Insect Systematics and Diversity, 2026, **10**, ixag016
DOI: https://doi.org/10.1093/isd/ixag016
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Evolution

Museomics reveals patterns of within-species divergence and diversification in *Troides* birdwings (Lepidoptera: Papilionidae)

Corné F. H. van der Linden*[,1,2], Eliette L. Reboud[3], Robin van Velzen[1], Joost van den Heuvel[4], Emmanuelle Chevalier[3], Marcel Eurlings[2], Mónica Guimarães Cruz[2], Théo Léger[5], Stephen L. Sutton[6,†], Patrick Verbaarschot[1], Rob de Vos[2], Niklas Wahlberg[7], M. Eric Schranz[1], Fabien L. Condamine[3], and Sabrina Simon[1]

[1]Biosystematics Group, Wageningen University & Research, Wageningen, The Netherlands
[2]Naturalis Biodiversity Center, Leiden, The Netherlands
[3]Institut des Sciences de l'Evolution de Montpellier, Université de Montpellier | CNRS | IRD | EPHE, Montpellier, France
[4]Laboratory of Genetics, Wageningen University & Research, Wageningen, The Netherlands
[5]Museum für Naturkunde, Leibniz-Institut fuer Evolutions- und Biodiversitaetsforschung, Berlin, Germany
[6]Institute for Tropical Biology & Conservation, University Malaysia Sabah, Kota Kinabalu, Malaysia
[7]Department of Biology, Lund University, Lund, Sweden

*Corresponding author. Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands (Email: corne@lindenature.com, corne.vanderlinden@wur.nl).

†Deceased, 31 January 2023.

Subject Editor: Dr. Manpreet Kohli

Sundaland, a subcontinent situated in the Indo-Malay Archipelago, is characterized by a dynamic climatic, geological, and geographical history and a correspondingly diverse biota. The distribution of this biota follows complex biogeographical patterns, with high levels of local endemism and frequent instances of hidden diversity—greater genetic differentiation than expected based on phenotype. Biogeographical studies on Sundaic insects are scarce, despite high species richness and complex distributions. By applying molecular "museomics" techniques to a species group of *Troides* birdwing butterflies, we generated inter- and intraspecific datasets comprising 39,485 to 59,905 SNP positions and complete mitogenomes from museum specimens collected between the 1890s and 1990s. These datasets allowed the investigation of taxonomic status, biogeography, and divergence timing within the Sundaland endemic *amphrysus* group. The origin of the group was estimated between the late Miocene and early Pliocene, and subsequent intraspecific divergence took place during the Pleistocene. Earlier-diverging taxa were found on Borneo, while later-diverging taxa had distributions centered in western Sundaland. Divergence between subspecies of later-diverging species complexes occurred <1 million years ago, overlapping with marine incursions on the Sunda Shelf. Javan populations showed comparatively high genetic differentiation from those on other Sundaic landmasses. We found genetic evidence of possible hidden species-level diversity within two species complexes. Contrastingly, *T. amphrysus euthydemus* (Fruhstorfer 1913) **syn. nov.** is synonymized with *T. amphrysus ruficollis* (Butler 1879), given the low divergence between these two species. Together, these results provide a perspective on *Troides* evolutionary dynamics and the role of Pleistocene glaciations in shaping patterns of Sundaland's insect diversity.

Keywords: biogeography, birdwings, phylogenomics, Pleistocene, Sundaland

Introduction

Southeast Asia is one of the world's megadiverse regions, hosting about a quarter of global biodiversity despite representing only 4% of the Earth's land surface (Myers et al. 2000, Mittermeier et al. 2005, Woodruff 2010, Corlett 2014). Situated within the region, the Indo-Malay Archipelago is particularly well-known for its exceptional biodiversity and hosts two global terrestrial biodiversity hotspots (Myers et al. 2000). The tropical position, complex geological history, and insularity of the Archipelago have contributed to the evolution of its megadiverse species assemblages, shaping patterns of high regional and local endemism in various clades (Woodruff 2010, Hall 2013, De Bruyn et al. 2014, Von Rintelen et al. 2017).

Received: 29 July 2025. **Revised:** 21 January 2026. **Accepted:** 1 March 2026

The western section of the Archipelago, located on the Sunda Shelf, persists as a highly fragmented assemblage of islands. However, for most of its geological history since its origin around 80 million years ago (Ma), the Sunda Shelf formed an exposed subcontinental landmass known as Sundaland (Hall and Morley 2004, Hall 2009, 2013, Sarr et al. 2019). Oscillations in global sea levels flooded (interglacial) and exposed (glacial) the Sunda Shelf during the late Pleistocene, culminating in the permanently flooded state from ~19,000 years ago to the present (Voris 2000, Cannon et al. 2009, Woodruff 2010, De Bruyn et al. 2014, Sarr et al. 2019). During glaciations, montane communities of flora and fauna were able to track the cooling climate downslope and correspondingly occupy larger ranges, with the opposite occurring during interglacial periods (Hewitt 2000, Qu et al. 2011).

These late Pleistocene climate and sea level dynamics are particularly thought to have had a major influence on diversification in Sundaland, which has been corroborated by biogeographic studies of various clades, including freshwater fishes, birds, and mammals (Mason et al. 2019, Cros et al. 2020, Husson et al. 2020, Sholihah et al. 2020). Sundaic insects show particularly high species richness and complex patterns of intraspecific diversity; however, their biogeography remains understudied. Recently diversified insect lineages with complex patterns of regional endemism are particularly suitable for investigating the role of Pleistocene dynamics on diversification in Sundaland.

Birdwing butterflies of the well-known genus *Troides* Hübner 1819 (Lepidoptera: Papilionidae) are an excellent example of such a recently diversified insect lineage. *Troides* birdwings have diversified into 21 species and 72 subspecies across their range, which straddles the Indo-Malay Archipelago from the Asian mainland to New Guinea (Tsukada and Nishiyama 1982, Matsuka 2001, Ohya 2001, Häuser et al. 2005, Condamine et al. 2015). Like other birdwing genera (*Trogonoptera* and *Ornithoptera*), *Troides* are host plant specialists feeding on species in the genera *Aristolochia* and *Thottea*, both of which belong to the birthwort family (Aristolochiaceae) (Nishida et al. 1993, Matsuka 2001). *Troides* are generally regarded as edge-tolerant canopy species and are typically strong flyers with representatives in most terrestrial habitats in Sundaland, from coastal mangroves to montane cloud forests (Haugum and Low 1979). The diversity within the genus shows significant geographical structuring and regional endemism, with many species being represented by allopatric subspecies.

Within *Troides*, four Sundaland endemic species form a clade known as the "*amphrysus*-group," these are *T. amphrysus* (Cramer 1782), *T. andromache* (Staudinger 1892), *T. cuneifera* (Oberthür 1879) and *T. miranda* (Butler 1869) (Matsuka 2001, Condamine et al. 2015). These four species show significant intraspecific morphological variation, with each species comprising several subspecific lineages, and we correspondingly refer to the four species as "species complexes" hereafter. These subspecies have mostly been described during the late 19th century, based on distinct wing color patterns in combination with allopatric distributions (Staudinger 1892, Fruhstorfer 1898). The *amphrysus*-group has 14 subspecies found on the four large Sundaland landmasses: Borneo, Sumatra, Java, and the Malay Peninsula (Fig. 1) (Tsukada and Nishiyama 1982, Matsuka 2001). Despite their popularity among naturalists over the past ~150 yr, taxonomic and systematic uncertainties persist within the *amphrysus*-group (Haugum and Low 1979, Tsukada and Nishiyama 1982). Their patchwork distribution and large morphological variation are an indication of the diversity that is found within each lineage, some of which may not be captured in current subspecies designations. Altogether, the *amphrysus*-group provides a suitable model for the study of within- and between-species diversification under the influence of late Pleistocene dynamics in Sundaland and allows the investigation of hidden diversity in an understudied biodiversity hotspot. The wealth of specimens present in natural history collections provides a superb opportunity to address such research questions.

Natural history collections harbor a vast catalog of Earth's biodiversity and are therefore key to investigating the origins of biodiversity (Bakker et al. 2020, Card et al. 2021). The genomic potential of these natural history collections is increasingly being unlocked through rapidly evolving high-throughput sequencing techniques combined with increasingly efficient bioinformatic pipelines for analyzing low-coverage fragmented genomes (Bakker et al. 2020, Call 2020, Card et al. 2021, Raxworthy and Smith 2021). Museomics, the use of high-throughput sequencing to obtain genomic data from museum specimens (Call 2020, Twort et al. 2021), can provide unprecedented genomic insights into evolutionary histories and intraspecific diversity contained in natural history collections.

In this study, we make use of *Troides* museum specimens collected over the last ~130 yrs to carry out a fine-scale biogeographical study of the clade and to genetically investigate the numerous subspecies and identify potential cryptic subspecies. With phylogenomic insights, we clarify morphological subspecies designations made by 19th-century naturalists, test for hidden diversity, and investigate the role of late Pleistocene glaciation dynamics in the diversification of Sundaland butterflies.

Materials and Methods

Taxon Sampling

Specimens were obtained from Naturalis Biodiversity Centre (RMNH, Leiden, The Netherlands) and the Museum für Naturkunde (MfN, Berlin, Germany). Multiple specimens per subspecies were included where possible to enable the detection of genetic variation. Accordingly, three to four specimens per subspecies were sampled within the *amphrysus*-group. In addition, one to two individuals were sampled for each of the selected outgroup species—*Troides hypolitus* (Cramer 1775), *Troides magellanus* (C. & R. Felder 1862), and *Troides rhadamantus* (Lucas 1835). This resulted in a total sampling of 32 specimens, covering 12 out of all 14 *amphrysus*-group subspecies occurring on the larger Sundaic landmasses (Fig. 1). Although we attempted to fully cover all *amphrysus*-group subspecies of the Greater Sunda Islands and the Malay Peninsula, we unfortunately were unable to incorporate *T. andromache nishikawai* Kobayashi, 1992, and *T. miranda hayamii* Kobayashi, 1991, both rare subspecies occurring in southern Borneo. "Minor" subspecies, with restricted ranges on small islands along the coasts of the Greater Sunda Islands, were omitted, as these subspecies are most likely derived from the subspecies present on the nearest larger landmass (similarly to mammals and birds, see eg, Mason et al. [2019], Ng et al. [2021], and Wu et al. [2022]). This is especially relevant for the numerous described subspecies of *T. amphrysus* on islands adjacent to Sumatra, eg *T. a. simeuluensis* (Ohya 1985). Specimens were identified by the main author on the basis of wing patterns and collection locality, using diagnostic characters mentioned in Haugum and Low (1979) and Tsukada and Nishiyama (1982).

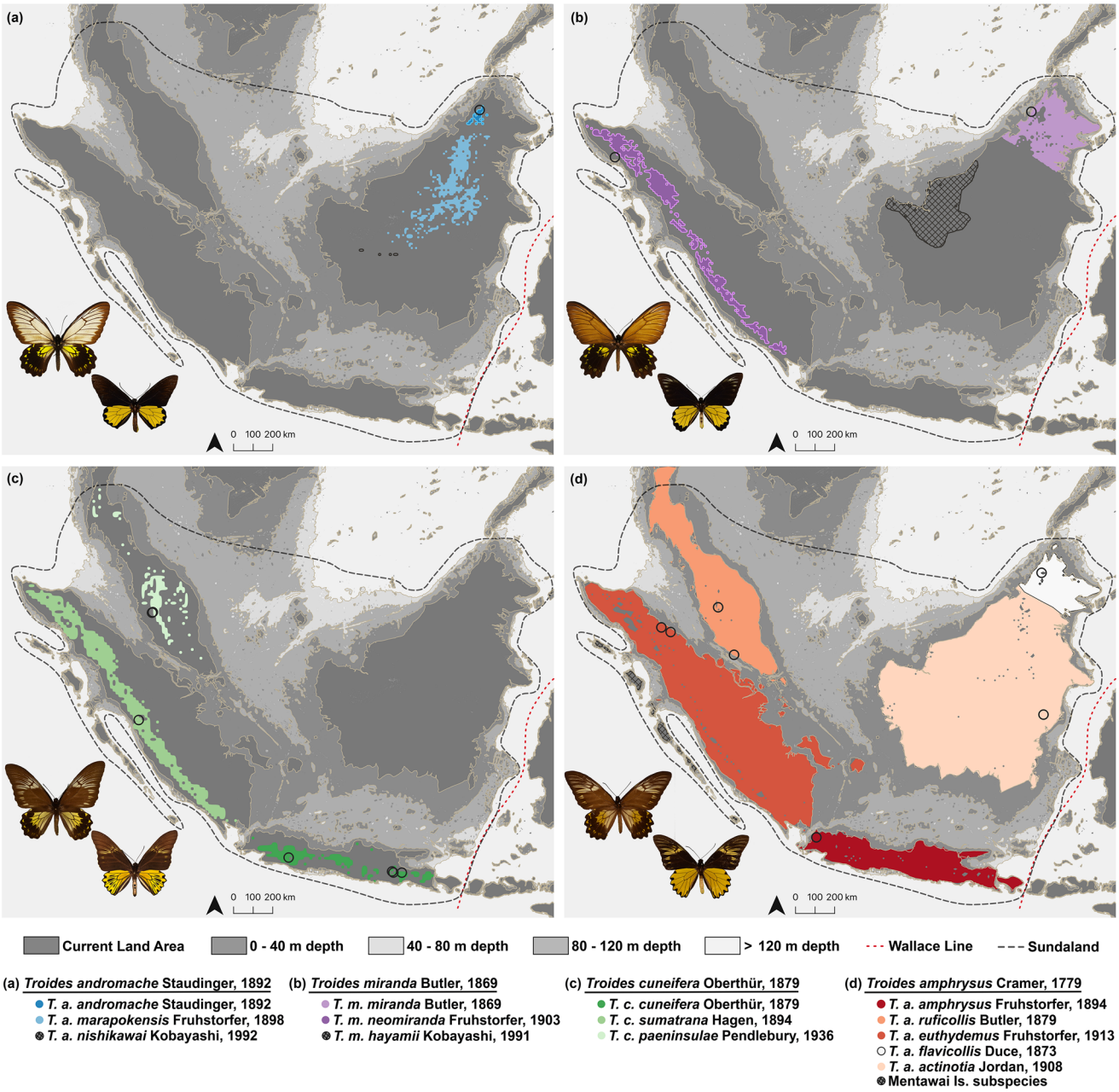

Fig. 1. Distribution maps of the different *amphrysus*-group taxa. Maps depict approximate subspecies distributions based on range limits and elevation preferences mentioned in Matsuka (2001), Haugum and Low (1979), and Tsukada and Nishiyama (1982), combined evidence was used to best approximate distribution limits. Circles indicate collection sites of samples with precise locality information (see Table 1). Bathymetry contours are shown at 40 m depth intervals, data sourced from GEBCO_2023 Grid (GEBCO Bathymetric Compilation Group 2023). For reference, approximate distributions of unsampled taxa are hashed. Males and females of nominate subspecies are shown; figures adapted from Naturalis Biodiversity Center & Natural History Museum (2014). Maps produced in QGIS 3.40.5 (QGIS.org 2025).

Sampled specimens were collected over a hundred-year interval, with the oldest specimen stemming from 1892 (*T. amphrysus flavicollis*, DNA voucher MFNLEP329)—around the time that the first subspecies in the group were being described—and the youngest specimen from 1990 (*T. cuneifera cuneifera*, specimen RMNH.INS.1047464). Different preservation methods were present within the selected taxa, although exact details are generally unclear for older specimens. Both papered specimens and pinned specimens were included in the sampled taxa. These are likely to have experienced different levels of degradation, since pinned specimens may have been relaxed prior to pinning, with this additional hydration potentially leading to additional DNA degradation, as mentioned in Mutanen et al. (2010) and Cho et al. (2016).

In addition to the specimens from museum collections, DNA sequences from whole-genome resequencing of recently collected specimens of each species in the *amphrysus*-group complex were incorporated into the dataset. These were downloaded from the publicly available sequence read archive (SRA BioProject PRJNA1013724) and belong to the following taxa: *Troides amphrysus amphrysus* (SAMN37296348), *T. cuneifera cuneifera* (SAMN37296350), *T. andromache andromache*

(SAMN37296349), and *T. miranda neomiranda* (SAMN37296351). Photographs of these four specimens were assessed and confirmed their original identifications. Sequences from four published draft genomes of outgroup taxa were also incorporated into the dataset; these were *Trogonoptera brookiana* (Wallace 1855), *Ornithoptera priamus* (Linnaeus 1758), *O. richmondia* (Gray 1853), and *Troides plateni* (Staudinger 1888) (Allio et al. 2020). A complete overview of the specimens (collection locality, collection year, and preservation method) is provided in Table 1.

