Clarifying the phylogeny and systematics of the recalcitrant tribe Leptocircini (Lepidoptera: Papilionidae) with whole-genome data

Eliette L Reboud, Benoit Nabholz, Emmanuelle Chevalier, Bérénice J Lafon, Marie-ka Tilak, Carlos G C Mielke, Adam M Cotton, Fabien L Condamine

Clarifying the phylogeny and systematics of the recalcitrant tribe Leptocircini (Lepidoptera: Papilionidae) with whole-genome data

Authors

Eliette L. Reboud[1]*, Benoit Nabholz[1,2], Emmanuelle Chevalier[1], Bérénice J. Lafon[1], Marie-ka Tilak[1], Carlos G. C. Mielke[3], Adam M. Cotton[4] & Fabien L. Condamine[1]*

Affiliations

[1] *Institut des Sciences de l'Evolution de Montpellier (Université Montpellier | CNRS | IRD | EPHE), Place Eugène Bataillon, 34095 Montpellier, France*

[2] *Institut Universitaire de France (IUF), Paris, France*

[3] *Caixa postal 1206, 84.145-000 Carambeí, Paraná, Brazil*

[4] *86/2 Moo 5, Tambon Nong Kwai, Hang Dong, Chiang Mai, Thailand*

Corresponding authors (*)

Eliette L. Reboud: eliette.reboud@umontpellier.fr

Fabien L. Condamine: fabien.condamine@gmail.com

Running head

Phylogenomic of Leptocircini (Papilionidae)

Abstract

Leptocircini is a dazzling tribe of Papilionidae, including dragontails, kite swallowtails, and swordtails. This tribe is widely distributed, notably throughout the tropics of Africa, Southeast Asia, and the Americas, making it a fascinating model in evolutionary biology. However, despite accounting for 25% of the global swallowtail butterfly diversity, Leptocircini have been surprisingly neglected in phylogenetic analyses. This has left unanswered questions about their taxonomy and systematics. Here, we present a new taxonomic working list for Leptocircini, featuring 162 valid species. Using a combination of long and short reads data, we produced five new reference genomes, and we generated highly covered and scaffolded whole genomes for 148 individuals to infer densely sampled phylogenetic hypotheses. Based on mitochondrial or thousands of nuclear genes and multiple phylogenetic approaches, a robust phylogenomic tree is recovered, representing ~90% of the known species, which allowed examination of several key phylogenetic hypotheses. We found the monotypic genus *Protographium* Munroe to be sister of genus *Graphium* Scopoli. Additionally, we found that subgenus *Paranticopsis* Wood-Mason and de Nicéville is nested within subgenus *Pathysa* Reakirt, which we found is likely attributed to an ancient gene flow. We therefore synonymize *Paranticopsis*, **syn.rest.** To keep a consistent approach to subgeneric classification across the tribe and family, we divided genus *Eurytides* Hübner into three subgenera: *Mimoides* Brown, *Eurytides sensu stricto*, and *Protesilaus* Swainson. This led to several taxonomic implications: *Asiographium* Möhn, **syn.rest.**, *Boreographium* Grishin, **syn.n.**, *Hyalaus* Grishin, **syn.n.**, and *Neographium* Möhn, **syn.n.** are synonymized with *Eurytides* (*Mimoides)*; and *Eurygraphium* Möhn, **syn.rest.** is synonymized with *Eurytides (Eurytides)*. Our analyses finally raised concerns about potential taxonomic inflation in two species-groups within *Graphium* and *Eurytides (Protesilaus)*. This study illuminates the clade's evolutionary history and paves the way for further research on this diverse group of charismatic butterflies.

Keywords: *Eurytides*, gene flow, *Graphium*, incomplete lineage sorting, phylogenomics, reference genome, scaffolding.

Introduction

The swallowtail butterflies (Papilionidae) of the tribe Leptocircini Kirby include species with a high diversity of wing shapes, sizes and colors, such as dragontails, kite swallowtails and swordtails. They comprise about 160 species distributed worldwide, which represent about a quarter of all Papilionidae (*ca*. 600 recognized species; Zakharov et al., 2004; Haüser et al., 2005; Condamine et al., 2012; Nakae, 2021). The phylogenetic relationships within tribe Leptocircini remain highly uncertain and have never been the subject of comprehensive and thorough phylogenetic studies, contrary to some well-studied swallowtail tribes (Papilionini: Ae, 1979; Hancock, 1983; Igarashi, 1984; Aubert et al., 1999; Caterino and Sperling 1999; Reed and Sperling 1999; Yagi et al., 1999; Zakharov et al., 2004; Condamine et al., 2013a, 2013b; Lewis et al., 2015; Wu et al., 2015; Owens et al., 2017; Condamine et al., 2023; Troidini: Silva-Brandão et al., 2005; Condamine et al., 2012, 2013a, 2015; Parnassiini: Nazari et al., 2007; Michel et al., 2008; Condamine et al., 2018; He et al., 2023; Tian et al., 2023).

Nevertheless, Leptocircini has been the subject of several morphological studies and revisions based mainly on genitalia and wing venation in the 1960s-2000s (e.g. Munroe, 1961; Hancock, 1983; Miller, 1987; Smith and Van-Wright, 2001; Makita et al., 2003; Hancock 2006). Recently, the genus *Graphium* Scopoli has been the focus of molecular work in localized parts of the phylogeny to describe new species, often based on the mitochondrial COI gene or a few genes (e.g. *Graphium (Pazala)* Moore: Hu et al., 2018, 2019; Zhang et al., 2020; *Graphium* (*Graphium)*: Page and Treadaway, 2014; Cotton et al., 2022). Leptocircini has also appeared in several attempts to reconstruct the global phylogeny of Papilionidae, but usually with a low sampling fraction (e.g. Makita et al., 2003 [25%]; Simonsen et al., 2011 [6.8%]; Page and Treadaway 2014 [24%]; Zhang et al., 2019 [29%]; Zhang et al., 2021 [6.8%]) and none of these phylogenies are sufficiently robust to to enable a firm classification to be constructed. The most comprehensive phylogenetic work dealing with family Papilionidae and

including Leptocircini was conducted by Allio et al. (2021a) and used seven mitochondrial and nuclear genes. However, the sampling of Leptocircini only included 42% of the tribe and the phylogenetic relationships were not robust enough to discuss the topology and its consistency with the existing literature. This apparent neglect is quite surprising for a group as studied and popular as Papilionidae, but probably stems from multiple explanations. Indeed, Leptocircini are known to be particularly difficult to collect in the field due to their still largely unknown behaviour and traits, seasonality, the speed and height of their flight (Collins and Morris, 1985; Smith and Vane-Wright, 2001; Nakae, 2021).

Leptocircini are distributed worldwide

Most of the species richness of Leptocircini is contained in genus *Graphium*, which alone includes ~110 of the 160 species of Leptocircini. As a typical component of the Old-World tropics, *Graphium* is composed of five subgenera including a species-rich African subgenus (*Arisbe* Hübner) and four Indomalayan-Australasian subgenera (*Graphium, Paranticopsis* Wood-Mason and de Nicéville, *Pathysa* Reakirt and *Pazala*) (**Fig. 1**). The Australian monotypic genus *Protographium* Munroe is thought to be sister to *Graphium* based on its morphology (Munroe, 1961; Munroe and Erlich 1960; Nakae 2021). Genus *Eurytides* Hübner is the second richest genus in Leptocircini with >40 species, and is exclusively found in the Americas. The systematics and taxonomy of both *Graphium* and *Eurytides* have undergone many changes and uncertainties, with numerous subgenera or genera having been proposed (see taxonomic details in the following paragraph). Here, we test all existing or unresolved subgeneric names, including latest clade-level studies (e.g. Smith and Vane-Wright, 2001; Mohn, 2002; Zhang et al., 2024) and global family-level assessments (Tyler et al., 1994; Nakae, 2021). It corresponds to five subgenera for *Graphium*: *Arisbe*, *Graphium*, *Paranticopsis*, *Pathysa* and *Pazala* (see details hereafter) and nine subgenera for *Eurytides*: *Asiographium*

Möhn, *Bellerographium* Möhn, *Boreographium* Grishin, *Eurygraphium* Möhn, *Eurytides* Hübner, *Hyalaus* Grishin, *Mimoides* Brown, *Neographium* Möhn and *Protesilaus* Swainson.

External to *Protographium* and these two diverse genera are two smaller genera containing three species each: (1) *Iphiclides* Hübner is the only Leptocircini genus with an Eurasian distribution; and (2) *Lamproptera* Gray, commonly called dragontail butterflies, is a South-east Asian clade with a divergent and peculiar morphology (**Fig. 1**).

Leptocircini have a complex and still unstable taxonomic history.

Graphium has been the subject of several morphological and molecular studies, attempting to understand its classification and evolutionary relationships. To our knowledge, all studies have shown genus *Graphium* to be monophyletic (e.g. Makita et al., 2003; Allio et al., 2021a). Despite several studies that upgraded some subgenera to genera (e.g. Niculescu, 1977; Igarashi, 1984; D'Abrera, 1982; Page and Treadaway, 2014), the subgeneric classification of genus *Graphium* is now commonly accepted (e.g. Munroe 1961; Collins and Morris 1985; Miller 1987; Parsons 1998; Racheli and Cotton 2009; Hardy and Lawrence 2017; Nakae 2021). However, the classification and validity of subgenera within *Graphium* remain highly controversial, despite multiple morphological and molecular analyses. *Graphium (Pazala)* is usually considered as monophyletic and valid and placed as the sister group to the other *Graphium* subgenera (Hancock, 1983; Miller 1987), but sometimes not (Makita et al., 2003; Allio et al., 2021a). However, *Graphium (Arisbe)*, *Graphium (Graphium)*, *Graphium (Paranticopsis)* and *Graphium (Pathysa)* have a more complex taxonomic history.

Saigusa et al. (1982) considered Asian *Graphium (Graphium)* as monophyletic with three species-groups: *sarpedon*, *agamemnon*, and *eurypylus*. However, no species of *Graphium (Arisbe)* were included in their analysis, which was later considered as a limitation, as Hancock

(1993) found that tailed *Arisbe* species were more closely related to *Graphium (Graphium)*. Hancock (1993) therefore divided *Graphium (Arisbe)* into non-tailed species (*Arisbe*) and tailed species (*antheus*-like) and included these latter in *Graphium (Graphium)*. Smith and Vane-Wright (2001) tested this morphological hypothesis in a cladistic framework, but were 'unable to resolve African and Oriental species-groups'. On the other hand, several studies, instead of questioning the monophyly *Graphium (Arisbe)*, have rather considered the monophyly of *Graphium (Graphium)* as doubtful, and particularly the placement of the *eurypylus* species-group. Indeed, the monophyly of *Graphium (Graphium)* was not supported in Makita et al. (2003), the *eurypylus* species-group being sister to *Graphium (Arisbe)*. Page and Treadaway (2003a) suggested a narrower definition of *Graphium (Graphium)*, including only certain species-groups (*sarpedon*, *codrus*, *macleayanus*, *wallacei* and *agamemnon* groups) and established a generic status for *Arisbe* and subgeneric status for *eurypylus*, *Pazala*, and *Pathysa* (all included in this broader genus *Arisbe*). Recently, Allio et al. (2021a) found a polyphyletic *Graphium (Graphium)*, but the *eurypylus* species-group was found to be sister to *Graphium (Pazala)* and the other *Graphium (Graphium)* species-groups, and not sister to *Graphium (Arisbe)*. Recent studies based on molecular data have been unable to resolve the issue, perhaps due to limited species representation.

Several studies considered *Graphium (Paranticopsis)* to be included in *Graphium (Pathysa)* (Munroe and Erlich, 1960; Hancock, 1983; Miller, 1987). Smith and Vane-Wright (2001) used *Graphium (Paranticopsis)* as a valid subgenus, but the study only included *G. macareus*, which showed relatedness to the African clade *Graphium (Arisbe)*, rather than *Graphium (Pathysa)* and *Graphium (Pazala)*. Later, Makita et al. (2003), Page and Treadaway (2003a, 2003b, 2004, 2014), Wilson et al. (2014) and Allio et al. (2021a) observed polyphyletic patterns or unresolved relationships between *Graphium (Paranticopsis)* and *Graphium (Pathysa)*, but their phylogenies were poorly supported.

A sister-group relationship between *Lamproptera* and *Graphium* was proposed by Hancock (1983), Igarashi (1984) and Miller (1987) following Ehrlich (1957) on the basis of larval and pupal morphological characters (Racheli and Cotton 2009). This relationship was recovered based on morphological characters in Smith and Vane-Wright (2001) (but without any *Eurytides* included) and based on morphological and molecular data in Page and Treadaway (2014). However, the relationship between *Lamproptera* and *Graphium* was described as unstable by Makita et al. (2003), as their molecular data showed that *Eurytides* and *Iphiclides* were more closely related to *Graphium* than *Lamproptera*. In addition, they did not formally question the 'traditional' *Lamproptera* + *Graphium* group that was proposed at the time. Simonsen et al. (2011) found that *Graphium* instead forms a sister group to the Neotropical Leptocircini (*Eurytides* , see below) and that *Iphiclides* and *Lamproptera* were early-diverging clades of Leptocircini, but with a non-robust relationship. Finally, Allio et al. (2020a, 2021a), found *Lamproptera* to be sister to *Iphiclides*, forming a sister clade to all other Leptocircini, but again with rather weak support even with genomic data.

The Neotropical Leptocircini have never been the focus of any well sampled study. Although they have appeared in several previously cited studies focused on *Graphium* (e.g. Makita et al., 2003; Page and Treadaway 2014), they were only represented by a few species and mainly included as outgroups. They also appear in several general studies of Papilionidae (Simonsen et al., 2011; Allio et al., 2021a), where up to 18 species were sampled (~44%, Allio et al., 2021a). Despite limited sampling, these studies suggest that Neotropical Leptocircini form a monophyletic group, though their internal classification remains complex.

Genus *Eurytides* was first described by Hübner in 1821. It was subsequently placed within genus *Papilio* (Rotschild and Jordan, 1906; Jordan, 1907-1908) or *Graphium* (Ford, 1944). Munroe (1961) reestablished *Eurytides* as a genus and divided it into *Eurytides (Protesilaus)* and *Eurytides (Eurytides)*. *Eurytides* was long thought to be closely related to the

Australian *Protographium leosthenes* (which Munroe [1961] also erected as a monospecific genus) and some *Eurytides* species were even placed in *Protographium* by Brown (1991) and Tyler et al. (1994). The two subgenera *Protesilaus* and *Eurytides* as defined by Munroe (1961), were later elevated to genera by Hancock (1983), downgraded again to subgenera by Miller (1987), and restored to genera by Brown (1991). Brown (1991) also described *Mimoides* as a genus including the Neotropical mimetic Leptocircini species that were previously placed by Munroe (1961) in *Eurytides (Protesilaus)*. Lately, Zhang et al. (2019) proposed downgrading *Mimoides* to a subgenus of *Eurytides* to account for its close relationship with *E. marcellus*.

Möhn (2002) removed Neotropical species from genus *Protographium* and proposed two new subgenera: *Eurytides* (*Neographium)* (*philolaus* species-group) and *Eurytides (Asiographium)* (monospecific, type-species *E. asius*). He also described two new subgenera of *Eurytides*: *Eurytides (Bellerographium)* (monospecific, type-species *E. bellerophon*) and *Eurytides (Eurygraphium)* (*thyastes* species-group). Lamas (2004) synonymized some of these names in *Protographium* (*Asiographium, Eurygraphium, Neographium*) or *Eurytides* (*Bellerographium)*. However, since the type specimen of *Protographium* is the Australian *P. leosthenes*, which appears to be unrelated to *Eurytides* (Zhang et al. 2019), this synonymy is likely incorrect. Finally, Zhang et al. (2021) proposed the monospecific subgenus *Eurytides (Boreographium)* (monospecific, type-species *E. marcellus*), and Zhang et al. (2024) proposed the monospecific subgenus *Eurytides (Hyalaus)* (monospecific, type-species *E. epidaus*), based on a limited sampling, to account for the proximity of these species with *Eurytides (Mimoides)* rather than with the *Eurytides (Neographium)* species.

A well-sampled phylogenomic dataset to resolve the tree of Leptocircini

To date, all phylogenetic studies of Leptocircini lack a high species sampling or are limited to morphological or a few molecular markers, which have resulted in highly unstable classifications and many unresolved relationships. It has long been shown that phylogenetic

studies based on a handful of molecular markers can result in poorly supported phylogenetic relationships due to a limited number of phylogenetically informative characters (Wiens and Penkrot 2002; Funk and Omland 2003; Wiens et al., 2010; Ross 2014; Mutanen et al., 2016). On the other hand, genomic data can alleviate some of the phylogenetic issues by expanding the number of characters from a few thousand to hundreds of thousands (Delsuc et al., 2005; Philippe and Blanchette 2007; Pennisi 2008), and help to produce better resolved phylogenetic trees and particularly robust backbones. For instance, using genome-scale data associated with a reference genome of *Papilio xuthus* (sequenced by Li et al., 2015), Allio et al. (2020a) were able to retrieve thousands of single-copy genes and inferred a robust phylogenetic backbone for swallowtail butterflies. Although this study lacked the sampling to constitute a complete species-level phylogeny, it represents a promising approach to combining reference genomes and fragmented whole genomes. Indeed, in the era of high-quality complete genomes that is flourishing (Formenti et al., 2022), the swallowtail butterflies are no exception. However, genus *Papilio* remains the most studied, with >30 whole genomes published since 2015 (e.g. Lu et al., 2019 and references therein), while the first high-quality genomes of Leptocircini were only published recently (*Lamproptera curius*, He et al., 2022; *Iphiclides podalirius*, Mackintosh et al., 2022). However, there is no reference genome for *Graphium* available yet, even though this genus represents the second most speciose genus in Papilionidae. Furthermore, a reference genome for *Graphium* can provide crucial data for the genomics of swallowtail butterflies since the expected size of their genomes is larger than that of phylogenetically related genera (Allio et al., 2020a).

In this study, we first establish a list of valid Leptocircini species based on previous molecular and morphological studies and then infer a robust and near complete species-level phylogeny for the tribe. We sequenced and assembled five new reference genomes of Leptocircini, including four *Graphium* species, and 136 *de novo* shotgun whole genomes,

which together allow us to extract thousands of nuclear genes and mitogenomes for around 88% of the total specific diversity of Leptocircini. Our study aims to provide a reference phylogenetic framework for assessing the monophyly of genera and subgenera as well as species relationships within Leptocircini. For this objective, we include all enigmatic and long-debated groups and species to be robustly placed in the swordtail tree of life, such as *Protographium leosthenes*, *Graphium phidias* and other species that will be key for testing diversification and biogeographic hypotheses (e.g. *Eurytides marcellus*), the evolution of mimicry (e.g. *Eurytides (Mimoides)*, *Graphium* (*Paranticopsis*)) and host-plant associations (e.g. *Graphium (Pazala)*, *Lamproptera*, *Iphiclides*), as well as rare and understudied species (e.g. *Graphium* (*Arisbe)*, *Eurytides*). On the basis of this dataset, we also explored the hypothesis of gene flow and incomplete lineage sorting to explain topological discordances between the species tree and gene trees.

