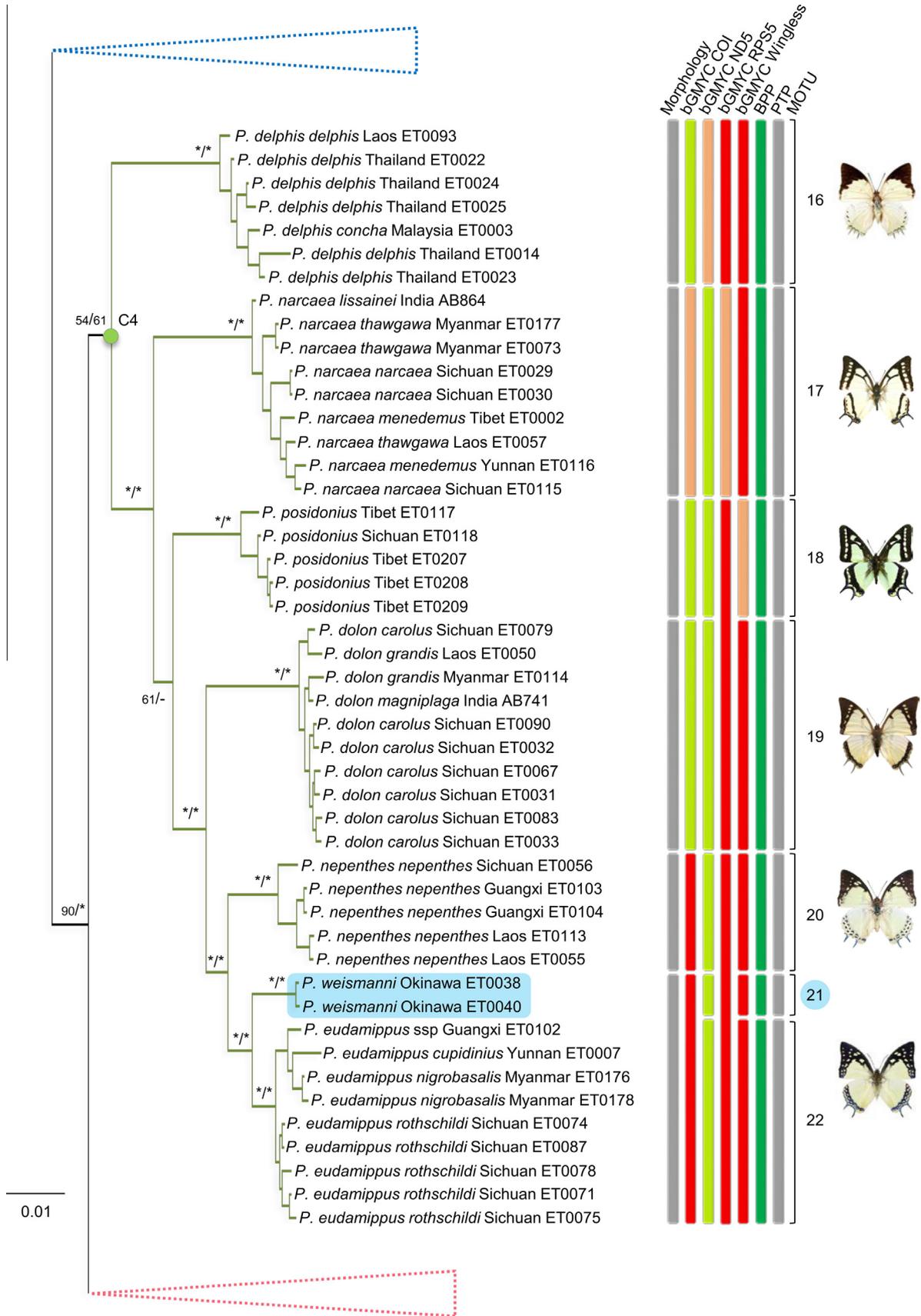


Fig. 3 (continued)

