Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives

Hamilton CA[1,2*], St Laurent RA[1], Dexter, K[1], Kitching IJ[3], Breinholt JW[1,4], Zwick A[5], Timmermans MJTN[6], Barber JR[7], Kawahara AY[1*]

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 USA

[2]Department of Entomology, Plant Pathology, & Nematology, University of Idaho, Moscow, ID 83844 USA

[3]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK

[4]RAPiD Genomics, 747 SW 2nd Avenue #314, Gainesville, FL 32601. USA

[5]Australian National Insect Collection, CSIRO, Clunies Ross St, Acton, ACT 2601, Canberra, Australia

[6]Department of Natural Sciences, Middlesex University, The Burroughs, London NW4 4BT, UK

[7]Department of Biological Sciences, Boise State University, Boise, ID 83725, USA

[*]Corresponding authors: hamiltonlab@uidaho.edu (CAH) or kawahara@flmnh.ufl.edu (AYK)

Abstract

Background

The silkmoths and their relatives constitute the ecologically and taxonomically diverse superfamily Bombycoidea, which includes some of the most charismatic species of Lepidoptera. Despite displaying some of the most spectacular forms and ecological traits among insects, relatively little attention has been given to understanding their evolution and the drivers of their diversity.

Results

To begin to address this problem, we created a new Bombycoidea-specific Anchored Hybrid Enrichment (AHE) probe set and sampled up to 571 loci for 117 taxa across all major lineages of the Bombycoidea, producing a well-supported phylogeny. The tree was overall consistent with prior morphological and molecular studies, although some taxa (e.g., *Arotros* Schaus) were misplaced in the Bombycidae and here formally transferred to Apatelodidae. We identified important evolutionary patterns (e.g., morphology, biogeography, and differences in speciation and extinction), and our analysis of diversification rates highlights the stark increases that exist within the Sphingidae (hawkmoths) and Saturniidae (wild silkmoths).

Conclusions

We postulate that these rate shifts are due to the well-documented bat-moth “arms race” and differences in selective pressures from insectivorous bats. The study establishes a backbone for future evolutionary, comparative, and taxonomic studies, and presents a modified DNA extraction protocol that allows Lepidoptera specimens to be readily sequenced from pinned natural history collections, succeeding in samples up to 30 years old. Our research highlights

the flexibility of AHE to generate genomic data from a wide range of museum specimens, both age and preservation method, and will allow researchers to tap into the wealth of biological data residing in natural history collections around the globe.

Keywords

Anchored Hybrid Enrichment, targeted sequence capture, phylogenomics, Bombycoidea, Lepidoptera, natural history collections

67 **Background**

68 The Bombycoidea include some of the most charismatic moths among all Lepidoptera.
69 This ecologically diverse superfamily comprises ten families, 520 genera, and 6,092 species [1].
70 Although widespread globally, the highest diversity of bombycoids occurs in the tropics and
71 subtropics. This diversity includes the most spectacular forms (i.e., range of body sizes and wing
72 shapes) and functions (i.e., mimicry, predator avoidance, flight capabilities, and feeding
73 strategies) in the Lepidoptera [2].

74 A number of bombycoids have become important contributors to human culture,
75 originally as economically important species for sericulture or as agricultural pests, but more
76 recently as model organisms for comparative studies of genetics, development, and physiology
77 [2]. Additionally, many lineages play important roles as pollinators ([3], [4], [5], [6], [7], [8], [9]),
78 or as indicators in biodiversity and habitat quality assessments [10]. Of the 10 families, three
79 contain species that have been used as model organisms (Bombycidae, Saturniidae, and
80 Sphingidae). Unfortunately, relationships among bombycoid families, and especially these three
81 families, have remained largely elusive. For example, Bombycidae have been considered the
82 sister lineage to either Saturniidae or Sphingidae ([11], [12], [13], [14], [15], [16], [17], [18]); but
83 see Breinholt and Kawahara [19]. Overall, relatively little attention has been given to the group
84 with regards to understanding their evolution.

85 Despite their charisma and intrigue, the lack of a robust phylogeny based on broad and
86 dense taxon sampling across the Bombycoidea is dramatically affecting our ability to answer
87 fundamental questions about the drivers of their diversity. Monophyly of the Bombycoidea has
88 been supported by six morphological synapomorphies [15], but Zwick [20] determined that only

89 two of these were systematically informative: one poorly understood thoracic character [21],
90 and one relating to the arrangement of the forewing vein (A. Zwick, unpublished). Recent
91 molecular studies of bombycoid systematics ([11], [13], [20]) have resulted in substantial
92 differences in terms of relationships from morphology-based phylogenetic hypotheses ([15],
93 [21], [22]). To date, nearly all molecular studies of bombycoids have included fewer than 20
94 protein-coding genes for ≤50 species (e.g., [11], [13], [23], [24]). Although these studies agreed
95 on the monophyly of the superfamily, many relationships among families, subfamilies, and
96 tribes remained unclear and were characterized by weak branch support or conflicting signal.
97 Modern phylogenomics (e.g., based on Anchored Hybrid Enrichment) has been shown to be
98 effective to resolve relationships among Lepidoptera at multiple taxonomic levels ([16], [25],
99 [26], [27], [28]) including the Bombycoidea ([16], [29]). However, those studies that utilize
100 phylogenomics to resolve inter- and intra-familial relationships within the Bombycoidea have
101 been limited by taxon sampling.
102 To better answer questions regarding bombycoid evolution, we applied Anchored
103 Hybrid Enrichment (AHE) targeted-sequencing phylogenomics [30]. We developed a new,
104 bombycoid-specific, AHE probe set (here called “BOM1”), redesigned from the “LEP1” probe set
105 of Breinholt et al. [16]. Our new probe set captures “legacy” Sanger sequencing-based loci that
106 were part of existing bombycoid molecular datasets ([11], [13], [24], [31]), enabling the merging
107 of older published datasets with those generated from the BOM1 probe set. To further improve
108 the ability to generate a dataset with greater taxon and locus sampling, we developed a new
109 method for extracting DNA from pinned natural history specimens, and show that our

110 extraction approach is successful in obtaining DNA sequence data for phylogenomics. In total,
111 our dataset resulted in 571 loci for 117 species across Bombycoidea.
112 We also use this opportunity to utilize the phylogeny to examine patterns of
113 diversification in the superfamily. Bombycoidea are well-known to have multiple different
114 ecological life-history strategies, especially for the Saturniidae and Sphingidae – two lineages
115 that harbor the majority of described bombycoid species ([29], [32], [33], [34]). The divergent
116 life-history strategies of these two families ([35], [36]) has likely played a major role in driving
117 their diversity. For example, the majority of hawkmoths feed as adults, seeking out nectar
118 resources during their relatively long lives (weeks to months). During this time, females
119 experience multiple mating events and retain the eggs internally for long periods to allow egg
120 maturation and host plant discovery [35]. This ecological strategy is significantly different from
121 saturniids ([23], [19], [37]), which depend entirely upon the resources acquired during the larval
122 period. Adult saturniids possess reduced or non-functional mouthparts and lay eggs almost
123 immediately after mating. Furthermore, these lineages possess a number of different traits
124 that appear to be anti-bat adaptations in response to echolocating bats, a lineage thought to
125 have arisen approximately 60 million years ago ([38], [39], [40], [41], [42]). Hawkmoths,
126 especially those in the Sphinginae and Macroglossinae, are strong fliers thought to have
127 evolved hearing organs and ultrasound producing organs capable of jamming bat sonar ([33],
128 [43]). Saturniids, in contrast, lack hearing organs and exhibit erratic evasive flight [35] and
129 hindwing tails that deflect bat echoes ([29], [34]). These traits firmly establish the relative
130 ecological roles of Saturniidae and Sphingidae as an important natural experiment from which

131 we can gather valuable information regarding the evolution of predator/prey interactions in
132 moths and their ultimate effects on the diversification process in general.
133
134 **Methods**
135 *The "BOM1" AHE probe set design*
136 Anchored Hybrid Enrichment (AHE) is a targeted-sequencing methodology designed to
137 capture hundreds of unique orthologous loci (i.e., single copy, phylogenetically-informative
138 markers) from across the genome, for resolving both shallow and deep-level evolutionary
139 relationships ([11], [13]). Probes are designed to anchor in conserved regions that are flanked
140 by variable regions randomly spread throughout the genome. This approach creates a diverse
141 set of informative loci that include exons, introns, intergenic, and conserved regions of the
142 genome. Targeted-sequencing approaches, like AHE and UCE (Ultraconserved Elements),
143 provide mechanisms whereby different researchers can confidently and effectively use the
144 same loci for independent projects, allowing for the combination of data across studies.
145 Breinholt et al. [16] constructed a Lepidoptera Agilent Custom SureSelect Target
146 Enrichment "LEP1" probe kit, designed for 855 loci. However, this probe set is not specific to
147 Bombycoidea, and does not include some of the traditional loci that have been used in
148 phylogenetics of Bombycoidea. In order to build a more Bombycoidea-specific AHE probe set
149 and phylogenomic dataset, we began by modifying the LEP1 kit, evaluating which loci were
150 most phylogenetically-informative within the superfamily, optimizing the set of probes to
151 recover these loci, and including 24 previously-sequenced Sanger-sequenced loci, such as e.g.,
152 CO1, CAD, DDC, period, wingless and others from [11], [13], [24], and [31], as well as eight

153 vision-related genes (see Supp. Table 1 for locus names). The probes for these vision-related
154 genes are based phototransduction genes mined from eye or head transcriptomes
155 (unpublished; generated by AYK), and are included for future analyses to investigate their
156 evolution across the superfamily.