Material and Methods

Species list, taxon sampling and sequencing strategy

We first established a taxonomic working list of Leptocircini species (**Table 1**). This species list takes into account some recent taxonomic lists and the results of recent studies (Hu et al., 2018, 2019; Zhang et al., 2020; Nakae, 2021; Cotton et al., 2022; Huang 2023; and this study). Due to the limited recent general phylogenetic studies conducted on Leptocircini, it is likely that they will continue to receive attention as a model clade for taxonomic studies, and we expect that many species boundaries will be tested with new genomic data, or at a populational level, which could ultimately revise this taxonomic working list.

Of the 162 species currently featured in the new taxonomic list presented here (including all of ‘valid’, ‘unconfirmed’ and ‘doubtful’ status), we sampled 147 species representing 90%

of the total diversity as follows: 143 species from this study (150 individuals, including 148 whole-genomes, using a combination of fresh and museum specimens), plus 4 species (mitogenomic data) from GenBank (**Table 1**). However, the total number of species might evolve with further systematic studies, as the validity of some species is still debated (e.g. conflicts within *Graphium (Pazala)*, Hu et al., 2018, 2019 *versus* Huang 2023). The missing species were essentially concentrated in *Graphium (Pazala)* (four missing) and *Graphium (Arisbe)* (six missing), some of them being extremely rare or of doubtful status (**Table 1**). We added 11 outgroup species representing several swallowtail genera to root the Leptocircini tree based on large-scale phylogenetic studies of Papilionidae (Condamine et al., 2012; Allio et al., 2020a, 2021): *Baronia brevicornis*, *Papilio xuthus*, *Parnassius apollo*, *Ornithoptera alexandrae*, *Teinopalpus imperialis*, *Meandrusa payeni*, *Bhutanitis thaidina*, *Sericinus montela*, *Luehdorfia chinensis*, *Hypermnestra helios* and *Archon apollinus*. The complete taxon sampling and accession numbers are provided in **Tables S1, S2**.

Sequencing was performed in ~50x short-reads Illumina sequencing based on previous estimates of genome sizes (Allio et al., 2020b; He et al., 2022). Among these sampled species, five reference genomes were produced with a combination of long reads and short reads data: *Iphiclides podalirius* (collected in the wild), *G. antheus* for *Graphium (Arisbe)*, *G. agamemnon* and *G. doson* for subgenus *Graphium (Graphium)*, and *G. antiphates* for subgenus *Graphium (Pathysa)*, all obtained from rearing.

DNA extractions and library preparations for whole genome sequencing

For museum and collection specimens (taking legs, rarely abdomens), DNA extractions were performed with the DNAeasy Blood and Tissue kit (QIAGEN), and library preparations followed the methods of Meyer and Kirscher (2010), but with slight modifications as described in Tilak et al. (2015). DNA quality and concentration and fragment length varied markedly between specimens. Final DNA purity and concentrations were measured using both Nanodrop

(Thermo Fisher, USA) and Qubit (Thermo Fisher, USA) and resulting libraries were analysed for size distribution by Agilent 2200 TapeStation. Illumina 150 bp paired-end sequencing was run on a NovaSeq 6000 instrument to obtain a genome depth-of-coverage of about 50x, hence varying from 20 to 60 Gb per library depending on the genus.

For reference genomes, DNA extractions were performed with butterflies killed and stored in a freezer at -20°C without any additional preservation product. Tissues from the thorax were used to extract high-molecular weight DNA. Following Reboud et al. (2023) who tested two different extraction methods, we used the Qiagen genomic DNA kit to obtain a better 260/230 ratio as estimated with Nanodrop assays guaranteeing DNA purity for long-read sequencing with Oxford Nanopore Technology (ONT). Final DNA purity and concentrations were measured using both Nanodrop (Thermo Fisher, USA) and Qubit (Thermo Fisher, USA). Whole-genome libraries were prepared using the resulting high-molecular-weight DNA as input for the Nanopore LSK-109 ligation kit (Oxford Nanopore Technologies, UK) following the manufacturer's protocol. Long-read (LR) sequencing was performed on a GridION device with two to six R9.4.1 flow cells. Remaining DNA extractions were sent to Novogene Europe (Cambridge, UK) for two short-read Illumina library preparations per genome. Libraries were generated using NEBNext DNA Library Prep Kit following manufacturer's recommendations, and indices were added to each library. Genomic DNA was randomly fragmented to a size of 350 bp by shearing, then DNA fragments were end-polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and the resulting libraries were analysed for size distribution by Agilent 2100 BioAnalyzer and quantified using real-time PCR. Since the genome sizes for *Graphium* was estimated to be about 1 Gb and *Iphiclides* about 390 Mb (Allio et al., 2020a), Illumina 150 bp paired-end sequencing was run on a NovaSeq 6000

297 instrument to obtain about 100 and 40 Gb, respectively, corresponding to a genome depth-of-
298 coverage of ~100x after combining the two libraries.

299

300 ***Assembling reference genomes***

301 For each reference genome (*G. agamemnon, G. antheus, G. doson, G. antiphates, I. podalirius*),
302 raw LR sequence data (fast5 files) were basecalled using Guppy 5.0.15 (Oxford Nanopore
303 Technology) with the super-high accuracy mode and a quality control of 10 (min_score 10).
304 Sequencing adapters were trimmed using Porechop 0.2.3
305 (https://github.com/rrwick/Porechop). Draft genome assemblies were performed with the LR
306 assembler Flye 2.8.3 (Kolmogorov et al., 2019) with default options. For *Graphium* genomes,
307 duplicated haplotigs and heterozygous contig overlaps were removed from the draft assembly
308 using purge_dups 1.2.5 (Guan et al., 2020) based on the LR depth (**Fig. 2**).

309 The Illumina raw reads were cleaned, filtered, and paired using fastp 20.0 (Chen et al.,
310 2018) with default options (**Fig. 2**). To improve base accuracy and reduce assembly errors, the
311 Flye draft assemblies were polished using short-reads (SR) with POLCA (Zimin and Salzberg,
312 2020) implemented in MaSuRCA 4.0.1 (Zimin et al., 2013). Assembly statistics were then
313 assessed using the gVolante2 platform (Nishimura et al., 2017) to retrieve the number and size
314 of contigs, the presence, completeness and duplication of BUSCO genes of the Lepidoptera
315 *odb10* database (Manni et al., 2021). We checked for possible contaminations using BlobTools
316 1.1.1 (Laetsch and Blaxter, 2017) set to the *ncbi* and *diamond* databases. We found no evidence
317 of artificial contamination coming from laboratory manipulation, but some contigs were clearly
318 identified as belonging to exogenous organisms such as symbionts. We removed all contigs
319 that belonged to plants or symbionts (*Wolbachia*, Ascomyota, Acetobacteraceae, Bacteroidota,
320 and Microsporidia).

We compared our genome of *I. podalirius* with the one published by Mackintosh et al. (2022) using the option *-cx asm5* of Minimap2 (Li, 2018). We excluded all alignments smaller than 10 kb and those with a mapping quality below 60 (*i.e.* the maximum) and we then calculated the average gap-compressed divergence (*i.e.*, counting consecutive gaps as one) provided by the program. We computed the BUSCO scores of both genomes using *compleasm* (Huang and Li, 2023)

Assembling shotgun whole genomes

The Illumina SR were filtered and paired with fastp 19.5-21.0 (Chen et al., 2018) using the default options (**Fig. 2**). The assemblies were performed by MEGAHIT 1.2.7-1.2.9 (Li et al., 2015) and submitted to the gVolante platform to obtain assembly statistics as well as assessing the proportion of genes found from the *odb10* database of Lepidoptera (BUSCO 5, Manni et al., 2021). The gVolante platform retrieves BUSCO genes classed in 'single copy', 'multicopy' and 'fragmented' categories. To improve the contiguity and the number of BUSCO genes recovered in these draft assemblies, they were scaffolded with the scaffold option of RagTag (Alonge et al., 2021) when a close reference genome was available (same species-group or same subgenus). Species of the genera *Eurytides* and *Protographium* did not have any close reference genome. In that case, only assemblies with less than 25% BUSCO completeness were scaffolded on the reference *Graphium agamemnon* to retrieve more BUSCO genes. To compare draft genome assemblies with scaffolded genome assemblies, the resulting assemblies were again submitted to gVolante. The BUSCO single copy genes in fasta nucleotide file of each individual were retrieved using BUSCOMP 0.13 (Edwards, 2019).

Assembling mitogenomes and phylogenetic reconstruction

For reference genomes, each *Graphium* individual, LR were mapped with Minimap 2.17 (Li 2018) on the mitogenome of *G. xenocles* (MZ394042) for *G. agamemnon*, *G. antiphates*, *G. doson*, and on the mitogenome of *G. leonidas* (FC1205, this study) for *G. antheus*. The reads that mapped with the reference were filtered by quality via SAMtools (Li et al., 2009) (*view -q 30*). For each individual, a subset of reads was created so that mitogenomes would have an expected depth of coverage of 200× (3.6 Mb). We used Flye 2.8.3 to assemble the mitogenomes and the resulting assemblies were given to MitoFinder 1.4 (Allio et al., 2020b) to annotate and extract thirteen protein-coding genes (ATP6, ATP8, COX1-3, CYTB, ND1-4, ND4L, ND5-6) and 2 non-coding genes (rrnL, rrnS). The complete mitogenome of *Iphiclides podalirius* was identified as a contig of the whole-genome assembly and was also selected for annotation by MitoFinder.

For shotgun assemblies, MEGAHIT draft-assemblies of each species were given to MitoFinder to extract and annotate the mitogenome. Some assemblies only displayed fragmented mitogenomes, pseudogenes, fungi contamination or cross-contamination. When the identification of 'real' mitogenomic contigs was not possible manually, GetOrganelle (Jin et al., 2020) was used with a close reference for mapping to reconstruct complete, or less fragmented, mitogenomes. This method was only used on difficult individuals as the computation time of GetOrganelle is longer and not necessary when MEGAHIT already reconstructed the full mitogenome. For every species, we used MitoFinder to annotate and extract genes. For missing species, we retrieved the COI-5P gene or complete mitochondrion from GenBank (see **Table S1** for more details).

We selected 11 outgroup species as previously described (see *Species list and taxon sampling* section), either using the mitogenome of the same individual as for the nuclear

analyses, or by taking the complete mitogenome of the same species on GenBank (see **Table S1**).

To account for frameshifts (both artificial and biological) and stop codons, sequences were aligned with MACSE 2.07 (Ranwez et al., 2018), setting an invertebrate mitochondrial genetic code (*-gc_def 5*), a frame-shift penalty of 5 (*-fs 5*) and an internal stop codon penalized by 10 (*-stop 10*). Maximum-likelihood (ML) phylogenetic inference was implemented with IQ-TREE 2.2.0 (Minh et al., 2020a) using ModelFinder to select the best-fit partition scheme and the best-fitting substitution model for each partition (*-m MFP+MERGE* option, Chernomor et al., 2016, Kalyaanamoorthy et al., 2017) applied to an initial subset of 41 possible partitions (the three codon-position of the 13 protein-coding genes and the 2 non-coding genes). Branch support of the ML tree was evaluated with 1,000 ultrafast bootstraps (UFBS; *-B 1000* option, Hoang et al., 2018).

Phylogenomic analyses with nuclear data

Phylogenomic analyses of genome-scale data were based on BUSCO genes. For each of the 5,286 lepidopteran core genes, we assessed the number of represented individuals, and applied two different sampling criteria on these genes to form two datasets. *Dataset 1* corresponds to all genes that gather at least 145 out of 159 individuals (91.2% of sampling for each gene), which corresponds to 1,402 genes. The *Dataset 2* was produced with a threshold of 50% (80/159 individuals) sampling for each gene, which corresponds to 4,525 genes.

The nucleotide sequences of the 1,402 genes (*Dataset 1*) were aligned with MAFFT 7.453 (Katoh and Standley, 2013). For each alignment, a gene tree was reconstructed with IQ-TREE with the best-fitting substitution model for each gene (*-m TESTNEW* option) and 1,000 UFBS. We used PhylteR (Comte et al., 2023) to detect potential outlier genes in our dataset,

applying two different settings for the 'k' parameter (*k*=4 and *k*=6), which controls the strength of outlier detection (**Fig. S1**). Four and three outliers were identified respectively, and two of these latter were not among the four outliers detected by *k*=4. We manually checked all the outlier gene trees. Finally, we decided to retain these few genes for the next stage of phylogenetic analysis 1) we were uncertain about the extent of artificial versus biological signals, 2) these cases represent only 0.028% (*k*=4: 4/1402) or 0.021% (*k*=6: 3/1402) of the entire dataset, and 3) the outlier genes detected did not consistently overlap between different *k* values. A supertree of the gene trees was produced with ASTRAL 5.7.7 (Zhang et al., 2018) following a tree-reconciliation (or 'supertree approach').

A supermatrix approach was also performed on both datasets. First, the nucleotide alignments of *Dataset 1* were concatenated into a single matrix. Maximum-likelihood phylogenetic inference was carried out with IQ-TREE. A first analysis was performed applying the GTR+I+G model to each of the 1402 partitions (*-m GTR+I+G*, one partition per gene). Second, we performed an analysis using ModelFinder to select the best-fit partition scheme and the best-fitting substitution model for each partition (*-m MFP+MERGE* option), on the initial subset of 1,402 possible partitions. Branch support was evaluated with 1,000 ultrafast bootstraps (UFBS; *-B* 1000 option). We performed ten independent IQ-TREE likelihood searches based on the best-fit partition scheme and substitution models to estimate the global log-likelihood score of the tree. *Dataset 1* was also analysed with an amino-acid supermatrix, and the phylogeny was reconstructed with IQ-TREE using the LG protein model (*-m LG+G4+F*, Le and Gascuel, 2008) and branch support evaluated with 1,000 UFBS. These same options were also used with a concatenated matrix of 4,525 amino acid genes (*Dataset 2*).

Finally, because traditional statistical measures of nodal support are prone to inflation in phylogenomic data sets (Minh et al., 2020b), we performed gene concordance factor (gCF)

and site concordance factor (sCF) analysis implemented in IQ-TREE 2 (Minh et al., 2020a), using respectively the gene trees of *Dataset 1*, and the nucleotide supermatrix of *Dataset 1*, in comparison with the consensus tree of the IQTREE_1402_NT analysis (**Table 2**).

Analysis of incomplete lineage sorting and gene flow

To investigate the hypotheses that incomplete lineage sorting (ILS) or gene flow (GF) have caused topological conflicts between the species tree and gene trees in *Graphium* (*Pathysa*) and *Graphium* (*Paranticopsis*) (see *Results*), we used Aphid (Galtier, 2023), a maximum-likelihood method that aims to quantify the sources of phylogenetic conflict via topology and branch length analysis of rooted three-species gene trees. Given a triplet of species A, B, C and outgroup species, and assuming a main topology ((A, B), C), Aphid classifies all gene trees into five categories: no event detected (the gene tree follows the main topology ((A, B), C) and its branching time), non-discordant ILS (same topology but branching times are higher than in the main topology), non-discordant GF (same topology but branching times are smaller than in the main topology), discordant ILS (noted ILSc ('c' for conflict), where topologies differ from the main topology, and branching times are higher), and discordant GF (GFc, topologies differ from the main topology and branching times are smaller). For the last two cases, Aphid also gives the 'discordant topology imbalance' associated with the event (predominance of ((A, C), B) or ((B, C), A)). The rationale is that GF events occur after population isolation and therefore result in shorter branch lengths, whereas ILS events are caused by ancestral polymorphisms older than population isolation. Aphid can therefore identify GF even when alternative topologies occur at similar frequencies, unlike the statistics derived from the ABBA/BABA test.

All gene trees that could be rooted with *Baronia brevicornis* were selected for the analyses (1,330 gene trees). To investigate the potential ILS and GF that could lead to the

observed paraphyly of *Graphium* (*Pathysa*) and polyphyly of *Graphium* (*Paranticopsis*), we focused our tests on five species-groups: the *aristeus* species-group (hereafter referred to as *"Pathysa 1"*), the *antiphates* species-group (*"Pathysa 2"*), the clade formed by *G. agetes* and *G. stratiotes* (*"Pathysa 3"*), the *macareus* species-group (*"Paranticopsis 1"*) and the *deucalion* species-group (*"Paranticopsis 2"*). Four configurations of triplets were tested, each requiring different A, B, C species (**Fig. 3**). In our case, A, B and C did not correspond to single species but to entire clades (species-groups). To account for the impact of selecting only one species as a representative of an entire species-group, the analysis of each triplet was repeated with five different A, B and C species randomly selected in each species-group (**Table S3**). For each series of triplets, we calculated the mean of the five replicates tested. In all analyses, *Graphium* (*Arisbe*) *antheus* was selected as outgroup of the triplets. Aphid was run with default options.

Results

Reference genomes

We assembled *de novo* genomes of five Leptocircini species representing the genera *Iphiclides* (*I. podalirius*), and *Graphium* with *Graphium (Arisbe)* (*Graphium antheus*), *Graphium (Pathysa)* (*G. antiphates*) and *Graphium (Graphium)* (*G. agamemnon, G. doson*) using LR and SR data (**Table 3**). Final depth of coverage estimates of LR assemblies ranged from 12 to 66× coverage, and SR polishing of these genomes from ~66 to 110× coverage. Overall, all assemblies had high BUSCO scores (between 97.3% and 98.8% complete BUSCO genes) but moderate N50 (70 kb - 1.1 Mb, **Table 3**). The sizes of the assembled genomes varied from 469 Mb (*I. podalirius*) to 2,052 Mb (*G. antiphates*). Two assemblies (*G. antheus*: 1,686 Mb; *G. antiphates*: 2,052 Mb) had higher size than expected in the literature, even after having removed the detected artificial duplication in the assembly through the *purge_dup* step and both assemblies still had >12% of duplicated BUSCO genes (**Table 3**). This could be due to highly heterozygous genomes but could not be corrected with other genome assemblers either. Our assembly of *I. posalirius* was larger than the one generated by Mackintosh et al. (2022) (469 Mb vs. 430 Mb) but both genomes were very similar in the number of BUSCO single-copy genes recovered (97.09% vs. 97.77%). The mapping performed with Minimap2 resulted in 393 Mb of alignment with a divergence of 0.77%.

Shotgun whole-genome assemblies

We assembled 136 *de novo* whole genomes of Leptocircini using Illumina SR data. Assembly with MEGAHIT resulted in draft assemblies with an average Complete single-copy genes BUSCO score of 57.9% (**Fig. 4**). The mean of Fragmented and Missing genes was 16.7% and 24.0%, respectively. Mean assemblage sizes vary greatly according to genus: 424 Mb for

Protographium, 489 Mb for *Iphiclides*, 839 Mb for *Lamproptera*, 839 Mb for *Eurytides*, and 1,093 Mb for *Graphium*. After a scaffolding step with RagTag with a phylogenetically close reference genome when available, the average complete single-copy genes BUSCO of all assemblies improved to 73.0% (**Fig. 4**) and that of Fragmented and Missing genes BUSCO score decreased to 10.8% and 15.3%, respectively.

