A diversification relay race from Caribbean-Mesoamerica to the Andes: Historical biogeography of *Xylophanes* hawkmoth

Xuankun Li[a, *], Chris A. Hamilton[a, b], Ryan St Laurent[a, c], Liliana Ballesteros-Mejia[d, e], Amanda Markee[a], Jean Haxaire[d], Rodolphe Rougerie[d], Ian J. Kitching[f], Akito Y. Kawahara[a, g, h, *]

[a] *McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 U.S.A.*
[b] *Department of Entomology, Plant Pathology & Nematology, University of Idaho, Moscow, ID, 83844 U.S.A.*
[c] *Smithsonian National Museum of Natural History, Department of Entomology, Washington, D.C. 20560 U.S.A.*
[d] *Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France*
[e] *CESAB, Centre de Synthèse et d'Analyse sur la Biodiversité, Montpellier, France*
[f] *Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.*
[g] *Department of Biology, University of Florida, Gainesville, FL 32611 U.S.A.*
[h] *Entomology and Nematology Department, University of Florida, Gainesville, FL 32608 U.S.A.*

Corresponding authors: XL: Xuankun.li@floridamuseum.ufl.edu; AYK: kawahara@flmnh.ufl.edu

Running title

Biogeography of *Xylophanes*

Abstract

The regions of the Andes and Caribbean-Mesoamerica are both hypothesized to be the cradle for many Neotropical lineages, but few studies have fully investigated the dynamics and interactions between Neotropical bioregions. The New World hawkmoth genus *Xylophanes* is the most taxonomically diverse genus in the Sphingidae, and the highest endemism and richness in the

Andes and Caribbean-Mesoamerica. We integrated phylogenomic and DNA barcode data and generated the first time-calibrated tree for this genus, covering 93.8% of the species diversity. We used event-based likelihood ancestral area estimation and biogeographical stochastic mapping to examine the speciation and dispersal dynamics of *Xylophanes* across bioregions. We also used trait-dependent diversification models to compare speciation and extinction rates of lineages associated with different bioregions. Our results indicate that *Xylophanes* originated in Caribbean-Mesoamerica in the late Miocene, and immediately diverged into five major clades. The current species diversity and distribution of *Xylophanes* can be explained by two consecutive phases. In the first phase, the highest *Xylophanes* speciation and emigration rates occurred in the Caribbean-Mesoamerica, and the highest immigration rates occurred in the Andes, whereas in the second phase the highest immigration rates were found in Amazonia, and the Andes had the highest speciation and emigration rates.

Keywords: Biogeography, DNA barcode, Neotropical, phylogenomic, Sphingidae

1. Introduction

The Neotropics is one of the most species-rich regions on Earth [1]. Biodiversity studies in the Neotropics have hypothesized both the Andes and Caribbean-Mesoamerica as cradle(s) for Neotropical lineages, but only a few studies have investigated the dynamics and interactions between bioregions [1,2]. The uplift of the Andes was a major event in the geological history of South America, providing barriers and opportunities for allopatric speciation and new ecological conditions for adaption and ecological speciation of animals (e.g. [2–4]) and flora (e.g. [5,6]). The Caribbean-Mesoamerican region has over 700 islands and a land bridge connecting two major continents [7]. Studies have identified this region as important for *in situ* speciation (e.g. [8,9]) with lineages originating in Caribbean-Mesoamerica and dispersing to South America (e.g. [7,10–12]). There are five biogeographic processes that describe Neotropical biodiversity: *cradle* (high speciation rate), *museum* (low extinction rate), *time-for-speciation* (early colonization), *sink* (as 'species-attractor', high immigration rate) [13], and *source* (high emigration rate) [1]. These processes play different but not necessarily mutually exclusive roles in shaping current biodiversity patterns.

With more than 120 described species, *Xylophanes* is the most species rich genus of hawkmoths [14,15]. The taxonomy has been carefully revised with the combination of

morphology and DNA barcode data (e.g. [16,17]). These moths are pollinators and strong fliers and thought to have high dispersal ability [18]. *Xylophanes* belongs to the mostly Old World subtribe Choerocampina [19,20], suggesting that *Xylophanes* may have dispersed to the New World via a jump dispersal event [19], or that the Choerocampina was already widely distributed globally prior to the origin of *Xylophanes* and the majority of the New World representatives then went extinct during the Neogene cooling [21].

Xylophanes is an ideal candidate to study dynamics and interactions between Neotropical bioregions because of the extraordinarily high species-level diversity for a hawkmoth lineage, the wide distribution of the genus covering all Neotropical bioregions, the relatively restricted geographical ranges of some individual species, and the excellent availability of DNA barcodes. We integrated phylogenomic and DNA barcode data to generate the first time-calibrated tree for *Xylophanes*, covering 93.8% of the species diversity. We used trait-dependent diversification models to compare speciation and extinction rates of lineages associated with different bioregions. We also used event-based likelihood ancestral area estimation, and biogeographical stochastic mapping to examine the speciation and dispersal dynamics among bioregions to provide insights into the evolutionary mechanisms underlying the diversification of *Xylophanes* in the Neotropics.