DNA Extraction

DNA was extracted from a single hindleg per specimen. For specimens with missing hindlegs, mesothoracic or forelegs were sampled. Legs were crushed in Eppendorf tubes using a clean scalpel. DNA extraction was carried out using the QIAamp DNA Micro kit (Qiagen) using the Qiagen protocol "Isolation of Genomic DNA from Tissues" modified by Twort et al. (2021) for museomics. These modifications were: (i) samples were incubated in lysis buffer overnight at 56 °C, (ii) samples were briefly centrifuged, and the supernatant was subsequently carried forward, and (iii) during DNA elution, samples were incubated on the filter column in elution buffer for 20 min before centrifugation (Twort et al. 2021). Extractions were carried out in an "ancient DNA" facility (Naturalis Biodiversity Centre, Leiden), to minimize potential contamination from other DNA sources. Other measures included working inside a flow cabinet, UV-sterilization of working surfaces, and the use of filter pipette tips. DNA concentration was measured for concentration and fragment size distribution using High Sensitivity D5000 ScreenTape Assays on the Agilent 4150 TapeStation System. All samples fulfilled the fragment length distribution criteria for library preparation (<250 bp), allowing us to directly proceed with library preparation.

Library Preparations and Sequencing

Whole-genome libraries were prepared for sequencing following the protocol by Meyer and Kircher (2010) with modifications according to Twort et al. (2021). Initially, we performed blunt-end repair to ensure homogeneous samples. During blunt-end repair, samples were incubated with a master mix consisting of: Tango buffer (1X), dNTP's (100 µM), ATP (1 mM), T4 PNK (0.5 U/µL), T4 DNA Polymerase (0.1 U/µL) and MilliQ. Incubation was carried out for 15 min at 25 °C followed by 5 min at 12 °C. Following blunt-end repair, a clean-up reaction was performed using filter columns (MinElute Reaction Cleanup Kit, Qiagen), and the samples were rinsed through three steps, first using PB buffer and then twice with PE buffer. Samples were removed from the column by incubation with elutriation buffer for 5 min at 37 °C. Afterwards, adapter ligation (P7 & P5, main constituents of the Oligo Hybridization Mix) was performed using T4 DNA Ligase Buffer (1X), PEG-400 (5%), T4 DNA Ligase (5 U/µL), MilliQ, and a freshly diluted Oligo Hybridization Mix (10 pmol) and incubated for 30 min at room temperature. Clean-up reactions were again performed as described above. Overhanging fragments were then filled in by incubation with a mix of Isothermal Reaction Buffer (1X), dNTP's (250 µM), Bst Polymerase, and MilliQ at 37 °C for 20 min, this reaction was subsequently terminated by incubating for 20 min at 80 °C. For each sample, 3 µL of library template was subsequently indexed with a unique combination of dual index primers and amplified in five separate PCR reactions using the following standardized protocol. After an initial denaturation of 2 min at 95 °C, samples were amplified by 15 cycles of the following: 15 s at 95 °C, 30 s at 60 °C and one min at 68 °C. The number of cycles was adjusted based on evaluation of the initial PCR reaction, where the number of cycles was increased or decreased to achieve approximately equimolar concentrations per sample. Finally, all six reactions per sample were pooled, and 15 µL of the pooled sample was sent off for sequencing on the Illumina NovaSeq 6000 system (Novogene, UK) with paired-end sequences of 150 bp, at a depth of 1.5 to 3 Gb. This shallow depth of sequencing was selected due to financial constraints. We assume a genome size between 330 and 350 Mb, based on published *Troides* genomes: 330 Mb *Troides helena* (He et al. 2022), 348 Mb *Troides oblongomaculatus* (Reboud et al. 2023), and 351 Mb *Troides aeacus* (Hong Kong Biodiversity Genomics Consortium 2024). Based on this assumption, sequencing coverage would roughly equate to 5 to 10 X. However, since low-coverage museomics genomes are often highly fragmented, a lower actual coverage is to be expected, particularly for more degraded samples. A detailed description of the protocol published by Twort et al. (2021) can be accessed online (https://doi.org/10.6084/m9.figshare.12927500).

Processing Sequence Data

Random ligation may occur during the adapter ligation step during library preparation and subsequently result in chimaera formation (Willerslev and Cooper 2005). Although not detrimental to the reliability of sequencing information, reads must therefore be treated as single-end to obtain accurate results (Rowe et al. 2011). As in other museomic studies (Call 2020, Twort et al. 2021), our samples consisted of highly degraded genomes and we therefore took the following precautions to identify paired-end and single-end reads in subsequent analyses.

Reads were cleaned with Trimmomatic v. 0.39 (Bolger et al. 2014) using parameters that have been used on museomic samples: LEADING: 3, TRAILING: 3, remove reads below 30 bp, quality measured using a sliding window of 4 bp and quality higher than 25 (Call et al. 2021, Twort et al. 2021). Adapter sequences were removed, and overrepresented sequences were investigated for the presence of contaminants using online nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and removed accordingly, this consisted only of primer contamination. Cleaned reads were output into paired end reads and unpaired for both forward and reverse sequence files. Overlapping paired-end reads, caused by sequencing fragments smaller than the insert size, were merged using PEAR v. 0.9.6 (Zhang et al. 2014) with default settings. This yielded a set of merged sequences and unmerged forward and reverse sequences for each sample. Altogether, this yielded three different cleaned read files: (i) the unmerged forward reads concatenated with the unpaired forward reads, (ii) the unmerged reverse reads concatenated with the unpaired reverse reads, and (iii) merged overlapping paired-end reads. Mitochondrial and nuclear genomes were subsequently processed and analyzed separately.

Mitochondrial Genome Assembly and Alignment

Mitochondrial genomes (mitogenomes) were assembled de novo for each specimen using MitoZ v. 3.6 (Meng et al. 2019) in

Table 1. Specimens accessed for this study. Indicated are sample number, specimen ID, collection locality, biogeographical sub-region, collection year or approximate timespan of collection and preservation method

Group	Species	Subspecies	Sample ID	Specimen ID	Locality verbatim	Locality	Biogeographical Sub-region	Collection year	Preservation method	Collection or Study Accession
In	*Troides amphrysus* Cramer 1779	*T. a. amphrysus* Fruhstorfer 1894	S4	RMNH. INS.1047702	Djonggol, Gn. Karang, Cibodas, Kabupaten/Bogor	Mount Karang, Jonggol	Java	1977	Papered	Naturalis Biodiversity Center, Leiden
		T. a. amphrysus Fruhstorfer 1894	S5	RMNH. INS.1047593	Djonggol, Gn. Karang, Cibodas, Kabupaten/Bogor	Mount Karang, Jonggol	Java	1977	Papered	Naturalis Biodiversity Center, Leiden
		T. a. amphrysus Fruhstorfer 1894	FC313	FC313	Indonesia, Bali Island	Bali island	Bali	2004	Papered	Bioproject Accession PRJNA1013724
		T. a. ruficollis Butler 1879	S16	MFNLEP328	West Malaysia, Cameron Highland, 6-900 m	Cameron Highlands	Malay Peninsula		Spread	Museum für Naturkunde, Berlin
		T. a. ruficollis Butler 1879	S20	MFNLEP332	Malacca, Eichhorn	Malacca	Malay Peninsula		Spread	Museum für Naturkunde, Berlin
		T. a. euthydemus Fruhstorfer 1913	S19	MFNLEP331	Westküste von Sumatra	West Coast	Sumatra		Spread	Museum für Naturkunde, Berlin
		T. a. euthydemus Fruhstorfer 1913	S28	RMNH. INS.326566	Tandjong Morawa. Serdang (N.O. Sumatra).	Tanjung Morawa	Sumatra		Spread	Naturalis Biodiversity Center, Leiden
		T. a. euthydemus Fruhstorfer 1913	S30	RMNH. INS.326548	E. coast of Sumatra Laut Tador	Laut Tador	Sumatra	1950	Spread	Naturalis Biodiversity Center, Leiden
		T. a. euthydemus Fruhstorfer 1913	S40	RMNH. INS.326587	Tandjong Morawa. Serdang (N.O. Sumatra).	Tanjung Morawa	Sumatra		Spread	Naturalis Biodiversity Center, Leiden
		T. a. euthydemus Fruhstorfer 1913	S14	MFNLEP326	Westküste von Sumatra	West Coast	Sumatra		Spread	Museum für Naturkunde, Berlin
		T. a. flavicollis Duce 1873	S17	MFNLEP329	Kina Balu 12 - 1560 m, N.O. Borneo	Gunung Kinabalu	Borneo	1893	Spread	Museum für Naturkunde, Berlin
		T. a. actinotia Jordan 1908	S18	MFNLEP330	S.O.Borneo Sultanat Kutei	Kutai	Borneo		Spread	Museum für Naturkunde, Berlin
		T. a. actinotia Jordan 1908	S33	RMNH. INS.326653	Borneo S.O.	South-East Borneo	Borneo		Spread	Naturalis Biodiversity Center, Leiden
	Troides cuneifera Oberthür 1879	*T. c. cuneifera* Oberthür 1879	S12	RMNH. INS.1047464	Gn Wayang	Gunung Wayang	Java	1990	Papered	Naturalis Biodiversity Center, Leiden
		T. c. cuneifera Oberthür 1879	S2	RMNH. INS.1048612	Oro Ombo	Oro-Oro Ombo	Java	1987	Papered	Naturalis Biodiversity Center, Leiden
		T. c. cuneifera Oberthür 1879	S3	RMNH. INS.1048603	Pujon Kidul	Pujon Kidul	Java	1987	Papered	Naturalis Biodiversity Center, Leiden
		T. c. cuneifera Oberthür 1879	S38	RMNH. INS.326655	Tengger [illegible] 130 m	Tengger Mountains	Java		Spread	Naturalis Biodiversity Center, Leiden
		T. c. cuneifera Oberthür 1879	FC320	FC320	Indonesia, Java, Mt. Wayana	Mount Wayana	Java	2007	Papered	Bioproject Accession PRJNA1013724

(*Continued*)

Table 1. Continued

Group	Species	Subspecies	Sample ID	Specimen ID	Locality verbatim	Locality	Biogeographical Sub-region	Collection year	Preservation method	Collection or Study Accession
		T. c. paeninsulae Pendlebury 1936	S24	MFNLEP336	Malaysia, Tapah	Tapah	Malay Peninsula		Spread	Museum für Naturkunde, Berlin
		T. c. paeninsulae Pendlebury 1936	S13	MFNLEP325	Malaysia, Tapah	Tapah	Malay Peninsula		Spread	Museum für Naturkunde, Berlin
		T. c. sumatrana Hagen 1894	S37	RMNH. INS.326654	Solok Padang. Bov.	Solok	Sumatra	1907	Spread	Naturalis Biodiversity Center, Leiden
	Troides andromache Staudinger 1892	*T. a. andromache* Staudinger 1892	S6	MFNLEP256	Kina Balu	Mount Kinabalu	Borneo		Spread	Museum für Naturkunde, Berlin
		T. a. andromache Staudinger 1892	S7	MFNLEP257	Kina Balu Gebirge 1500 m	Mount Kinabalu	Borneo		Spread	Museum für Naturkunde, Berlin
		T. a. andromache Staudinger 1892	S9	RMNH. INS.1283435	Borneo Noord Kina Baloe Gebergte	Mount Kinabalu	Borneo		Spread	Naturalis Biodiversity Center, Leiden
		T. a. andromache Staudinger 1892	FC1047	FC1047	Indonesia, Borneo, Crocker Range, Mt. Kinabalu	Mount Kinabalu	Borneo	2002	Papered	Bioproject Accession PRJNA1013724
		T. a. marapokensis Fruhstorfer 1898	S22	MFNLEP334	Nord Borneo Marapack	Mount Merapok	Borneo		Spread	Museum für Naturkunde, Berlin
		T. a. marapokensis Fruhstorfer 1898	S8	MFNLEP258	Nord Borneo, Gebirge, Marapock	Mount Merapok	Borneo		Spread	Museum für Naturkunde, Berlin
	Troides miranda Butler 1869	*T. m. miranda* Butler 1869	S36	RMNH. INS.326717	Noord Borneo Kina Baloe	Mount Kinabalu	Borneo	1907	Spread	Naturalis Biodiversity Center, Leiden
		T. m. neomiranda Fruhstorfer 1903	FC1046	FC1046	Indonesia, Sumatra, Brastagi	Brastagi	Sumatra	1976	Spread	Bioproject Accession PRJNA1013724
		T. m. neomiranda Fruhstorfer 1903	S15	MFNLEP327	Sumatra	Sumatra	Sumatra		Spread	Museum für Naturkunde, Berlin
		T. m. neomiranda Fruhstorfer 1903	S21	MFNLEP333	Sumatra	Sumatra	Sumatra		Spread	Museum für Naturkunde, Berlin
		T. m. neomiranda Fruhstorfer 1903	S31	RMNH. INS.326722	N. W. Sumatra W. Atjeh Seumanjam Meulaboh	West Aceh, Seumanjam Meulaboh	Sumatra	1954	Spread	Naturalis Biodiversity Center, Leiden
Out	*Troides magellanus* C. & R. Felder 1862	*Troides m. magellanus* C. & R. Felder 1862	S27	RMNH. INS.327160	Bohol Bilar	Bilar	Philippines	1966	Spread	Naturalis Biodiversity Center, Leiden
		Troides m. magellanus C. & R. Felder 1862	S29	RMNH. INS.327158	Bohol Bilar	Bilar	Philippines	1966	Spread	Naturalis Biodiversity Center, Leiden
	Troides hypolitus Cramer 1775	*Troides h. hypolitus* Cramer 1775	S11	RMNH. INS.1036109	Ambon.	Ambon	Molucca		Papered	Naturalis Biodiversity Center, Leiden
	Troides rhadamantus Lucas 1838	*Troides rhada-mantus* Lucas 1838	S26	MFNLEP338	Mindoro	Mindoro	Philippines		Spread	Museum für Naturkunde, Berlin

(*Continued*)

Table 1. Continued

Group	Species	Subspecies	Sample ID	Specimen ID	Locality verbatim	Locality	Biogeographical Sub-region	Collection year	Preservation method	Collection or Study Accession
	Troides plateni Staudinger 1888	*Troides plateni* Staudinger 1888	FC2806	FC2806	Philippines, Palawan	Palawan	Philippines	2016	Papered	Bioproject Accession PRJNA533298
	Ornithoptera richmondia Gray 1852	*Ornithoptera richmondia* Gray 1852	1938	1938	Australia	Australia	Australia	1978	Papered	Bioproject Accession PRJNA533298
	Ornithoptera priamus Linnaeus 1758	*Ornithoptera priamus* Linnaeus 1758	1919	1919	Papua New Guinea	Papua New Guinea	Papua New Guinea	1987	Papered	Bioproject Accession PRJNA533298
	Trogonoptera brookiana Wallace 1855	*Trogonoptera brookiana* Wallace 1855	2023	2023	Malaysia	Malaysia	Malaysia	2003	Papered	Bioproject Accession PRJNA533298

"quick mode" using default settings with the following modifications: read length 150, insert size 150, k-mer size 59 79 99 119 141. The assembled gene sequences were aligned using the L-INS-I algorithm in MAFFT v. 7.471 (Katoh and Standley 2013). Aligned sequences were subsequently filtered for sections of random similarity or ambiguous alignments using Aliscore v. 2.0 (Kück et al. 2010), and these were subsequently removed using Alicut v. 2.3 (Kück 2016). The remaining sequences were concatenated into a partitioned matrix consisting of 13 protein-coding genes, two rRNA genes, and 22 tRNA genes using FASconCAT-G v. 1.05 (Kück and Longo 2014). This dataset will from hereon be referred to as "birdwing_mito."

Inferring Genotype Likelihoods and SNP

The cleaned reads of each sample were mapped to the chromosome-level genome assembly of *Troides aeacus* (CUHK_Taea_v1; GenBank accession no.: GCA_033220335.2), a closely related *Troides* species (Hong Kong Biodiversity Genomics Consortium 2024). Reads were mapped to the reference genome using BWA-MEM2 v. 2.2.1 (Vasimuddin et al. 2019) and sorted using SAMtools v. 1.6 (Danecek et al. 2021). Duplicate reads were marked and removed using Picard MarkDuplicates in the Picard Toolkit v. 3.0.0 (Broad Institute 2019), and the de-duplicated files were indexed using SAMtools. Reads were realigned locally around insertions and deletions using GATK v. 3.8 (McKenna et al. 2010) tools RealignerTargetCreator and IndelRealigner. To account for low base qualities due to DNA degradation, the base quality scores were rescaled using MapDamage2.0 v. 2.2.2 (Jónsson et al. 2013) using the *T. aeacus* assembly as reference.

Genotype likelihoods were imputed from the rescaled files using ANGSD 0.941 (Korneliussen et al. 2014) with the *T. aeacus* assembly as reference and the following parameters: *-minMapQ 30 -minQ 30 -remove_bads 1 -trim 0 -only_proper_pairs 0 -baq 1 -C 50 -doMaf 1 -doMajorMinor 1 -skipTriAllelic -doCounts 1 -dumpCounts 2 -doGeno 32 -GL 1 -doGlf 2 -doPost 1 -postCutoff 0.33 -snp_pval 1e-6 -minMaf 0.05 -nThreads 20*. Imputed genotype likelihoods in beagle file format were separated per chromosome. Since genotype likelihoods could not be used in all downstream analyses, SNPs were hard called when the likelihood for an allele at a position ≥0.8, following the approach of Laine et al. (2023). SNPs were called in Beagle v. 5.4 (Browning et al. 2021) using gprobs2beagle.jar and subsequently converted to variant call format using beagle2vcf.jar.