Phylogenomics of Leptocircini

All genome-scale phylogenetic analyses provided very similar phylogenetic trees, with identical backbones, but differed in branch length estimates and relationships in some species-groups (**Fig. 5**, **Figs. S1, S3, S4, S5**).

All analyses recovered high mean and median branch support (**Fig. S6**). For instance, the IQ-TREE analysis of *Dataset 1* (supermatrix of 1,402 nucleotide genes for a total matrix of 3.5 Mb, 'IQTREE_1402_NT') showed a mean branch support of 99.5 and a median of 100, with 94.2% of nodes having maximal branch support (**Fig. S6**).

Both IQTREE_1402_NT analyses (*GTR+I+G* and *MPF+MERGE* options) produced trees identical in topology and almost identical in branch length (mean branch length $1.848e^{-2}$ vs. $1.832e^{-2}$, **Figs. 5, S7, S8**), but the *GTR+I+G* analysis was 11.8 times faster (38h vs. 448h). The *MFP+MERGE* analysis was selected as the reference topology (**Fig. 5**). The number of final partitions in this IQ-TREE analysis was reduced to 183 (compared to the initial 1402 possible partitions). All log-likelihood values from the 10 tree searches were consistent, ranging between -54,638,088.158 and -54,638,088.200, varying only at the decimal level. The resulting trees were identical in topology, and the branch length variations were infinitesimal, with differences ranging between $-3e^{-6}$ and $2e^{-7}$ (**Fig S9**). For the mitochondrial analysis, the number of final partitions was reduced to 14 partitions (compared to the initial 41 partitions).

Overall, all maximum likelihood trees of the nucleotide *Dataset 1* (IQTREE_1402_NT, **Fig. 5**; ASTRAL_1402_NT, **Fig. S2**), the amino-acid *Dataset 2* (IQTREE_4525_AA, **Fig. S3**) and the mitochondrial analysis (IQTREE_15_Mito, Fig. S5) recovered strong support for the monophyly of all genera: *Iphiclides* ($UFBS_{NT}$/$UFBS_{AA}$/LPP = 100/100/1), *Lamproptera* (100/100/1), *Eurytides* (100/100/1), and *Graphium* (100/100/1). *Protographium* was recovered as sister to *Graphium* with maximal support (100/100/1), except in the mitochondrial analysis ($UFBS_{mito}$= 95).

Considering the former higher-level systematics (**Fig. 5**), nine subgeneric names have been proposed for genus *Eurytides* (e.g. Mohn 2002; Lamas, 2004; Nakae, 2021; Zhang et al. 2021, 2024). *Eurytides (Asiographium)* and *Eurytides (Bellerographium)* were strongly nested within *Eurytides (Neographium)* ($UFBS_{NT}$/$UFBS_{AA}$/LPP = 100/100/1) and *Eurytides (Eurytides)* (100/100/1) respectively. *Eurytides (Eurygraphium)* was found monophyletic and sister to *Eurytides (Eurytides)* (100/100/1) and *Eurytides (Protesilaus)* was found monophyletic and sister to this previous clade (100/100/1). *Eurytides (Hyalaus)* was found sister to all *Eurytides (Mimoides)* (100/100/1) and *Eurytides (Boreographium)* was found sister to this previous clade (100/100/1). Finally, *Eurytides (Neographium)* was found sister to the clade formed by *Mimoides+Hyalas+Boreographium* (100/100/1).

Within genus *Graphium*, *Graphium (Pazala)*, *Graphium (Graphium)* and *Graphium (Arisbe)* were recovered as monophyletic with strong support ($UFBS_{NT}$/$UFBS_{AA}$/LPP = 100/100/1 for the crown node of each subgenus). However, *Graphium (Pathysa)* as delimited by Nakae (2021) was found to be paraphyletic in all analyses (100/100/1, **Fig. 5**, **Fig. S10**), and *Graphium (Paranticopsis)* was found to be polyphyletic in all analyses (100/97/1, see **Fig. 5**). Indeed, the latter was separated into two distinct clades: the *deucalion* species-group and the *macareus* species-group which was sister to *G. agetes* (*Pathysa*) + *G. stratiotes* (*Pathysa*),.

Overall, topological conflicts between all phylogenetic analyses were mostly limited to shallow nodes with short internode lengths within well-supported clades or deeper nodes with low branch support, especially in *Eurytides (Protesilaus)* (**Fig. S11**), the species *idaeoides* and *encelades* within *Graphium (Pathysa/Paranticopsis)* (**Fig. S10**), the *adamastor* species-group (*Graphium (Arisbe)*, **Figs. 5, S2, S3, S4, S5**), and in *Graphium (Graphium)* (e.g. the *doson* and *sarpedon* species-groups, **Figs. 5, S2, S3, S4, S5**).

Concordance factors

Gene and site concordance factors (gCF and sCF) had systematically lower values than UFBS (see **Fig. S12** for comparison of $UFBS_{NT}$/gCF/sCF). The median and mean values were respectively 77.42 and 68.49 for the gCF, and 55.80 and 56.48 for the sCF. Several clades had overall poor gCF and sCF values (**Fig. 6**, **Figs. S12, S13**), such as the sister relationship between *Lamproptera* and *Iphiclides*; or deep nodes in *Graphium (Arisbe)*, *Graphium (Graphium)* and *Graphium (Pathysa/Paranticopsis)*; or recent nodes within species-groups such as the *Eurytides* (*Protesilaus*) clade, the *adamastor* species-group in *Graphium (Arisbe)*, and the *sarpedon* species-group in *Graphium (Graphium)*.

Incomplete lineage sorting and gene flow in Pathysa-Paranticopsis

Graphium (*Pathysa)* was found to be paraphyletic and *Graphium (Paranticopsis)* polyphyletic: the *macareus* species-group (*"Paranticopsis 1"*) was found to be sister to *G. agetes* and *G. stratiotes* (*"Pathysa 3"*) and they were separated from the *deucalion* species-group (*"Paranticopsis 2"*). In addition, this topology is associated with a high rate of discordant gene tree topologies that could be explained by GF or ILS. Our sampling of *Graphium (Pathysa)* and *Graphium (Paranticopsis)* species is almost complete (except the very rare *Graphium*

549 *euphratoides*, assumed to be in the *"Pathysa 2"* species-group based on morphology) and all
550 species possess more than 90% of the 1,402 genes. Our data thus provides a suitable framework
551 for testing the hypotheses of GF and ILS between *Graphium (Paranticopsis)* and *Graphium*
552 *(Pathysa)*.

553 Aphid was run on four different series of triplets. The total number of genes classified
554 by Aphid was very similar between triplets and within replicates (>1000, **Table S3**). When the
555 observed topology was set as the main topology, *i.e.* triplet nb 1 ((“*Pathysa 3"*, “*Paranticopsis*
556 *1"*), “*Paranticopsis 2"*), GF was detected between “*Paranticopsis 1"* and “*Paranticopsis 2"*
557 (46.7% of GF, and 56.9% of imbalance toward (“*Paranticopsis 1"*, “*Paranticopsis 2"*) in these
558 46.7%, **Table 4**). When the tested topology was set to triplet nb 2 ((“*Paranticopsis 1"*,
559 “*Paranticopsis 2"*), “*Pathysa 3"*), a high proportion of GF was detected between
560 “*Paranticopsis 1"* and “*Pathysa 3"* (which was expected as it is supported by the ‘main
561 topology’). In that case, the proportion of ILS between “*Paranticopsis 1"* and “*Pathysa 3"* was
562 not particularly higher than in the first triplet. Finally, when “*Pathysa 2"* was included in the
563 triplet (triplets nb 3 and nb 4, **Table 3**), GF was detected between *"Pathysa 2"* and the two
564 other taxa of the triplet, with a slight imbalance for GF between *"Pathysa 2" and Paranticopsis*
565 (55.8 and 58.4% of imbalance in triplet n°3 and n°4 respectively, **Table 4**). This GF was higher
566 when the *Paranticopsis* clade was represented by *"Paranticopsis 2"* than by “*Paranticopsis*
567 *1"* (33.4% in triplet nb 3 against 25.7% in triplet nb 4; **Table 4**).

568

Discussion

Whole genome assemblies of Leptocircini

Our study provides the first robust and near complete phylogenomic framework for tribe Leptocircini, thanks to the *de novo* assemblies of 150 individuals (143 species) including five reference genomes. The final phylogeny represents more than 90% of total Leptocircini richness. Our genome of *Iphiclides podalirius* generated using Oxford Nanopore sequencing (FC536) was larger compared to the PacBio-sequenced genome published by Mackintosh et al. (2022) (468 Mb vs. 430 Mb), whereas both samples were coincidentally collected in the same town near Montpellier (Saint-Martin-de-Londres, France). Despite the size difference, both genomes show highly similar gene completeness, with comparable BUSCO scores. The discrepancy in assembly size might be linked to assembly differences (*e.g.*, more haploid contigs in FC536 or less efficient assembly of complex repeat-rich regions in Mackintosh et al. 2022) or biological factors (*e.g.*, variation in transposable element content). Interestingly, Reboud et al. (2023) assembled four genomes of *Ornithoptera alexandrae* (two per population) with long and short reads data using the same pipeline and found slight differences in the genome size between the two populations (~321 Mb in the lowland population vs. ~326 Mb in the highland population), suggesting some variation in genome size within Papilionidae species can occur. Sequence divergence between the assemblies of *Iphiclides podalirius* was 0.77%, which is consistent with heterozygosity levels observed in other Papilionidae populations investigated so far (which range from 0.1% to over 1% (Mackintosh et al., 2019; Reboud et al., 2023; Marino et al., 2023). We have produced the first reference genomes for genus *Graphium*, and despite considerable sequencing efforts (e.g. *Graphium antiphates*, 114 Gb of long reads), the assembly of highly contiguous genomes of *Graphium* species remains a challenge. Their genomes appear to be highly heterozygous, and contain many repeats, further complicating the assembly step. Nevertheless, these new reference genomes allowed RagTag to scaffold and

recover around 20% more genes than the draft assemblies of the shotgun whole genomes, even recovering genes from the 'missing' BUSCO category. The closer the reference used was, the better the scaffolding step, and in our case, we still lack a good *Eurytides.* reference genome to scaffold the Neotropical Leptocircini species. With an average of 73% complete single-copy BUSCO genes, the scaffolded genomes provided robust data for reconstructing the phylogeny and will surely be useful for further genome studies such as molecular evolution (e.g. dN/dS as in Allio et al., 2021a).

Phylogeny and global taxonomic revision of Leptocircini

Overall, genus-level relationships are fairly consistent with previous studies. We recovered *Protographium* as sister to all *Graphium*, in agreement with the results of Munroe (1961) and Zhang et al. (2019). *Lamproptera* and *Iphiclides* are found to be sisters, although both branch supports and gCF indicate it is a complicated node to resolve, which is consistent with Condamine et al. (2012), Allio et al. (2020a, 2021a), but not consistent with Makita et al. (2003), Simonsen et al. (2011), Page and Treadaway (2014), and Kawahara et al. (2023).

Our phylogeny is the first to show a robust and well-sampled phylogenetic framework of *Eurytides*. Following global assessments for the family (Tyler et al. 1994; Nakae 2021) and clade-based studies (Möhn 2002; Lamas, 2004; Zhang et al. 2021, 2024), nine subgenera of *Eurytides* were considered in this study. *Eurytides* (*Neographium*) is not monophyletic because of *Eurytides (Asiographium).* In addition, *Eurytides marcellus* and *E. epidaus* are more closely related to *Eurytides (Mimoides)* than other *Eurytides (Neographium)* species. The case of *E. marcellus* was previously discussed in Zhang et al. (2021) who proposed the monospecific *Eurytides (Boreographium)* Grishin (type-species *E. marcellus*), based on a reduced species sampling, to accommodate the non-monophyly of *Eurytides (Neographium)*. Following this work and notably adding *E. epidaus* in the sampling, Zhang et al. (2024) proposed the monospecific subgenus *Eurytides (Hyalaus)* Grishin (type-species *E. epidaus*) to again address

619 the non-monophyly of *Eurytides (Neographium)*. Here, in view of the dense global sampling
620 that this study provides, in order to maintain consistency of the subgeneric level within the
621 entire tribe (and family) and to avoid the description of too many taxonomic entities
622 (monospecific and/or paraphyletic subgenera), we propose merging all species of the clade
623 *Asiographium+Boreographium+Neographium+Hyalaus+Mimoides* into *Eurytides*
624 *(Mimoides)*, thus placing *Eurytides (Boreographium)*, *Eurytides (Hyalaus)*, *Eurytides*
625 *(Asiographium)* and *Eurytides (Neographium)* as synonyms of subgenus *Eurytides (Mimoides)*
626 (*Boreographium* **syn. nov.**, *Asiographium* **syn. rest.**, *Hyalaus* **syn. nov** and *Neographium* **syn.**
627 **nov.**).

628 *Eurytides* (*Eurygraphium)* (the *thyastes* species-group) is found sister to *Eurytides*
629 *(Eurytides)*. Lamas (2004) synonymized *Eurygraphium* within *Protographium* but instead we
630 synonymize it with *Eurytides*.. (*Eurygraphium* **syn. rest.**). Lamas (2004) also synonymized
631 *Bellerographium* Möhn within *Eurytides (Eurytides)*, which we confirm with this study.
632 Finally, in agreement with Nakae (2021) and Zhang et al. (2024), we confirm the subgeneric
633 status of *Eurytides (Protesilaus)*. As a result, *Eurytides* is divided in three subgenera: *Eurytides*
634 *(Eurytides)*, *Eurytides (Mimoides)* and *Eurytides (Protesilaus)*.

635 The relationships within genus *Graphium* are very consistent with Miller (1987), but
636 differ from Smith and Vane-Wright (2001), Makita et al. (2003), Page and Treadaway (2014)
637 and Allio et al. (2021a), especially regarding the strong support for monophyly of *Graphium*
638 *(Graphium)*.

639 *Graphium phidias* has had an ambiguous placement in the literature. It was long placed
640 within *Graphium (Paranticopsis)*, due to its similar wing pattern (Munroe, 1961, and with
641 caution, e.g. Racheli and Cotton, 2009; Page and Treadaway, 2014). It was found to be sister
642 to the *eurypylus* species-group in Makita et al. (2003), and sister to *Graphium (Arisbe)*,
643 *Graphium (Paranticopsis)* and *Graphium (Pathysa)* in Allio et al. (2021a). It was placed in

644 *Graphium (Graphium)* in Nakae (2021). Here, *G. phidias* is strongly supported as sister to the
645 *eurypylus* species-group, with a rather long branch, suggesting an early divergence from this
646 group.

647 *Graphium (Graphium)* is usually divided into three species-groups: the *eurypylus*,
648 *sarpedon* and *agamemnon* species-groups (Munroe, 1961; Saigusa et al., 1982; Makita et al.,
649 2003; Racheli and Cotton 2009; Page and Treadaway 2014). However, the *sarpedon* species-
650 group is not recovered as monophyletic, as the *agamemnon* species-group is nested within the
651 clade, dividing it into the *sarpedon* and the *macleayanus* species-groups. However, the short
652 branches separating these three clades (as well as their low gCF supports) suggest a rapid
653 evolution into different species-groups, and it would not be surprising if these relationships
654 were to change.

655 The inclusion of *Graphium (Paranticopsis)* within *Graphium (Pathysa)* has long been
656 suspected in the literature (e.g. Munroe and Ehrlich, 1960; Hancock 1983; Miller 1987; Makita
657 et al., 2003; Page and Treadaway, 2014), but has never been clearly demonstrated or strongly
658 supported. Our study provides strong support for this relationship, by placing *G. stratiotes* and
659 *G. agetes* as the sister clade of the *macareus* species-group, making *Graphium (Paranticopsis)*
660 polyphyletic. An ambiguous placement of *G. agetes* and *G. stratiotes* with species of the
661 *Graphium (Paranticopsis)* group had been found several times in previous studies (e.g. Makita
662 et al., 2003; Wilson et al., 2014), but was never strongly supported. Wilson et al. (2014) found
663 that two subspecies of *G. agetes* were either sister of *Graphium (Paranticopsis)* or *Graphium*
664 *(Pathysa)* species-groups, but this relied on a poorly resolved maximum parsimony tree based
665 on 28S rRNA sequences. Interestingly, Inayoshi (2023) reports the morphology of *G. agetes*
666 male genitalia supports its relationship with 'the mimetic species of the *macareus* group'.
667 Whatever the complexity of relationships within the clade *Pathysa+Paranticopsis,* (i.e. the
668 polyphyly of *Graphium (Paranticopsis)* being real or not, see below), *Graphium (Pathysa)*

would always remain paraphyletic. We therefore propose to retain only *Graphium (Pathysa)* Reakirt, and synonymize *Graphium (Paranticopsis)* Wood-Mason and de Nicéville (**syn. rest.**).

Finally, our tree confirms the monophyly of *Graphium (Arisbe)* and strongly supports the delineation of *Graphium (Arisbe)* as a subgenus of *Graphium* restricted to all the Afrotropical taxa, and not present in Asia.

Ancient gene flow could explain discordant phylogenetic relationships.

In our study, we found a strong phylogenetic discordance within the *Pathysa s.l.* clade, with low sCF and gCF values at key nodes of the different species-groups. According to Minh et al., (2020), low gCF values are difficult to interpret as they may result from weak phylogenetic signals in individual loci, or from discordance between gene trees resulting from GF or ILS. Here, the small branches recovered suggest rapid radiation from this clade that could reinforce unresolved or erroneous genetic trees.

The distinctive distribution of traits within this clade (shift to mimicry being discordant with the observed topology, see **Fig. 7**) led us to investigate the possibility of ILS and GF in the history of this clade. Indeed, two sets of morphological traits are observed in this clade. On the one hand, *Pathysa* (as circumscribed by Nakae, 2021) have tailed wings and a wing pattern with dark stripes on a pale background (**Fig. 7**). On the other hand, *Paranticopsis* as circumscribed by Nakae (2021) mimic various Nymphalidae in subfamily Danainae (mainly *Ideopsis*, *Tirumala*, *Parantica* and *Euploea*) that are tailless with a mostly black wing pattern (**Fig. 7**). The '*Pathysa* traits' are common in other distant parts of the tree. For instance, the striped wing pattern with tails is largely distributed across the phylogeny of Leptocircini. (*Protographium leosthenes, Iphiclides, Graphium (Protesilaus), Graphium (Pazala)* and some *Graphium (Arisbe*) and even in other Papilionidae tribes (Papilionini, Sericinini). Therefore, it

693 is assumed that these traits are likely to be the ancestral traits, whereas mimetic *Paranticopsis*
694 would be considered as the 'derived traits' (Hancock, 1983). Our consensus phylogeny i.e. the
695 'observed' topology (**Fig. 5**, **Fig. 7**, **Fig. S10**) indicates that *Graphium (Paranticopsis)* is nested
696 within *Graphium (Pathysa)*, causing *Graphium (Pathysa)* to be paraphyletic. Although this
697 was expected based on previous works and literature (e.g. Munroe and Ehrlich, 1960; Hancock,
698 1983; Miller, 1987; Smith and Vane-Wright, 2001), the polyphyly of *Paranticopsis* appears
699 surprising given the rather complex derived traits shared by both *Graphium (Paranticopsis)*
700 species-groups (**Fig. 7**).

