Systematic Entomology

Systematic Entomology (2020), DOI: 10.1111/syen.12433

A deeper meaning for shallow-level phylogenomic studies: nested anchored hybrid enrichment offers great promise for resolving the tiger moth tree of life (Lepidoptera: Erebidae: Arctiinae)

NICOLAS J. DOWDY[1,2,3], SHANNON KEATING[4], ALAN R. LEMMON[5], EMILY M. LEMMON[6], WILLIAM E. CONNER[3], CLARE H. SCOTT CHIALVO[7], SUSAN J. WELLER[8], REBECCA B. SIMMONS[9], MELISSA S. SISSON[9] and JENNIFER M. ZASPEL[1,2,4]

[1]Department of Zoology, Milwaukee Public Museum, Milwaukee, WI, U.S.A., [2]Department of Entomology, Purdue University, West Lafayette, IN, U.S.A., [3]Department of Biology, Wake Forest University, Winston-Salem, NC, U.S.A., [4]Department of Biology, Marquette University, Milwaukee, WI, U.S.A., [5]Department of Scientific Computing, Florida State University, Tallahassee, FL, U.S.A., [6]Department of Biological Science, Florida State University, Tallahassee, FL, U.S.A., [7]Department of Biology, Appalachian State University, Boone, NC, U.S.A., [8]University of Nebraska State Museum, University of Nebraska-Lincoln, Lincoln, NE, U.S.A. and [9]Department of Biology, University of North Dakota, Grand Forks, ND, U.S.A.

Abstract. Anchored hybrid enrichment (AHE) has emerged as a powerful tool for uncovering the evolutionary relationships within many taxonomic groups. AHE probe sets have been developed for a variety of insect groups, though none have yet been shown to be capable of simultaneously resolving deep and very shallow (e.g., intraspecific) divergences. In this study, we present NOC1, a new AHE probe set (730 loci) for Lepidoptera specialized for tiger moths and assess its ability to deliver phylogenetic utility at all taxonomic levels. We test the NOC1 probe set with 142 individuals from 116 species sampled from all the major lineages of Arctiinae (Erebidae), one of the most diverse groups of noctuoids (>11 000 species) for which no well-resolved, strongly supported phylogenetic hypothesis exists. Compared to previous methods, we generally recover much higher branch support (BS), resulting in the most well-supported, well-resolved phylogeny of Arctiinae to date. At the most shallow-levels, NOC1 confidently resolves species-level and intraspecific relationships and potentially uncovers cryptic species diversity within the genus *Hypoprepia*. We also implement a 'sensitivity analysis' to explore different loci combinations and site sampling strategies to determine whether a reduced probe set can yield results similar to those of the full probe set. At both deep and shallow levels, only 50–175 of the 730 loci included in the complete NOC1 probe set were necessary to resolve most relationships with high confidence, though only when the more rapidly evolving sites within each locus are included. This demonstrates that AHE probe sets can be tailored to target fewer loci without a significant reduction in BS, allowing future studies to incorporate more taxa at a lower per-sample sequencing cost. NOC1 shows great promise for resolving long-standing taxonomic issues and evolutionary questions within arctiine lineages, one of the most speciose clades within Lepidoptera.

Correspondence: Nicolas J. Dowdy, Department of Zoology, Milwaukee Public Museum, 800 W Wells Street, Milwaukee, Wisconsin, 53233, U.S.A.
E-mail: njdowdy@gmail.com

Introduction

The development of next-generation sequencing methods has facilitated the production and growth of genomic resources for a wide variety of nonmodel organisms. These massive datasets allow systematists to utilize hundreds or thousands of molecular markers for phylogenetic reconstruction with the aim of reconciling relationships that were previously unresolved, poorly supported and/or incongruent between analyses. Many phylogenomic studies of animals utilize transcriptomic datasets (Hittinger *et al.*, 2010; Hedin *et al.*, 2012; Kawahara & Breinholt, 2014; Wickett *et al.*, 2014; Garrison *et al.*, 2016; Bazinet *et al.*, 2017) or hybrid enrichment (Hodges *et al.*, 2007; Gnirke *et al.*, 2009) to target, isolate and sequence designated regions of the genome, depending on the scope of the evolutionary hypotheses being tested. Transcriptome-based methods are sensitive to specimen condition, requiring carefully preserved or fresh tissues, limiting their practical use in constructing large phylogenies (Ozsolak & Milos, 2011). Hybrid enrichment techniques, such as Ultraconserved Elements (UCE; Faircloth *et al.*, 2012) and Anchored Hybrid Enrichment (AHE; Lemmon *et al.*, 2012; Lemmon & Lemmon, 2013) are capable of utilizing tissues stored using standard methods including ethanol preservation and even dried, museum material (Blaimer *et al.*, 2016; St. Laurent *et al.*, 2018). These two methods differ mainly in the genomic targets of their probe designs (Lemmon & Lemmon, 2013). UCE focuses on ultra-conserved regions of genomic DNA, which are conserved across taxa at deep phylogenetic scales (Faircloth *et al.*, 2012; McCormack *et al.*, 2012). AHE also captures homologous DNA sequences shared at deep scales but aims to target relatively less-conserved regions using variable probes that represent the sequence diversity in the group under study (Lemmon *et al.*, 2012). In the context of arthropod phylogenetics, these and other large genomic datasets have provided resolution across diverse lineages such as Myriapoda (Fernández *et al.*, 2018), Insecta (Misof *et al.*, 2014), Diptera (Young *et al.*, 2016), Hymenoptera (Peters *et al.*, 2017), Coleoptera (Shin *et al.*, 2017; Zhang *et al.*, 2018b; McKenna *et al.*, 2019), Neuroptera (Winterton *et al.*, 2018) and Arachnida (Hamilton *et al.*, 2016), to name a few.

Among Lepidoptera, transcriptomic data have proven useful in reconstructing framework phylogenies for deep divergences at the level of order (Bazinet *et al.*, 2013; Kawahara & Breinholt, 2014) and within (Scott Chialvo *et al.*, 2018). A recent study by Breinholt *et al.* (2018) combined transcriptomic data with data derived from an AHE probe set ('Lep1') to resolve relationships within Lepidoptera at both deep and shallow taxonomic levels. This approach was mostly effective for resolving phylogenetic relationships at the superfamily and interfamilial levels, but the extent to which Lep1 can address evolutionary affinities within certain groups remains unclear (e.g., Johns *et al.*, 2018). A notably difficult group for Lep1 to resolve was the Noctuoidea, the largest superfamily of Lepidoptera, which contains >25% of all lepidopteran species including many agricultural pests and species which perform important ecosystem services such as pollination (Mitchell *et al.*, 2006; van Nieukerken *et al.*, 2011; Howard & Barrows, 2014). Lep1 combined with additional transcriptomic and genomic data (>2.5 Mbp in total) analysed with a maximum likelihood (ML) approach generated poor branch support (<70%) for all examined interfamilial relationships within the Noctuoidea (Breinholt *et al.*, 2018). Recently, the Lep1 probe set was applied within the Erebinae, a subfamily of Erebidae, one of the largest families within Noctuoidea (Homziak *et al.*, 2018). Only 658 loci of the total 855 within the Lep1 probe set were recovered for this group. While this study recovered the most well-resolved phylogeny of Erebinae to date, many relationships still lacked strong statistical support, making the placement of major groups uncertain (Homziak *et al.*, 2018). One strategy to deal with this is to create more taxon-specialized AHE probe sets. Using the Lep1 probe set as a foundation, Espeland *et al.* (2018) created the 'BUTTERFLY1.0' probe set by combining the loci within Lep1 with high capture success within Papilionoidea with additional reference sequences to further improve loci capture rate. The BUTTERFLY1.0 kit was further specialized for Hesperiidae in a later study ('BUTTERFLY1.1') (Toussaint *et al.*, 2018). A similar approach was taken to produce the Bombycoidea-specific probe set 'BOM1' (Hamilton *et al.*, 2019). This probe kit specialization has generally resulted in an increased number of captured loci as well as an overall improved view of evolutionary relationships within the targeted groups (Hamilton *et al.*, 2019). In this study, we produce a new AHE probe set (NOC1) to improve support at both deep and shallow taxonomic levels within an important group of Lepidoptera. Producing robust results at all taxonomic levels is necessary for conducting comparative studies among closely related taxa, particularly in lineages where resources for examining trait evolution are rich.

One such noctuoid lineage is the tiger moths, a subfamily of erebid moths encompassing at least 11 000 species in 750 genera worldwide (Watson & Goodger, 1986; Weller *et al.*, 2009; van Nieukerken *et al.*, 2011) with new genera and species continually being described, revealing hidden diversity within the subfamily (e.g., Vincent *et al.*, 2014; Pinheiro & Duarte, 2016; Joshi *et al.*, 2017; Schmidt & Sullivan, 2018; Volynkin *et al.*, 2018). Moths in this subfamily are known for their bright coloration and mimicry (Fig. 1) as well as their complex defensive and mating strategies. Consisting of both generalist and specialist feeders, tiger moths often utilize toxic plants (e.g., Asteraceae, Boraginaceae, Fabaceae, Apocynaceae) and lichens as hosts (Fig. 2A,B). Many species have been shown to engage in pharmacophagy, wherein adults and/or larvae actively seek out and sequester secondary metabolites from their hosts for purposes other than nutrition (Boppré, 1981; Conner & Jordan, 2009). These sequestered toxins can be utilized for defence against vertebrate and invertebrate predators during both the adult and larval stages, self-medication against parasitism (Singer *et al.*, 2009) and/or attracting and protecting their mates (Conner *et al.*, 2000). Tiger moths are also well-known for their ability to signal their toxicity to bird and bat predators with aposematic wing patterns (Conner, 2009; Rojas *et al.*, 2019) and ultrasonic clicks (Blest *et al.*, 1963; Dunning, 1967; Barber *et al.*, 2009; Dowdy & Conner, 2016). In certain species, these clicks are capable of disrupting, or 'jamming' bat echolocation (Fullard *et al.*, 1979;

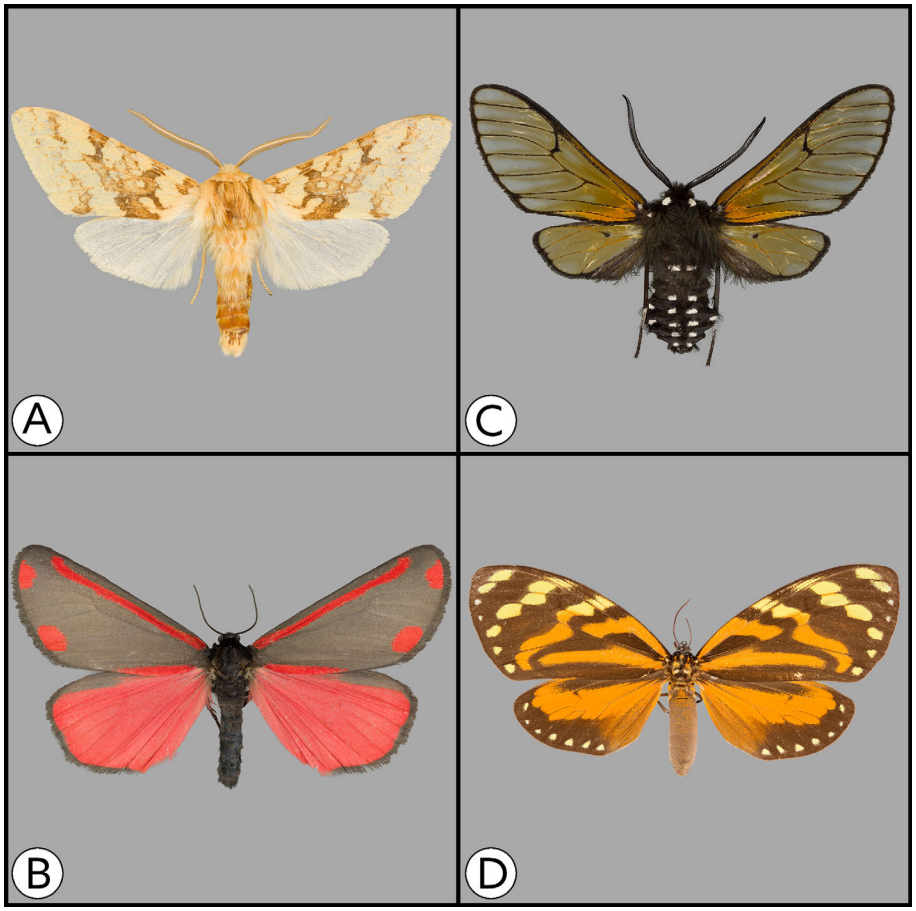

Fig. 1. Exemplar Adult Arctiinae. Members of Arctiinae exhibit a diverse array of visual patterns, including cryptic coloration (A; *Lophocampa maculata*), highly contrasting aposematic coloration (B, *Tyria jacobaeae*), clear-winged wasp mimicry (C; *Dasysphinx volatilis*) and butterfly mimicry (D; *Chetone angulosa*). https://doi.org/10.6084/m9.figshare.9995324 [Colour figure can be viewed at wileyonlinelibrary.com].