2. Materials and methods

(a) Taxon sampling

In the present study, 150 taxa were selected for phylogenetic analysis, including (i) 136 operational taxonomic units (OTUs), comprising both described species and subspecies, and the Barcode Index Numbers [22] that we consider as representing currently unrecognized (sub)species of *Xylophanes*; and (ii) 14 non-*Xylophanes* outgroups (Appendix S1). Anchored Hybrid Enrichment (AHE) data were newly generated for 57 taxa (53 ingroup and four outgroup) using the BOM1 Agilent Custom SureSelect probe set [23]. Cytochrome c oxidase subunit I (CO1) barcode data of all ingroup taxa were added to our phylogeny, including 28 sequences generated here and 109 generated as part of a global DNA barcoding campaign for Sphingidae (Appendix S1); all DNA barcodes are publicly available from BOLD dataset DS-XYLOPHY1 (DOI:XXXX). Specimens were identified by co-authors AYK, IJK, JH, and, RR using morphology and CO1 barcodes. Ingroups sampled represent 136 of the 145 known *Xylophanes*

OTUs and 93.8% of species diversity. Details of the sampled taxa can be found in Appendices S1, S2.

(b) DNA extraction and the BOM1 anchored hybrid enrichment probe set

DNA extractions for 63 taxa were conducted for the present study. These extractions were used for AHE sequencing, following the methods outlined in [23], and targeting 571 loci across Bombycoidea, including the CO1 barcode. AHE, library preparation, hybrid enrichment, and sequencing were carried out at RAPiD Genomics (Gainesville, FL, U.S.A.). CO1 barcodes of seven samples were amplified by PCR using the LCO/HCO universal insect primers [24] and NEB Long Taq DNA polymerase (New England BioLabs, Ipswich, Suffolk, UK) at the McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History (MGCL) (Appendix S1), and sequenced at Eurofins Genomics (Louisville, KY, U.S.A.). All DNA extracts are stored at −80°C in the molecular collection of the MGCL, in Gainesville, FL, U.S.A.

(c) Data set preparation

For BOM1 'probe' regions with phylogenomic data (68 species: 10 transcriptomes, and 58 AHE sequences), we followed the assembly steps outlined in [25]. Raw reads were assembled with Trim Galore! v.0.4.0 (bioinformatics.babraham.ac.uk). Orthology was determined using the *Bombyx mori* genome [26] as reference using with NCBI blastn [27]. Cross-contamination checks were conducted with USEARCH [28]. Cleaned sequences were aligned in MAFFT v. 7.245 [29], and isoform consensus sequences were generated using FASconCAT-G 1.02 [30]. We used a long branch detection protocol to investigate possible non-orthologous sequences followed [31] (details see Appendix S2). Individual locus information was summarized using AMAS [32] and loci with < 60% taxon coverage (41 taxa) were excluded. In total, 482 loci were assembled across 68 taxa.

New DNA barcode sequences were either captured with the AHE probe set (in 21 taxa), or Sanger sequenced (in 7 taxa); their identification was verified using BOLD Identification Engine to rule out contamination issues or misidentification. All these DNA barcode sequences were checked for 5' to 3' direction, then aligned in MAFFT v. 7.245 [29] together with the 109 sequences downloaded from BOLD. All nucleotide sequences were then manually examined in AliView [33] to ensure correct reading frame and corresponding amino acid alignments. Cleaned

MSAs of each locus were concatenated using Phyx v. 1.1 [34] to generate a matrix with 150 taxa and 483 loci (122,100 aligned nucleotides).

(d) Phylogenetic analysis

Phylogenetic analyses using an ML approach with 60 separate heuristic searches were carried out in IQ-TREE v. 2.1.2 [35]. The matrix was partitioned by locus, and the best partitioning scheme determined by allowing merging of partitions ('-MFP+MERGE' command) and utilizing the Bayesian Information Criterion (BIC). Details on parameter settings see Appendix S2. Node supports were computed via 1000 ultrafast bootstrap ('-B 1000' command) replicates [36,37], and SH-aLRT ('-alrt 1000' command) [38]. We refer to support as "strong" if SH-aLRT ≥ 80 and UFBoot ≥ 95, and "moderate" if SH-aLRT ≥ 80 or UFBoot ≥ 95, following [36].

(e) Divergence Time Estimation

Divergence time estimation was implemented in a Bayesian framework using BEAST v1.10.4 [39]. We used SortaDate [40] to reduce the nucleotide alignment to a computationally tractable matrix (50 loci), and used PartitionFinder2 [41] to partition the reduced matrix (details see Appendix S2). The reduced concatenated data matrix was imported into BEAUTi (BEAST package). Substitution models were unlinked among partitions, and clock and tree models were linked. We applied both an uncorrelated relaxed molecular clock model [42] and an exponential prior. We also tested two different tree priors, Yule (pure birth) and birth-death for each partition schemes. We used a fixed cladogram based on the best topology selected from the previous IQ-TREE analyses. Nine nodes were constrained with uniform distributions based on the 95% confidence interval (CI) given in [43]. Details on calibration nodes selected is provided in Appendix S3.

