

Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment

Elsa Call,^{1,5,✉} Christoph Mayer,^{2,✉} Victoria Twort,^{1,3,✉} Lars Dietz,^{2,✉} Niklas Wahlberg,^{1,✉} and Marianne Espeland^{4,✉}

¹Department of Biology, Lund University, 22362 Lund, Sweden, ²Statistical Phylogenetics and Phylogenomics, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany, ³University of Helsinki, Finnish Natural History Museum, Luomus, Helsinki, Finland, ⁴Arthropoda Department, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany, and ⁵Corresponding author, e-mail: elsa.call.fr@gmail.com

Subject Editor: Marko Mutanen

Received 25 May 2020; Editorial decision 27 October 2020

Abstract

Billions of specimens can be found in natural history museum collections around the world, holding potential molecular secrets to be unveiled. Among them are intriguing specimens of rare families of moths that, while represented in morphology-based works, are only beginning to be included in genomic studies: Pseudobistonidae, Sematuridae, and Epicopeiidae. These three families are part of the superfamily Geometroidea, which has recently been defined based on molecular data. Here we chose to focus on these three moth families to explore the suitability of a genome reduction method, target enrichment (TE), on museum specimens. Through this method, we investigated the phylogenetic relationships of these families of Lepidoptera, in particular the family Epicopeiidae. We successfully sequenced 25 samples, collected between 1892 and 2001. We use 378 nuclear genes to reconstruct a phylogenetic hypothesis from the maximum likelihood analysis of a total of 36 different species, including 19 available transcriptomes. The hypothesis that Sematuridae is the sister group of Epicopeiidae + Pseudobistonidae had strong support. This study thus adds to the growing body of work, demonstrating that museum specimens can successfully contribute to molecular phylogenetic studies.

Key words: Museomics, museum sample, target enrichment, phylogenomics, Lepidoptera

Over 3 billion specimens are estimated to be found in natural history museum collections around the world, representing one of the most important biobanks in the world (Duckworth et al. 1993, Suarez and Tsutsui 2004, Chapman 2005). Until recently, this vast amount of biological resource was mainly used for morphological studies because the DNA from these specimens was thought to be too degraded to be used for molecular studies (Shapiro and Hofreiter 2012). Due to this, DNA work has, for a long time, been limited to species for which freshly collected samples could be obtained, while molecular work from collections was restricted to Sanger sequencing of short fragments of DNA (Hajibabaei et al. 2006, Lozier and Cameron 2009, Strutzenberger et al. 2012, Hebert et al. 2013, Cameron et al. 2016). Moreover, the methods were often destructive for the specimens (Hajibabaei et al. 2006, Strutzenberger et al. 2012, Hebert et al. 2013). Recently, high-throughput sequencing technologies have made the DNA in museum specimens more accessible, either through whole-genome sequencing (Cong et al. 2017, Sproul and Maddison 2017, Allio et al. 2019, Li et al. 2019, Zhang et al. 2019) or through genome reduction methods (Suchan et al. 2016, Breinholt et al. 2018, Toussaint et al. 2018). These advanced sequencing approaches have opened up a new field

with great potential for studying the evolutionary history of taxa that are difficult to collect: museomics.

The family Epicopeiidae is a small Asian family of Lepidoptera represented by 25 species (Minet 2002, Wei and Yen 2017, Zhang et al. 2020). Many of them are large diurnal species mimicking butterflies in the families Papilionidae and Pieridae. The history of the family has been dynamic. Epicopeiidae had originally been described to harbor only one genus *Epicopeia* Westwood, 1841 (Laithwaite and Whalley 1975). The pierid-like moths *Nossa* Kirby, 1892, were previously assigned to the family Epipleminidae (now considered a subfamily of Uraniidae), but then were rightly placed in Epicopeiidae by Fletcher (1979) and later confirmed by Minet (1983, 1986). In latter studies, Minet (1983, 1986) added five genera to Epicopeiidae: *Amana* Walker, 1855; *Chatamla* Moore, 1881; *Parabraxas* Leech, 1897; *Psychostrophia* Butler, 1877; and *Schistomitra* Butler, 1881. In 2002, Minet described two new genera, *Deuweia* and *Burmeia*. Finally, in 2017, the number of genera increased to 10 with the description of *Mimapora* by Wei and Yen. The family was thought to be related to Drepanidae and was placed in the superfamily Drepanoidea (Minet 2002), until recent molecular data suggested that they are in fact related to the superfamily

Geometroidea (Regier et al. 2009, Bazinet et al. 2013, Rajaei et al. 2015). The sister group of Epicopeiidae has been suggested to be the recently described Pseudobistonidae (Rajaei et al. 2015, Wang et al. 2019). Minet (2002) studied the relationships of genera within Epicopeiidae based on 34 morphological characters obtained from the head, thorax, pregenital abdomen, and male genitalia. He found that *Deuveia* was sister to the rest of Epicopeiidae and that the relationships of the other genera were relatively clear (Fig. 1, left side). However, the position of *Amana* was not stable; it was either sister to *Chatamla* + *Parabraxas* or sister to a clade containing *Chatamla*, *Parabraxas*, *Schistomitra*, *Nossa*, and *Epicopeia*.

The first attempt to infer the phylogeny of the family based on genetic markers was done by Wei and Yen (2017). They used sequence data for three gene regions (*COI*, *EF-1 α* , and *28S*) and 14 species. Their study was mainly focused on describing a new genus, *Mimaporina*, but they sampled widely throughout the family. The results of their analyses are highly incongruent with those of Minet (2002), but showed poor or no support on many branches. Wei and Yen (2017) showed that *Epicopeia* and *Nossa* likely are paraphyletic, and they were not able to resolve the relationships of the new genus *Mimaporina* with any confidence (Fig. 1).

Recently, Zhang et al. (2020) used PCR-generated baits to infer a multilocus phylogenetic hypothesis for Epicopeiidae based on 18 species and 94 loci. Their results were highly congruent with Minet's (2002) results based on morphology and also found that *Epicopeia* and *Nossa* both were paraphyletic with regard to each other. In addition to using fresh specimens, Zhang et al. (2020) used older specimens with some degree of success, although they were able to recover a significantly smaller number of loci from the older specimens.

Epicopeiidae species are generally rare and difficult to collect, as they are mainly distributed in areas that are not easy to access, nevertheless they can be found in natural history museums. Here we investigate the use of target enrichment (TE) methods to study the phylogenetic relationships of this family of Lepidoptera based

only on museum specimens. Genome reduction methods, such as TE, aim to sequence only specific segments of the genome. In the case of highly fragmented genomes (e.g., museum specimens), such genome reduction methods might be a very useful way of gathering data for phylogenetic studies. To study phylogenetic relationships among species, one usually analyzes an a priori known set of genetic markers, e.g., a set of single-copy, protein-coding, homologous genes. By targeting specific genes of interest, the TE method can be particularly relevant for phylogenetic studies. However, it has generally been thought that such reduction methods require good-quality DNA from fresh or properly stored tissue (Lemmon and Lemmon 2013, Jones and Good 2016). Regardless, TE methods have been used successfully on stored DNA extractions (Faircloth et al. 2012, McCormack et al. 2013), as well as museum specimens (Bi et al. 2013, Cruz-Dávalos et al. 2017, St Laurent et al. 2018).