Corcoran *et al.*, 2009; Corcoran & Conner, 2012; Conner & Corcoran, 2012; Fig. 2E–G). Some have even co-opted their acoustic defence for use as a courtship signal (Conner, 1987; Sanderford & Conner, 1995; Simmons & Conner, 1996; Sanderford *et al.*, 1998; Fig. 2C,D). The variety of communication strategies within Arctiinae makes it a tractable system for understanding the evolution of both inter- and intraspecific signalling, as well as the co-option of traits to serve multiple functions.

The hyperdiversity of tiger moths appears to be the result of an extremely rapid adaptive radiation, which took place as recently as 25–45 million years ago during the late Eocene and early Oligocene (Sohn *et al.*, 2012; Toussaint *et al.*, 2012; Kawahara *et al.*, 2019). As an example, despite being 2–4 times younger than the most speciose butterfly family (Nymphalidae), the tiger moths contain nearly twice the number of species (van Nieukerken *et al.*, 2011; Espeland *et al.*, 2018). Recent diversity estimates suggest that tiger moths may be one of the most diverse subfamilies of Lepidoptera, rivalling both the generic and species diversity of many lepidopteran families and superfamilies, as well as being among the most rapidly speciating clades (van Nieukerken *et al.*, 2011). Accounting for the evolutionary relationships among members of such a hyperdiverse, relatively young and quickly evolving clade within the Lepidoptera using traditional morphological and molecular tools has been a major challenge.

Historically, cladistic analyses based on morphological characters have supported the monophyly of the Arctiinae (Jacobson & Weller, 2002). However, there are still many taxa within the subfamily with uncertain phylogenetic placements. This is likely due to difficulties in establishing morphological homologies across early-branching lineages, repeated reduction of wing venation and examination of traits in life stages that are poorly known or unavailable for most taxa. Many Arctiinae also exhibit convergent evolution of aposematic wing patterns and associated wing vein reduction, which has contributed to considerable confusion among historical classification schemes (see Kitching & Rawlins, 1998; Jacobson & Weller, 2002; Weller *et al.*, 2009 for a full review of taxonomic history).

Recent molecular studies based on 8–9 genetic markers have provided strong support for the monophyly of the subfamily

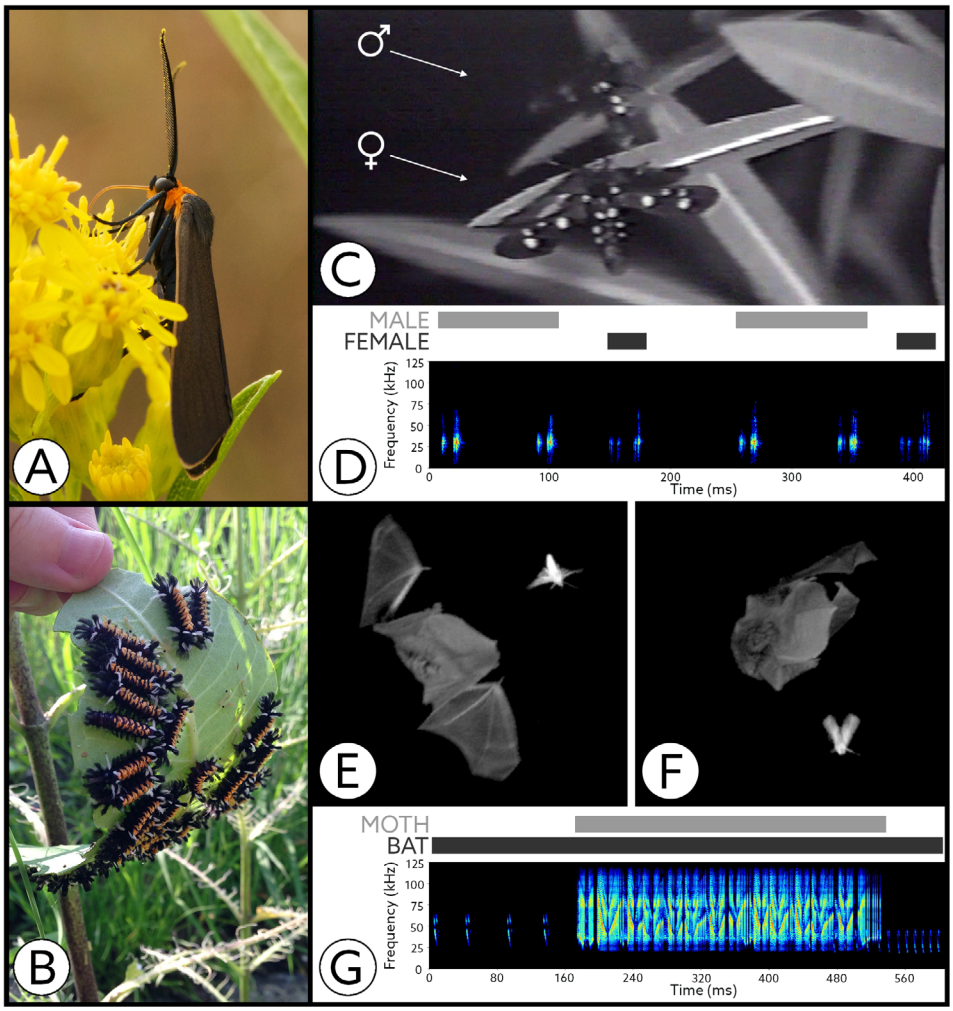

Fig. 2. Exemplar Arctiine Behaviors. Many tiger moths defend themselves with sequestered toxins, acquired either through adult pharmacophagy (A; *Cisseps fulvicollis* on *Solidago* sp.) or larval uptake from host plants (B; *Euchaetes egle* on *Asclepias syriaca*). Adult moths can also produce sound during courtship (C, D), as an advertisement to predators of their distastefulness, and/or to disrupt the echolocation cries of their bat predators (E–G). https://doi.org/10.6084/m9.figshare.9995327 [Colour figure can be viewed at wileyonlinelibrary.com].

and tribes therein, yet many of the shallower taxonomic groupings have lacked strong support (Wahlberg & Wheat, 2008; Zahiri *et al.*, 2012; Zaspel *et al.*, 2014; Zenker *et al.*, 2016). Results from these studies demonstrate that the traditional markers used to resolve relationships within Lepidoptera are not effective at resolving shallow divergences within Arctiinae, particularly within groups with high species diversity (e.g., lithosiines and phaegopterines, clades with an estimated diversity of more than 3000 species each). Many phylogenetic studies within tiger moths have examined the relationships at the subtribal and intergeneric levels using morphological and/or molecular evidence, but produced results that were weakly supported (Simmons & Weller, 2001; Simmons & Weller, 2002; Weller *et al.*, 2004; DaCosta & Weller, 2005; DaCosta *et al.*, 2006; Zaspel & Weller, 2006; Scott & Branham, 2012; Simmons *et al.*, 2012; Scott *et al.*, 2014), and in some cases incongruent (Scott & Branham, 2012 vs Scott *et al.*, 2014). In the most recent molecular study within the subfamily, Rönkä *et al.* (2016) used eight molecular markers to survey relationships among a subset of genera within the subtribe Arctiina. This study proposed 33 genus-level synonymies, putting it at odds with prior morphological evidence supporting distinct genera (Ferguson, 1985). Work examining the interspecific relationships within *Grammia* based on mitochondrial DNA (mtDNA) relative to other nonmolecular, 'morpho-ecological' traits suggests that in some cases, the often-utilized mtDNA markers may be inadequate for species delineation and species-tree estimation when used in isolation (Schmidt & Sperling, 2008). Species-level phylogenetic studies within arctiines are relatively scarce, with most being narrowly focused on either the generic placement of individual species (Vincent *et al.*, 2014), pairwise divergences of sister species using mtDNA barcodes (Vincent *et al.*, 2009) or clarifying species boundaries (Weller *et al.*, 2004).

Extreme species and trait diversity with relevance to chemical and behavioural ecology studies make tiger moths an attractive system for studying comparative evolution within a phylogenetic context. Despite this, no robust phylogenetic hypotheses currently exist for most lineages in the subfamily.

Herein, we use the subfamily Arctiinae as a case study to demonstrate NOC1's effectiveness at outperforming traditional molecular markers at all taxonomic levels within a diverse noctuoid subfamily, with an emphasis on the tribal-, subtribal-, and species-levels. To examine NOC1's performance at resolving traditionally intractable subtribal relationships, we greatly expanded the taxon sampling, particularly within the subtribe Phaegopterina, a clade which has had a difficult taxonomic history (reviewed in Vincent & Laguerre, 2014). To test the capabilities of NOC1 at resolving more recent divergences, we included multiple congeners for several genera. We also assess NOC1 at the very shallow species-complex level with multiple individuals from two species of *Hypoprepia* Hübner, a relatively small group of tiger moths (5 species) widely distributed in North America. Adult *H. fucosa* and *H. miniata* are brightly pigmented and have a bold aposematic pattern, presumably advertising the presence of distasteful lichen-derived phenolics sequestered via larval lichenivory (Rawlins, 1984). Previous attempts have been made to clarify identity through subspecies designations; yet, in certain regions, such as along the Gulf of Mexico, these designations have proven to be inconsistent and adult coloration appears to be of little diagnostic value.

Finally, we perform a 'sensitivity analysis' at a variety of taxonomic levels spanning from deep to very shallow to determine whether future iterations of the NOC1 probe set could be optimized to balance robust statistical results with additional taxon sampling through reduced per taxon sequencing costs. By utilizing loci with robust, multilevel phylogenetic utility, the investigation of trait evolution within this fascinating group of moths will finally be possible.

Materials and methods

Taxon sampling

2008 A complete list of taxa sampled, their localities, and read recovery data for each is provided in File S1. Voucher specimens are currently housed at the Milwaukee Public Museum. In total, we sampled 142 individuals from 116 species and 77 genera of Arctiinae, with an additional 3 species from 3 genera sampled from outgroup taxa derived from the erebid subfamilies Lymantriinae and Aganainae.

Tribal-level dataset

In order to examine the performance of our probe set to resolve the deepest divergences among the major tiger moth lineages we have included multiple representatives from all four currently recognized tribes within the Arctiinae. Genus and species-level percent coverage values are rough estimates. Our sampling focused on the two largest tribes, the Arctiini (55 genera, ∼10%; 87 species, ∼2%) and the Lithosiini (19 genera, ∼4%; 23 species, ∼1%), but also included members of the monogeneric tribe Amerilini (1 genus, 100%; 3 species, ∼5%) and the old world tribe Syntomini (2 genera, ∼3%; 3 species, ∼0.3%).

Subtribal-level dataset

The taxonomic sampling for our subtribal-level analysis focused on resolving relationships within the so-called 'PPCE' clade to clarify these subtribal relationships (Jacobson & Weller, 2002). The PPCE clade includes members of each of the major subclades, or 'generic affinities' of Phaegopterina (29 genera, 46 species) (Forbes, 1960; Weller *et al.*, 2009) as well as members of Pericopina (2 genera, 2 species), Ctenuchina (7 genera, 10 species) and Euchromiina (4 genera, 4 species). In addition, we sampled 23 species representing 19 genera from among 4 subtribes within the lichen-feeding tribe of tiger moths (Lithosiini).