Three independent runs of each clock scheme and tree prior combination were run to check for convergence. The subsampled trees were used to summarize the maximum clade credibility tree by TreeAnnotator [44], with median heights as node heights. In order to identify the best tree prior and clock scheme combination, path sampling and stepping stone sampling [45–47] were performed as part of all BEAST analyses. Details on parameter settings see Appendix S2.

(f) Ancestral area estimation

Distribution areas for each *Xylophanes* species (see Appendix S1) for the BioGeoBEARS analyses were assessed by IJK and RR (see Appendix S2). We recognize six bioregions that best account for the distribution of the species within the genus: (A) North America, (B) Caribbean-Mesoamerica, (C) Amazonia, (D) Dry diagonal, (E) Andes, and (F) Atlantic Forest (Figure 1a), which largely follows widely accepted scheme of [48].

We performed an event-based likelihood ancestral area estimation using BioGeoBEARS [49]. Three models were used: (1) DEC (Dispersal Extinction Cladogenesis; [50]); (2) DIVALIKE (a likelihood-based implementation of dispersal vicariance analysis, originally parsimony based; [51]); and (3) BAYAREALIKE (a likelihood implementation of BayArea, originally Bayesian; [52]). All models were also evaluated under a constrained analysis (M1), in which we considered palaeogeographical events that occurred in the past 11 Myr over two time slices (11 – 7 Ma and 7 Ma to present; formation of the Panama Isthmus, Acre system, and the orogeny of Andes) and geographical distance variation (Table 1B, Figure 2), for a total of six scenarios. The Akaike Information Criterion (AIC, [53]) and the corrected Akaike Information Criterion (AICc, [54]) were calculated. The chronogram from BEAST was used for this analysis after exclusion of outgroup taxa and of two *Xylophanes* OTUs lacking distribution data (*X. hannemanni* sp1 and *X. hannemanni* sp2, Appendix S1), thus leaving 134 ingroup taxa.

(g) Dispersal and speciation rates through time

To account for missing taxa and uncertainties related to divergence time estimation and ancestral areas we used simulated trees and carried out 100 biogeographical stochastic mappings (BSMs; [55]) for each of the new trees. In all, 100,000 pseudoreplicated biogeographical histories were simulated to estimate the number of dispersal events and *in situ* speciation events (details see Appendix S2). We used the DEC model implemented in BioGeoBEARS [49] to infer geographic range evolution of lineages and performed the analysis without any constraints to decrease artificial influence. We followed [56] to calculate *in situ* speciation rates as $\lambda_X(t1) = s_X(t1)/L_X(t0)$. We followed [1] to calculate the colonization rates as $c_{XtoY}(t1) = d_{XtoY}(t1)/Br(t1)$. In addition, we calculated emigration rates as $E_X(t1) = df_X(t1)/Br(t1)$ and immigration rates as $I_X(t1) = dt_X(t1)/Br(t1)$ (details see Appendix S2).

(h) State-dependent speciation and extinction

We applied 33 GeoHiSSE models [57] in the R package HiSSE [58] to study the effects of distribution on *Xylophanes* diversity. Four regions with high *Xylophanes* diversity (Amazonia, Andes, Atlantic Forest, and Caribbean-Mesoamerica) were tested. Thirty-three models (adopted and modified from [57]; Appendix S4) were fitted, with the pruned chronogram and distribution characters modified from the ancestral area estimation analyses (Appendix S1). Species were coded as endemic to the target region (state 1) or absent from the target region (state 2) or distributed in the target region and other regions (state 0). Details on sampling fractions for each bioregion see Appendix S2. Finally, model-averaged diversification rates were mapped based on Akaike weights [57]. All phylogenetic and biogeographic analyses were conducted on the University of Florida HiPerGator High Performance Computing Cluster (http://www.hpc.ufl.edu/).

3. Results

The best partitioning schemes combined loci into 24 partitions (Appendix S5). *Xylophanes* was recovered as monophyletic (SH-aLRT/UFBoot: 100/99) and five major subclades were identified (Figure 1b). Support for 36.6% and 38.6% of nodes were strong or moderate respectively. Best partitioning schemes for three different initial partition strategies for the BEAST analyses are listed in Appendix S5. The preferred BEAST analysis was identified with marginal likelihood estimations (Table 2), which supported a birth-death tree prior with 11 unlinked molecular clocks. Divergence time estimation results reveal an origin of *Xylophanes* in the late Miocene at 8.6 Ma (95% highest posterior density = 7.8 – 9.5 Ma), immediately followed by the divergence of the five major clades around 7.7 Ma (7.0 – 8.6 Ma) (Appendix S6).

The biogeographic model DEC, with the M1 constrained analysis, yielded the highest likelihood among all six models tested (Table 3, Figure 1b). This model recovered a Caribbean-Mesoamerica origin for the genus and for three of the five major clades, and Caribbean-Mesoamerica had the longest time for speciation (supporting the time-for-speciation hypothesis).

The simulation analysis recovered a pattern that is strongly consistent with the results of the ancestral area estimation analysis (Figure 2, Appendix S7). Caribbean-Mesoamerica is the largest source for species dispersal to all other areas, followed by the Andes and Amazonia. Dispersal to the Andes from Caribbean-Mesoamerica occurred on average 18.3 times, the highest

of all dispersal types. Amazonia is the largest sink followed by the Andes and the Atlantic Forest, and the Andes is the largest source for Amazonia with an average of 15.5 dispersal events (Figure 2, Appendix S7).