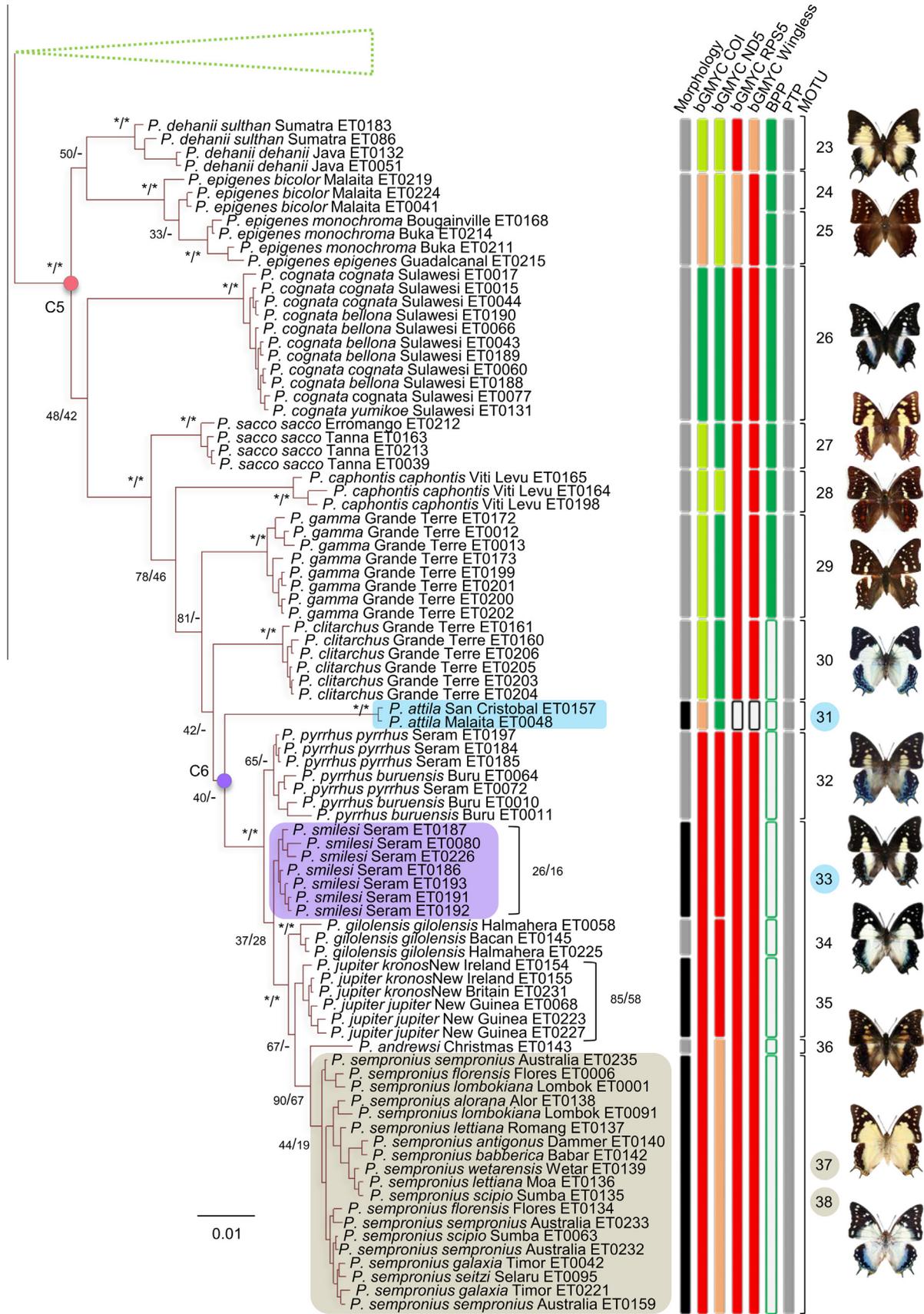


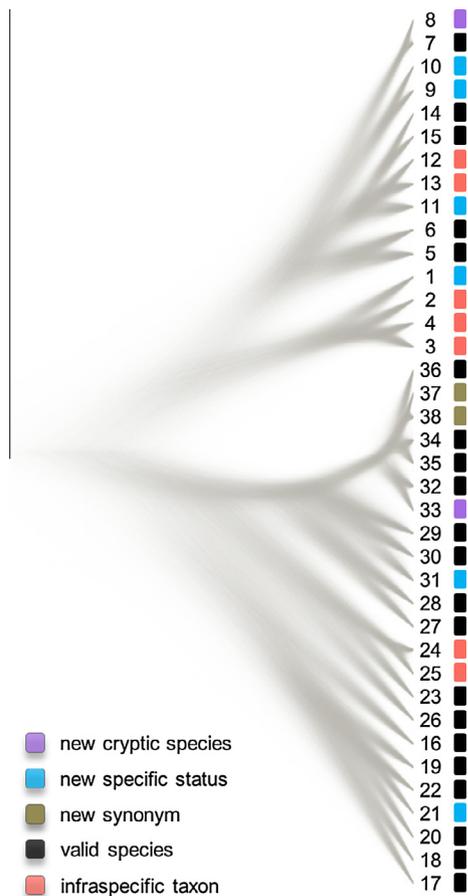
Fig. 3 (continued)

**Table 1**

Molecular composition of the dataset and molecular coverage of the different taxonomic subsets.

	Specimens	Length (bp)	Missing data (%)	GC content (%)	MOTUs	Species
CO1	196 (95.6%)	471	5.2	27.9	38 (100%)	33 (100%)
ND5	196 (95.6%)	417	2.2	20.3	38 (100%)	33 (100%)
Rps5	125 (61.0%)	573	5.3	42.3	32 (84.2%)	29 (87.9%)
Wingless	116 (56.6%)	396	1.7	24.8	33 (86.8%)	30 (90.9%)
Full dataset	205	1857	26.3	25.6	38 (100%)	33 (100%)

Notes: MOTU, Molecular Operational Taxonomic Unit here referring to the 38 different putative molecular clusters found using the different species delimitation methods as highlighted in Fig. 3; Species, here referring to the all described species of *Polyura* including the species described and populations raised to species level in this study.



**Fig. 4.** \*BEAST cloudogram of candidate species trees. Left side: All post-burn-in posterior probability trees resulting from the \*BEAST species tree analysis comprising the 38 putative species with corresponding MOTU numbers. Areas where the majority of topologies are in agreement are shown in dark and correspond to well-supported clades.

### 3.2. Species delimitation

Across the entire tree, the number of molecular species-level MOTUs varies depending on the method used, ranging from 13 to 37 versus 26 sampled species *sensu* Smiles (1982) (Figs. 3 and 5). Overall, the bGMYC analyses delineated the smallest numbers of species whereas BPP and PTP methods split the tree into larger numbers of putative species (Fig. 5).

The use of bGMYC with CO1 or ND5 yielded very similar results although clade support was generally higher with ND5. Among the only highly supported clades recovered by the method, three are

intra-specific clades which are discussed below. The bGMYC analyses conducted on the Rps5 and Wingless gene trees gave contrasting results. The analysis based on the Rps5 gene tree yielded the lowest number of MOTUs with clear inconsistencies among other methods (Fig. 5). The analysis based on the Wingless gene tree delivered reasonably close results from the other methods. Both bGMYC analyses based on nuclear gene trees resulted in poorly supported MOTUs.

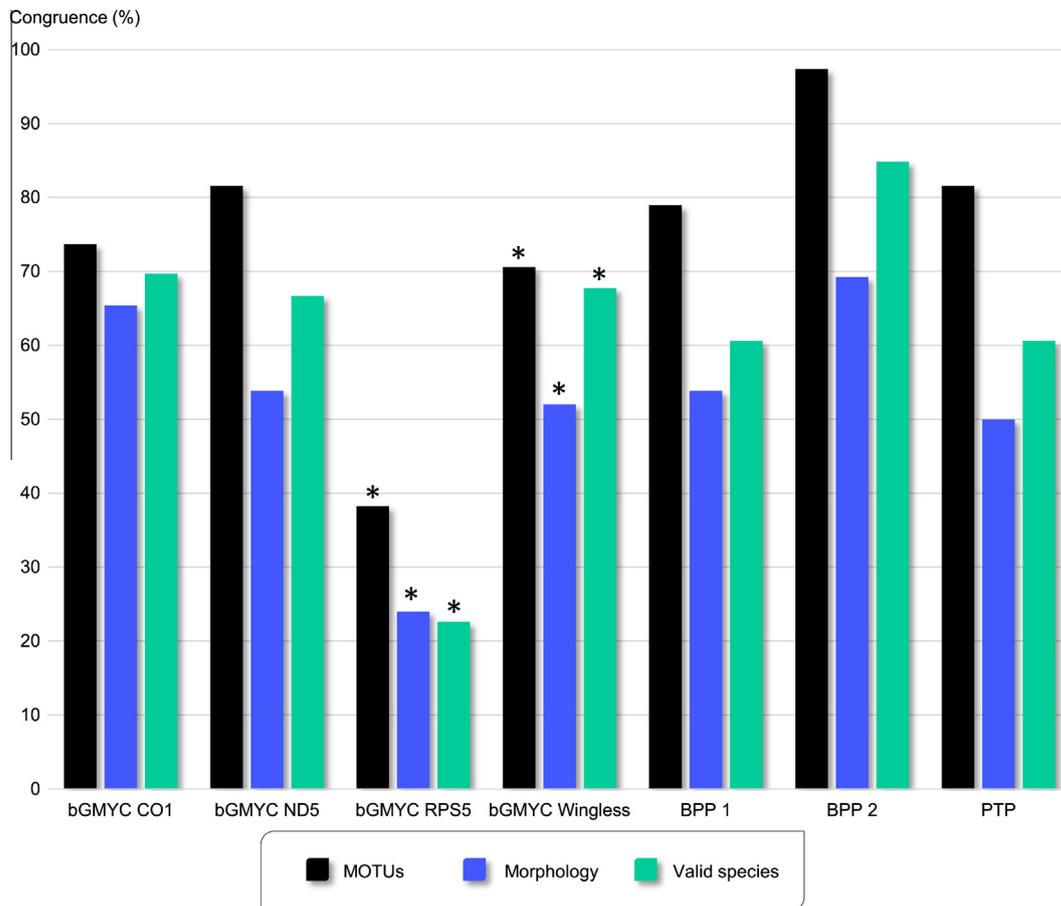
The BPP method based on a \*BEAST species tree recovered the largest number of putative species with high support for each of them. Although BPP was the method that delineated the maximal number of species in our case it actually cross-validated most species from Smiles (1982). The use of extremely loose priors for the input parameters of the models proved to give highly stable results except in the case of the *P. pyrrhus* complex. Both analyses with a large ancestral population size resulted in the same results whereas the one based on a small ancestral population size delivered slightly different results. The only discrepancy between these two sets of analyses is localized in the *P. pyrrhus* complex. The analyses with a large ancestral population size recovered all the taxa tested as valid species except *P. galaxia* (MOTU 38) and *P. sempronius* (MOTU 37) that were lumped, whereas the analyses based on a small ancestral population size recovered the entire *P. pyrrhus* complex as one species also including *P. clitarchus* (MOTU 30). Globally, this method was highly congruent with the other ones and especially with PTP. The latter performed very well at delineating taxa from Smiles (1982) although it did not detect *P. arja*.