157 To determine the informative loci for BOM1, phylogenetically-informative loci were
158 identified by examining the sequence variation of 56 bombycoid species from across the
159 taxonomic breadth of the superfamily (55 AHE samples, from either Breinholt et al. [16] or
160 generated for this study, plus loci mined from the *Bombyx mori* reference genome [44]). These
161 samples were sequenced and processed using the LEP1 probe kit and Breinholt et al. [16]
162 bioinformatics pipeline [45]. Individual gene trees were generated for each of the 855 loci,
163 using Maximum Likelihood in RAxML v8.2 [46], under a GTRGAMMA model of evolution with
164 100 non-parametric bootstraps for node support. For each tree, phylogenetic informativeness
165 was calculated using PhyDesign ([47], [48], [49]). Parsimony informative characters and the
166 number of segregating sites were calculated using the R packages 'phyloch' [50] and 'ape' [51]
167 respectively. Additionally, each phylogeny was scrutinized visually to determine whether
168 branch lengths and topological patterns appeared realistic (i.e., no significant outliers present).
169 Loci deemed to be phylogenetically uninformative, or those that were capturing poorly (<60%
170 of sampled species represented for a locus), were excluded from the probe set. The final BOM1
171 probe kit comprised 571 loci, 539 of which came from the original LEP1 kit.

172

173 *Taxon sampling*

174 To build a backbone phylogeny of the Bombycoidea, we sampled all major lineages (i.e.,
175 families, subfamilies, and tribes), with the exception of three rare, species-poor groups:
176 subfamily Munychryiinae (Anthelidae) and tribes Sataspedini and Monardini (Sphingidae), for
177 which representative samples were unavailable for DNA sequencing. Sampled lineages were
178 chosen because: 1) they were appropriate representatives of the taxonomic group needed for
179 the analysis (i.e., good morphological and evolutionary representative of a tribe); and 2) they
180 were accessible for use in phylogenomics. In total, 115 ingroup Bombycoidea species from 97
181 genera were included in the phylogenetic analysis, as well as two Lasiocampidae outgroups –
182 the sister lineage to the bombycoids (Supp. Table 2).
183 Specimens were obtained from fieldwork, historically preserved dry collections (Supp.
184 Table 1), and molecular tissue collections. Field-collected specimens were stored in ≥95%
185 ethanol, RNAlater, or papered and dried with silica gel. Genomic DNA was extracted using
186 OmniPrep Genomic DNA Extraction Kits (G-Biosciences, St. Louis, MO, USA) and DNeasy Blood
187 and Tissue Kits (Qiagen, Valencia, CA, USA). DNA concentration was evaluated through agarose
188 gel electrophoresis and fluorometry using a Qubit 2.0 (Invitrogen, Thermo Fisher Scientific,
189 Carlsbad, CA, USA). Library preparation, hybridization enrichment, and Illumina HiSeq 2500
190 sequencing (PE100) was carried out at RAPiD Genomics (Gainesville, FL, USA). Specimen wing
191 vouchering and tissue storage methods follow Cho et al. [52]. All DNA extracts and specimens
192 preserved in ethanol, RNAlater, or those freshly papered were stored at -80°C at the Florida
193 Museum of Natural History, McGuire Center of Lepidoptera and Biodiversity (MGCL).
194
195 *DNA extraction protocol for museum specimens*

196 We evaluated the efficacy of obtaining DNA from historical museum specimens because
197 there is great interest in understanding the feasibility of this approach for use in phylogenetics
198 and systematics ([27], [53], [54], [55]). The Lepidoptera specimens evaluated herein, were
199 "field-pinned" (never rehydrated), "papered" (stored in an envelope and kept dry since
200 collected, thus not rehydrated or pinned), and "traditionally-pinned" specimens (dried,
201 rehydrated, and subsequently pinned) – the historically most common method of Lepidoptera
202 specimen storage. Collecting dates ranged from 1987 to 2017 (Supp. Table 2). Samples were
203 not initially intended to be preserved for molecular sequencing and information about potential
204 contaminants and/or extraction inhibitors (i.e. fumigation compounds, other chemicals) was
205 unavailable. In some cases there was little soft-tissue to extract from within the abdomens,
206 having degraded over time. Being that these samples were not kept in molecular-grade
207 conditions, many were contaminated with fungal and bacterial growth, which could be
208 identified visually by the spores left on the bodies or by the smell of decay. These factors along
209 with the amount of sclerotized tissue present and the fact that the abdomens used needed to
210 stay intact (not homogenized) for dissecting purposes, made extracting good quality genomic
211 DNA challenging.
212 Our extraction method, detailed in the Supplemental information, attempts to account
213 for several factors: the amount of degraded tissue, the presence of eggs, the relative fat
214 content, and the overall abdomen size. Many commercial DNA extraction kits on the market
215 (including the Omni Prep kit used in this study) recommend using 10 mg-20 mg of well
216 preserved, soft tissue for the extraction process. Given that the museum specimens used had
217 been desiccated for many years, a number of abdomens had little to no visible internal soft

218 tissue remaining. To digest the remaining material in solution, we increased the ratio of
219 proteinase K to lysis buffer. Companies that produce DNA extractions kits know, as can be seen
220 by their specific extraction kits, the relative fat content of samples is an issue because lipid-rich
221 tissue can interfere with the digestion of the soft tissue as well as change the chemistry of the
222 DNA isolation buffers. Specimens that appeared to be "greasy" or seemed to have an oily film
223 on the abdomen were not used for extraction. The overall size of the abdomen was used to
224 estimate the amount of lysis buffer needed in order to sufficiently submerge the abdomen and
225 reach the available soft tissue. The amount of buffer also reflected the volume of reagents
226 needed for the remainder of the DNA isolation process. Lastly, the modified protocol also
227 allows a user to easily prepare the genitalia for taxonomic or morphological taxonomic work as
228 these structures remain undamaged.
229
230 *Bioinformatics*
231 The bioinformatics pipeline of Breinholt et al. [16] was used to clean and assemble raw
232 Illumina reads for each AHE locus. The pipeline uses a probe-baited iterative assembly that
233 extends beyond the probe region, checks for quality and cross contamination due to barcode
234 leakage, removes paralogs, and returns a set of aligned orthologs for each locus and taxon of
235 interest. To accomplish these tasks, the pipeline uses the *Bombyx mori* genome [44], and an
236 AHE reference library, which in this study was the BOM1 reference library.
237 Loci for phylogenetic analysis were selected by applying a cutoff of ≥40% sampled taxa
238 recovery (i.e., for a locus to be included in the analysis, the locus had to be recovered in at least
239 40% of the sampled taxa). The pipeline evaluates density and entropy at each site of a

240 nucleotide sequence alignment. We elected to trim with entropy and density cutoffs only in
241 “flanking” regions, allowing the “probe” region (exon) to be converted into amino acid
242 sequences. For a site (outside of the probe region) to remain, that site must pass a 60% density
243 and 1.5 entropy cutoff, rejecting sites that fail these requirements. A higher first value (60)
244 increases the coverage cutoff (e.g., a site is kept if 60% of all taxa are represented at that site).
245 A higher second value (1.5) increases the entropy cutoff (i.e., entropy values represent the
246 amount of saturation at a site); sites with values higher than 1.5 possess higher saturation and
247 are thus deleted). AliView v1.18 [56] was used to translate to amino acids, check for frame
248 shifts, recognize and remove stop codons, and edit sequencing errors or lone/dubious indels.
249 Because flanking sequences are generally non-coding and sites have been deemed homologous
250 (see [16]), these flanking sequences, before and after the probe regions, were separated from
251 the exons, then combined and treated together as an independent partition. Due to the
252 filtering steps in the bioinformatics pipeline (i.e., site orthology, and density and saturation
253 evaluation), the flanking partition can be viewed as a SNP supermatrix, where each site is
254 homologous, but uninformative sites, saturated sites, or sites with large amounts missing data
255 have been removed.