Two consecutive phases are identified from the results. In the first phase (9.5 to ~2 Ma), dispersal from Caribbean-Mesoamerica was the highest, and the Andes was the greatest sink (Figures 3, 4). The *in situ* speciation rate in Caribbean-Mesoamerica was also the highest among all areas, although the rate experienced a sharp decrease until 5 Ma (Figure 4). During the second phase (~2 Ma to present), emigration and *in situ* speciation rates of the Andes exceeded Caribbean-Mesoamerica and became the highest (Figures 3, 5) while Amazonia simultaneously became the greatest sink (Figure 3). Diversification and immigration/emigration patterns of *Xylophanes* followed a relay race-like pattern such that these important evolutionary dynamics shifted geospatially and temporally. A relay race of diversification and emigration was from Caribbean-Mesoamerica to the Andes, and a relay race of immigration was from the Andes to Amazonia.

GeoHiSSE models with the highest likelihood for Caribbean-Mesoamerica and the Andes had two rate classes, with or without extinction (Appendix S8). Lineages endemic to Caribbean-Mesoamerica had low relative speciation rate (0.13) and a low extinction rate ($2.51e^{-8}$), resulting in a low net-diversification rate (0.13) (Appendix S9). In contrast, lineages endemic to the Andes had a high relative speciation rate (0.56) and a high relative extinction rate (0.06), resulting in a relatively high net-diversification rate (0.5) (Appendix S9). Slightly lower net-diversification rates are found in endemic Amazonia and Atlantic Forest lineages (Appendix S10).

4. Discussion

Our study is the first comprehensive phylogenetic analysis of *Xylophanes* hawkmoths and presents the most likely biogeographic scenario for their diversification. We recovered all 14 previously recognized species groups as monophyletic, among which 12 were strongly or moderately supported (Figure 1b, Appendix S6). In trait-dependent diversification models, constrained analyses (M1) always yielded higher likelihoods than unconstrained analyses (M0) (Table 3), indicating that the dispersal was not hampered by the still open Panama Isthmus but facilitated by the developing northern Andes, while the presence of the Acre system decreased

the dispersal rate. Our historical biogeography reconstructions reveal that *Xylophanes* originated in Caribbean-Mesoamerica in the late Miocene, setting the stage for a relay race of temporally and spatially shifting diversification patterns that occurred over the next eight million years. The "baton" of high diversification and emigration rates in Caribbean-Mesoamerica and the Andes which led to successively high immigration rates in the Andes and Amazonia, helps to explain the high diversity of *Xylophanes* in each of these major biogeographical regions of the Tropical Americas. Overall, we show that the high diversity of *Xylophanes* in tropical regions does not fall into a single, simple category of cradle, museum, time-for-speciation, sink, or source over time, but that these vagile moths, which evolved during complex geological processes, likewise diversified and dispersed dynamically.

A Caribbean-Mesoamerica origin is unusual among Neotropical Lepidoptera, as previous research generally recover origins in either the historically stable Amazonia or the dynamic orogeny of the Andes (e.g. [2,59] but see [10,60,61]). Our results show that Caribbean-Mesoamerica is the largest source (Figure 2, Appendix S7), and the cradle (during the first phase) for *Xylophanes* (Figure 4). Similar scenarios have also been found in other Neotropical insect groups (e.g. [9,11,60]). During a second phase, a low speciation rate is found in Caribbean-Mesoamerica lineages (Figure 4), and our GeoHiSSE tests support the "museum" hypothesis of Caribbean-Mesoamerica for *Xylophanes* (Appendix S9). In contrast, a study of the Neotropical butterfly tribe Brassolini (2) found a high and increasing speciation rate in Caribbean-Mesoamerica lineages. This difference between two lepidopteran groups maybe due to Caribbean-Mesoamerica being an unstable environment for Brassolini, but a stable environment for *Xylophanes*. Because 1) *Xylophanes* is much younger than Brassolini (8.6 vs. 38 Ma) and experienced less geographical dynamism within Caribbean-Mesoamerican; 2) *Xylophanes* are thought to be more vagile than most butterflies [18], which resulted in fewer instances of isolation for ancient *Xylophanes* in Caribbean-Mesoamerica compared to brassolines butterflies. For *Xylophanes*, Caribbean-Mesoamerican played as a stable environment during the second phase, therefore, low speciation and extinction rates are expected [62,63]. Our results support the idea that an area considered variable for some species might be seen as stable for others [63]. The dynamics of dispersal rate from Caribbean-Mesoamerica to the Andes is coincident with the intense orogeny of the Andes during the Miocene-Pliocene and early Quaternary, when new ecological niches arose and facilitated colonization and diversification [13,64].

