Systematic
Entomology

Systematic Entomology (2019), DOI: 10.1111/syen.12365

Out of the Orient: Post-Tethyan transoceanic and trans-Arabian routes fostered the spread of Baorini skippers in the Afrotropics

EMMANUEL F.A. TOUSSAINT[1,2], ROGER VILA[3], MASAYA YAGO[4], HIDEYUKI CHIBA[5], ANDREW D. WARREN[2], KWAKU ADUSE-POKU[6,7], CAROLINE STORER[2], KELLY M. DEXTER[2], KIYOSHI MARUYAMA[8], DAVID J. LOHMAN[6,9,10] and AKITO Y. KAWAHARA[2]

[1]Natural History Museum of Geneva, Geneva, Switzerland, [2]Florida Museum of Natural History, University of Florida, Gainesville, FL, U.S.A., [3]Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain, [4]The University Museum, The University of Tokyo, Bunkyo-ku, Japan, [5]B. P. Bishop Museum, Honolulu, HI, U.S.A., [6]Biology Department, City College of New York, City University of New York, New York, NY, U.S.A., [7]Biology Department, University of Richmond, Richmond, VA, U.S.A., [8]9-7-106 Minami-Ôsawa 5 chome, Hachiôji-shi, Japan, [9]Ph.D. Program in Biology, Graduate Center, City University of New York, New York, NY, U.S.A. and [10]Entomology Section, National Museum of the Philippines, Manila, Philippines

Abstract. The origin of taxa presenting a disjunct distribution between Africa and Asia has puzzled biogeographers for more than a century. This biogeographic pattern has been hypothesized to be the result of transoceanic long-distance dispersal, Oligocene dispersal through forested corridors, Miocene dispersal through the Arabian Peninsula or passive dispersal on the rifting Indian plate. However, it has often been difficult to pinpoint the mechanisms at play. We investigate biotic exchange between the Afrotropics and the Oriental region during the Cenozoic, a period in which geological changes altered landmass connectivity. We use Baorini skippers (Lepidoptera, Hesperiidae) as a model, a widespread clade of butterflies in the Old World tropics with a disjunct distribution between the Afrotropics and the Oriental region. We use anchored phylogenomics to infer a robust evolutionary tree for Baorini skippers and estimate divergence times and ancestral ranges to test biogeographic hypotheses. Our phylogenomic tree recovers strongly supported relationships for Baorini skippers and clarifies the systematics of the tribe. Dating analyses suggest that these butterflies originated in the Oriental region, Greater Sunda Islands, and the Philippines in the early Miocene *c.* 23 Ma. Baorini skippers dispersed from the Oriental region towards Africa at least five times in the past 20 Ma. These butterflies colonized the Afrotropics primarily through trans-Arabian geodispersal after the closure of the Tethyan seaway in the mid-Miocene. Range expansion from the Oriental region towards the African continent probably occurred via the *Gomphotherium* land bridge through the Arabian Peninsula. Alternative scenarios invoking long-distance dispersal and vicariance are not supported. The Miocene climate change and biome shift from forested areas to grasslands possibly facilitated geodispersal in this clade of butterflies.

Correspondence: Emmanuel F.A. Toussaint, Natural History Museum of Geneva, Geneva, Switzerland. E-mail: emmanuel.touss1@gmail.com

Introduction

The remarkable geological features of the Old World tropics, such as the Indo-Australian Archipelago (IAA), the Himalayan mountain range, and the Deccan traps are testament to the dynamic geological evolution that took place throughout the Cenozoic (Seton *et al.*, 2012). The collision of the Asian, Australian and Pacific plates resulted in the assemblage of Wallacea (Hall, 2013) and orogeny of New Guinea (Toussaint *et al.*, 2014), allowing the Asian and Australian biotas to connect through shallow marine straits and ephemeral land bridges between myriad tropical islands (Voris, 2000). The collision of the rifting Indian plate with the Asian plate in the Paleocene (Hu *et al.*, 2016), triggered the orogeny of the Himalayas (Zhang *et al.*, 2012), and connected the relic Gondwanan stock with the Laurasian biota. The closure of the Tethyan Ocean in the Oligocene (Pirouz *et al.*, 2017) triggered the orogeny of the Zagros mountain chain along the Alpine–Himalayan orogenic belt, thereby connecting African and Asian biotas by a land bridge (i.e. the *Gomphotherium* land bridge; Rögl, 1998) through the Arabian Peninsula (Harzhauser *et al.*, 2007; Berra & Angiolini, 2014). Disentangling the impact of these major geological rearrangements on the evolution of the regional biota is necessary to understand the mechanisms of lineage diversification in the Old World tropics.

Of particular interest are clades of organisms disjunctly distributed in Asia and Africa, because these allow testing hypotheses related to dispersal, vicariance and extinction with respect to the geology of the African, Arabian, Asian and Indian plates. The distribution of such clades has been explained by four biogeographic hypotheses. The first hypothesis, transoceanic long-distance dispersal (LDD) between Asia and Africa, has been invoked to explain disjunct distributions. Under this hypothesis (H_1), the colonization of distant areas from a source geographic range is accomplished directly and is not the result of range expansion followed by widespread extinction/range contraction. A second hypothesis posits range expansion with trans-Arabian dispersal (H_2, also known as the '*Gomphotherium* land bridge hypothesis'; Harzhauser *et al.*, 2007) followed by regional extinction in the Arabian Peninsula. The *Gomphotherium* land bridge hypothesis is based on geological evidence that the eastern part of the Tethys Ocean (i.e. Neotethys) that once separated Africa and Asia, closed as a result of the Arabian plate moving northwards and colliding with the Asian plate in the Oligocene *c*. 27 Ma, engendering the orogeny of the Zagros mountain chain in the process (Pirouz *et al.*, 2017). A third hypothesis is derived from the 'boreotropical dispersal hypothesis'. This suggests that lineages expanded their geographic ranges in higher latitudes through boreotropical forests that existed in the Eocene and Oligocene before climate cooling triggered the wane of boreotropical forests, in turn promoting ecological vicariance (Wolfe, 1975). Recent studies (e.g. Couvreur *et al.*, 2011; Sánchez-Ramírez *et al.*, 2015) suggest that a similar process could be invoked for Africa–Asia disjunctions, where range expansion could have happened after the closure of the Tethys Ocean via now-vanished tropical forests in the Arabian Peninsula (Ghazanfar & Fisher, 1998; Griffin, 2002). This hypothesis, which we coin 'Arabotropical forest dispersal' (H_3), implies that range expansion would had to have happened between the emergence of these habitats and their disappearance, triggered by the onset of a cooler climate in the late Eocene and Oligocene (Zachos *et al.*, 2001). As a result, this hypothesis predicts range expansions resulting in widespread ancestral distributions followed by vicariance triggered by the aridification of the Arabian Peninsula and associated floristic reconfiguration (Ghazanfar & Fisher, 1998; Griffin, 2002). The last hypothesis proposes that organisms rafted on the Gondwanan Indian plate ('biotic ferry hypothesis' or 'out-of-India hypothesis'; Datta-Roy & Karanth, 2009) and subsequently diversified in Asia. Under this hypothesis (H_4), lineages occurring in Gondwana during the Cretaceous and therefore in the landmass comprising Africa, India and Madagascar at the time, would have been transported via the rafting Indian plate across the Indian Ocean before docking in the Oriental region in the Eocene *c*. 60 Ma (Hu *et al.*, 2016). Although this hypothesis was initially supported by phylogenetic patterns, it has since been largely rejected by numerous studies of unrelated plants and animals that use molecular dating approaches (Toussaint *et al.*, 2016). A related mechanism is the use of the Indian rafting plate as a 'stepping stone', whereby taxa dispersed from the Oriental region to insular India before its final docking and then dispersed towards other Gondwanan landmasses (Toussaint & Short, 2018). Both processes imply ancient African and Oriental divergences. There are few empirical studies focusing on animal taxa that investigate the evolution of African–Asian disjunctions in a phylogenetic framework with the aim of testing these competing hypotheses.