Materials and Methods

Taxon Sampling

Specimens were taken from the collection at the Zoological Research Museum Alexander Koenig (ZFMK, Bonn, Germany). We sampled 16 available species of Epicopeiidae, including at most four specimens per species. In addition, we sampled two species of Sematuridae (*Anurapteryx interlineata* and *Mania empedocles*) and two specimens of Pseudobistonidae (*Pseudobiston pimratana*) to investigate the relationships between these three families. In total, 33 museum specimens collected between 1892 and 2001 were included (Table 1). The oldest sample is a *Parabraxas davidi* (Oberthür, 1885) specimen from 1892, whereas the most recent one is *Parabraxas flavomarginaria* (Leech, 1897) from 2001 (Table 1). We were not able to acquire samples of the genera *Chatamla* (Moore, 1881), *Burmeia* (Minet, 2002), *Mimaporina* (Wei and Yen, 2017), or *Amana* (Walker, 1855), or samples of *Heracula discivitta*,

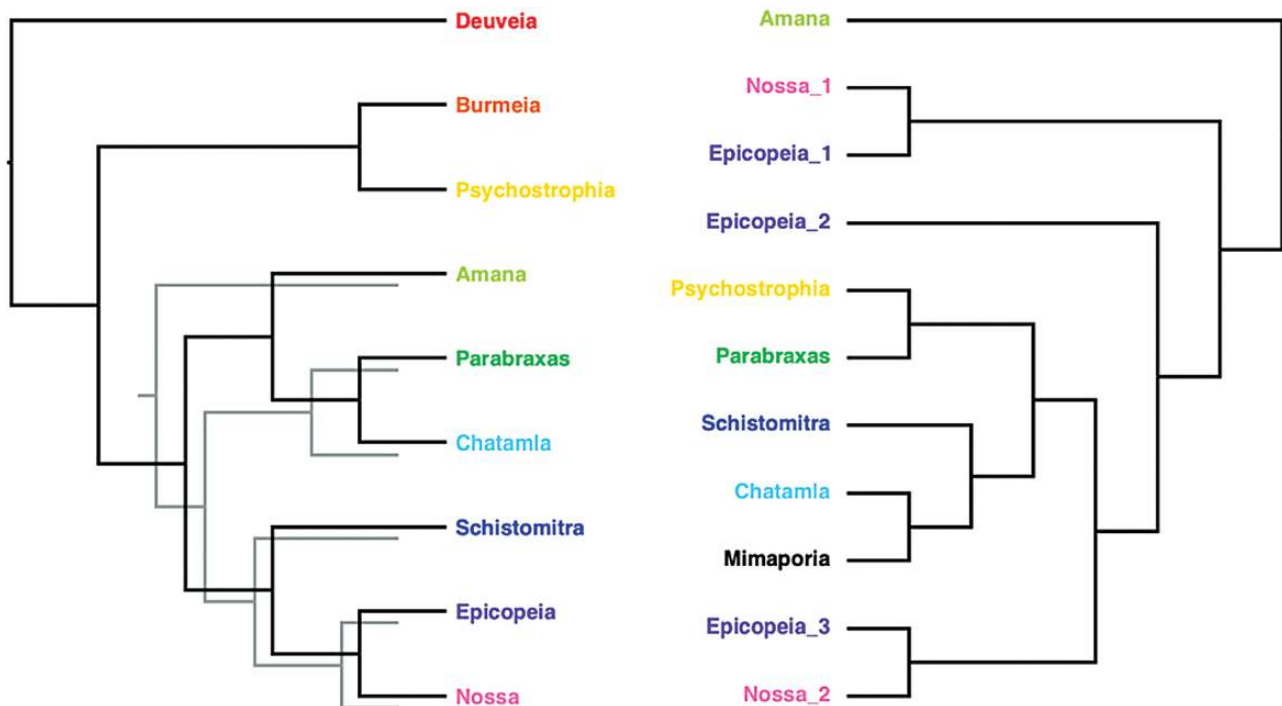


Fig. 1. Simplified representation of Epicopeiidae phylogenetic relationships according to Minet (2002) (left) and Wei and Yen (2017) (right). Each genus has a specific color. Minet's alternative hypothesis about the position of *Amana* is represented by gray lines.

Table 1. Number of raw recovered loci and selected loci per specimen

Family	Species	Specimen	Collection year	Raw loci	Loci found in at least 20 specimens	Reference
Epicopeiidae	<i>Deuveia banghaasi</i> (Hering, 1932)	S35	1936	12	11	This study
	<i>D. banghaasi</i>	S37	1936	666	353	This study
	<i>Epicopeia hainseii</i> (Holland 1889)	S51	1932	549	327	This study
	<i>E. hainseii</i>	S53	1932	936	373	This study
	<i>E. hainseii</i> (Moore, 1874)	S43	1951	1,063	374	This study
	<i>E. hainseii</i>	S45	2001	1,383	376	This study
	<i>E. philenora</i> (Westwood, 1841)	S55	1937	736	358	This study
	<i>E. philenora</i>	S57	1938	467	306	This study
	<i>E. polydora</i> (Westwood, 1841)	S47	1992	1,270	378	Mayer et al. (2021)
	<i>E. polydora</i>	S49	1932	9	8	This study
	<i>Nossa moorei</i> (Elwes, 1890)	S11	1931	6	5	This study
	<i>N. moorei</i>	S13	1931	210	185	This study
	<i>N. nagaensis</i> (Elwes, 1890)	S9	1991	1	365	This study
	<i>N. nelcinna</i> (Moore, 1875)	S3	1932	0	0	This study
	<i>N. palaeartica</i> (Staudinger, 1887)	S5	1989	1	1	This study
	<i>N. palaeartica</i>	S7	1990	1,202	375	This study
	<i>N. palaeartica chinensis</i>	S1	1937	4	3	This study
	<i>Parabraxas davidi</i> (Oberthür, 1885)	S17	1892	516	330	Mayer et al. (2021)
	<i>P. davidi</i>	S19	1957	1	1	This study
	<i>P. davidi</i>	S21	1906	0	0	This study
	<i>P. flavomarginaria</i> (Leech, 1897)	S23	2001	982	351	This study
	<i>P. nigromacularia</i> (Leech, 1897)	S25	1999	1,215	378	This study
	<i>Psychostrophia endoi</i> (Inoue, 1992)	S27	1995	1,275	373	This study
<i>P. melanargia</i> (Butler, 1877)	S39	1956	1,115	372	This study	
<i>P. melanargia</i>	S41	1934	780	362	This study	
<i>P. nymphidiaria</i> (Oberthür, 1893)	S31	1938	848	367	This study	
<i>P. nymphidiaria</i>	S33	1946	739	364	Mayer et al. (2021)	
<i>P. picaria</i> (Leech, 1897)	S29	2001	1,219	375	This study	
<i>Schistomitra funeralis</i> (Butler, 1881)	S15	1966	151	131	This study	
Pseudobistonidae	<i>Pseudobiston pinratanai</i> (Inoue, 1994)	S2	1999	232	170	Mayer et al. (2021)
	<i>P. pinratanai</i>	S63	1999	890	361	Mayer et al. (2021)
Sematuridae	<i>Anurapteryx interlineata</i> (Walker, 1854)	S61	?	928	376	Mayer et al. (2021)
	<i>Mania empedocles</i> (Cramer, 1782)	S59	1960	431	283	This study

which was recently moved to the family Pseudobistonidae (Wang et al. 2019). Details for all the specimens included can be found on Zenodo (doi:10.5281/zenodo.3769000).

Sample Preparation and DNA Extractions

We used a semidestructive approach, i.e., we removed the abdomen for DNA extraction without grinding the tissue, thus preserving the genitalia for future preparation (Hundsdoerfer and Kitching 2010). Genitalia dissections are routinely done for Lepidoptera by boiling abdomens in KOH to remove soft tissue, thus destroying the DNA in the process, so our approach is less destructive than what is normally done. For large specimens (like *Nossa* or *Epicopeia*), the abdomen was cut in half above the genitalia to ensure that they fit inside 1.5-ml Eppendorf tubes. Abdomens were first soaked in 180- μ l H₂O, for about 5 min, to rehydrate tissues. Water was removed before starting DNA extractions. Samples were lysed at 56°C overnight shaking with 350 rpm (by using a thermomixer) for approximately 12–18 h. We used the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and followed the standard DNA extraction protocol for tissues, with the following modifications: we included an RNase-digestion step and eluted the DNA in MilliQ water. Finally, DNA concentration of each sample was quantified using a Quantus Fluorometer (Promega, Madison, WI), and fragment lengths were measured with a Fragment Analyzer (Advanced Analytical, now Agilent Technologies Inc., Santa Clara, CA).