To ensure only autosomes were used in downstream analyses, the *T. aeacus* assembly was mapped to the chromosome assembly of *Papilio machaon* (ilPapMach1.1; GenBank accession no.: GCA_912999745.1) (Lohse et al. 2022) using minimap2 v. 2.28 (Li 2018) with default settings. None of the assembled chromosomes of the *T. aeacus* assembly mapped to the *P. machaon* W-chromosome, indicating the absence of this chromosome in the *T. aeacus* assembly. *T. aeacus* chromosome CM067268.1 mapped to the Z-chromosome and was accordingly omitted from downstream analyses. The remaining autosome vcf files were indexed and concatenated into one vcf using BCFtools v. 1.20 (Danecek et al. 2021).

Curation of Genotype Likelihoods and SNP Datasets

Two separate datasets were created, one containing all samples and a second including only members of the *amphrysus*-group. These two datasets were filtered interactively using the packages

SNPfiltR v. 1.0.1 (DeRaad 2022) and vcfR v. 1.15.0 (Knaus and Grünwald 2017) in R v. 4.4.2 (R Core Team 2024). Loci with a coverage greater than twice the median coverage (≥7.334) were removed as these positions are likely repeat regions from multiple sites across the genome. Subsequently, sites below a minor allele count of two were removed from the dataset to only retain sites that were present in at least two individuals. Thereafter, a strict missing data filter of 20% was applied to the remaining sites. Altogether, this ultimately yielded two sets of loci, one set of 39,485 loci across all samples and a second set of 59,905 sites across members of the *amphrysus*-group. From each of these two sets a genotype likelihood and a SNP dataset were created, ultimately yielding: (i) a 39,485 site genotype likelihoods dataset, hereafter referred to as "birdwing_GL," (ii) a 39,485 site SNP dataset, hereafter referred to as "birdwing_SNP," (iii) a 59,905 site genotype likelihoods dataset, hereafter referred to as "*amphrysus*-group_GL," and (iv) a 59,905 site SNP dataset, hereafter referred to as "*amphrysus*-group_SNP."

Genetic Structure

Mitochondrial Haplotypes

Haplotype networks were constructed to examine the geographical variation in mitochondrial divergence within each *amphrysus*-group species using nucleotide sequences of the cytochrome oxidase subunit 1 (CoI) and cytochrome b (CytB). Haplotypes of each gene were inferred for the *amphrysus*-group and a minimum spanning network was constructed for the haplotypes of each gene with the functions "haplotype" and "msn" in the R package pegas (Paradis 2010).

Nuclear Genetic Structure

The *amphrysus*-group_GL dataset into subsets of each species complex, whereafter genetic structure among the individuals in each species complex was inferred using the admixture and principal component analyses in PCAngsd v. 1.36.1 (Meisner and Albrechtsen 2018). For admixture analysis PCAngsd automatically selects the best-fit value for k to describe genetic structure components among the individuals analyzed. The resulting clustering assignment was used to visualize patterns of genetic structure in R with ggplot2 v. 3.5.1 (Wickham 2016).

Phylogenetic Reconstructions

Mitogenome Phylogeny

Mitochondrial phylogenetic relationships were inferred using the birdwing_mito dataset. The best scoring partition-specific substitution and rate models (Chernomor et al. 2016) were identified using ModelFinder in IQ-TREE v. 2.1.3 for each gene partition in the mitogenome alignment using the specifications "-AUTO" and "-m MF" (Kalyaanamoorthy et al. 2017, Minh et al. 2020). All default substitution models were evaluated: JC69 (Jukes and Cantor 1969), F81 (Felsenstein 1981), K80 (Kimura 1980), HKY (Hasegawa et al. 1985), TN (Tamura and Nei 1993), TNe, K81 (Kimura 1981), K81u, TPM2, TPM2u, TPM3, TPM3u, TIM, TIMe, TIM2, TIM2e, TIM3, TIM3e, TVM, TVMe, SYM (Zharkikh 1994), and GTR (Tavaré 1986). Base frequencies were empirically counted from the alignments (+F). Substitution models were tested with the default rate heterogeneity models (+G, +R, +I and combinations thereof: +I+G & +I+R) (Yang 1994, Gu et al. 1995, Yang 1995, Soubrier et al. 2012). Multiple FreeRate categories were also tested for by setting -cmin 4 and -cmax 15. Model quality was evaluated based on maximum likelihood (ML) and the best-fit model was selected for each partition using the corrected Akaike Information Criterion (AICc, Hurvich and Tsai [1989]). Subsequently, the partitioned alignment and the best-scoring partition models were used to initiate 50 likelihood tree-search analyses; 25 with a random starting tree and 25 guided using the maximum-parsimony tree obtained from ModelFinder. From this set the final maximum-likelihood tree was selected, and branch support was assessed using a non-parametric bootstrap analysis with 100 replicates mapped onto the ML tree.

Nuclear SNP Phylogeny

The birdwing_SNP dataset was used to infer a nuclear phylogenetic topology under the multi-species coalescent (MSC) model using SVDquartets (Chifman and Kubatko 2014) in Paup* v. 4.0a (Swofford 2002). The dataset was converted to nexus format using the python script "vcf2phylip.py" v. 2.0 (Ortiz 2019). All samples were treated as separate taxa, and *Trogonoptera brookiana* was specified as the outgroup in the analysis based on earlier phylogenetic studies (eg Condamine et al. [2015]). Quartets were exhaustively sampled and assembled using the assembly algorithm "QFM." The consensus tree was constructed using the MSC tree model specification and bootstrapped using standard bootstrapping within PAUP*.

Estimation of Divergence Times

Bayesian inferences were performed in BEAST 2.7.7 (Bouckaert et al. 2019) to estimate a temporal framework for the inter- and intraspecific relationships under the MSC model in the package snapper v1.1.4 (Stoltz et al. 2021). The XML files were prepared for BEAST using the birdwing_SNP dataset and the "*snapp_prep.rb*" ruby script (https://github.com/mmatschiner/snapp_prep), which implements the divergence time estimation method described in Stange et al. (2018). The alignment was converted to phylip format using vcf2phylip.py and automatically trimmed by snapp_prep.rb to retain only sites useable by SNAPPER—sites that have only missing data in one or more species were omitted—resulting in a subset of 3,794 bi-allelic loci suitable for divergence time estimation. Samples were assigned to taxa based on three criteria: (i) existing subspecies descriptions (distinct phenotypes with allopatric ranges—Haugum and Low (1979)), (ii) monophyly in mitochondrial and genomic SNP phylogenies, and (iii) discernible genomic clusters revealed by PCAngsd. Accordingly, the following taxa were lumped for the dating analysis: *T. andromache andromache* + *T. andromache marapokensis* (as *T. andromache s.l.*), *T. amphrysus euthydemus* + *T. amphrysus ruficollis* (as *T. amphrysus ruficollis s.l.*) and *T. amphrysus actinotia* + *T. amphrysus flavicollis* (as *T. amphrysus flavicollis s.l.*). Since there is no fossil record available for birdwings (*Trogonoptera*, *Ornithoptera*, and *Troides*), node calibrations could not be directly specified using fossil information. Therefore, we relied on divergence time estimates from a recent comprehensive phylogeny of Papilionidae containing ~400 species calibrated by three papilionid fossils from Allio et al. (2021) to set informative node age calibrations. Prior age constraints were specified by setting mean, minimal, and maximal ages to equal those from the 95% highest posterior density (HPD) obtained in Allio et al. (2021). Age calibrations were specified for three nodes, namely: (i) the crown age of birdwings between 17.5018 and 51.9644 Ma, (ii), the crown age of *Troides* and *Ornithoptera* between 13.7414 and 41.0473 Ma, and (iii) the crown age of

Troides between 9.4524 and 29.7289 Ma (Allio et al. 2021). Accordingly, the following priors were assigned to those nodes: (i) lognormal (offset 13.02, mean 17.3, standard deviation 0.605), (ii) lognormal (offset 10.5, mean 13.3, standard deviation 0.625), and (iii) lognormal (offset 6.352, mean 10.85, standard deviation 0.56). The strict clock model was used, as this is the only clock model currently available in SNAPPER. The analysis was initiated using the consensus tree from SVDquartets as the starting tree topology. We specified runs of 3,000,000 Markov chain Monte Carlo (MCMC) generations, sampling every 1,500 generations, ultimately yielding 1,800 posterior trees after removing 10% of posterior as burn-in. Two independent analyses were carried out, and we assessed the convergence of each run by evaluating the effective sample size (posterior ESS, >200) and posterior distributions using Tracer v. 1.7.2 (Rambaut et al. 2018). Posterior trees from the two separate runs were combined using LogCombiner v. 2.7.7 (Bouckaert et al. 2019), and one maximum-clade credibility tree was generated using TreeAnnotator v. 2.7.7 using common ancestor heights (Bouckaert et al. 2019).

Biogeographical Analysis

We estimated ancestral ranges on the dated maximum-clade credibility tree using the R package BioGeoBEARS 1.1.3 (Matzke 2018) to study the historical biogeography of the *amphrysus*-group. Outgroups from the genera *Trogonoptera* and *Ornithoptera* were omitted from the dataset for this analysis since they are from species groups that are incompletely represented by our taxon selection, and including incompletely sampled clades may bias the biogeographical analyses (Matzke 2018). For this reason, only the closely related outgroups within *Troides* were retained in the analysis. Interpreting the biogeographical estimations below the species level is not straightforward, and node reconstructions within species can lead to misrepresentation of biogeographical patterns, as population divergence may not follow a strict tree-like model. Furthermore, it is difficult to discern ancestral ranges based on current ranges, especially if the terrain has undergone extensive rearrangement in the intervening time interval. We therefore encoded the ranges at the level of broadly accepted biogeographical regions—Wallacea, Sundaland, the Philippines and Palawan—to infer ancestral ranges within *Troides*. The number of maximum areas a species could exist in was set to 4, which translates to a distribution in all biogeographical regions. The time-calibrated tree was trimmed to one tip per species complex and used together with the geographical states to evaluate the fit of two biogeographical models. These were the Dispersal-Extinction-Cladogenesis (DEC; [Ree and Smith 2008]) with and without the parameter + *J*, which models founder-event speciation (thought to be appropriate in an island context; Condamine et al. 2015). The best-fit model was selected by comparing AICc scores between models. All R-based analyses were performed in R v. 4.4.2 (R Core Team 2024). Subspecies range maps were drawn based on the range extent provided in Matsuka (2001) using QGIS v. 3.40.5 (QGIS.org 2025).

Species Delimitation Analyses

The birdwing_mito dataset was trimmed to include only *amphrysus*-group samples and used as input for Assemble Species by Automated Partitioning (ASAP), a single-locus distance-based delimitation algorithm (Puillandre et al. 2021). Several distance metrics were evaluated to ensure consistency of results (uncorrected p-distance, JC69, and K80 corrected distances with ts/tv 2). ASAP was accessed through its web portal: https://bioinfo.mnhn.fr/abi/public/asap/asapweb.html.

Additionally, a tree-based delimitation method, the Bayesian implementation of Poisson Tree Processes (bPTP) (Zhang et al. 2013), was carried out on the phylogeny obtained with the birdwing_mito dataset. The analysis was performed on the bPTP webserver (https://species.h-its.org/) and constrained to include only *Troides* samples. The number of MCMC generations was set to 100,000 generations, sampling to every 1,000 generations. Four independent analyses were performed, and convergence of the MCMC chains was assessed after discarding the initial 10% of the posterior as burn-in by visually examining the bPTP trace log.

To carry out species delimitation under the MSC model, we conducted a stepping stone analysis using the package Model_Selection v. 1.6.2 with the SNAPPER package in BEAST 2.7.7 (Bouckaert et al. 2019). To provide computationally feasible datasets, the *amphrysus*-group_SNP dataset was trimmed using the *snapp_prep.rb* script into two subsets of randomly selected loci, one set of 500 and another of 1,000 bi-allelic loci (only sites that are present across all taxa in the analysis were retained). Using these reduced datasets, delimitation analyses were set up with three taxonomic hypotheses: (i) the traditional taxonomy, with each species complex recognized as a species, (ii) traditional taxonomy but with *T. miranda miranda* and *T. miranda neomiranda* recognized as full species, and (iii) each genetically validated subspecies—based on the criteria used for the estimation of divergence times—as full species. The mean for lambda (λ), the yule birth rate prior, was calculated using the number of species and the variation in the alignment using the script *yule.py* v. 3 (https://github.com/joaks1/pyule) and the specification in the XML with the gamma distribution (alpha 2, beta 2.5). The mean for theta (θ), the coalescent rate prior, was selected by calculating the average divergence within species complexes and specified in the XML with the gamma distribution (alpha 2, beta 0.019). The path sampling analyses were set up using 48 steps, each with runs of 3,000,000 MCMC generations, sampling every 3,000 generations, ultimately yielding 752 likelihoods per "step" after removing 25% of posterior as burn-in. We assessed the convergence of each run by evaluating the effective sample size (ESS, >200) and posterior distributions using Tracer v. 1.7.2 (Rambaut et al. 2018). Likelihood log files were analyzed using the PathSampleAnalyser tool of the Model Selection package provided through the AppLauncher in BEAST 2.7.7 (Bouckaert et al. 2019). We then used Bayes Factor Delimitation (BFD) to evaluate the taxonomic schemes as described in Leaché et al. (2014). Where marginal likelihoods (Baele et al. 2012) of different competing taxonomies are estimated and used to rank support for each taxonomic scenario using Bayes factors (Kass and Raftery 1995).

Results

Genetic Structure

Mitochondrial Haplotypes

Mitochondrial haplotype networks (Fig. 2) illustrate genetic differences within and between species complexes. Results from both genes examined (CoI and CytB) depict similar patterns of

intraspecific variation between subspecies pairs (Fig. 2). In general, haplotypes clustered closely together based on geographical proximity and subspecies affiliation. Mitochondrial divergence between subspecies showed a high degree of variation across species complexes. Additionally, haplotype networks revealed marked differences in genetic variation within subspecies, with subspecies represented by more than one sample either sorted into multiple (eg *T. miranda neomiranda*, Fig. 2) or one single haplotype (eg *T. andromache marapokensis*, Fig. 2).

The *T. andromache* species complex, here represented by *T. andromache marapokensis* and *T. andromache andromache*, showed different mitochondrial haplotypes occurring in relatively close geographical proximity (Figs 1 and 2). The haplotypes of the two subspecies were separated by a low number of substitutions (3–4 Cytb & CoI) but did not show overlap. *T. andromache marapokensis* was represented by a single haplotype for both genes investigated. Multiple haplotypes were recovered for *T. andromache andromache*, with levels of intra-subspecies divergence comparable to its divergence with the conspecific *T. andromache marapokensis*.

Mitochondrial haplotypes of the representatives of the *T. miranda* species complex cluster into two distinct groups consisting of the *T. miranda neomiranda* (Sumatra) and *T. miranda miranda* (northern Borneo) subspecies. *T. miranda neomiranda* showed a low number of substitutions between individuals (Fig. 2; 1–5 Cytb & 2 CoI) and some shared haplotypes and a high number of substitutions with respect to *T. miranda miranda* (Fig. 2; 38 Cytb & 32 CoI). These subspecies are found in distant parts of the range (Fig. 1).

Mitochondrial haplotype networks of the species complex *T. cuneifera* reveal a high amount of differentiation between *T. cuneifera cuneifera* and the other two subspecies, *T. cuneifera sumatrana* and *T. cuneifera paeninsulae* (Fig. 2; 30 Cytb & 28 CoI). Differentiation between *T. cuneifera paeninsulae* and *T. cuneifera sumatrana* shows a lower number of substitutions for both genes investigated (Fig. 2; 5 Cytb & 6 CoI). Individuals of *T. cuneifera cuneifera* differed from each other with three substitutions for Cytb and 1–4 substitutions for CoI (Fig. 2A and B). The two individuals of *T. cuneifera paeninsulae* differed from each other with only one substitution for both genes examined (Fig. 2A and B).

T. amphrysus subspecies showed the least divergence between their haplotypes, with haplotype sharing occurring among multiple subspecies. Furthermore, the number of substitutions between haplotypes of the different subspecies remained low (Fig. 2A and B, 1–9 Cytb & 1–8 CoI), despite the large distances between the distributions of the subspecies (Fig. 1). Peninsular and Sumatran subspecies *T. amphrysus ruficollis* and *T. amphrysus euthydemus* showed a high degree of overlap and shared haplotypes for both genes investigated. Some individuals of *T. amphrysus euthydemus* differed with one substitution (Fig. 2A and B). In the CoI haplotype network, *T. amphrysus amphrysus* also shared this haplotype, while it received its own, closely related haplotype in the CytB network (Fig. 2A and B). Bornean subspecies *T. amphrysus flavicollis* and *T. amphrysus actinotia* had unique haplotypes and were more closely affiliated with the Javan *T. amphrysus amphrysus* than with the large, shared haplotype consisting predominantly of *euthydemus* and *ruficollis* (Fig. 2A and B).