Baorini grass skippers (Lepidoptera, Hesperiidae, Hesperiinae) are widespread butterflies distributed from Madagascar, the Mascarenes, Africa and southern Europe, through the Arabian Peninsula, and to the Oriental region and the IAA (Warren *et al.*, 2009). The tribe currently includes 17 genera (Fan *et al.*, 2016; Zhu *et al.*, 2016; de Jong & Coutsis, 2017) and *c*. 100 species, which are mostly distributed in the Oriental region and the IAA, with a few African species in geographically widely distributed genera or species-rich clades (Fan *et al.*, 2016). *Parnara* Moore comprises several Oriental/Indo-Australian species and two species in Madagascar, the Mascarenes and southeastern coastal Africa. *Zenonia* Evans is endemic to Africa and comprises three widespread species, and *Gegenes* Hübner has three African endemic species and another more widely distributed species in the Oriental and Palearctic regions. The mostly Oriental and Indo-Australian genus, *Pelopidas* Walker, comprises two species with geographic ranges extending from the IAA to continental Africa and Madagascar. The four genera *Afrogegenes* de Jong & Coutsis, *Borbo* Evans, *Brusa* Evans and *Larsenia* Chiba, Fan & Sáfián are wholly African (except *Borbo* that is also found in Asia) with multiple endemic species, some restricted to Madagascar. The other genera, *Baoris* Moore, *Caltoris* Swinhoe, *Iton* de Nicéville, *Polytremis* Mabille, *Prusiana* Evans, *Pseudoborbo* Lee, *Tsukiyamaia* Zhu, Chiba & Wu, *Zinaida* Evans and *Zenonoida* Fan & Chiba, are distributed from the Oriental region, including India, to the Australian region and some Pacific islands in the Solomon Islands and Vanuatu

(Braby, 2000). A few recent molecular systematic studies have clarified some phylogenetic relationships of Baorini (Jiang *et al.*, 2013; Fan *et al.*, 2016; Zhu *et al.*, 2016; Li *et al.*, 2017; Tang *et al.*, 2017), but these studies included a relatively small number of loci (≤ 5) and species (≤ 50), and have been unable to confidently place genera, making it difficult to assess their biogeographic history. In the present study, we use a phylogenomic approach to sequence 13 gene fragments, including those most commonly used in butterfly phylogenetics (Wahlberg & Wheat, 2008). We then combine this newly generated dataset with other available data to infer a robust time-tree of nearly 70 Baorini species and test whether the distribution of Afrotropical taxa can be linked to one of the four biogeographic hypotheses described earlier. We use a hypothetico-deductive framework to test whether the phylogenomic pattern, divergence time estimates and ancestral range estimation are consistent with one or more of the hypotheses. We do not test the Indian biotic ferry hypothesis, because divergence time evidence suggests that Baorini are much younger than the invoked geological events (Espeland *et al.*, 2018).

Materials and methods

Taxon sampling and molecular biology

Samples of 32 Baorini species were collected in the field (Appendix S1) and placed in glassine envelopes or directly in $>95\%$ ethanol. In the laboratory, DNA was extracted from abdomens or legs using an OmniPrep™ DNA extraction kit (G-Biosciences, Saint-Louis, MO, U.S.A.) following Toussaint *et al.* (2018). Quantified DNA extracts were submitted to RAPiD Genomics (Gainesville, FL, U.S.A.) for library preparation, hybridization enrichment, and sequencing. Random mechanical shearing of DNA was conducted with an average size of 300 bp followed by an end-repair reaction and ligation of an adenine residue to the 3′-end of the blunt-end fragments to allow ligation of barcode adapters and PCR amplification of the library. Following library construction, solution-based target enrichment of Agilent (Santa Clara, CA, U.S.A.) SureSelect probes was conducted in a pool containing 16 libraries. These libraries were enriched with the SureSelect Target Enrichment System for Illumina (San Diego, CA, U.S.A.) Paired-End Multiplexed Sequencing Library protocol. Paired-end sequencing of 150 bp reads was accomplished with an Illumina HiSeq.

We used the BUTTERFLY2.0 probe set (Kawahara *et al.*, 2018) to capture the following 13 loci: acetyl-CoA (ACOA, 1020 bp), carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD, 1854 bp), catalase (CAT, 1290 bp), cytochrome oxidase *c* subunit 1 (CO1, 1341 bp), dopa decarboxylase (DDC, 702 bp), elongation factor 1 alpha (EF1A, 1059 bp), glyceraldhyde-3-phosphate dehydrogenase (GAPDH, 606 bp), hairy cell leukaemia protein 1 (HCL, 633 bp), isocitrate dehydrogenase (IDH, 708 bp), malate dehydrogenase (MDH, 681 bp), ribosomal protein S2 (RPS2, 471 bp), ribosomal protein S5 (RPS5, 555) and wingless (WGL, 240 bp).

We incorporated sequences of 47 additional Baorini species available on GenBank and the Barcode of Life Database, which were included in recent systematic studies of the tribe (Fan *et al.*, 2016; Zhu *et al.*, 2016; Tang *et al.*, 2017). We included sequence data for the mitochondrial ribosomal 16S (533 bp), and nuclear ribosomal 18S (378 bp) and 28S (857 bp). When possible, we built chimeras between newly sequenced specimens and other specimens of the same species for which other sequence data were available to improve sequence coverage for terminals. We included 19 skipper outgroups from all subfamilies and major clades (Toussaint *et al.*, 2018) to allow the use of secondary calibrations. All sequence data were imported into GENEIOUS R8.1.8 (Biomatters, New-Zealand), cleaned and aligned separately for each locus using MUSCLE (Edgar, 2004). The alignments of ribosomal loci were straightforward and did not require fine-tuning, probably because of the relative youth of this taxon. Gene trees were inferred using FASTTREE 2.1.5 (Price *et al.*, 2010) and inspected for contamination and sequence quality. Trees were rooted with *Macrosoma hyacinthina* Warren (Hedylidae), as the Hedylidae are the sister group to the Hesperiidae (Espeland *et al.*, 2018; Toussaint *et al.*, 2018). The final matrix included 89 species, including 70 Baorini, and 16 concatenated loci, totalling up to 12 928 aligned nucleotides.

Anchored hybrid enrichment data assembly and clean-up

We used the pipeline for anchored phylogenomics of Breinholt *et al.* (2018) to create a matrix from raw Illumina reads. Paired-end Illumina data were cleaned with TRIM GALORE! v.0.4.0 (www.bioinformatics.babraham.ac.uk), allowing a minimum read size of 30 bp, and we removed bases with a Phred score < 20. The 13 loci that were sequenced using the BUTTERFLY2.0 kit (Kawahara *et al.*, 2018) were assembled with iterative baited assembly (Breinholt *et al.*, 2018), in which only reads with both forward and reverse reads that passed filtering were included. Assembled reads from the probe region were blasted against the *Danaus plexippus* Linnaeus (Nymphalidae) reference genome and BLAST results were used for single hit and orthology filtering. Following Breinholt *et al.* (2018), the loci were screened for orthology with a single hit threshold of 0.9 and genome mapping. The identified orthologous sequences were screened for contamination by identifying and removing sequences that were nearly identical at the family and genus levels (Breinholt *et al.*, 2018). Sequences from each locus were separately aligned with MAFFT v.7.245 (Katoh & Standley, 2013) and concatenated with FASCONCAT-G 1.0.4 (Kück & Meusemann, 2010). ALISCORE 2.2 (Kück *et al.*, 2010) was used to check for saturation and sites that appeared to evolve randomly. The individual locus alignments and files used for phylogenetic inference in this study (see later) are available on Dryad (https://doi.org/10.5061/dryad.6qp7p13).

Datasets, model partitioning and phylogenetic analysis

We simultaneously estimated the best-partitioning scheme and corresponding models of nucleotide substitution in IQ-TREE 1.6.9 (Nguyen *et al.*, 2015) using MODELFINDER

(Kalyaanamoorthy *et al.*, 2017) with the greedy algorithm and based on the corrected Bayesian information criterion (BIC). The optimal models of nucleotide substitution were determined across all available models in IQ-TREE, including the FreeRate model (+R; Soubrier *et al.*, 2012), which relaxes the assumption of gamma distributed rates. The 13 protein-coding loci were divided by codon positions (39 partitions), and the three nonprotein-coding loci were left as independent partitions (three partitions). To find the most likely tree, 100 maximum likelihood (ML) searches were conducted in IQ-TREE, with two calculations of nodal support: ultrafast bootstrap (UFBoot) and Shimodaira–Hasegawa approximate likelihood ratio (SH-aLRT) tests. We generated 1000 replicates for UFBoot (Hoang *et al.*, 2018) and SH-aLRT (Guindon *et al.*, 2010). To reduce the risk of overestimating branch supports with UFBoot due to severe model violations, we used hill-climbing nearest neighbour interchange (NNI) to optimize each bootstrap tree (Hoang *et al.*, 2018). All phylogenomic analyses were conducted on the University of Florida HiPerGator High Performance Computing Cluster (www.hpc.ufl.edu). When discussing branch support, we refer to 'strong' support as SH-aLRT ≥ 80 and UFBoot ≥ 95, and 'moderate' support as SH-aLRT ≥ 80 or UFBoot ≥ 95.