Library Preparation, TE, and Sequencing

There is still no consensus on how the DNA in museum specimens is best accessed. Here we used TE, a genome reduction approach. TE methods use probes, designed to target specific regions of the genome (Breinholt et al. 2018, Toussaint et al. 2018, Espeland et al. 2019). In the case of phylogenetic studies, this approach has the main advantage to recover exactly the loci of interest, and as long as a probe kit exists for a group, no previous knowledge about the genomes of the group of interest is required. This approach follows three major steps: 1) bait design, 2) libraries preparations and sequencing, and 3) filtering and processing of the data.

Regarding the bait design, new genes were selected and added to the Butterfly1.0 kit by Espeland et al. (2018). Mayer et al. (2021) designed hybrid enrichment baits with BaitFisher software version 1.2.8 (Mayer et al. 2016). A bait length of 120 bp was specified with a clustering threshold of 0.15, and a tiling design of 3 baits per bait region with an overlap of 60 bp for two consecutive baits resulting in bait regions with a total length of 240 bp. Individual coding sequences (CDS) from *Danaus plexippus* (Linnaeus), *Melitaea cinxia* (Linnaeus), *Heliconius melpomene* (Linnaeus), *Papilio glaucus* (Linnaeus), *Plutella xylostella* (Linnaeus), *Bombyx mori* (Linnaeus), and *Manduca sexta* (Linnaeus) were used as references. The LepZFMK1.0 kit includes 2,954 probe regions in different CDS regions belonging to 1,754 genes and is compatible with BUTTERFLY1.0 (Espeland et al. 2018) and partially compatible

with LEP1 (Breinholt et al. 2018). For more details on the kits, see Mayer et al. (2021). In many cases, multiple exons of single genes were targeted when they were long enough.

Library preparation was performed at the Zoological Research Museum Alexander Koenig (Bonn, Germany). Most of our samples contained less than 100-ng genomic DNA, which is the needed concentration according to standard protocol, but we included them anyway. With the Fragment Analyzer we found that many fragments of our samples were around 140 bp; therefore, no fragmentation was necessary for these samples. Other samples with higher quality and longer fragments were fragmented with Bioruptor PICO sonicator (Diagenode, Seraing, Belgium) to obtain DNA fragments with an approximate length of 350 bp.

We repaired the DNA with NEBNext FFPE DNA Repair Mix (NEB, Ipswich, United Kingdom), following the manufacturer's protocol. We purified the reactions with Agencourt AMPure XP beads with a ratio of (1:3). We quantified the resulting libraries with Quantus Fluorometer (Promega) and quality checked with a Fragment Analyzer (Advanced Analytical, now Agilent Technologies Inc.).

We proceeded to the enrichment and captured steps with the Agilent SureSelect XT2 protocol, with additional modification following Bank et al. (2017). Enrichment and sequencing were done at StarSEQ GmbH (Mainz, Germany) on Illumina Nextseq 500 Systems with a read length of 150 bp. Exons found in at least 20 of the 33 specimens (with an average of 254 loci per specimen) were used for downstream phylogenetic analyses (Table 1). Sequencing data is available at the NCBI under Bioproject PRJNA684488.

Data Clean up and Assembly

Reads were trimmed with fastq-mcf (Aronesty 2011) using default parameters to remove adapters and low-quality regions. Data cleaning and assembly was done using the iterated bait assembly (IBA) pipeline (Breinholt et al. 2018) with default parameters, except that the paired gap length was set to 100 (-g 100). Genomic sequences of the target regions from *D. plexippus* were used as a reference for the IBA pipeline. In brief, reads similar to the reference sequence were identified with USEARCH (Edgar 2010) and assembled

with Bridger (Chang et al. 2015). The resulting assembly was then used as a reference sequence for another run of USEARCH, and this process was repeated three times.

Alignments

The loci were aligned using the FFT-NS-i algorithm with two iterations in MAFFT v.7 (Katoh and Standley 2013) prior to phylogenetic analyses. Alignments were trimmed to the probe regions by using TrimAl (Capella-Gutierrez et al. 2009), with the options '-gapthreshold' and '-conservative'. These commands were implemented to remove gaps. The alignment cleanup was performed with HmmerCleaner (Di Franco et al. 2019), which allows the detection and removal of primary sequence errors in multiple alignments. We used the commands '-costs' and '--noX' and defined the four costs as follows: -0.15, -0.08, 0.15, and 0.45. We subsequently manually checked for frame shifts, gaps, and codon positions. Finally, alignments containing less than 20 samples (excluding references) were discarded from the downstream analysis. The final filtered data set consisted of 378 genes.

Screening Available Genomes and Transcriptomes

Additional 19 taxa were added to our data set by mining available genomes and transcriptomes, including one epicopeiid and one sematurid (Table 2). Twelve of the transcriptomes were from the superfamily Geometroidea, the remaining ones were from other macroheteroceran superfamilies. Raw reads were downloaded from the NCBI Sequence Read Archive (Leinonen et al. 2011). Reads were first processed to remove low-quality regions ($Q < 30$), adapters and homopolymer stretches using Cutadapt 1.4.1 (Martin 2011; minimum read length 50 bp) and Prinseq 0.20.4 (Schmieder and Edwards 2011), respectively. De novo assembly was carried out with Trinity 2.0.6 (Grabherr et al. 2011, Haas et al. 2013), with default parameters, including a minimum contig length of 100 bp and a minimum kmer coverage of 5.

Identification of the 378 genes was carried out with a BLAST approach. A reference sequence set was created from the TE alignments from one representative per gene. A tblastn (Gertz et al. 2006) search of the reference set against the transcriptomes (*e*-value threshold

Table 2. List of the 19 available transcriptomes and genomes added to this study

Family	Subfamily	Species	Source/accession numbers
Bombycidae	Bombycinae	<i>Bombyx mori</i> (Linnaeus, 1758)	SilkDB
Crambidae	Crambinae	<i>Chilo suppressalis</i> (Walker, 1863)	LepBase v4
Erebidae	Arctiinae	<i>Callimorpha dominula</i> (Linnaeus, 1758)	SRR1191023
Epicopeiidae	—	<i>Epicopeia hainseii</i> (Holland, 1889)	SRR1021610
Geometridae	Larentiinae	<i>Operophtera brumata</i> (Linnaeus, 1758)	LepBase v4
	Ennominae	<i>Biston betularia</i> (Linnaeus, 1758)	SRR1021599
		<i>Biston suppressaria</i> (Guenée, 1858)	SRR1777716
		<i>Ectropis obliqua</i> (Prout, 1915)	SRR3056076
		<i>Macaria distribuararia</i> (Hubner, 1825)	SRR1299213
	Geometrinae	<i>Chlorosea margaretararia</i> (Sperry, 1944)	SRR1021603
		<i>Nemoria lixaria</i> (Guenée, 1858)	SRR1299347
	Sterrhinae	<i>Idaea eremiata</i> (Hulst, 1887)	SRR1021615
Noctuidae	Hadeninae	<i>Spodoptera frugiperda</i> (Smith, 1797)	SRR3406055
Notodontidae	Nystaleinae	<i>Notoplusia minuta</i> (Druce, 1900)	SRR1299746
Pyrilidae	Phycitinae	<i>Amyelois transitella</i> (Walker, 1863)	LepBase v4
Sematuridae	—	<i>Mania lunus</i> (Linnaeus, 1758)	SRR1299318
Sphingidae	Sphinginae	<i>Manduca sexta</i> (Linnaeus, 1763)	LepBase v4
Uraniidae	Uraniinae	<i>Lyssa zampa</i> (Butler, 1869)	SRR1299769
	Epileminae	<i>Calledapteryx dryopterata</i> (Grote, 1868)	SRR1021601

Published transcriptomes have their SRA accession numbers listed. SRA, NCBI Sequence Read Archive.