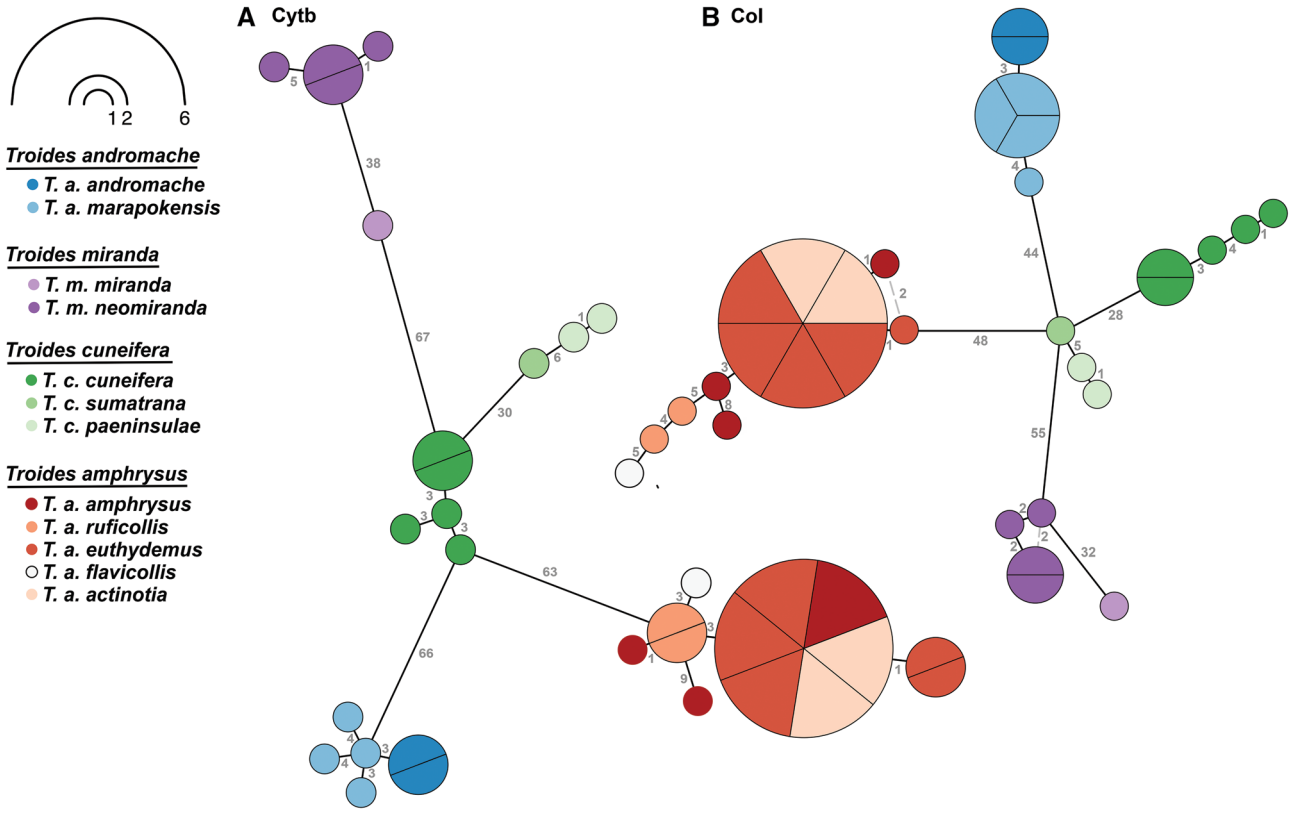

Fig. 2. Minimum spanning networks of mitochondrial haplotypes found within the *amphrysus*-group. Haplotypes are based on coding sequences of A) cytochrome b (CytB) and B) cytochrome oxidase subunit 1 (COI) sequences. Number of substitutions between different haplotypes is given in grey. Piechart colors confer subspecies identity and size is relative to the number of individuals belonging to that haplotype (ranging between 1 and 6).

Nuclear Genetic Structure

PCAngsd revealed $k = 2$ as the best-fit clustering scheme to describe intraspecific variation within each species complex (Fig. 3). As in mitochondrial haplotypes, geographic patterns and subspecies affiliation highly coincided with patterns of nuclear genetic structure (Supplementary Figs. S1 and S2A–D).

T. andromache individuals investigated revealed a gradation between the two dominant genetic components across both subspecies in the species complex (Fig. 3A). *T. andromache andromache* individuals were characterized by high degrees of one genetic component (0.87% to –0.99%), while *T. andromache marapokensis* individuals were more dominantly represented by the other genetic component (0.99% to 0.77%). Ordination clearly separated the two subspecies into distinct clusters (Supplementary Fig. S2A).

Within *T. miranda* the Sumatran subspecies *T. miranda neomiranda* was characterized by a different genetic component than its Bornean counterpart *T. miranda miranda* (Fig. 3B and Supplementary Fig. S2B). Only one individual of *neomiranda* showed a small proportion (5%) of the genetic component typical for the *T. miranda miranda* individual.

Genetic clustering revealed two distinct genetic components within species complex *T. cuneifera*, with Javan subspecies *T. cuneifera cuneifera* predominantly being characterized by one genetic component and sharing this to a low degree (9%) with the sampled individual of the Sumatran subspecies *T. cuneifera sumatrana* (Fig. 1C). This individual was further characterized by a high proportion (91%) of the genetic component present in the Peninsular subspecies subspecies *T. cuneifera paeninsulae*. The latter subspecies was characterized by a single genetic component and did not overlap from the dominant component in *T. cuneifera cuneifera*. Each subspecies was clearly separated from the others in the ordination analysis (Supplementary Fig. S2C).

in divergence time estimation. Bornean representatives *T. amphrysus flavicollis* and *T. amphrysus actinotia* were characterised by the presence of a single genetic component and did not exhibit genetic structure between them, leading us to lump these into one overarching Bornean taxon "*T. amphrysus flavicollis s. l.*" in divergence time estimation.

Phylogenetic Relationships

We were able to infer interspecific relationships among the studied taxa with a high degree of confidence (Fig. 4). Relationships between and within outgroup taxa were well-resolved and were supported by high bootstrap percentage (BP) values (100%). We recovered the *amphrysus*-group as a monophyletic lineage consisting of four divergent clades corresponding to the four species complexes *T. andromache*, *T. cuneifera*, *T. miranda*, and *T. amphrysus* (BP = 100%). *T. andromache* formed the earliest diverging group within the *amphrysus*-group and was confidently placed outside of the monophyletic clade consisting of *T. cuneifera*, *T. miranda*, and *T. amphrysus* (BP = 100%). Mitogenome and nuclear SNP phylogenies revealed a difference in the relationships between the *T. cuneifera*, *T. miranda*, and *T. amphrysus* complexes. *T. cuneifera* was inferred as the sister lineage to the clade containing *T. miranda* and *T. amphrysus* in the mitogenome phylogeny (Fig. 4A, BP = 100%). In the nuclear SNP tree, *T. miranda* was inferred as sister lineage to the clade consisting of *T. cuneifera* and *T. amphrysus* with high confidence (Fig. 4B, BP = 100%).

Intraspecific relationships showed variable support across the species complexes, and not all currently described subspecies were recovered as monophyletic (Fig. 4). Within *T. andromache*, subspecies *marapokensis* and the nominate subspecies *andromache* were polyphyletic in both the mitogenome and nuclear SNP phylogenies (BP= 66% to 44%). In *T. cuneifera*, specimens from subspecies *cuneifera* clustered together, forming a well-supported monophyletic clade in both phylogenies (BP = 100%). This clade was sister to a monophyletic clade of subspecies *paeninsulae* and *sumatrana*, in which *paeninsulae* was monophyletic with respect to the single *sumatrana* specimen (BP = 97%–100%). The two subspecies of *T. miranda* were reciprocally monophyletic, subspecies *neomiranda* was represented by multiple specimens in our dataset and formed a monophyletic sister clade to the single *miranda* specimen in the analysis (BP = 100%). Intraspecific relationships within *T. amphrysus* showed the highest degree of incongruence with the current subspecies taxonomy, the mitogenome phylogeny revealed short intraspecific branches, and none of the subspecies were recovered as monophyletic groups with high bootstrap support (Fig. 4A). The nuclear SNP phylogeny inferred Bornean subspecies *T. amphrysus flavicollis* and *T. amphrysus actinotia* as the sister lineage to the *T. amphrysus amphrysus* subspecies of Java and the subspecies *T. amphrysus euthydemus* and *T. amphrysus ruficollis* of Sumatra and the Malay Peninsula (Fig. 4B, BP = 100%). *T. amphrysus amphrysus* was recovered as a monophyletic clade with high confidence (BP = 100%). Subspecies *T. amphrysus euthydemus* and *T. amphrysus ruficollis* do not appear to be reciprocally monophyletic, and relationships between these two taxa could not be resolved with confidence (BP = 26% to 79%).

Divergence Time Analyses

Bayesian divergence time analyses recovered age estimates for the divergences of interspecific relationships and intraspecific relationships between monophyletic subspecies identified *a priori* using genotype clustering and ML phylogenies (Fig. 5). The two independent analyses converged after 300 million generations with high posterior ESS values (ESS = 342 to 401). The crown age of birdwings was estimated to be around 23.54 Ma (95% HPD = 17.09 to 31.68 Ma). The divergence of the genera *Ornithoptera* from *Troides* was estimated to have occurred around 23.51 Ma (95% HPD = 16.91 to 31.48 Ma). The most recent common ancestor of *Troides* was placed at 12.25 Ma (95% HPD = 8.91 to 16.40 Ma), and, subsequently, the *amphrysus*-group diverged from other *Troides* around 10.02 Ma (95% HPD = 7.36 to 13.51 Ma). *T. rhadamantus* and *T. plateni* diverged from each other approximately 0.139 Ma (95% HPD = 0.02 to 0.26 Ma). Within the *amphrysus*-group, species complexes diverged approximately between 7 and 6 Ma. *T. andromache* was the first lineage to diverge from the other three species complexes around 7.35 Ma (95% HPD = 5.27 to 9.88 Ma). *T. miranda* diverged from the monophyletic clade consisting of *T. cuneifera* and *T. amphrysus* around 6.96 Ma (95% HPD = 4.99 to 9.36 Ma). The Bornean subspecies *miranda* diverged from the Sumatran *neomiranda* approximately 2.05 Ma (95% HPD = 1.39 to 2.79 Ma). *T. cuneifera* split from *T. amphrysus* around 6.32 Ma (95% HPD = 4.55 to 8.54 Ma). Javan subspecies *cuneifera* diverged from Sumatran and Peninsular counterparts 0.76 Ma (95%

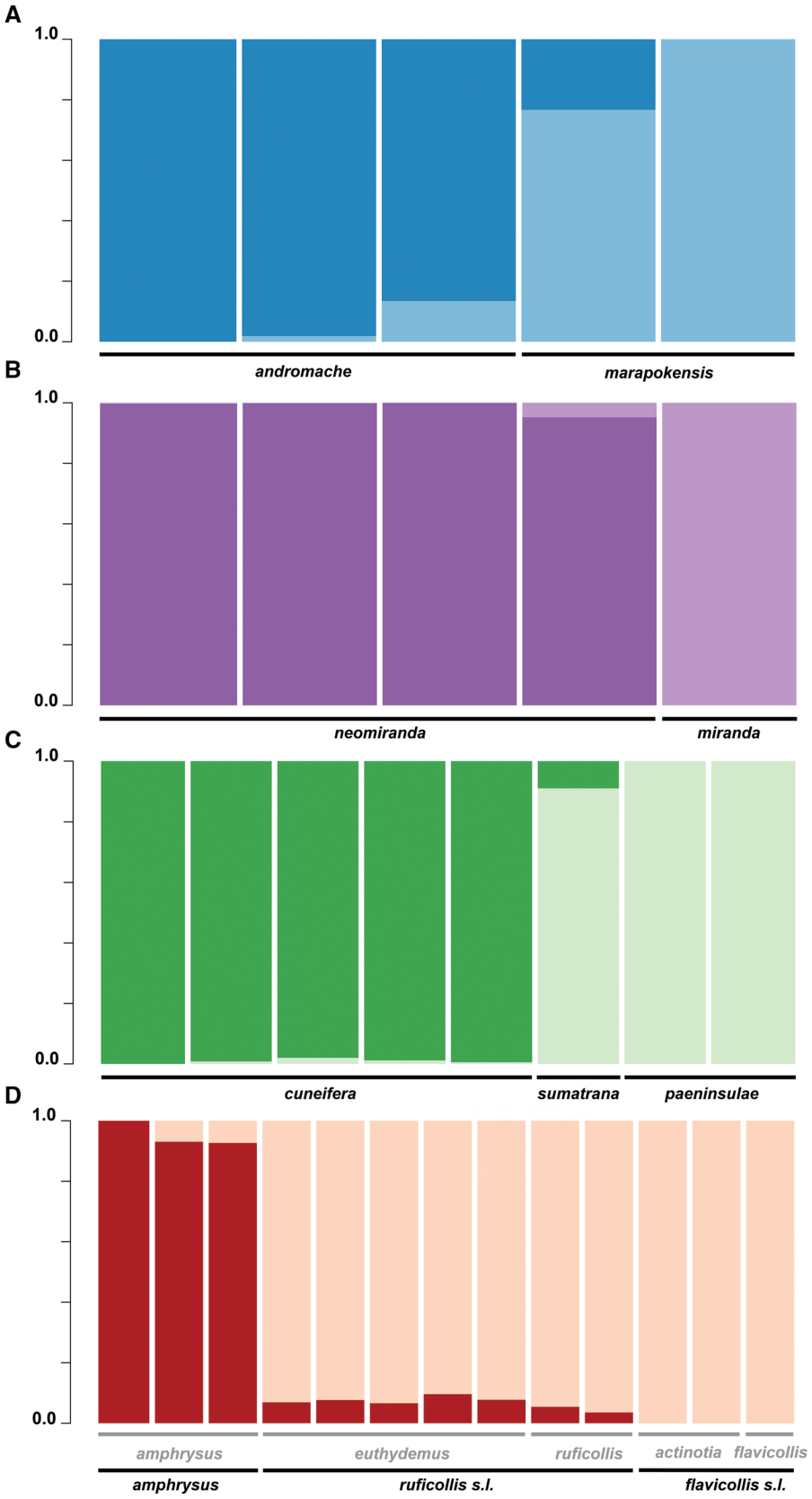

Fig. 3. Genomic composition as inferred by PCAngsd using the *amphrysus*-group_GL dataset. Best number of clusters (*k*) and percentage cluster assignment as inferred by PCAngsd for each of the species complexes: A) *Troides andromache*, B) *Troides miranda*, C) *Troides cuneifera*, and D) *Troides amphrysus*. Colors of genomic components were selected on the basis of the taxon they were most closely associated with as per the color scheme used in Figs 1 and 2.

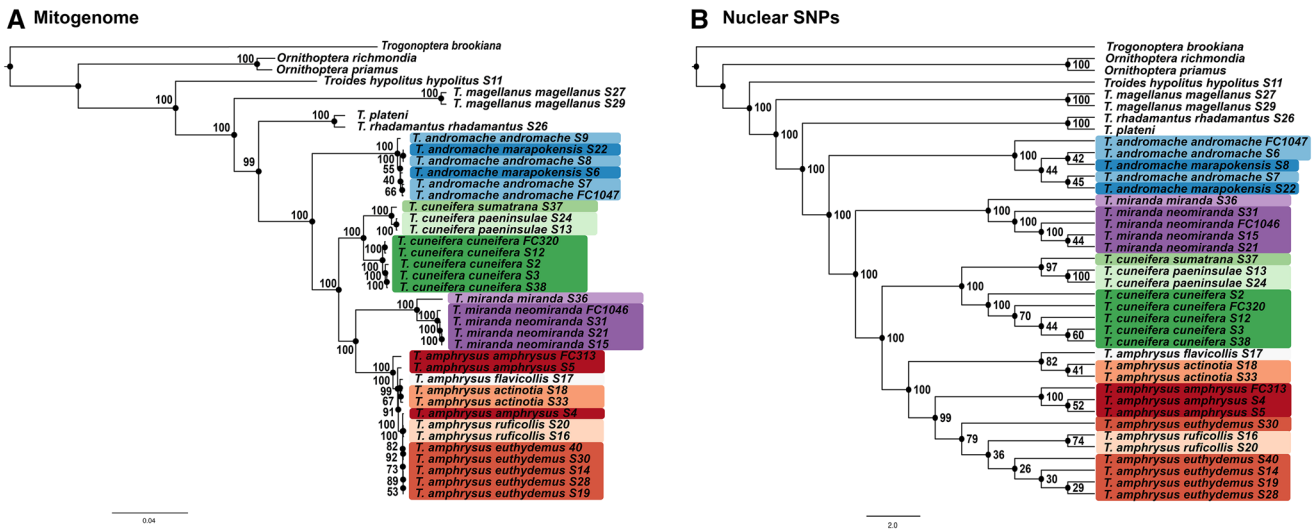

Fig. 4. A) Maximum-likelihood phylogeny of the *amphrysus*-group and outgroup taxa, as inferred by IQTREE 2 from the birdwing_mito dataset. Percentage bootstrap support is given per node. B) Maximum-likelihood phylogeny of the *amphrysus*-group and outgroup taxa. Topology inferred using the multi-species coalescent model with SVDquartets in PAUP* with the birdwing_SNP dataset. This phylogeny is only topological and does not have branch lengths. Percentage bootstrap support is given per node.