We also performed an additional set of analyses based on anchored hybrid enrichment (AHE) data alone, excluding all information from GenBank to ensure that chimeric constructions and molecular matrix incompleteness were not biasing the phylogenetic inference. This reduced dataset comprised 50 taxa, including 31 Baorini representatives. The methods were identical to the ones detailed earlier for the full dataset. The best-partitioning scheme and corresponding models of nucleotide substitution were simultaneously estimated in IQ-TREE using MODELFINDER based on the BIC, and 100 ML searches were conducted to avoid local optima with computation of 1000 UFBoot replicates and 1000 SH-aLRT tests.

Divergence time estimation

We estimated divergence times in a Bayesian framework with BEAST 1.10.1 (Suchard *et al.*, 2018). The best-partitioning scheme and models of substitution were selected in PARTITIONFINDER2 (Lanfear *et al.*, 2016) using the greedy algorithm and the BIC across all models included in BEAST (option models = beast). The dataset was partitioned *a priori*, similar to the MODELFINDER analysis (see earlier). To take into account the importance of clock partitioning, we implemented: (i) a single clock for all partitions; (ii) two clocks, one for the mitochondrial partition and one for the nuclear partitions; and (iii) one clock for first and second codon positions of protein-coding genes, one clock for third codon positions of protein-coding genes, and one clock for each ribosomal gene for a total of five clocks. We assigned a Bayesian lognormal relaxed clock model to the different clock partitions. We also tested different tree models by using a Yule or a birth–death model in different analyses. Clock rates were set with an approximate continuous time Markov chain rate reference prior (Ferreira & Suchard, 2008). All analyses consisted of 100 million generations with a parameter and tree sampling every 5000 generations. We estimated marginal likelihood estimates (MLEs) for each analysis using path-sampling and stepping-stone sampling (Baele *et al.*, 2012, Baele *et al.*, 2013), with 1000 path steps, and chains running for one million generations with a log-likelihood sampling every 1000 cycles.

Two hesperiid fossils, *Pamphilites abdita* Scudder and *Protocoeliades kristenseni* de Jong, were candidates to be included as fossil calibrations for this study (de Jong, 2017). However, preliminary testing in BEAST revealed that the age estimates obtained with these two fossils were largely underestimated compared with the ages from Espeland *et al.* (2018) or Chazot *et al.* (2019). This discrepancy is partly due to the difficulty of placing the oldest described skipper fossil into an extant taxon: *Protocoeliades kristenseni* from Denmark, dated at *c.* 55 Ma (de Jong, 2016). This fossil has been described as belonging to the Coeliadinae based on morphology (de Jong, 2016). However, because relationships among Coeliadinae genera and their monophyly are uncertain, any placement of the fossil elsewhere than on the stem of Coeliadinae remains contentious (de Jong, 2016, 2017). Therefore, because Coeliadinae is sister to the remainder of Hesperiidae (Espeland *et al.*, 2018; Toussaint *et al.*, 2018), the fossil would only calibrate a minimum age of 55 Ma for crown Hesperiidae, while Espeland *et al.* (2018) estimated this node to be *c.* 76 Ma and the split between Hesperiidae and Hedylidae to be *c.* 106 Ma, and Chazot *et al.* (2019) estimated these nodes to *c.* 65 Ma and *c.* 99 Ma. Because Chazot *et al.* (2019) has a more extensive taxon sampling, including Baorini genera, we chose this study for secondary calibrations over the more robust phylogenomic tree of Espeland *et al.* (2018). We constrained the node corresponding to the split between Hesperiidae and Hedylidae with a uniform prior encompassing the 95% credibility interval (lower = 80.73, upper = 119.22) estimated for this node in Chazot *et al.* (2019). We also constrained the crown of Hesperiidae (lower = 55.8, upper = 78.1) and the node corresponding to all Baorini excluding *Parnara* (lower = 10.3, upper = 17.2).

Ancestral range estimation

We used the R package BIOGEOBEARS 1.1.1 (Matzke, 2018) to estimate ancestral ranges in Baorini. Because the validity of model comparison in BIOGEOBEARS is debated (Ree & Sanmartín, 2018), the analyses were only performed under the dispersal–extinction–cladogenesis (DEC) model (Ree, 2005; Ree & Smith, 2008). The DEC model specifies instantaneous transition rates between geographical ranges along the branches of a given phylogeny. The likelihoods of ancestral states at cladogenesis events in the phylogeny are then estimated in a ML framework, with probabilities of range transitions as a function of time. We used the BEAST maximum clade credibility tree from the best analysis (see the Results section) with outgroups pruned. The geographic distribution of Baorini was extrapolated from the literature (Chiba & Eliot, 1991; de Jong & Treadaway, 1993; Braby, 2000; Zhang *et al.*, 2010; Huang, 2011). In order to test the three hypotheses related to the colonization of Africa, we

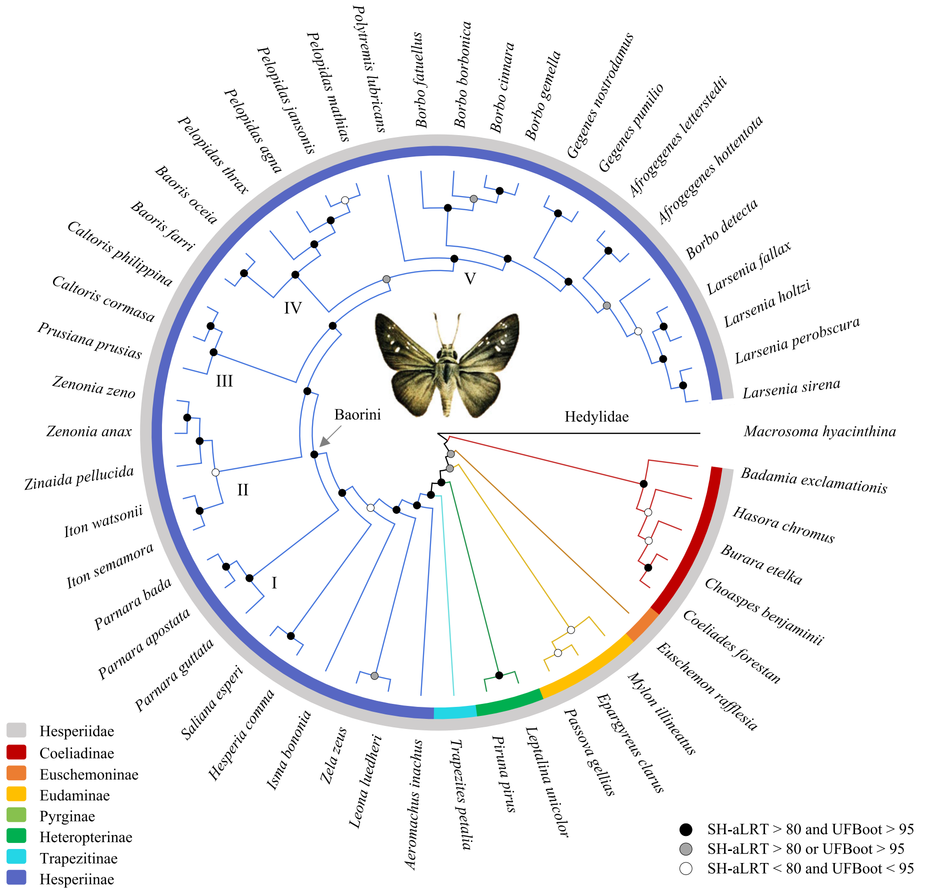

Fig. 1. Maximum likelihood (ML) tree of Baorini grass skippers based on anchored phylogenomics. Best-scoring ML phylogenetic tree inferred in IQ-TREE based on the concatenated dataset of 13 loci comprised in the BUTTERFLY2.0 kit (Kawahara *et al.*, 2018). The five main Baorini clades are labelled I–V. A historical drawing of *Pelopidas mathias* is presented from the book *Indian Insect Life* by Harold Maxwell-Lefroy (1909). SH-aLRT, Shimodaira–Hasegawa approximate likelihood ratio test; UFBoot, ultrafast bootstrap test. [Colour figure can be viewed at wileyonlinelibrary.com].

used the following areas in the BIOGEOBEARS analyses: Western Palearctic (P) from southern Europe to Turkey; Africa, Madagascar and Mascarenes (F); Arabian Peninsula (A); Oriental (O) from Iran to Japan and from China to the Malaysian Peninsula; Greater Sunda Islands and Philippines (G), including Borneo, Java, Palawan, Sumatra and satellite islands; Wallacea (W), including the Lesser Sunda Islands, Moluccas and Sulawesi, including satellite islands; and Australia and New Guinea (N), including satellite islands.