256 Of the 115 bombycoid and two outgroup specimens, 110 were sequenced directly using
257 AHE target capture sequencing, of which 68 were sequenced using the BOM1 and 42 using the
258 LEP1 kit. Seven specimens had their AHE loci probe regions mined from previously sequenced
259 transcriptomes or the *B. mori* genome (Supp. Table 2). These specimens did not have flanking
260 data because of nature of transcriptome data. All specimens were processed using either the
261 ‘Bmori’ (for LEP1) or ‘BOM1’ (for BOM1) reference libraries. Previous Breinholt et al., [16] a

262 scripts are available in Dryad [45]. Instructions on how to use the pipeline and additional scripts
263 that were not part of Breinholt et al. [16] are provided in Dryad (will update upon acceptance).
264
265 *Phylogenetics*
266 Lepidoptera AHE probe sets comprise highly-conserved coding probe regions (i.e.,
267 exons) and more variable, generally non-coding flanking regions (e.g., introns or intergenic
268 regions) located on either side of the probe region [16]. We evaluated both nucleotide and
269 amino acid datasets to examine phylogenetic signal and the role that saturation may play in the
270 data (see [19]). With the Breinholt et al. [16] pipeline and scripts [45], three datasets were built
271 for phylogeny inference: 1) AA = an amino acid supermatrix composed of translated probe
272 region loci; 2) Pr+Fl = a probe + flanking supermatrix; and 3) ASTRAL = the individual loci from
273 the Pr+Fl supermatrix, used for individual gene tree inference and then species tree estimation
274 to evaluate the potential effects of deep coalescence. Additional analyses (i.e., probe region-
275 only, as nucleotides) were investigated, but are not reported here because their outcomes did
276 not differ from those reported.
277 Concatenated supermatrices were assembled using FASconCAT-G v1.02 [57].
278 Phylogenetic inference was performed in a maximum likelihood (ML) framework using IQ-TREE
279 MPI multicore v1.5.3 [58]. For both nucleotide and amino acid datasets, the '–m TEST'
280 command was used in IQ-TREE to perform a search for the most appropriate model of amino
281 acid or nucleotide substitution. For all inferences, we performed 1000 random addition
282 sequence (RAS) replicates, and 1000 replicates each for both ultrafast bootstraps (UFBS) ('–bb'
283 command) and SH-aLRT tests ('-alrt' command). The SH-like approximate likelihood ratio test

284 (SH-aLRT) estimates branch support values that have been shown to be as conservative as the
285 commonly used non-parametric bootstrap values [59]. SH-aLRT and bootstrap values tend to
286 agree for data sets with strong phylogenetic signal (i.e., datasets with loci that are sufficiently
287 large in number of bases, and tips that share sufficient divergence between sequences).
288 Disagreements in branch support are thought to arise as a consequence of small sample size,
289 insufficient data, or saturated divergence levels (see [60]). We classified nodes as "robust" if
290 they were recovered with support values of UFBS ≥ 95 and SH-aLRT ≥ 80 ([59], [60]).
291 Because concatenation can be misleading when there are high levels of incomplete
292 lineage sorting or deep coalescence [61], we assessed the impact of potential gene-tree
293 discordance ([62], [63], [64]) by inferring a phylogeny for each individual locus, using IQ-TREE
294 under the same parameters as above (i.e., probe and single partition of flanking loci were
295 modeled by site). Species tree estimation was performed in ASTRAL-III [65]. ASTRAL is a
296 computationally efficient and statistically consistent (under the multi-species coalescent)
297 nonparametric method that takes input gene trees and estimates a highly accurate species
298 tree, even when there is a high level of incomplete lineage sorting (or deep coalescence) [66].
299 The use of ASTRAL is also an informative "data exploration" exercise with phylogenomic
300 datasets, providing valuable information regarding the level of general tree discordance across
301 your set of gene trees, and the potential presence of incomplete lineage sorting/deep
302 coalescence that should be investigated further. To evaluate node support on the species tree,
303 we used the ASTRAL support values (ASV) – local posterior probabilities that are more precise
304 than evaluating bootstrap values across a set of input trees [67]. ASTRAL support values were
305 determined to be "robust" if nodes were recovered with local posterior probabilities ≥ 0.95. All

306 pipeline steps and phylogenomic analyses were conducted on the University of Florida
307 HiPerGator HPC (http://www.hpc.ufl.edu/). All alignment FASTA files, loci information, partition
308 files, tree files, and other essential data files used for phylogenetic inference are available as
309 supplementary materials on Dryad (will update upon acceptance).
310
311 *Rogue taxon & outlier locus analyses*
312 We investigated whether our molecular data included rogue taxa or outlier loci that
313 were potentially influencing our phylogenetic results. A rogue taxon analysis was carried out
314 using the online version of RogueNaRok ([68], [69]), http://rnr.h-its.org/, on the 1000 ultrafast
315 bootstrap trees and the consensus tree from the Pr+Fl supermatrix. Outlier taxa and loci
316 analyses were carried out using Phylo-MCOA v.1.4 (PMCoA) [70] in R (on the 650 gene trees
317 used in the ASTRAL analysis). No rogue taxa or loci were found, and therefore we did not prune
318 any taxa or loci from subsequent analyses.
319
320 *Diversification rate analyses*
321 As an initial investigation into why some bombycoid lineages are more diverse than
322 others, we examined and quantified how diversification rates (the interplay between speciation
323 and extinction) have changed over time. Simply calculating species diversity per clade and
324 assuming extant diversity is a true indicator of increases in diversification rate could produce
325 significant biases in one's interpretations due to some charismatic lineages receiving more
326 taxonomic effort than their "boring" sister lineages (see [71], [72], [73], [74]). We therefore
327 applied BAMM [75] and 'BAMMtools' [76] to infer the number and location of

328 macroevolutionary rate shifts across our phylogeny, and visualize the 95% credible set of shift
329 configurations.

330 The fossil record of Lepidoptera, especially of the Bombycoidea, is poor [77]. Prior
331 lepidopteran studies that have included fossils in dating analyses have been scrutinized for
332 incorrect fossil identification or placement on phylogeny [78]. Because Lepidoptera fossils are
333 limited and characters are difficult to discern, we decided not to conduct a dating analysis for
334 this study. Instead, the ML best tree was converted into a relative-rate scaled ultrametric tree
335 using the 'chronopl' command in the R package 'ape' [51]. This approach produces a tree
336 whose branches are scaled to evolutionary rates, not a dated tree, and provides a way to
337 understand evolutionary changes over relative "time" of the group being investigated.

338 For the first time, quantifiable rates of diversification were calculated for the
339 Bombycoidea. This is important, because whether a taxonomic group possesses more described
340 species than another related lineage, does not mean they have diversified "more". A larger
341 number of described species could simply be due to the taxonomic effort, a well-known bias in
342 bombycoids, with the Sphingidae and Saturniidae representing the charismatic groups that
343 most bombycoid taxonomists have historically worked. To account for non-random missing
344 speciation events, we quantified the percentage of taxa sampled within each family and
345 incorporated these in the form of branch-specific sampling fractions. Sampling fractions were
346 based on the updated superfamily numbers calculated from Kitching et al. [1]. Informed priors,
347 based on our sampling and phylogeny, were determined using 'setBAMMpriors' in BAMMtools.
348 The MCMC chain was run for 100 million generations, sampling every 1000 generations.

