Molecular Phylogenetics and Evolution 127 (2018) 600–605

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetics of moth-like butterflies (Papilionoidea: Hedylidae) based on a new 13-locus target capture probe set

Akito Y. Kawahara[a,b,c,*], Jesse W. Breinholt[a,d], Marianne Espeland[a,e], Caroline Storer[a], David Plotkin[a,b], Kelly M. Dexter[a], Emmanuel F.A. Toussaint[a], Ryan A. St Laurent[a,c], Gunnar Brehm[f], Sergio Vargas[g], Dimitri Forero[g], Naomi E. Pierce[h], David J. Lohman[i,j,k]

[a] Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[b] Entomology and Nematology Department, University of Florida, Gainesville, FL 32611, USA
[c] Department of Biology, University of Florida, Gainesville, FL 32611, USA
[d] RAPiD Genomics, 747 SW 2nd Avenue, IMB#14, Gainesville, FL 32601, USA
[e] Arthropoda Department, Zoological Research Museum Alexander Koenig, Adenauer Allee 160, Bonn 53113, Germany
[f] Institute of Zoology and Evolutionary Biology with Phyletic Museum, Friedrich-Schiller-University Jena, Jena 07743, Germany
[g] Laboratorio de Entomología, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia
[h] Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA
[i] Biology Department, City College of New York, New York, NY 10031, USA
[j] Ph.D. Program in Biology, Graduate Center, City University of New York, New York, NY 10016, USA
[k] Entomology Section, National Museum of the Philippines, Manila 1000, Philippines

ARTICLE INFO

Keywords:
Anchored hybrid enrichment
Exon capture
Lepidoptera systematics
Papilionoidea
Phylogenomics

ABSTRACT

The Neotropical moth-like butterflies (Hedylidae) are perhaps the most unusual butterfly family. In addition to being species-poor, this family is predominantly nocturnal and has anti-bat ultrasound hearing organs. Evolutionary relationships among the 36 described species are largely unexplored. A new, target capture, anchored hybrid enrichment probe set ('BUTTERFLY2.0') was developed to infer relationships of hedylids and some of their butterfly relatives. The probe set includes 13 genes that have historically been used in butterfly phylogenetics. Our dataset comprised of up to 10,898 aligned base pairs from 22 hedylid species and 19 outgroups. Eleven of the thirteen loci were successfully captured from all samples, and the remaining loci were captured from ≥94% of samples. The inferred phylogeny was consistent with recent molecular studies by placing Hedylidae sister to Hesperiidae, and the tree had robust support for 80% of nodes. Our results are also consistent with morphological studies, with *Macrosoma tipulata* as the sister species to all remaining hedylids, followed by *M. semiermis* sister to the remaining species in the genus. We tested the hypothesis that nocturnality evolved once from diurnality in Hedylidae, and demonstrate that the ancestral condition was likely diurnal, with a shift to nocturnality early in the diversification of this family. The BUTTERFLY2.0 probe set includes standard butterfly phylogenetics markers, captures sequences from decades-old museum specimens, and is a cost-effective technique to infer phylogenetic relationships of the butterfly tree of life.

1. Introduction

Day-flying butterflies have fascinated researchers and enthusiasts for centuries. The 36 described species of moth-like butterflies (Hedylidae) have received little attention, mainly because they are predominantly nocturnal and restricted to Central and South America (Scoble, 1986, 1990b; Kawahara et al., 2018). Most prior studies have focused on their behavior (Yack and Fullard, 2000), neuroethology (Yack et al., 2007a) or vision (Yack et al., 2007b), and the family still lacks a published genome (Triant et al., 2018). Hedylids have several moth-like features, such as clubless antennae and nocturnal adult activity, which led many authors to believe that they were not butterflies but belonging to the family Geometridae (Prout, 1910). Despite the similarity of these two families, Geometridae possess a pair of bat ultrasound-detecting tympanal organs on the abdomen, whereas hedylids have hearing organs on their wings (Yack and Fullard, 2000). Hedylids also have larvae that have more pairs of prolegs than geometrids (Scoble, 1986). Due to many similarities that hedylids share with other Lepidoptera, the phylogenetic position of Hedylidae in the Lepidoptera remained largely uncertain. Recent molecular phylogenetic studies

* Corresponding author at: Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA.
E-mail address: kawahara@flmnh.ufl.edu (A.Y. Kawahara).

https://doi.org/10.1016/j.ympev.2018.06.002
Received 27 March 2018; Received in revised form 25 May 2018; Accepted 2 June 2018
Available online 11 June 2018
1055-7903/

have provided strong support for the Hedylidae belonging in the Papilionoidea (butterflies), as the sister-group to Hesperiidae (Heikkilä et al., 2012; Espeland et al., 2018; Breinholt et al., 2018; Kawahara and Breinholt, 2014).