701 The question is then: Is the distribution of traits we observe in the phylogeny the result
702 of 1) convergence of the derived traits (or of the ancestral traits), 2) ILS from ancient
703 polymorphism of ancestral and derived traits in the common ancestor of "*Pathysa 3"* and
704 *Paranticopsis*, or 3) GF between the different clades of *Graphium (Paranticopsis)* and/or
705 *Graphium (Pathysa)*, including the gene coding for the considered traits or not? Another way
706 of phrasing this problem is: Are the regions of the genome shared by "*Pathysa 3"* and
707 "*Paranticopsis 1"* the results of a speciation event or the result of a massive GF?

708 Although basic in appearance, these different scenarios were overall very difficult to
709 validate or even to distinguish. Yet, several hypotheses could be ruled out by our analyses.
710 First, the hypothesis that the observed topology is the 'true' topology and that *Graphium*
711 *(Paranticopsis)* derived traits are the results of a convergence gained twice in the phylogeny is
712 unlikely because of the large number of discordant topologies in the gene trees. Second, the
713 hypothesis that the observed distribution of traits is the result of ILS is not very likely because
714 of the rather low ILS supports found in all triplets. Finally, the pattern of GF and their direction
715 was the most predominant, but also the most complex to decipher.

716 When the 'observed' topology was set as the 'main' topology, i.e. the tree
717 (("*Pathysa 3", "Paranticopsis 1"*), *"Paranticopsis 2"*), a high proportion of GF was detected

between "*Paranticopsis 1"* and "*Paranticopsis 2"*. This supports the scenario of a GF from "*Paranticopsis 2"* to "*Paranticopsis 1"*, and that this GF include the derived traits (**Fig. 8C**). In that case, the last event of speciation would have been between "*Pathysa 3"* and "*Paranticopsis 1"* (as represented in the observed topology). Thereby, when the tested topology is set to ("*Pathysa 3"*, ("*Paranticopsis 1"*, *"Paranticopsis 2"*)), if the last speciation event was in reality between "*Pathysa* 3" and "*Paranticopsis 1"*, we should have found a large number of trees supporting ILS between "*Pathysa 3"* and "*Paranticopsis 1"*. This was not clearly the case, therefore we could not totally exclude the scenario in which "*Paranticopsis 1"* and "*Paranticopsis 2"* was the last event of speciation and that massive GF happened between "*Pathysa 3"* and "*Paranticopsis 1"* (**Fig. 8D**). These two scenarios could probably be distinguished by comparing the age (or absolute divergence) of the most recent nodes in the two types of topologies (*Paranticopsis* monophyletic or not), but this would probably be very difficult because of the overall very small branch length between our three clades. Finally, we found some GF between "*Pathysa 2"* and "*Pathysa 3"*, meaning that ancestral traits could also have been transferred to "*Pathysa 3"* (**Fig. 8E**). This could not really be excluded, but it is overall less likely because the amount of GF detected is smaller than in the two first cases.

The case of *Graphium (Pathysa)* and *Graphium (Paranticopsis)* is an example of the difficulty in classifying and interpreting the origin of discordant topologies in a genomic dataset. This confirms that one of the main challenges of modern phylogenomics is to manage the heterogeneity of evolutionary histories contained and revealed in the multitude of markers used (Richards et al., 2018; Scornavacca et al., 2020; Dong et al., 2022). Furthermore, our case is likely to be particularly difficult to decipher because we are testing discordances for phylogenetic scales that are deeper than simple species pairs, which probably blurs the dominant signal by the long history and possible multiplicity of evolutionary processes between the different clades (Zhang et al., 2020; Dong et al., 2022) or by the potential existence

743 of ghost lineages (Tricou et al., 2022). In any case, our example includes such small branches
744 in the inferred consensus tree that it suggests a rapid radiation of this clade and thus
745 undoubtedly reinforces the proportion of unresolved or difficult-to-classify trees for Aphid
746 (Galtier, 2023). Here, gene ontology would likely help to understand how and when the genes
747 encoding the derived or ancestral trait would have been transferred.

748 ***Further work in species-groups is needed***

749 Apart from *Graphium (Pathysa)*, concordance factor analyses (gCF and sCF) revealed poor
750 support in several species-groups, such as *Eurytides (Protesilaus)*, *Graphium* (*Arisbe)* and
751 some *Graphium (Graphium)* (**Fig. 6, Figs. S12, S13**). These were usually more 'recent nodes'
752 that could also reflect ongoing GF and recent ILS.

753 In the *adamastor* species-group (in *Arisbe*), most gCF values range between 7% to 39%
754 (8 nodes, gCF: mean = 36.0, median = 31.4), which means that in all these cases, the main
755 topology is at best retrieved in only a third of the gene tree topologies, whereas the last value
756 of the index, called gDFP (including all trees that are not either the main topology or the two
757 alternatives of the quartet *i.e.* paraphyly of the quartet and polytomy topologies) was usually
758 the most predominant alternative for the *adamastor* species-group (8 nodes, gDFP: mean =
759 51.3, median = 54.3). It is difficult to draw conclusions based only on this information, but it
760 shows that in this species-group, the 1,402 genes generally have insufficient information to
761 separate species clearly, or are not informative enough to distinguish or resolve well the
762 'species' of the *adamastor* species-group. Yet, the same genes are informative in other parts of
763 the tree, such as the rest of *Graphium (Arisbe)* that do not have so many irresolutions (whole
764 tree gCF: mean = 68.49, median = 77.42; gDFP: mean = 18.618, median = 9.735, **Fig. S13**). It
765 could reflect an ongoing GF between these different 'species', and could also mean that they
766 might not all deserve a species status. Previous reviews of this particular species-group already
767 warned about potential over-splitting in this group (Smith and Vane-Wright, 2001; Hancock,

2006). Furthermore, some of their putative species are even missing such as *G. aurivilliusi*, *G. kigoma* and *G. poggianus*, *G. ucalegonides* and *G. rileyi* (now rather considered subspecies of *G. fulleri*), *G. olbrechtsi* and *G. abri* (**Table 1**).

In *Eurytides (Protesilaus)*, the same pattern of low gCF values and strong prevalence of alternative topologies and polytomies were found (12 nodes, gCF: mean = 39.8, median = 27.9; gDFP: mean = 52.0, median = 58.9). Indeed, the observed topologies of this clade in our different analyses were very unstable (**Fig. S11**), and this clade probably needs robust population-level sampled phylogeny and taxonomic revisions (as the current species are largely sympatric cryptic species). At least one species tested here was found to be invalid: *E exiguus* (second paratype in Winhard (2018) was sequenced) is a synonym of *E. glaucolaus leucas* (**syn. nov.**) (**Table 1**). This synonymy was also confirmed by examination of male genitalia, which are identical.

In *Graphium (Pazala)*, our phylogenomic tree based on the nuclear genome lacks a few recently described species (*G. confucius*, *G. daiyuanae*, *G. sichuanica*, *G. wenlingae* only in COI, **Fig. S5**). In *Graphium (Graphium)*, our sampling was conservative in the *sarpedon* species-group by only selecting *G. anthedon* for the group *anthedon/isander/choredon* group from Cotton et al. (2022). For both cases, the degree of resolution was low despite our conservative sampling (*Graphium (Pazala)*, 8 nodes: gDFP: mean = 19.5, median = 18.5; *sarpedon*, 5 nodes, gDFP: mean = 35.1 median = 35.5).

We are concerned that further splitting in these groups will result in the same scenario, which may lead to never-resolved phylogenies. The back-and-forth between new description and synonymizing increases confusion and taxonomic instability (e.g. in subgenus *Pazala*, Hu et al., 2018; Huang, 2023), which should be avoided when taxonomic changes are not strongly supported (Christenhusz 2020). Within the *sarpedon* group, genetic structure is expected in a widely distributed clade with several subspecies described from different islands, but as Huang

(2020) pointed out, populations can maintain strong structuring without leading to speciation, and phylogeny as measures of population structure do not appear to be good predictors of speciation, even with many loci (Sukumaran and Knowles, 2017). Indeed, there may be structure in the genomics of the species, but such a population structure does not necessarily mean that they are different species or even that speciation is ongoing.

The best practice when testing taxonomic hypotheses is to be able to support a new revision with several approaches, ideally combining strong genetic, morphological and biological/ecological evidence (even if the latter is not the easiest to obtain or subject to marked and observable differences) on a large and representative sampling of the different populations and possible contact zones (Mutanen 2005; Padial et al., 2010; Schlick-Steiner et al., 2010; Yeates et al., 2011; Sangster, 2014). It is important to rely on metrics that have been shown to be informative for delimiting species within the group studied (Mutanen, 2005; Tóth and Varga, 2011; Mikitová et al., 2021; Wingert, 2022). In the case of butterflies, variations in wing size and colour pattern, or genitalia size, are expected according to environmental variations, while marked differences in genitalia shape and structure may correspond to prezygotic barriers and species differentiation (Mutanen, 2005; Mutanen and Pretorius, 2007, Tóth and Varga, 2011; Mikitová et al., 2021). For genetic evidence, it has been shown that DNA barcoding has its limitations and should generally not be considered as solid and sufficient evidence (Moritz and Cicero, 2004; Will et al., 2005). Nonetheless, comparative methods integrating a large number of genetic and metric markers can be used (Will et al., 2005; Galtier et al., 2019; Arias-Cárdenas et al., 2024; Khan et al., 2024; Wingert et al., 2024). Galtier (2019) proposed comparing the genetic differentiation for pairs of species to be delimited with pairs of species well recognized by the community as a basis for comparison. The genetic differentiation can be estimated by pairwise calculation of indices of population polymorphism structure (such as F_{ST}), indices of genetic differentiation (D_a, D_{xy}; Nei 1987; Fraisse et al., 2021; De Jode et al.,

2023), as well as analyses of population structure (PCA, admixture, structure). This approach can help to determine these species complexes as objectively and integratively as possible, as was recently performed for mammals such as aardwolf and armadillos (Allio et al., 2021b; Barthe et al., 2024).

On the generic and subgeneric status within Leptocircini

While the concept of genus is fundamental in biology and paleontology (Allmon, 1992), its definition remains subjective and inconsistent across different taxa. Genera are generally understood as groups of species that share specific characteristics, often indicating close evolutionary relationships, or in the case of monotypic genera, emphasizing the uniqueness of a species (Condamine et al. 2023). However, there are no universally accepted criteria for determining genus boundaries, beyond the requirement for monophyly supported by synapomorphies. Attempts to standardize the genus concept have proposed several guidelines, including the number of species contained within the genus, group compactness, divergence time, distinctness from related taxa, and degree of confidence in the assessment of its phylogenetic relationships (e.g. Ashlock and Mayr, 1991; Talavera et al., 2012; Dorchin et al., 2018; Sigward et al., 2018; Nakahara et al., 2020). However, one of the most arbitrary problems in systematics is still striking a balance between these standards while preserving taxonomic stability. In this context, practical considerations play a crucial role, as the elevation of subgenera to full genera can obscure broader phylogenetic relationships and complicate communication for non-specialists. The *International Code of Zoological Nomenclature* (ICZN, 1999, 2012) emphasizes taxonomic stability as a guiding practice, and this should be carefully considered in any potential taxonomic revision.

While this new phylogenomic study of Leptocircini provided an opportunity to reevaluate et revised genera and subgenera delineation within the tribe, the question of rehabilitating some subgenera into genera could be raised, particularly as some of them have

already undergone taxonomic back-and-forth in this respect. Here, we propose that the elevation of subgenera to genera is not warranted at this time (for both *Graphium* and *Eurytides*), for several reasons. First, although some subgenera of *Graphium* have been used as generic names in certain publications, most studies retain these taxa at the subgeneric level, and we advocate for this approach (e.g. Munroe 1961; Collins and Morris 1985; Miller 1987; Parsons 1998; Racheli and Cotton 2009; Hardy and Lawrence 2017; Nakae 2021). Classification within *Eurytides* has been more variable, but the most recent studies have used subgenus levels (Nakae, 2021; Zhang et al. 2021, 2024). Second, while some authors have argued for the recognition of genera based on crown ages (e.g., Avise and Johns, 1999; Talavera et al., 2012), divergence times can vary widely between clades. For instance, genera and subgenera in Leptocircini represent much older lineages (Allio et al., 2021a) compared to other butterfly families like Nymphalidae, where younger divergence times are more common (Chazot et al., 2021), thus making Papilionidae rather an exception in Papilionoidea with "old genera". As a contrary example, Talavera et al. (2012), in their work on *Polyommatus* blue butterflies, designated a divergence age of 4–5 million years to define genera. Such a time frame is clearly not applicable to Papilionidae or Leptocircini, as no compact, uniform groups appear at this age—they are much older. The evolutionary history of Papilionidae, characterized by slower diversification rates and smaller species richness compared to other families, further supports a more conservative approach to genus-level recognition. In fact, differences in crown ages might also likely reflect variations in species diversification of Papilionoidea families. Such a disparity in evolutionary timelines between different butterfly families is expected and also reflected in species diversity, with Nymphalidae comprising more than 6,400 species while Papilionidae has only ~640 species, suggesting that the taxonomic groups within these families have experienced different evolutionary histories and radiation patterns. Therefore, it seems reasonable for the age of taxonomic groups—and the ranks of

subgenus and genus—to differ across families. Furthermore, the use of divergence times would be delicate to apply here because our phylogeny is not dated. Attempts to apply divergence times from le last recent dated phylogeny of Papilionidae (Allio et al. 2021a) would be delicate too because there are significant differences in phylogenetic topologies for Leptocircini between the two studies, and such differences in phylogenetic relationship could imply significant differences in ages of major Leptocircini lineages. Finally, we think it is important to consider this matter of genus delineation with a global approach to the taxonomy of other tribes within Papilionidae, such as Papilionini, Troidini, and Parnassiini and to maintain consistency within the family. For the genus *Papilio* specifically, this issue has been recently addressed by Condamine et al. (2023). In addition, several genera in Troidini and Parnassiini contain subgenera that, while not always well known, do not warrant genus-level recognition based on molecular and morphological data. For instance, current subgenera within *Parnassius* or *Ornithoptera* are not suitable as genera without clear morphological criteria, and elevating them as such would create unnecessary and disputable complexity. This suggests that more comprehensive analyses are necessary before any taxonomic elevation can be justified.

In summary, while revising the taxonomic ranks within Leptocircini could be valuable, we believe it is essential to approach this within the broader framework of Papilionidae taxonomy. A comprehensive, well-sampled, and robustly dated phylogeny of the entire family would provide a more solid basis for such revisions. Until then, maintaining the current classification system, with subgenera recognized at their existing rank, ensures consistency and stability across the family.

Data availability

All genomic data related to the project have been deposit on GenBank (BioProject PRJNA1131164). See Table S1 and Table S2 for accession number details of each genome (BioSample and raw DNA sequenced data (SRA) numbers). All alignments, gene trees, tree files and assembly scripts are available on the Figshare: doi.org/10.6084/m9.figshare.27195603

Acknowledgments

We thank Marianne Espeland for handling the manuscript and providing comments, along with Keith R. Willmott and Niklas Wahlberg, whose constructive reviews improved the study. We would like to thank Thomas Wanchet from *La Ferme aux Papillons* for sending us fresh *Graphium* individuals from his breeding activity. We thank Rodolphe Rougerie and Jérôme Barbut for access to the collections of the Muséum National d'Histoire Naturelle in Paris (France) and Gerardo Lamas for access to the collections of the Museo de Historia Natural in Lima (Peru). We also thank James Stewart for providing some *Arisbe* specimens and Walter Winhard for providing a paratype sample of *Eurytides exiguus*. We are grateful to Amandine Magdeleine for her early work on molecular experiments, to Daphné Navratil for her help in exploring the complexity of *Graphium* genomes, and Rémi Allio for his advice on genome assemblies. We thank Nicolas Galtier and Mathilde Barthe for discussions on the implementation and interpretation of Aphid analyses. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (project GAIA, agreement no. 851188).

Conflicts of interest statement

The authors declare no conflict of interest.

ORCID

Eliette L. Reboud: 0000-0001-90794649

Adam M. Cotton : 0000-0002-4852-903X

Benoit Nabholz: 0000-0003-0447-1451

Fabien L. Condamine: 0000-0003-1673-9910

References

Ae SA. 1979. The phylogeny of some Papilio based on interspecific hybridization data. *Syst. Entomol*. **4**:1-16.

Allio R, Scornavacca C, Nabholz B, Clamens AL, Sperling FA, Condamine FL. 2020a. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. *Syst. Biol*. **69**:38-60.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020b. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol. Ecol. Res*. **20**:892–905.

Allio, R., Nabholz, B., Wanke, S., Chomicki, G., Pérez-Escobar, O.A., Cotton, A.M., Clamens, A.-L., Kergoat, G.J., Sperling, F.A.H., Condamine, F.L. 2021a. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. *Nat. Commun*. **12**:354.

Allio R, Tilak MK, Scornavacca C, Avenant NL, Kitchener AC, Corre E, *et al.*, 2021b. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. *eLife* **10**:e63167

Allmon WD. 1992. Genera in paleontology: definition and significance. *Historical Biology*, **6**:149-158.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. **23**:1-19.

Arias-Cárdenas A, Barrientos LS, Pardo-Diaz C, Paz A, Crawford AJ, Salazar C. 2024. Taxonomic inflation and a reconsideration of speciation in the Andes: the case of the high-elevation tree frog *Dendropsophus molitor* (Anura: Hylidae). *Zoological Journal of the Linnean Society*, **200**:763-775.

Ashlock PD, Mayr E. 1991. *Principles of Systematic Zoology, second ed*. MacGraw-Hill. Inc., New York.

Aubert J, Legal L, Descimon H, Michel F. 1999. Molecular phylogeny of swallowtail butterflies of the tribe Papilionini (Papilionidae, Lepidoptera). *Mol. Phylogenet. Evol.* **12**:156–167.

Avise, J. C., & Johns, G. C. 1999. Proposal for a standardized temporal scheme of biological classification. *Proc Natl Acad Sci U S A*. **96**:7358–7363.

Barthe M, Rancilhac L, Arteaga MC, Feijó A, Tilak MK, Justy F, Loughry WJ, McDonough CM, de Thoisy B, Catzeflis F, Guillaume Billet G, Hautier L, Nabholz N, Delsuc, F. 2024. Exon capture museomics deciphers the nine-banded armadillo species complex and identifies a new species endemic to the Guiana Shield. *Syst. Biol*. syae027.