Considering the relative geological stability of the African and Asian regions in the late Cenozoic, we only took into account the dynamic geological history of the IAA by designing two time slices with differential dispersal rate scalers. TS1 (25–15 Ma) corresponds to the period predating the acceleration of orogenies in the Philippines archipelago, Wallacea and New Guinea (Yumul Jr *et al.*, 2008; Hall, 2013; Toussaint *et al.*, 2014), and TS2 (15 Ma to present) corresponds to the above-mentioned events, more active formation of the IAA and intensifying sea-level fluctuations in the Plio-Pleistocene (Voris, 2000). The dispersal rate scaler values were selected according to terrain and water body positions throughout the timeframe of the age of the tribe (Appendix S2). The maximum number of areas per ancestral state was fixed to five. To reduce the computational burden, ancestral states corresponding to unrealistic disjunct areas were removed manually from the list of possible states.

Results

Phylogenetic relationships

The phylogenetic tree based on the reduced dataset (AHE data only) with the highest likelihood ($\ln L = -113\,733.176$) from the 100 searches performed in IQ-TREE is presented in Fig. 1 (see Appendix S3 for the full tree, including outgroups), and the one based on the full dataset ($\ln L = -126\,096.182$) is presented in Fig. 2. As the phylogenies inferred in ML with both datasets gave similar branching patterns, we only discuss the ones recovered with the greater taxon sampling and presented in Fig. 2. The inferred tree is well-resolved with more than 85% of nodes that recovered moderate or strong support. The phylogeny inferred in BEAST under Bayesian inference (BI)

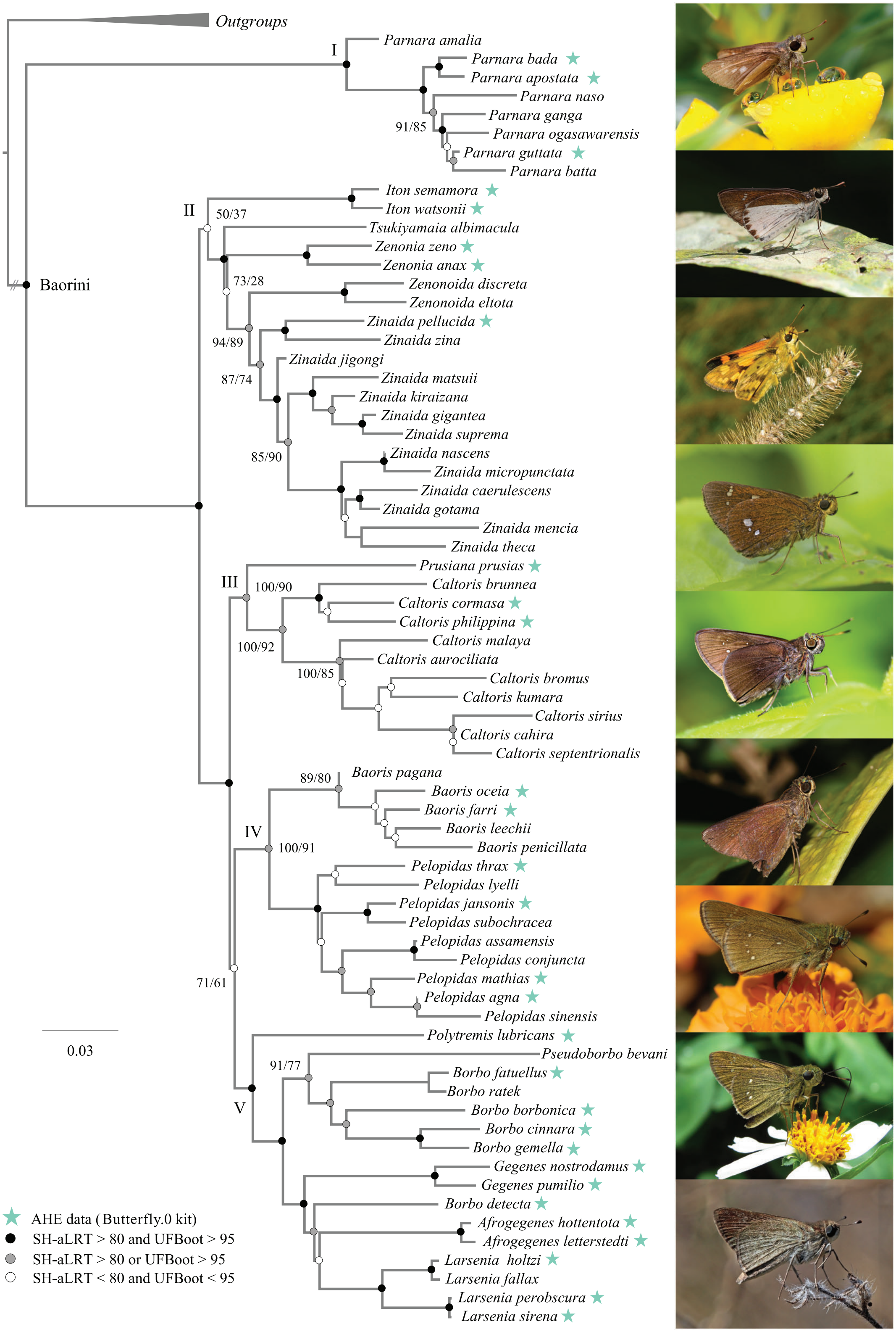

Fig. 2. Legend on next page [Colour figure can be viewed at wileyonlinelibrary.com].

gave a similar topology (Fig. 2; Appendix S3) and a greater number of nodes with strong support values [posterior probability (PP) ≥ 0.95]. *Hesperia* Fabricius and *Saliana* Evans are recovered as sister to Baorini with strong support in both ML and BI (UFBoot = 100, SH-aLRT = 100, PP = 1.0). Baorini is recovered as monophyletic with strong support (UFBoot = 100, SH-aLRT = 100, PP = 1.0) and is split into five clades. *Parnara* is recovered in clade I and as sister to the remainder of the tribe with strong support (UFBoot = 100, SH-aLRT = 100, PP = 1.0). Clade II, which comprises the genera *Iton*, *Tsukiyamaia*, *Zenonia*, *Zenonoida* and *Zinaida*, is recovered with weaker support in ML (UFBoot = 37, SH-aLRT = 50) but with strong support in BI (PP = 1.0). Clade III is recovered with moderate to strong support (UFBoot = 90, SH-aLRT = 100, PP = 0.99) and comprises *Prusiana* as sister to *Caltoris*. In clade IV, *Baoris* is recovered as sister to *Pelopidas* with moderate to strong support (UFBoot = 91, SH-aLRT = 100, PP = 0.99). Finally, in clade V, *Polytremis lubricans* Herrich-Schäffer is sister to a clade comprising the monotypic *Pseudoborbo*, the polyphyletic *Borbo*, *Gegenes* and the sister genera *Afrogegenes* and *Larsenia*. Species of Baorini distributed in Africa, Madagascar, or the Mascarenes are recovered in different clades across the tree.