A limited number of systematic studies have been conducted on Hedylidae. Most are morphological classifications that are not based on a phylogenetic analysis. One of the first major taxonomic studies was by Prout (1932), who used wing pattern to divide the group into the genera *Hedyle* Guenée, *Lasiopates* Warren, *Macrosoma* Hübner, *Phellinodes* Guenée, and *Venodes* Guenée. Scoble (1986) evaluated Prout's genera and examined their morphology in greater detail, including that of the genitalia and immature stages. Although Scoble agreed that some hedylids could be placed into subgroups based on wing pattern similarity, he determined that all species are morphologically similar enough to justify treating them as a single genus, *Macrosoma*. He postulated that species in the family constitute a monophyletic group, although he did not conduct a formal phylogenetic analysis. Lourido (2011) conducted a cladistic analysis of Hedylidae, in which she coded 111 morphological characters from the 36 recognized species. Her work proposed a new classification to revalidate *Hedyle* Guenée and *Phellinodes* Guenée that, together with *Macrosoma* Hübner, would comprise three hedylid genera. However, relationships among many species were not resolved due to low branch support and unresolved polytomies, and only three outgroups were included. Furthermore, the study was not formally published, and therefore the proposed taxonomic changes are not valid under Article 9.12 of the Code (ICZN, 1999). Here, we test the monophyly of Hedylidae, examine relationships among its species, and test the hypothesis that nocturnality evolved once from diurnality in the family, using a molecular dataset that was built from a newly-developed anchored hybrid enrichment probe set. This target capture kit includes 13 loci that were carefully chosen to overlap with most existing Lepidoptera molecular datasets. We tested the capture success of this probe set on 22 Hedylidae species and outgroups sampled from six butterfly families and one moth family.

2. Methods

2.1. Probe design: the BUTTERFLY2.0 anchored hybrid enrichment kit

The BUTTERFLY2.0 probe set captures 13 gene regions including those most commonly included in studies on butterfly phylogenetics (*e.g.*, Wahlberg et al., 2016; Mitter et al. 2011; Wahlberg et al., 2005; Wahlberg and Wheat, 2008; Simonsen et al., 2011; Heikkilä et al., 2012; Kozak et al., 2015; Table 1). The 13 loci are also part of the 425 locus BUTTERFLY1.0 probe set of Espeland et al. (2018). The choice of the 13 loci was motivated by the need for a cost-effective method to generate comparatively large molecular datasets that overlap in gene sequence with prior studies and yield strong phylogenetic resolution at different taxonomic levels. The reliance on a smaller number of loci per sample reduces the price significantly compared to larger probe sets (*e.g.*, BUTTERFLY1.0, Espeland et al., 2018; or LEP1, Breinholt et al., 2018), mostly at the sequencing stage. In order to improve probe set capture, we examined sequences that captured genomic DNA fragments from all 207 butterfly species in the study of Espeland et al. (2018) and used a 90% centroid cutoff in USEARCH (Edgar, 2010) to define capture regions. Exon boundaries were based on the genomes of *Bombyx mori, Danaus plexippus, Heliconius melpomene*, and *Melitaea cinxia* in the Ensembl Metazoa database (Kersey et al., 2015).

The 658 bp barcoding region of COI was included in the BUTTERFLY2.0 kit for its phylogenetic utility, and because it serves as an important locus for species identification in Lepidoptera (Hebert et al., 2003). We also included this locus because it can facilitate merging of phylogenetic datasets for future projects (Hajibabaei et al., 2007), and because it can help assess contamination among samples (Shen et al., 2013). Sequences of each locus were aligned separately with default commands in MAFFT v.7.0.1 (Katoh and Standley, 2013) and the alignments used as the template for probe design. Probes of 120 bp were placed across each template region with 2x tiling. To avoid overcapture of the COI gene, we only designed two probes, placed 144 bp apart, that appeared to vary the least across the alignment. We summarize our probe and flanking region capture in R, using ggplot2 (Wickham, 2009) (Fig. 1).

Table 1
The 13 loci of the BUTTERFLY2.0 probe set, with sequence lengths (bp) and summary statistics for the probe and flanking regions of each gene.

Locus no.	Gene name	Probe (P)	Flank (F)	P + F
L1	Cytochrome *c* oxidase I (COI)	658	415	1073
L2	Thiolase	1024	43	1067
L3	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD)	1569	426	1995
L4	Catalase (CAT)	1300	173	1473
L5	Dopa decarboxylase (DDC)	579	150	729
L6	Elongation factor 1 alpha (EF1-a)	1049	508	1557
L7	Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	599	392	991
L8	Hairy cell leukemia protein 1 (HCL)	633	209	842
L9	Isocitrate dehydrogenase (IDH)	708	298	1006
L10	Malate dehydrogenase (MDH)	718	268	986
L11	Ribosomal protein S2 (RPS2)	472	191	663
L12	Ribosomal protein S5 (RPS5)	554	221	775
L13	Wingless	239	361	600
	Total	9827	3930	13,757
	Average	756	302	1058