State-dependent speciation and extinction analyses identified the Andes as a cradle of *Xylophanes* diversity, but differences in speciation/extinction rates were driven by an unmeasured, hidden state. This result is unsurprising as ecological speciation has been shown to play an important role in insect diversification (e.g. [61,65]). The Andes is the second-largest source of *Xylophanes*, with most immigrations taking place from Caribbean-Mesoamerica during the Pliocene (Figure 2, Appendix S7). Similar high immigration rates in Andean lineages have been found in other study systems, such as in birds and butterflies [59,66,67]. The *in situ* speciation rate of the Andes also surpassed Caribbean-Mesoamerica in the mid to late Quaternary (Figures 4, 5). The Andes orogeny has been shown to increase both the immigration rate and the *in situ* speciation rate of other lineages [13,64]. As the source in mid to late Quaternary, most of the emigrations from the Andes dispersed to Amazonia (Figures 2, 3, Appendix S7). Our study indicates that Amazonia is the largest sink and its high diversity was likely driven by immigration rather than *in situ* speciation. Amazonia as the largest sink has not been formally reported so far, although several studies have detected dispersal from the Andes to Amazonia (e.g. [59,64,66,67]). Our study offers a new perspective of Lepidoptera evolution in the Americas, where an incredibly diverse, widespread lineage of moths has undergone discontinuous, yet connected, periods of evolutionary dynamism in geographically separate regions with distinct topographic histories. The complex pattern of *in situ* diversification, with bouts of significant emigration and immigration events to colonize vast new areas of concurrently evolving landscapes, paints a vibrant picture of relatively recent events that shaped the largest radiation of hawkmoths on the planet. Furthermore, our results provide the foundation to understand the Neotropical component of the evolution of the mostly Old World distributed Choerocampina subtribe, paving the way for future research that could uncover evolutionary parallels in the Old World by studying moths ecologically similar to *Xylophanes*, but which underwent wholly different biogeographic dynamics.

References

1. Antonelli A, Zizka A, Carvalho FA, Scharn R, Bacon CD, Silvestro D, Condamine FL. 2018 Amazonia is the primary source of Neotropical biodiversity. *Proc. Natl Acad. Sci. USA* **115**, 6034–6039. (https://doi.org/10.1073/pnas.1713819115)

2. Matos-Maraví P, Wahlberg N, Freitas AVL, Devries P, Antonelli A, Penz CM. 2021 Mesoamerica is a cradle and the Atlantic Forest is a museum of Neotropical butterfly diversity: insights from the evolution and biogeography of Brassolini (Lepidoptera: Nymphalidae). *Biol. J. Linn Soc.* **133**, 704–724. (https://doi.org/10.1093/biolinnean/blab034)

3. Alfaro JWL, Cortés-Ortiz L, Di Fiore A, Boubli JP. 2015 Comparative biogeography of Neotropical primates. *Mol Phylogenet Evol.* **82**, 518–529. (https://doi.org/10.1016/j.ympev.2014.09.027)

4. Rojas D, Warsi OM, Dávalos LM. 2016 Bats (Chiroptera: Noctilionoidea) challenge a recent origin of extant neotropical diversity. *Syst. Biol.* **65**, 432–448. (https://doi.org/10.1093/sysbio/syw011)

5. Antonelli A, Sanmartín I. 2011 Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. *Syst. Biol.* **60**, 596–615. (https://doi.org/10.1093/sysbio/syr062)

6. Givnish TJ et al. 2014 Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol Phylogenet Evol.* **71**, 55–78. (https://doi.org/10.1016/j.ympev.2013.10.010)

7.   Crews SC, Esposito LA. 2020 Towards a synthesis of the Caribbean biogeography of terrestrial arthropods. *BMC Evol Biol.* **20**, 1–27. (https://doi.org/10.1186/s12862-019-1576-z)

8.   Rovito SM, Wake DB, Papenfuss TJ, Parra-Olea G, Muñoz-Alonso A, Vásquez-Almazán CR. 2012 Species formation and geographical range evolution in a genus of Central American cloud forest salamanders (Dendrotriton). *J. Biogeogr.* **39**, 1251–1265. (https://doi.org/10.1111/j.1365-2699.2012.02696.x)

9.   Beza-Beza CF, Jiménez-Ferbans L, McKenna DD. 2021 Historical biogeography of New World passalid beetles (Coleoptera, Passalidae) reveals Mesoamerican tropical forests as a centre of origin and taxonomic diversification. *J. Biogeogr.* **48**, 2037–2052. (https://doi.org/10.1111/jbi.14134)

10.   Matos-Maraví P, Núñez Águila R, Peña C, Miller JY, Sourakov A, Wahlberg N. 2014 Causes of endemic radiation in the Caribbean: Evidence from the historical biogeography and diversification of the butterfly genus *Calisto* (Nymphalidae: Satyrinae: Satyrini). *BMC Evol Biol.* **14**, 199. (https://doi.org/10.1186/s12862-014-0199-7)

11.   Zhang G, Basharat U, Matzke N, Franz NM. 2017 Model selection in statistical historical biogeography of Neotropical insects—The *Exophthalmus* genus complex (Curculionidae: Entiminae). *Mol Phylogenet Evol.* **109**, 226–39. (https://doi.org/10.1016/j.ympev.2016.12.039)