Divergence times and biogeography

Dating analyses using different numbers of clocks and tree models estimated similar ages with broadly overlapping credibility intervals (Table 1). Based on the MLE comparison (Table 1), the analysis with two clocks and a Yule model was selected for further analyses. The results of this analysis are presented in Fig. 3 (see Appendix S4 for the full chronogram), along with results of the DEC model in BIOGEOBEARS (DEC, $\ln L = -176.24$). The BIOGEOBEARS analysis suggested an origin in the Oriental region and Greater Sunda Islands/Philippines, with multiple independent colonization events of Africa, Madagascar and the Mascarenes, but also of regions east of Wallace's Line in the Miocene (Appendix S5). We recover an origin of Baorini in the early Miocene *ca.* 23 Ma in the Oriental region and Greater Sunda Islands/Philippines. Based on this result, we estimate range expansion across several areas in *Parnara* (clade I) toward the IAA and Australian region. This initial range expansion in *Parnara* is followed by a second range expansion towards Africa, Madagascar and the Mascarenes in the late Miocene to the early Pliocene. We infer a range contraction in the remainder of Baorini, a lineage that shows an Oriental origin in the mid-Miocene (Fig. 3). Within clade II, we estimate range expansion towards the Western Palearctic, Africa, Madagascar and the Mascarenes from the Oriental region in the Miocene with a potential vicariance between *Zenonia* and *Zinaida* (Fig. 3). We estimate an origin of clade III in the Greater Sunda Islands/Philippines, followed by range contraction in the Oriental region in *Caltoris*. The ancestor of *Prusiana* colonized Wallacea sometime in the Miocene or Plio-Pleistocene. We also infer range expansion towards the Greater Sunda Islands/Philippines in *Caltoris*, followed by range expansion towards Wallacea and the Australian region in the Miocene or Pliocene. In clade IV, the ancestor of *Pelopidas* is inferred to have expanded its range from the Oriental region to Wallacea in the late Miocene, allowing a subsequent range expansion towards Australia and New Guinea in *P. lyelli*. A similar range expansion is estimated in the sister pair *P. agna* Moore/*P. mathias* Fabricius. The colonization of Africa, Madagascar and the Mascarenes by the ancestors of *P. mathias* and *P. thrax* Hübner is estimated to have occurred in the late Miocene to Plio-Pleistocene through independent range expansions. We estimate a mid-Miocene range expansion from the Oriental region towards a vast region encompassing the Arabian Peninsula, the Western Palearctic region, Africa, Madagascar and the Mascarenes, the Oriental region and the Greater Sunda Islands/Philippines in clade CV. The Arabian Peninsula was also colonized from Africa in the past 5 Ma in *Afrogegenes* (Fig. 3).

Discussion

Phylogenetics and taxonomy of Baorini

Our phylogenomic analysis recovers a well-resolved phylogenetic hypothesis for Baorini skippers, especially in the BEAST analysis. Zhu *et al.* (2016) recovered *Parnara* as sister to the remainder of Baorini, a relationship consistent with our results (Figs 1–3) but at odds with the conclusions of Fan *et al.* (2016), who recovered *Larsenia* as sister to the remainder of Baorini, including *Parnara*. Unlike the results of Fan *et al.* (2016), *Larsenia* is nested within clade V and these relationships are strongly supported in our analyses, although the exact placement of *Larsenia* within this clade is uncertain (Figs 1–3). The placement of other genera in the studies of Zhu *et al.* (2016) and Fan *et al.* (2016) were only partially consistent with our results, albeit with weak support. These previous studies were based on few genetic markers (three and five loci, respectively), resulting in inconsistencies among studies. Our inferences consistently recover five clades with strong support that are somewhat consistent with the current tribal classification. The only taxonomic discrepancy in our inferences

Fig. 2. Maximum likelihood (ML) tree of Baorini grass skippers based on an extended dataset. Best-scoring ML phylogenetic tree inferred in IQ-TREE based on the concatenated dataset of 16 loci, including ribosomal loci recovered from GenBank. The five main Baorini clades are labelled I–V. Green stars indicate species for which data were generated in this study, and others for which data were already available are not labelled. Photographs of Baorini species are given on the right-hand side of the figure. From top to bottom: *Parnara ganga* (credit: Antonio Giudici), *Iton semamora* (credit: Antonio Giudici), *Zenonia zeno* (credit: Geoffrey Prosser), *Zenonoida discreta* (credit: Antonio Giudici), *Caltoris cahira* (credit: Antonio Giudici), *Baoris farri* (Jee & Rani Nature Photography), *Pelopidas mathias* (credit: Laitche), *Borbo cinnara* (credit: Hsu Hong Lin), *Gegenes pumilio* (credit: Charles James Sharp). AHE, anchored hybrid enrichment; SH-aLRT, Shimodaira–Hasegawa approximate likelihood ratio test; UFBoot, ultrafast bootstrap.

Table 1. Comparison of BEAST divergence time analyses.

Analysis	Relaxed clocks	Tree model	SS MLE	PS MLE	Baorini	Clade I	Clade II	Clade III	Clade IV	Clade V
A1	1	Yule	−127 312.248	−127 312.461	22.519 (18.277–26.284)	5.373 (3.378-8.031)	13.906 (11.661–16.214)	12.215 (10.048–14.222)	10.715 (8.625–12.701)	12.250 (10.189–14.071)
A2	1	Birth–death	−127 319.429	−127 320.4659	21.629 (17.145–25.808)	5.017 (3.137-7.547)	13.230 (10.716–15.774)	11.506 (9.246–13.682)	10.059 (7.924–12.198)	11.566 (9.366–13.621)
A3	2	Yule	−127 146.005	−127 146.829	22.839 (19.543–26.577)	5.868 (3.193-10.337)	15.105 (13.193–16.643)	13.139 (11.137–14.880)	11.323 (9.496–13.107)	13.282 (11.447-14.824)
A4	2	Birth–death	−127 155.605	−127 155.341	22.676 (18.794–26.654)	5.282 (2.872–9.538)	14.369 (11.995–18.373)	12.362 (10.014–14.395)	10.780 (8.695–12.674)	12.551 (10.456–14.435)
A5	5	Yule	−127 256.942	−127 256.656	22.258 (18.489–25.943)	5.246 (3.263–7.751)	13.736 (11.478–15.978)	12.189 (10.170–14.113)	10.400 (8.515–12.235)	12.130 (10.253–13.938)
A6	5	Birth–death	−127 261.944	−127 261.822	21.755 (17.937–25.543)	4.908 (3.057–7.324)	13.129 (10.827–15.351)	11.627 (9.604–13.619)	9.897 (8.004–11.774)	11.604 (9.624–13.460)

PS, path sampling; SS, stepping-stone sampling; MLE, marginal likelihood estimate from the path sampling and stepping-stone analyses conducted in BEAST; the values in the 'clade' columns are median ages in Ma estimated in BEAST along with 95% credibility intervals.

is the polyphyly of *Borbo*. Fan *et al.* (2016) described *Larsenia* as comprising a number of exclusively African species of *Borbo*, because their analysis recovered this clade as sister to all Baorini. Based on our reconstruction, *Larsenia* is one of the most derived clades within Baorini (Figs 1–3). It is unclear what should be done regarding *Borbo detecta* Trimen, as it is recovered as sister to *Afrogegenes* and *Larsenia*. Additional taxon sampling would be needed to propose taxonomic changes in this clade.

Origins of Baorini skippers

The placement of Baorini within Hesperiinae remains unclear due to the staggering diversity of the subfamily (> 2000 described species) and the lack of a comprehensive phylogeny for this group. However, recent phylogenomic evidence (Toussaint *et al.*, 2018) suggests that Baorini might be part of a large phylogenetic grade, thereby rendering difficult the inclusion of outgroups that might inform the estimation of ancestral ranges (Toussaint & Short, 2018). Our BIOGEOBEARS analysis recovers an origin of Baorini in the Oriental region plus Greater Sunda Islands/Philippines. We deliberately chose a broader definition of the Oriental region to avoid losing biogeographic signal while retaining the power to test hypotheses relative to the mechanisms for the colonization of Africa. Ancestors of Baorini probably originated in an area climatically and biologically similar to the modern Oriental region, but a finer-scale biogeographical and ecological study of specific clades would be needed to test this hypothesis. Colonization of Wallacea and Sahul (i.e. the Australian region) would have been difficult before *c*. 15–20 Ma, except perhaps through LDD (Hall, 2013); the inferred ancestral ranges of early diverging Baorini in the Oriental region are therefore unsurprising.