2.2. Taxon sampling and species identification

We sampled 22 of the 36 described hedylid species, and included multiple specimens per species when possible. We also included 19 outgroups, representing all butterfly families and one moth, resulting in a final dataset of 48 specimens representing 41 species. Hedylid species were identified using the morphological descriptions of Scoble (1990a, b) and Lévêque (2007), and by comparison with original type material. A complete catalogue with images of all analyzed species and their type specimens is provided in Supplementary File 1. Initial identifications were cross-checked using the COI barcode sequences from this dataset with specimens from the curated collection of the McGuire Center for Lepidoptera and Biodiversity (MGCL, Gainesville, FL). COI sequences were blasted against the COI reference databases of BOLD (Ratnasingham and Hebert, 2007) and GenBank (Benson et al., 2012), which comprise > 300 sequences representing ~20 described *Macrosoma* species (www.boldsystems.org, last accessed on March 15, 2018). Two specimens in our analysis, *Macrosoma hedylaria* (Sample ID: FG120122) and *Troides rhadamantus* (Sample ID: RF140032) had the 13 loci extracted from published transcriptomes (Kawahara and Breinholt, 2014; Breinholt et al., 2018). Twelve butterflies were sequenced with the BUTTERFLY1.0 probe set (Espeland et al., 2018), and added to the dataset. These 12 samples had their sequences trimmed to the 13 loci that were included in the present study. Although there were other *Macrosoma* sequences published (*e.g.*, Wahlberg et al., 2005), we did not include these because their identifications could not be confirmed. Complete lists of species, their code names, and the type of tissue used in DNA extractions can be found in Supplementary Tables 1 and 2.

2.3. DNA extraction

DNA was extracted from ethanol-preserved tissues using an OmniPrep™ DNA extraction kit (G-Biosciences) following the methods outlined in Espeland et al. (2018). When vouchering alcohol-preserved specimens, we used the protocol of Cho et al. (2016). Since we were able to obtain AHE probe and flanking sequence data from dried

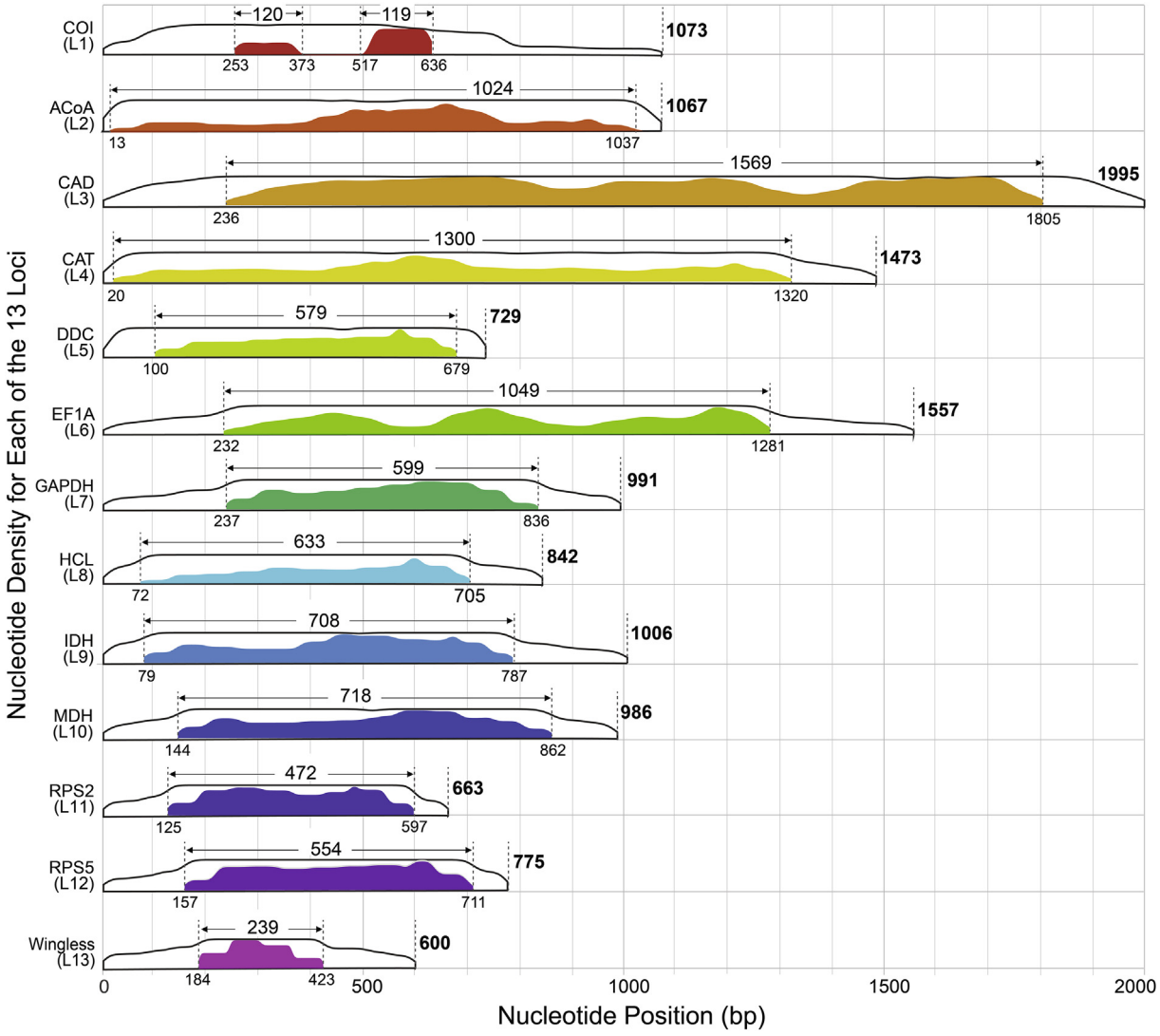

Fig. 1. The 13 loci in the BUTTERFLY2.0 kit showing their length, probe distribution, and density (capture success). The y-axis shows the amount of nucleotide capture density at each site, scaled to have the same maximum height for each gene. Colored areas indicate nucleotide densities of the BUTTERFLY2.0 probes for that locus; white areas under the curve are nucleotide densities for the 33 samples included in this study that used the BUTTERFLY2.0 kit. Sequence regions of the locus outside of the colored areas are the flanks. Numbers refer to a particular site (bp) along that gene. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

specimens in previous studies (Rubin et al., 2018; St Laurent et al., 2018), we tested whether the BUTTERFLY2.0 kit would also capture DNA from dried specimens. Twenty-two dried specimens that had been in the MGCL collection for decades were included in the present study (Supplementary Tables 1 and 2). For dry tissues, we followed modifications made to the OmniPrep™ extraction protocol described in St Laurent et al. (2018). DNA was extracted from abdomens and/or legs removed from these specimens.