12.   Weir JT, Bermingham E, Miller MJ, Klicka J, González MA. 2008 Phylogeography of a morphologically diverse Neotropical montane species, the Common Bush-Tanager (*Chlorospingus ophthalmicus*). *Mol Phylogenet Evol.* **47**:650–64. (https://doi.org/10.1016/j.ympev.2008.02.004)

13.   Chazot N et al. 2016 Into the Andes: multiple independent colonizations drive montane diversity in the Neotropical clearwing butterflies Godyridina. *Mol Ecol.* **25**, 5765–5784. (https://doi.org/10.1111/mec.13773)

14.   Kitching IJ, Rougerie R, Zwick A, Hamilton CA, St Laurent RA, Naumann S, Mejia LB, Kawahara AY. 2018 A global checklist of the Bombycoidea (Insecta: Lepidoptera). *Biodivers Data J.* **6**, e22236. (https://doi.org/10.3897/BDJ.6.e22236)

15.   Correa-Carmona Y, Giusti A, Haxaire J, Rougerie R, Kitching IJ. 2021 Three new species of the *Xylophanes crotonis* species-group (Lepidoptera: Sphingidae) from Colombia and a

neotype designation for *Xylophanes aristor*. *Eur J Entomol.* **118**, 64–81. (https://doi.org/10.14411/eje.2021.009)

16. Vaglia T, Haxaire J, Kitching I a NJ, Meusnier I, Rougerie R. 2008 The *Xylophanes neoptolemus* and *loelia* species-groups (Lepidoptera: Sphingidae). *Zootaxa* **36**, 18–36.

17. Haxaire J, Mielke CGC. 2020 A revised and annotated checklist of the Brazilian Sphingidae with new records, taxonomical notes, and description of one new species (Lepidoptera Sphingidae). *Eur Entomol.* **11**, 101–187.

18. Beck J, Kitching IJ, Linsenmair KE. 2006 Wallace's line revisited: Has vicariance or dispersal shaped the distribution of Malesian hawkmoths (Lepidoptera: Sphingidae)? *Biol J Linn Soc.* **89**, 455–468. (https://doi.org/10.1111/j.1095-8312.2006.00686.x)

19. Kawahara AY, Mignault AA, Regier JC, Kitching IJ, Mitter C. 2009 Phylogeny and biogeography of hawkmoths (Lepidoptera: Sphingidae): Evidence from five nuclear genes. *PLoS One* **4**, e5719. (https://doi.org/10.1371/journal.pone.0005719)

20. Kawahara AY, Barber JR. 2015 Tempo and mode of antibat ultrasound production and sonar jamming in the diverse hawkmoth radiation. *Proc. Natl Acad. Sci. USA* **112**, 6407–6412. (https://doi.org/10.1073/pnas.1416679112)

21. Amorim DS, Oliveira SS, Henao-Sepúlveda AC. 2018 A new species of *Eumanota* Edwards (Diptera: Mycetophilidae: Manotine) from Colombia: evidence for a pseudogondwanan pattern. *Am. Mus. Novit.* **3915**, 1–19. (https://doi.org/10.1206/3915.1)

22. Ratnasingham S, Hebert PDN. 2013 A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS One* **8**, e66213. (https://doi.org/10.1371/journal.pone.0066213)

23. Hamilton CA, St Laurent RA, Dexter K, Kitching IJ, Breinholt JW, Zwick A, Timmermans MJTN, Barber JR, Kawahara AY. 2019 Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives. *BMC Evol Biol.* **19**, 182. (https://doi.org/10.1186/s12862-019-1505-1)

24. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol.* **3**, 294–299.

25. Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. 2018 Resolving relationships among the megadiverse butterflies and moths with a novel

pipeline for anchored phylogenomics. *Syst Biol.* **67**, 78–93.
(https://doi.org/10.1093/sysbio/syx048)

26. Xia Q et al. 2004 A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**, 1937–1940. (https://doi.org/10.1126/science.1102210)

27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421. (https://doi.org/10.1186/1471-2105-10-421)

28. Edgar RC. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. (https://doi.org/10.1093/bioinformatics/btq461)

29. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* **30**, 772–780. (https://doi.org/10.1093/molbev/mst010)

30. Kück P, Longo GC. 2014 FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool.* **11**, 81. (https://doi.org/10.1186/s12983-014-0081-x)

31. Li X et al. 2021 Phylogeny of gracillariid leaf-mining moths: evolution of larval behaviour inferred from phylogenomic and Sanger data. *Cladistics* 1–24. (https://doi.org/10.1111/cla.12490)

32. Borowiec ML. 2016 AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660. (https://doi.org/10.7717/peerj.1660)

33. Larsson A. 2014 AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278. (https://doi.org/10.1093/bioinformatics/btu531)

34. Brown JW, Walker JF, Smith SA. 2017 Phyx: Phylogenetic tools for unix. *Bioinformatics* **33**, 1886–1888. (https://doi.org/10.1093/bioinformatics/btx063)

35. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020 IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol.* **37**, 1530–1534. (https://doi.org/10.1093/molbev/msaa015)

36. Minh BQ, Nguyen MAT, von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**, 1188–1195. (https://doi.org/10.1093/molbev/mst024)

37. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2:

Improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518–522. (https://doi.org/10.1093/molbev/msx281)