349 Convergence diagnostics was assessed using the R package 'coda' [79]. The first 20% of runs
350 were discarded as burn-in.
351
352 **Results**
353 Our abdomen soaking approach proved relatively consistent with high enough yield of
354 target DNA to proceed with AHE sequencing. Datasets constructed for phylogeny inference
355 contained the following number of loci, sequence length, and model specifications: 1) AA = 579
356 loci, 48,456 amino acid residues, modeled by locus; 2) Pr+Fl = 649 probe loci + one flanking
357 locus (261,780 bp), each probe locus was modeled by site, flanking data was maintained as a
358 single partition and modeled by site; and 3) ASTRAL = 650 loci, each locus modeled by site.
359 Differences in loci number between the AA dataset and the Pr+Fl dataset are the result of loci
360 being removed from the AA phylogenetic inference due to a lack of variation at the amino acid
361 level for those loci (i.e., the probe regions were highly conserved with no variation across taxa).
362 Increasing the number of loci and taxon sampling significantly improved our
363 understanding of Bombycoidea relationships. The inferred relationships are generally
364 consistent across the three phylogenetic inferences that we performed (AA, Pr+Fl, ASTRAL),
365 with all major backbone relationships robustly supported, and all bombycoid families *sensu*
366 Zwick [20] and Zwick et al. [13] were recovered as monophyletic. Due to the methodological
367 approach (i.e., the treatment of different data types) and the more biologically realistic and
368 parsimonious explanation of the topology (see Systematics section in Supplemental Materials),
369 our preferred phylogeny is the tree generated from the "probe + flanking" (Pr+Fl) dataset. All
370 family-level placements of genera *sensu* Kitching et al. [1] were supported, with the exception

371 of *Arotros* Schaus, a genus long considered to be an epiine bombycid [80]. *Arotros* is clearly
372 nested within Apatelodidae in our tree (Fig. 1; Supp. Fig. 1) and is hereby transferred to
373 Apatelodidae (see supplemental text for additional details on this taxonomic change). The
374 family-level rearrangements of Zwick [20] and Zwick et al. [13] were also recovered in our
375 phylogenetic results: the placement of taxa formerly classified in the "Lemoniidae" (e.g.,
376 Lemaire & Minet [22]) are recovered within Brahmaeidae; the Apatelodidae are distinct from
377 the Bombycidae and Phiditiidae; and the broader concept of Endromidae, which includes taxa
378 formerly placed within the "Mirinidae" and Bombycidae (e.g., Lemaire & Minet [22]), are
379 recovered as monophyletic.

380 We find broad congruence with the major groupings designated by Zwick [20], though
381 internal relationships within these groups did not exactly match previously published trees.
382 Historically, the most problematic familial placement in the superfamily has been the
383 Bombycidae *sensu stricto*. Phylogenetic studies that were based on a handful of gene regions
384 (e.g., [13], [18], [20]), placed this family either as sister to the Saturniidae or to the Sphingidae
385 (reviewed in [19]), albeit without strong support for either. Our study clearly places the
386 Bombycidae as the sister lineage to the Saturniidae + Sphingidae (the 'SBS' group – coined by
387 Zwick et al. [13]), as seen in trees from the AA and Pr+Fl datasets, but not in the ASTRAL tree
388 (Supp. Trees 1-3) – an outcome that mirrors traditional Sanger sequencing studies based on few
389 loci, where individual gene trees can lack the phylogenetic signal of supermatrices. The
390 'CAPOPEM' group (Carthaeidae, Anthelidae, Phiditiidae, and Endromidae – coined by Regier et
391 al. [11]) is recovered in all three analyses, although in the ASTRAL inference this group is nested
392 within the clade containing the Sphingidae, Bombycidae and Saturniidae (Supp. Fig. 2).

393 Interfamilial relationships within the CAPOPEM clade were not robustly supported in Regier et
394 al. [11] or Zwick et al. [13], but our AHE-based trees confidently solidify relationships within this
395 group. Zwick et al. [13] recovered the Old World Endromidae as sister to the Australian
396 Anthelidae + (Australian Carthaeidae + Neotropical Phiditiidae), however, our Pr+Fl analysis
397 supports the Anthelidae as sister to Endromidae + (Carthaeidae + Phiditiidae) instead.
398 Effectively, Anthelidae and Endromidae swap places within CAPOPEM when comparing Zwick et
399 al. [13] and our Pr+Fl phylogeny. Conversely, our AA results illustrate yet a different picture, in
400 which Phiditiidae are sister to Anthelidae + (Carthaeidae + Endromidae), mirroring Regier et al.
401 [11]. This discordance is important to note considering the disjunct geographic distribution of
402 these families. Among these families, only Phiditiidae is found in (and endemic to) the New
403 World, with each of the other families being restricted to the Old World. Lastly, while
404 intrafamilial relationships in the SBS and CAPOPEM groups change depending on which of our
405 datasets are used to infer the phylogeny (Fig. 1, Supp. Figs. 1-3, Supp. Trees 1-3), the 'BALE'
406 group (Brahmaeidae, Apatelodidae, and Eupterotidae – coined by Regier et al. [11]) is
407 recovered in all three analyses.

408 When we investigate intrafamilial relationships, the Bombycidae (excluding *Arotros*,
409 mentioned above; see Supplemental text) possesses distinct and monophyletic subfamilies
410 Bombycinae (Old World) and Epiinae (New World; Fig. 1; Supp. Fig. 1). Within the Eupterotidae,
411 the Striphnopteryginae and Janinae are monophyletic, but others are not (Fig. 1; Supp. Fig. 1).
412 Within the Saturniidae, the Pr+Fl and AA topologies are similar to Regier et al. [31] (Fig. 1 &
413 Supp. Fig. 3), where the Oxyteninae are sister to the rest of the family, followed by the
414 Cercophaninae. Arsenurinae is the sister lineage to the Hemileucinae and the Ceratocampinae,

415 while Salassinae is the sister lineage to the Saturniinae. A major difference in subfamily
416 relationships is the placement of Agliinae, which was either sister to the Salassinae +
417 Saturniinae (Pr+Fl) or to the Arsenurinae (AA, also [13] and [31]). The ASTRAL inference is quite
418 different, placing the Ceratocampinae as the sister lineage to a clade containing the Agliinae,
419 Arsenurinae + Hemileucinae, and a Saturniinae + Salassinae clade (Supp. Fig. 3), albeit with low
420 branch support. Within the Sphingidae, our topologies (both Pr+Fl and AA) are largely
421 congruent, but differ slightly from Kawahara et al. [81], Zwick et al. [13], and Kawahara and
422 Barber [33]. Smerinthinae, which now excludes *Langia* (Kitching et al [1]) is monophyletic and
423 the Sphinginae are polyphyletic (Fig. 1; Supp. Fig. 1). This outcome is robustly supported in the
424 Pr+Fl and ASTRAL trees, although internal relationships slightly differ between the two trees.
425 The placement of the Macroglossinae, Langiinae, and Sphinginae are robustly supported in the
426 AA and Pr+FI tree, but low internal node support obfuscates relationships within Smerinthinae.
427 Among the other families, a number of subfamilies and genera, as defined by the most
428 current classification of the Bombycoidea [1], are not monophyletic (Fig. 1; Supp. Fig. 1). Within
429 Eupterotidae, the Eupterotinae and the "Ganisa group" are not monophyletic. The Eupterotinae
430 are rendered paraphyletic due to the placements of Panacelinae and Striphnopteryginae. The
431 Striphnopteryginae genus *Phiala* is paraphyletic with respect to *Lichenopteryx*, and the Ganisa
432 group is polyphyletic due to the traditional inclusion of the genus *Neopreptos* [82]. Within the
433 Anthelidae, *Anthela* is paraphyletic due to the placement of *Nataxa* and *Pterolocera*, a finding
434 congruent with Zwick [20]. Within Sphingidae, Sphinginae is polyphyletic due to the placement
435 of *Pentateucha* as sister to the Langiinae, and the Smerinthinae genus *Polyptychus* is
436 paraphyletic.

A number of factors can lead to the appearance of a taxonomic group being more "diverse" than other sister lineages. For example, when simply looking at numbers of described species, taxonomic bias in interest and effort could substantially affect our understanding. This is why it is essential to formally test diversification rates. While there are more described species of Saturniidae and Sphingidae than the rest of the Bombycoidea, that doesn't mean those two families are necessarily more evolutionarily diverse. For the first time, actual quantifiable rates of diversification were calculated for the Bombycoidea, by evaluating how the interplay between speciation and extinction has changed over relative time. At the family level, a major diversification rate shift occurs along the lineage leading to the Saturniidae and Sphingidae (Fig. 1). The effective sample size of the log-likelihood was 8475.169 and the effective sample size of the number of shift events present in each sample was 18141.4. The 95% credible set of rate shift configurations sampled with BAMM can be seen in Supp. Fig. 4. Unfortunately, due to the poor Bombycoidea fossil record and the limited computational approaches currently available to adequately estimate extinction rates, we are unable to discern whether the differences that exist are due to increases or decreases in speciation, although species diversity numbers, based on the classification of Kitching et al. [1] agree with our diversification results. The number of described species for each of the ten bombycoid families, according to Kitching et al. [1] is: Anthelidae (94 spp.); Apatelotidae (182 spp.); Bombycidae (202 spp.); Brahmaeidae (68 spp.); Carthaeidae (1 sp.); Endromidae (70 spp.); Eupterotidae (396 spp.); Phiditiidae (23 spp.); Saturniidae (3,454 spp.); Sphingidae (1,602 spp.).