Quantified DNA extracts were submitted to RAPiD Genomics (Gainesville, FL, USA) for library preparation, hybridization enrichment and sequencing. Random mechanical shearing of DNA was conducted with an average size of 300 bp followed by an end-repair reaction and ligation of an adenine residue to the 3′-end of the blunt-end fragments to allow ligation of barcoded adapters and PCR-amplification of the library. Following library construction, solution-based target enrichment of Agilent SureSelect probes was conducted in a pool containing 16 libraries. These libraries were enriched with the SureSelect Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol. The enriched libraries were multiplexed with other projects and sequenced for paired-end 100-bp reads with an Illumina HiSeq 3000 to generate approximately 150,000 paired reads per-sample (Supplementary Table 3).

2.4. Data assembly and cleanup

We used a pipeline for anchored phylogenomics (Breinholt et al., 2018) to create a data matrix from raw Illumina reads. Paired-end Illumina data were cleaned with Trim Galore! ver. 0.4.0 (www.bioinformatics.babraham.ac.uk), allowing a minimum read size of 30 bp, to remove bases with a Phred score below 20. Loci were assembled with iterative baited assembly (IBA; Breinholt et al., 2018) in which only reads with both a forward and reverse read that passed filtering were included. Assembled reads from the probe region were blasted against the *D. plexippus* reference genome and BLAST results were used for single hit and orthology filtering. Loci were screened for orthology with a single hit threshold of 0.9 and genome mapping following Breinholt et al. (2018). The identified orthologous sequences were screened for contamination by identifying and removing sequences that were nearly identical at the family and genus level (Breinholt et al., 2018). Loci were aligned with MAFFT v. 7.0.1 (Katoh

and Standley, 2013) and loci were concatenated with FASconCAT-G 1.0.4 (Kück and Meusemann, 2010). Aliscore 2.2 (Kück et al., 2010) was used to check for saturation and sites that appeared to evolve randomly.

2.5. Datasets, model partitioning and phylogenetic analysis

Two molecular data matrices were constructed for this study. Dataset 1 is a concatenated dataset of all 13 probe regions. Dataset 2 is a concatenated dataset of the 13 probe regions plus the flanking regions from each of the 13 loci (see Breinholt et al., 2018 for a discussion on probe and flanking regions). We estimated the best partitioning scheme and best model of nucleotide substitution with ModelFinder (Kalyaanamoorthy et al., 2017), using a partition input file that that divided the probe region by codon position. For Dataset 2, the 'head' and 'tail' flanking regions of all genes were treated as a single, separate division appended to the concatenated probe region. To find the most likely tree, 100 ML searches were done on both datasets in IQ-TREE v.1.5.3 (Nguyen et al., 2015), along with two calculations of nodal support: ultrafast bootstrap (UFBoot) and SH-aLRT tests. We generated 1000 replicates for UFBoot ("–bb" command) (Minh et al., 2013) and SH-aLRT ("-alrt" command) (Guindon et al., 2010). These measures have proven useful to assess the confidence of relationships (Minh et al., 2013), and have been applied in other phylogenetic studies that used AHE data (*e.g.*, St Laurent et al., 2018; Toussaint et al., 2018). We also generated 100 standard bootstrap (BS) replicates as a third measure of support for Dataset 2.

To account for possible incomplete lineage sorting, a species tree for Dataset 1 was also inferred using the coalescent summary method in ASTRAL-III v 5.5.9 (Mirarab and Warnow, 2015). ML gene trees inferred in IQ-TREE with best-fit models of nucleotide substitution selected for the ML analyses were used as input for coalescent summary methods. For ASTRAL, we calculated using the normalized quartet score (Sayyari and Mirarab, 2016), which is an overall measure of the amount of discordance in the gene trees (proportion of input gene tree quartet trees satisfied by the species tree). We also calculated the quartet support (percentage of quartets in gene trees that agree with a branch), to get an estimate of branch support. All phylogenomic analyses were conducted on the University of Florida HiPerGator High Performance Computing Cluster (www.hpc.ufl.edu).