Two paraphyletic species were found for which all different clades were not clearly delineated by the molecular species delimitation methods. First, we find in both BI and ML that *P. athamas* (MOTUs 12 and 13) is paraphyletic due to the inclusion of *P. arja* (MOTU 14) and *P. hebe* (MOTU 15) in a derived position. Among the three clades recovered and tested, only the specimens from India are clearly delineated whereas the two remaining clades are not supported by the bGMYC analyses. Second, *P. jupiter* (MOTUs 31, 33 and 35) is recovered as polyphyletic within the *pyrrhus* complex with three distinct and well-delineated entities. Populations from the Solomon Islands form a distinct clade in the *P. pyrrhus* complex and populations from Seram also form a distinct clade.

Overall the species delimitation methods gave comparable results and were mostly congruent (Fig. 5). The main discrepancies were found with bGMYC where the analyses based on the Rps5 gene tree delivered radically different results compared to all the other methods including the bGMYC analysis of the Wingless gene tree. However the congruence of bGMYC with other methods when excluding the analysis based on Rps5 was good. The congruence score between PTP and BPP (when assuming a small initial population size) is particularly high (94%), an unexpected similarity between a discovery and a validation method resting on extremely different priors and models. Finally, the congruence scores of the different models (Fig. 5) with the morphological species retained by Smiles (1982) vary from 24% (bGMYC Rps5) up to 69% (BPP with large ancestral population size) demonstrating the need for a substantial part of the clades to be revised taxonomically. Such revision should be conducted within the generalized species concept (de Queiroz, 2007) by combining the different results of the species delimitation methods with additional lines of evidence such as morphological features and geographic distributions.

### 3.3. Nuclear haplotype network

The haplotype network inferred in SplitsTree and based on a subset of the *athamas* species complex yielded well-resolved and robust relationships between the different putative species



**Fig. 5.** Performance of the different molecular species delimitation methods. Histogram presenting the results of the different species delimitation methods used on the *Polyura* datasets. The following values are presented: proportion of MOTUs recovered compared to the total amount of MOTUs delineated (38 MOTUs), proportion of morphospecies (26 species) from Smiles (1982) recovered, and proportion of valid species recognized in this study (33 species) recovered. BPP 1 summarizes the results of the models with large initial population size and BPP 2 with small initial population size. The asterisks indicate that for these two methods (bGMYP with either the Rps5 or the Wingless gene tree) some MOTUs were lacking and therefore the calculations were made accordingly.

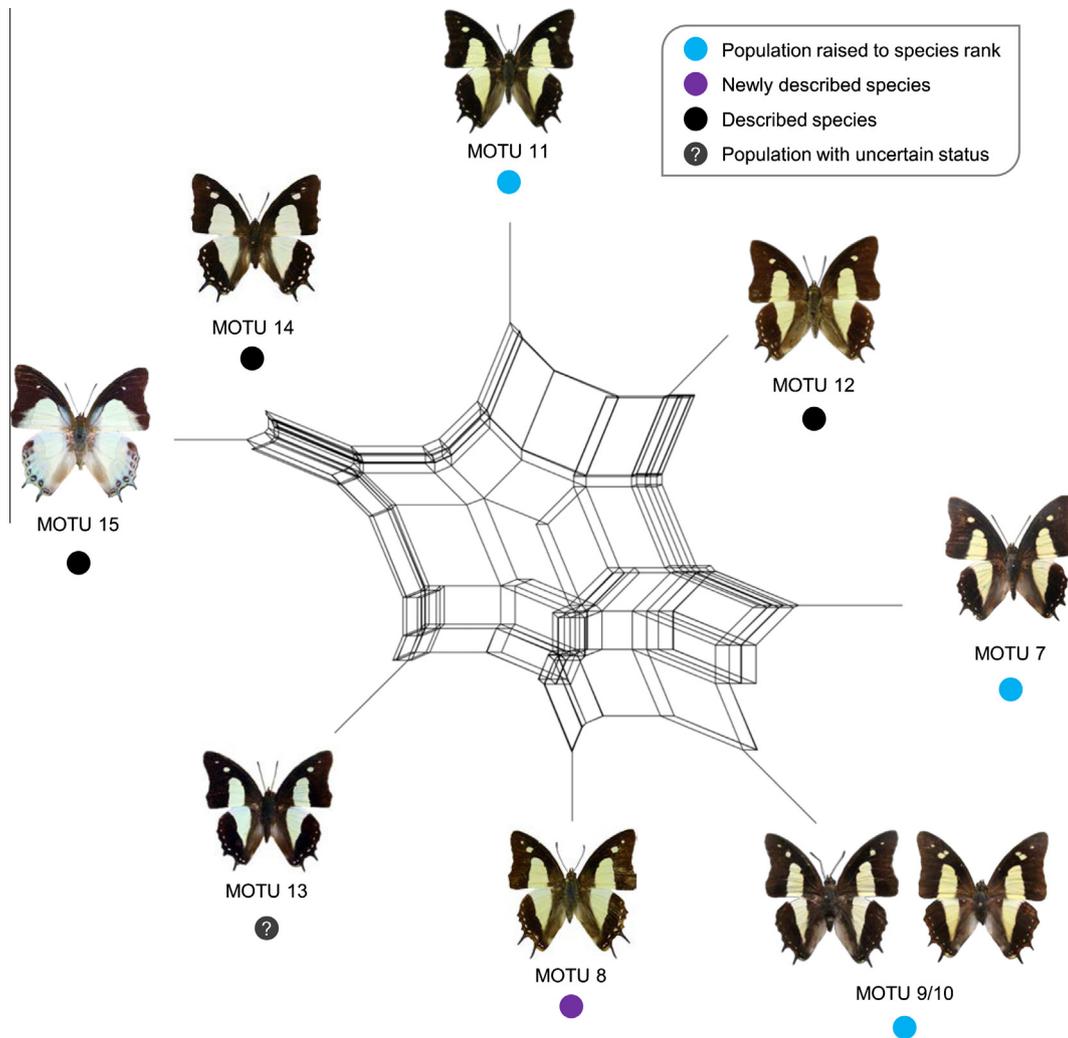
(Fig. 6). Overall we find a clear structuration between all MOTUs except for the Easternmost populations of *Polyura agraria* sensu Smiles (1982). Both representatives of the Lesser Sunda Islands (MOTU 10) and Sulawesi (MOTU 9) are found together highlighting the lack of nuclear differentiation between these two MOTUs.