Discussion

459 Within the past decade, a few studies using molecular sequence data have attempted to
460 resolve the phylogenetic relationships of the Bombycoidea ([11], [13], [20]). None have
461 corroborated the earlier morphology-based hypotheses ([15], [22]). In order to establish a
462 backbone for future evolutionary, comparative, and taxonomic studies, we sampled exemplars
463 from all major lineages in the superfamily and used Anchored Hybrid Enrichment (AHE)
464 phylogenomics to provide a robust phylogeny of the superfamily based on the largest
465 taxonomic and molecular sampling to date.

466 To achieve dense taxonomic coverage, we included many samples from pinned
467 specimens in natural history collections. We modified previous DNA extraction protocols to
468 increase DNA yield for high-throughput sequencing and modified the AHE probe set developed
469 by Breinholt et al. [16] to more efficiently recover phylogenetically-informative loci within the
470 Bombycoidea. To allow more flexibility in the use of the data (e.g., integration of these samples
471 with "legacy" datasets and CO1 for species identification), we added a selection of traditional
472 Sanger-sequenced loci. To assist future usage of the BOM1 and other AHE probe sets, we
473 expanded the bioinformatic protocol of Breinholt et al. [16] to include step-by-step instructions
474 for running the bioinformatics pipeline.

475 Our phylogeny largely reinforces the results of earlier molecular works, although
476 phylogenomics finally brings into focus those long problematic relationships, while also
477 identifying important topological where some subfamilies and genera are not recovered as
478 monophyletic. When comparing with trees published in previous works, topological
479 discordances are likely the product of increased locus sampling, which provided significantly
480 more phylogenetic information, as well as morphological homoplasy or convergence that likely

481 obscured the true placement of certain taxonomic groups. For example, our results provide a
482 well-supported placement of the historically troublesome family Bombycidae *sensu stricto*, as
483 the sister lineage to Saturniidae + Sphingidae. An earlier phylogenomic study [19] had provided
484 some evidence in support of this relationship, but its taxon sampling was very limited.
485 Our study highlights how morphological convergence in Bombycoidea has confused our
486 understanding of their evolution. For example, the bombycid genus *Rotunda*, endemic to the
487 Old World, and the apatelodid genus *Arotros*, endemic to the New World, are clearly unrelated
488 (Fig. 1; Supp. Fig. 1). However, they are both astonishingly similar in body size, wing shape, and
489 phenotypic appearance (see Supplemental Information: Systematics). Such results imply that
490 there may be adaptive advantages to evolving particular wing shape and size, as found in other
491 bombycoid lineages such as in the Saturniidae [29]. Furthermore, within the CAPOPEM clade,
492 we see a biogeographical distribution that, based on the AA and ASTRAL topologies, appears to
493 be a more biologically realistic and parsimonious explanation (than the Pr+Fl topology) of the
494 evolutionary history due to the placement of the Neotropical Phiditiidae as sister to the
495 remaining and Old World CAPOPEM families (Supp. Fig. 2). In the future, increased taxon
496 sampling will likely help bring better resolution to both the backbone and internal bombycoid
497 relationships. These types of findings, wherein different dataset types answer different
498 questions, highlight the importance of evaluating phylogenetic data in different ways because
499 phylogenetic signal could be hiding in phylogenomic datasets.
500 From an evolutionary viewpoint, one of the most interesting results came from the first
501 attempt to quantify the diversification rates across the Bombycoidea, in particular, the dramatic
502 shift in diversification rates leading to the Sphingidae and Saturniidae lineages. While maybe

503 not “surprising”, because of the number of species described, this had never been quantified
504 before in this group. We know that simple numbers of described species does not mean
505 necessarily that they have diversified “more” than other closely related lineages. A number of
506 factors could cause one taxonomic group to appear as if there are more species than another
507 related lineage. Perhaps the most important of these is simply due to the taxonomic effort that
508 has historically been applied to a group, a bias that can certainly be found in the bombycoids,
509 with Sphingidae and Saturniidae representing highly charismatic groups where most bombycoid
510 taxonomists have worked.

511 These findings are fascinating because the Sphingidae and Saturniidae have contrasting
512 life-history strategies, such as larvae feeding on different kinds of food plants, adults having the
513 ability nectar source [35], and anti-bat strategies ([29], [32], [33], [34]) which are reflected in
514 their morphology (e.g., body sizes and wing shapes) and behavior, including flight speed and
515 maneuverability [35]. In an “arms race”, as has been shown between moths and bats, bat
516 predation selectively removes unfit lineages from the environment, thus increasing the speed
517 of evolution of these surviving lineages. As new traits evolve that can be used to effectively
518 evade the predator, subsequent release from the predatory pressure provides the
519 opportunities to diversify ecologically and behaviorally. Although we did not conduct a
520 divergence time estimation analysis in the present study, the origin of the SBS group has been
521 postulated to be approximately 50 mya [83], with Sphingidae originating soon thereafter [33].
522 Insectivorous bats are thought to have originated roughly 60 mya, and the diversification of
523 Sphingidae and Saturniidae around that time suggests that the incredible taxonomic diversity

524 within these two families could be in part due to bat-related selection pressures resulting in
525 diverse anti-bat traits.
526 The hawkmoth and silk moth evolutionary story is certainly more complex than simply
527 reflecting their interactions with bats. Although we postulate that the differences in
528 diversification rates are correlated with bat predation, it is possible that these rate shifts are
529 due to other factors, such as ecological specialization or shifts in host plant usage, both as
530 larvae and adults. Amassing and collating behavioral and ecological datasets for the tips of the
531 Bombycoidea Tree of Life for macroevolutionary comparative investigations is essential to
532 furthering our understanding of this diverse, global superfamily, as well as understanding how
533 bats, ecological traits, and/or biogeographical history may or may not have shaped their
534 diversity. At this time, large trait datasets do not exist for these groups, but are currently being
535 worked on to be included in future trait-dependent diversification analyses with much more
536 complete sampling, at the genus and species level across these families, to truly explore the
537 drivers of Bombycoidea diversity. This research establishes some hypotheses to be further
538 tested when more complete sampling of the Bombycoidea has been completed, and robust
539 trait datasets have been collected.
540
541 **Conclusions**
542 Our study finally brings into focus long problematic bombycoid relationships and
543 establishes a backbone for future evolutionary, comparative, and taxonomic studies. Our
544 modified DNA extraction protocol allows Lepidoptera specimens to be readily sequenced from
545 pinned natural history collections, and highlights the flexibility of AHE to generate genomic data

546 from a wide range of museum specimens, both age and preservation method. By allowing
547 researchers to tap into the wealth of biological data residing in natural history collections
548 around the globe, these types of methodologies (e.g., DNA from museum specimens and
549 targeted sequencing capture) will provide the opportunities for us to continually add to our
550 understanding of Lepidoptera and Bombycoidea evolution, as well as refinine our
551 understanding of relationships across the Tree of Life.
552
553 **Declarations**
554 **Ethics approval and consent to participate**
555 Not applicable
556
557 **Consent for publication**
558 Not applicable
559
560 **Availability of data and material**
561 The datasets supporting the results of this article are available in Dryad Data Repository
562 (will update upon acceptance with the unique persistent identifier and hyperlink to datasets in
563 http:// format). Supplemental Information includes four figures, one table, plus the full
564 museum specimen DNA extraction protocol, systematic information on Bombycoidea families,
565 bioinformatics pipeline scripts and instructions, alignment FASTA files (nucleotides, amino
566 acids, supermatrices, and individual loci), partition files, tree files, and other essential data files

567 that were inputted or outputted from the phylogenetic inference and analysis. Previous
568 Breinholt et al., [16] bioinformatics scripts are available in Dryad [45].
569
570 **Competing interests**
571 The authors declare that they have no competing interests.
572
573 **Authors' Information**
574 ORCID ID - Hamilton: 0000-0001-7263-0755; St Laurent: 0000-0001-6439-5249; Dexter: 0000-
575 0002-5136-6083; Kitching: 0000-0003-4738-5967; Breinholt: 0000-0002-3867-2430; Zwick:
576 0000-0002-7532-1752; Timmermans: 0000-0002-5024-9053; Barber: 0000-0003-3084-2973;
577 Kawahara: 0000-0002-3724-4610
578
579 **Funding**
580 This work was partly supported by NSF grants [IOS-1121807, JRB], [IOS-1121739, AYK], [DBI-
581 1349345, AYK], [DBI-1601369, AYK], [DEB-1557007, AYK], and [PRFB-1612862, CAH]; a NERC
582 grant [NE/P003915/1, IJK]; National Geographic Society grants [CRE 9944-16, JRB]; and
583 postdoctoral support from the Florida Museum of Natural History (to CAH).
584
585 **Authors' Contributions**
586 CAH, AYK, JWB, IJK, MT, AZ designed the study. CAH and JWB analyzed data. CAH, AYK, RAS
587 wrote the paper. CAH, AYK, JRB, KD, IJK, RAS, AZ collected, identified, and processed specimens.
588 All authors discussed results and contributed to the writing of the manuscript.