Bollino, M., & Bouyer, T. 2024. A new subspecies of *Graphium bouyeri* Cipolla, 2021 and notes on taxonomic status of *Graphium biokoensis* Gauthier, 1984 (Lepidoptera, Papilionidae). Faunitaxys, 12:1-6.

Brown Jr KS. 1991. Neotropische Segelfalter: ihre Praeimaginalstadien nebst Beschreibung einer neuen Gattung *Mimoides* n. gen. (Papilionidae: Graphiini). *Entomol.* **101**:373-392.

Caterino MS, Sperling FAH. 1999. *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Mol. Phylogenet. Evol.* **11**:122–137.

Chazot N, Condamine FL, Dudas G, Peña C, Kodandaramaiah U, Matos-Maraví P, Aduse-Poku K, Elias M, Warren AD, Lohman DJ, *et al*. (2021). Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. *Nat. Comm.* **12**:5717.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884-i890.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol*. **65**:997-1008.

Christenhusz MJ. 2020. On species concepts, phylogenetics and the science of natural history—three current issues facing taxonomy. *Megataxa* **1**:67-72.

Cipolla A., 2021a. – Description d'une nouvelle espèce africaine de Graphium Scopoli, 1777 (Lepidoptera, Papilionidae). Entomologia Africana, 26(1): 49-56.

Cipolla A., 2021b. – Corrigendum - Graphium bouyeri n. sp. Cipolla, 2021 (Lepidoptera, Papilionidae). Entomologia Africana, 26 (1) Supplément: 1.

Collins NM, Morris MG. 1985. *Threatened Swallowtail Butterflies of the World: the IUCN Red Data Book.* IUCN.

Comte A, Tricou T, Tannier E, Joseph J, Siberchicot A, Penel S, Allio R, Delsuc F, Dray S, de Vienne DM. 2023. PhylteR: efficient identification of outlier sequences in phylogenomic datasets. *Mol. Biol. Evol*. **40**:msad234.

Condamine FL. Allio R, Reboud EL, Dupuis JR, Toussaint EF, Mazet N, Hu S-J, Lewis DS, Kunte K, Cotton AM, *et al.,* 2023. A comprehensive phylogeny and revised taxonomy illuminate the origin and diversification of the global radiation of *Papilio* (Lepidoptera: Papilionidae). *Mol. Phylogenet. Evol.* **183**:107758.

Condamine FL, Rolland J, Höhna S, Sperling FA, Sanmartín I. 2018. Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies. *Syst. Biol*. **67**:940-964.

Condamine FL, Sperling FAH, Kergoat GJ, 2013a. Global biogeographic pattern in swallowtail diversification demonstrates alternate colonization routes in Northern and Southern Hemispheres. *J. Biogeogr*. **40**:9–23.

Condamine FL, Toussaint EFA, Cotton AM, Genson GS, Sperling FAH, Kergoat GJ. 2013b. Fine-scale biogeographic and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics* **29**:88–111.

Condamine FL, Sperling FAH, Wahlberg N, Rasplus J-Y, Kergoat GJ. 2012. What caused the latitudinal gradient of species diversity in swallowtail butterflies? *Ecol. Lett*. **15**:267–277.

Cotton AM, Doleck T, Zhang XIN, Inayoshi Y, Lohman DJ, Hu SJ. 2022. *Graphium septentrionicolus* Page & Treadaway, 2013 (Lepidoptera: Papilionidae) is a distinct species. *Zootaxa* **5154**:211-224.

d'Abrera B. 1982. *Butterflies of the Oriental Region, Part I Papilionidae, Pieridae & Danaidae*. Hill House, Victoria.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**:361–375.

de Jode A, Le Moan A, Johannesson K, Faria R, Stankowski S, Westram AM, Butlin RK, Rafajlović M, Fraïsse C. 2023. Ten years of demographic modelling of divergence and speciation in the sea. *Evol. Appli.* **16**:542-559

Dong W, Li E, Liu Y, Xu C, Wang Y, Liu K, Cui K, Sun J, Zhang Z, *et al.*, 2022. Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC Biol.* **20**:92.

Domagala PJ, Gonzales JM. 2021. Notes on two specimens of the rare swallowtail *Eurytides iphitas* Hübner, [1821] (Papilionidae) from the Friedrich Wilhelm Niepelt collection in the Museum of Natural History, University of Wroclaw, Powland. *The Journal of the Lepidopterists' Society*. **75**:65-69.

Dorchin A, Lopez-Uribe MM, Praz CJ, Griswold T, Danforth BN. 2018. Phylogeny, new generic-level classification, and historical biogeography of the *Eucera* complex (Hymenoptera: Apidae). *Mol. Phylogenet. Evol.* **119**:81–92.

Edwards RJ. 2019. BUSCOMP: BUSCO compilation and comparison–Assessing completeness in multiple genome assemblies. *F1000Research* **8**:995.

Ehrlich PR. 1957. The Morphology, Phylogeny and Higher Classification of the Butterflies (Lepidoptera: Papilionoidea). Doctoral dissertation, University of Kansas.

Ford EB. 1944. Studies on the chemistry of pigments in the Lepidoptera, with reference to their bearing on systematics. 4. The classification of the Papilionidae. *Trans. roy. entomol. Soc. London.* **94**: 201-223

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, *et al.*, 2022. The era of reference genomes in conservation genomics. *Trends Ecol. Evol.* **37**:197–202.

Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire E, Simon A, Galtier N, Duret L, *et al.*, 2021. DILS: Demographic inferences with linked selection by using ABC. *Mol. Ecol. Res.* **21**:2629-2644.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* **34**:397-423.

Galtier N. 2019. Delineating species in the speciation continuum: A proposal. *Evol. Appli.* **12**:657-663.

Galtier N. 2024. An approximate likelihood method reveals ancient gene flow between human, chimpanzee and gorilla. *Peer Community Journal*. **4**:e3. https://doi.org/10.1101/2023.07.06.547897

Grice H, Nunez-Bustos E, Mega NO, Dias FMS, Rosa A, Freitas AVL *et al*. 2019. *Euryades corethrus* (amended version of 2018 assessment). IUCN: e. T160549A145166527.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**:2896-2898.

Hancock DL. 1993. Origins and evolution of the Afrotropical Papilionidae (Lepidoptera). *Arnoldia Zimbabwe* **9**:557-583.

Hancock DL. 2006. Notes on the afrotropical species of *Graphium* Scopoli (Lepidoptera: Papilionidae). *Metamorphosis* **17**:6-19.

Hancock DL. 1983. Classification of the Papilionidae (Lepidoptera): a phylogenetic approach. *Smithersia* **2**:1–48.

Hardy PB, Lawrence JM. 2017. *Field guide to butterflies of the Philippines* (p. 488). Siri Scientific Press.

Häuser CL, de Jong R, Lamas G, Robbins RK, Smith C, Vane-Wright RI. 2005. Papilionidae revised GloBIS⁄GART species checklist (2nd draft). Available at: http://www.globis.insects-online.de/species&tree_h=.Papilionidae&sub=yes&tree_status=plus&tree_seq=0.

He B, Zhao Y, Su C, Lin G, Wang Y, Li L, Ma J, Yang Q, Hao J. 2023. Phylogenomics reveal extensive phylogenetic discordance due to incomplete lineage sorting following the rapid radiation of alpine butterflies (Papilionidae: *Parnassius*). *Syst. Entomol.* **48**:585–599.

He JW, Zhang R, Yang J, Chang Z, Zhu LX, Lu SH, Xie F-A, Mao J-L, Dong Z-W, Li G-C. 2022. High-quality reference genomes of swallowtail butterflies provide insights into their coloration evolution. *Zool. Res*. **43**:367-379.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**:518-522.

Hu SJ, Cotton AM, Condamine FL, Duan K, Wang RJ, Hsu YF, Zhang X & Cao J. 2018. Revision of *Pazala* Moore, 1888: the *Graphium* (*Pazala*) *mandarinus* (Oberthür, 1879) group, with treatments of known taxa and descriptions of new species and new subspecies (Lepidoptera: Papilionidae). Zootaxa, 4441, 401–446. https://doi.org/10.11646/zootaxa.4441.3.1

Hu SJ, Condamine FL, Monastyrskii A, Cotton A. 2019. A new species of the *Graphium* (*Pazala*) *mandarinus* group from Central Vietnam (Lepidoptera: Papilionidae). *Zootaxa* **4554**:286-300.

Hu, S.J., Cotton, A.M., Lamas, G., Duan, K. & Zhang, X. 2023. Checklist of Yunnan Papilionidae (Lepidoptera: Papilionoidea) with nomenclatural notes and descriptions of new subspecies. *Zootaxa*, 5362:1–69. https://doi.org/10.11646/zootaxa.5362.1.1

Huang H. 2023. New or little known butterflies from China – 5 (Lepidoptera: Papilionidae, Pieridae, Nymphalidae, Lycaenidae et Hesperiidae). *Neue Entomologische Nachrichten* **82**:93-214.

Huang N, Li H. 2023. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics*. **39**:btad595.

Huang JP. (2020). Is population subdivision different from speciation? From phylogeography to species delimitation. *Ecol. Evol.* **10**:6890-6896.

ICZN. 1999. International Code of Zoological Nomenclature, 4th edition. xxix, 306 pp. The International Trust for Zoological Nomenclature, London.

ICZN. 2012. Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. *ZooKeys.* 219:1–10.

Igarashi S. 1984. The classification of the Papilionidae mainly based on the morphology of their immature stages. *Trans. Lepidopterist Soc. Japan* **34**:41–96.

Inayoshi Y. 2021. A check list of butterflies in Indo-China: Chiefly from Thailand, Laos & Vietnam. Available from: http://yutaka.it-n.jp (accessed August 1, 2023)

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**:1-31.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Meth.* **14**:587-589.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**:772–780.

Kawahara AY, Storer C, Carvalho APS, Plotkin DM, Condamine FL, Braga MP, *et al.*, 2023. A global phylogeny of butterflies reveals their evolutionary history, ancestral hosts and biogeographic origins. *Nat. Ecol. Evol.* **7**:903-913.

Khan M, Joshi M, Espeland M, Huemer P, Lopez-Vaamonde C, Mutanen M. (2024). Patterns of speciation in a parapatric pair of Saturnia moths as revealed by target capture. *Molecular Ecology*. **33**:e17194.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotech.* **37**:540-546.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**:1287.

Lamas 2004. Papilionidae. pp. 87-98 in: Lamas, G. (Ed.), Checklist: Part 4A. Hesperioidea - Papilionoidea. In: Heppner, J. B. (Ed.), *Atlas of Neotropical Lepidoptera*. Volume 5A. Gainesville, Association for Tropical Lepidoptera; Scientific Publishers

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**:1307-1320.

Lewis DS, Sperling FAH, Nakahara S, Cotton AM, Kawahara AY, Condamine FL. 2015. Role of Caribbean Islands in the diversification and biogeography of Neotropical *Heraclides* swallowtails. *Cladistics* **31**:291–314.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.

Li X, Fan D, Zhang W, Liu G, Zhang L, Zhao L, Fang X, Chen L, Dong Y, Chen Y, *et al.*, 2015. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat. Commun.* **6**:8212.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**:1674-1676.

Lu S, Yang J, Dai X, Xie F, He J, Dong Z, *et al.*, 2019. Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis. *GigaScience* **8**:giz128.

Mackintosh A, Laetsch DR, Hayward A, Charlesworth B, Waterfall M, Vila R, Lohse K. 2019. The determinants of genetic diversity in butterflies. *Nat Commun*. **10**:3466.

Mackintosh A, Laetsch DR, Baril T, Ebdon S, Jay P, Vila R, *et al.*, 2022. The genome sequence of the scarce swallowtail, *Iphiclides podalirius*. *G3:Genes Genomes Genet.* **12**:jkac193.

Makita H, Shinkawa T, Kondo K, Xing L, Nakazawa T. 2003. Phylogeny of the *Graphium* butterflies inferred from nuclear 28S rDNA and mitochondrial ND5 gene sequences. *Lepidoptera Sci*. **54**:91-110.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**:4647-4654.

Marino A, Reboud EL, Chevalier E, Tilak M-K, Contreras-Garduño J, Nabholz B, Condamine FL. 2023. Genomics of the relict species *Baronia brevicornis* sheds light on its demographic history and genome size evolution across swallowtail butterflies. *G3: Genes, Genomes, Genetics.* **13**:jkad239.

Maza-Elvira J., & Maza-Elvira R.G. 2022. An Illustrated Atlas Of Mexican Butterflies: Papilionidae. Mexico: Colección de la Maza de Lepidoptera de México, 152 pp.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* **6**:t5448.

Michel F, Rebourg C, Cosson E, Descimon H. 2008. Molecular phylogeny of Parnassiinae butterflies (Lepidoptera: Papilionidae) based on the sequences of four mitochondrial DNA segments. *Ann. Soc. Entomol. France* **44**:1-36.

Mikitová B, Šemeláková M, Panigaj Ľ. 2021. Morphological variability of *Argynnis paphia* (Lepidoptera: Nymphalidae) across different environmental conditions in eastern Slovakia. *Biologia*, **76**:2941-2956.

Miller JS. 1987. Phylogenetic studies in the Papilioninae (Lepidoptera: Papilionidae). *Bull. Am. Mus. Nat. Hist.* **186**:365–512.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020a. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**:1530-1534.

Minh BQ, Hahn MW, Lanfear R. 2020b. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* **37**:2727-2733.

Möhn E. 2002. Papilionidae VIII. *Baronia, Euryades, Protographium, Neographium* und *Eurytides*. Goecke & Evers, Keltern. 15 pp. vol. 14, p. ls 1-36. Schmetterlinge der Erde.Tagfalter. Teil 5.

Moritz C, Cicero C. 2004. DNA barcoding: promise and pitfalls. *PLoS biology*. **2**:e354.

Munroe E. 1961. The classification of the Papilionidae (Lepidoptera). *Memoirs Entomol. Soc. Canada* **92**:5-51.

Munroe E, Ehrlich PR. 1960. Harmonization of concepts of higher classification of the Papilionidae. *J. Lepid. Soc.* **14**:169-175.

Mutanen M. 2005. Delimitation difficulties in species splits: a morphometric case study on the *Euxoa tritici* complex (Lepidoptera, Noctuidae). *Syst. Entomol.* **30**:632-643.

Mutanen M, Pretorius E. 2007. Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: a comparison of moth genitalia. *Syst. Entomol*. **32**:371-386.

Mutanen M, Kivelä SM, Vos RA, Doorenweerd C, Ratnasingham S, Hausmann A, *et al.*, 2016. Species-level para-and polyphyly in DNA barcode gene trees: strong operational bias in European Lepidoptera. *Syst. Biol.* **65**:1024-1040.

Nakae M. 2021. Papilionidae of the World. Roppon-Ashi Entomological Books.

Nakahara S, Lamas G, Willmott K, Espeland M. 2020. Description of a new genus and species for a common and widespread Amazonian satyrine butterfly (Lepidoptera: Nymphalidae: Satyrinae: Satyrini). *PeerJ.* **8**:e10324.

Nazari V, Zakharov EV, Sperling FAH. 2007. Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. *Mol. Phylogenet. Evol.* **42**:131-156.

1182 Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press.

1183 Niculescu EV. 1977. Un nouveau genre de Graphiini: *Klinzigia* n. gen. *Bull. Soc. Entomol.*
1184 *Mulhouse* **33**:51–52.

1185 Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment
1186 of genome and transcriptome assemblies. *Bioinformatics* **33**:3635-3637.

1187 Owens H, Lewis DS, Dupuis JR, Clamens AL, Sperling FAH, Kawahara AY, Guralnick RP,
1188 Condamine FL. 2017. The latitudinal diversity gradient in New World swallowtail
1189 butterflies is caused by contrasting patterns of out-of- and into-the-tropics dispersal.
1190 *Glob. Ecol. Biogeogr.* **26**:1447–1458.

1191 Padial JM, Miralles A, De la Riva I, Vences M 2010. The integrative future of taxonomy. *Front.*
1192 *Zool.* **7**:1-14.

1193 Page MGP, Treadaway CG, 2003a. Papilionidae of the Philippine Islands I. Descriptions of
1194 New Subspecies and Changes in Classification. *Bauer & Frankenbach, Butterflies of the*
1195 *World*, Supplement 8. 1–6.

1196 Page MGP, Treadaway CG. 2003b. Bauer & Frankenbach, Butterflies of the World. part 17.
1197 Papilionidae IX. Papilionidae of the Philippine Islands. Goecke & Evers, Keltern.

1198 Page MGP, Treadaway CG 2004. Papilionidae of the Philippine Islands II. In: Bauer E,
1199 Frankenbach, T eds Butterflies of the world, Supplement 8. Keltern: Goecke & Evers, 7–
1200 58.

1201 Page MG, Treadaway CG. 2014. Revisional notes on the *Arisbe eurypylus* species group
1202 (Lepidoptera: Papilionoidea: Papilionidae). *Stuttgarter Beiträge zur Naturkunde: A,*
1203 *Neue Serie* **7**:253-284.

1204 Parsons M. 1998. *The Butterflies of Papua New Guinea: Their Systematics and Biology*.
1205 Academic Press.

1206 Pennisi E. 2008. Building the tree of life, genome by genome. *Science* **5884**:1716-1717.

1207 Philippe H, Blanchette M. 2007. Overview of the first phylogenomics conference. *BMC Evol.*
1208 *Biol.* **7**:S1.

1209 Racheli T, Cotton AM. 2009. Papilionidae part 1. – In: Bozano, G. C. (ed.), Guide to the
1210 Butterflies of the Palearctic Region. — Omnes Artes, Milano, 70 pp.

1211 Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the
1212 alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol.*
1213 *Evol.* **35**:2582-2584.

1214 Reboud EL, Nabholz B, Chevalier E, Tilak MK, Bito D, Condamine FL. 2023. Genomics,
1215 population divergence, and historical demography of the world's Largest and endangered
1216 butterfly, the Queen Alexandra’s birdwing. *Genome Biol. Evol.* **15**:evad040.

1217 Reed RD, Sperling FAH. 1999. Interaction of process partitions in phylogenetic analysis: an
1218 example from the swallowtail butterfly genus *Papilio*. *Mol. Biol. Evol.* **16**:286–297.

Richards EJ, Brown JM, Barley AJ, Chong RA, Thomson RC. 2018. Variation across mitochondrial gene trees provides evidence for systematic error: how much gene tree variation is biological? *Syst. Biol.* **67**:847-860.

Ross HA. 2014. The incidence of species-level paraphyly in animals: a re-assessment. *Mol. Phylogenet. Evol.* **76**:10-17.

Rothschild W, Jordan K. 1906. A Revision of the American Papilios. *Novitates Zoologocae.* **13**:411-752.

Saigusa T, Nakanishi A, Shima H, Yata O. 1982. Phylogeny and geographical distribution of the swallowtail subgenus *Graphium* (Lepidoptera: Papilionidae). *Entomologia Generalis* 59-69.

Sangster G. 2014. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. *Biol. Rev.* **89**:199-214.