38. 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. *Syst Biol.* **59**, 307–321. (https://doi.org/10.1093/sysbio/syq010)

39. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016. (https://doi.org/10.1093/ve/vey016)

40. Smith SA, Brown JW, Walker JF. 2018 So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* **13**, e0197433. (https://doi.org/10.1371/journal.pone.0197433)

41. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017 Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* **34**, 772–773. (https://doi.org/10.1093/molbev/msw260)

42. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (https://doi.org/10.1371/journal.pbio.0040088)

43. Kawahara AY et al. 2019 Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl Acad. Sci. USA* **116**, 22657–22663. (https://doi.org/10.1073/pnas.1907847116)

44. Rambaut A, Drummond AJ. 2015 TreeAnnotator v1. 8.2. MCMC Output Anal.

45. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011 Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Syst Biol.* **60**, 150–160. (https://doi.org/10.1093/sysbio/syq085)

46. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko A V. 2012 Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* **29**, 2157–2167. (https://doi.org/10.1093/molbev/mss084)

47. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013 Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol.* **30**, 239–243.

(https://doi.org/10.1093/molbev/mss243)

48. Morrone JJ. 2014 Biogeographical regionalisation of the neotropical region. *Zootaxa.*
3782, 1–110. (https://doi.org/10.11646/zootaxa.3782.1.1)

49. Matzke NJ. 2014 Model selection in historical biogeography reveals that founder-event
speciation is a crucial process in island clades. *Syst Biol.* **63**, 951–970.
(https://doi.org/10.1093/sysbio/syu056)

50. Ree RH, Smith SA. 2008 Maximum likelihood inference of geographic range evolution by
dispersal, local extinction, and cladogenesis. *Syst Biol.* **57**, 4–14.
(https://doi.org/10.1080/10635150701883881)

51. Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of
historical biogeography. *Syst Biol.* **46**, 195–203. (https://doi.org/10.1093/sysbio/46.1.195)

52. Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP. 2013 Bayesian analysis of
biogeography when the number of areas is large. *Syst Biol.* **62**, 789–804.
(https://doi.org/10.1093/sysbio/syt040)

53. Burnham KP, Anderson DR. 1998 Practical use of the information-theoretic approach. In:
Model selection and inference. New York, NY. 75–117. (https://doi.org/10.1007/978-1-
4757-2917-7_3)

54. Burnham KP, Anderson DR. 2002 Model selection and multimodel inference: a practical
information-theoretic approach. 2nd ed. 485 pp.

55. Dupin J, Matzke NJ, Särkinen T, Knapp S, Olmstead RG, Bohs L, Smith SD. 2017
Bayesian estimation of the global biogeographical history of the Solanaceae. *J. Biogeogr.*
44, 887–899. (https://doi.org/10.1111/jbi.12898)

56. Xing Y, Ree RH. 2017 Uplift-driven diversification in the Hengduan Mountains, a
temperate biodiversity hotspot. *Proc. Natl Acad. Sci. USA* **114**, E3444–E3451.
(https://doi.org/10.1073/pnas.1616063114)

57. Caetano DS, O'Meara BC, Beaulieu JM. 2018 Hidden state models improve state-
dependent diversification approaches, including biogeographical models. *Evolution* **72**,
2308–2324. (https://doi.org/10.1111/evo.13602)

58. Beaulieu JM, O'Meara BC. 2016 Detecting hidden diversification shifts in models of trait-
dependent speciation and extinction. *Syst Biol.* **65**, 583–601.
(https://doi.org/10.1093/sysbio/syw022)

59. De-Silva DL, Elias M, Willmott K, Mallet J, Day JJ. 2016 Diversification of clearwing butterflies with the rise of the Andes. *J. Biogeogr.* **43**, 44–58. (https://doi.org/10.1111/jbi.12611)

60. Lewis DS, Sperling FAH, Nakahara S, Cotton AM, Kawahara AY, Condamine FL. 2015 Role of Caribbean Islands in the diversification and biogeography of Neotropical *Heraclides* swallowtails. *Cladistics* **31**, 291–314. (https://doi.org/10.1111/cla.12092)

61. Toussaint EFA, Dias FMS, Mielke OHH, Casagrande MM, Sañudo-Restrepo CP, Lam A, Morinière J, Balke M, Vila R. 2019 Flight over the Proto-Caribbean seaway: Phylogeny and macroevolution of Neotropical Anaeini leafwing butterflies. *Mol Phylogenet Evol.* **137**, 86–103. (https://doi.org/10.1016/j.ympev.2019.04.020)

62. Stebbins GL. 1974 Flowering plants. Evolution above the species level. Cambridge, MA, USA: Harvard University Press.

63. Thais V, O'Meara BC, Beaulieu JM. 2021 Retiring "cradles" and "museums" of biodiversity. *Am Nat.* (https://doi.org/10.1086/717412)

64. Hoorn C et al. 2010 Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**, 927–931. (https://doi.org/10.1126/science.1194585)

65. Matos-Maraví PF, Peña C, Willmott KR, Freitas AVL, Wahlberg N. 2013 Systematics and evolutionary history of butterflies in the "*Taygetis* clade" (Nymphalidae: Satyrinae: Euptychiina): Towards a better understanding of Neotropical biogeography. *Mol Phylogenet Evol.* **66**, 54–68. (https://doi.org/10.1016/j.ympev.2012.09.005)