Acknowledgements

We thank Seth Bybee, Deborah Glass, Yash Sondhi, Jamie Theobald, and Emmanuel Toussaint. We are indebted to Samantha Epstein, John Heppner, Geena Hill, Nicholas Homziak, Jacqueline Miller, Charles Mitter, Kim Mitter, Lary Reeves, Andrei Sourakov, and Andrew Warren, who provided specimens for this project, contributed help in identifying or organizing specimens, or assisted the project in other ways (e.g., lab work or intellectually). The authors would also like to thank the peer-review process for improving the manuscript. The authors acknowledge University of Florida Research Computing for providing computational resources and support that contributed to the research results reported in this publication. URL: http://researchcomputing.ufl.edu

References

1. Kitching I, Rougerie R, Zwick A, St Laurent R, Ballesteros Mejia L, Kawahara A, et al. A global checklist of the Bombycoidea (Insecta: Lepidoptera). BDJ. 2018;6:e22236–13.

2. Roe AD, Weller SJ, Baixeras J, Brown J, Cummings MP, Davis DR, et al. Evolutionary framework for Lepidoptera model systems. In: Goldsmith M, Marec F, editors. Molecular Biology and Genetics of the Lepidoptera. 2010. pp. 1–24.

3. Gregory D. Hawkmoth pollination in the genus *Oenothera*. Aliso A Journal of Systematic and Evolutionary Botany. 1963;5:357–84.

609 4. Haber WA, Frankie GW. A Tropical Hawkmoth Community- Costa Rican Dry Forest
610 Sphingidae. Biotropica. 1989;21:155–72.

611 5. Nilsson LA, Jonsson L, Rason L, Randrianjohany E. Monophily and pollination mechanisms in
612 *Angraecum arachnites* Schltr. (Orchidaceae) in a guild of long-tongued hawk-moths
613 (Sphingidae) in Madagascar. Biological Journal of the Linnean Society. 1985;26:1–19.

614 6. Nilsson LA, Rabakonandrianina E, Pettersson B. Exact tracking of pollen transfer and mating
615 in plants. Nature. 1992;360:666–8.

616 7. Wasserthal LT. The pollinators of the Malagasy star orchids *Angraecum sesquipedale*, *A.*
617 *sororium* and *A. compactum* and the evolution of extremely long spurs by pollinator shift.
618 Botanica Acta. 1997;110:343–59.

619 8. Nilsson LA. Deep flowers for long tongues. Trends In Ecology & Evolution. 1998;13:259–60.

620 9. Raguso RA, Willis MA. Synergy between visual and olfactory cues in nectar feeding by naïve
621 hawkmoths, *Manduca sexta*. Animal Behaviour. 2002;64:685–95.

622 10. Beck J, Kitching IJ, Linsenmair KE. Effects of habitat disturbance can be subtle yet significant:
623 biodiversity of hawkmoth-assemblages (Lepidoptera: Sphingidae) in Southeast-Asia.
624 Biodiversity and Conservation. 2006;15:465–86.

625 11. Regier JC, Cook CP, Mitter C, Hussey A. A phylogenetic study of the “bombycoid complex”
626 (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of
627 macrolepidopteran phylogeny. Systematic Entomology. 2008;33:175–89.

628 12. Meusemann K, Reumont von BM, Simon S, Roeding F, Strauss S, Kück P, et al. A
629 phylogenomic approach to resolve the Arthropod Tree of Life. Mol Biol Evol. 2010;27:2451–64.

630 13. Zwick A, Regier JC, Mitter C, Cummings MP. Increased gene sampling yields robust support
631 for higher-level clades within Bombycoidea (Lepidoptera). Systematic Entomology. 2011;36:31–
632 43.

633 14. Kim MJ, Kang AR, Jeong HC, Kim K-G, Kim I. Reconstructing intraordinal relationships in
634 Lepidoptera using mitochondrial genome data with the description of two newly sequenced
635 lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). Mol
636 Phylogenet Evol. 2011;61:436–45.

637 15. Minet J. The Bombycoidea: Phylogeny and higher classification (Lepidoptera, Glossata).
638 Entomologica Scandinavica. 1994;25:63–88.

639 16. Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. Resolving relationships
640 among the megadiverse butterflies and moths with a novel pipeline for anchored
641 phylogenomics. Syst Biol. 2018;67:78–93.

642 17. Timmermans MJTN, Lees DC, Simonsen TJ. Towards a mitogenomic phylogeny of
643 Lepidoptera. Mol Phylogenet Evol. 2014;79:169–78.

644 18. Wang X, Chen Z-M, Gu X-S, Wang M, Huang G-H, Zwick A. Phylogenetic relationships among
645 Bombycidae s.l. (Lepidoptera) based on analyses of complete mitochondrial genomes.
646 Systematic Entomology. 2018;20:537–9.

647 19. Breinholt JW, Kawahara AY. Phylotranscriptomics: saturated third codon positions radically
648 influence the estimation of trees based on next-gen data. Genome Biol Evol. 2013;5:2082–92.

649 20. Zwick A. Molecular phylogeny of Anthelidae and other bombycoid taxa (Lepidoptera:
650 Bombycoidea). Systematic Entomology. 2008;33:190–209.

651 21. Minet J. Tentative reconstruction of the ditrysian phylogeny (Lepidoptera, Glossata).
652 Entomologica Scandinavica. 1991;22:69–95.

653 22. Lemaire C, Minet J. 18. The Bombycoidea and their relatives. In: Kristensen NP, editor.
654 Lepidoptera, Moths and Butterflies. vol. 1. Evolution, Systematics, and Biogeography. Handbook
655 of Zoology. Walter de Gruyter, Berlin; 1998. pp. 322–53.

656 23. Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, et al. A large-scale,
657 higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and
658 butterflies). PLoS ONE. 2013;8:e58568–23.

659 24. Regier JC, Zwick A, Cummings MP, Kawahara AY, Cho S, Weller S, et al. Toward
660 reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an
661 initial molecular study. BMC Evolutionary Biology. 2009;9:280–21.

662 25. Espeland M, Breinholt J, Willmott KR, Warren AD, Vila R, Toussaint EFA, et al. A
663 comprehensive and dated phylogenomic analysis of butterflies. CURBIO. 2018;28:770–5.

664 26. Johns CA, Toussaint EFA, Breinholt JW, Kawahara AY. Origin and macroevolution of micro-
665 moths on sunken Hawaiian Islands. Proc. Biol. Sci. 2018;285:20181047–10.

666 27. St Laurent RA, Hamilton CA, Kawahara AY. Museum specimens provide phylogenomic data
667 to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae).
668 Systematic Entomology. 2018;43:729–61.

669 28. Toussaint EFA, Breinholt JW, Earl C, Warren AD, Brower AVZ, Yago M, et al. Anchored
670 phylogenomics illuminates the skipper butterfly tree of life. BMC Evolutionary Biology.
671 2018;18:1–11.

672 29. Rubin JJ, Hamilton CA, McClure CJW, Chadwell BA, Kawahara AY, Barber JR. The evolution of
673 anti-bat sensory illusions in moths. Science Advances. 2018;4.

674 30. Lemmon AR, Emme SA, Lemmon EM. Anchored Hybrid Enrichment for massively high-
675 throughput phylogenomics. Syst Biol. 2012;61:727–44.

676 31. Regier JC, Grant MC, Mitter C, Cook CP, Peigler RS, Rougerie R. Phylogenetic relationships of
677 wild silkmoths (Lepidoptera: Saturniidae) inferred from four protein-coding nuclear genes.
678 Systematic Entomology. 2008;33:219–28.