HPD = 0.52 to 1.03 Ma), which subsequently diverged 0.41 Ma (95% HPD = 0.25 to 0.58 Ma) into the Sumatran *sumatrana* and the Peninsular *paeninsulae*. In the most recently diversified species complex within the group, *T. amphrysus*, the nominate subspecies *amphrysus* split from the Bornean and Sumatran & Peninsular clades approximately 0.52 Ma (95% HPD = 0.36 to 0.70 Ma). Subsequently, the Bornean clade of *flavicollis s. l.* (*T. a. flavicollis* & *T. a. actinotia*) diverged around 0.17 Ma (95% HPD = 0.11 to 0.24 Ma) from the *T. a. ruficollis s. l.* clade (*T. a. euthydemus* & *T. a. ruficollis*).

Ancestral Range Estimation

Model comparison based on AICc scores for the two tested models yielded DEC as the most likely model for the biogeographical history of the *amphrysus*-group (Table 2). Under the DEC model, the ancestral ranges of the *amphrysus*-group were most likely shaped by vicariance (Fig. 6). The earliest diverging species in the analysis, *T. hypolitus*, is distributed in Wallacea. The crown node of the investigated phylogeny received a combined ancestral range composed of Wallacea, the Philippines, and Sundaland. The combination of the Philippines and Palawan was inferred as the most likely ancestral range of *T. rhadamantus* and *T. plateni* (Fig. 6). Internal nodes of taxa with extant distributions in the Philippines and Sundaland received a combined ancestral range of these two areas (Fig. 6). The crown node and all internal nodes the *amphrysus*-group were assigned to Sundaland, as this is also the extant distribution of the group (Fig. 6).

Species Delimitation

The distance-based delimitation algorithm ASAP reported six species-level partitions within the *amphrysus*-group, which were stable regarding distance metric choice (simple distance $P = 1.40e{-}01$, JC69 $P = 1.16e{-}01$, K80 $P = 1.01e{-}01$) (Fig. 7, Supplementary Fig. S3, Supplementary Table S1–S3). The tree-based bPTP analysis yielded the same species-level designations (Fig. 7, Supplementary Table S4). These six species-level lineages correspond to: (i) *T. andromache*, (ii) *T. cuneifera cuneifera*, (iii) *T. cuneifera paeninsulae* + *T. cuneifera sumatrana*, (iv) *T. miranda neomiranda*, (v) *T. miranda miranda*, and (vi) *T. amphrysus* (Fig. 7). The MSC BFD method yielded the classical taxonomy of each species complex as a species-level lineage as the best-fitting taxonomic model (Fig. 7, Table 3, Supplementary Table S5).

Taxonomic Changes

The taxon *Ornithoptera amphrysus euthydemus* Fruhstorfer, 1913, is here considered to be a junior synonym of *Ornithoptera ruficollis* Butler, 1879, as analysis reveal a lack of genetic structure in mitochondrial haplotypes (Fig. 2) and genomic components (Fig. 3, Supplementary Fig. S2D), in combination with the short branch lengths and non-monophyly in mitochondrial and nuclear phylogenetic analyses (Fig. 4). Therefore, *euthydemus* Fruhstorfer, 1913 is considered to be junior synonym of *ruficollis* Butler, 1879.

***Troides amphrysus* (Cramer, [1779])**

Papilio amphrysus Cramer, [1779]; Uitl. Kapellen 3 (17-21): 43, pl. 219, f. A; TL: Java

ssp. *ruficollis* (Butler, 1879)

Ornithoptera ruficollis Butler, 1879; Trans. Linn. Soc. Lond. (2) 1 (8): 552; TL: Malacca

= *Ornithoptera amphrysus euthydemus* Fruhstorfer, 1913 **syn. nov.**; Dt. ent. Z. Iris 27 (3): 133; TL: Sumatra

Discussion

We studied the patterns of divergence in a well-known clade of *Troides* birdwing butterflies by drawing on the specimen record amassed in natural history museums over a 100-yr period of biological exploration. Through the application of high-throughput sequencing to obtain whole-genome short-read resequencing data from museum specimens (museomics) to butterfly specimens collected between the 1890s and 1990s, we gathered nuclear and mitochondrial data spanning the geographical range of the *amphrysus*-group. In doing so, we were able to reconstruct the phylogenetic relationships, timing of divergence, historical biogeography, and intraspecific diversity and in turn provide a new

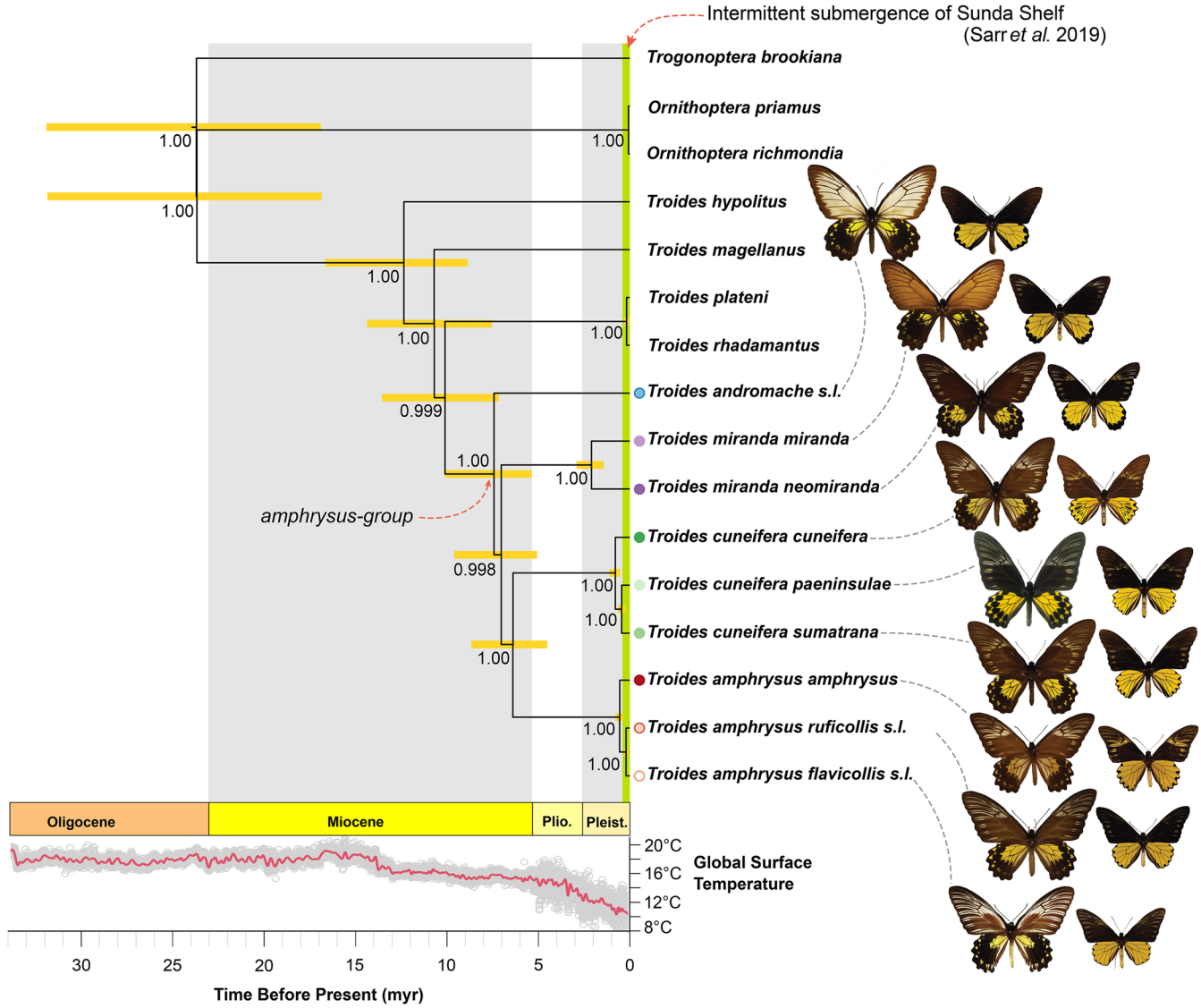

Fig. 5. Time-calibrated phylogeny of the *amphrysus*-group and selected outgroup taxa. Bars depict 95% HPD intervals of posterior estimates of divergence times. Scale indicated both as millions of years before the present and as a geologic time scale. Phylogenetic relationships and time calibrations inferred with Bayesian Inference in SNAPPER using 3,794 bi-allelic loci of the birdwing_SNP dataset. Selected subspecies are shown, specimen figures adapted from Naturalis Biodiversity Center & Natural History Museum (2014).

Table 2. Model comparison overview of the two ancestral range estimation models (DEC & DEC + *J*) evaluated using BioGeoBEARS for the *amphrysus*-group and selected outgroup taxa

	Log likelihood	Number of parameters	D	E	J	AICc
DEC	–9.53	2	1.00E–12	1.00E–12	0	24.27
DEC+J	–9.33	3	1.00E–12	4.00E–10	0.028	27.32

perspective on within- and between-species divergence in the megadiverse region Sundaland based on an iconic insect clade.

Phylogeny and Timing of Birdwing Evolution

Intergeneric and Interspecific Relationships

Interspecific and intergeneric relationships recovered in our study are in high agreement with earlier studies on the phylogenetic relationships of birdwings and Papilionidae in general (Figs 4 and 5) (Condamine et al. 2015, Allio et al. 2020, Allio et al. 2021). *Trogonoptera* and *Ornithoptera* are well-supported as separate genera and form the outgroup to *Troides*. The monophyly of the *amphrysus*-group has previously been recognized based on morphological (Haugum and Low 1979) and molecular studies (Condamine et al. 2015), and has been confirmed here using genome-scale data (Figs 4 and 5).

Within the *amphrysus*-group, we recovered two alternative topologies of the relationships between the three species complexes *T. cuneifera*, *T. miranda*, and *T. amphrysus* for the

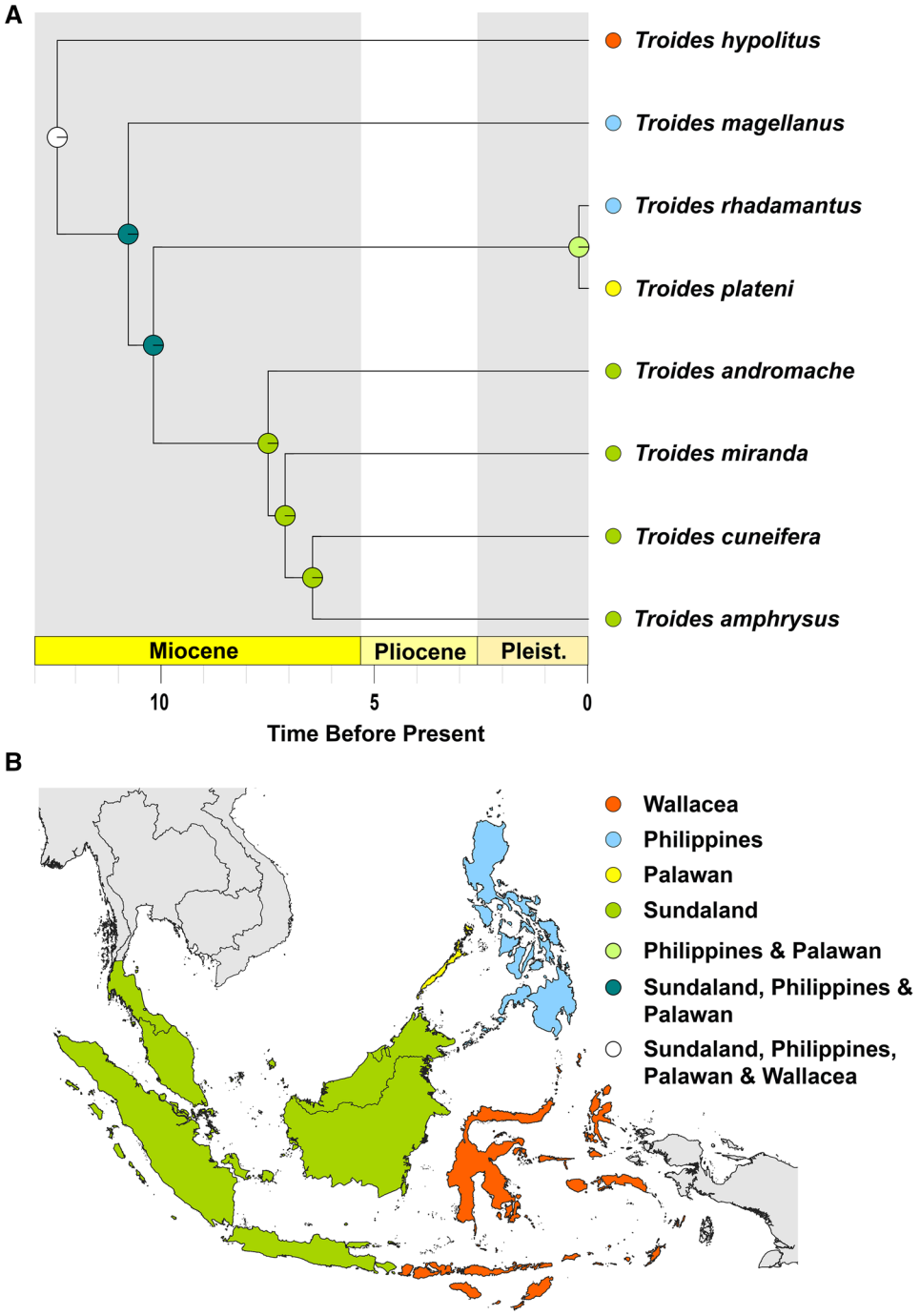

Fig. 6. A) Time-calibrated phylogeny with ancestral range estimates of the *amphrysus*-group and selected outgroup taxa as proposed under the DEC model in BioGeoBEARS. Scale indicated both as millions of years before the present and as a geologic time scale. Pie charts indicate the probability of ancestral ranges of a certain node. Range colors indicated on the phylogeny and inset map refer to: orange (Wallacea), blue (Philippines), yellow (Palawan), and green (Sundaland). Colored circles outside the phylogeny indicate extant ranges and piecharts at ancestral nodes represent ancestral ranges mixed colors represent mixed ancestral ranges. B) Current geographical distribution of land areas in the region of interest, colors correspond to the range colors as described above, grey areas are extralimital.

phylogenies obtained from the mitochondrial genome versus the nuclear genomes (Fig. 4). In the mitochondrial topology, *T. miranda* was placed as the sister taxon of *T. amphrysus* (Fig. 4A), whilst in the phylogeny obtained using the nuclear genome dataset, the inferred sister taxon of *T. amphrysus* was *T. cuneifera* (Fig. 4B). The sister lineage to *T. amphrysus* was the only node for which discordance was observed in the relationships between the investigated taxa. Condamine et al. (2015) recover *T. andromache* and *T. cuneifera* as a sister pair to *T. miranda* and *T. amphrysus*, which they respectively recover as sister lineages, based on three mitochondrial and two nuclear loci. Discordance between the mitochondrial and nuclear phylogenies could be due to various processes, including sex-biased dispersal, *Wolbachia*-mediated selective sweeps, adaptive

Table 3. Marginal likelihood estimates of a stepping stone analysis of 500 SNP subset of the *amphrysus*-group_SNP dataset using SNAPPER with 48 steps, alpha value of 0.3 and a 25% burnin

Taxonomic hypothesis	Species	MLE	BF	Rank
Classical taxonomy	4	–6,399.25	-	1
Classical + *neomiranda*	5	–6,936.77	1,075.04	2
All monophyletic subspecies	9	–6,975.19	1,151.88	3

MLE stands for marginal likelihood estimate and BF for Bayes Factor.

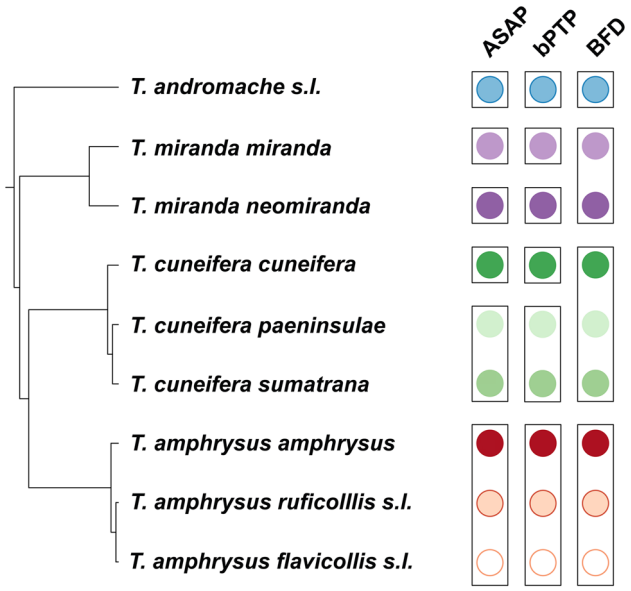

Fig. 7. Species delimitations based on three species delimitation analyses: Assemble Species by Automated Partitioning (ASAP) with the mitogenome alignment, Bayesian implementation of Poisson Tree Processes (bPTP) using the time-calibrated phylogeny and Bayes Factor Delimitation (BFD) using 500 and 1,000 SNPs in SNAPPER (the same partitions were obtained with both SNP datasets). Colored circles indicate taxonomic affiliation, and black outlines indicate species partitions proposed per analysis.

introgression, and incomplete lineage sorting, which are widespread among animal taxa, eg Toews and Brelsford (2012). The driver behind such discordance in the *amphrysus*-group is uncertain and requires further investigation. Disparate phylogenetic signals due to mito-nuclear discordance can hinder accurate reconstruction of patterns in speciation and biogeography. In this study, we therefore rely on the combined evidence from a large amount of SNP positions from across the genome to reconstruct phylogenetic relationships, timing of divergence, and biogeographical patterns with greatest accuracy.