Species-level dataset

We sampled multiple individuals of *H. fucosa* ($N = 8$) and *H. miniata* ($N = 11$) across their ranges (eastern North America from southern Canada to the Gulf of Mexico), including multiple specimens from certain localities. We also included specimens matching the phenotype of the northern North American subspecies *H. f. tricolor* ($N = 2$) and one currently unplaced specimen denoted here as '*Hypoprepia sp*.' from Louisiana ($N = 1$; VB11). We placed an emphasis on southern populations, which exhibit the most phenotypic diversity.

DNA extractions

Genomic DNA was extracted from excised thoracic tissue or 1–2 legs from specimens using the DNeasy tissue extraction kit (QIAGEN, Valencia, CA, U.S.A.) and following the manufacturer's protocol for animal tissues. DNA concentration was evaluated for each sample using a Qubit fluorometer (Life Technologies, Inc). DNA quality was determined by electrophoresis on a 1% agarose gel according to suggested protocols for the Anchored Laboratory of Phylogenomics (http://anchoredphylogeny.com/).

Probe design

Lepidopteran AHE target loci were previously identified by Breinholt *et al.* (2018), who developed an enrichment probe kit representing a dispersed set of lepidopteran reference species. Here, we improve the efficiency by which these targets can be enriched from noctuoid samples using two types of genomic resources for this group: (i) low coverage whole genome

sequence data and (ii) assembled transcriptome data. Low coverage whole genome sequence data (1x to 15x coverage) were collected for 10 individuals from eight species and three genera of erebid noctuoids (File S2). In short, indexed libraries were prepared from extracted DNA following Lemmon *et al.* (2012) and sequenced on two Illumina Hi-Seq lanes with a PE-150 protocol with C-bot clustering (total data = 91Gb). After quality filtering and demultiplexing the reads, the overlapping reads were merged following Rokyta *et al.* (2012). Merged reads were then mapped to the 855 AHE loci of Breinholt *et al.* (2018), using *Bombyx mori* as a reference. For each of the erebid reference individuals, the read best matching to each locus (minimum 55% similarity) was then used as a seed in an extension assembly in which the merged reads were used to extend the seed into adjacent regions (see Hamilton *et al.*, 2016 for methodological details). Between 53% and 76% of the loci were recovered.

We also utilized assembled transcriptome data from 31 individuals from 29 species and 25 genera of erebid noctuoids (File S2). For each AHE locus, the transcript from each species best matching to the AHE *B. mori* reference (minimum 55% similarity) was isolated and aligned with the extended whole-genome sequences (see above) using MAFFT v7.023b (Katoh & Standley, 2013). Alignments were then manually inspected in Geneious R9 (2015) (Biomatters Ltd. Kearse *et al.*, 2012) and trimmed down to well-aligned regions. These typically corresponded to whole exons, which could be identified through comparison of the transcriptome and whole-genome data. Poorly aligned and aberrant sequences were also removed. Alignments with less than 50% representation across the 41 references were removed from further consideration. This filter reduced the target set to 651 target loci that are shared with the Lep1 probe set (Breinholt *et al.*, 2018). After searching for the presence of common 60-mers to ensure that no target loci overlapped, the alignments were evaluated for repetitive elements, which were masked (see Hamilton *et al.*, 2016 for Methodological details). Finally, probes were tiled uniformly along each locus, at a tiling density of 2× per individual. This produced 130 747 probes, which were reduced to 32 533 probes after thinning to remove identical and very similar probes.

Library preparation

Library preparation and read data processing of the extracted DNA were completed following Prum *et al.* (2015) at the Center for Anchored Phylogenomics at Florida State University (www.anchoredphylogeny.com). Genomic DNA was sonicated to a fragment size of ∼200–600 bp via a Covaris E220 Focused-ultrasonicator. Libraries were prepared and indexed using a modified protocol from Meyer & Kircher (2010). Indexed samples were pooled in equal quantities, and the pools were enriched using an Agilent Custom SureSelect kit (Agilent Technologies) with AHE probes designed for Noctuoidea: Erebidae (i.e., NOC1). Sequencing was done on 3 PE150 Illumina HiSeq2500 lanes at the Translational Science Laboratory, College of Medicine, Florida State University. A lane was composed of approximately 50 samples.

Read assembly

To increase read accuracy and length, paired reads were merged before assembly, following Rokyta *et al.* (2012). Reads were mapped to the probe regions using *Virbia aurantiaca* (Arctiinae: Arctiini), *Hypoprepia fucosa* (Arctiinae: Liithosiini), *Calyptra minuticornis*, and *C. thalictri* (Calpinae: Calpini) as references. After mapping the reads to references, a quasi de novo assembly approach was used to extend the assembly into flanking regions (Prum *et al.*, 2015; Hamilton *et al.*, 2016). Read files were traversed repeatedly until no additional mapped reads were produced. Following read assembly, consensus bases were called from assembly clusters either as ambiguous or unambiguous bases, depending on probability of sequencing error. Assembly contigs based on fewer than 109 reads were removed to mitigate effects of rare sequencing errors and low-level contamination.

Orthology assessment

For each locus, orthology was determined following procedures described in Hamilton *et al.* (2016). A pairwise distance matrix among homologs was calculated using an alignment-free approach and used to cluster sequences with a neighbour-joining algorithm. This allowed the assessment of whether gene duplication occurred prior to or following the basal divergence of the clade. Duplication following basal divergence usually results in two clusters, one of which contains only a subset of the taxa. These were removed from further analysis if they contained fewer than 53 taxa.

Alignment and trimming

Sequences in each orthologous cluster were first aligned using MAFFT v7.023b (Katoh & Standley, 2013), then trimmed and masked following the procedure established in Hamilton *et al.* (2016). Sites with the same character in >60% of sequences were considered 'conserved'. A 20 bp sliding window was then moved across the alignment and regions with <14 characters matching the common base at the corresponding conserved site were masked. Sites with <70 unmasked bases were removed. Finally, the masked alignments were inspected by eye and regions considered obviously misaligned or paralogous were removed. After the bioinformatics filtering process, we obtained a dataset composed of 730 orthologous loci (420 574 bp; Mean Locus Length = 575 bp, 95% CI: [537 bp, 613 bp]), with 9.53% missing data ('-') or ambiguous bases ('N'). On average, the probes were recovered among >94% of taxa included in this study. Additional details about the loci within the NOC1 probe set are given in File S3 and individual locus alignments are given in File S4.

Phylogenetic analyses

Using the 730 orthologous loci with good alignments (420 574 bp), we estimated phylogenies for both individual gene trees and a concatenated dataset partitioned by locus using maximum likelihood under RAxML-HPC v.8 with the default rapid hill-climbing search algorithm, a GTRGAMMA substitution model, and 100 or 1000 bootstrap (BS) replicates for individual gene trees and the concatenated tree, respectively. The concatenated alignment used is available in File S5. We utilized a python wrapper script to generate the gene trees (available at https://github.com/dportik/Phylo_Wrapper_Scripts) and XSEDE (Stamatakis, 2014) on the CIPRES cluster (Miller *et al.*, 2010). We then used the individual gene trees to estimate a species tree using ASTRAL-III v 5.6.2 (Mirarab *et al.*, 2014; Zhang *et al.*, 2018a), a summary method based on the multispecies coalescent model. Support for the ASTRAL tree was generated by multilocus bootstrapping (Seo, 2008) using the bootstrap files for each gene tree generated by RAxML.

To examine the utility of the probe set at an extremely shallow taxonomic level, we sampled from the *Hypoprepia* species complex, including samples for the species *H. fucosa*, *H. miniata*, *H. inculta* and *Ptychoglene coccinea*. For these taxa, we constructed gene and species trees utilizing the same methods as described for our complete analysis. We compared the phylogenetic performance of our probe set to DNA barcode regions previously sequenced and published for *Hypoprepia* species. We used the program MITObim (Hahn *et al.*, 2013) to reconstruct the cytochrome oxidase subunit 1 (CO1) barcode sequence from the processed read data for 22 individuals representing different geographic locations within the *H. fucosa-miniata* complex. These regions were extracted and aligned to previously published sequences (Zahiri *et al.*, 2014; Adamowicz, 2015; Zahiri *et al.*, 2017, downloaded from NCBI GenBank; see File S1 for information about these sequences), and analysed using RAxML. The phylogenetic trees from these analyses are available as supplemental material (File S6).

Sensitivity analysis

The impact of locus sampling and site rate variation across deep and shallow taxonomic levels of the Arctiinae was assessed by performing a sensitivity analysis based on a method discussed in Buddenhagen *et al.*, 2016. This method subsamples loci from the hypothesized species tree according to rate variation and gene tree distance. This subsampling alters the data from which trees are inferred and tests the robustness of phylogenetic hypotheses by examining the stability of topologies between subsamples (Edwards, 2016). High evolutionary rates might cause site saturation, which can impede phylogenetic inference among distantly related taxa. However, more variable sites may also provide the phylogenetic signal necessary to reconstruct evolutionary relationships among closely related taxa. Subsampling the data allowed us to test the effect of per site evolutionary rate and gene tree heterogeneity on species tree support. This procedure tests the robustness of the relationships generated, but also examines the ability of the loci to return congruent results when altered to include slower or faster sites.

We first binned the individual loci by site rates using the program Tree Independent Generation of Evolutionary Rates (TIGER; Cummins & McInerney, 2011). TIGER categorizes the sites of a locus based on site disagreement, a proxy for molecular evolution rate. The sites are organized into bins such that the first bin contains the constant sites, and the highest-valued bin contains the most rapidly evolving sites. The remaining sites are placed into bins by splitting the rates into equal partitions (Cummins & McInerney, 2011). Each locus in the dataset was independently analysed and binned. We used the default setting of 10 bins, but given the conservative nature of these loci, there were relatively few sites sorted into the highest valued bins (i.e., most sites were sorted into bins 1–8). In the case of loci with a higher proportion of rapidly evolving sites (i.e., sites sorted into bins 9–10), the most conservative 2 or 3 bins were combined into one bin, such that the most conservative bin always contained some site variation and there were always 8 total bins. We used the program AMAS (Borowiec, 2016) to create seven new loci subsets by sequentially concatenating bins 1–8 (e.g., bins $1+2$, $1+2+3$, $1+2+3+4$, etc.). As more bins are concatenated, less conserved sites are added to the alignment. Maximum likelihood trees were constructed for each new alignment using the same parameters as above. The pairwise distance among the trees was estimated using treeCMP (Bogdanowicz *et al.*, 2012) using the triple metric (Critchlow *et al.*, 1996). This method requires a rooted tree, so a single outgroup taxon, *Asota ficus* (Erebidae: Aganainae), was used to root the subsampled trees. Loci that lacked this taxon were dropped from further analysis. The remaining binned alignment trees were plotted in multidimensional space using the R function cmdscale (R Core Team, 2019) to calculate the Euclidean distance of each subsampled tree to their average centre. Greater distances indicate a tree as being a greater outlier. This distance was used to rank loci into seven inclusion sets, with each successive set containing more outlying trees: 50 loci, 100 loci, 175 loci, 225 loci, 300 loci, 375 loci and 475 loci. This loci-ranking step was performed separately on each binning subset, such that the top 50 loci in the 2-bin subset were not necessarily the same as the top 50 loci in the 8-bin subset. Forty-nine final sets were created from the combination of 7 binning sets and 7 locus-inclusion sets using a Biopython script (Cock *et al.*, 2009). New binned alignment trees were generated in RAxML, and ASTRAL-III 5.6.2 was run in parallel on these utilizing a custom Python wrapper to generate an ASTRAL species tree (File S7). The loci were also concatenated to make maximum likelihood trees in RAxML. We used the RAxML -b flag to utilize the bootstrap files of the 49 loci sets to generate support values for our hypothesized species trees and assess the consistency of relationships as locus sites and binned alignment trees were subsampled. We used the R packages ape (Paradis *et al.*, 2004) and ggtree (Yu *et al.*, 2017) to visualize trees.