Hypothesis testing: African colonization of Baorini skippers

Our divergence time estimation supports H_2 (*Gomphotherium* land bridge), as the African continent was mostly colonized by range expansion (e.g. ancestors of *P. mathias* and *P. thrax* in clade IV; Fig. 3), probably taking advantage of the *Gomphotherium* land bridge (Rögl, 1998). Consequently, the BIOGEOBEARS analysis rejects H_1 (transoceanic LDD), although some dispersal events were probably required to explain, for instance, the transition from the Oriental region to a larger range also comprising the Western Palearctic and the African continent in clade II (Fig. 3). Both our dating and BIOGEOBEARS analysis reject H_3 (Arabotropical forest dispersal). For the ancestors of Baorini to take advantage of tropical forests, as found in plants and fungi (Couvreur *et al.*, 2011; Sánchez-Ramírez *et al.*, 2015), they would have had to colonize the Arabian Peninsula region in the Oligocene when these forests still existed. Instead, our BIOGEOBEARS estimate suggests that the ancestors of Baorini originated in the Oriental region in the Miocene. By the time Baorini skippers began to disperse to other regions

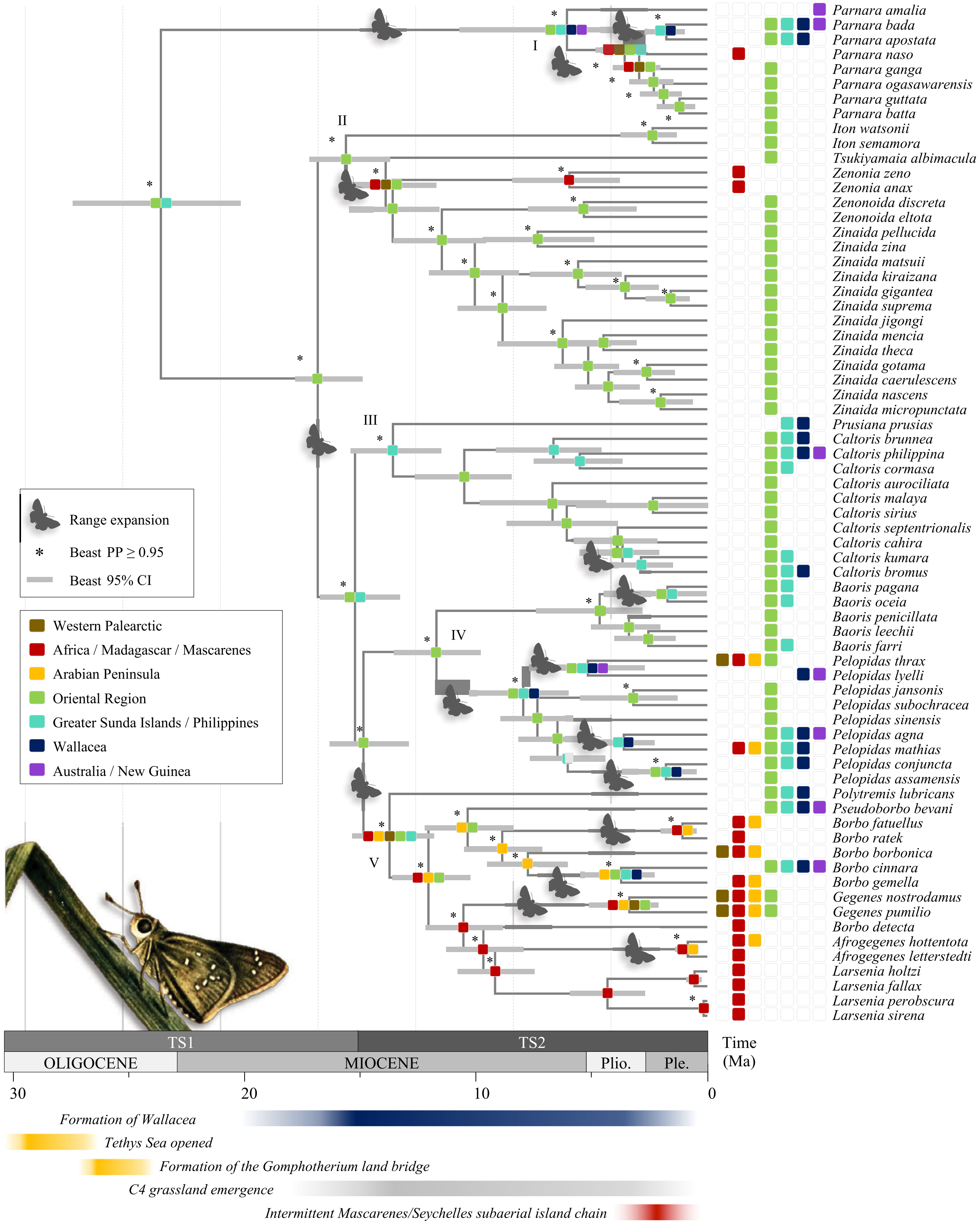

Fig. 3. Legend on next page. [Colour figure can be viewed at wileyonlinelibrary.com].

west of the Oriental region, the boreotropical forests had disappeared due to a decrease in global temperature (Zachos *et al.*, 2001). It is therefore unlikely that these butterflies would have taken advantage of these ecosystems to colonize the African continent.

Biogeographic history of Baorini skippers

The first colonization event towards Africa is found in *Parnara*, where a range expansion event allowed the ancestor of *Parnara*, except *P. amalia* Semper, to disperse from a widespread range in the Palearctic, Oriental region, and IAA toward eastern Africa. *Parnara naso* Fabricius is one of only two *Parnara* species to live in Africa, and its distribution is limited to Madagascar and the Mascarenes (Chiba & Eliot, 1991); the other African *Parnara*, *P. monasi* Trimen, is restricted to eastern coastal Africa. The current distribution of these two species is estimated to be the result of range expansion in the Western Palearctic (H_1). We argue that LDD (H_2) is a possible alternative scenario because of the peculiar distribution of *Parnara* skippers in southeastern coastal Africa. When considering LDD as an alternative hypothesis to the range expansion pattern recovered by the DEC model (Fig. 3) for the colonization of Africa in *Parnara*, it is possible that now-sunken islands between India and the Mascarenes facilitated such dispersal, acting as stepping stones during periods of low sea level (e.g. Shapiro *et al.*, 2002; Bradler *et al.*, 2015; Warren *et al.*, 2010). Bathymetric fluctuations have occurred throughout the Cenozoic, with lowstands of up to 50 m below present sea level in the last 5 Ma (Miller *et al.*, 2005). Such fluctuations coincide with colonization of Africa by the ancestor of *P. naso* (and probably *P. monasi*), possibly by using islands that were subaerial at the time as stepping stones, including, for instance, Saya de Malha (Warren *et al.*, 2010). African *Brusa* (two extant species), which were not sampled in our study, could represent an additional colonization event of Africa. Including *P. monasi* and generating more data for these two African species will be critical to understanding the underlying mechanism of colonization in *Parnara*.