10e-5) was carried out. The resulting BLAST output was used to extract the coding regions from each assembly using a set of open access python scripts from Dr. C. Peña (PyPhylogenomics, <https://github.com/carlosp420/PyPhyloGenomics>). The extracted sequences were aligned to the existing alignment with MAFFT 7.266 (Katoh and Standley 2013) using the ‘add fragments’ and ‘auto’ options, to preserve existing gaps in the alignment and choose the most appropriate alignment strategy, respectively. The resulting alignments were manually screened to ensure accurate alignment and frame preservation.

Phylogenetic Analyses

To partition our data set, we calculated the relative rates of evolution for each site in the alignment using TIGER (Cummins and McInerney 2011) and created partitions using the RatePartitions algorithm (Rota et al. 2018). We tested a range of d values (1.1, 1.5, 2.0, 3.0, and 4.0), which affects the number of partitions, and calculated the Bayesian information criteria (BIC) values for each partitioning scheme in PartitionFinder2 (Guindon et al. 2010, Frandsen et al. 2015, Lanfear et al. 2017). The partitioning scheme with the highest BIC value was found for $d = 2.0$, which resulted in 14 subsets.

Using the optimal partitioning scheme, we inferred the phylogenetic relationships with IQ-TREE 1.6.10 (Nguyen et al. 2015, Chernomor et al. 2016) under the maximum likelihood (ML) criterion. We used the model finding option in IQ-TREE (Kalyaanamoorthy et al. 2017) to find the optimal model for each partition. To investigate the robustness of our inferences, we used 1,000 ultrafast bootstraps (-bb; Hoang et al. 2018) and 1,000 replicates for SH-aLRT (-alrt; Guindon et al. 2010), which is the minimum recommended number.

Results

Genes

We recovered a total of 2,131 raw loci. From our total of 33 specimens, two (6%) provided no data: *Nossa nelcimna* (S3) and *P. davidi* (S21). Six specimens provided only 1–12 raw loci (18%); for 16 specimens, we obtained between 150 and 1,000 loci (48%); finally, nine specimens gave more than 1,000 loci, with a maximum of 1,383 loci recovered (27%; Table 1, Fig. 2).

There is a positive correlation between the collection date of the specimens and the number of recovered loci ($\rho = 0.46$, $P = 0.008$;

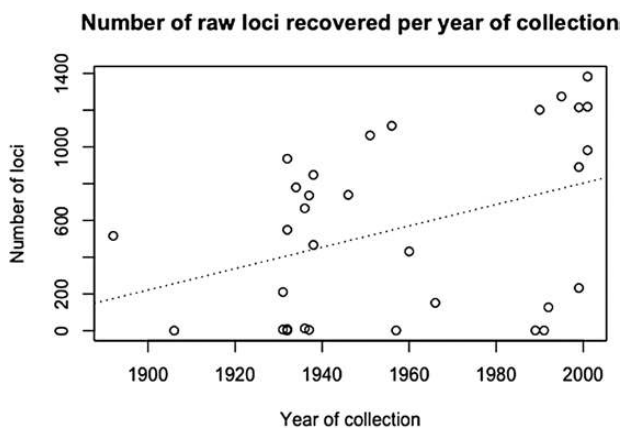


Fig. 2. Number of raw loci recovered for each sample per year of collection. The dashed line is for reference and represents the trend. Plot made on R.

Fig. 2). As expected, the younger a specimen is, the more loci we can recover from it. However, there is a lot of variation, meaning some recently collected specimens can give fewer loci than specimens collected a long time ago. This is, e.g., the case in two specimens of *P. davidi*. We recovered 516 raw loci from the older of the two, collected in 1892, whereas the more recent one (1957) provided only a single raw locus.

We obtained on average 254 loci and a median of 353 loci per specimen (Table 1). For our phylogenetic analyses, we first used all the 31 specimens that produced some data, including the 6 from Mayer et al. (2021). The samples *Epicopeia philenora* (S57) and *Nossa palaeartica* (S5) appeared to be contaminated as their phylogenetic position in preliminary analyses were highly doubtful, and thus they were excluded from the rest of our analyses.

Our final data set comprised 37 species, including 20 species sequenced for this study and 17 outgroup species with published transcriptomes. The data matrix included 378 nuclear loci (327 genes), for a total alignment of 134,881 base pairs. The average length of the 378 loci involved in this study is 367 bp.

Model Selection and Phylogenetic Analyses

The ML analyses for the different models tested gave the same phylogenetic relationships, and there were no conflicting nodes. The taxon data set, extended with 17 outgroup species, analyzed in IQ-TREE resulted in a highly supported ML tree (Fig. 3). We also performed the same phylogenetic analyses where we excluded specimens with less than 10 loci, and we obtain the same topology (Supp Material 1 [online only]), indicating that the necessarily somewhat limited data recovered from old specimens are of sufficient quality for phylogenetic analysis. Although our data set gave strong support for many of the branches, the relationships among the Noctuoidea, Bombycoidea, and Geometroidea were weakly supported.

The monophyly of Epicopeiidae is strongly supported, and the sister group is Pseudobistonidae, with Sematuridae being sister to these two, also with strong support (SH-like = 100, UFBoot = 100). Within Epicopeiidae, almost all relationships are strongly supported, with the exception of the position of *Schistomitra funeralis* (SH-like = 54.8, UFBoot = 81). Relationships of genera are congruent with Minet (2002) and Zhang et al. (2020), i.e., *Deuweia* is sister to the rest of Epicopeiidae, with *Psychostrophia* branching off next, then *Schistomitra*, and finally *Parabraxas* being sister to a clade containing paraphyletic *Epicopeia* and *Nossa* (Fig. 3).

Within genera, species for which two or more individuals were included were mainly monophyletic, with the exception of *Epicopeia hainseii* and *Epicopeia polydora*, which were intermixed in a clade with very short branches (Fig. 3). The branch leading to *P. davidi* has weak support values (66.1/94), and this species appears to be genetically very closely related to *P. flavomarginaria*. In addition, *Nossa moorei* is not genetically differentiated from *Nossa nagaensis*, while being morphologically very similar (Fig. 3).

Discussion

Phylogenetic Relationships

Within Epicopeiidae, our results strongly support and are almost entirely congruent with the relationships suggested by Minet (2002) and Zhang et al. (2020) and thus highly incongruent with the results of Wei and Yen (2017). We find *Deuweia* to be sister to the rest of Epicopeiidae, with the monophyletic *Psychostrophia* being sister to the rest of all taxa excluding *Deuweia* (Fig. 3). Wei and Yen (2017) found *Parabraxas* to be sister to *Psychostrophia*, but our