## 4. Discussion

### 4.1. Consistency assessment across molecular species delimitation methods

Overall the different molecular species delimitation methods yielded consistent results except for the bGMYP analyses based on the Rps5 gene tree that yielded a substantially lower number of putative species (Figs. 3 and 5). If the analysis based on the Wingless gene tree is mostly in agreement with other methods, the results of the two analyses based on nuclear gene trees are rather poorly supported and in the case of the Rps5 gene tree seem extremely dubious especially in clade C4. However, we believe that the results of these two analyses should be taken with caution as they were based on dramatically reduced taxonomic subsets of the complete dataset due to the impossibility to sequence nuclear genes for most Museum specimens. As a result, both gene trees were not only lacking a large fraction of the geographic sampling for each MOTU but were also completely lacking some MOTUs therefore rendering error prone the branch length estimation and

clustering of specimens. These biases in addition to a lower level of genetic variation in these nuclear genes highlight the need for properly geographically and taxonomically sampled gene trees to obtain reliable bGMYP estimates in an empirical framework. Hence, we chose not to further discuss the results of these two analyses shown in Figs. 3 and 5. Nevertheless, it seems that Wingless might be a good candidate gene for species delimitation studies unlike Rps5 that might be too conserved to properly infer the breaking point between speciation and Coalescent events (Fig. 5). Overall, the GMYP approach proved to be much more conservative in a Bayesian framework than its original implementation (Pons et al., 2006) that is widely used and can engender oversplit results (see Talavera et al., 2013 for a discussion). Here, we find it difficult to meet the threshold of robustness ( $\geq 0.95$ ) because of the use of randomly selected posterior probability trees with different branch lengths. However, when looking at the maximum credibility clades, bGMYP analyses (except the ones based on the Rps5 gene tree) delivered results comparable to the ones obtained with other methods and recovered most morphologically delineated species from Smiles (1982). Although GMYP and bGMYP approaches are from far the most commonly used methods to infer species boundaries, our results show discrepancies depending on the gene fragment used. Overall our results indicate that molecular species delimitation methods based on the multilocus dataset outperform the bGMYP approaches to consistently delineate well-supported MOTUs (Figs. 3 and 5). Yet, even though the GMYP approach with the single or multiple thresholds has



**Fig. 6.** Bootstrap nuclear haplotype networks of the *athamas* species complex. Networks based on a concatenated alignment of the two nuclear genes with a unique representative of each MOTU. Each haplotype network from the 1000 bootstrap replicates is shown to highlight the robustness of the inferred relationships.

been shown to overestimate putative species numbers (see Miralles and Vences, 2013; Hamilton et al., 2014; Lecocq et al., 2015 for recent empirical examples), we find here that its implementation in a Bayesian framework is much more reliable as it yields somewhat comparable results with PTP and BPP (Fig. 5). However, one should keep in mind the pitfalls associated with the use of single locus datasets to infer species boundaries (Knowles and Carstens, 2007; O'Meara, 2010). Satler et al. (2013) used the bGMYC approach on their trapdoor spider CO1 dataset and concluded that it severely overestimated the species richness in this clade indicating that this method might be unsuitable in some cases. We also find that bGMYC analyses based on nuclear markers are less efficient to recover comparable species richness estimates, likely due to lower levels of genetic information in comparison with mitochondrial DNA. This caveat is alleviated by the use of multimarker datasets allowing to capture enough phylogenetic information to properly estimate the transition between speciation and coalescent events. Here, both PTP and BPP deliver comparable species richness estimates and recover most of the described species as well as valid new ones (Figs. 3 and 5). These results support the growing body of literature suggesting the efficiency to delineate valid species with BPP (Yang and Rannala, 2010; Zhang et al., 2011; Camargo et al., 2012; Rannala and Yang, 2013) and PTP (Zhang et al., 2013; Tang et al., 2014).

#### 4.2. Molecular species delimitation implications

Among the *P. athamas* group, specimens of *P. schreiber* from the Philippines (MOTU 1) are clearly delineated. Interestingly the subspecies *P. schreiber praedicta* from Palawan is not recovered as being part of this clade. This pattern is supported by the geological affiliation of Palawan with the Sunda shelf whereas the remainder of the Philippines has an independent and highly complex geological origin (Hall, 2012, 2013). The remainder of the *P. schreiber* populations form a very widespread clade distributed from India to Borneo, a case already documented for several butterfly species of the region (Wilson et al., 2013). Morphologically, Philippines specimens are extremely close to the ones from other regions except for a more slender wing shape. As a result we decided to recognize all Philippine populations of *P. schreiber* as an endemic species for which the name *P. luzonica* Rothschild stat. rev. is applicable. In a more derived position of the *P. athamas* group, Thai and Myanmar populations of *P. agraria* (MOTU 8) are also clearly delineated (Fig. 3). These populations form a distinct genetic and geographic entity although being morphologically similar to the *P. agraria* specimens. The same applies to the populations of *P. agraria* from the Wallacea (MOTUs 9 and 10). As a result, we raise these two distinct populations to species level. The name *P. alphius* stat. rev. is available for the Wallacean species.