We infer a Miocene range expansion event from the Oriental region to the Western Palearctic and Africa in clade II. Although a colonization route through the Palearctic region and into northern Africa is possible, it is equally likely that Baorini skippers dispersed through the Arabian Peninsula (see later). Our analysis suggests range expansion through the Arabian Peninsula and Africa in clade V. This pattern is consistent with the colonization of Africa by *Pelopidas* and is largely compatible with H_1. All of these colonization events seem to postdate the closure of the Tethys Ocean, a geological suture that was, until recently, believed to be of early Miocene age (e.g. Rögl, 1998; Bosworth *et al.*, 2005; Harzhauser *et al.*, 2007). New geological evidence based on magnetostratigraphy and strontium isotope analyses of sedimentation in the High Zagros points to a slightly older origin of this land bridge, probably in the mid-Oligocene, *c.* 27 Ma (Pirouz *et al.*, 2017). According to H_1, the closure of the eastern part of the Tethys Ocean (i.e. Neotethys) in the mid-Oligocene permitted the formation of the *Gomphotherium* land bridge, a land corridor between Africa and the Oriental region (Rögl, 1998). Studies of unrelated taxa have suggested that this passage was used as a means of dispersal. For instance, Bourguignon *et al.* (2017) hypothesized that multiple clades of termites (Isoptera) used this bridge to disperse between the two continents. They argued that because these termite clades were fungus-growers, geodispersal was more likely than LDD. Tamar *et al.* (2018) also recognized the role played by the *Gomphotherium* land bridge in allowing dispersal to and from Africa in a study on the biogeography of *Uromastyx* Merrem lizards (Squamata, Agamidae). In particular, they argued that the aridification of North Africa and the Arabian Peninsula in the early Miocene might have favoured dispersal in this clade. Miletinae butterflies are also suggested to have dispersed into Africa via this corridor in the early Miocene (Kaliszewska *et al.*, 2015). Other butterfly studies have shed light on patterns consistent with the *Gomphotherium* land bridge hypothesis (Aduse-Poku *et al.*, 2009, 2015; Sahoo *et al.*, 2018). In the case of Baorini skippers, the potential passage of the *Gomphotherium* land bridge is dated between *c.* 5 and 15 Ma in three different clades. When considering the credibility intervals associated with these estimates, the crossing of the *Gomphotherium* land bridge could be combined with more humid conditions in the Arabian Peninsula (Griffin, 2002), as well as the emergence of C_4-grassland ecosystems in the eastern Mediterranean and African regions as early as the mid-Miocene, followed by their dominance in the late Miocene (Edwards *et al.*, 2010). All Baorini skipper larvae are Poaceae-feeders and we hypothesize that the synchronous emergence of large grasslands and savannas in this key region might have facilitated dispersal towards Africa, but also fostered diversification in Africa as suggested for other Lepidoptera (Kergoat *et al.*, 2012; Toussaint *et al.*, 2012). Based on the co-occurrence of a land bridge, appropriate host plants and a humid climate on the Arabian Peninsula and North Africa in the Miocene, the disjunct distribution of Baorini skippers between Africa and Asia seems to stem principally from range expansion from the Oriental region rather than transoceanic dispersal toward Africa.

Fig. 3. Bayesian divergence time estimates and historical biogeography of Baorini skippers. Median age time tree from the BEAST analysis using two Bayesian relaxed clocks and a Yule speciation tree model; 95% credibility intervals are shown as grey bars on all nodes. Nodal support values are posterior probabilities (PP), and nodes with PP > 0.95 are labelled with an asterisk. The biogeographic estimation of ancestral ranges from the dispersal–extinction–cladogenesis analysis performed in BIOGEOBEARS is presented with the most likely state at each node. Range expansion and long-distance dispersal events are indicated following the inserted caption. The major geological events that occurred in the past 40 Ma and which are relevant to the evolution of Baorini are presented under the geological timescale. A historical drawing of *Pelopidas mathias* is presented from the book *Indian Insect Life* by Harold Maxwell-Lefroy (1909).

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Taxon sampling. Table of the specimens used in this study with taxonomic, geographic and genomic details.

Appendix S2. BIOGEOBEARS dispersal rate scaler matrices. Tables of the dispersal rate scaler matrices of the two time slices used in the BIOGEOBEARS analyses.

Appendix S3. IQ-TREE phylogenetic relationships. Complete phylogenetic tree inferred in IQ-TREE (best ML score) based on the concatenated dataset of 16 loci, including ribosomal loci recovered from GenBank.

Appendix S4. BEAST divergence times. Median age time tree from the BEAST analysis using two Bayesian relaxed clocks and a Yule speciation tree model; 95% credibility intervals are shown on all nodes. Nodes with $PP > 0.95$ are labelled with an asterisk.

Appendix S5. Results of the BIOGEOBEARS analyses. Detailed results of the BIOGEOBEARS analyses with relative probabilities of ancestral ranges and parameter estimates for dispersal and extinction.

Acknowledgements

We thank Niklas Wahlberg, Thomas Simonsen and an anonymous reviewer for insightful comments and suggestions that contributed to improving this manuscript. We thank Antonio Giudici, Geoffrey Prosser, Jee & Rani Nature Photography, Laitche, Hsu Hong Lin and Charles James Sharp, who provided photographs that were illustrated in this study. The University of Florida's HiPerGator provided computational resources and technical support. EFAT acknowledges UF Health Shands Hospital for assistance, and Marius Toussaint and Claire Toussaint for inspiring this project on the 19th of May 2018. DJL acknowledges support from DEB-1541557 and 1120380. AYK acknowledges funding from DEB-1541500 and 1541560. RV acknowledges funding for sample collecting from grant CGL2016-76322-P (AEI/FEDER, UE). The authors declare there are no conflicts of interest.

References

Aduse-Poku, K., Brattström, O., Kodandaramaiah, U., Lees, D.C., Brakefield, P.M. & Wahlberg, N. (2015) Systematics and historical biogeography of the old-world butterfly subtribe Mycalesina (Lepidoptera: Nymphalidae: Satyrinae). *BMC Evolutionary Biology*, **15**, 167.

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. *Molecular Phylogenetics and Evolution*, **53**, 463–478.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A. & Alekseyenko, A.V. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Baele, G., Li, W.L.S., Drummond, A.J., Suchard, M.A. & Lemey, P. (2013) Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, **30**, 239–243.

Berra, F. & Angiolini, L. (2014) The evolution of the Tethys region throughout the Phanerozoic: a brief tectonic reconstruction. In: *Petroleum Systems of the Tethyan Region, AAPG Memoir*, (Eds by L. Marlow, C. Kendall and L. Yose), Vol. **106**, pp. 1–27. AAPG, Tulsa, Oklahoma.

Bosworth, W., Huchon, P. & McClay, K. (2005) The Red Sea and Gulf of Aden basins. *Journal of African Earth Sciences*, **43**, 334–378.

Bourguignon, T., Lo, N., Šobotník, J. *et al.* (2017) Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. *Molecular Biology and Evolution*, **34**, 589–597.

Braby, M.F. (2000) *Butterflies of Australia: their Identification, Biology and Distribution*. CSIRO Publishing, Clayton, Australia.

Bradler, S., Cliquennois, N. & Buckley, T.R. (2015) Single origin of the Mascarene stick insects: ancient radiation on sunken islands? *BMC Evolutionary Biology*, **15**, 196.

Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L. & Kawahara, A.Y. (2018) Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Systematic Biology*, **67**, 78–93.

Chazot, N., Wahlberg, N., Freitas, A.V.L. *et al.* (2019) Priors and posteriors in Bayesian timing of divergence analyses: the age of butterflies revisited. *Systematic Biology* in press.

Chiba, H. & Eliot, J.N. (1991) A revision of the genus *Parnara* Moore (Lepidoptera, Hesperiidae), with special reference to the Asian species. *Lepidoptera Science*, **42**, 179–194.

Couvreur, T.L., Pirie, M.D., Chatrou, L.W., Saunders, R.M., Su, Y.C., Richardson, J.E. & Erkens, R.H. (2011) Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography*, **38**, 664–680.

Datta-Roy, A. & Karanth, K.P. (2009) The out-of-India hypothesis: what do molecules suggest? *Journal of Biosciences*, **34**, 687–697.

de Jong, R. & Coutsis, J.G. (2017) A re-appraisal of *Gegenes* Hübner, 1819 (Lepidoptera, Hesperiidae) based on male and female genitalia, with the description of a new genus, *Afrogegenes*. *Tijdschrift voor Entomologie*, **160**, 41–60.

de Jong, R. & Treadaway, C.G. (1993) The Hesperiidae (Lepidoptera) of The Philippines. *Zoologische Verhandelingen*, **288**, 1–115.

de Jong, R. (2016) Reconstructing a 55-million-year-old butterfly (Lepidoptera: Hesperiidae). *European Journal of Entomology*, **113**, 423–428.

de Jong, R. (2017) Fossil butterflies, calibration points and the molecular clock (Lepidoptera: Papilionoidea). *Zootaxa*, **4270**, 1–63.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edwards, E.J., Osborne, C.P., Strömberg, C.A., Smith, S.A. & C4 Grasses Consortium (2010) The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science*, **328**, 587–591.

Espeland, M., Breinholt, J., Willmott, K.R. *et al.* (2018) A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, **28**, 770–778.

Fan, X., Chiba, H., Huang, Z., Fei, W., Wang, M. & Sáfián, S. (2016) Clarification of the phylogenetic framework of the tribe Baorini (Lepidoptera: Hesperiidae: Hesperiinae) inferred from multiple gene sequences. *PLoS One*, **11**, e0156861.

Ferreira, M.A. & Suchard, M.A. (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics*, **36**, 355–368.