2.6. Ancestral state reconstruction

Butterflies are predominantly diurnal, but many hedylid species are nocturnal. We tested the hypothesis that there was a single transition from diurnal to nocturnal activity in adult Hedylidae. We performed ancestral state reconstructions (ASR) of hedylid diel activity on the ML tree from Dataset 2 using parsimony in Mesquite 3.40 (Maddison and Maddison, 2018) and also applying a Bayesian method (stochastic character mapping) with the 'make.simmap' command in the R package Phytools v. 06–44 (Revell, 2012). The former ASR method allows multiple states to be assigned to a single species (*e.g.*, a species that is both diurnal and nocturnal), enabling more precise character scoring, whereas the latter method is generally considered more robust (Bollback, 2006). Diel activity was scored as 'diurnal' and/or 'nocturnal' for the parsimony ASR, and was scored as 'some diurnal records' or 'no diurnal records' for the Bayesian ASR. Ten thousand stochastic maps were generated when running 'make.simmap' for the Bayesian ASR, and a symmetric model was used (*i.e.*, forward and reverse character state transitions were assumed to have equal rates).

3. Results and discussion

3.1. The BUTTERFLY2.0 probe set and sequence capture success

A total of 14,433 probes, tiled across 13 loci, were included in the Agilent SureSelect Target Enrichment kit. The average number of raw paired-end reads per sample was 176,598. The minimum and maximum number of paired-end reads recovered were 63,975 and 260,592 respectively, across all taxa sequenced using the BUTTERFLY2.0 probe set (Fig. 1; Supplementary Table 3). The average probe region length was 756 bp, and flanking regions averaged 302 bp. Two of the longest loci were CAD and CAT, which had probes spanning 1569 and 1300 bp of the gene, respectively. These two loci reached up to 1995 and 1473 bp when the flanking regions were included (Table 1). Eleven loci were captured successfully from 100% of the taxa extracted using the BUTTERFLY2.0 probe set, and the remaining two loci, COI and wingless, were captured from 94% and 97% of taxa, respectively (Supplementary Table 2). The BUTTERFLY2.0 loci, probe design information, probe-only gene trees, and the concatenated probe + flank alignment are available via Dryad (DOI: 10.5061/dryad.t5sd8c8). See Supplementary Table 4 for a list of partitions and models applied to the two datasets.

3.2. Phylogenetic analysis and evolution of diel activity in Hedylidae

We sampled 22 hedylid species and show that Hedylidae is sister to Hesperiidae (Figs. 2, S1, S2) with strong support (BS/UFBoot/SH-aLRT = 100; Figs. 2, S1, S3), a result congruent with other molecular phylogenetic studies (*e.g.,* Regier et al. 2009; Mutanen et al. 2010; Heikkilä et al., 2012; Kawahara and Breinholt, 2014; Breinholt et al., 2018; Espeland et al., 2018; Toussaint et al., 2018). There was no topological difference between the ML trees from Datasets 1 and 2 (Figs. S1, S2), and both analyses resulted in robust support for most relationships (22 and 23 nodes with > 95% UFBoot, 18 and 20 nodes with > 95% SH-aLRT, for Datasets 1 and 2 respectively; Figs. S1, S2).

In all analyses conducted in this study, *Macrosoma tipulata* Hübner is the sister species to all remaining hedylids with strong support (BS/UFBoot/SH-aLRT = 100; Figs. 2, S1). This result is congruent with Prout (1910; 1932), Scoble (1986) and Lourido (2011) who noted that *M. tipulata* is morphologically distinct from the remaining species in the family. Within that subclade, the lineage containing *M. semiermis* (Prout) diverges from the remaining hedylids; this position is also well-supported in all analyses (BS/UFBoot/SH-aLRT = 100; Figs. 2, S1) and congruent with the morphological analysis of Lourido (2011). The remainder of the Hedylidae (Clade A) were also well-supported (BS/UFBoot/SH-aLRT = 100; Figs. 1, S1), and this clade was divided into two clades (B, C) with strong support (BS/UFBoot/SH-aLRT = 100; Figs. 1, S1). All of the species in our dataset that were originally placed in *Phellinodes* by Prout (1932) are contained within Clade A. However, relationships among species in Clade A are not congruent with Prout's divisions. For instance, Prout's *Lasiopates* (containing the single species *M. hyacinthina*) is nested within Clade A, sister to *M. hedylaria*, with strong support. Prout further classified these species into two sections (*Hyphedyle* and *Phellinodes*) based on presence/absence of a modified, semi-translucent region of the male hindwing. These two subgenera do not correspond to Clades B and C. Lourido's (2011) concept of *Phellinodes* also corresponds, very roughly, with Clade A. However, as with Prout's concept, internal relationships of Lourido's *Phellinodes* are notably different. For example, Lourido placed *M. subornata* (Warren) sister to all other *Phellinodes*, whereas in our analysis, *M. subornata* is placed within Clade C, sister to *M. muscerdata* (Felder and Rogenhofer).

Scoble (1990b) hypothesized that *M. lamellifera, M. rubedinaria,* and *M. ustrinaria* (which were all part of Prout's *Phellinodes* (section *Hyphedyle*)) are closely related. All of these species fall into Clade C, although they do not form a monophyletic group. Lourido (2011) concluded that species that are grouped into Clade B in our phylogeny constitute a monophyletic group, but our results show that species that fall into Clade C form a paraphyletic grade with weak branch support. Clades B and C are characterized by species that lack a conspicuously notched forewing apex and tend to have a dark-bordered hindwing (Clade B), or have a conspicuously notched forewing apex (Clade C; Supplementary File 1). The ASTRAL tree based on the coalescent was