Internode certainty

Low phylogenetic support can be caused by poor phylogenetic signal and/or conflicting signals among the loci used for species tree inference. To test the effects of conflicting phylogenetic signal within the data set, we calculated the quadripartition internode certainty score (QP-IC; Zhou *et al.*, 2019) of internodes within the species trees. We calculated the QP-IC scores using the program QuartetScores, a quartet-based measure for examining incongruence within a set of phylogenetic trees (available at https://github.com/lutteropp/QuartetScores). We used the -r flag to map scores to the concatenated maximum likelihood tribal level species tree and the ASTRAL and concatenated maximum likelihood *Hypoprepia* tree, providing the individual gene trees as input to test the support of individual internodes within the species tree. This metric corrects for impartial gene trees wherein taxa are absent from some gene trees. This analysis provides insights into the congruency of individual phylogenetic relationships among the loci used to construct the species tree. QP-IC scores vary between −1 and 1, with higher values indicating that more individual genes recover the same internal branch as the reference tree (i.e., there is less conflict among gene trees), whereas values closer to 0 indicate that there is greater conflict (i.e., stronger support for one or more alternative topologies). Negative scores indicate that individual genes recover an alternative internode more often than the one given in the reference tree. Additionally, we calculated QP-IC for the 8-bin subset from the sensitivity analysis to examine how incongruence varied as more loci were included.

Results

Topological performance–tribal-level

Our analysis indicates that Arctiinae is a monophyletic group under both RAxML and ASTRAL (BS = 100/100%) (Fig. 3, File S8; for rectangularized tree see Files S6 and S9). Both analyses recovered a robust tribal topology comprised of the 4 tribes Lithosiini, Amerilini, Syntomini and Arctiini (I, II, III, and IV from Fig. 3, respectively). The Lithosiini (I) are sister to a clade composed of Amerilini (II), Syntomini (III) and Arctiini (IV) (BS = 100/100%). Amerilini (II) is sister to the clade comprised of Syntomini (III) and Arctiini (IV) (BS = 100/100%).

Topological performance–subtribal-level

Within the Arctiini, we recovered a well-supported, monophyletic PPCE-group utilizing RAxML as well as ASTRAL (BS = 100/100%) (Fig. 3, File S8). For branch lengths associated with the RAxML analysis, see File S10. The PPCE clade was found to be sister with members of the clade comprising the Callimorphina, Nyctemerina, Spilosomina, Arctiina, (5, 6, 7, 8 from Fig. 3, respectively) and a clade composed of *Utetheisa* + *Mangina* (BS = 100/100%). Within the PPCE clade, we found support for a monophyletic Pericopina, Euchromiina and Ctenuchina (9, 10, 11 from Fig. 3, respectively; All BS = 100/100%). Pericopina was recovered as sister with the remaining members of the PPCE clade (BS = 100/100%). Ctenuchina and Euchromiina were recovered as sisters (BS = 100/100%). However, our results indicate that the Phaegopterina is paraphyletic with respect to Ctenuchina + Euchromiina (BS = 100/100%). The PPCE clade was found to be composed of five subgroups including Pericopina, Ctenuchina + Euchromiina and three other non-sister clades traditionally classified together as Phaegopterina (A, B, C from Fig. 3). Our probe set provided strong support for nearly all branches in the PPCE clade (62 PPCE taxa, 61 nodes) with a few exceptions. The placement of *Pseudepimolis syrissa* relative to the *Bertholdia* + *Melese* clade and the *Idalus* + *Symphlebia* + *Amaxia* clade was uncertain under both RAxML and ASTRAL (BS = 68/37%). There was also some uncertainty in the relationships among three subclades of the phaegopterine 'Clade C' containing *Phaegoptera*, *Leucanopsis* and *Pachydota*, respectively. Our results provide moderate support under RAxML that the *Phaegoptera*-containing clade is most closely related to the *Pachydota*-containing clade, while ASTRAL only weakly supported this topology (BS = 75/33%). The clade comprised the Callimorphina, Nyctemerina, Spilosomina, Arctiina and *Utetheisa* + *Mangina* was found to be monophyletic (BS = 100/100%). Our probe set provided strong support for nearly all branches within this clade (27 taxa, 26 nodes). However, the Spilosomina were recovered to be paraphyletic, with respect to the taxon *Hypercompe laeta*, which was found to be more closely related with *Virbia*, ostensibly representing the Arctiina (BS = 91/91%).

Within the Lithosiini, our probe set provided strong support for the monophyly of all four sampled subtribes, including Nudarina, Ascalina, Lithosiina and Cisthenina (1, 2, 3, 4 from Fig. 3, respectively; BS = 100/100%). Relationships within all subtribes received strong support under RAxML (BS = 97–100%). Within each subtribe, ASTRAL and RAxML largely produced congruent topologies. However, ASTRAL recovered *Balbura dorsisigna* as sister with *Cisthene martini* with weak support (BS = 50%), whereas RAxML strongly supported *B. dorsisigna* as sister with the remainder of the Cisthenina, excluding *Hypoprepia* and *Ptychoglene* (BS = 100%). Within the Nudarina, ASTRAL placed the taxon *Schistophleps albida* with lower confidence compared to RAxML (BS = 66% vs 100%). While support for a monophyletic Lithosiina + Acsalina was strong (BS = 100/100%), the branch uniting these lineages with the Cisthenina received mixed support from RAxML and ASTRAL (BS = 53/100%).

Topological performance–species-level

Both the RAxML and ASTRAL analyses strongly supported the monophyly of at least two distinct species within the *Hypoprepia fucosa-miniata* species complex (Figs. 4 and 5). For branch lengths associated with the RAxML and ASTRAL analyses of *Hypoprepia*, see File S11 and File S12, respectively. Both

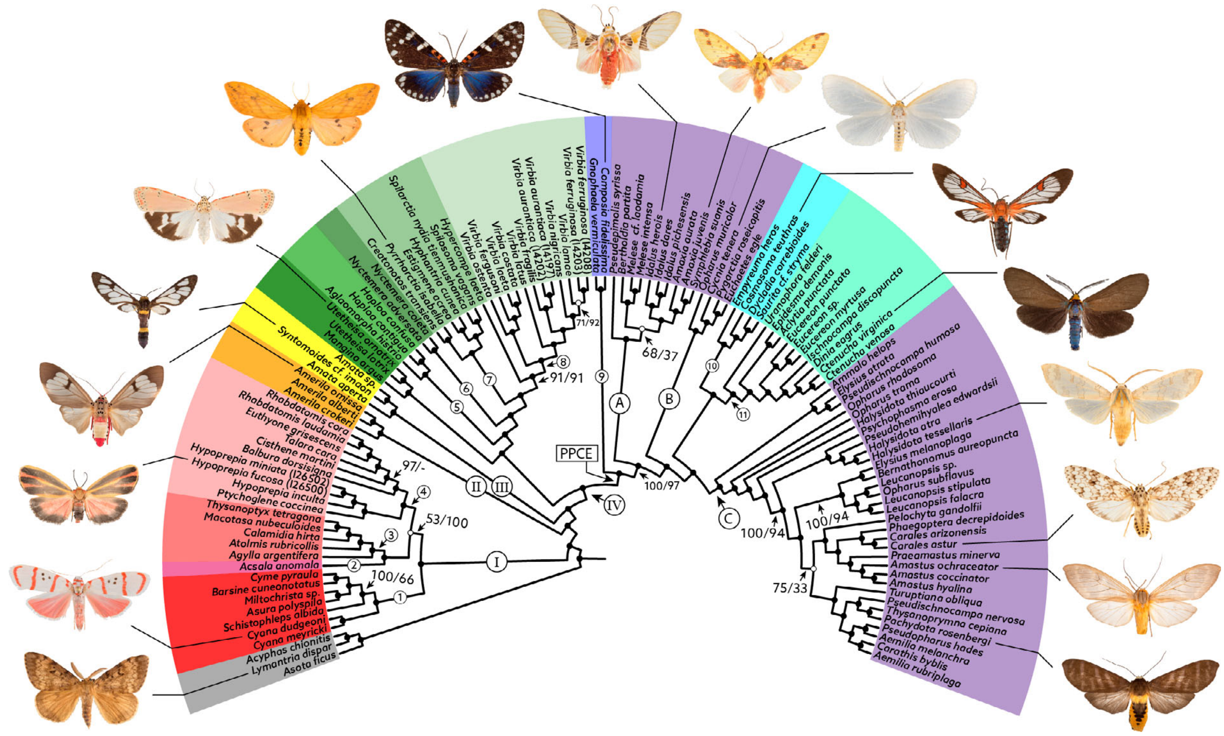

Fig. 3. Phylogenetic hypothesis for the subfamily Arctiinae (Noctuoidea: Erebidae) based on supermatrix analysis (RAxML), with outgroup taxa. Clades representing subtribes are colored. Outgroups are colored grey. Bootstrap branch support (BS) for RAxML and ASTRAL reconstructions are separated by '/' and their branches are indicated by arrows. Branches leading to nodes with black circles had >75% RAxML BS and those with white circles had ≤75% RAxML BS. Nodes without printed BS values have ≥98% RAxML and ASTRAL support. A '-' indicates that an alternative topology was recovered with ASTRAL. 'PPCE' denotes the location of the PPCE clade. 'A', 'B', and 'C' denote three subclades of the paraphyletic subtribe Phaegopterina. Only one representative of *Hypoprepia fucosa* and *H. miniata* was included for illustrative purposes. Tribes and subtribes are denoted with the following symbols. Tribes: (I) Lithosiini, (II) Amerilini, (III) Syntomini, (IV) Arctiini; Subtribes: (1) Nudarina, (2) Acsalina, (3) Lithosiina, (4) Cisthenina, (5) Callimorphina, (6) Nyctemerina, (7) Spilosomina, (8) Arctiina, (9) Pericopina, (10) Euchromiina and (11) Ctenuchina, (A–C) Phaegopterina. https://doi.org/10.6084/m9.figshare.9995330 [Colour figure can be viewed at wileyonlinelibrary.com].

analyses further divided *H. miniata* into two well-supported subclades: 'Subclade 1' comprised individuals derived from the upper midwestern and northeastern United States (see Figs. 4 and 5D) and 'Subclade 2' from the southeastern United States (see Figs. 4 and 5E), including one specimen ('*Hypoprepia sp.*') possessing a darker, diminutive phenotype that cannot confidently be assigned to a currently recognized species. These subgroupings reflect significant differences in external phenotype. Individuals in Subclade 1 are larger and red in colour, whereas individuals from Subclade 2 exhibit a diminutive and predominantly yellow phenotype. Our comparison of AHE-derived DNA barcode data with sequences available via BOLD (40 individuals across its geographic range; Ratnasingham & Hebert, 2007) clustered all individuals of *Hypoprepia* into either *H. fucosa* or *H. miniata* appropriately with good support (File S13). However, DNA barcode data alone lacked the power to resolve the strongly supported subclades within *H. miniata* recovered by the NOC1 probe set.

The *H. fucosa* clade ($N = 10$) was also well-supported in both our analyses. ASTRAL strongly supported a midwestern United States clade, including individuals classified into the subspecies *H. f. tricolor*, though support for relationships within this clade were weak (Figs. 4 and 5A,B). RAxML also tended recover this midwestern clade, apart from *H. fucosa* I26500 from Indiana, which was placed with samples from Louisiana, albeit with low support. Only RAxML placed the two specimens of *H. f. tricolor* as sister to each other with strong confidence.

Sensitivity analysis

At both deep and shallow levels, only a relatively small number (e.g., 50–175 loci) of the 730 loci included in the complete NOC1 probe set were necessary to resolve most relationships, though this was only true when the more rapidly evolving sites of the loci were included (e.g., 7–8 bins). The compositions of these loci subsets are given in File S14. In almost all cases, the inclusion of more variable sites (e.g., 8 bin site-binning strategy) had a strong effect on increasing BS, whereas including only lower-valued bins (i.e., more-conserved sites) often led to failure to resolve branches. The branch uniting the entire Arctiinae as well as the branch uniting Lithosiina and Acsalina were notable exceptions, exhibiting better BS with the 7-bin site-binning strategy, particularly with ASTRAL (Fig. 6).