Populations from Thailand and Myanmar form a distinct cryptic lineage and are described as a new species under the name *P. paulettae* nov. sp. while populations from India are kept under the name *P. agraria*. These three populations are monophyletic with respect to the examined genes and likely represent recent independently evolving lineages presenting no clear morphological variation (Smiles, 1982). This kind of allopatric speciation might be difficult to highlight, but the use of multiple marker approaches to delineate such independently evolving lineages has been shown to be very efficient in the past (Huemer and Mutanen, 2012). Finally, Indian specimens of *P. athamas* (MOTU 11) are found in a well-delineated clade recovered in all analyses. We raise Indian populations to species-level with the available name *P. bharata* stat. rev. We have gathered a relatively comprehensive geographical sampling for *P. athamas sensu Smiles (1982)* except for the easternmost populations in the Philippines and as a result we suggest that the clear demarcation of Indian specimens from the rest of the representatives is not an artefact driven by sampling bias (Irwin, 2002). Morphological characters of the wings in the *athamas* group were not retained as diagnostic characters by Smiles (1982). These are highly variable and therefore render the delimitation of species tedious and error-prone. Interestingly, the conflict between morphological and molecular characters to delineate species was already recognized in the study of Aduse-Poku et al. (2009) investigating African *Charaxes* species relationships. In the context of recent diversification events, it might be difficult to distinguish between phenotypic variability and specific morphological divergences and therefore molecular characters might be more reliable to identify species-level taxa and diagnose them.

In the *P. eudamippus* group, the subspecies *P. eudamippus weismanni* (MOTU 21) is found as a separate lineage. This taxon has the northernmost distributional range among *Polyura*, in the Ryukyu archipelago. This assemblage of islands is of recent tectonic origin with late connections to the continent before sea-level raised in the Pleistocene (Kimura, 2000). In their study Long et al. (2006) suggested that Taiwanese *P. eudamippus* could be derived from continental populations via glacial land bridge colonization in the past thousands of years. However, the genetic demarcation we observe between *P. eudamippus weismanni* and the remainder of *P. eudamippus* is unlikely to be the result of such a short period of time and we hypothesize that the former is likely the result of a more ancient dispersal event out of China. This is supported by a greater morphological deviation of this subspecies compared to the rest of *P. eudamippus* populations including the easternmost ones in Taiwan (Smiles, 1982). *P. eudamippus* from the Ryukyu archipelago is recognized as a valid species with the available name *P. weismanni* stat. rev.

Among the *P. pyrrhus* group *sensu lato*, most species are recovered by the species delimitation methods except for *P. epigenes* and within the *pyrrhus* complex (Fig. 3). The investigation of species boundaries in *P. epigenes* revealed the potential existence of two separate lineages on different islands of the Solomons. The subspecies *P. epigenes bicolor* recently described from the island of Malaita (Turlin and Sato, 1995), is found to be paraphyletic with very low support in BI and as monophyletic with moderate support in ML. bGMYC consistently failed to recover *P. epigenes bicolor* as a valid species when BPP and PTP provided strong evidence for it. In the light of our results a taxonomic reassessment of this lineage might be needed but additional taxon sampling and better gene coverage are needed. Within the *P. pyrrhus* complex, *P. jupiter sensu Smiles (1982)* is recovered as polyphyletic with three clades delineated by the phylogenetic analyses but with different levels of support by species molecular delimitation methods. Specimens from the Solomons (MOTU 31) are found well delineated. This taxon represents an endemic with rather deviant morphology presenting a larger habitus than species from the *pyrrhus* complex

recognized by Smiles (1982) and diagnostic morphological features on the upper- and underside of the wings. As a result, we raise this taxon to species rank under the available name *P. attila* stat. rev. The placement of *P. attila* stat. rev. reveals a more complex evolutionary history for the group with possibly an early colonization of Pacific islands as illustrated by the branching of *P. sacco*, *P. caphontis* and *P. gamma* in the *P. pyrrhus* group *sensu lato*. The colonization of New Guinea and the Moluccas out of Pacific clades in a westward configuration would be in line with recent studies on the paleogeography of the region (Hall, 2012, 2013) and the origin and timing of Melanesian clade diversification (Toussaint et al., 2014). The remainder of the *P. pyrrhus* complex showcases a much more puzzling pattern because most species delimitation methods disagree with morphology but also between each other. Based on strong morphological evidence (Smiles, 1982; Turlin and Sato, 1995), we argue that it would not be parsimonious to lump all extant species of this complex into one valid species. Moreover, our phylogenetic reconstructions clearly disclose a fine geographic structuration of populations in this group despite presenting moderate nodal support. Considering the validity of each extant species, two problems remain; (i) specimens of *P. jupiter sensu Smiles (1982)* from Seram (MOTU 33) are found in a distant clade, and (ii) specimens from *P. galaxia* (MOTU 38) and *P. sempronius* (MOTU 37) *sensu Smiles (1982)* are found in a same clade without any structuration. Originally populations of *P. jupiter sensu Smiles (1982)* from Seram were described as an aberration of *P. jupiter* with which they share a common morphology but that allows an easy separation from the sympatric *P. pyrrhus*. In order to keep *P. jupiter* monophyletic and reach a balanced decision, we describe the populations from Seram as a new species under the name *P. smilesi* sp. nov.. Finally, *P. galaxia* (MOTU 38) and *P. sempronius* (MOTU 37) *sensu Smiles (1982)* are consistently found as one species across the different methods used and the branching of the multiple specimens clearly indicates that these two taxa are conspecific. The geographic distribution of both taxa also supports the view of a single widespread species ranging from Lombok to Lord Howe Island about 600 km East of the Australian mainland, and encompassing most Lesser Sunda Islands and the entire coastal region of Australia from West to South-East (Figs. 1 and 2). We therefore synonymize *P. galaxia* with *P. sempronius* in order to reflect our results. *Polyura andrewsi* from Christmas Island could possibly belong to *P. sempronius* as the westernmost representative of this widespread species but additional data is needed. Overall the *P. pyrrhus* complex is a geographically highly structured species complex with a distributional range encompassing Australia, the Moluccas, the New Guinean archipelago, the Solomons and Christmas Island. The shallow genetic divergence between members of this group would also be here in line with the late geological assemblage of parts of Melanesia and Wallacea and the great dispersal ability of these insects.

#### 4.3. Description of new species and taxonomic reassessment

Family: Nymphalidae Rafinesque, 1815

Subfamily: Charaxinae Guenée, 1865

Genus: *Polyura* Billberg, 1820

*Polyura paulettae* Toussaint **sp. n.**

LSID: urn:lsid:zoobank.org:act:6C9FA666-6D52-4B3C-BAC3-63DCF5B269E6

Species page: [http://species-id.net/wiki/Polyura\\_paulettae](http://species-id.net/wiki/Polyura_paulettae)

Corresponding molecular operational taxonomic unit (MOTU): MOTU 8 (Fig. 3).

Types: Holotype (Appendix 2, ZSM – Zoologische Staatssammlung München, Germany): Female from Wang Chin District, Phrae Province, Thailand, IV 1982, a dry-pinned specimen with voucher label ET27 and red HOLOTYPE label. Paratypes (ZSM): three

dry-pinned specimens with voucher label ET52 (male from the same locality as the holotype), ET61 (male from Shan States, Myanmar), ET94 (male from North Sagaing, Myanmar) and red PARATYPE label.