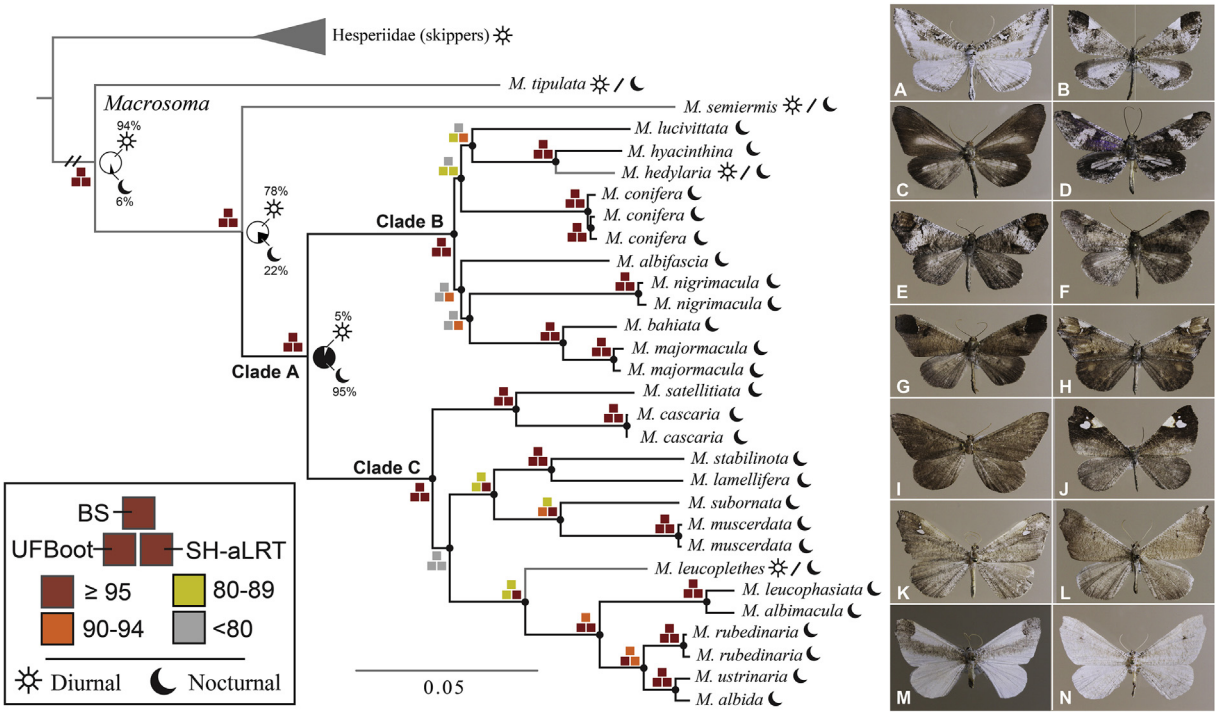

Fig. 2. Maximum likelihood tree of Hedylidae (Dataset 2) showing evolution of diel activity patterns. Colored squares indicate measures of bootstrap (BS), ultrafast bootstrap (UFBoot), and SH-aLRT support. Ancestral nodes show diel probabilities (circles at nodes) obtained from a Bayesian ASR analysis. Grey branches correspond to species that have been recorded to be active during the daytime. Images on the right correspond to the following species: A) *M. tipulata,* B) *M. semiermis,* Clade B: C) *M. lucivittata,* D) *M. hyacinthina,* E) *M. albifascia,* F) *M. nigrimacula,* G) *M. majormacula,* Clade C: H) *M. cascaria,* I) *M. lamellifera,* J) *M. subornata,* K) *M. muscerdata,* L) *M. rubedinaria,* M) *M. leucophasiata,* N) *M. ustrinaria.* The horizontal bar at the bottom of the ML tree indicates the number of substitutions per site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

topologically congruent with the ML trees, although particular nodes, especially in Clade B, were not strongly supported, indicating some gene tree discordance in the dataset (Fig. S4).

Based on the lack of consistent morphological traits across Prout's generic definitions, Scoble (1986) synonymized all genera, and recognized *Macrosoma* as the senior synonym. Lourido (2011) recognized three genera in Hedylidae, but the revalidation of *Hedyle* and *Phellinode* in that work was not formally published, and thus only *Macrosoma* is considered valid. Since *M. tipulata* is the type species and is sister to the remainder of the Hedylidae (Fig. 2), we retain the classification of Scoble (1986) and recognize one genus, *Macrosoma*, for the family.

Butterflies are one of the few groups of Lepidoptera in which the adults of most species are diurnal. Among butterflies, Hedylidae is unusual in being the only family in which the majority of species are nocturnal (Scoble and Aiello, 1990; Kawahara et al., 2018). We examined the evolution of diel activity in Hedylidae to test the hypothesis that there was a single transition to nocturnality from an ancestrally diurnal butterfly. Both of our ancestral state reconstructions show that the hedylid ancestor, like other butterflies, exhibited diurnal activity, followed by a transition to strict nocturnality in Clade A (Figs. 2, S5, S6). However, the parsimony reconstruction, which allows for multiple character state assignments to a single species, shows a more gradual transition, with an intermediate step of both diurnal and nocturnal behavior (Fig. S6). *Macrosoma leucoplethes* (Prout) and *M. hedylaria* (Warren) have been observed flying both during the day and at night (N. Homziak pers. comm.; S. Vargas pers. comm.) and these two species appear to have each undergone independent reversals to partial diurnality. *Macrosoma heliconiaria,* a noctunal species, is in the same species group as *M. semiermis*, which is both nocturnal and diurnal (Prout, 1932; Scoble, 1990a). *Macrosoma heliconiaria* has bat ultrasound-detecting hearing organs on their wings (Yack and Fullard, 2000), but hearing sensitivity has never been studied in any other hedylid, including the day-flying species. Given the general morphological similarity among *Macrosoma* species (Scoble, 1990b), it is likely that other species in the genus may also have ultrasound-detecting ears. Future studies should examine whether day-flying species have different hearing organ morphology or reduced auditory sensitivity, similar to other moth species that inhabit bat-free environments (Fullard et al., 2004).