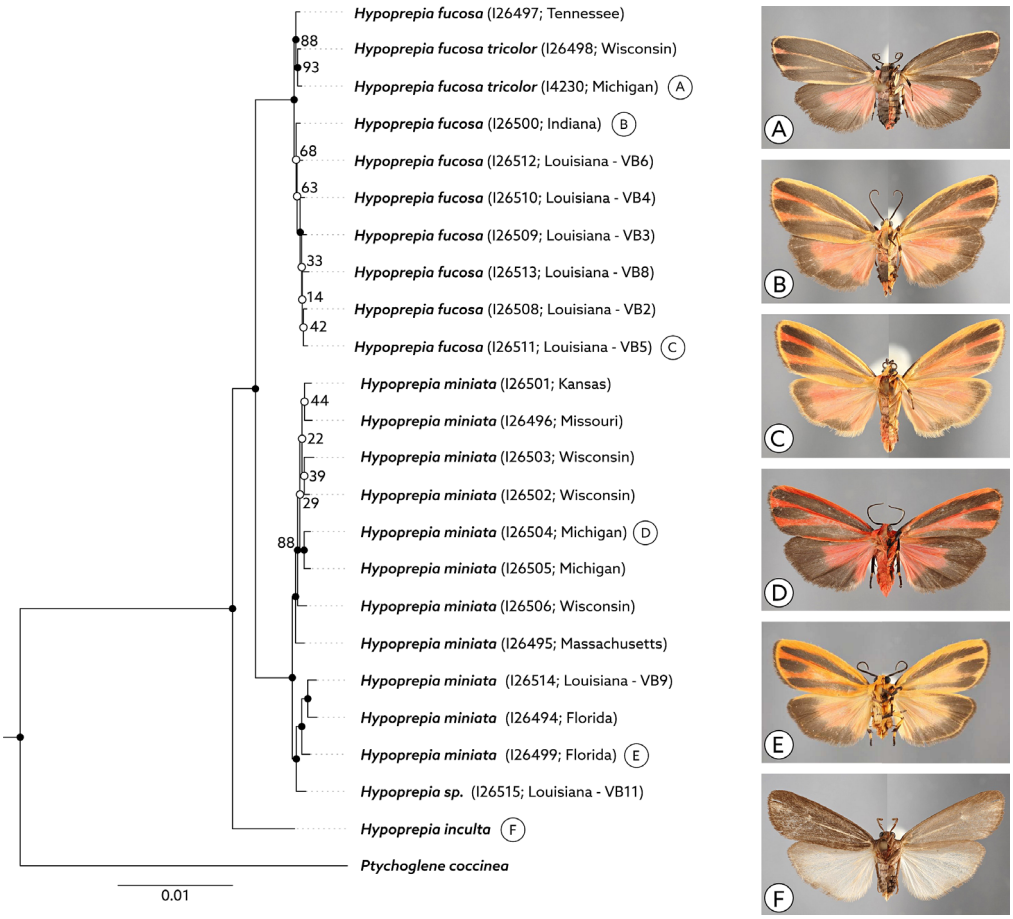

Fig. 4. RAxML species tree of the *Hypoprepia* species complex. The tree was estimated from a concatenated dataset of loci successfully enriched among the taxa. BS is indicated on each node (black circle: >75%; white circle: ≤75%). Nodes without printed BS values have ≥98% support. Dorsal (left) and ventral (right) images provided for common phenotypes (A–F). https://doi.org/10.6084/m9.figshare.9995333 [Colour figure can be viewed at wileyonlinelibrary.com].

Sensitivity analysis–subtribal-level

At the subtribal level, the overall trends between RAxML and ASTRAL were similar (Fig. 6). All but two subtribal nodes required only the 50 or 100 loci subset to receive strong support when analysed with either RAxML or ASTRAL. Phaegopterina Clade A required at least 225 loci to robustly place using ASTRAL, compared to only 100 loci using RAxML. The strongly supported placement of Acsalina as sister to Lithosiina required at least 175 loci using ASTRAL, compared to only 100 loci using RAxML.

The branch joining Cisthenina and (Lithosiina, Acsalina) is not well-supported for most loci and bin combinations (Fig. 6). Relative to RAxML, ASTRAL more strongly supported (Nudariina, [Cisthenina, (Lithosiina, Acsalina)]) (hereafter 'recovered topology') across all loci subsets. Under both methods, this arrangement was most strongly supported only when very few loci were included (50 loci subset) and when many loci (475 loci subset and full dataset) were included (File S15). Intermediately sized combinations of loci (i.e., 100–375 loci subsets), analysed with RAxML and ASTRAL, supported [(Cisthenina, Nudariina), (Lithosiina, Acsalina)] (hereafter 'alternative topology') more strongly than either analysis supported the recovered topology (File S15). This is despite the 53% and 100% bootstrap support for the recovered topology from the phylogenetic analyses utilizing the complete alignment of all recovered loci (Fig. 3, File S8).

We also found that intergeneric relationships with weak to moderate BS often exhibited either a plateau in BS as the number of loci included were increased (File S16A, B) or equally weak support at all bin and loci combinations (File S16C, D).

Sensitivity analysis–species-level

ASTRAL and RAxML produced very different topologies within the genus *Hypoprepia*, making direct comparison difficult. Where the two methods were similar, RAxML (File S17) produced less-conservative and less-consistent BS across binning strategies and loci subsets compared to ASTRAL

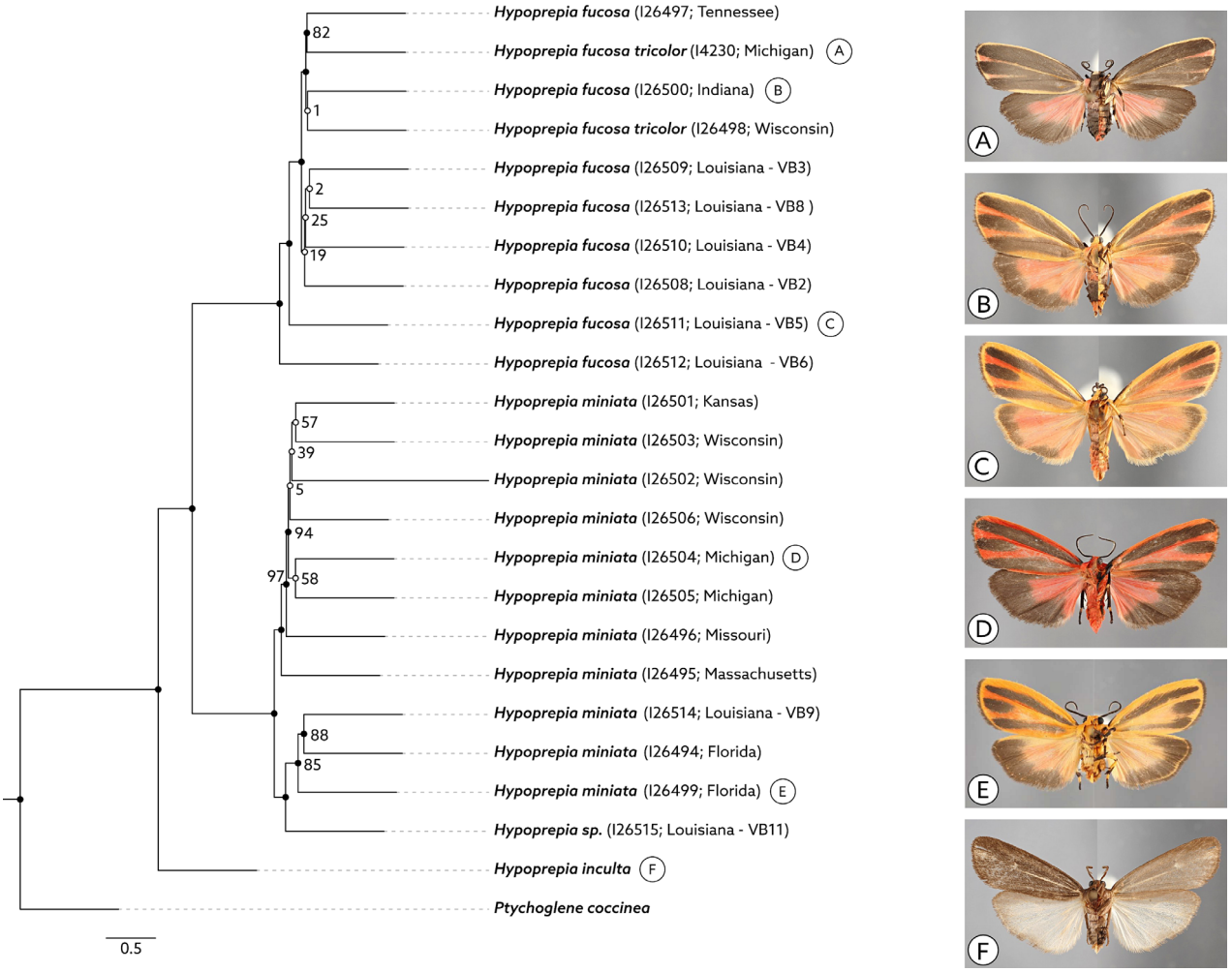

Fig. 5. ASTRAL species tree of the species complex. The tree was estimated from gene trees of loci successfully enriched among the taxa. BS is indicated on each node (black circle: >75%; white circle: ≤75%). Nodes without printed BS values have ≥98% support. Dorsal (left) and ventral (right) images provided for common phenotypes (A–F). https://doi.org/10.6084/m9.figshare.9995336 [Colour figure can be viewed at wileyonlinelibrary.com].

(File S18). This was also observed among species-relationships within the genus *Virbia* (File S19). Infraspecific relationships within *H. miniata* were more similar between ASTRAL and RAxML analyses than those within *H. fucosa*. With ASTRAL, as few as 175 loci strongly supported both *H. miniata* Subclade 1 and Subclade 2 as distinct and separable clades (File S18). RAxML also recovered this relationship but required the largest loci subset (475 loci) to do so confidently (File S17). Generally, fewer loci were needed to recover strong support values for the ASTRAL topology as compared with RAxML, which required more data to produce strong support values.

Internode certainty

The average QP-IC score for the concatenated maximum likelihood tree of all taxa was 0.329 (File S10). Quartet-based IC scores can be open to interpretation (Zhou *et al.*, 2019), making it difficult to objectively evaluate these scores, however, this result may be indicative of moderate levels of phylogenetic incongruence of the gene trees relative to the species tree. The only negative scores, indicating an alternative topology was more prevalent among the gene trees, corresponded to branches descended from the root and may be a technical artifact. Branches with BS < 98 always had low QP-IC scores (<0.01), indicating a high level of incongruence, whereas 46% (46/113) of highly supported branches (BS = 100) had QP-IC scores >0.5 suggesting low incongruence (File S20, File S9). The lack of highly negative values suggests there are no strongly supported alternative relationships incongruent to the species tree results. Instead, the gene trees may contain a moderate to weak degree of conflicting signal, but nonetheless the species tree remains highly supported.