66. Beckman EJ, Witt CC. 2015 Phylogeny and biogeography of the New World siskins and goldfinches: Rapid, recent diversification in the Central Andes. *Mol Phylogenet Evol.* **87**, 28–45. (https://doi.org/10.1016/j.ympev.2015.03.005)

67. Dantas SM, Weckstein JD, Bates JM, Krabbe NK, Cadena CD, Robbins MB, Valderrama E, Aleixo A. 2016 Molecular systematics of the new world screech-owls (*Megascops*: Aves, Strigidae): biogeographic and taxonomic implications. *Mol Phylogenet Evol.* **94**, 626–634. (https://doi.org/10.1016/j.ympev.2015.09.025)

Tables

Table 1. Dispersal rates matrix between each pair of biogeographical area considered and for the two time slices used in our historical biogeography analysis.

Table 2. Marginal likelihood estimate (MLE) scores for various BEAST analyses performed for this study, and estimated ages (in Ma) for *Xylophanes* crown nodes for each tree prior/clock scheme in BEAST.

Table 3. Results of the BioGeoBEARS analyses.

Figures

Figure 1. (a). Map of the America and the defined bioregions based on distribution data of *Xylophanes*. The delineated bioregions are mainly based on (48). (b). Ancestral area estimates for *Xylophanes* under the dispersal–extinction–cladogenesis model and constrained dispersal rates (DEC, M1). The estimation was performed with BioGeoBEARS, based on the chronogram generated using BEAST shown in **Appendix S4**. Scale is in Ma. Distribution of each species is mapped to the right of the chronogram. A single most probable ancestral area is mapped at each node. Maps below the scale are modified from Hoorn et al. (2010) showing palaeogeographical models of two time slices used in the constrained analysis. Left: 7–11 Ma, Panama Isthmus open, Acre system present, and northern Andes undeveloped; right: 7 Ma to the present, Panama Isthmus closed, and northern Andes developed.

Figure 2. Summary of major *Xylophanes* dispersal events, average number of dispersal events between two areas based on 100,000 biogeographical stochastic mappings under the DEC model in BioGeoBEARS. (a) The highest emigration and the highest immigration of each area are summarized on the map. The width and shape of lines represent the estimated average number of dispersal event, coloured arrow represent dispersal source area. (b) Bar chart showing average number of emigration and immigration events of each area. Complete average dispersal events between each type shown in **Appendix S5**.

Figure 3. Dispersal rates through time based on 100,000 biogeographical stochastic mappings under the DEC model in BioGeoBEARS. Rates are displayed for selected pairs of areas. Continuous lines are the median values; coloured ribbons are the lower and upper quartiles (0.25 and 0.75 quantiles).

Figure 4. Within-area dispersal and speciation rates through time of Caribbean-Mesoamerica based on 100,000 biogeographical stochastic mappings under the DEC model in BioGeoBEARS. Continuous lines are the median values; coloured ribbons are the lower and upper quartiles (0.25 and 0.75 quantiles).

Figure 5. Within-area dispersal and speciation rates through time of Amazonia, Andes, and Atlantic Forest based on 100,000 biogeographical stochastic mappings under the DEC model in BioGeoBEARS. Continuous lines are the median values; coloured ribbons are the lower and upper quartiles (0.25 and 0.75 quantiles).

Supporting Information

Appendix S1. Taxa sampled in the present study, with distribution coding results, number of loci generated, number of loci used in the phylogeny, record identifiers (SampleIDs) in dataset DS-XYLOPHY1 (DOI:XXXX) of the Barcode of Life Datasystems (www.boldsystems.org), and GenBank accession numbers for DNA barcode sequences.

Appendix S2. Detailed materials and methods.

Appendix S3. Secondary calibrations used in divergence time estimation.

Appendix S4. Set of 33 models used for GeoHiSSE analyses (modified from Caetano et al., 2018).

Appendix S5. Partitioning scheme used for phylogeny and BEAST analyses.

Appendix S6. Time-calibrated species tree of *Xylophanes* based on AHE data, using BEAST. Scale is in Ma. Bars depict the 95% highest posterior probability density of each estimate. Pale blue node diamonds on the chronogram represent minimum age constraints for those lineages. Number around nodes are SH-aLRT (left) and UFboot (right).

Appendix S7. The average number of dispersal events of each type based on 100,000 biogeographical stochastic mappings under the DEC model in BioGeoBEARS. The row headings represent source areas, and the column headings represent destination areas.

Appendix S8. GeoHiSSE models with Akaike weight higher than 0.2 in each region.

Appendix S9. Boxplots representing diversification rate distribution (from left to right: speciation rates, extinction rates and net-diversification rates) obtained from GeoHiSSE analyses with Caribbean-Mesoamerica (upper) and Andes (lower) as target regions.

Appendix S10. Boxplots representing diversification rate distribution (from left to right: speciation rates, extinction rates and net-diversification rates) obtained from GeoHiSSE analyses with Amazonia (upper) and Atlantic Forest (lower) as targeting regions.