Hedylidae are the least taxonomically diverse and most geographically restricted butterfly family. They are only found in tropical Central and South America, whereas all other butterfly families are cosmopolitan in distribution and have at least one or two orders of magnitude more species. Hedylidae is estimated to have split from its sister taxon, Hesperiidae, at 106 Ma (95% HPD: 80–129 Ma), but extant species did not start to diversify until around 29 Ma (95% HPD: 18–42 Ma) (Espeland et al., 2018). This suggests that the lineage persisted for tens of millions of years before extant species began diversifying in the mid- to late Paleogene. Interestingly, some species of moth-like butterflies are very widespread whereas others are only known from small geographical ranges (Lourido, 2011; Scoble, 1990b). Future investigations will add missing species to our dataset and analyze results in a temporal framework to study the biogeographic history of Hedylidae, and improve understanding of the mechanisms of lineage diversification at a regional scale. Based on the capture success of the BUTTERFLY2.0, we predict that this probe set will be useful for future studies on the butterfly tree of life.

Acknowledgements

We thank the members of the Kawahara Lab and curators and collection managers of the Florida Museum of Natural History, McGuire Center for Lepidoptera and Biodiversity. Stacey Huber took photographs of specimens in Fig. 2. Geoff Martin (NHM) and Patricia Gentili-Poole (USNM) allowed access to type specimens and photographs. Jesse Barber and members of his lab (Boise State University) accompanied

field trips and helped collect some specimens that were sequenced. RAPiD Genomics (Gainesville, FL, USA) is acknowledged for conducting library preps and DNA sequencing. The University of Florida HiPerGator provided computational resources. AYK, DJL, JWB, ME and NEP acknowledge funding from DEB 1541500, 1541557, and 1541560; financial support for images included in this paper were provided by DBI 1601369 to AYK.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.06.002.