The *Hypoprepia* species tree had moderate QP-IC scores for the deeper branches except for the split between *H. fucosa* and *H. miniata*, which had a score 0.080 (File S11 and File

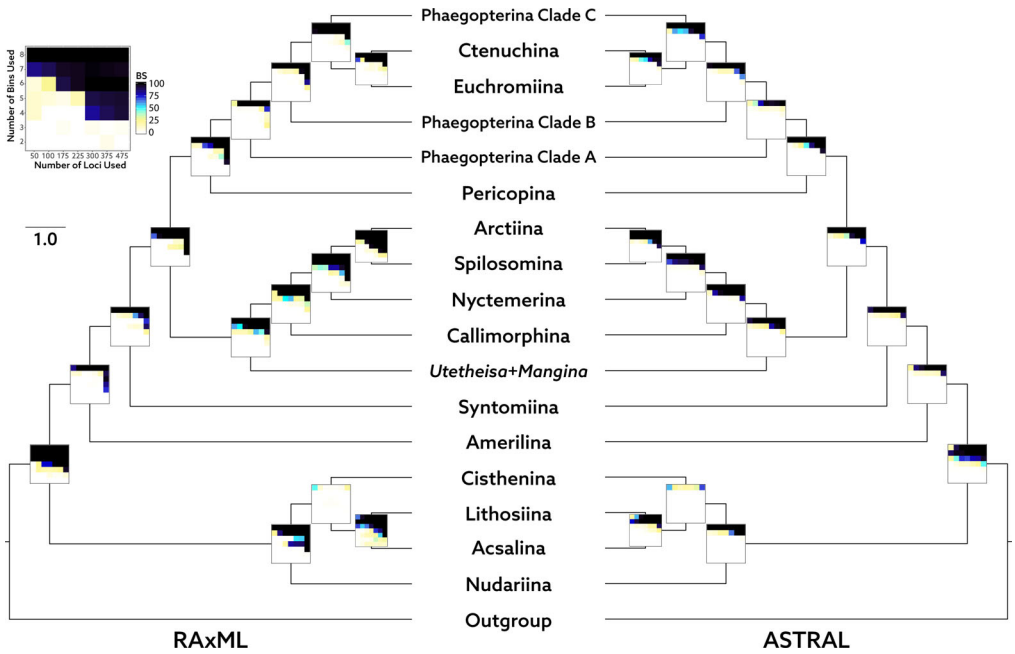

Fig. 6. NOC1 Subtribal Sensitivity Analysis. RAxML (left) and ASTRAL (right) topologies of subtribal relationships within the Arctiinae. Heatmaps indicate branch supports (color encoded; black = 100%, blue = 75%, teal = 50%, yellow = 25%, white = 0% BS) for a given clade for phylogenies constructed under different combinations of loci subsets (x-axis) and site-binning strategies (y-axis). https://doi.org/10.6084/m9.figshare.9995339 [Colour figure can be viewed at wileyonlinelibrary.com].

S12). Divergences between the intraspecific clades of *H. miniata* and *H. fucosa* had higher scores (0.310 for the initial *H. fucosa* divergence and 0.286 for the *H. miniata* divergence), but shallower intraspecific relationships within geographic areas had much lower scores. These incongruent phylogenetic signals may explain the different intraspecific relationships observed between the RAxML (Fig. 4, File S11) and ASTRAL trees (Fig. 5, File S12).

QP-IC scores varied across the 8-bin loci subsets, though nearly 80% of branches (95/119) received higher QP-IC as more genes were added, with the 300 loci subset yielding the highest score for 56% of branches (67/119) (File S20). Branches with BS < 98 had lower average QP-IC across the loci subsets (QP-IC < 0.1), and branches with poor support (BS < 75) tended to switch between negative and positive QP-IC, indicating that the support among the gene trees shifted between incongruent topologies for that particular relationship. There was no clear difference in QP-IC between deeper- and shallower-level relationships.

Discussion

Topological performance

Recent efforts to confidently resolve relationships within the Arctiinae utilizing a genetic marker set of eight genes, even when paired with a massive taxonomic sampling of nearly 300 species, have not been sufficient (Zenker *et al.*, 2016). Our results indicate that the NOC1 probe set is well-suited to producing robust phylogenetic hypotheses, even in a group which has challenged traditional methodologies. Unlike traditional molecular markers, our probe design robustly resolves relationships at each taxonomic level we sampled. However, to address the evolutionary relationships within extremely diverse groups like the Arctiinae, a strong taxonomic sampling scheme for a given hypothesis is still a requirement.

Our NOC1 probe set recovered results at deep taxonomic levels (e.g., tribal, subtribal) that are largely congruent with the results of previous molecular studies (Zahiri *et al.*, 2011; Zahiri *et al.*, 2012; Zaspel *et al.*, 2014; Zenker *et al.*, 2016; Scott Chialvo *et al.*, 2018). However, we also recovered novel and well-supported relationships within the Arctiinae which have eluded previous studies utilizing traditional markers, placing arctiine taxonomy, particularly at the subtribal level, in a potential state of taxonomic flux (Zaspel *et al.*, 2014; Rönkä *et al.*, 2016; Zenker *et al.*, 2016). Due to inadequate sampling, we cannot make definitive recommendations to resolve these issues here, but we identify a few trends from this work. As indicated in many previous studies, the subtribe Phaegopterina is not a monophyletic group (Jacobson & Weller, 2002; Zaspel *et al.*, 2014; Zenker *et al.*, 2016). To resolve the paraphyly of Phaegopterina with respect to Ctenuchina and Euchromiina, we suggest that the group may need to be split into at least three clades (Fig. 3; Subclades 'A', 'B', 'C'). These new groups resemble the historical 'generic affinities' within Phaegopterina: a '*Eupseudosoma* group' (likely coincident with Subclade A), a

'*Euchaetes* group' (coincident with Subclade B) and a '*Halysidota* group' (coincident with Subclade C) (Forbes, 1960). Prior molecular studies hinted at a paraphyletic Phaegopterina as well, but with much weaker support compared to our results (Zaspel *et al.*, 2014; Zenker *et al.*, 2016). However, we retain 'Phaegopterina' in its traditional sense until more taxa are sampled and diagnostic synapomorphies of the resulting monophyletic groups can be identified. Our results strongly suggest that Ctenuchina and Euchromiina are monophyletic and sister subtribes. Previous support for this finding has varied from weak (BS = 30% and 41%, respectively; Zenker *et al.*, 2016; Posterior probability = 0.83 and 0.86, respectively; Simmons *et al.*, 2012) to relatively strong (BS = 98% and 75%, respectively; Zaspel *et al.*, 2014). Despite a somewhat different taxon sampling scheme (50% and 30% generic-level overlap, respectively), our results within Lithosiini are consistent with recent work based on traditional markers as well as transcriptome data, at least within subtribes (Zaspel *et al.*, 2014; Scott Chialvo *et al.*, 2018). The relationships among subtribes within the Lithosiini were markedly different, however, particularly compared to results based on transcriptomic evidence. This transcriptomic data moderately supported (BS = 87%) the topology recovered in the sensitivity analysis utilizing the intermediately sized loci subsets (i.e., 175–375 loci) of the NOC1 probe set [i.e., (Nudariina, Cisthenina), (Lithosiina)], rather than the topology we recovered utilizing the full probe set {i.e., [Nudariina, (Cisthenina, Lithosiina)]} (Scott Chialvo *et al.*, 2018). The transcriptome-based study sampled a greater diversity of Lithosiini, indicating that a denser sampling of taxa and/or more loci with more conserved sites or less gene tree discordance may be necessary to help clarify the relationships among subtribes within Lithosiini.

Our probe set appears well-suited to resolving intergeneric relationships within subtribes. Nearly all generic relationships were fully resolved and strongly supported using NOC1 by both RAxML and ASTRAL approaches, with only a few exceptions. This is notable as only about 40–55% of intrageneric relationships from prior studies based on traditional markers obtained moderate to good statistical support (BS > 75%) (Zaspel *et al.*, 2014; Rönkä *et al.*, 2016; Zenker *et al.*, 2016). However, troublesome taxonomic assignments may be lurking where taxonomy and topology disagree. We have discovered several polyphyletic 'trashcan' genera that require future revisionary work. These genera are *Elysius*, *Opharus*, *Aemilia*, *Halysidota* and *Pseudischnocampa*. Some of these genera have historically challenged taxonomists. *Opharus* has previously been recommended for revisionary studies and several species within *Aemilia* and *Halysidota* have been classified as '*sensu lato*', indicating uncertainty in their generic assignments (Watson & Goodger, 1986).

The generic compositions of the various subtribes recovered here are generally consistent with previous genetic studies; however, some differences exist (Zaspel *et al.*, 2014; Zenker *et al.*, 2016). Within the Callimorphina, the genus *Utetheisa* has traditionally been difficult to place. The genus has been recovered within the Callimorphina, within the PPCE clade, or within a clade sister to the Arctiini depending on the data utilized and analysis performed (Zaspel *et al.*, 2014; Zenker *et al.*, 2016). Our results strongly support a novel topology of *Utetheisa* as sister to the clade containing the Callimorphina, Nyctemerina, Spilosomina and Arctiina. Our results indicate that the genus *Virbia* (currently Arctiina) likely requires reclassification as a member of Spilosomina (e.g., Ferguson, 1985; Lafontaine & Schmidt, 2010; Vincent & Laguerre, 2014). Reassignment of *Virbia* into Spilosomina is also supported by previous genetic studies (Zaspel *et al.*, 2014; Rönkä *et al.*, 2016; Zenker *et al.*, 2016). Under this scenario, the placement of the subtribe Arctiina remains uncertain, as no other putative members of Arctiina were included in this study.

The strong performance of the probe set at the shallowest taxonomic levels was surprising given that AHE-based loci are designed to be highly conserved, though one other study in squamates also found AHE to be capable of resolving intraspecific relationships (Brandley *et al.*, 2015). NOC1 recovered well-supported relationships among the three species and one subspecies of *Hypoprepia* included, as well as most of the relationships among 10 species within the genus *Virbia*. Under RAxML and ASTRAL analyses, species and intraspecific relationships were captured with strong statistical support in many cases. ASTRAL produced results that were generally as well-supported as RAxML, though the topologies were different in a few cases. Gene tree discordance due either to incomplete lineage sorting (ILS) or hybrid introgression can create conflict between individual gene trees and the species tree. Modelling gene tree incongruence is known to be more statistically consistent under coalescence-based species tree estimation (e.g., ASTRAL) compared to concatenation-based analyses (e.g., RAxML) (Rannala & Yang, 2003). Notably, however, RAxML clustered samples of *H. f. tricolor* together, whereas ASTRAL and COI did not. It is not clear which scenario is more likely, since the description of *H. f. tricolor* was based only on their dark phenotype. Major differences in genitalia or other characteristics are yet to be clearly outlined in support of this subspecies and the possibility that *H. f. tricolor* simply represents a dark, northern form of *H. fucosa* cannot be ruled out. NOC1 provided more robust intraspecific relationships than those provided by COI alone, although at a higher sequencing cost. Our results suggest that DNA barcoding based on only a single, short marker (i.e., COI, ~650 bp) are in some cases insufficient for resolving shallow-level divergences, potentially obfuscating cases of speciation.

We believe additional taxonomic sampling will likely help resolve the few remaining clades whose placement remains uncertain, as our sampling here was relatively sparse for certain taxonomic groups. For example, we included six species in five genera within the phaegopterine 'Subclade A', but we estimate this group contains approximately 800 species in 80 genera (i.e., <1% species and <6.5% generic coverage). The largest gaps exist within the Syntomini, the Lithosiini and the clade comprised *Utetheisa* + *Mangina*, Callimorphina, Arctiina and Spilosomina. Greater efforts should also be taken to encompass the diversity of these groups as well as taxa from under-studied geographic regions. Such sampling will ensure that our knowledge about arctiine relationships is not biased towards commonly encountered or commonly sampled groups.

Sensitivity analysis

As the cost of DNA sequencing has decreased, researchers have used ever-larger DNA data sets to estimate species trees. Having an adequate number of characters from which to build phylogenies is important, but our results confirm that there can be diminishing returns on the phylogenetic resolution per DNA base pair, not all loci are equally informative, and not all sites within a locus contribute equally to phylogenetic signal. Our sensitivity analysis indicated that only 50–175 loci of the 730 loci included in the NOC1 probe set were necessary to resolve most relationships within the Arctiinae. In most cases, the 8-bin binning strategy produced the most statistically robust results. Increasing the proportion of less conservative sites within each locus (i.e., moving from lower-valued to higher-valued binning strategies) tended to increase BS. This indicates that the faster-evolving sites, possibly contained within the more variable regions flanking the targeted probe regions, are providing the necessary signal to resolve relationships rather than creating phylogenetic conflict via site saturation (e.g., see Breinholt *et al.*, 2018 and St. Laurent *et al.*, 2018 for discussion of flanking regions).