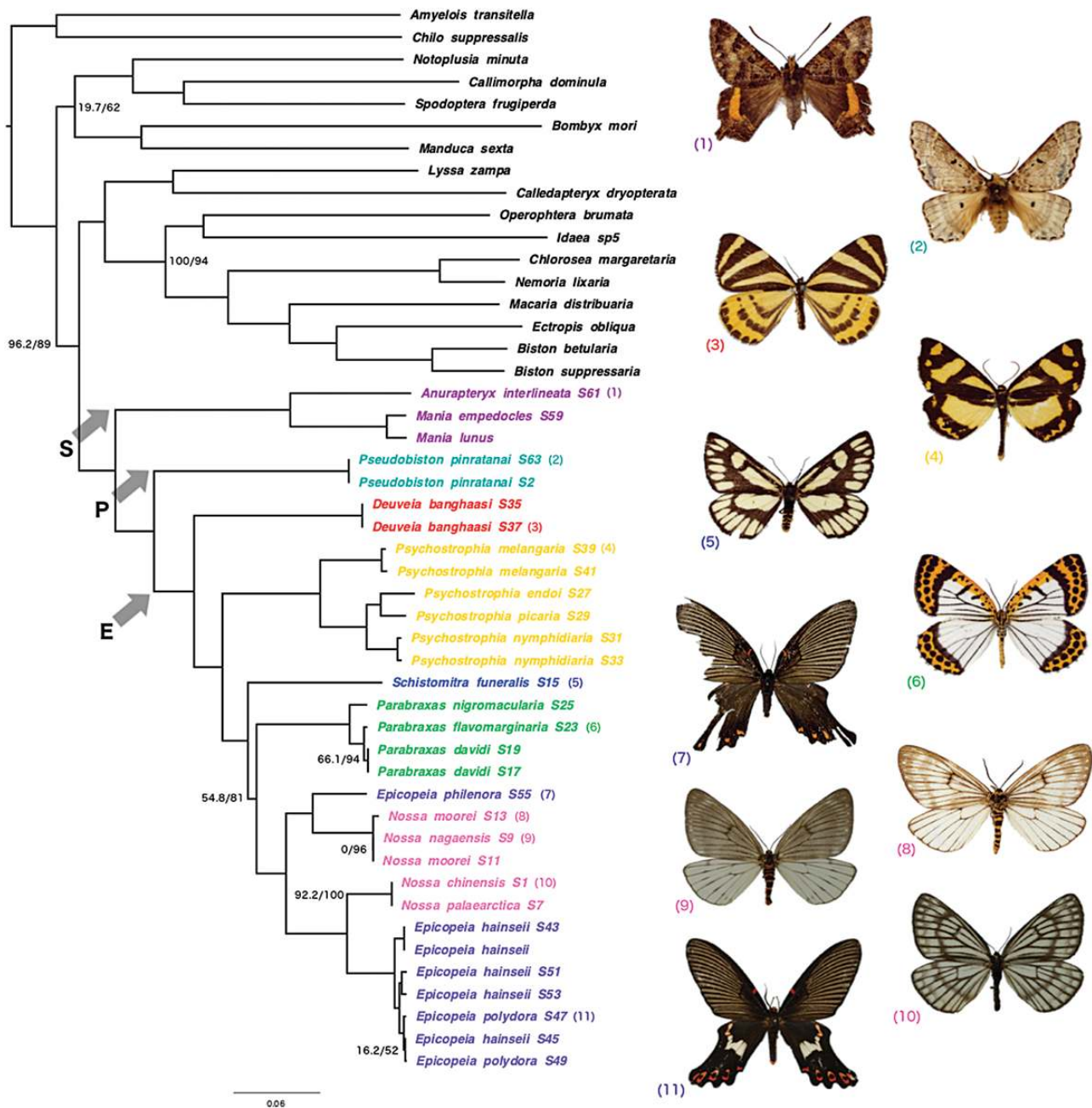


Fig. 3. Phylogenetic tree from maximum likelihood analysis of 36 taxa based on 378 loci. If the support values are not displayed on the branch, it means it is equal to 100/100. When displayed, numbers are the SH-aLRT support (%)/ultrafast bootstrap support (%). The images are representative species (indicated with numbers; not to scale). The three families are represented by an arrow and a letter. S, Sematuridae; P, Pseudobistonidae; and E, Epicopeiidae.

results place *Parabraxas* in a clade with *Schistomitra* and (*Epicopeia* + *Nossa*) with strong support.

The position of *S. funeralis* (which has 131 loci in our dataset) was incongruent with the hypothesis by Minet (2002), but with low support. In our study, we found *Schistomitra* to be the sister group of *Parabraxas* + (*Epicopeia* + *Nossa*) (Fig. 3), whereas Minet (2002) found it to be the sister group of *Epicopeia* + *Nossa* (Fig. 1), and Zhang et al. (2020) found it to be sister to *Parabraxas* + *Chatamla*. Wei and Yen (2017) found it to be sister to *Chatamla* + the newly described genus *Mimapora*, and this clade to be closer to *Parabraxas* + *Psychostrophia* than to *Epicopeia* + *Nossa* (Fig. 1). However, we are not able to confidently resolve the relationships of *Schistomitra*,

Parabraxas, and (*Epicopeia* + *Nossa*). Our study does not include the taxa *Amana*, *Chatamla*, or *Mimapora*, which are all potentially related to *Schistomitra* and *Parabraxas* (Minet, 2002). All four genera, *Schistomitra*, *Amana*, *Chatamla*, and *Mimapora*, are currently being considered to be monotypic, and their relationships based on morphology are somewhat enigmatic (Minet 2002, Wei and Yen 2017). Zhang et al. (2020) did include all four genera, and they were able to resolve their phylogenetic positions with confidence.

As in Zhang et al. (2020), we find that *Nossa* and *Epicopeia* are paraphyletic with regard to each other. Indeed, *E. philenora* appears to be the sister group to *N. moorei* and *N. nagaensis*, whereas *N. palaeartica* comes out as related to *E. hainseii* and *E. polydora*.

Furthermore, these relationships are well supported. Minet (2002) also finds the two genera to be closely related and sharing six apomorphic character states, despite being superficially quite distinct with *Epicopeia* species tending to mimic papilionids, and *Nossa* species tending to mimic pierid species (Fig. 3). Clearly, these two genera need to be studied in more detail by including all 12 described species. It is possible that the genera should be synonymized, in which case *Epicopeia* would have priority. Also, we found *E. hainseii* and *E. polydora* to be genetically inseparable based on our dataset. In contrast, Zhang et al. (2020) find these two taxa to be completely separate, with *E. polydora* being sister to *N. moorei*, in a similar position to our *E. philenora*. Zhang et al. (2020) did not sample *E. philenora*, but *E. polydora* and *E. philenora* are morphologically very similar, suggesting that our sequences may be contaminants.

The paraphyly of *Epicopeia* and *Nossa* is surprising. These two genera are morphologically superficially very different, with *Epicopeia* species showing distinct tails on the hindwings, whereas *Nossa* species lack these tails. Indeed, Minet separated these two genera on morphological characters, including their genitalia (Minet 2002). However, one should keep in mind that *Epicopeia* are mimicking species of butterflies in the genera *Papilio* and *Byasa* (that have tails on the hindwings), whereas *Nossa* is thought to mimic species of Pieridae (that do not have tails; Wei and Yen 2017, Zhang et al. 2020). It has been considered that mimicry might be one of the causes for the rapid divergence of phenotypes (Turner 1976, Counterman et al. 2010, Kozak et al. 2015). Thus, in further work, we need to investigate this aspect by including more species and individuals of *Epicopeia* and *Nossa*.

Within Epicopeiidae, specimens with few loci explain most branches with low support (the exception being *Schistomitra* described above). When we removed the four specimens with less than 10 loci (see Table 1) from our analyses, the relationships do not change, while the support greatly improved to reach the maximum value of 100/100 on some branches, like for *N. moorei* and *N. nagaensis*, or for the relationships between *Epicopeia hainseii* and *E. polydora* (Supp Material 1 [online only]). This would indicate that specimens with few loci are only affecting the support values, but not the general topology.

Here we obtain strong support for the hypothesis that Sematuridae is the sister group of Epicopeiidae + Pseudobistonidae. Even with few representatives for Sematuridae and Pseudobistonidae, the support for this hypothesis is compelling (100/100) and in line with previous studies (Rajaei et al. 2015, Kawahara et al. 2019, Wang et al. 2019). Furthermore, we confirmed that Epicopeiidae is monophyletic with regard to Pseudobistonidae, strengthening the case for the latter family.