Scornavacca C, Delsuc F, Galtier N. 2020. *Phylogenetics in the Genomic Era.* No commercial publisher. Authors open access book. 568 pp.

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annu. Rev. Entomol.* **55**:421-438.

Sigward JD, Sutton MD, Bennett KD. 2018. How big is a genus? Towards a nomothetic systematics. *Zool. J. Linn. Soc*. **183**:237–252.

Silva-Brandao KL, Freitas AVL, Brower AV, Solferini VN. 2005. Phylogenetic relationships of the New World Troidini swallowtails (Lepidoptera: Papilionidae) based on COI, COII, and EF-1α genes. *Mol. Phylogenet. Evol.* **36**:468-483.

Simonsen TJ, Zakharov EV, Djernaes M, Cotton AM, Vane-Wright RI, Sperling FAH. 2011. Phylogenetics and divergence times of Papilioninae (Lepidoptera) with special reference to the enigmatic genera *Teinopalpus* and *Meandrusa*. *Cladistics* **27**:113-137.

Smith CR, Vane-Wright RI. 2001. A review of the afrotropical species of the genus *Graphium* (Lepidoptera: Rhopalocera: Papilionidae). *Bull. Nat. Hist. Mus. Lond. (Ent.)* **70**:503-719.

Sukumaran J, Knowles LL. 2017. Multispecies coalescent delimits structure, not species. *Proc Natl Acad Sci U S A.* **114**:1607-1612.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ER, *et al.,* 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr. Biol.* **32**:111-123.

Talavera G, Lukhtanov VA, Pierce NE, Vila R. 2012. Establishing criteria for higher-level classification using molecular data: the systematics of *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). *Cladistics*. **29**:166–192.

Tian X, Mo S, Liang D, Wang H, Zhang P. 2023. Amplicon capture phylogenomics provides new insights into the phylogeny and evolution of alpine *Parnassius* butterflies (Lepidoptera: Papilionidae). *Syst. Entomol.* **48**:571–584.

Tilak MK, Justy F, Debiais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP. 2015. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. *Cons. Genet. Resour.* **7**:37-40.

Tóth JP, Varga Z. 2011. Inter-and intraspecific variation in the genitalia of the 'Melitaea phoebe group'(Lepidoptera, Nymphalidae). *Zoologischer Anzeiger-A Journal of Comparative Zoology*, **250**:258-268.

Tricou T, Tannier E, de Vienne DM. 2022. Ghost lineages highly influence the interpretation of introgression tests. *Syst Biol.* **71:**1147-1158.

Tyler HA, Brown KS, Wilson KH. 1994. Swallowtail butterflies of Americas: a study in biological dynamics, ecological diversity, biosystematics, and conservation. Scientific Publishers, Inc., Gainesville

Wiens JJ, Penkrot TA. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Syst. Biol.* **51**:69-91.

Wiens JJ, Kuczynski CA, Stephens PR. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biol. J. Linn. Soc.* **99**:445-461.

Wingert BD. 2022. Delimitation and identification of crescent butterflies (Nymphalidae: *Phyciodes*) in Alberta using molecular and morphological techniques.

Wingert BD, Campbell EO, Acorn JH, Sperling FA. 2024. Genomic integrity of *Phyciodes* butterfly species in a region of contact (Lepidoptera: Nymphalidae). *Insect Systematics and Diversity*, **8**:4.

Winhard W. 2018. Eine neue Tagfalterart und neue Tagfalterunterarten aus Kolumbien und Ecuador (Lepidoptera: Papilionidae, Pieridae et Nymphalidae). *Neue Entomologische Nachrichten*. **75**:143-148.

Will KW, Mishler BD, Wheeler QD. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* **54**:844-851

Wilson JJ, Karen-Chia HM, Sing KW, Sofian-Azirun M. 2014. Towards resolving the identities of the *Graphium* butterflies (Lepidoptera: Papilionidae) of Peninsular Malaysia. *J. Asia-Pacific Entomol.* **17**:333-338.

Wu LW, Yen SH, Lees DC, Lu CC, Yang PS, Hsu YF. 2015. Phylogeny and historical biogeography of Asian *Pterourus* butterflies (Lepidoptera: Papilionidae): A case of intercontinental dispersal from North America to East Asia. *PLoS One* **10**:e0140933.

Yagi T, Sasaki G, Takebe H, 1999. Phylogeny of Japanese papilionid butterflies inferred from nucleotide sequences of the mitochondrial ND5 gene. *J. Mol. Evol.* **48**:42–48.

Yeates DK, Seago A, Nelson L, Cameron SL, Joseph LEO, Trueman JW. 2011. Integrative taxonomy, or iterative taxonomy? *Syst. Entomol.* **36**:209-217.

Zakharov EV, Caterino MS, Sperling FAH. 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Syst. Biol.* **53**:193-215.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**:15-30.

Zhang C, Zhang T, Luebert F, Xiang Y, Huang CH, Hu Y, *et al.,* 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* **37**:3188-3210.

Zhang HH, Cotton AM, Condamine FL, Wang RJ, Hsu YF, Duan K, *et al.,* 2020. Revision of *Pazala* Moore, 1888: The *Graphium* (*Pazala*) *alebion* and *G.* (*P.*) *tamerlanus* groups, with notes on taxonomic and distribution confusions (Lepidoptera: Papilionidae). *Zootaxa* **4759**:77-97.

Zhang J, Cong Q, Shen J, Opler PA, Grishin NV. 2019. Changes to North American butterfly names. *Taxonomic Rep. Intl. Lepido. Survey* **8**:2.

Zhang J, Cong Q, Shen J, Opler PA, Grishin NV. 2021. Genomics-guided refinement of butterfly taxonomy. *Taxonomic Rep. Intl. Lepido. Survey* **9**:3.

Zhang J, Cong Q, Shen J, Song L, Grishin NV. 2024. Taxonomic advances driven by the genomic analysis of butterflies. *Taxonomic Rep. Intl. Lepido. Survey* **11**:7.

Zimin AV, Salzberg S.L. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput. Biol.* **16**:e1007981.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* **29**:2669-2677.

Figures

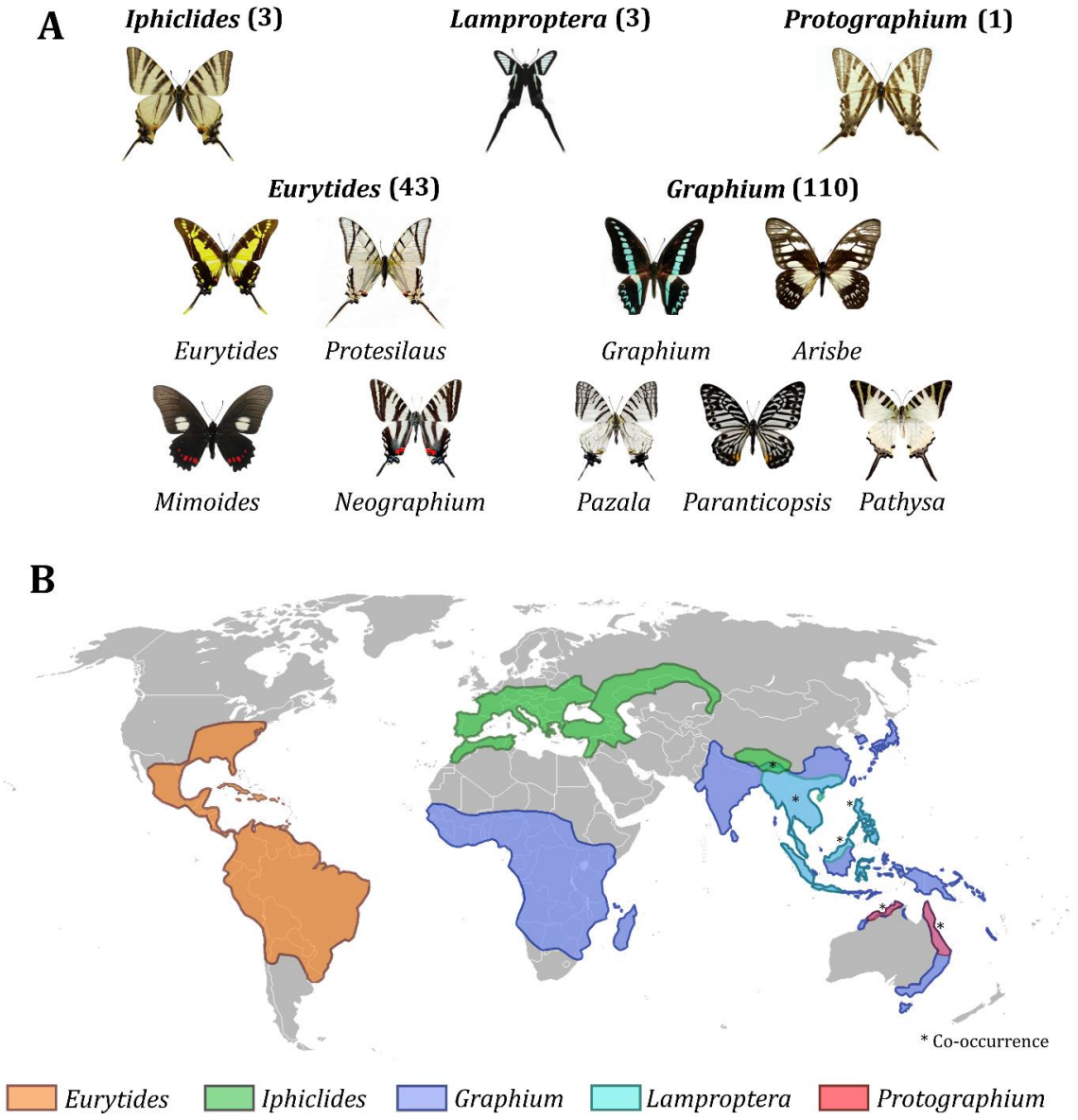


Fig. 1: Overview of Leptocircini and their global distribution. **A.** Habitus of Leptocircini genera (bold font) and subgenera (regular font) following the systematic treatment of Nakae (2021), *i.e.* monospecific subgenera of *Eurytides* are not represented (see Fig. 5 for complete subgeneric classification). The species richness is indicated for each genus between parentheses (refer to Table 1). **B.** Distribution pattern of Leptocircini genera. The distribution map of each genus is drawn approximately according to the literature and *iNaturalist* occurrences (www.inaturalist.org) and GBIF (https://www.gbif.org/fr/). Butterfly pictures: *Iphiclides podalirius, Lamproptera curius, Protographium leosthenes, Eurytides (E.) thyastes, E. (Protesilaus) telesilaus, E. (Neographium) marcellus, E. (Mimoides) ariarathes, Graphium (Arisbe) adamastor, G. (G.) milon, G. (Paranticopsis) delessertii, G. (Pazala) mandarinus, G. (Pathysa) antiphates.*

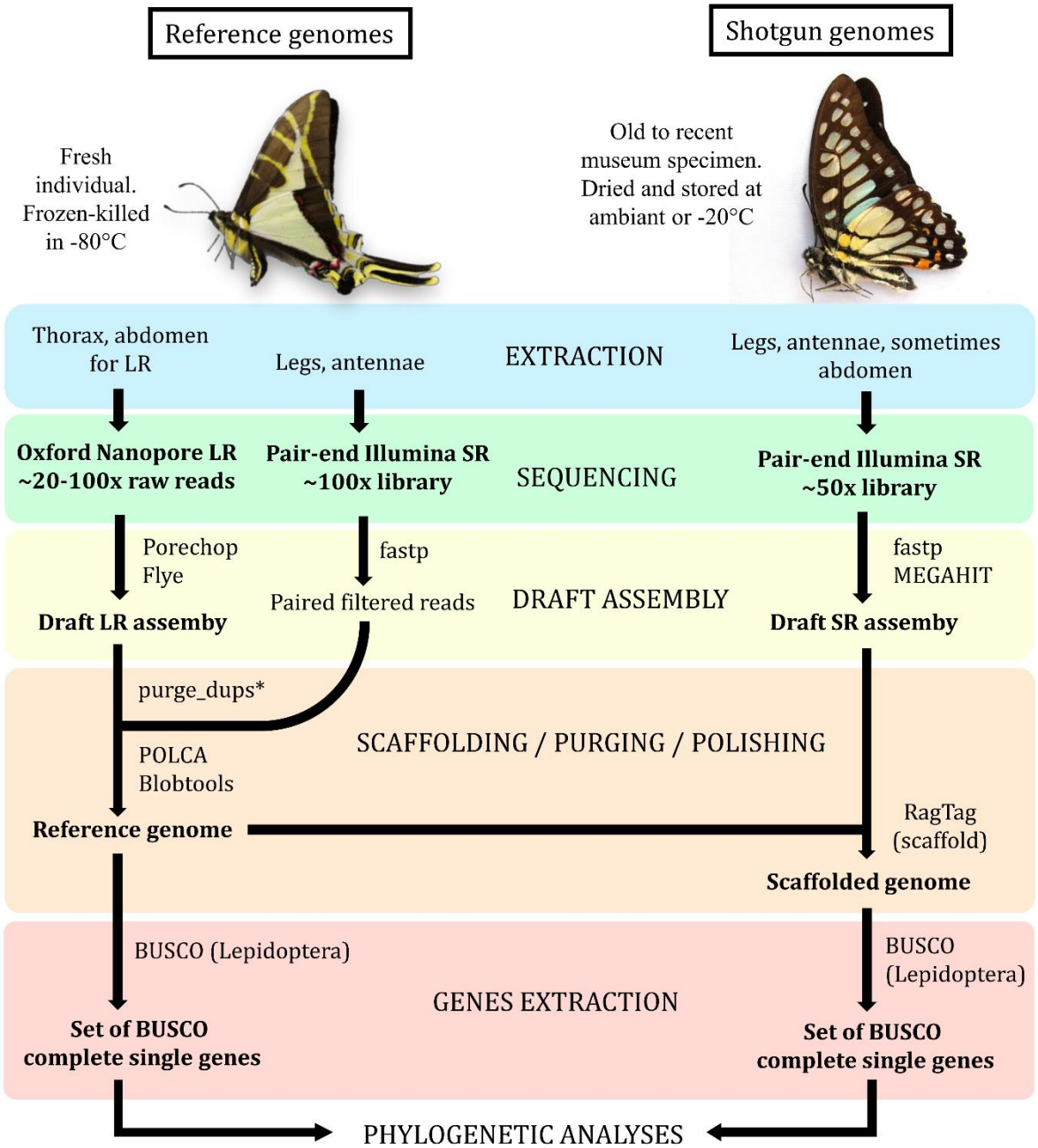


Fig. 2: Conceptualization of the sequencing and assembly pipeline for reference genomes (left) and shotgun genomes (right) used to construct phylogenomic datasets of nuclear data, based on long reads (LR) and short reads (SR) data.

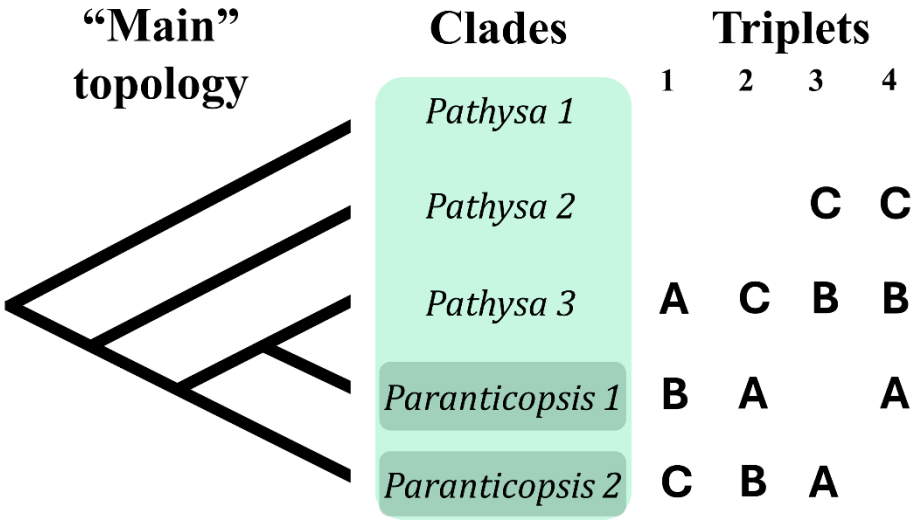
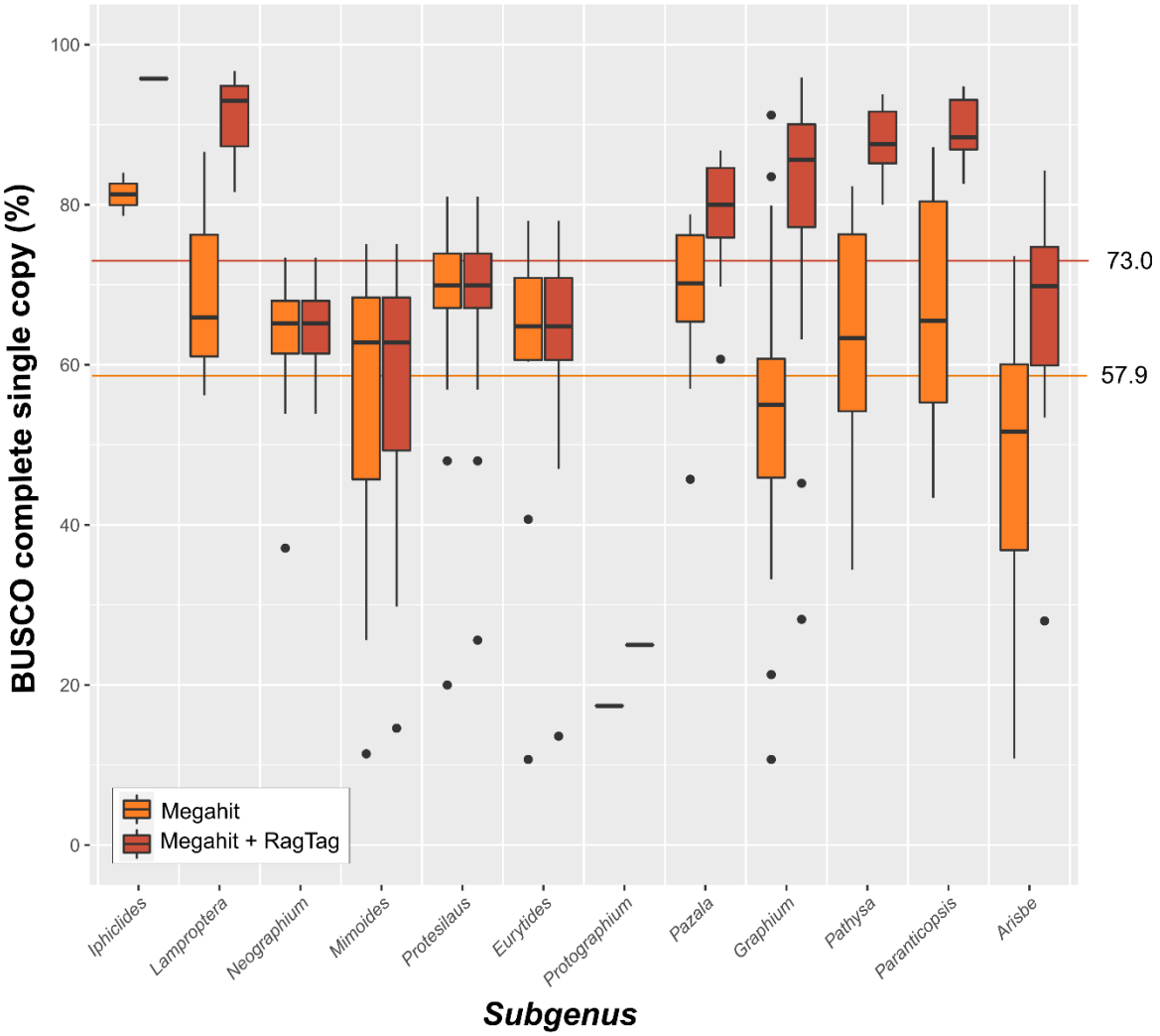


Fig. 3: Definition of the four series of triplets for the Aphid analyses. On the left is the observed phylogenetic relationships between the different clades, considered as the 'main' topology. On the right are the four tested series of ABC triplets for Aphid. The first conforms to the main topology, the second assumes that "*Paranticopsis 1*" and "*Paranticopsis 2*" shared parts of their genome because of a speciation event. The third and fourth allow testing hypotheses of gene flow or ILS with "*Pathysa 2*".