Generally, fewer loci were needed at deeper taxonomic levels, whereas more were needed at very shallow levels or in clades, which had relatively sparse taxonomic sampling. The majority of subtribal relationships were recovered with strong statistical support with the smallest loci subset (i.e., 50 loci) by including more rapidly evolving sites (i.e., 8-bin strategy). Among the subtribes of the Lithosiini, the sensitivity analysis results suggest that either a large number of loci are necessary to resolve subtribal relationships and/or some loci may be contributing false information. Based on our analysis of QP-IC, it seems likely that gene tree discordance is playing a role in the uncertainty of this node, particularly since all loci subsets except for the 50 loci set were found to have QP-IC scores that favoured an alternative topology joining Cisthenina and Nudarina as sister groups. The generally low to moderate QP-IC scores throughout the tree indicate some degree of conflicting phylogenetic signal exists within the data set, but so long as enough phylogenetic signal is present, most of the relationships remained highly supported. Indeed, many phylogenomic data sets report high levels of incongruency among gene tree (Salichos & Rokas, 2013; Jarvis *et al.*, 2014), but nonetheless manage to produce well-supported species trees. Gene tree discordance appears to a contributing factor for our inability to resolve other relationships as well, such as the divergence between *Virbia lamae* and *V. ferruginosa* or between the *Bertholdia* + *Melese* clade and the *Idalus* + *Symphlebia* + *Amaxia* clade. Increased taxonomic sampling might be one way to overcome this. By breaking up long-branch attraction with more thorough sampling, this gene tree discordance may be reduced. While the sensitivity analysis returned high support for most relationships when 175 loci were used, the QP-IC scores suggested incongruence was lowest when 300 loci were used.

The few remaining poorly supported relationships at more shallow levels exhibited a BS plateau or weak support at all bin and loci combinations (File S16). This may be an additional signature of poor taxonomic sampling, as increasing the number of loci and increasing the proportion of more quickly evolving sites did not affect BS in these cases. The species-level phylogeny of *Hypoprepia*, demonstrates the effectiveness of the NOC1 probe set at robustly uncovering potential cases of cryptic or incipient speciation with as few as 175 loci. Despite this, we retain the members of *H. miniata* Subclade 2 as *H. miniata* and our unnamed species of *Hypoprepia* as '*Hypoprepia* sp.' until diagnostic morphological characters are identified.

These findings demonstrate that the selection of AHE probes can be reduced in size and/or fine-tuned for a specific array of taxa or research question. This is significant, as future studies can use tailored probe sets to sequence smaller DNA datasets from more taxa, while obtaining topologies and statistical support similar to those derived from larger DNA datasets. This 'tailored probe set' approach could reduce the sequencing cost per individual, rendering phylogenetic studies based on resources like NOC1 more economical and require less computational time. It should be possible for researchers working with large probe sets in other systems to perform a similar analysis and better optimize the trade-off between sequencing and sampling.

Here we have used one of the most diverse subfamilies within Noctuoidea to demonstrate the effectiveness of NOC1 to resolve both deep and shallow relationships within the lineage. Results from the Lep1 probe set indicate that nucleotide completeness and total number of captured loci is higher for taxa that are more closely related to the focal taxa used to develop the probe sets (Fig. 3 from Breinholt *et al.*, 2018). Producing augmented probe sets that are focused on a lineage of interest (e.g., BUTTERFLY1.0, BUTTERFLY 1.1 and BOM1 derived from the more general Lep1) more efficiently recovers phylogenetically informative loci and/or loci capture efficiency (Espeland *et al.*, 2018; Toussaint *et al.*, 2018; Hamilton *et al.*, 2019). Comparing the application of Lep1 within the Erebinae (Noctuoidea: Erebidae) to our results applying NOC1 to the Arctiinae, we recovered more loci (730 vs 658), longer mean locus length (575 bp vs 320 bp) and a longer concatenated total length (420 574 bp vs 210 484 bp) (Homziak *et al.*, 2018). While we have not explicitly tested the effectiveness of NOC1 across the entire Noctuoidea, NOC1 may also perform well within other noctuoid lineages, generating new insights into the evolution of a lineage which comprises >25% of all Lepidoptera (van Nieukerken *et al.*, 2011), includes major agricultural pests, and serves critical functions in terrestrial ecosystems such as herbivory, pollination, and as a food source for many predators (Mitchell *et al.*, 2006; Zahiri *et al.*, 2011).

Conclusion

Our results demonstrate that NOC1 can be used to robustly address taxonomic and evolutionary questions at multiple divergence levels using readily available tissues preserved using standard methods. This study is the first to construct a robust phylogenetic hypothesis of relationships within the Arctiinae at all taxonomic levels. This was accomplished through the

generation of a conserved probe set of 730 loci derived from AHE methods, including utilization of more variable sites that may be derived from the probes' less-conserved flanking regions. We also demonstrate that tailored probe sets composed of fewer probes could be produced, allowing future studies to incorporate more taxa at a lower per-taxon cost without significant reduction in phylogenetic statistical support. These results will help place the convolution of chemical, acoustic and other behavioural adaptations exhibited by this unique and diverse group into an evolutionary framework with implications for resolving long-standing taxonomic quagmires within the Arctiinae.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

File S1. Taxon Sampling Table. The first sheet contains information about the taxa used for Anchored Hybrid Enrichment, including collection localities and read recovery for each taxon. The second sheet contains information about the samples used in the analysis of COI, including information for the samples obtained from GenBank. https://doi.org/10.6084/m9.figshare.9994646.

File S2. Probe Design Resources. Contains a table summarizing the genomics and transcriptomic resources that were used to produce the novel probes within the NOC1 probe set. https://doi.org/10.6084/m9.figshare.11987523.

File S3. Probe Set Details. Contains information about which probes were recovered from each taxon ('Probe Recovery Profile'), the number of taxa recovered by each probe ('Probe Recovery Summary Stats'), and various probe statistics ('Probe Statistics'; e.g., length, GC content). https://doi.org/10.6084/m9.figshare.9994652.

File S4. Individual Loci Alignments. Individual alignments of each of the 729 recovered loci in the NOC1 probe kit in PHYLIP format. https://doi.org/10.6084/m9.figshare.9994685.

File S5. Complete DNA Alignment. Contains the complete alignment of nucleotide data for all taxa and all loci used to construct phylogenies in PHYLIP format. https://doi.org/10.6084/m9.figshare.9994691.

File S6. NEXUS Tree Files. Contains all the phylogenies constructed in this study in NEXUS format. https://doi.org/10.6084/m9.figshare.9994703.

File S7. ASTRAL Wrapper Script. Python script used to parallelize ASTRAL-III, allowing for multiple concurrent runs. https://doi.org/10.6084/m9.figshare.9994715.

File S8. Phylogenetic hypothesis for the subfamily Arctiinae (Noctuoidea: Erebidae) based on ASTRAL analysis, with outgroup taxa. Clades representing subtribes are coloured. Outgroups are coloured grey. Bootstrap branch support (BS) for RAxML and ASTRAL reconstructions are separated by '/' and their branches are indicated by arrows. Branches leading to nodes with black circles had >75% ASTRAL BS and those with white circles had ≤75% ASTRAL BS. Nodes without printed BS values have ≥98% RAxML and ASTRAL support. A '-' indicates that an alternative topology was recovered with RAxML. 'PPCE' denotes the PPCE clade. 'A', 'B', and 'C' denote three subclades of the paraphyletic subtribe Phaegopterina. Only one representative of *Hypoprepia fucosa* and *H. miniata* was included for illustrative purposes. Tribe and subtribe symbols follow Fig. 3. https://doi.org/10.6084/m9.figshare.9994718.

File S9. Labeled edges of RAxML topology. ID numbers of each edge of the RAxML topology are given to facilitate referencing particular branches. Values correspond to the edge numbers given in File S20. https://doi.org/10.6084/m9.figshare.9994790.

File S10. Internode Certainty for the phylogenetic hypothesis of the subfamily Arctiinae (Noctuoidea: Erebidae) based on supermatrix analysis (RAxML). Quadripartition internode certainty (QP-IC) was calculated for each internal branch using the program QuartetScores utilizing the gene trees as input. A high QP-IC score (close to 1) suggests a specific branch is present in high frequency among the gene tree set, whereas a low score (closer to 0) indicates a lower frequency of that branch. A negative score indicates an alternative, incongruent relationship is found at higher frequency among the gene trees compared to the branch present in the reference tree. Bootstrap values and QP-IC scores are annotated on each internal branch. Branch lengths are proportional to number of nucleotide substitutions. https://doi.org/10.6084/m9.figshare.9994724.

File S11. Internode Certainty for *Hypoprepia* RAxML tree. Quadripartition internode certainty (QP-IC) was calculated for each internal branch using the program QuartetScores utilizing the gene trees as input. A high QP-IC score (close to 1) suggests a specific branch is present in high frequency among the gene tree set, whereas a low score (closer to 0) indicates a lower frequency of that branch. A negative score indicates an alternative, incongruent relationship is found at higher frequency among the gene trees compared to the branch present in the reference tree. Bootstrap values and QP-IC scores are annotated on each internal branch. Branch lengths are proportional to number of nucleotide substitutions. https://doi.org/10.6084/m9.figshare.9994733.

File S12. Internode Certainty for *Hypoprepia* ASTRAL tree. Quadripartition internode certainty (QP-IC) was calculated

for each internal branch using the program QuartetScores utilizing the gene trees as input. A high QP-IC score (close to 1) suggests a specific branch is present in high frequency among the gene tree set, while a low score (closer to 0) indicates a lower frequency of that branch. A negative score indicates an alternative, incongruent relationship is found at higher frequency among the gene trees compared to the branch present in the reference tree. Bootstrap values and QP-IC scores are annotated on each internal branch. Branch lengths are proportional coalescent except for external branches, which have arbitrary values of 0.1. https://doi.org/10.6084/m9.figshare.9994742.

File S13. DNA barcode tree of *Hypoprepia*. CO1 barcodes (~700 bp) were extracted from our AHE data (red labels) and combined with BOLD barcodes (black labels) from across the geographic range of *Hypoprepia*. The BOLD ID or our internal ID as well as abbreviated locality is given in parentheses. BS is indicated on each node (black circle: ≥10%; white circle: <10%). Nodes without printed BS values have <10% support. Branch lengths are proportional to number of nucleotide substitutions. Samples I26511 and I26496 were excluded due to insufficient recovery of the CO1 region from the AHE data. https://doi.org/10.6084/m9.figshare.9994745.

File S14. Reduced Probe Sets. Contains the reduced probe sets for the 8-bin binning strategy which produced well-supported phylogenies (i.e., 50 loci, 100 loci, 175 loci, 225 loci) ('8bin_reduced_probe_sets'), as well as all reduced sets (i.e., 50, 100, 175, 225, 300, 375, 475 loci) for all binning strategies (2–8 bins) ('all_sensitivity_analysis_sets'). https://doi.org/10.6084/m9.figshare.9994748.

File S15. Sensitivity Analysis Comparing Two Alternative Topologies of Lithosiini. Two alternative topologies for subtribal relationships within the Lithosiini are presented. Heatmaps follow Fig. 6. Both RAxML (left) and ASTRAL (right) analyses favored the 'main topology' (upper tree; Fig. 3, File S8) only in the smallest (50) and largest (475) loci subsets used in this analysis. The 'alternative topology' (lower tree) is favored in analyses utilizing the intermediate probe subset sizes (100–375 loci). https://doi.org/10.6084/m9.figshare.9994751.

File S16. Sensitivity Analysis Comparing Performance of RAxML and ASTRAL among clades with low BS values. Sensitivity analysis results from RAxML and ASTRAL are compared for clades containing branches with low BS. Heatmaps follow Fig. 6. As the number of loci increased (increasing x-axis), the BS either plateaued (A, B) or were poorly supported in most combinations of site-binning strategies and loci subsets (C, D). https://doi.org/10.6084/m9.figshare.9994757.

File S17. Sensitivity Analysis of RAxML within the genus *Hypoprepia*. Strong support exists for both *H. miniata* Subclade 1 and Subclade 2 as distinct and separable clades, though the recovery of a monophyletic *H. miniata* Subclade 1 required the largest loci subset (475 loci) to do so confidently. Heatmaps follow Fig. 6. https://doi.org/10.6084/m9.figshare.9994760.

File S18. Sensitivity Analysis of ASTRAL within the genus *Hypoprepia*. As few as 175 loci strongly support both *H. miniata* Subclade 1 and Subclade 2 as distinct and separable clades. Heatmaps follow Fig. 6. https://doi.org/10.6084/m9.figshare.9994763.