Placing Divergence in a Temporal Framework

Divergence time estimation inferred the divergence time between the genera *Trogonoptera* and *Ornithoptera* + *Troides* to lie around 23.54 Ma (Fig. 5), which is corroborated by Condamine et al. (2015). The divergence of *Ornithoptera* and *Troides* is placed soon thereafter at 23.51 Ma (Fig. 5). The short internode distance is surprising, as other studies have inferred a greater divergence time interval between these branching events (Condamine et al. 2015, Allio et al. 2020). Despite acknowledging that clock rates may vary between different genera, or even species, we were restricted to the strict clock model, as it is currently the only available clock model for use with SNP data in BEAST (Stange et al. 2018). This short internode distance could be due to different clock rates within the different genera, resulting in artificial forcing together of divergence events if clock rates differ between branches.

Intraspecific relationships, especially shallow intraspecific relationships, are generally difficult to resolve using phylogenetic methods due to the interaction of phylogenetic and population genetic processes at play during speciation (Mirarab et al. 2021). Some, but not all, of these processes are accommodated for by careful model selection (Mirarab et al. 2021). Although divergence-dated "specimen trees" are commonplace in the literature, such trees commonly violate model assumptions for speciation and coalescent-based tree priors and may correspondingly provide inaccurate divergence time estimates (Ritchie et al. 2017). The MSC provides the most stable model to discern relationships within recently diversified clades, as it accommodates different genealogical histories within one species tree (Rannala and Yang 2003). The variation of genetic differentiation between different subspecies lineages (Figs 2, 3, 4, and 5) highlights that some lineages within the *amphrysus*-group may be further along the speciation continuum than others. Whilst acknowledging that population genetic processes may interfere with phylogenetic signal among subspecies lineages, we believe that our selection of taxa with allopatric distributions (Fig. 1), distinct haplotypes (Fig. 2), distinct genomic structure (Fig. 3, Supplementary Fig. S2) and monophyly in both mitochondrial (Fig. 4A) and nuclear trees (Fig. 4B) provides a valid set of taxa for divergence time estimation.

Late Miocene–Early Pliocene Speciation and Biogeography in *Troides*

Arrival in Sundaland

Model comparisons in BioGeoBEARS revealed the DEC model as the most likely for the biogeography of the investigated taxa (Table 2). Ancestral range reconstructions revealed a westward dispersal trend from Wallacea, via the Philippines to Sundaland (Fig. 6). The ancestral node could not be attributed to any of the geographical areas with high certainty, instead receiving a combination of all four states as the most likely range (Fig. 6). Shared ancestral nodes of species with a Philippine—Palawan distribution and those with a Sundaic distribution received a combination of these ranges as the most likely ancestral range. The ancestral nodes for these taxa were attributed to a combined distribution in Sundaland and the Philippines (Fig. 6). The lack of a jump dispersal parameter in the best-fit model explains why such a vicariance-based biogeographic ancestry is proposed for the investigated taxa. Such a vicariance-based biogeographical history seems unlikely given the prevailing geological evidence of biogeographical patterns in the region. Geological evidence indicates that land fragments of Wallacea, the Philippines, Palawan, and Sundaland were not connected as one large land area at any point in their geological history (Hall 2017). This would preclude a biogeographical history driven by vicariance. Wallacea and the Philippines are derived from different composite origins from different plate fragments

that behave as microplates (Hall 2017). Moreover, Palawan is a microplate that has originated from East Asia and moved southwards independently reaching the Philippines and Sundaland (Hall 2017). Given this geological evidence, "jump dispersal" between different land areas, rather than vicariance, would be the most likely biogeographical process to shape distributions. Combining the present-day ranges of the group with the geological evidence available for the region, we discuss an alternative biogeographical history for the studied taxa. The westward spread of the *amphrysus*-group into Sundaland from Wallacea via the Philippines and Palawan is corroborated by a comprehensive study on the biogeography of the birdwing genera *Trogonoptera*, *Ornithoptera*, and *Troides* (Condamine et al. 2015). Sundaland was likely intermittently connected with the Philippines via Borneo along an arc of volcanic islands in the Sulu Sea during the late Miocene (~10 Ma) (Hall 2013). This may have functioned as the entry point to Sundaland for ancestral representatives of the *amphrysus*-group from their former range in the Philippines and Wallacea (Condamine et al. 2015). At this time Borneo, Sumatra, and the Malay Peninsula were connected by a land bridge running across the central Sunda Shelf (Hall 2013), perhaps facilitating colonization of the western Sundaic landmasses. Future analyses with more taxa, and perhaps with a realistic approximation of the difficulty of dispersal between different ranges, may provide a more biogeographical reconstruction that is more congruent with the biological reality.

Biogeography Within Sundaland

Earlier-diverging species complexes *T. andromache* and *T. miranda* have a Bornean-Sumatran present-day distribution while later diverging species complexes *T. cuneifera* and *T. amphrysus* have a present-day distribution centered in western Sundaland (Fig. 1). The Borneo-centered distribution of early diverging species complexes could imply Borneo as an evolutionary source within Sundaland, as it is one of the older landmasses, has the largest land-area with the greatest expanse of highlands, and remained relatively climatically stable throughout its geological history (Hall 2013). These properties may have allowed lineages that went extinct elsewhere to persist on Borneo. In a meta-analysis of divergence times, De Bruyn et al. (2014) identified Borneo as a major evolutionary hotspot within Southeast Asia, with Borneo showing a diverse range of organisms having significantly higher levels of *in situ* diversification, and correspondingly, higher levels of emigration than any other Sundaic landmass (De Bruyn et al. 2014). The importance of Borneo as a source and center of diversity for many Sundaic taxa has been coined the "Out-of-Borneo" hypothesis and is applicable to a wide range of taxa (De Bruyn et al. 2014, Grismer et al. 2016, Williams et al. 2017, Flury et al. 2021). Our results corroborate the "Out-of-Borneo" hypothesis, with Borneo likely serving as a stepping stone in the colonization of western Sundaland.

The species complexes that make up the *amphrysus*-group are inferred to have diverged within Sundaland 7-6 Ma (Fig. 5).This divergence time interval corresponds with speciation in other Sundaic clades, such as freshwater crabs (Klaus et al. 2013), toads (Grismer et al. 2016), rough-skinned skinks (Karin et al. 2017), bulbuls (Manawatthana et al. 2017), and tree squirrels (Den Tex et al. 2010, Hinckley et al. 2020) and coincides with marked changes in paleoclimate and paleogeography in the Southeast Asian region (Hall 2009, 2013, Holbourn et al. 2018). During the mid- to late Miocene, Borneo became a higher and larger landmass, and mountain ranges rapidly emerged in Sumatra and West Java (Hall 2013). Extensive lowland and highland areas were present in Borneo, the Malay Peninsula, and Sumatra, with Java emerging rapidly from approximately 7 Ma onwards (Hall 2013). Additionally, intense climatic changes took place during this timeframe because of late Miocene cooling and the intensification of the Southeast Asian winter monsoon 7 to 5.5 Ma (Holbourn et al. 2018).

Pleistocene Divergence and Biogeography Within Sundaland

Pleistocene Divergence Within the *amphrysus*-Group

Following the initial divergence in the late Miocene—early Pliocene, no new extant lineages arose for approximately 4 million years until a second interval of divergence, which is inferred to have taken place during the Pleistocene between 2.05 and 0.17 Ma, which gave rise to the currently recognized subspecies (Fig. 5). Pleistocene glaciations have globally impacted evolutionary processes, not only in high-latitude regions but also in the tropics (Hewitt 2000, Hewitt 2004, Cros et al. 2020, Garg et al. 2020, Garg et al. 2022). Glaciations allowed montane communities to periodically track the cooling climate downslope and correspondingly occupy larger ranges, with the opposite occurring during interglacial periods (Hewitt 2000, Qu et al. 2011). Glaciations likely facilitated population expansions of montane taxa (Garg et al. 2020) and thereby allowed gene flow between previously disjunct populations (Hewitt 2000, Qu et al. 2011, Capurucho et al. 2018). In addition, glaciation-driven shifts in vegetation zones are expected to have a strong impact on distributions of phytophagous insects through their own microclimatic niche requirements and tight ecological links with their host plants.

Subspecies divergence during the Pleistocene highlights the role of paleoclimate dynamics on recent population divergence in Sundaic insects (Fig. 5). Concordantly with other species groups in Sundaland, *amphrysus*-group divergence times do not follow a "one-size-fits-all" pattern. Like birds such as bulbuls and babblers (Cros et al. 2020) and mammals, such as colugos and mouse deer (Mason et al. 2019) and tree squirrels (Den Tex et al. 2010, Hinckley et al. 2020), some *Troides* lineages show relatively deep interspecific divergence across the Greater Sunda Islands (eg *T. miranda*), whilst others are only weakly diverged (eg *T. amphrysus*). Ecological characteristics may drive such patterns, such as shown for edge-tolerant versus forest-dwelling birds (Cros et al. 2020) and mountain versus lowland tree squirrels (Hinckley et al. 2020). Different degrees of ecological specialization occur within the *amphrysus*-group, and the clade contains ecological generalists but also lowland and montane specialists. These different ecological traits sometimes vary within species complexes where closely related taxa have evolved as specialists of montane (eg *Troides miranda neomiranda*) or lowland (eg *Troides miranda miranda*) habitats. However, in general, most *amphrysus*-group species complexes occur across a relatively broad elevation gradient (Matsuka 2001) and unfortunately, detailed ecological descriptions are lacking for most subspecies. Unfortunately, this hampers a clear investigation of intraspecific divergence patterns in relation to ecological specialization within the *amphrysus*-group.

Investigation of montane divergence may be aided by sampling more densely and sampling across elevational gradients of known montane specialists, eg *T. andromache*.

In both species complexes with a representative on Java (*T. cuneifera* and *T. amphrysus*), subspecies occurring on Java form a sister lineage to the other subspecies lineages (Figs 2 to 5). This corroborates a more general pattern of Java showing the greatest degree of faunal dissimilarity to the Malay Peninsula, Borneo, and Sumatra, which has been long recognized at the species level for birds and mammals (Wallace 1869, 1911) and at the population level across a wide range of the Sundaic vertebrate fauna (Leonard et al. 2015, Mason et al. 2019, Lim et al. 2020, Wu et al. 2022). Multiple drivers for this pattern have been suggested. For example, the Malay Peninsula, Borneo, and Sumatra were more frequently connected during times of lower sea levels than Java was connected to these landmasses (Voris 2000). Furthermore, Javan ecosystems are more divergent to those on the other Sundaic landmasses. Indeed, Java is relatively recent and has more volcanic activity (Hall 2009, 2013) and receives lower average rainfall (Whitten et al. 1997). Which correspondingly translates to different vegetations, especially in the central and eastern regions of the island (Whitten et al. 1997). Together, the lower degree of connectivity across Sunda Shelf land bridges and more distinct ecological conditions on Java most likely explain the greater genetic differentiation of its *Troides* populations. On the other hand *T. amphrysus flavicollis s. l.* of Borneo and *T. amphrysus ruficollis s. l.* of Sumatran and the Malay Peninsula are most closely related to each other. This biogeographical connection between Borneo and Sumatra is well-known for different taxonomic groups within Sundaland (Husson et al. 2020).

Sea-level Fluctuations and the *amphrysus*-Group

During glaciations, global sea levels decreased as water became locked in expanding polar ice sheets, with local sea levels in the Sundaland region generally decreasing to 40 to 60 m (up to 120 m) below present sea level, thereby intermittently exposing the Sunda Shelf (Voris 2000, Hall and Morley 2004, De Bruyn et al. 2014). The most recent scenario, based on re-evaluated geological and biogeographical evidence by Sarr et al. (2019), indicates that the Sunda Shelf may have been continuously exposed until ~400,000 years ago and intermittently during glaciations from then onwards (green band, Fig. 5). When Sundaland was exposed, evergreen rainforests are thought to have covered most of its plains, eventually forming an expanse of rainforest habitat (Cannon et al. 2009, Woodruff 2010, Raes et al. 2014). Such extensive rainforests, of both lowland and upland vegetation zones, are thought to currently exist in a state of refuge (Cannon et al. 2009, Woodruff 2010, De Bruyn et al. 2014).

Meta-analysis of intraspecific divergence times in different (largely vertebrate) clades revealed broad biogeographical support for divergences between 400,000 and 18,000 years ago, coinciding with periodical flooding of the Sunda Shelf (Husson et al. 2020). Our phylogenetic evidence indicates recent intraspecific divergence within the species complexes *T. cuneifera* and *T. amphrysus*, with divergence time intervals that broadly overlap with the sea level fluctuations on the Sunda Shelf (Fig. 5). *T. cuneifera paeninsulae* from the Malay Peninsula and *T. cuneifera sumatrana* from Sumatra diverged between 580,000 and 250,000 years ago. Javan *T. amphrysus* populations diverged approximately 700,000 to 360,000 years ago from Bornean and western Sundaic subspecies *T. amphrysus flavicollis s. l.* and *T. amphrysus ruficollis s. l.*, which respectively diverged from one another between approximately 240,000 to 110,000 years ago (Fig. 5). These divergence times coincide with the onset of major marine incursions reported by Sarr et al. (2019), which may have been a significant driver of population divergence (Fig. 5). Divergence between taxa of the Malay Peninsula and Sumatra, and between Sumatra and Borneo, appears particularly congruent with the widespread divergence scenario for Sundaic taxa as described in Husson et al. (2020). Marine incursions occurred repeatedly, and at least four inundation cycles may have taken place after 400,000 years ago (Sarr et al. 2019); this may have resulted in cyclical reproductive isolation and secondary contact between populations.

We have no reason to believe that overwater dispersal continuously takes place between *amphrysus*-group populations on the larger landmasses of Sundaland given the divergence time estimates (Fig. 4) and the genetic distance (Fig. 2) between allopatric subspecies in the different species complexes. The only taxa for which overwater dispersal may take place are subspecies *T. amphrysus euthydemus* and *T. amphrysus ruficollis*, which are highly similar genetically (Figs 2 and 3) and both occupy relatively low-lying areas that are only separated by the narrow Strait of Malacca (Fig. 1). However, due to the particularly shallow depth of this section of the Sunda Shelf, a land bridge connection via the Riau Archipelago was most likely frequently exposed for long durations, even at slightly lower sea levels (Voris 2000). The proximity and frequent land bridge connection may correspondingly be the cause of the close genetic affinity between these two populations, for which reproductive isolation may have been as recent as after the Last Glacial Maximum (~21,000 years ago).

Particularly for lowland taxa, sea level fluctuations are expected to have resulted in losses of habitat and connectivity. The most recently diverged lineage, *T. amphrysus*, is the only species complex of which subspecies are known to occur from sea level to the lower montane zone and on numerous islands off the coasts of the Greater Sunda Islands (Haugum and Low 1979, Tsukada and Nishiyama 1982, Matsuka 2001). Changes to the connectivity of lowland ecosystems due to marine incursions on the Sunda Shelf may therefore have had a particularly significant impact on patterns in this species complex.

Increasing the geographical coverage of this wide-ranging, subspecies-rich, and recently diversified "supertramp" species (*sensu* Diamond [1974]) may reveal additional insights into the role of sea level fluctuations on its diversification. Here, Sunda Shelf islands may provide additional biogeographical clues by harboring intermediate forms between eastern and western Sundaland populations. Recently, several studies demonstrated that populations of birds and mammals on such islands show genetic affinity with the landmasses they were most recently connected with during Pleistocene glaciations, which does not necessarily correspond to the nearest landmass, in contrast to classic island biogeographical theory (Mason et al. 2019, Garg et al. 2022). Such patterns could perhaps also apply to populations in the *T. amphrysus* species complex, and sampling of island forms occurring on islands between Borneo and the western Sundaic landmasses such as the Natuna Islands would be particularly insightful.

Hidden Diversity Revealed?

Currently, recognized subspecies of the *amphrysus*-group have generally been described based on coloration and allopatric distribution ranges, eg Staudinger (1892). However, color patterns within the *amphrysus*-group may be relatively plastic and vary considerably between closely related subspecies whilst showing overlap between co-occurring but more distantly related taxa, possibly governed by Müllerian mimicry associations (Haugum and Low 1979). Nevertheless, results from phylogenomic analyses closely match the 19th and early 20th century classifications within the *amphrysus*-group. Cryptic diversity—phenotypically identical but widely genetically differentiated lineages—is generally not present within the species group.