679 32. Barber JR, Kawahara AY. Hawkmoths produce anti-bat ultrasound. Biology Lett.
680 2013;9:20130161–1.

681 33. Kawahara AY, Barber JR. Tempo and mode of antibat ultrasound production and sonar
682 jamming in the diverse hawkmoth radiation. Proceedings of the National Academy of Sciences.
683 2015;112:6407–12.

684 34. Barber JR, Leavell BC, Keener AL, Breinholt JW, Chadwell BA, McClure CJW, et al. Moth tails
685 divert bat attack: evolution of acoustic deflection. Proceedings of the National Academy of
686 Sciences. 2015;112:2812–6.

687 35. Janzen D. Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. Oxford
688 surveys in evolutionary biology. 1984;1:140.

689 36. Tammaru T, Haukioja E. Capital breeders and income breeders among Lepidoptera:
690 consequences to population dynamics. Oikos. 1996;77:561–4.

691 37. Kawahara AY, Breinholt JW. Phylogenomics provides strong evidence for relationships of
692 butterflies and moths. P Roy Soc B-Biol Sci. 2014;281:20140970–0.

693 38. Jones KE, Bininda-Emonds ORP, Gittleman JL. Bats, clocks, and rocks: diversification patterns
694 in Chiroptera. Evolution. 2005;59:2243–55.

695 39. Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. A molecular phylogeny
696 for bats illuminates biogeography and the fossil record. Science. 2005;307:580–4.

697 40. Miller-Butterworth CM, Murphy WJ, O'Brien SJ, Jacobs DS, Springer MS, Teeling EC. A family
698 matter: conclusive resolution of the taxonomic position of the long-fingered bats, *Miniopterus*.
699 Mol Biol Evol. 2007;24:1553–61.

700 41. Shi JJ, Rabosky DL. Speciation dynamics during the global radiation of extant bats. Evolution.
701 2015;69:1528–45.

702 42. Lei M, Dong D. Phylogenomic analyses of bat subordinal relationships based on
703 transcriptome data. Scientific Reports. 2016;6:27726.

704 43. Barber JR, Conner WE. Acoustic mimicry in a predator-prey interaction. Proceedings of the
705 National Academy of Sciences. 2007;104:9331–4.

706 44. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Bin Li, et al. A draft sequence for the genome of the
707 domesticated silkworm (*Bombyx mori*). Science. 2004;306:1937–40.

708 45. Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. Data from: Resolving
709 relationships among the megadiverse butterflies and moths with a novel pipeline for Anchored
710 Phylogenomics. Dryad Digital Repository; 2017.

711 46. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
712 phylogenies. Bioinformatics. 2014.

713 47. Mayrose I, Graur D, Ben-Tal N, Pupko T. Comparison of site-specific rate-inference methods
714 for protein sequences: empirical Bayesian methods are superior. Mol Biol Evol. 2004;21:1781–
715 91.

716 48. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Springer, New
717 York, NY; 2005. pp. 125–81.

718 49. López-Giráldez F, Townsend JP. PhyDesign: an online application for profiling phylogenetic
719 informativeness. BMC Evolutionary Biology. 2011;11:152.

720 50. Heibl C. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic
721 software packages. 2013.

722 51. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language.
723 Bioinformatics. 2004;20:289–90.

724 52. Cho S, Epstein SW, Mitter K, Hamilton CA, Plotkin D, Mitter C, et al. Preserving and
725 vouchering butterflies and moths for large-scale museum-based molecular research. PeerJ.
726 2016;4:e2160.

727 53. Blaimer BB, Lloyd MW, Guillory WX, Brady SG. Sequence capture and phylogenetic utility of
728 genomic Ultraconserved Elements obtained from pinned insect specimens. PLoS ONE.
729 2016;11:e0161531–20.

730 54. Mikheyev AS, Zwick A, Magrath MJL, Grau ML, Qiu L, Su YN, et al. Museum genomics
731 confirms that the Lord Howe Island stick insect survived extinction. CURBIO. 2017;27:3157–
732 3161.e4.

733 55. Ruane S, Austin CC. Phylogenomics using formalin-fixed and 100+ year-old intractable
734 natural history specimens. Molecular Ecology Resources. 2017;17:1003–8.

735 56. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets.
736 Bioinformatics. 2014.

737 57. Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment
738 preparations concerning phylogenetic studies. Frontiers in Zoology. 2014;11.

739 58. Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ. IQ-TREE: a fast and effective stochastic
740 algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

741 59. Minh BQ, Nguyen MAT, Haeseler von A. Ultrafast approximation for phylogenetic bootstrap.
742 Mol Biol Evol. 2013;30:1188–95.

743 60. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and
744 methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML
745 3.0. Syst Biol. 2010;59:307–21.

746 61. Mendes FK, Hahn MW. Why concatenation fails near the anomaly zone. Syst Biol.
747 2017;67:158–69.

748 62. Maddison W. Gene trees in species trees. Syst Biol. 1997;46:523–36.

749 63. Slowinski J, Page R. How should species phylogenies be inferred from sequence data? Syst
750 Biol. 1999;48:814–25.

751 64. Edwards SV. Is a new and general theory of molecular systematics emerging? Evolution.
752 2009;63:1–19.

753 65. Zhang C, Sayyari E, Mirarab S. ASTRAL-III: increased scalability and impacts of contracting
754 low support branches. RECOMB International Workshop on Comparative Genomics. 2017. pp.
755 53–75.

756 66. Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many
757 hundreds of taxa and thousands of genes. Bioinformatics. 2015;31:i44–i52.

758 67. Sayyari E, Mirarab S. Anchoring quartet-based phylogenetic distances and applications to
759 species tree reconstruction. BMC Genomics. BioMed Central; 2016;17:101–13.

760 68. Aberer AJ, Stamatakis A. A simple and accurate method for rogue taxon identification. IEEE
761 International Conference on Bioinformatics and Biomedicine BIBM. 2011. pp. 118–22.

762 69. Aberer AJ, Krompass D, Stamatakis A. Pruning rogue taxa improves phylogenetic accuracy:
763 an efficient algorithm and webservice. Syst Biol. 2012;62:162–6.

764 70. de Vienne DM, Ollier S, Aguileta G. Phylo-MCOA: a fast and efficient method to detect
765 outlier genes and species in phylogenomics using multiple co-inertia analysis. Mol Biol Evol.
766 2012;29:1587–98.

767 71. Dial KP, Marzluff JM. Nonrandom Diversification Within Taxonomic Assemblages.
768 Systematic Zoology. 1989;38:26–37.

769 72. Doyle JA, Donoghue MJ. Phylogenies and Angiosperm Diversification. Paleobiology.
770 1993;19:141–67.

771 73. Sanderson MJ, Donoghue MJ. Shifts in Diversification Rate with the Origin of Angiosperms.
772 Science. 1994;264:1590–3.

773 74. Sanderson MJ, Donoghue MJ. Reconstructing shifts in diversification rates on phylogenetic
774 trees. Trends In Ecology & Evolution. 1996;11:15–20.

775 75. Rabosky DL. Automatic detection of key innovations, rate shifts, and diversity-dependence
776 on phylogenetic trees. PLoS ONE. 2014;9:e89543.

76. Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, et al. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution. 2014;5:701–7.

77. Sohn JC, Labandeira CC, Davis DR, Mitter C. An annotated catalog of fossil and subfossil Lepidoptera (Insecta: Holometabola) of the world. Zootaxa. 2012,1-132.

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

79. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6:7–11.

80. Schaus W. Familie Bombycidae (Epiinae). In: Seitz A, editor. Die Gross-Schmetterlinge der Erde. Kernen, Stuttgart; 1929. pp. 675–711.