1343

1344 **Fig. 4:** Effect of scaffolding fragmented shotgun whole genomes on the recovery of BUSCO genes.
1345 Initial taxonomy followed the systematic treatment of Nakae (2021). Boxplots of BUSCO single
1346 complete genes values of all assembly (*n*=147 specimens) within a subgenus, before (orange) and after
1347 (red) the scaffolding. Horizontal coloured lines correspond to the mean BUSCO single complete score
1348 of all individuals before (orange) and after (red) the scaffolding. Draft assemblies of subgenera
1349 *Neographium*, *Mimoides*, *Protesilaus* and *Eurytides* were only scaffolded when the initial BUSCO
1350 score was under 25%.

1351

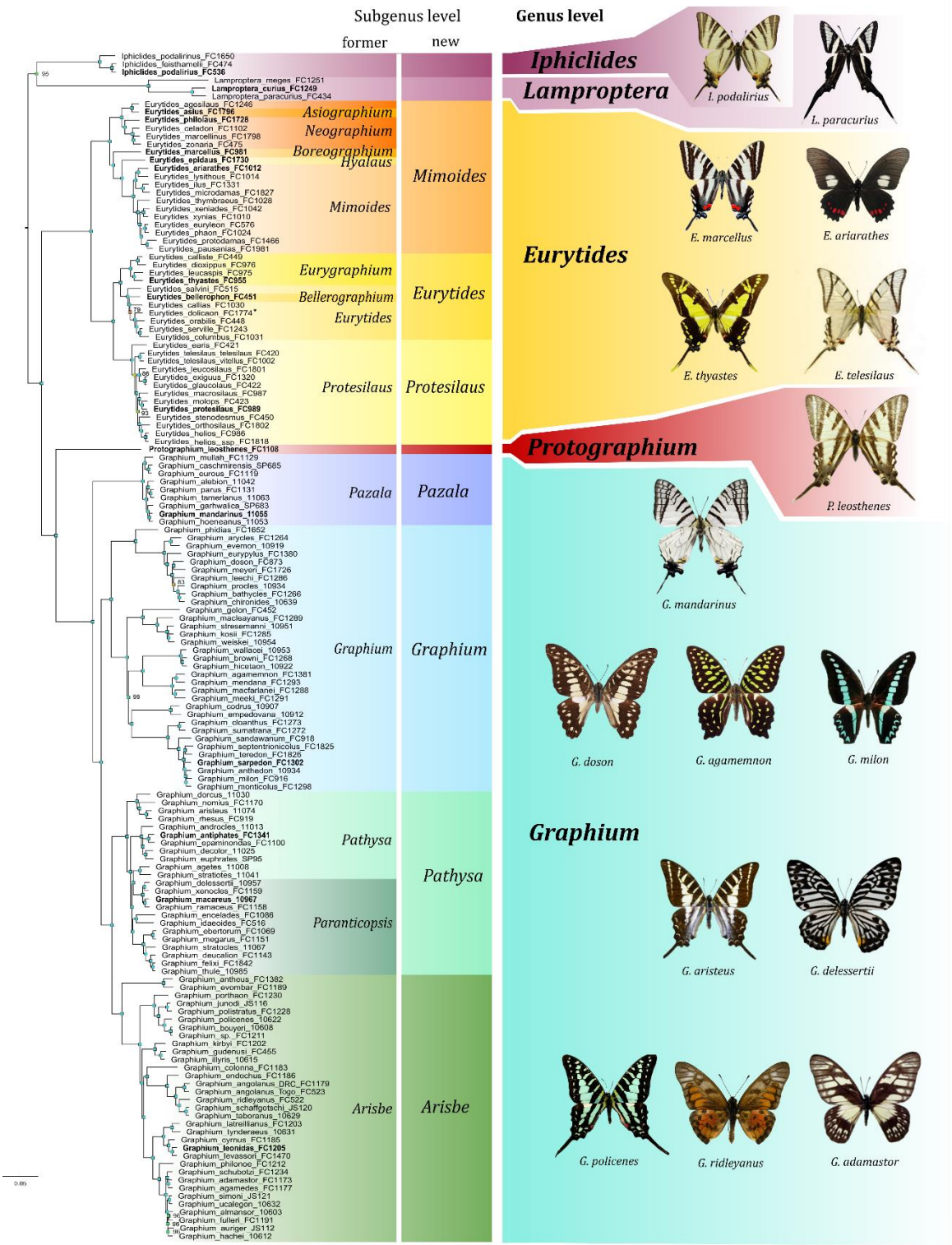


Fig. 5: Phylogenomic relationships of Leptocircini (*IQTREE_1402_NT*). The phylogeny has been reconstructed with a supermatrix approach of 1,402 single-copy genes and a maximum-likelihood inference in IQ-TREE. Light blue squares at nodes correspond to a bootstrap value equal to 100. Values <100 are indicated. Genera are in italic bold font, and subgenera in regular italic. Species in bold front are type species of subgenera. *The type species of *Eurytides (Eurytides)* is *Eurytides iphitas*, which is closely related to *Eurytides dolicaon* (Zhang et al. 2024). Illustration credit: *L. paracurius* (Adam Cotton), *G. mandarinus* (Hu et al., 2019), *G. milon, G. aristeus* (Alex Dumchus), *G. ridleyanus* (CC-BY-NC Thomas Desloges), *G. doson* (Ariane Chotard).

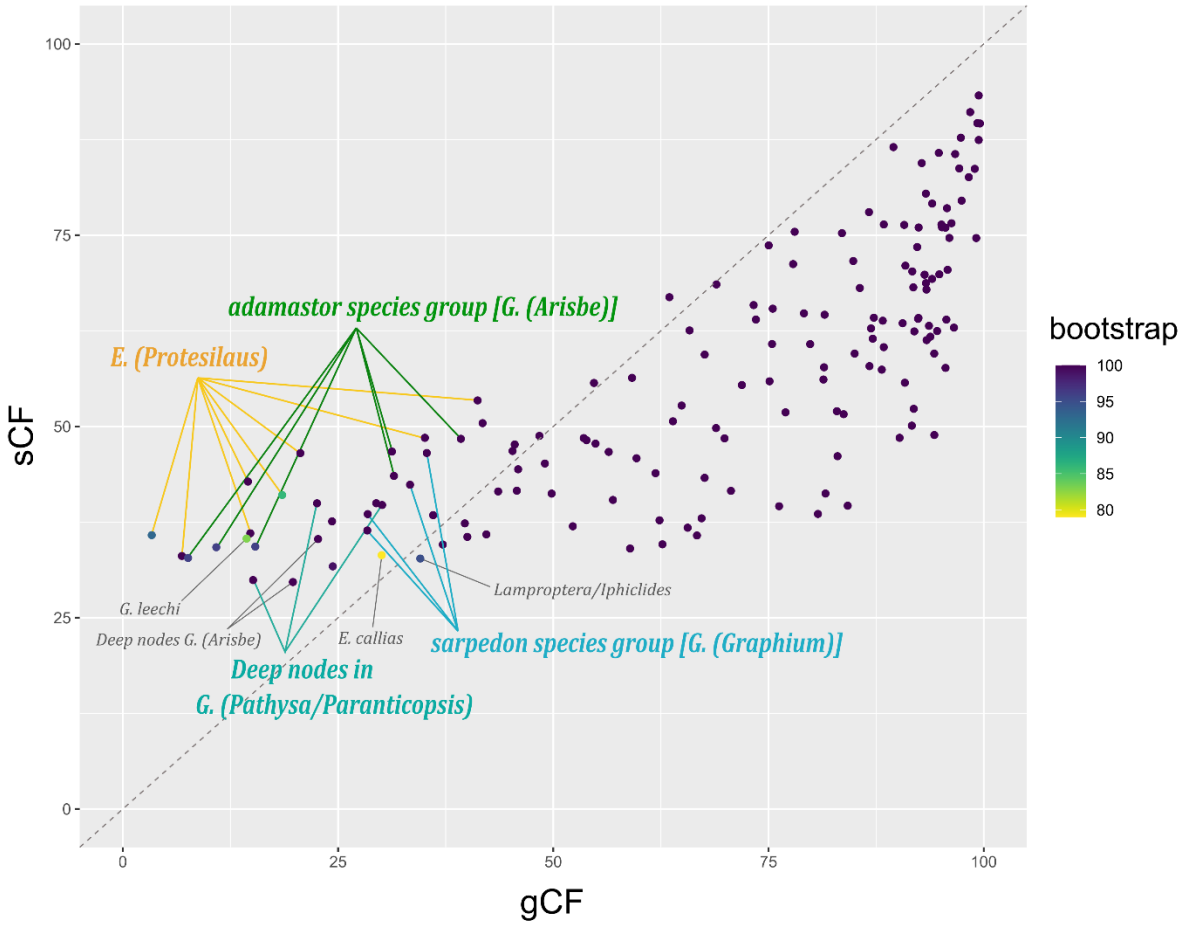


Fig. 6: Values of gCF and sCF values. Each point represents a node in the phylogeny and is coloured by its ultrafast bootstrap (UFBS) value. The dashed line corresponds to the identity line. Several clades or taxa with both poor gCF and sCF values are pointed by coloured or grey lines. *Eurytides* is abbreviated by *E.* and *Graphium* by *G.* in species names.

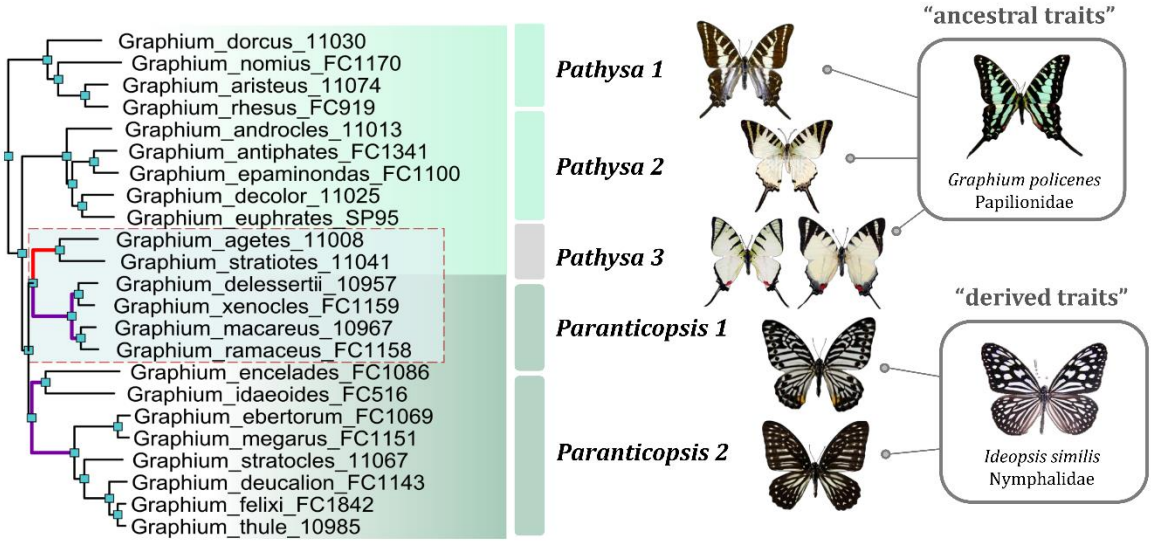


Fig. 7: Phylogenetic relationships of the clade *Graphium (Pathysa/Paranticopsis)* (*IQTREE_1402_NT* analysis). The different species-groups are delimited and named. Each species-group is linked to its phenotypic traits, corresponding either to the 'ancestral traits' (striped wing pattern and tail), or the 'derived traits' (mimetic wing pattern). Branch supports (ultrafast bootstrap) are indicated at nodes.

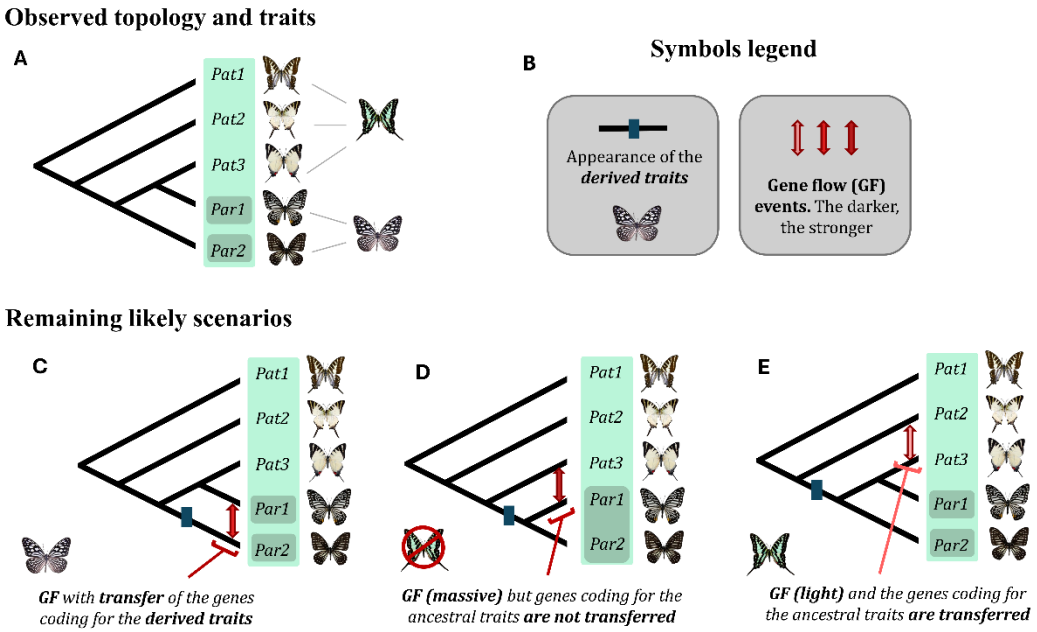


Fig. 8: A. Observed topology and trait distribution between the different species-groups of *Graphium (Pathysa)* and *Graphium (Paranticopsis)*. B. Significance of the symbols used in the illustration of the remaining likely scenarios. C, D and E. Scenarios that could explain the observed topology and trait distribution. *Pat = Pathysa*, *Par = Paranticopsis*. Note that in C, the appearance of derived traits could have also happened along the *Par1* branch, and be transferred to *Par2*.

Tables

Table 1: Taxonomic working list for Leptocircini species. The list includes subgenera that are currently recovered as monophyletic in molecular phylogenies, and species that belong to each subgenus. In the comment section, species in blue were sampled in our study and species in red were unsampled.

Genus *Eurytides* Hübner, [1821]	**Sampling**	**Taxonomic status**	**Comment**
Subgenus *Eurytides* Hübner, [1821]			
Eurytides (*Eurytides*) *bellerophon* (Dalman, 1823) ***comb. rev.***	Sampled	Valid	
Eurytides (*Eurytides*) *callias* (Rothschild & Jordan, 1906)	Sampled	Valid	
Eurytides (*Eurytides*) *calliste* (Bates, 1864) ***comb. rev.***	Sampled	Valid	
Eurytides (*Eurytides*) *columbus* (Kollar, 1849) ***stat. rev.***	Sampled	Valid	We propose a ***stat. rev.*** based on morphological, nuclear and mitochondrial divergence
Eurytides (*Eurytides*) *dioxippus* (Hewitson, [1856]) ***comb. rev.***	Sampled	Valid	
Eurytides (*Eurytides*) *dolicaon* (Cramer, 1775)	Sampled	Valid	
Eurytides (*Eurytides*) *iphitas* Hübner, [1821]	Unsampled	Valid	Probably extinct (Grice et al., 2019, Domagala and Gonzales 2021). Last specimen seen in 1977
Eurytides (*Eurytides*) *leucaspis* (Godart, 1819) ***comb. rev.***	Sampled	Valid	
Eurytides (*Eurytides*) *orabilis* (Butler, 1872)	Sampled	Valid	
Eurytides (*Eurytides*) *salvini* (Bates, 1864)	Sampled	Valid	
Eurytides (*Eurytides*) *serville* (Godart, [1824])	Sampled	Valid	
Eurytides (*Eurytides*) *thyastes* (Drury, [1782]) ***comb. rev.***	Sampled	Valid	*E. t. marchandii* is treated as a separate species in Möhn (2002) and Maza Elvira & Maza Elvira (2022). This remains to be genetically assessed.
Subgenus *Mimoides* Brown, 1991			
Eurytides (*Mimoides*) *agesilaus* (Guérin-Méneville, 1835) ***comb. nov.***	Sampled	Valid	The species *E. oberthueri* (Rothschild & Jordan, 1906) is a supposed natural hybrid of *E.agesilaus* x *E.philolaus*
Eurytides (*Mimoides*) *anaxilaus* (C. Felder & R. Felder, 1865) ***comb. nov.***	Unsampled	Valid	