The first attempt to resolve the position of Pseudobistonidae was made when the family was described by Rajaei et al. (2015) to accommodate *P. pinratanai*. Rajaei et al. (2015) found the family to be the sister group of Epicopeiidae. Recently, the position of Pseudobistonidae was corroborated with the addition of another species in the family: *H. discivitta* (Wang et al. 2019). However, Wang et al. (2019) only included three Epicopeiidae and two Sematuridae species. Furthermore, the support for the branches leading to these three families was quite low, e.g., the branch supporting Sematuridae as the sister group of Epicopeiidae + Pseudobistonidae had a bootstrap value of 33. Zhang et al. (2020) include *Heracula* in their dataset and find it to be sister to Epicopeiidae with strong support; thus, it would appear that Pseudobistonidae is indeed the sister lineage to Epicopeiidae, with Sematuridae being sister to these two.

Old Material and Contamination

We see a tendency for old museum specimens to yield fewer loci than the more recently collected ones (Fig. 2). Overall, the older a specimen is, the lower the chances are to get DNA out of it with the TE approach. Nevertheless, some old specimens provide more loci than younger ones. For instance, for the two specimens of *P. davidi*, the older, collected in 1892, provided 516 raw loci, whereas the younger, collected in 1957, provided only a single raw locus. There is no clear explanation for these kinds of outliers, but they might be due to different treatments during their curation (Espeland et al. 2010, Burrell et al. 2015, Vaudo et al. 2018). Unfortunately, neither a proper record of this kind of treatment nor how specimens have been collected and curated are usually available, making it impossible here to infer what other factors than age can affect the quality of DNA. Regardless, even if the tendency is, as expected, that older samples have less and poorer DNA quality, it remains a trend. Therefore, we should not discount these specimens just because they are old, as they can still turn out to be real genetic treasure troves.

Unfortunately, two specimens were definitely contaminated, *E. philenora* (S57) and *N. palaeartica* (S5), and therefore were not analyzed further. If they had been of good quality, they could have helped us to confirm the position of *E. philenora* in the case of S57, as well as the separation of *Nossa* in two groups with *N. moorei* + *N. nagaensis* on one side and *N. palaeartica* (S5) on the other side. In addition, our *E. polydora* specimens were found to be genetically identical to *E. hainseii*, in stark contrast to Zhang et al. (2020). One of our specimens (S47) yielded 1,270 raw loci (Table 1), suggesting large amounts of DNA in the extract. The two species cannot be confused morphologically (see doi:10.5281/zenodo.3769000). Clearly, this needs to be investigated in more detail, but for the moment, we do not have a good explanation for these results.

The Importance of Museomics

Since their creation, natural history museums have been an essential source of biological knowledge and resources for both the scientific community and the public (Duckworth et al. 1993, Suarez and Tsutsui 2004). These collections of biological specimens are vital for the study of systematics, global climate change research, biological invasion studies, as well as for many other scientific disciplines (Bi et al. 2013, Bradley et al. 2014, Bakker et al. 2020). Curated specimens in museums have several advantages compared with collecting fresh specimens; they can be easy to access, most of them are identified, and often possess information such as the date of collection and the location. Moreover, nowadays, researchers in biology and ecology face many challenges before being able to sample in the field. These issues can be monetary (e.g., lack of funding), stochastic events (inaccessibility of species of interest, adverse weather conditions, pandemic etc.), but also administrative difficulties, with bureaucratic hurdles being erected at an increasing pace (Neumann et al. 2018). Natural history museums also contain extinct taxa, rare and challenging to collect species, which can be a crucial asset to studies. However, until recently, this vast amount of biological resources was mainly used for morphological studies because the DNA from these specimens was thought to be too degraded to be used for molecular studies (Shapiro and Hofreiter 2012). Due to this, DNA work has for a long time mainly been limited to species for which freshly collected samples could be obtained, whereas DNA work from collections has been limited to sequencing short fragments DNA (Hajibabaei et al. 2006, Lozier and Cameron 2009, Strutzenberger et al. 2012, Hebert et al. 2013, Cameron et al. 2016).

We have taken advantage of recent advances in sequencing technologies, which have opened up access to genomic data of museum specimens. Within the past few years, various studies emerged applying these methods on a wide variety of species: from birds (Anmarkrud and Lifjeld 2017, Cloutier et al. 2018) and mammals (Fabre et al. 2014, Hawkins et al. 2016) to insects (Kanda et al. 2015, Sproul and Maddison 2017), and plants (Zedane et al. 2016, Silva et al. 2017). Part of these studies used whole-genome sequencing (Kanda et al. 2015, Zedane et al. 2016, Sproul and Maddison 2017, Cloutier et al. 2018), whereas the others employed diverse genome reduction methods, such as exon capture (Bi et al. 2013) and TE (Hawkins et al. 2016). Although these studies used different kinds of sequencing methods, they focus on very distinct scientific questions: from systematics (Silva et al. 2017), to the origin and diversification of a taxon (Fabre et al. 2014), to population genomics (Bi et al. 2013).

Here, we used a genome reduction method, TE, on curated museum specimens of rare and challenging to collect moth species, to refine our knowledge of their phylogenetic relationships. We managed to recover on average 566 nuclear loci per species using the TE method. The present study also shows that it is possible to extract substantial amounts of DNA sequence data from specimens collected up to 127 yr ago. Hence, our study contributes to the field of museomics, demonstrating the application of this sequencing method on museum specimens, increasing the value of such specimens even further. Museomics opens a window to the past, providing possibilities for testing new hypotheses and for casting new light on old ones.

Conclusion

In summary, we conducted a phylogenetic analysis on small and rare families of Lepidoptera, using museum specimens. We successfully sequenced samples that were collected between 1892 and 2001. By utilizing a TE approach, we were able to recover between 150 and 1,383 loci per specimen for 75% of our samples. From all these raw loci, we used 378 genes—present in at least 20 samples—to reconstruct a phylogenetic hypothesis based on ML analysis of 37 taxa. This analysis corroborates, with strong support, the hypothesis that Sematuridae are the sister group of Epicopeiidae + Pseudobistonidae. Within Epicopeiidae, our study finds *Deuweia* as sister group of the rest of Epicopeiidae genera. The position of *Schistomitra* is incongruent with the central hypothesis suggested by Minet (2002) for this family; however, the support for this branch is low. The low support for this branch might be explained in our study by the lack of some genera (*Amana*, *Chatamla*, and *Mimapura*). Indeed, these taxa may help to clarify the phylogenetic position of *Schistomitra*, as seen in Zhang et al. (2020). Although we showed that *Psychostrongia* and *Parabraxas* are monophyletic, we also found that *Nossa* and *Epicopeia* are paraphyletic. Overall, the genera of Epicopeiidae require more work to reveal their phylogenetic relationships.

Museum collections represent a varied and essential biobank of samples for studying the diversity on earth. The availability of specimens, not only rare but also extinct, within worldwide museum collection is a fantastic asset. Nowadays, sequencing techniques are powerful enough to allow scientists to recover DNA from old museum specimens. This is the beginning of an exciting era for molecular studies. Our study makes its contribution to the field of museomics by successfully demonstrating that researchers can use museum samples at a molecular level for phylogenetic studies. Consequently, this study is paving the way for more molecular work using museum specimens.

Supplementary Data

Supplementary data are available at *Insect Systematics and Diversity* online.

Supplementary Material 1. Phylogenetic tree from ML analysis of 36 taxa based on 378 loci, specimens with less than 10 loci were excluded.

Acknowledgments

We are thankful to Claudia Eitzbauer for help with ordering the kit and to Sandra Kukowka for assistance in the molecular lab. We highly appreciate the effort of everyone depositing samples at the ZFMK. The study was funded by the Zoological Research Museum Alexander Koenig, and received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 6422141.