References

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2012. GenBank. Nucleic Acids Res. 41, D36–D42.
Bollback, J.P., 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinf. 7, 88.
Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L., Kawahara, A.Y., 2018. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. Syst. Biol. 67, 78–93.
Cho, S., Epstein, S.W., Mitter, K.T., Hamilton, C.A., Plotkin, D., Mitter, C., Kawahara, A.Y., 2016. Preserving and vouchering butterflies and moths for large-scale museum-based molecular research. PeerJ 4, e2160.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.
Espeland, M., Breinholt, J.W., Willmott, K.R., Warren, A.D., Vila, R., Toussaint, E.F.A., Maunsell, S.C., Aduse-Poku, K., Talavera, G., Eastwood, R., Jarzyna, M., Reis, L., Guralnick, R., Lohman, D.J., Pierce, N.E., Kawahara, A.Y., 2018. A comprehensive and dated phylogenomic analysis of butterflies. Curr. Biol. 28, 770–778.
Fullard, J.H., Ratcliffe, J.M., Soutar, A.R., 2004. Extinction of the acoustic startle response in moths endemic to a bat-free habitat. J. Evol. Biol 17, 856–861.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A., 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23, 167–172.
Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270, 313–321.
Heikkilä, M., Kaila, L., Mutanen, M., Pena, C., Wahlberg, N., 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proc. R. Soc. B-Biol. Sci. 279, 1093–1099.
ICZN, 1999. International Code of Zoological Nomenclature. 4. The International Trust for Zoological Nomenclature, London.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kawahara, A.Y., Breinholt, J.W., 2014. Phylogenomics provides strong evidence for relationships of butterflies and moths. Proc. R. Soc. Lond. B 281, 20140970.
Kawahara, A.Y., Plotkin, D., Hamilton, C.A., Gough, H., St Laurent, R.A., Owens, H.L., Homziak, N.T., Barber, J.R., 2018. Diel behavior in moths and butterflies: a synthesis of data illuminates the evolution of temporal activity. Org. Divers. Evol. 18, 13–27.
Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., 2015. Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 44, D574–D580.
Kozak, K.M., Wahlberg, N., Neild, A.F.E., Dasmahapatra, K.K., Mallet, J., Jiggins, C.D., 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst. Biol. 64, 505–524.
Kück, P., Meusemann, K., 2010. FASconCAT: convenient handling of data matrices. Mol. Phylogenetics Evol. 56, 1115–1118.
Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wagele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7, 1–12.
Lévêque, A., 2007. Une nouvelle espèce de *Macrosoma* Hübner, 1818 (Lepidoptera Hedylidae). L'Entomologiste 63, 97–101.
Lourido, G.M., 2011. Revisão taxonômica e análise filogenética da familia neotropical Hedylidae (Lepidoptera, Hedyloidea). Tese de douturado. INPA/UFAM, Manaus.
Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3.40 http://mesquiteproject.org.
Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.
Mirarab, S., Warnow, T., 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52.
Mitter, K.T., Larsen, T.B., De Prins, W., Collins, S., Vande Weghe, G., Safian, S., Zakharov, E.V., Hawthorne, D.J., Kawahara, A.Y., Regier, J.C., 2011. The butterfly subfamily Pseudopontiinae is monobasic: marked genetic diversity and morphology reveal three new species of *Pseudopontia* (Lepidoptera: Pieridae). Syst. Ent. 36, 139–163.
Mutanen, M., Wahlberg, N., Kaila, L., 2010. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. Proc. R. Soc. Lond. B 277, 2839–2848.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Prout, L.B., 1910. Lepidoptera Heterocera, Fam. Geometridae, Subfam. Oenochrominae. In: Wytsman, P. (Ed.), Genera Insectorum. Belgium, Brussels, pp. 1–119.
Prout, L.B., 1932. The American Geometridae. In: In: Seitz, A. (Ed.), The Macrolepidoptera of the World Vol. 8 Alfred Kernen, Stuttgart.
Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: The barcode of life data system (www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364.
Regier, J.C., Zwick, A., Cummings, M.P., Kawahara, A.Y., Cho, S., Weller, S., Roe, A., Baixeras, J., Brown, J.W., Parr, C., Davis, D.R., Epstein, M., Hallwachs, W., Hausmann, A., Janzen, D.H., Kitching, I.J., Solis, M.A., Yen, S.H., Bazinet, A.L., Mitter, C., 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. BMC Evolutionary Biol. 9, 280.
Revell, L.J., 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223.
Rubin, J.J., Hamilton, C.A., McClure, C.J.W., Chadwell, B.A., Kawahara, A.Y., Barber, J.R. The evolution of anti-bat sensory illusions in moths, Sci. Adv. (in press).
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.
Scoble, M.J., 1986. The structure and affinities of the Hedyloidea: a new concept of the butterflies. Bull. Br. Museum (Nat. History) 53, 251–286.
Scoble, M.J., 1990a. A catalogue of the Hedylidae (Lepidoptera: Hedyloidea), with descriptions of two new species. Insect Syst. Evol. 21, 113–119.
Scoble, M.J., 1990b. An identification guide to the Hedylidae (Lepidoptera: Hedyloidea). Guía de identificación para los Hedylidae (Lepidoptera: Hedyloidea). Entomologica Scandinavica 21, 121–158.
Scoble, M.J., Aiello, A., 1990. Moth-like butterflies (Hedylidae: Lepidoptera): a summary, with comments on the egg. J. Nat. Hist. 24, 159–164.
Shen, Y.Y., Chen, X., Murphy, R.W., 2013. Assessing DNA barcoding as a tool for species identification and data quality control. Plos One 8.
Simonsen, T.J., Zakharov, E.V., Djernaes, M., Cotton, A.M., Vane-Wright, R., Sperling, F.A., 2011. Phylogenetics and divergence times of Papilioninae (Lepidoptera) with special reference to the enigmatic genera *Teinopalpus* and *Meandrusa*. Cladistics 27, 113–137.
St Laurent, R.A., Hamilton, C.A., Kawahara, A.Y., 2018. Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonidae). Syst. Entomol. http://dx.doi.org/10.1111/syen.12301. (in press).
Toussaint, E.F.A., Breinholt, J.W., Earl, C., Warren, A.D., Brower, A.V.Z., Yago, M., Dexter, K.M., Espeland, M., Pierce, N.E., Lohman, D.J., Kawahara, A.Y., 2018. Anchored phylogenomics illuminates the skipper butterfly tree of life. BMC Evolutionary Biol. http://dx.doi.org/10.1186/s12862-018-1216-z. (in press).
Triant, D.A., Cinel, S.D., Kawahara, A.Y., 2018. Lepidoptera genomes: Current knowledge, gaps and future directions. Curr. Opin. Insect Sci. 25, 99–105.
Wahlberg, N., Braby, M.F., Brower, A.V.Z., de Jong, R., Lee, M.-M., Nylin, Pierce, N., Sperling, F.A., Vila, R., Warren, A.D., Zakharov, E., 2005. Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proc. R. Soc. B 272, 1577–1586.
Wahlberg, N., Wheat, C.W., 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. Syst. Biol. 57, 231–242.
Wahlberg, N., Peña, C., Ahola, M., Wheat, C.W., Rota, J., 2016. PCR primers for 30 novel gene regions in the nuclear genomes of Lepidoptera. Zookeys 596, 129–141.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Yack, J.E., Kalko, E.K.V., Surlykke, A.M., 2007a. Neuroethology of ultrasonic hearing in nocturnal butterflies (Hedyloidea). J. Comparat. Physiol. A 193, 577–590.
Yack, J.E., Johnson, S.E., Brown, S.G., Warrant, E.J., 2007b. The eyes of *Macrosoma* sp. (Lepidoptera: Hedyloidea): a nocturnal butterfly with superposition optics. Arthropod Struct. Dev. 36, 11–22.
Yack, J.E., Fullard, J.H., 2000. Ultrasonic hearing in nocturnal butterflies. Nature 403, 265–266.