Eurytides (*Mimoides*) *ariarathes* (Esper, 1788)	Sampled	Valid	
Eurytides (*Mimoides*) *asius* (Fabricius, 1781) ***comb. nov.***	Sampled	Valid	
Eurytides (*Mimoides*) *celadon* (Lucas, 1852) ***comb. nov.***	Sampled	Valid	
Eurytides (*Mimoides*) *epidaus* (Doubleday, [1846]) ***comb. nov.***	Sampled	Valid	*E. epidaus* was split into four species (*E.epidaus, E.fenochionis, E.tlahuica and E.tepicus*) by Maza Elvira & Maza Elvira (2022) based on wing pattern and genitalia. This remains to be genetically assessed.
Eurytides (*Mimoides*) *euryleon* (Hewitson, [1856])	Sampled	Valid	
Eurytides (*Mimoides*) *ilus* (Fabricius, 1793)	Sampled	Valid	
Eurytides (*Mimoides*) *lysithous* (Hübner, [1821])	Sampled	Valid	
Eurytides (*Mimoides*) *marcellinus* (Doubleday, [1845]) ***comb. nov.***	Sampled	Valid	
Eurytides (*Mimoides*) *marcellus* (Cramer, 1777) ***comb. nov.***	Sampled	Valid	
Eurytides (*Mimoides*) *microdamas* (Burmeister, 1878)	Sampled	Valid	
Eurytides (*Mimoides*) *pausanias* (Hewitson, 1852)	Sampled	Valid	
Eurytides (*Mimoides*) *phaon* (Boisduval, 1836)	Sampled	Valid	
Eurytides (*Mimoides*) *philolaus* (Boisduval, 1836) ***comb. nov.***	Sampled	Valid	*E. p. xanticles* could be a species
Eurytides (*Mimoides*) *protodamas* (Godart, 1819)	Sampled	Valid	
Eurytides (*Mimoides*) *thymbraeus* (Boisduval, 1836)	Sampled	Valid	*E. t. aconophos* is treated as a separate species in Maza Elvira & Maza Elvira (2022). This remains to be genetically assessed.
Eurytides (*Mimoides*) *xeniades* (Hewitson, 1867)	Sampled	Valid	
Eurytides (*Mimoides*) *xynias* (Hewitson, 1875)	Sampled	Valid	
Eurytides (*Mimoides*) *zonaria* (Butler, 1869) ***comb. nov.***	Sampled	Valid	
Subgenus *Protesilaus* Swainson, [1832]			
Eurytides (*Protesilaus*) *aguiari* (D'Almeida, 1937)	Unsampled	Valid	
Eurytides (*Protesilaus*) *earis* (Rothschild & Jordan, 1906)	Sampled	Valid	
Eurytides (*Protesilaus*) *exiguus* (Winhard, 2018)	Sampled	***syn. nov.***	We propose a ***syn. nov.***, as a synonym of *E. glaucolaus leucas*, also supported by identical genitalia
Eurytides (*Protesilaus*) *glaucolaus* (Bates, 1864)	Sampled	Valid	

Eurytides (*Protesilaus*) *helios* (Rothschild & Jordan, 1906)	Sampled	Valid	
Eurytides (*Protesilaus*) *leucosilaus* (Zikán, 1937)	Sampled	Uncertain	Sometimes considered as subspecies of *E. molops*, but monophyly not recovered in this study. Might be a synonym of *E. glaucolaus*
Eurytides (*Protesilaus*) *macrosilaus* (Gray, [1853])	Sampled	Valid	
Eurytides (*Protesilaus*) *molops* (Rothschild & Jordan, 1906)	Sampled	Valid	
Eurytides (*Protesilaus*) *orthosilaus* (Weymer, 1899)	Sampled	Valid	
Eurytides (*Protesilaus*) *protesilaus* (Linnaeus, 1758)	Sampled	Valid	
Eurytides (*Protesilaus*) *stenodesmus* (Rothschild & Jordan, 1906)	Sampled	Valid	
Eurytides (*Protesilaus*) *telesilaus* (C. Felder & R. Felder, 1864)	Sampled	Valid	
Genus *Graphium* Scopoli, 1777	**Sampling**	**Taxonomic status**	**Comment**
Subgenus *Arisbe* Hübner, [1819]			
Graphium (*Arisbe*) *abri* Smith & Vane-Wright, 2001	Unsampled	Doubtful	Described as a species from two specimens. Have never been genetically assessed. Could be a subspecies or melanic form of *G. adamastor* or *G. agamedes*, or a natural hybrid (Hancock, 2006).
Graphium (*Arisbe*) *adamastor* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Arisbe*) *agamedes* (Westwood, 1842)	Sampled	Valid	
Graphium (*Arisbe*) *almansor* (Honrath, 1884)	Sampled	Valid	
Graphium (*Arisbe*) *angolanus* (Goeze, 1779)	Sampled	Valid	
Graphium (*Arisbe*) *antheus* (Cramer, 1779)	Sampled	Valid	
Graphium (*Arisbe*) *auriger* (Butler, 1876)	Sampled	Valid	
Graphium (*Arisbe*) *aurivilliusi* (Seeldrayers, [1897])	Unsampled	Doubtful	Uncertain type locality. Only three specimens known. Could be a natural hybrid of *G. adamastor* x *G. schubotzi* (Hancock 2006).
Graphium (*Arisbe*) *biokoensis* Gauthier, 1984	Unsampled	Unconfirmed	From Bioko Island. Recent revision in Cipolla (2021a, 2021b) and Bollino and Bouyer (2024). Might be a subspecies of *G. policenes*

Graphium (*Arisbe*) *bouyeri* Cipolla, 2021	Sampled	Valid	Recent revision in Cipolla (2021a, 2021b) and Bollino and Bouyer (2024). Relationship with *G. biokoensis* remains to be genetically tested.
Graphium (*Arisbe*) *colonna* (Ward, 1873)	Sampled	Valid	
Graphium (*Arisbe*) *cyrnus* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Arisbe*) *endochus* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Arisbe*) *evombar* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Arisbe*) *fulleri* (Grose-Smith, 1883)	Sampled	Valid	
Graphium (*Arisbe*) *gudenusi* (Rebel, 1911)	Sampled	Valid	
Graphium (*Arisbe*) *hachei* (Dewitz, 1881)	Sampled	Valid	
Graphium (*Arisbe*) *illyris* (Hewitson, 1873)	Sampled	Valid	
Graphium (*Arisbe*) *junodi* (Trimen, 1893)	Sampled	Valid	
Graphium (*Arisbe*) *kigoma* Carcasson, 1964	Unsampled	Unconfirmed	A subspecies of *almansor* or *poggianus* (Hancock 2006)
Graphium (*Arisbe*) *kirbyi* (Hewitson, 1872)	Sampled	Valid	
Graphium (*Arisbe*) *latreillianus* (Godart, 1819)	Sampled	Valid	
Graphium (*Arisbe*) *leonidas* (Fabricius, 1793)	Sampled	Valid	
Graphium (*Arisbe*) *levassori* (Oberthür, 1890)	Sampled	Valid	
Graphium (*Arisbe*) *liponesco* (Suffert, 1904)	Unsampled	Valid	
Graphium (*Arisbe*) *morania* (Angas, 1849)	Unsampled	Valid	
Graphium (*Arisbe*) *olbrechtsi* Berger, 1950	Unsampled	Unconfirmed	Might be a subspecies of *G. schubotzi*. Insufficient data
Graphium (*Arisbe*) *philonoe* (Ward, 1873)	Sampled	Valid	
Graphium (*Arisbe*) *poggianus* (Honrath, 1884)	Unsampled	Valid	
Graphium (*Arisbe*) *policenes* (Cramer, 1775)	Sampled	Valid	
Graphium (*Arisbe*) *policenoides* (Holland, 1892)	partial Mito only	Valid	
Graphium (*Arisbe*) *polistratus* (Grose-Smith, 1889)	Sampled	Valid	

Graphium (*Arisbe*) *porthaon* (Hewitson, [1865])	Sampled	Valid	
Graphium (*Arisbe*) *ridleyanus* (White, 1843)	Sampled	Valid	
Graphium (*Arisbe*) *schaffgotschi* (Niepelt, 1927)	Sampled	Valid	
Graphium (*Arisbe*) *schubotzi* (Schultze, 1913)	Sampled	Valid	
Graphium (*Arisbe*) *simoni* (Aurivillius, [1899])	Sampled	Valid	
Graphium (*Arisbe*) *taboranus* (Oberthür, 1886)	Sampled	Valid	
Graphium (*Arisbe*) *tynderaeus* (Fabricius, 1793)	Sampled	Valid	
Graphium (*Arisbe*) *ucalegon* (Hewitson, [1865])	Sampled	Valid	
Subgenus *Graphium* Scopoli, 1777			
Graphium (*Graphium*) *agamemnon* (Linnaeus, 1758)	Sampled	Valid	
Graphium (*Graphium*) *anthedon* (C. Felder & R. Felder, 1864)	Sampled	Valid	*G. a. isander* and *G. a choredon* were considered subspecies in this study, because of the subspecies status and paraphyly of *G. a. crudus* in Cotton et al., 2022
Graphium (*Graphium*) *arycles* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Graphium*) *bathycles* (Zinken, 1831)	Sampled	Valid	
Graphium (*Graphium*) *browni* (Godman & Salvin, 1879)	Sampled	Valid	
Graphium (*Graphium*) *chironides* (Honrath, 1884)	Sampled	Valid	
Graphium (*Graphium*) *cloanthus* (Westwood, [1841])	Sampled	Valid	
Graphium (*Graphium*) *codrus* (Cramer, 1777)	Sampled	Valid	
Graphium (*Graphium*) *doson* (C. Felder & R. Felder, 1864)	Sampled	Valid	
Graphium (*Graphium*) *empedovana* (Corbet, 1941)	Sampled	Valid	
Graphium (*Graphium*) *eurypylus* (Linnaeus, 1758)	Sampled	Valid	
Graphium (*Graphium*) *evemon* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Graphium*) *gelon* (Boisduval, 1859)	Sampled	Valid	
Graphium (*Graphium*) *hicetaon* (Mathew, 1886)	Sampled	Valid	
Graphium (*Graphium*) *kosii* Müller & Tennent, 1999	Sampled	Valid	

Graphium (*Graphium*) *leechi* (Rothschild, 1895)	Sampled	Valid	
Graphium (*Graphium*) *macfarlanei* (Butler, 1877)	Sampled	Valid	
Graphium (*Graphium*) *macleayanus* (Leach, 1814)	Sampled	Valid	
Graphium (*Graphium*) *meeki* (Rothschild & Jordan, 1901)	Sampled	Valid	
Graphium (*Graphium*) *mendana* (Godman & Salvin, 1888)	Sampled	Valid	
Graphium (*Graphium*) *meyeri* (Hopffer, 1874)	Sampled	Valid	
Graphium (*Graphium*) *milon* (C. Felder & R. Felder, 1865)	Sampled	Valid	
Graphium (*Graphium*) *monticolus* (Fruhstorfer, 1896)	Sampled	Valid	
Graphium (*Graphium*) *phidias* (Oberthür, 1896)	Sampled	Valid	
Graphium (*Graphium*) *procles* (Grose-Smith, 1887)	Sampled	Valid	
Graphium (*Graphium*) *sandawanum* Yamamoto, 1977	Sampled	Valid	
Graphium (*Graphium*) *sarpedon* (Linnaeus, 1758)	Sampled	Valid	
Graphium (*Graphium*) *septentrionicolus* Page & Treadaway, 2013	Sampled	Valid	
Graphium (*Graphium*) *stresemanni* (Rothschild, 1915)	Sampled	Valid	
Graphium (*Graphium*) *sumatrana* (Hagen, 1894)	Sampled	Valid	
Graphium (*Graphium*) *teredon* (C. Felder & R. Felder, 1864)	Sampled	Valid	
Graphium (*Graphium*) *wallacei* (Hewitson, 1858)	Sampled	Valid	
Graphium (*Graphium*) *wayabulaensis* Hanafusa, 1998	Unsampled	Unconfirmed	=*batjanensis* Okano, 1984 *nomen nudum*. Generally considered as a valid species close to *G. stresemanni*.
Graphium (*Graphium*) *weiskei* (Ribbe, 1900)	Sampled	Valid	
Subgenus *Pathysa* Reakirt, [1865]			
Graphium (*Pathysa*) *agetes* (Westwood, 1843)	Sampled	Valid	
Graphium (*Pathysa*) *androcles* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Pathysa*) *antiphates* (Cramer, 1775)	Sampled	Valid	
Graphium (*Pathysa*) *aristeus* (Stoll, 1780)	Sampled	Valid	
Graphium (*Pathysa*) *decolor* (Staudinger, 1888)	Sampled	Valid	

Graphium (*Pathysa*) *delessertii* (Guérin-Méneville, 1839) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *deucalion* (Boisduval, 1836) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *dorcus* (de Haan, 1840)	Sampled	Valid	
Graphium (*Pathysa*) *ebertorum* Koçak, 1983 ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *encelades* (Boisduval, 1836) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *epaminondas* (Oberthür, 1879)	Sampled	Valid	
Graphium (*Pathysa*) *euphrates* (C. Felder & R. Felder, 1862)	Sampled	Valid	
Graphium (*Pathysa*) *euphratoides* (Eimer, 1889)	Unsampled	Unconfirmed	
Graphium (*Pathysa*) *felixi* (Joicey & Noakes, 1915) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *idaeoides* (Hewitson, 1855) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *macareus* (Godart, 1819) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *megarus* (Westwood, 1844) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *nomius* (Esper, 1799)	Sampled	Valid	
Graphium (*Pathysa*) *ornatus* (Rothschild, 1895)	Unsampled	Unconfirmed	
Graphium (*Pathysa*) *ramaceus* (Westwood, 1872) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *rhesus* (Boisduval, 1836)	Sampled	Valid	
Graphium (*Pathysa*) *stratiotes* (Grose-Smith, 1887)	Sampled	Valid	
Graphium (*Pathysa*) *stratocles* (C. Felder & R. Felder, 1861) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *thule* (Wallace, 1865) ***comb. rev.***	Sampled	Valid	
Graphium (*Pathysa*) *xenocles* (Doubleday, 1842) ***comb. rev.***	Sampled	Valid	
Subgenus *Pazala* Moore, 1888			
Graphium (*Pazala*) *alebion* (Gray, [1853])	Sampled	Valid	
Graphium (*Pazala*) *confucius* Hu, Duan & Cotton, 2018	Mito NCBI only	Unconfirmed	Subspecies of *G. mandarinus* in Huang (2023), reinstated as a species in Hu et al. (2023) on morphological and biological grounds.

Graphium (*Pazala*) *daiyuanae* Hu, Zhang & Cotton, 2018	COI-5P NCBI only	Unconfirmed	Subspecies of *G. mandarinus* in Huang (2023)
Graphium (*Pazala*) *eurous* (Leech, 1893)	Sampled	Valid	
Graphium (*Pazala*) *garhwalica* (Katayama, 1988)	Sampled	Valid	
Graphium (*Pazala*) *hoeneanus* Cotton & Hu, 2018	Sampled	Unconfirmed	Subspecies of *G. sichuanica* in Huang (2023). Unconfirmed status as *G. sichuanica* was not sampled in this study.
Graphium (*Pazala*) *mandarinus* (Oberthür, 1879)	Sampled	Valid	
Graphium (*Pazala*) *mullah* (Alphéraky, 1897)	Sampled	Valid	
Graphium (*Pazala*) *paphus* (de Nicéville, 1886)	Mito only	Valid	
Graphium (*Pazala*) *parus* (de Nicéville, 1900)	Sampled	Valid	
Graphium (*Pazala*) *sichuanica* (Koiwaya, 1993)	Unsampled	Valid	
Graphium (*Pazala*) *tamerlanus* (Oberthür, 1876)	Sampled	Valid	
Graphium (*Pazala*) *wenlingae* Hu, Cotton & Monastyrskii, 2019	COI-5P NCBI only	Unconfirmed	Subspecies of *G. mandarinus* in Huang (2023)

Genus *Iphiclides* Hübner, [1819]	**Sampling**	**Taxonomic status**	**Comment**
Iphiclides feisthamelii (Duponchel, 1832)	Sampled	Valid	
Iphiclides podalirinus (Oberthür, 1890)	Sampled	Valid	
Iphiclides podalirius (Linnaeus, 1758)	Sampled	Valid	

Genus *Lamproptera* Gray, 1832	**Sampling**	**Taxonomic status**	**Comment**
Lamproptera curius (Fabricius, 1787)	Sampled	Valid	
Lamproptera meges (Zinken, 1831)	Sampled	Valid	
Lamproptera paracurius Hu, Zhang & Cotton, 2014	Sampled	Valid	

Genus *Protographium* Munroe, 1961	Sampling	Taxonomic status	Comment
Protographium leosthenes (Doubleday, 1846)	Sampled	Valid	

Table 2: Summary of the five different phylogenomic analyses and their abbreviations. The asterisk corresponds to the analysis from which the matrix and the resulting topology were used in the concordance factors analysis.

	Mitogenome	***Dataset 1***	***Dataset 2***
Nucleotide	IQTREE_15_Mito	IQTREE_1402_NT*, ASTRAL_1402_NT	
Amino acid		IQTREE_1402_AA	IQTREE_4525_AA

Table 3 Assembly statistics for the five *de novo* reference genomes of Leptocircini. LR = long reads, SR = short reads. For BUSCO scores, abbreviations stand for S: Simple, D: Duplicated, F: Fragmented, M: Missing.

	Iphiclides podalirius	***Graphium***			
		G. agamemnon	***G. antheus***	***G. antiphates***	***G. doson***
	FC536	**FC1381**	**FC1382**	**FC1341**	**FC873**
Raw data sequenced (Gb) (LR + SR)	40.0 + 20.0	31.1 + 103.1	30.9 + 103.8	101.9 + 105.3	20.6 + 71.1
Final mean coverage (LR + SR)	66x + 96x	19x + 110x	12x + 68x	32x + 68x	14x + 105x
Assembly size (bp)	468,946,765	960,973,235	1,685,715,539	2,052,007,334	989,610,395
Number of contigs	2,755	8,469	22,326	28,705	36,745
N50 (bp)	1,107,841	236,847	149,287	173,984	70,611
Max length (bp)	13,306,297	2,403,220	1,528,824	9,291,578	1,128,770
Nucleotide assembly BUSCO score (%)	S:97.0; D:1.8; F:0.3; M:0.9	S:96.5; D:1.9; F:0.8; M:0.8	S:85.1; D:13.1; F:0.9; M:0.9	S:85.8; D:12.3; F:0.9; M:1.0	S:93.2; D:4.1; F:1.5; M:1.2

Table 4. Summary of Aphid results for the four series of triplets within *Graphium (Paranticopsis)* and *Graphium (Pathysa)*. Each value is the mean of five replicates for each series of triplet. Values indicated in the 'Posterior imbalance GFc triplet' column correspond to the percentage of the predominant discordant topology (displayed in parenthesis) over the other one. ILS = incomplete lineage sorting, GF = gene flow. ILSc and GFc are for discordant topologies. Par = *Paranticopsis*, Pat = *Pathysa*.

Tested triplet	Number of analyzed genes	Events (%)					Posterior imbalance GFc triplet (%)
		none	ILS	GF	ILSc	GFc	
1 - A=Pat3; B=Par1; C=Par2	1034	23.3	3.4	12.6	14.1	46.7	((BC)A): 56.9
2 - A=Par1; B=Par2; C=Pat3	1038	22.5	4.5	6.4	13.7	52.9	((AC)B): 63.4
3 - A=Par2; B=Pat3; C=Pat2	1014	29.0	5.9	15.1	16.5	33.4	((AC)B): 55.8
4 - A=Par1; B=Pat3; C=Pat2	1023	31.1	5.9	20.5	16.7	25.7	((AC)B): 58.4