References Cited

- Allio, R., C. Scornavacca, B. Nabholz, A.-L. Clamens, F. A. H. Sperling, and F. L. Condamine. 2019. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. *Syst. Biol.* 69: 38–60.
- Anmarkrud, J. A., and J. T. Lifjeld. 2017. Complete mitochondrial genomes of eleven extinct or possibly extinct bird species. *Mol. Ecol. Resour.* 17: 334–341.
- Aronesty, E. 2011. Fastq-mcf sequence quality filter, clipping and processor. (Com/p/ea-utils/wiki/fastqmcfc).
- Bakker, F. T., A. Antonelli, J. A. Clarke, J. A. Cook, S. V. Edwards, P. G. P. Ericson, S. Faurby, N. Ferrand, M. Gelang, and R. G. Gillespie, et al. 2020. The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ* 8: e8225.
- Bank, S., M. Sann, C. Mayer, K. Meusemann, A. Donath, L. Podsiadlowski, A. Kozlov, M. Petersen, L. Krogmann, and R. Meier, et al. 2017. Transcriptome and target DNA enrichment sequence data provide new insights into the phylogeny of vespid wasps (Hymenoptera: Aculeata: Vespidae). *Mol. Phylogenetics Evol.* 116: 213–226. doi:10.1016/j.ympev.2017.08.020.
- Bazinet, A. L., M. P. Cummings, K. T. Mitter, and C. W. Mitter. 2013. Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. *PLoS One* 8: e82615.
- Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next-generation museum population genomics. *Mol. Ecol.* 22: 6018–6032.
- Bradley, R. D., L. C. Bradley, H. J. Garner, and R. J. Baker. 2014. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. *BioScience* 64: 1150–1158.
- Breinholt, J. W., C. Earl, A. R. Lemmon, E. M. Lemmon, L. Xiao, and A. Y. Kawahara. 2018. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Syst. Biol.* 67: 78–93.
- Burrell, A. S., T. R. Disotell, and C. M. Bergery. 2015. The use of museum specimens with high-throughput DNA sequencers. *J. Hum. Evol.* 79: 35–44.
- Cameron, S. A., H. C. Lim, J. D. Lozier, M. A. Duennes, and R. Thorp. 2016. Test of the invasive pathogen hypothesis of bumble bee decline in North America. *Proc. Natl. Acad. Sci. USA* 113: 4386–4391.
- Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Chang, Z., G. Li, J. Liu, Y. Zhang, C. Ashby, D. Liu, C. L. Cramer, and X. Huang. 2015. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* 16: 30.
- Chapman, A. D. 2005. Uses of primary species-occurrence data. *Global Biodiversity Information Facility, Copenhagen, Denmark.*

- Chernomor, O., A. von Haeseler, and B. Q. Minh. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65: 997–1008.
- Cloutier, A., T. B. Sackton, P. Grayson, S. V. Edwards, and A. J. Baker. 2018. First nuclear genome assembly of an extinct moa species, the little bush moa (*Anomalopteryx didiformis*). bioRxiv, doi:10.1101/262816, 9 February 2018, preprint.
- Cong, Q., J. Shen, D. Borek, R. K. Robbins, P. A. Opler, Z. Otwinowski, and Grishin. V. 2017. When coi barcodes deceive: complete genomes reveal introgression in hairstreaks. *Proc. R. Soc. B Biol. Sci. USA* 284: 20161735.
- Counterman, B. A., F. Araujo-Perez, H. M. Hines, S. W. Baxter, C. M. Morrison, D. P. Lindstrom, R. Papa, L. Ferguson, M. Joron, and R. H. ffrench-Constant, et al. 2010. Genomic hotspots for adaptation: the population genetics of Müllerian Mimicry in *Heliconius erato*. *PLoS Genet.* 6: e1000796.
- Cruz-Dávalos, D. I., B. Llamas, C. Gaunitz, A. Fages, C. Gamba, J. Soubrier, P. Librado, A. Seguin-Orlando, M. Pruvost, A. H. Alfarihan, et al. 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* 17: 508–522.
- Cummins, C. A., and J. O. McInerney. 2011. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Syst. Biol.* 60: 833–844.
- Di Franco, A., R. Poujol, D. Baurain, and H. Philippe. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* 19: 21.
- Duckworth, W. D., H. H. Genoways, and C. L. Rose. 1993. Preserving natural science collections: chronicle of our environmental heritage. *Mammology Papers: University of Nebraska State Museum No. 271*. National Institute for the Conservation of Cultural Property, Washington, DC. p 153.
- Edgar, R. 2010. Usearch. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA. (<https://www.osti.gov/science/cinema/biblio/1137186>).
- Espeland, M., M. Irestedt, K. A. Johanson, M. Åkerlund, J.-E. Bergh, and M. Källersjö. 2010. Dichlorvos exposure impedes extraction and amplification of DNA from insects in museum collections. *Front. Zool.* 7: 2.
- Espeland, M., J. Breinholt, K. R. Willmott, A. D. Warren, R. Vila, E. F. A. Toussaint, S. C. Maunsell, K. Aduse-Poku, G. Talavera, and R. Eastwood, et al. 2018. A comprehensive and dated phylogenomic analysis of butterflies. *Curr. Biol.* 28: 770–778.e5.
- Espeland, M., J. W. Breinholt, E. P. Barbosa, M. M. Casagrande, B. Huertas, G. Lamas, M. A. Marín, O. H. H. Mielke, J. Y. Miller, S. Nakahara, et al. 2019. Four hundred shades of brown: higher level phylogeny of the problematic Euptychiina (Lepidoptera, Nymphalidae, Satyrinae) based on hybrid enrichment data. *Mol. Phylogenet. Evol.* 131: 116–124.
- Fabre, P.-H., J. T. Vilstrup, M. Raghavan, C. Der Sarkissian, E. Willerslev, E. J. P. Douzery, and L. Orlando. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biol. Lett.* 10: 20140266–20140266.
- Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61: 717–726.
- Fletcher, D. S. 1979. Geometroidea: apoprogonidae, Axiidae, Callidulidae, Cyclidiidae, Drepanidae, Epicopeiidae, Epiplemidae, Geometridae, Pterothysanidae, Sematuridae, Thyatiridae, Uraniidae, p. 243. *In* I.W.B. Nye (ed.), *The generic names of moths of the world*, Vol. 3. British Museum (Natural History), London, UK.
- Frandsen, P. B., B. Calcott, C. Mayer, and R. Lanfear. 2015. Automatic selection of partitioning schemes for phylogenetic analyses using iterative *k*-means clustering of site rates. *BMC Evol. Biol.* 15: 13.
- Gertz, E. M., Y.-K. Yu, R. Agarwala, A. A. Schäffer, and S. F. Altschul. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.* 4: 41.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, and Q. Zeng, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644–652.
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59: 307–321.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, and M. Lieber, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8: 1494–1512.
- Hajibabaei, M., M. A. Smith, D. H. Janzen, J. J. Rodriguez, J. B. Whitfield, and P. D. N. Hebert. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* 6: 959–964.
- Hawkins, M. T. R., C. A. Hofman, T. Callicrate, M. M. McDonough, M. T. N. Tsuchiya, E. E. Gutiérrez, K. M. Helgen, and J. E. Maldonado. 2016. In-solution hybridization for mammalian mitogenome enrichment: Pros, cons and challenges associated with multiplexing degraded DNA. *Mol. Ecol. Resour.* 16: 1173–1188.
- Hebert, P. D. N., J. R. Dewaard, E. V. Zakharov, S. W. J. Prosser, J. E. Sones, J. T. A. McKeown, B. Mantle, and J. La Salle. 2013. A DNA ‘barcode blitz’: rapid digitization and sequencing of a natural history collection. *PLoS One* 8: e68535.
- Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35: 518–522.
- Hundsdoerfer, A. K., and I. J. Kitching. 2010. A method for improving DNA Yield from century-plus old specimens of large lepidoptera while minimizing damage to external and internal abdominal characters. *Arthropod Syst. Phylog.* 68: 151–155.
- Jones, M. R., and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* 25: 185–202.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14: 587–589.
- Kanda, K., J. M. Pflug, J. S. Sproul, M. A. Dasenko, and D. R. Maddison. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.
- Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780.
- Kawahara, A. Y., D. Plotkin, M. Espeland, K. Meusemann, E. F. A. Toussaint, A. Donath, F. Gimmich, P. B. Frandsen, A. Zwick, and M. Reis, et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. USA* 116: 22657–22663.
- Kozak, K. M., N. Wahlberg, A. F. E. Neild, K. K. Dasmahapatra, J. Mallet, and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Syst. Biol.* 64: 505–524.
- Laithwaite, E. R., and P. E. S. Whalley. 1975. *Dictionary of butterflies and moths in color*. Michael Joseph, London, United Kingdom. (<http://agris.fao.org/agris-search/search.do?recordID=US201300521786>).
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34: 772–773.
- Leinonen, R., H. Sugawara, and M. Shumway. 2011. The sequence read archive. *Nucl. Acids Res.* 39: D19–D21.
- Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. *Ann. Rev. Ecol. Evol. Syst.* 44: 99–121.
- Li, W., Q. Cong, J. Shen, J. Zhang, W. Hallwachs, D. H. Janzen, and N. V. Grishin. 2019. Genomes of skipper butterflies reveal extensive convergence of wing patterns. *Proc. Natl. Acad. Sci. USA* 116: 6232–6237.
- Lozier, J. D., and S. A. Cameron. 2009. Comparative genetic analyses of historical and contemporary collections highlight contrasting demographic histories for the bumble bees *Bombus pensylvanicus* and *B. impatiens* in Illinois. *Mol. Ecol.* 18: 1875–1886.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17: 10–12.
- Mayer, C., M. Sann, A. Donath, M. Meixner, L. Podsiadlowski, R. S. Peters, M. Petersen, K. Meusemann, K. Liere, and J. W. Wägele, et al. 2016.