Diagnosis: A cryptic species of *P. agraria sensu Smiles (1982)* found in the *P. athamas* group. Very similar to *P. agraria* and *P. alphius* with which it shares a more elongated wing shape than the rest of the *athamas* group. Allopatric from the two other species of the *agraria* group (Fig. 2). Molecular diagnostic characters compared to other species of the *agraria* species complex are shown in Appendix 3.

Description: Males and females share a same morphological appearance. Abdomen dark brown on the upperside and beige on the underside. Wing upperside dark chocolate brown becoming lighter toward the bases of both wings. Clearly demarcated greenish discal band found on both wings commencing on vein M3 in the forewing and ending approximately at vein 2A in the hindwing. Two subapical spots of the same greenish color in cell M1 and R5 of the forewing. Hindwing with orange admarginal spots above which is found a series of 7–8 white submarginal spots. Center of the tails light blue. Wing underside pinkish brown becoming darker toward the outer margins. Series of greyish chevrons with an external black line from the forewing cell R5 to Cu1b where a large blackish spot is found. Greenish discal band similar to the one found on the upperside but slightly paler and surrounded by a narrow orange-brown band. Spot of the same color as the discal band is found in cell M1 of the forewing. Hindwing tails are blue-centered. Series of parallel submarginal white and black spots on the hindwing.

Etymology: Named after the first author's grandmother Paule "Paulette" Toussaint, passionate butterfly admirer and collector.

Distribution: Currently known from Myanmar and Northern Thailand.

*Polyura smilei* Toussaint **sp. n.**

*Eriboea jupiter* aberration *rectifascia* Talbot, 1920 (unavailable name, introduced at infrasubspecific level).

LSID: urn:lsid:zoobank.org:act:276D1F53-C639-4FC3-B790-21A7CE686BA1.

Species page: [http://species-id.net/wiki/Polyura\\_smilesi](http://species-id.net/wiki/Polyura_smilesi).

Corresponding molecular operational taxonomic unit: MOTU 33 (Fig. 3).

Types: Holotype (Appendix 2, MZB – Museum Zoologicum Bogoriense, at the Indonesian Institute of Sciences LIPI, Division of Zoology, Cibinong, Indonesia): Female from Seram, Indonesia, X 1969, dry-pinned specimen with voucher ET187 and red HOLOTYPE label. Paratypes (ZSM): three dry-pinned specimens from the same locality as the holotype, with voucher ET191 (male), ET192 (male), ET193 (female) and red PARATYPE label.

Diagnosis: Morphologically very similar to *P. jupiter* with less gray–blue scaling on the hindwing. Black lines on the anal veins of the hindwing underside narrow whereas exaggeratedly thick in *P. pyrrhus*. Molecular diagnostic characters compared to the sister clade in the gene alignments (codon position): CO1: 145, A (1st); ND5: 345, T (3rd); 360, C (3rd).

Description: Males and females share a same morphological appearance. Abdomen dark brown on the upperside and beige on the underside. Wing upperside dark brown becoming brown toward the bases of both wings. Cream submarginal spots on the forewing with postdiscal spots of the same color in cells M1 and R5. Cream discal band in the forewing from cell Cu1a until the inner margin and in the hindwing from the coastal margin to cell Cu1b. In the hindwing the discal band is bordered with a very light gray–blue scaling. Two large cream discal spots in cells M2 and M3 above this discal band. Hindwing with a series of 7–8 blueish submarginal spots. One large orange spot at the tornus. Center of the tails light blue. Wing underside light orange–brown becoming

darker toward the outer margins. White submarginal narrow band with an additional parallel white narrow band proximally. Discal band similar to the one on the upperside but almost white and surrounded by a narrow black band proximally. Hindwing tails blue-centered and admarginals light orange. Black submarginal ocelli bordered with blueish white. Three blood-red postdiscal lunules with a proximal pale blue border and two black margins in cells M3, Cu1a and Cu1b. Lunules in cells R5, M1 and M2 orange with a light row of scales distally in place of the black margin. Veins 2a and 3a thinly overlaid with black.

Etymology: Named after Robert Leslie Smiles who conducted a stunning revision of *Polyura* Nawab butterflies in 1982 and without whom this study would have never been possible.

Distribution: Endemic to Seram.

We additionally propose the following taxonomic changes following our results with updated distributional ranges and habitus pictures presented in Fig. 2:

*Polyura alphius* Staudinger, 1886, **stat. rev.: species propria**

*Polyura attila* Grose-Smith, 1889, **stat. rev.: species propria**

*Polyura bharata* Felder & Felder, 1867, **stat. rev.: species propria**

*Polyura luzonica* Rothschild, 1899 (originally described in *Eulepis*), **stat. rev.: species propria**

*Polyura weismanni* Fritze, 1894, **stat. rev.: species propria**

*Polyura sempronius* Fabricius, 1793 (originally described in *Papilio*) = *Polyura galaxia* Butler, 1866, **syn. n.**

#### 4.4. Best practice recommendations

In this study we highlight the efficiency of using multiple methods of molecular species delimitations on a multimarker dataset in order to inform taxonomic decisions. It seems crucial that discovery and validation methods be used in concert to provide a robust comparative framework. The field of molecular species delimitation is burgeoning with new methods that can be used and compared to this regard (Yang and Rannala, 2010; Ence and Carstens, 2011; Puillandre et al., 2012; Reid and Carstens, 2012; Satler et al., 2013; Zhang et al., 2013; Grummer et al., 2014). With the reduction of costs to generate DNA sequence data, molecular species delimitations should be at the core of taxonomy. We believe that modern taxonomy should encapsulate a combination of several lines of evidence (e.g. morphology, ecology, behavior, geographic distributions, divergence times) and molecular comparative approaches to have an objective way of delineating species rather than using subjective and/or a unique criterion. We strongly advocate as best practice the formal description of species delineated using such an integrative approach. Species in general should be considered hypotheses to revisit when additional sampling and/or data is available. Although morphological diagnostic characters are desirable when it comes to describe biodiversity, the lack of such information should not hamper descriptions of taxa that are suggested to be valid species based on other data sources. We believe that by applying these methods in a scientific, objective and repeatable manner, it should be possible to improve the pace and soundness of biodiversity inventory.