We applied three different types of species delimitation analyses to the dataset, on nuclear and mitochondrial datasets, to assess if hidden diversity—greater genetic differentiation than expected based on phenotypic differences—occurs among the investigated taxa. ASAP, the distance-based analysis, and the tree-based analysis bPTP, were applied to the mitochondrial data and delimited the same six species-level lineages (Fig. 7, Supplementary Tables S3 and S4). The MSC method, BFD, applied to SNP subsets of 500 and 1000 SNP, did not reveal the same pattern and supported the classical taxonomy instead (Table 3, Supplementary Table S5). This could indicate that classical taxonomy best reflects the population genetic processes among the studied taxa or that the subsampled SNP sets did not contain sufficient signal to discern more recent divergence between taxa. However, the congruence between the first two methods indicates that the *amphrysus*-group may putatively be represented by more species-level lineages rather than the originally described four (Fig. 7).

The results from tree- and distance-based species delimitation analyses identified two putative species within the *T. miranda* species complex, which corresponds with marked differentiation in phenotype and ecology (Fig. 7). Female *T. miranda miranda* and *T. miranda neomiranda* differ markedly in the coloration of forewings, *T. miranda neomiranda* females having more strikingly patterned forewings in comparison to the more uniformly colored *T. miranda miranda* females (eg insets, Fig. 5). Haugum and Low (1979) describe significant differences in male genitalia between *T. miranda miranda* and *T. miranda neomiranda*, especially noting a different degree of serration on the ventral hook of the harpe. Not only do the two subspecies occur on two—currently—distantly separated areas (Borneo and Sumatra) (Fig. 1), their habitat and altitudinal preferences are also very different. *T. miranda miranda* is known predominantly as a lowland species, even frequently recorded in sea-level habitats such as mangroves and lowland rainforest, and is rarely recorded at higher elevations and unknown from elevations exceeding 500 m above sea level (Haugum and Low 1979). On the contrary, *T. miranda neomiranda* is a high-altitude species, generally thought to be restricted to mountain ranges around 1,000 m above sea level (Haugum and Low 1979, Tsukada and Nishiyama 1982). Both Haugum and Low (1979) and Tsukada and Nishiyama (1982) have hinted at possible species-level differentiation. And along with the multiple ecological and morphological distinctions, the evidence presented in this study provides the first line of genetic evidence for a respective species status for *T. miranda miranda* and *T. miranda neomiranda*. All analyses conducted on mitogenome and nuclear datasets reveal a high level of differentiation between the two subspecies with the greatest degree of differentiation encountered within the *amphrysus*-group species complexes (Figs 2, 3, 4, 5 **and** Supplementary Fig. S2B). However, we remain cautious because a single specimen for *T. miranda miranda* has been included. Furthermore, the inclusion of the rarely collected *T. miranda hayamii*, the southernmost representative of the *T. miranda* complex in Borneo, would shed additional light on the taxonomic status, divergence, and biogeographical patterns of this species complex.

The second case of putative species-level differentiation was uncovered in the *T. cuneifera* species complex, with tree- and distance-based delimitation methods indicating a species-level split between *T. cuneifera cuneifera* and *T. cuneifera paeninsulae* + *T. cuneifera sumatrana* (Fig. 7). We found a high degree of genomic differentiation between these two putative species (Fig. 3, Supplementary Fig. S2C). However, the potential species-level differentiation within *T. cuneifera* is less obvious with respect to phenotypic differentiation, as phenotypes of the two subspecies are similar and may be influenced by Müllerian mimicry associations with co-occurring *Troides* species. With respect to male genitalia, the clasper and harpe of *T. cuneifera cuneifera* differ slightly from those of *T. cuneifera paeninsulae* and *T. cuneifera sumatrana* (Haugum and Low 1979). Observations on the ecology of the subspecies, such as habitat requirements and host-plant associations, are scant and cannot at present provide additional evidence for taxonomical considerations within this species complex. On the contrary, *T. cuneifera paeninsulae* and *T. cuneifera sumatrana* are highly similar, and multiple authors have questioned the validity of their respective subspecies status eg Haugum and Low (1979). However, genetic distance, genomic composition, and phylogenomic evidence do indicate some degree of genomic differentiation between these subspecies, and their status as distinct subspecies remains valid from this perspective (Figs 2, 3, and 4).

Aside from these two putative species-level splits indicated by some of the delimitation methods, subspecies designations seemed to reflect the genetic differentiation between the different taxa examined relatively accurately. For some subspecies we do not recover enough population genomic or phylogenetic signal to justify subspecies designations and lump taxa for divergence time and biogeographical analyses. One such case is the *T. andromache* complex, where *T. andromache andromache* and *T. andromache marapokensis* are found in proximity in northern Borneo. This species complex exhibits a (low) degree of variation in mitochondrial haplotype (Fig. 2), and nuclear data reveal dominance of different genomic components in the two subspecies (Fig. 3, Supplementary Fig. S2A). However, this low degree of differentiation is combined with non-monophyly in both mitogenome and nuclear phylogenies (Fig. 4). Together, this evidence does not support a scenario whereby the two populations have been isolated from each other for sufficient time to treat them as reciprocally monophyletic subspecies. We correspondingly refrained from dating the divergence of these two subspecies. Another subspecies pair, *T. amphrysus euthydemus* and *T. amphrysus ruficollis*, shows near-complete overlap in mitochondrial haplotypes (Fig. 2), high similarity in genomic components (Fig. 3, Supplementary Fig. S2D), and non-monophyly in both mitogenome and nuclear phylogenies (Fig. 4). This is a strong indication that respective subspecies status may not accurately reflect the level

of divergence between the two taxa. This has been recognized previously, with several authors noting a lack of constant morphological characters to separate the Peninsular and Sumatran populations (eg d'Abrera [1975]), however, their distinct status has persisted since their description. The lack of distinction between Peninsular and Sumatran representatives is not completely unsurprising, as *T. amphrysus* is known to have a relatively broad elevational and geographical range, along with the frequent and recent land connection between Sumatra and the Malay Peninsula during periods of lower sea-level (Voris 2000). The paucity of morphological distinctions in conjunction with the lack of genetic differentiation recovered by this study indicates that *T. amphrysus euthydemus* (Fruhstorfer, 1913) **syn. nov.** is a junior synonym of *T. amphrysus ruficollis* (Butler 1879).

The ability of species delimitation methods to differentiate genetic clusters into species-level lineages varies between approaches and the speciation model (Carstens et al. 2013). Comparison of several species delimitation methods to form a consensus delimitation scheme is generally advocated (Carstens et al. 2013, Suárez-Villota et al. 2018, Jin et al. 2020, Pozzi et al. 2020). Together, the different delimitation methods reveal a plausible species partition scheme among the studied taxa (Fig. 7); however, it remains difficult to weigh different analyses against one another. Molecular species delimitation requires critical evaluation and should not be seen as a replacement to traditional taxonomic treatment, as stressed by many developers of species delimitation algorithms (eg Puillandre et al. (2012)). Classical taxonomical revision of the *amphrysus*-clade should be carried out to validate suggested species-level splits within the *T. cuneifera* and *T. miranda* species complexes. Such work should encompass all subspecies lineages and be carried out in an integrative taxonomic framework that considers morphological, ecological, and genomic lines of evidence to investigate species boundaries.

Conclusion

The complex distribution patterns of island biota have been a source of fascination to evolutionary biologists since Alfred Russel Wallace founded the field of biogeography (Wallace 1855). In Sundaland, complex geography and geological and paleoclimatic processes have shaped evolution in a way that exemplifies this fascination to the extreme. In this study we applied museomics to a well-known and diverse lineage of butterflies from one of the world's most famous archipelagos. Using phylogenomic methods, even specimens collected ~150 years ago by 19th-century naturalists could be successfully incorporated into a phylogenetic framework and thereby contribute to our understanding of the biogeographic and speciation history of the group. Here, we provide evidence for intraspecific biogeographic processes in Sundaic insects and thereby address a knowledge gap in our understanding of biogeographical patterns in this biodiversity hotspot. After initially originating in the late Miocene, the formation of *amphrysus*-group species complexes dates to the Pliocene and within-species divergence to the Pleistocene, reaffirming the role of this geological period in shaping the Sundaic biota. Intraspecific divergence patterns within the *amphrysus*-group are highly congruent with those of other taxonomic groups in Sundaland and provide support for the prevailing biogeographical hypotheses between the different Sundaic landmasses. Furthermore, through species delimitation analyses, we provide the first holdfasts for describing hidden species within the group and suggest detailed taxonomical examination of the *T. cuneifera* and *T. miranda* complexes.

Acknowledgements

We would like to thank Viola Richter (Museum für Naturkunde, Berlin) for providing the Museum für Naturkunde specimens used in this study. We would also like to express our thanks to the team of volunteers—"de Papillotters"—who have digitized butterfly collections at Naturalis Biodiversity Center and made them available for research. Furthermore, we are grateful for the help of Elsa Call, Freek Bakker, Harald Letsch, Siavash Mirarab and Chris Wheat, who provided a wealth of advice and guidance regarding their respective areas of expertise. We would also like to thank Martin Partridge of the Swallowtail & Birdwing Butterfly Trust for sharing relevant literature. We would also like to acknowledge Djunijanti Peggie (Bogor Zoology Museum, Bogor), Ulmar Grafe (Universiti Brunei Darussalam, Brunei Darussalam), and Wei Song Hwang (Lee Kong Chian Natural History Museum, Singapore) for searching for *T. andromache* specimens in the collections under their care. We are grateful for comments by Tobias Bader, Esther Blommert, Mikay Breet, Klaas Bouwmeester, Nora Damaris Pasquali Medici de Biron, Liana Greenburg, Iris Haverkorn, Sybren Kooistra, Thijmen van der Loop, and Isabel Lugtenburg on early versions of the manuscript. Additionally, we would like to thank Bart Burger for his suggestions on components of the analysis and Isabela Pombo Geertsma for her suggestions to improve the final manuscript and figures. We also thank Siva Selvanayagam for his assistance with uploading annotated mitogenomes to ENA. We also thank both the Uyttenboogaart-Eliasen Stichting and the Koninklijke Hollandsche Maatschappij der Wetenschappen for their recognition of this work.

Nomenclature

This paper and the nomenclatural acts it contains have been registered in Zoobank (www.zoobank.org), the official register of the International Commission on Zoological Nomenclature. The LSID (Life Science Identifier) number of the publication is: urn: lsid: zoobank.org: pub : 6427A570-6BE8-49DE-A30E-8FAD220EEEEA

Author Contributions

Corné van der Linden (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Writing—original draft [lead], Writing—review & editing [lead]), Eliette Reboud (Data curation [equal], Formal analysis [equal], Resources [equal], Writing—review & editing [equal]), Robin van Velzen (Methodology [equal], Supervision [equal], Writing—review & editing [equal]), Joost van den Heuvel (Methodology [equal], Supervision [equal], Writing—review & editing [equal]), Emmanuelle Chevalier (Resources [equal], Writing—review & editing [equal]), Marcel Eurlings (Methodology [equal], Resources [equal], Writing—review & editing [equal]), Mónica Guimarães Cruz (Resources [equal], Writing—review & editing [equal]), Théo Léger (Resources [equal], Writing—review & editing [equal]), Stephen Sutton (Conceptualization [equal], Writing—review & editing

[equal), Patrick Verbaarschot (Methodology [equal], Resources [equal], Supervision [equal], Writing—review & editing [equal]), Rob de Vos (Resources—[equal], Writing—review & editing [equal]), Niklas Wahlberg (Methodology [equal], Resources [equal], Writing—review & editing [equal]), M. Eric Schranz (Project administration [equal], Supervision [equal], Writing—review & editing [equal]), Fabien Condamine (Funding acquisition [equal], Methodology [equal], Resources [equal], Writing—review & editing [equal]), and Sabrina Simon (Conceptualization [equal], Funding acquisition [equal], Project administration [equal], Resources [equal], Supervision [equal], Writing—review & editing [equal])

Supplementary Material

Supplementary material is available at *Insect Systematics and Diversity* online.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project GAIA, agreement no. 851188) and under the European Union's Horizon Europe Framework Programme (project ORION, agreement no. 101170500) to F.L.C. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Conflicts of Interest

None declared.

Data Availability

Raw sequences and mitogenome assemblies are available European Nucleotide Archive (ENA) and the National Center for Biotechnology Information (NCBI) under accession number PRJEB64430. Alignments and variant call files are available at Zenodo at https://doi.org/10.5281/zenodo.15855028.

References

Allio R, Nabholz B, Wanke S, et al. 2021. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. *Nat. Commun.* 12:354. https://doi.org/10.1038/s41467-020-20507-3

Allio R, Scornavacca C, Nabholz B, et al. 2020. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. *Syst. Biol.* 69:38–60. https://doi.org/10.1093/sysbio/syz030

Baele G, Lemey P, Bedford T, et al. 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

Bakker FT, Antonelli A, Clarke JA et al. 2020. The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ.* 8:e8225. https://doi.org/10.7717/peerj.8225

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouckaert R, Vaughan TG, Barido-Sottani J, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Broad Institute. 2019. Picard Toolkit. Version 3.0.0. GitHub Repository. Broad Institute. https://broadinstitute.github.io/picard/

Browning BL, Tian X, Zhou Y, et al. 2021. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* 108:1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Call E. 2020. *The age of museomics: how to get genomic information from museum specimens of Lepidoptera* [doctoral thesis]. Lund University. p. 9-41. https://www.lunduniversity.lu.se/lup/publication/97808862-3009-4f30-adb5-bbb3c4eb88f0

Call E, Mayer C, Twort V, et al. 2021. Museomics: phylogenomics of the moth family Epicopeiidae (Lepidoptera) using target enrichment. *Insect Syst Divers.* 5. https://doi.org/10.1093/isd/ixaa021

Cannon CH, Morley RJ, Bush ABG. 2009. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc. Natl. Acad. Sci. U S A.* 106:11188–11193. https://doi.org/10.1073/pnas.0809865106

Capurucho JMG, Ashley MV, Ribas CC, et al. 2018. Connecting Amazonian, Cerrado, and Atlantic forest histories: Paraphyly, old divergences, and modern population dynamics in tyrant-manakins (Neopelma/Tyranneutes, Aves: Pipridae). *Mol. Phylogenet. Evol.* 127:696–705. https://doi.org/10.1016/j.ympev.2018.06.015

Card DC, Shapiro B, Giribet G, et al. 2021. Museum genomics. *Annu. Rev. Genet.* 55:633–659. https://doi.org/10.1146/annurev-genet-071719-020506

Carstens BC, Pelletier TA, Reid NM, et al. 2013. How to fail at species delimitation. *Mol. Ecol.* 22:4369–4383. https://doi.org/10.1111/mec.12413

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65:997–1008. https://doi.org/10.1093/sysbio/syw037

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics.* 30:3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Cho S, Epstein SW, Mitter K et al. 2016. Preserving and vouchering butterflies and moths for large-scale museum-based molecular research. *PeerJ.* 4:e2160. https://doi.org/10.7717/peerj.2160

Condamine FL, Toussaint EF, Clamens A-L, et al. 2015. Deciphering the evolution of birdwing butterflies 150 years after Alfred Russel Wallace. *Sci. Rep.* 5:11860. https://doi.org/10.1038/srep11860

Corlett R. 2014. *The ecology of tropical East Asia*, Oxford University Press

Cros E, Chattopadhyay B, Garg KM, et al. 2020. Quaternary land bridges have not been universal conduits of gene flow. *Mol. Ecol.* 29:2692–2706. https://doi.org/10.1111/mec.15509

d'Abrera B. 1975. *Birdwing butterflies of the world*. Lansdowne Press.

Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience.* 10:giab008. https://doi.org/10.1093/gigascience/giab008

De Bruyn M, Stelbrink B, Morley R, et al. 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian Biodiversity. *Syst. Biol.* 63:879–901. https://doi.org/10.1093/sysbio/syu047.