- BaitFisher: a software package for multispecies target DNA enrichment probe design. *Mol. Biol. Evol.* 33: 1875–1886.
- Mayer, C., L. Dietz, E. Call, S. Kukowka, S. Martin, and M. Espeland. 2021. Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens. *Syst. Entomol.* doi:10.1111/syen.12481.
- McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66: 526–538.
- Minet, J. 1983. Étude morphologique et phylogénétique des organes tympaniques des Pyraloidea. 1 – Généralités et homologues (Lep. Glossata). *Ann. La Société Entomol. France* 19: 175–207.
- Minet, J. 1986. Ébauche d'une classification moderne de l'ordre des Lépidoptères. *Alexandria* 14: 291–313.
- Minet, J. 2002. The epicopeiidae: phylogeny and a redefinition, with the description of new taxa (Lepidoptera: Drepanoidea). *Ann. La Société Entomol. France* 38: 463–487.
- Neumann, D., A. V. Borisenko, J. A. Coddington, C. L. Häuser, C. R. Butler, A. Casino, J. C. Vogel, G. Haszprunar, and P. Giere. 2018. Global biodiversity research tied up by juridical interpretations of access and benefit sharing. *Organ. Divers. Evol.* 18: 1–12.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32: 268–274.
- Rajaci, H., C. Greve, H. Letsch, D. Stünning, N. Wahlberg, J. Minet, and B. Misof. 2015. Advances in Geometroidea phylogeny, with characterization of a new family based on *Pseudobiston pimatanai* (Lepidoptera, Glossata). *Zool. Scr.* 44: 418–436.
- Regier, J. C., A. Zwick, M. P. Cummings, A. Y. Kawahara, S. Cho, S. Weller, A. Roe, J. Baixeras, J. W. Brown, and C. Parr, et al. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evol. Biol.* 9: 280.
- Rota, J., T. Malm, N. Chazot, C. Peña, and N. Wahlberg. 2018. A simple method for data partitioning based on relative evolutionary rates. *PeerJ* 6: e5498.
- Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
- Shapiro, B., and M. Hofreiter. 2012. *Ancient DNA: methods and protocols*. Humana Press, New York, NY.
- Silva, C., G. Besnard, A. Piot, J. Razanatsoa, R. P. Oliveira, and M. S. Vorontsova. 2017. Museomics resolve the systematics of an endangered grass lineage endemic to north-western Madagascar. *Ann. Bot.* 119: 339–351.
- Sproul, J. S., and D. R. Maddison. 2017. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Mol. Ecol. Resour.* 17: 1183–1201.
- St Laurent, R. A., C. A. Hamilton, and A. Y. Kawahara. 2018. Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae). *Syst. Entomol.* 43: 729–761.
- Strutzenberger, P., G. Brehm, and K. Fiedler. 2012. DNA barcode sequencing from old type specimens as a tool in taxonomy: a case study in the diverse genus *Eois* (Lepidoptera: Geometridae). *PLoS One* 7: e49710.
- Suarez, A. V., and N. D. Tsutsui. 2004. The value of museum collections for research and society. *BioScience* 54: 66.
- Suchan, T., C. Pitteloud, N. S. Gerasimova, A. Kostikova, S. Schmid, N. Arrigo, M. Pajkovic, M. Ronikier, and N. Alvarez. 2016. Hybridization capture using RAD Probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One* 11: e0151651.
- Toussaint, E. F. A., J. W. Breinholt, C. Earl, A. D. Warren, A. V. Z. Brower, M. Yago, K. M. Dexter, M. Espeland, N. E. Pierce, and D. J. Lohman, et al. 2018. Anchored phylogenomics illuminates the skipper butterfly tree of life. *BMC Evol. Biol.* 18: 101.
- Turner, J. R. G. 1976. Adaptive radiation and convergence in subdivisions of the butterfly genus *Heliconius* (Lepidoptera: Nymphalidae). *Zool. J. Linn. Soc.* 58: 297–308.
- Vaudo, A. D., M. L. Fritz, and M. M. López-Urbe. 2018. Opening the door to the past: accessing phylogenetic, pathogen, and population data from museum curated bees. *Insect Syst. Divers.* 2: 4. doi:10.1093/isd/ixy014.
- Wang, H., J. D. Holloway, N. Wahlberg, M. Wang, and S. Nylin. 2019. Molecular phylogenetic and morphological studies on the systematic position of *Heracula discivitta* reveal a new subfamily of Pseudobistonidae (Lepidoptera: Geometroidea). *Syst. Entomol.* 44: 211–225.
- Wei, C.-H., and S.-H. Yen. 2017. *Mimaporina*, a new genus of Epicopeiidae (Lepidoptera), with description of a new species from Vietnam. *Zootaxa* 4254: 537.
- Zedane, L., C. Hong-Wa, J. Murienne, C. Jeziorski, B. G. Baldwin, and G. Besnard. 2016. Museomics illuminate the history of an extinct, paleoendemic plant lineage (Hesperelaea, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biol. J. Linn. Soc.* 117: 44–57.
- Zhang, J., Q. Cong, J. Shen, E. Brockmann, and N. Grishin. 2019. Genomes reveal drastic and recurrent phenotypic divergence in firetip skipper butterflies (Hesperiidae: Pyrrhopyginae). *Proc. R. Soc. B Biol. Sci.* 286: 20190609.
- Zhang, Y., S. Huang, D. Liang, H. Wang, and P. Zhang. 2020. A multilocus analysis of Epicopeiidae (Lepidoptera, Geometroidea) provides new insights into their relationships and the evolutionary history of mimicry. *Mol. Phylogenet. Evol.* 149: 106847.