## 5. Conclusion

Molecular species delimitation methods offer a tantalizing opportunity to accelerate the discovery of biodiversity on our planet. This is especially true for cryptic species complexes that host a substantial fraction of this unknown species richness that cannot be discovered with traditional morphology-based taxonomic approaches. Here, using molecular species delimitation techniques in addition to geographic and morphological data, we unveil new species in a group of tropical emblematic butterflies occurring in

some of the most threatened regions of Earth. We argue that the proper use of molecular species delimitation methods might have at least two cardinal implications; (i) with an accelerated rate of species extinction and the issue of traditional taxonomic description, these methods represent an increasingly efficient and objective tool that taxonomists should embrace in order to enhance the linkage with formal descriptions (Riedel et al., 2013; Pante et al., 2015), and (ii) while anthropogenic destruction of habitats is greatly impacting the sustainability of known and unknown biodiversity, especially in the tropics, showcasing an unsuspected richness of flagship organisms can help capture the attention of conservation planners in order to preserve this ecological legacy.

## Acknowledgments

The authors would like to thank two anonymous reviewers as well as the associate editor Stephen Cameron for fruitful comments that contributed to improve an earlier version of this article. This work was supported by German Science Foundation (DFG) Grants BA2152/6-1, 7-1, 11-1 and 20-1. The Indian work was supported by a Ramanujan Fellowship to KK. The Indian states of Kerala, Nagaland and West Bengal are acknowledged for giving permission to conduct scientific research (Permit numbers: WL10-3781/2012, CWL/GEN/240/522, 3748(3)/WL/4R-1/13). The Department of Environment and Conservation, Papua New Guinea is acknowledged for giving permission to conduct scientific research (Permit numbers: 013066, 013183, 013339).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2015.05.015>.

## References

- Aduse-Poku, K., Vingerhoedt, E., Wahlberg, N., 2009. Out-of-Africa again: a phylogenetic hypothesis of the genus *Charaxes* (Lepidoptera: Nymphalidae) based on five gene regions. *Mol. Phylogenet. Evol.* 53 (2), 463–478.
- Astrin, J.J., Stüben, P.E., Misof, B., et al., 2012. Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation. *Mol. Phylogenet. Evol.* 63 (1), 1–14.
- Bauer, A.M., Parham, J.F., Brown, R.M., et al., 2011. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proc. Royal Soc. B: Biol. Sci.* 278 (1705), 490–492.
- Billberg GJ (1820) *Enumeratio insectorum*. Holmiae, 138pp.
- Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., et al., 2006. Global biodiversity conservation priorities. *Science* 313 (5783), 58–61.
- Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W., Hebert, P.D., 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *J. Lepidopterists Soc.* 61 (3), 138–153.
- Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W., Hebert, P.D., 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proc. Natl. Acad. Sci.* 105 (17), 6350–6355.
- Burns, J.M., Janzen, D.H., Hallwachs, W., Hajibabaei, M., Hebert, P.D., 2010. Genitalia, DNA barcodes, larval facies, and foodplants place the mimetic species *Neoxeniades molion* in *Rhinton* (Hesperiidae: Hesperinae). *J. Lepidopterists Soc.* 64 (2), 69–78.
- Camargo, A., Morando, M., Avila, L.J., Sites, J.W., 2012. Species delimitation with ABC and other Coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae). *Evolution* 66 (9), 2834–2849.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., et al., 2012. Biodiversity loss and its impact on humanity. *Nature* 486 (7401), 59–67.
- Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. *Mol. Ecol.* 22 (17), 4369–4383.
- Chacón, I.A., Montero-Ramírez, J., Janzen, D.H., et al., 2012. A new species of *Opsiphanes*. Doubleday, [1849] from Costa Rica (Nymphalidae: Morphinae: Brassolini), as revealed by its DNA barcodes and habitus. *Bullet. Allyn Museum* 166, 1–15.
- de Queiroz, K., 2007. Species concepts and species delimitation. *Syst. Biol.* 56 (6), 879–886.
- Dincă, V., Lukhtanov, V.A., Talavera, G., Vila, R., 2011. Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nature Commun.* 2, 324.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29 (8), 1969–1973.
- Dumas, P., Barbut, J., Le Ru, B., Silvain, J.F., Clamens, A.L., d'Alençon, E., Kergoat, G.J., 2015. Phylogenetic molecular species delimitations unravel potential new species in the pest genus *Spodoptera* Guenée, 1852 (Lepidoptera, Noctuidae). *PLoS ONE* 10 (4), e0122407.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 (5), 1792–1797.
- Ence, D.D., Carstens, B.C., 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Resources* 11 (3), 473–480.
- Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52 (5), 665–673.
- Fujita, M.K., Leaché, A.D., 2011. A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proc. Royal Soc. B: Biol. Sci.* 278 (1705), 493–495.
- 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 (9), 480–488.
- Grummer, J.A., Bryson, R.W., Reeder, T.W., 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Syst. Biol.* 63 (2), 119–133.
- Hall, R., 2012. Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. *Tectonophysics* 570, 1–41.
- Hall, R., 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J. Limnol.* 72 (2), 1–17.
- 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. Phylogenet. Evol.* 71, 79–93.
- Hausmann, A., Hebert, P., Mitchell, A., et al., 2009. Revision of the Australian *Oenochroma vinaria* Guenée, 1858 species-complex (Lepidoptera, Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. *Zootaxa* 2239, 1–21.
- Hebert, P.D., Cywinska, A., Ball, S.L., 2003. Biological identifications through DNA barcodes. *Proc. Royal Soc. London. Series B: Biol. Sci.* 270 (1512), 313–321.
- Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27 (3), 570–580.
- Henning, S., 1989. *The Charaxinae butterflies of Africa*. Aloe Books, Johannesburg.
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42 (2), 182–192.
- Hooper, D.U., Adair, E.C., Cardinale, B.J., et al., 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486 (7401), 105–108.
- Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 21 (6), 1123–1133.
- Huemer, P., Mutanen, M., 2012. Taxonomy of spatially disjunct alpine *Teleiopsis albifemorella* s. lat. (Lepidoptera: Gelechiidae) revealed by molecular data and morphology – how many species are there. *Zootaxa* 3580, 1–23.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23 (2), 254–267.
- Irwin, D.E., 2002. Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56, 2383–2394.
- Jörger, K.M., Schrödl, M., 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front. Zool.* 10 (10.1186), 1742–9994.
- Kergoat, G.J., Toussaint, E.F.A., Capdevielle-Dulac, C., Clamens, A.-L., Ong'amo, G., Conlong, D., Van den Berg, J., Cugala, D., Pallangyo, B., Mubenga, O., Chipabika, G., Ndemah, R., Sezoulin, M., Bani, G., Molo, R., Catalayud, P.-A., Kaiser, L., Silvain, J.-F., Le Ru, B.P., 2015. Integrative taxonomy reveals six new species related to the Mediterranean corn stalk borer *Sesamia nonagrioides* (Lefebvre) (Lepidoptera, Noctuidae, Sesamina). *Zoological Journal of the Linnean Society*, in press.
- Kimura, M., 2000. Paleogeography of the Ryukyu Islands. *Tropics* 10 (1), 5–24.
- Knowles, L.L., Carstens, B.C., 2007. Delimiting species with-out monophyletic gene trees. *Syst. Biol.* 56, 887–895.
- Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29 (6), 1695–1701.
- Le Ru, B.P., Capdevielle-Dulac, C., Toussaint, E.F.A., Conlong, D., Van den Berg, J., Pallangyo, B., Ong'amo, G., Chipabika, G., Mole, R., Overholt, W.A., Cuda, J.P., Kergoat, G.J., 2014. Integrative taxonomy of *Acrapex* stem borers (Lepidoptera: Noctuidae: Apameini): combining morphology and Poisson Tree Process analyses. *Invertebr. Syst.* 28 (5), 451–475.
- Le Ru, B., Capdevielle-Dulac, C., Conlong, D., Pallangyo, B., Van den Berg, J., Ong'amo, G., Kergoat, G.J., 2015. A revision of the genus *Conicofrontia* Hampson (Lepidoptera, Noctuidae, Apameini, Sesamiina), with description of a new species: new insights from morphological, ecological and molecular data. *Zootaxa* 3925 (1), 056–074.
- Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc. Royal Soc. B: Biol. Sci.* 277 (1697), 3071–3077.
- Lecocq, T., Dellicour, S., Michez, D., Dehon, M., Dewulf, A., De Meulemeester, T., Brasero, N., Valterová, I., Rasplus, J.-Y., Rasmont, R., 2015. Methods for species