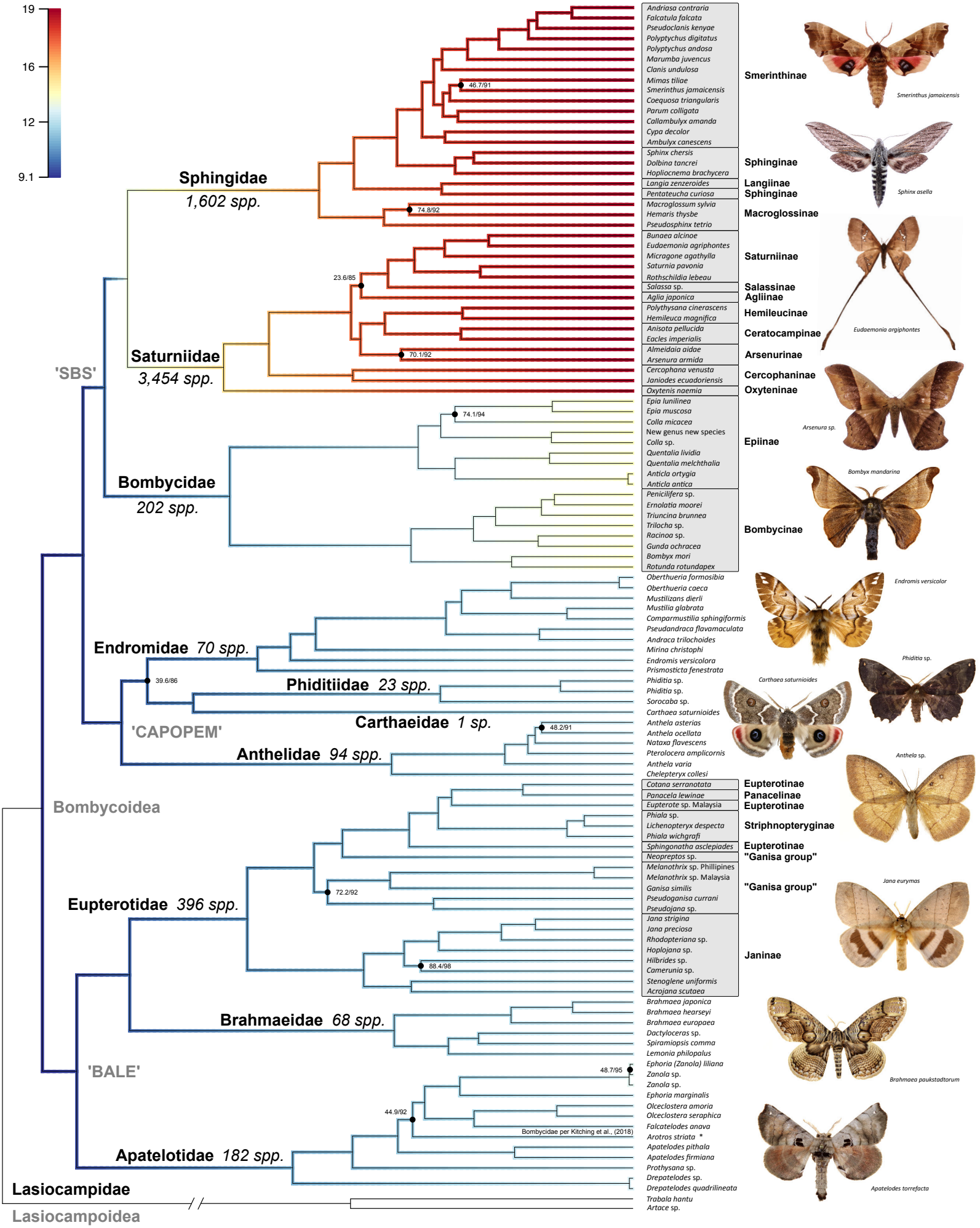
81. Kawahara AY, Mignault AA, Regier JC, Kitching IJ, Mitter C. Phylogeny and Biogeography of Hawkmoths (Lepidoptera: Sphingidae): Evidence from Five Nuclear Genes. Knapp M, editor. PLoS ONE. 2009;4:e5719.

82. Nässig WA, Oberprieler RG. An annotated catalogue of the genera of Eupterotidae (Insecta, Lepidoptera, Bombycoidea). Senckenbergiana biologica. 2008;88:53–80.

83. Misof B, Liu S, Meusemann K, Peters RS, Donath A. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346:763–7.

Figures

796 Figure 1. Maximum likelihood tree of Bombycoidea, based on 650 AHE loci. All nodes are
797 supported by ≥95% UFBoot and ≥80% SH-aLRT values unless otherwise noted. Branch color
798 indicates the estimated diversification rate, with warmer colors representing lineages with
799 higher rates. Major taxonomic groups such as families and subfamilies are labeled. Photographs
800 represent species in lineages sampled in the phylogeny. Species diversity, based on Kitching et
801 al. [1], are noted next to families.
802


19
16
12
9.1
Sphingidae
1,602 spp.
Saturniidae
3,454 spp.
Bombycidae
202 spp.
Endromidae 70 spp.
Phiditiidae 23 spp.
Carthaeidae 1 sp.
Anthelidae 94 spp.
Eupterotidae 396 spp.
Brahmaeidae 68 spp.
Apatelotidae 182 spp.
Lasiocampidae
'SBS'
'CAPOPEM'
'BALE'
Bombycoidea
Lasiocampoidea
46.7/91
74.8/92
23.6/85
70.1/92
74.1/94
39.6/86
48.2/91
72.2/92
88.4/98
48.7/95
44.9/92
Bombycidae per Kitching et al., (2018)
Andriasa contraria
Falcatula falcata
Pseudoclanis kenyae
Polyptychus digitatus
Polyptychus andosa
Marumba juvencus
Clanis undulosa
Mimas tiliae
Smerinthus jamaicensis
Coequosa triangularis
Parum colligata
Callambulyx amanda
Cypa decolor
Ambulyx canescens
Sphinx chersis
Dolbina tancrei
Hopliocnema brachycera
Langia zenzeroides
Pentateucha curiosa
Macroglossum sylvia
Hemaris thysbe
Pseudosphinx tetrio
Bunaea alcinoe
Eudaemonia agriphontes
Micragone agathylla
Saturnia pavonia
Rothschildia lebeau
Salassa sp.
Aglia japonica
Polythysana cinerascens
Hemileuca magnifica
Anisota pellucida
Eacles imperialis
Almeidaia aidae
Arsenura armida
Cercophana venusta
Janiodes ecuadoriensis
Oxytenis naemia
Epia lunilinea
Epia muscosa
Colla micacea
New genus new species
Colla sp.
Quentalia lividia
Quentalia melchthalia
Anticla ortygia
Anticla antica
Penicilifera sp.
Ernolatia moorei
Triuncina brunnea
Trilocha sp.
Racinoa sp.
Gunda ochracea
Bombyx mori
Rotunda rotundapex
Oberthueria formosibia
Oberthueria caeca
Mustilizans dierli
Mustilia glabrata
Comparmustilia sphingiformis
Pseudandraca flavamaculata
Andraca trilochoides
Mirina christophi
Endromis versicolora
Prismosticta fenestrata
Phiditia sp.
Phiditia sp.
Sorocaba sp.
Carthaea saturnioides
Anthela asterias
Anthela ocellata
Nataxa flavescens
Pterolocera amplicornis
Anthela varia
Chelepteryx collesi
Cotana serranotata
Panacela lewinae
Eupterote sp. Malaysia
Phiala sp.
Lichenopteryx despecta
Phiala wichgrafi
Sphingonatha asclepiades
Neopreptos sp.
Melanothrix sp. Phillipines
Melanothrix sp. Malaysia
Ganisa similis
Pseudoganisa currani
Pseudojana sp.
Jana strigina
Jana preciosa
Rhodopteriana sp.
Hoplojana sp.
Hilbrides sp.
Camerunia sp.
Stenoglene uniformis
Acrojana scutaea
Brahmaea japonica
Brahmaea hearseyi
Brahmaea europaea
Dactyloceras sp.
Spiramiopsis comma
Lemonia philopalus
Ephoria (Zanola) liliana
Zanola sp.
Zanola sp.
Ephoria marginalis
Olceclostera amoria
Olceclostera seraphica
Falcatelodes anava
Arotros striata *
Apatelodes pithala
Apatelodes firmiana
Prothysana sp.
Drepatelodes sp.
Drepatelodes quadrilineata
Trabala hantu
Artace sp.
Smerinthinae
Sphinginae
Langiinae
Sphinginae
Macroglossinae
Saturniinae
Salassinae
Agliinae
Hemileucinae
Ceratocampinae
Arsenurinae
Cercophaninae
Oxyteninae
Epiinae
Bombycinae
Eupterotinae
Panacelinae
Eupterotinae
Striphnopteryginae
Eupterotinae
"Ganisa group"
"Ganisa group"
Janinae
Smerinthus jamaicensis
Sphinx asella
Eudaemonia argiphontes
Arsenura sp.
Bombyx mandarina
Endromis versicolor
Phiditia sp.
Carthaea saturnioides
Anthela sp.
Jana eurymas
Brahmaea paukstadtorum
Apatelodes torrefacta