Ghazanfar, S.A. & Fisher, M. (1998) *Vegetation of the Arabian Peninsula*, p. 365. Kluwer Academic Publishers, Dordrecht.

Griffin, D.L. (2002) Aridity and humidity: two aspects of the late Miocene climate of North Africa and the Mediterranean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **182**, 65–91.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Hall, R. (2013) The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *Journal of Limnology*, **72**, 1.

Harzhauser, M., Kroh, A., Mandic, O., Piller, W.E., Göhlich, U., Reuter, M. & Berning, B. (2007) Biogeographic responses to geodynamics: a key study all around the Oligo–Miocene Tethyan Seaway. *Zoologischer Anzeiger-A Journal of Comparative Zoology*, **246**, 241–256.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Le, S.V. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518–522.

Hu, X., Garzanti, E., Wang, J., Huang, W., An, W. & Webb, A. (2016) The timing of India-Asia collision onset–facts, theories, controversies. *Earth-Science Reviews*, **160**, 264–299.

Huang, H. (2011) Notes on the genera *Caltoris* Swinhoe, 1893 and *Baoris* Moore, [1881] from China (Lepidoptera: Hesperiidae). *Atalanta*, **42**, 201–220.

Jiang, W., Zhu, J., Song, C., Li, X., Yang, Y. & Yu, W. (2013) Molecular phylogeny of the butterfly genus *Polytremis* (Hesperiidae, Hesperiinae, Baorini) in China. *PLoS One*, **8**, e84098.

Kaliszewska, Z.A., Lohman, D.J., Sommer, K. *et al.* (2015) When caterpillars attack: biogeography and life history evolution of the Miletinae (Lepidoptera: Lycaenidae). *Evolution*, **69**, 571–588.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kawahara, A.Y., Breinholt, J.W., Espeland, M. *et al.* (2018) Phylogenetics of moth-like butterflies (Papilionoidea: Hedylidae) based on a new 13-locus target capture probe set. *Molecular Phylogenetics and Evolution*, **127**, 600–605.

Kergoat, G.J., Prowell, D.P., Le Ru, B.P. *et al.* (2012) Disentangling dispersal, vicariance and adaptive radiation patterns: a case study using armyworms in the pest genus *Spodoptera* (Lepidoptera: Noctuidae). *Molecular Phylogenetics and Evolution*, **65**, 855–870.

Kück, P. & Meusemann, K. (2010) FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution*, **56**, 1115–1118.

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W. & Misof, B. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, **7**, 10.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Li, Y.P., Gao, K., Yuan, F., Wang, P. & Yuan, X.Q. (2017) Molecular systematics of the butterfly tribe Baorini (Lepidoptera: Hesperiidae) from China. *Journal of the Kansas Entomological Society*, **90**, 100–108.

Matzke, N.J. (2018) *BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts*. version 1.1.1. GitHub.

Miller, K.G., Kominz, M.A., Browning, J.V. *et al.* (2005) The Phanerozoic record of global sea-level change. *Science*, **310**, 1293–1298.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

Pirouz, M., Avouac, J.P., Hassanzadeh, J., Kirschvink, J.L. & Bahroudi, A. (2017) Early Neogene foreland of the Zagros, implications for the initial closure of the Neo-Tethys and kinematics of crustal shortening. *Earth and Planetary Science Letters*, **477**, 168–182.

Price, M.N., Dehal, P.S. & Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.

Ree, R.H. (2005) Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution*, **59**, 257–265.

Ree, R.H. & Sanmartín, I. (2018) Conceptual and statistical problems with the DEC+ J model of founder-event speciation and it comparison with DEC via model selection. *Journal of Biogeography*, **45**, 741–749.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Rögl, F. (1998) Palaeogeographic considerations for Mediterranean and Paratethys ceaways (Oligocene to Miocene). *Annalen des Naturhistorischen Museums in Wien*, **99**, 279–310.

Sahoo, R.K., Lohman, D.J., Wahlberg, N. *et al.* (2018) Evolution of *Hypolimnas* butterflies (Nymphalidae): out-of-Africa origin and *Wolbachia*-mediated introgression. *Molecular Phylogenetics and Evolution*, **123**, 50–58.

Sánchez-Ramírez, S., Tulloss, R.E., Amalfi, M. & Moncalvo, J.M. (2015) Palaeotropical origins, boreotropical distribution and increased rates of diversification in a clade of edible ectomycorrhizal mushrooms (*Amanita* section Caesareae). *Journal of Biogeography*, **42**, 351–363.

Seton, M., Müller, R.D., Zahirovic, S. *et al.* (2012) Global continental and ocean basin reconstructions since 200 Ma. *Earth-Science Reviews*, **113**, 212–270.

Shapiro, B., Sibthorpe, D., Rambaut, A. *et al.* (2002) Flight of the dodo. *Science*, **295**, 1683–1683.

Soubrier, J., Steel, M., Lee, M.S., Der Sarkissian, C., Guindon, S., Ho, S.Y. & Cooper, A. (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution*, **29**, 3345–3358.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J. & Rambaut, A. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4**, vey016.

Tamar, K., Metallinou, M., Wilms, T., Schmitz, A., Crochet, P.A., Geniez, P. & Carranza, S. (2018) Evolutionary history of spiny-tailed lizards (Agamidae: *Uromastyx*) from the Saharo-Arabian region. *Zoologica Scripta*, **47**, 159–173.

Tang, J., Huang, Z., Chiba, H., Han, Y., Wang, M. & Fan, X. (2017) Systematics of the genus *Zinaida* Evans, 1937 (Hesperiidae: Hesperiinae: Baorini). *PLoS One*, **12**, e0188883.

Toussaint, E.F.A., Condamine, F.L., Kergoat, G.J., Capdevielle-Dulac, C., Barbut, J., Silvain, J.F. & Le Ru, B.P. (2012) Palaeoenvironmental shifts drove the adaptive radiation of a noctuid stemborer tribe (Lepidoptera, Noctuidae, Apameini) in the Miocene. *PLoS One*, **7**, e41377.

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**, 4001.

Toussaint, E.F.A., Fikáček, M. & Short, A.E. (2016) India–Madagascar vicariance explains cascade beetle biogeography. *Biological Journal of the Linnean Society*, **118**, 982–991.

Toussaint, E.F.A. & Short, A. (2018) Transoceanic stepping–stones between cretaceous waterfalls? The enigmatic biogeography of pantropical *Oocyclus* cascade beetles. *Molecular Phylogenetics and Evolution*, **127**, 416–428.

Toussaint, E.F.A., Breinholt, J.W., Earl, C. *et al.* (2018) Anchored phylogenomics illuminates the skipper butterfly tree of life. *BMC Evolutionary Biology*, **18**, 101.

Voris, H.K. (2000) Maps of Pleistocene Sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.

Wahlberg, N. & Wheat, C.W. (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**, 231–242.

Warren, A.D., Ogawa, J.R. & Brower, A.V. (2009) Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Systematic Entomology*, **34**, 467–523.

Warren, B.H., Strasberg, D., Bruggemann, J.H., Prys-Jones, R.P. & Thébaud, C. (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics*, **26**, 526–538.

Wolfe, J.A. (1975) Some aspects of plant geography of the northern hemisphere during the late cretaceous and tertiary. *Annals of the Missouri Botanical Garden*, **62**, 264–279.

Yumul, G.P. Jr, Dimalanta, C.B., Maglambayan, V.B. & Marquez, E.J. (2008) Tectonic setting of a composite terrane: a review of the Philippine Island arc system. *Geosciences Journal*, **12**, 7–17.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–693.

Zhang, Y.L., Xue, G.X. & Yuan, F. (2010) Descriptions of the female genitalia of three species of *Caltoris* (Lepidoptera: Hesperiidae: Baorini) with a key to the species from China. *Proceedings of the Entomological Society of Washington*, **112**, 576–584.

Zhang, J., Santosh, M., Wang, X., Guo, L., Yang, X. & Zhang, B. (2012) Tectonics of the northern Himalaya since the India–Asia collision. *Gondwana Research*, **21**, 939–960.

Zhu, J.Q., Chiba, H. & Wu, L.W. (2016) *Tsukiyamaia*, a new genus of the tribe Baorini (Lepidoptera, Hesperiidae, Hesperiinae). *ZooKeys*, **555**, 37–55.

Accepted 2 April 2019