- delimitation in bumblebees (Hymenoptera, Apidae, *Bombus*): towards an integrative approach. *Zool. Scripta* 44 (3), 281–297.
- Long, Y., Wan, H., Yan, F., et al., 2006. Glacial effects on sequence divergence of mitochondrial COII of *Polyura eudamippus* (Lepidoptera: Nymphalidae) in China. *Biochem. Genet.* 44 (7–8), 359–375.
- Miralles, A., Vences, M., 2013. New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLoS One* 8 (7), e68242.
- Mittermeier, R.A., Gil, P.R., Hoffman, M., et al., 2004. Hotspots revisited: Earth's biologically richest and most threatened ecoregions. CEMEX, Mexico City, Mexico.
- Müller, C.J., Wahlberg, N., Beheregaray, L.B., 2010. 'After Africa': the evolutionary history and systematics of the genus *Charaxes* Ochseneheimer (Lepidoptera: Nymphalidae) in the Indo-Pacific region. *Biol. J. Linnean Soc.* 100 (2), 457–481.
- O'Meara, B.C., 2010. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* 59, 59–73.
- O'Meara, B.C., Ané, C., Sanderson, M.J., Wainwright, P.C., 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60 (5), 922–933.
- Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. Review: the integrative future of taxonomy. *Front. Zool.* 7, 1–14.
- Pante, E., Schoelink, C., Puillandre, N., 2015. From integrative taxonomy to species description: one step beyond. *Syst. Biol.* 64 (1), 152–160.
- Pons, J., Barraclough, T.G., Gomez-Zurita, J., et al., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609.
- Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* 21 (8), 1864–1877.
- 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 (1), 245–253.
- Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoS ONE* 8 (8), e66213.
- 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 (1), 196.
- Riedel, A., Sagata, K., Suhardjono, Y.R., Tänzler, R., Balke, M., 2013. Integrative taxonomy on the fast track-towards more sustainability in biodiversity research. *Front. Zool.* 10 (1), 15.
- Ronquist, F., Teslenko, M., van der Mark, P., et al., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61 (2012), 539–542.
- Satler, J.D., Carstens, B.C., Hedin, M., 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Syst. Biol.* 62 (6), 805–823.
- Smiles, R.L., 1982. The taxonomy and phylogeny of the genus *Polyura* Billberg (Lepidoptera: Nymphalidae). *Bullet. British Museum (Natural History) (Entomology)* 44, 115–237.
- Stamatakis, A., 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22 (21), 2688–2690.
- Talavera, G., Dincă, V., Vila, R., 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods Ecol. Evol.* 4 (12), 1101–1110.
- 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 Ecol. Evol.* 5 (10), 1086–1094.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* 18 (2), 70–74.
- Thomas, C.D., Cameron, A., Green, R.E., et al., 2004. Extinction risk from climate change. *Nature* 427 (6970), 145–14.
- Toussaint, E.F.A., Hall, R., Monaghan, M.T. et al., 2014. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications*, 5, ncomms5001.
- Turlin, B., 2005. Nymphalidae X, *Charaxes* I (Vol. 22). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2007a. Nymphalidae XII, *Charaxes* II (Vol. 25). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2007b. Nymphalidae XIV, *Charaxes* III (Vol. 28). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2009. Nymphalidae XVII, *Charaxes* IV (Vol. 32). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2011. Nymphalidae XIX, *Charaxes* V (Vol. 34). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2013. Nymphalidae XXII, *Charaxes* VI (Vol. 38). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., 2014. Nymphalidae XXIV, *Palla* and *Euxanthe* (Vol. 40). In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.
- Turlin, B., Sato, H., 1995. Description of new subspecific taxa in the genus *Polyura* Billberg (Lepidoptera, Nymphalidae) from the Moluccas and Solomons Archipelagos. *Futao*, 20, 20pp.
- Wahlberg, N., Wheat, C.W., 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst. Biol.* 57 (2), 231–242.
- Wang, R.J., Wan, H., Lei, G.C., 2003. DNA sequence divergence of mitochondrial cytochrome oxidase II of *Polyura narcaea* in China. *Acta Scientiarum Naturalium Universitatis Pekinensis* 39 (6), 775–779.
- Wang, R., Wan, H., Long, Y., Lei, G., Li, S., 2004. Phylogenetic analysis of *Polyura* in China inferred from mitochondrial CO II sequences (Lepidoptera: Nymphalidae). *Acta Entomol. Sin.* 47 (2), 243–247.
- Wardle, D.A., Bardgett, R.D., Callaway, R.M., Van der Putten, W.H., 2011. Terrestrial ecosystem responses to species gains and losses. *Science* 332 (6035), 1273–1277.
- Williams, K.J., Ford, A., Rosauer, D.F. et al., 2011. Forests of East Australia: the 35th biodiversity hotspot. In: *Biodiversity Hotspots*. Springer, Berlin Heidelberg, pp. 295–310.
- Wilson, J.J., Sing, K.W., Sofian-Azirun, M., 2013. Building a DNA barcode reference library for the true butterflies (Lepidoptera) of Peninsula Malaysia: what about the subspecies? *PLoS one* 8 (11), e79969.
- Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. USA* 107 (20), 9264–9269.
- Zhang, C., Zhang, D.X., Zhu, T., Yang, Z., 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Syst. Biol.* 60 (6), 747–761.
- Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29 (22), 2869–2876.
- Zwickl, D.J., 2008. GARLI (Genetic algorithm for rapid likelihood inference), version 0.96. URL <<http://www.bio.utexas.edu/faculty/anjitise/garli/garli.html>>.