Den Tex R-J, Thorington R, Maldonado JE, et al. 2010. Speciation dynamics in the SE Asian tropics: putting a time perspective on the phylogeny and biogeography of Sundaland tree squirrels, Sundasciurus. *Mol. Phylogenet. Evol.* 55:711–720. https://doi.org/10.1016/j.ympev.2009.12.023

DeRaad DA. 2022. snpfiltr: an R package for interactive and reproducible SNP filtering. *Mol. Ecol. Resour.* 22:2443–2453. https://doi.org/10.1111/1755-0998.13618

Diamond JM. 1974. Colonization of exploded volcanic islands by birds: the supertramp strategy. *Science.* 184:803–806. https://doi.org/10.1126/science.184.4138.803

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–376. https://doi.org/10.1007/BF01734359

Flury JM, Haas A, Brown RM, et al. 2021. Unexpectedly high levels of lineage diversity in Sundaland puddle frogs (Dicroglossidae: Occidozyga Kuhl and van Hasselt, 1822). *Mol. Phylogenet. Evol.* 163:107210. https://doi.org/10.1016/j.ympev.2021.107210

Fruhstorfer H. 1898. Neue Papilio-Formen aus dem Malayischen Archipel. *Mitt. Mus. Natkd. Berl, Dtsch. Entomol.* 1898:419–430. https://doi.org/10.1002/mmnd.47918980308

Garg KM, Chattopadhyay B, Koane B, et al. 2020. Last glacial maximum led to community-wide population expansion in a montane songbird radiation in highland Papua New Guinea. *BMC Evol. Biol.* 20:82. https://doi.org/10.1186/s12862-020-01646-z

Garg KM, Chattopadhyay B, Cros E, et al. 2022. Island biogeography revisited: museomics reveals affinities of shelf island birds determined by bathymetry and paleo-rivers, not by distance to mainland. *Mol. Biol. Evol.* 39:msab340. https://doi.org/10.1093/molbev/msab340

GEBCO Bathymetric Compilation Group. 2023. The GEBCO_2023 Grid - a continuous terrain model of the global oceans and land. 2023/04/18. https://doi.org/10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b

Grismer LL, Wood , PLJr., Aowphol A, et al. 2016. Out of Borneo, again and again: biogeography of the Stream Toad genus Ansonia Stoliczka (Anura: Bufonidae) and the discovery of the first limestone cave-dwelling species. *Biol. J. Linn. Soc.* 120:371–395. https://doi.org/10.1111/bij.12886

Gu X, Fu YX, Li WH. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12:546–557. https://doi.org/10.1093/oxfordjournals.molbev.a040235

Hall R. 2009. Southeast Asia's changing palaeogeography. *Blum. - j. plant. tax. and Plant. geog.* 54:148–161. https://doi.org/10.3767/000651909X475941

Hall R. 2013. The palaeogeography of Sundaland and Wallacea since the late Jurassic. *J. Limnol.* 72:e1. https://doi.org/10.4081/jlimnol.2013.s2.e1

Hall R. 2017. Southeast Asia: new views of the geology of the Malay Archipelago. *Annu. Rev. Earth Planet. Sci.* 45:331–358. https://doi.org/10.1146/annurev-earth-063016-020633

Hall R, Morley CK 2004. Sundaland basins. In: P Clift, W Kuhnt, P Wang, et al. editors. *Continent-Ocean interactions within East Asian marginal seas*. American Geophysical Union. p. 55–85. https://doi.org/10.1029/149GM04

Hasegawa M, Kishino H, Yano T-a 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174. https://doi.org/10.1007/BF02101694

Haugum J, Low AM. 1979. *A monograph of the birdwing butterflies, troides; amphrysus & haliphron groups*. Scandinavian Science Press Ltd

Häuser CL, Holstein J, Steiner A. 2005. *The global butterfly information system*. https://doi.org/10.48580/dfrdl-394

He JW, Zhang R, Yang J, et al. 2022. High-quality reference genomes of swallowtail butterflies provide insights into their coloration evolution. *Zool. Res.* 43:367–379. https://doi.org/10.24272/j.issn.2095-8137.2021.303

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature.* 405:907–913. https://doi.org/10.1038/35016000

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* 359:183–195. https://doi.org/10.1098/rstb.2003.1388

Hinckley A, Hawkins MTR, Achmadi AS, et al. 2020. Ancient divergence driven by geographic isolation and ecological adaptation in forest dependent Sundaland tree squirrels. *Front. Ecol. Evol.* 8:1–18. https://doi.org/10.3389/fevo.2020.00208

Holbourn AE, Kuhnt W, Clemens SC, et al. 2018. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. *Nat. Commun.* 9:1584. https://doi.org/10.1038/s41467-018-03950-1

Hong Kong Biodiversity Genomics Consortium 2024. Chromosomal-level genome assembly of golden birdwing Troides aeacus (Felder & Felder, 1860). *Gigabyte* 2024:1–14. https://doi.org/10.46471/gigabyte.122

Hurvich CM, Tsai C-L. 1989. Regression and time series model selection in small samples. *Biometrika.* 76:297–307. https://doi.org/10.1093/biomet/76.2.297

Husson L, Boucher FC, Sarr A-C, et al. 2020. Evidence of Sundaland's subsidence requires revisiting its biogeography. *J. Biogeogr.* 47:843–853. https://doi.org/10.1111/jbi.13762

Jin M, Zwick A, Ślipiński A, et al. 2020. Museomics reveals extensive cryptic diversity of Australian prionine longhorn beetles with implications for their classification and conservation. *Syst. Entomol.* 45:745–770. https://doi.org/10.1111/syen.12424

Jónsson H, Ginolhac A, Schubert M, et al. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* 29:1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jukes TH, Cantor CR. 1969. Chapter 24 - Evolution of protein molecules'. In: HN Munro editor. *Mammalian protein metabolism*. Academic Press. p. 21–132

Kalyaanamoorthy S, Minh BQ, Wong TK, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods.* 14:587–589. https://doi.org/10.1038/nmeth.4285

Karin BR, Das I, Jackman TR et al. 2017. Ancient divergence time estimates in *Eutropis rugifera* support the existence of Pleistocene barriers on the exposed Sunda Shelf. *PeerJ.* 5:e3762. https://doi.org/10.7717/peerj.3762.

Kass RE, Raftery AE. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795. https://doi.org/10.1080/01621459.1995.10476572

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

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120. https://doi.org/10.1007/BF01731581.

Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. U S A.* 78:454–458. https://doi.org/10.1073/pnas.78.1.454

Klaus S, Selvandran S, Goh JW, et al. 2013. Out of Borneo: neogene diversification of Sundaic freshwater crabs (Crustacea: Brachyura: Gecarcinucidae: Parathelphusa). *J. Biogeogr.* 40:63–74. https://doi.org/10.1111/j.1365-2699.2012.02771.x

Knaus BJ, Grünwald NJ. 2017. vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17:44–53. https://doi.org/10.1111/1755-0998.12549

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics.* 15:356. https://doi.org/10.1186/s12859-014-0356-4

Kück P. 2016. *ALICUT: a Perlscript which cuts ALISCORE identified RSS. Version 2.3*. Department of Bioinformatics, Zoologisches Forschungsmuseum A. Koenig (ZMFK), Bonn, Germany, version 2.0

Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* 11:81–88. https://doi.org/10.1186/s12983-014-0081-x

Kück P, Meusemann K, Dambach J, et al. 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* 7:10–12. https://doi.org/10.1186/1742-9994-7-10

Laine VN, Sävilammi T, Wahlberg N, et al. 2023. Whole-genome analysis reveals contrasting relationships among nuclear and mitochondrial genomes between three sympatric bat species. *Genome Biol. Evol.* 15:evac175. https://doi.org/10.1093/gbe/evac175

Leaché AD, Fujita MK, Minin VN, et al. 2014. Species delimitation using genome-wide SNP data. *Syst. Biol.* 63:534–542. https://doi.org/10.1093/sysbio/syu018

Leonard JA, den Tex R-J, Hawkins MTR, et al. 2015. Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. *J. Biogeogr.* 42:871–879. https://doi.org/10.1111/jbi.12465

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Lim HC, Shakya SB, Harvey MG, et al. 2020. Opening the door to greater phylogeographic inference in Southeast Asia: comparative genomic study of five codistributed rainforest bird species using target capture and historical DNA. *Ecol. Evol*. 10:3222–3247. https://doi.org/10.1002/ece3.5964

Lohse K, Hayward A, Laetsch DR, et al. 2022. The genome sequence of the common yellow swallowtail, *Papilio machaon* (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations. *Wellcome Open Res*. 7:261. https://doi.org/doi.org/10.12688/wellcomeopenres.18119.1

Manawatthana S, Laosinchai P, Onparn N, et al. 2017. Phylogeography of bulbuls in the genus Iole (Aves: Pycnonotidae). *Biol. J. Linn. Soc*. 120:931–944. https://doi.org/10.1093/biolinnean/blw013

Mason VC, Helgen KM, Murphy WJ. 2019. Comparative phylogeography of forest-dependent mammals reveals paleo-forest corridors throughout Sundaland. *J. Hered*. 110:158–172

Matsuka H. 2001. トリバネチョウ生態図鑑: *Natural history of birdwing butterflies*, 松香出版

Matzke NJ. 2018. BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts. Version 1.1.1. GitHub published on GitHub on November 6, 2018

McKenna A, Hanna M, Banks E, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*. 210:719–731. https://doi.org/10.1534/genetics.118.301336

Meng G, Li Y, Yang C, et al. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res*. 47:e63. https://doi.org/10.1093/nar/gkz173

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc*. 2010:pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Minh BQ, Schmidt HA, Chernomor O, et al. 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

Mirarab S, Nakhleh L, Warnow T. 2021. Multispecies coalescent: theory and applications in phylogenetics. *Annu. Rev. Ecol. Evol. Syst*. 52:247–268. https://doi.org/10.1146/annurev-ecolsys-012121-095340

Mittermeier R, Robles-Gil P, Hoffmann M, 2005. *Hotspots revisited: earth's biologically richest and most endangered ecoregions*. Conservation International

Mutanen M, Wahlberg N, Kaila L. 2010. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proc. R. Soc. Lond. Ser. B: Biol. Sci*. 277:2839–2848. https://doi.org/10.1098/rspb.2010.0392

Myers N, Mittermeier RA, Mittermeier CG, et al. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858. https://doi.org/10.1038/35002501

Natural History Museum. 2014. *Dataset: Collection specimens. Resource: Specimens*. Natural History Museum Data Portal (data.nhm.ac.uk)

Ng DYJ, Švejcarová T, Sadanandan KR, et al. 2021. Genomic and morphological data help uncover extinction-in-progress of an unsustainably traded hill myna radiation. *Ibis*. 163:38–51. https://doi.org/10.1111/ibi.12839

Nishida R, Weintraub JD, Feeny P, et al. 1993. Aristolochic acids from *Thottea* spp. (Aristolochiaceae) and the osmeterial secretions of Thottea-feeding Troidine swallowtail larvae (Papilionidae). *J. Chem. Ecol*. 19:1587–1594. https://doi.org/10.1007/BF00984899

Ohya T. 2001. Checklist of birdwing butterfly. In: H Matsuka editor. *Natural history of birdwing butterflies*. Matsuka Shuppan. p. 346–348

Ortiz EM. 2019. *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis*. Version 2.0. Zenodo

Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*. 26:419–420. https://doi.org/10.1093/bioinformatics/btp696

Pozzi L, Penna A, Bearder SK, et al. 2020. Cryptic diversity and species boundaries within the *Paragalago zanzibaricus* species complex. *Mol. Phylogenet. Evol*. 150:106887.

Puillandre N, Brouillet S, Achaz G. 2021. ASAP: assemble species by automatic partitioning. *Mol. Ecol. Resour.* 21:609–620. https://doi.org/10.1111/1755-0998.13281.

Puillandre N, Lambert A, Brouillet S, et al. 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol*. 21:1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

QGIS.org. 2025. QGIS Version 3.40.5. QGIS Association. http://www.qgis.org

Qu Y, Luo X, Zhang R, et al. 2011. Lineage diversification and historical demography of a montane bird *Garrulax elliotii*—implications for the Pleistocene evolutionary history of the eastern Himalayas. *BMC Evol. Biol*. 11:174. https://doi.org/10.1186/1471-2148-11-174

R Core Team. 2024. *R: a language and environment for statistical computing. Version 4.4.2*. R Foundation for Statistical Computing

Raes N, Cannon CH, Hijmans RJ, et al. 2014. Historical distribution of Sundaland's Dipterocarp rainforests at quaternary glacial maxima. *Proc. Natl. Acad. Sci. U S A*. 111:16790–16795. https://doi.org/10.1073/pnas.1403053111

Rambaut A, Drummond AJ, Xie D, et al. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol*. 67:901–904. https://doi.org/10.1093/sysbio/syy032

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*. 164:1645–1656. https://doi.org/10.1093/genetics/164.4.1645

Raxworthy CJ, Smith BT. 2021. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol*. 36:1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Reboud EL, Nabholz B, Chevalier E, et al. 2023. Genomics, population divergence, and historical demography of the world's largest and endangered butterfly, the queen Alexandra's birdwing. *Genome Biol. Evol*. 15: 1–19. https://doi.org/10.1093/gbe/evad040

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

Ritchie AM, Lo N, Ho SYW. 2017. The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. *Syst. Biol*. 66:413–425. https://doi.org/10.1093/sysbio/syw095

Rowe KC, Singhal S, Macmanes MD, et al. 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Mol. Ecol. Resour.* 11:1082–1092. https://doi.org/10.1111/j.1755-0998.2011.03052.x

Sarr AC, Husson L, Sepulchre P, et al. 2019. Subsiding Sundaland. *Geology*. 47:119–122. https://doi.org/10.1130/G45629.1

Sholihah A, Delrieu-Trottin E, Sukmono T, et al. 2020. Disentangling the taxonomy of the subfamily Rasborinae (Cypriniformes, Danionidae) in Sundaland using DNA barcodes. *Sci. Rep*. 10:2818. https://doi.org/10.1038/s41598-020-59544-9

Soubrier J, Steel M, Lee MSY, et al. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol*. 29:3345–3358. https://doi.org/10.1093/molbev/mss140

Stange M, Sánchez-Villagra MR, Salzburger W, et al. 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. *Syst. Biol*. 67:681–699. https://doi.org/10.1093/sysbio/syy006

Staudinger O. 1892. Ornithoptera Andromache n. sp. *Dtsch. Entomol. Z* 5:393–394

Stoltz M, Baeumer B, Bouckaert R, et al. 2021. Bayesian inference of species trees using diffusion models. *Syst. Biol*. 70:145–161. https://doi.org/10.1093/sysbio/syaa051

Suárez-Villota EY, Quercia CA, Díaz LM, et al. 2018. Speciation in a biodiversity hotspot: phylogenetic relationships, species delimitation, and divergence times of Patagonian ground frogs from the *Eupsophus roseus* group (Alsodidae). *PLoS ONE*. 13:e0204968. https://doi.org/10.1371/journal.pone.0204968

Swofford D. 2002. PAUP. Phylogenetic analysis using parsimony (and other methods). Version 4.0b10

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol*. 10:512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tavaré S. 1986. Location providence some probabilistic and statistical problems in the analysis of DNA sequences'. In: RM Miura, editor. *Lectures on mathematics in the life sciences*. American Mathematical Society. p. 57–86

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol*. 21:3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Tsukada E, Nishiyama Y. 1982. *Butterflies of the South East Asian islands. English Edition*. Plapac Co. Ltd

Twort VG, Minet J, Wheat CW, et al. 2021. Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life. *Syst. Entomol*. 46:926–937. https://doi.org/10.1111/syen.12503

Vasimuddin M, Misra S, Li H, et al. 2019. Efficient architectureaware acceleration of bwa-mem for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE. p. 314–324. https://doi.org/10.1109/IPDPS.2019.00041

Von Rintelen K, Arida E, Häuser C. 2017. A review of biodiversity-related issues and challenges in megadiverse Indonesia and other Southeast Asian countries. *Res Ideas Outcomes*. 3:e20860. https://doi.org/10.3897/rio.3.e20860

Voris HK. 2000. Special paper 2: maps of Pleistocene Sea Levels in Southeast Asia: shorelines, river systems and time durations. *J. Biogeogr*. 27:1153–1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x

Wallace AR. 1855. XVIII.— On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History* 16: 184–196. https://doi.org/10.1080/037454809495509

Wallace AR. 1869. *The Malay archipelago: the land of the Orang-utan, and the bird of paradise. A narrative of travel, with studies of man and nature*. Macmillan & Co. Ltd

Wallace AR. 1911. '*XVII Borneo, Java, and the Philippines*'. *In: Island life or the phenomena and causes of insular faunas and floras, including a revision and attempted solution of the problem of geological climates 3rd*. Macmillan & Co. Ltd. p. 373–385

Whitten T, Soeriaatmadja RE, Afiff SA. 1997. *The ecology of Java and Bali*. Oxford University Press

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag

Willerslev E, Cooper A. 2005. Review paper. ancient DNA. *Proc. Biol. Sci*. 272:3–16. https://doi.org/10.1098/rspb.2004.2813

Williams EW, Gardner EM, Harris , RIII, et al. 2017. Out of Borneo: biogeography, phylogeny and divergence date estimates of Artocarpus (Moraceae). *Ann. Bot*. 119:611–627. https://doi.org/10.1093/aob/mcw249

Woodruff DS. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodivers. Conserv*. 19:919–941. https://doi.org/10.1007/s10531-010-9783-3

Wu MY, Lau CJ, Ng EYX, et al. 2022. Genomes from historic DNA unveil massive hidden extinction and terminal endangerment in a tropical Asian songbird radiation. *Mol. Biol. Evol*. 39:msac189. https://doi.org/10.1093/molbev/msac189

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol*. 39:306–314. https://doi.org/10.1007/BF00160154

Yang Z. 1995. A space-time process model for the evolution of DNA sequences. *Genetics*. 139:993–1005. https://doi.org/10.1093/genetics/139.2.993

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

Zhang J, Kobert K, Flouri T, et al. 2014. PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics*. 30:614–620. https://doi.org/10.1093/bioinformatics/btt593

Zharkikh A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol*. 39:315–329. https://doi.org/10.1007/BF00160155.