File S19. Sensitivity Analysis Comparing Performance of RAxML and ASTRAL within the genus *Virbia*. Sensitivity analysis results from RAxML and ASTRAL are compared for a subset of the genus *Virbia*. Heatmaps follow Fig. 6. ASTRAL produced BS values that were much more consistent, albeit more conservative, across site-binning strategies and loci subsets. https://doi.org/10.6084/m9.figshare.9994781.

File S20. QP-IC Scores Table. Contains the QP-IC scores of branches for each locus combination of the 8-bin binning strategy, as well as the complete data set. The most consistent locus is that which produced the largest absolute QP-IC across all loci. Tips do not have QP-IC scores. See File S9 for a mapping of edge numbers to the RAxML phylogeny. https://doi.org/10.6084/m9.figshare.9994784.

Acknowledgements

The authors would like to thank Michelle Kortyna and Sean Holland at the Center for Anchored Phylogenomics for assistance with data collection and analysis. We thank Akito Kawahara and the 1KITE consortium for allowing access to genomic resources for the probe design. We would also like to thank Dr. Santiago F. Burneo at Pontificia Universidad Católica del Ecuador for his assistance in acquiring permits for fieldwork in Ecuador and field assistants. Drs. Thomas Walla and Lee Dyer were also instrumental in obtaining some of the material used in this study. We are grateful to Andrea Vargas and Andrea Vallejo for their invaluable field assistance collecting material. We greatly appreciate Dr. Harold Greeney and José Simbaña of the Yanayacu Biological Station for providing a fantastic environment for fieldwork. Assistance with fieldwork and specimen acquisition was graciously provided by Vladimir Kononenko, Vernon Brou Jr., Ring Cardé, James Adams, and Reza Zahiri. Finally, we would like to thank Jason Ksepka for the image used in Fig. 2A, Fritz Flohr Reynolds for the image used in Fig. 2B, André Poremski for assistance capturing images used in Figs 1 and 3, and File S8, and Timothy J. Anderson for images used in Figs. 4 and 5. This work was supported by an NSF grant (DEB-0919185) to J. M. Zaspel and S. J. Weller, a National Geographic Exploration Grant to

J. M. Zaspel, A. R. Lemmon, and E. M. Lemmon, an NSF CSBR (DBI-1561448) grant to J. M. Zaspel, an NSF grant (IOS-0951160) to W. E. Conner and an NSF PRFB grant (DBI-1811897) to N. J. Dowdy. We would also like to acknowledge a Wake Forest University Pilot Research grant to W. E. Conner and Purdue University College of Agriculture startup funds to J. M. Zaspel. The authors declare no conflicts of interest.

REFERENCES

Adamowicz, S.J. (2015) International barcode of life: evolution of a global research community. *Genome*, **58**, 151–162.

Barber, J.R., Chadwell, B.A., Garrett, N., Schmidt-French, B. & Conner, W.E. (2009) Naive bats discriminate arctiid moth warning sounds but generalize their aposematic meaning. *The Journal of Experimental Biology*, **212**, 2141–2148.

Bazinet, A.L., Cummings, M.P., Mitter, K.T. & Mitter, C.W. (2013) Can RNA-seq resole the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. *PLoS One*, **8**, e82615.

Bazinet, A.L., Mitter, K.T., Davis, D.R., Van Nieukerken, E.J., Cummings, M.P. & Mitter, C. (2017) Phylotranscriptomics resolves ancient divergences in the Lepidoptera. *Systematic Entomology*, **42**, 82–93.

Blaimer, B.B., Lloyd, M.W., Guillory, W.X. & Brady, S.G. (2016) Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS One*, **11**, e0161531.

Blest, A.D., Collett, T.S. & Pye, J.D. (1963) The generation of ultrasonic signals by a New World Arctiid Moth. *Proceedings of the Royal Society of London B: Biological Sciences*, **158**, 196–207.

Bogdanowicz, D., Giaro, K. & Wróbel, B. (2012) TreeCmp: comparison of trees in polynomial time. *Evolutionary Bioinformatics*, **8**, S11657.

Boppré, M. (1981) Adult Lepidoptera 'feeding' at withered *Heliotropium* plants (Boraginaceae) in East Africa. *Ecological Entomology*, **6**, 449–452.

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, **4**, e1660.

Brandley, M.C., Bragg, J.G., Singhal, S. *et al.* (2015) Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian Eugongylus group scincid lizards. *BMC Evolutionary Biology*, **15**, 62.

Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L. & Kawahara, A.Y. (2018) Resolving relationships among the megadiverse butterflies and with a novel pipeline for anchored phylogenomics. *Systematic Biology*, **67**, 78–93.

Buddenhagen, C., Lemmon, A.R., Lemmon, E.M. *et al.* (2016) Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *BioRxiv*, 086298. https://doi.org/10.1101/086298

Cock, P.J., Antao, T., Chang, J.T. *et al.* (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.

Conner, W.E. (1987) Ultrasound: its role in the courtship of the arctiid moth, *Cycnia tenera*. *Experientia*, **43**, 1029–1031.

Conner, W.E. (2009) *Tiger Moths and Woolly Bears: Behavior, Ecology, and Evolution of the Arctiidae*, p. 303. Oxford University Press, New York.

Conner, W.E. & Corcoran, A.J. (2012) Sound strategies: the 65-million-year-old Battle between bats and insects. *Annual Review of Entomology*, **57**, 21–39.

Conner W.E., & Jordan A.T. 2009. From armaments to ornaments: the relationship between chemical defense and sex in tiger moths. Tiger Moths and Woolly Bears Behavior, Ecology, and Evolution of the Arctiidae (ed. by Conner W.E.). New York: Oxford University Press. pp. 155–172.

Conner, W.E., Boada, R., Schroeder, F. & Eisner, T. (2000) Chemical defense: bestowal of a nuptial alkaloidal garment by a male moth on its mate. *Proceedings of the National Academy of Sciences*, **97**, 14406–14411.

Corcoran, A.J. & Conner, W.E. (2012) Sonar jamming in the field: effectiveness and behavior of a unique prey defense. *The Journal of Experimental Biology*, **215**, 4278–4287.

Corcoran, A.J., Barber, J.R. & Conner, W.E. (2009) Tiger moth jams bat sonar. *Science*, **325**, 325–327.

Critchlow, D.E., Dennis, K.P. & Qian, C. (1996) The triples distance for rooted bifurcating phylogenetic trees. *Systematic Biology*, **45**, 323–334.

Cummins, C.A. & McInerney, J.O. (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. *Systematic Biology*, **60**, 833–844.

DaCosta, M.A. & Weller, S.J. (2005) Phylogeny and classification of Callimorphini (Lepidoptera: Arctiidae: Arctiinae). *Zootaxa*, **1025**, 1–94.

DaCosta, M.A., Larson, P., Donahue, J.P. & Weller, S.J. (2006) Phylogeny of milkweed tussocks (Arctiidae: Arctiinae: Phaegopterini) and its implications of evolution of ultrasound communication. *Annals of the Entomological Society of America*, **99**, 723–742.

Dowdy, N.J. & Conner, W.E. (2016) Acoustic aposematic and evasive action in select chemically defended Arctiine (Lepidoptera: Erebidae) species: nonchalant or not? *PLoS One*, **11**, e0152981.

Dunning, D.C. (1967) Warning sounds of moths. *Zeitschrift fur Tierpsychologie*, **25**, 9–138.

Edwards, S.V. (2016) Phylogenomic subsampling: a brief review. *Zoologica Scripta*, **45**(S1), 63–74.

Espeland, M., Breinholt, J., Willmott, K.R. *et al.* (2018) A comprehensive and dated Phylogenomic analysis of butterflies. *Current Biology*, **28**, 770–778. https://doi.org/10.1016/j.cub.2018.01.061.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.

Ferguson, D.C. (1985) Contributions toward reclassification of the world genera of the tribe Arctiini, part 1: introduction and a revision of the *Neoarctia-Grammia* group (Lepidoptera: Arctiidae: Arctiinae). *Entomography*, **3**, 181–275.

Fernández, R., Edgecombe, G.D. & Giribet, G. (2018) Phylogenomics illuminates the backbone of the Myriapoda tree of life and reconciles morphological and molecular phylogenies. *Scientific Reports*, **8**, 83. https://doi.org/10.1038/S51598-017-18562-w.

Forbes, W.T.M. (1960) *Lepidoptera of New York and Neighboring States Part IV*, *Ithaca, NY: Cornell University Agricultural Experiment Station*, *Memoir*, **371**, 1-188.

Fullard, J.H., Fenton, M.B. & Simmons, J.A. (1979) Jamming bat echolocation: the clicks of arctiid moths. *Canadian Journal of Zoology*, **57**, 647–649.

Garrison, N.L., Rodriguez, J., Agnarsson, I. *et al.* (2016) Spider phylogenomics: untangling the spider tree of life. *PeerJ*, **4**, e1719.

Geneious version R9 created by Biomatters. n.d. (2015) URL http://www.geneious.com/.

Gnirke, A., Melnikov, A., Maguire, J. *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.

Hahn, C., Bachmann, L. & Chevreux, B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129–e129.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M. & Bond, J.E. (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, **16**, 212.

Hamilton, C.A., Laurent, R.A.S., Dexter, K. *et al.* (2019) Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives. *BMC Evolutionary Biology*, **19**, 182.

Hedin, M., Starrett, J., Akhter, S., Schönhofer, A.L. & Shultz, J.W. (2012) Phylogenomic resolution of Paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLoS One*, **7**, e42888.

Hittinger, C.T., Johnston, M., Tossberg, J.T. & Rokas, A. (2010) Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences*, **107**, 1476–1481. https://doi.org/10.1073/pnas.0910449107.

Hodges, E., Xuan, Z., Balija, V. *et al.* (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.

Homziak, N.T., Breinholt, J.W., Branham, M.A., Storer, C.G. & Kawahara, A.Y. (2018) Anchored hybrid enrichment phylogenomics resolves the backbone of erebine moths. *Molecular Phylogenetics and Evolution*, **131**, 99–105.

Howard, A.F. & Barrows, E.M. (2014) Self-pollination rate and floral-display size in *Asclepias syriaca* (common milkweed) with regard to floral-visitor taxa. *BMC Evolutionary Biology*, **14**, 144.

Jacobson, N.L. & Weller, S.J. (2002) *A Cladistic Study of the Arctiidae (Lepidoptera) by Using Characters of Immatures and Adults. Thomas Say Publications in Entomology*, p. 98. Entomological Society of America, Lanham.

Jarvis, E.D., Mirarab, S., Aberer, A.J. *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Johns, C.A., Toussaint, E.F.A., Breinholt, J.W. & Kawahara, A.Y. (2018) Origin and macroevolution of micro-moths on sunken Hawaiian islands. *Proceedings of the Royal Society B*, **285**, 20181047.

Joshi, R., Singh, N., Kirti, J.S., Volynkin, A.V. & Bucsek, K. (2017) Two new species of *Miltochrista* from India (Lepidoptera, Erebidae, Arctiinae). *Zootaxa*, **4238**, 445–450.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kawahara, A.Y. & Breinholt, J.W. (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **281**, 1–8.

Kawahara, A.Y., Plotkin, D., Espeland, M. *et al.* (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences*, **116**(45), 201907847.

Kearse, M., Moir, R., Wilson, A., *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kitching, I.J., & Rawlins, J.E. 1998. The Noctuoidea. Handbook of Zoology. Volume IV Arthropoda: Insecta. Part 35. Lepidoptera, Moths and Butterflies. Volume 1. Evolution, Systematics, and Biogeography (ed. by Kristensen N.P.). New York, Walter de Gruyter. pp. 355-401.

Lafontaine, D. & Schmidt, C. (2010) Annotated check list of the Noctuoidea (Insecta, Lepidoptera) of North America north of Mexico. *ZooKeys*, **40**, 1–239.

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research*, **22**, 746–754.

McKenna, D.D., Shin, S., Ahrens, D. *et al.* (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences*, 201909655.

Meyer, M. & Kircher, M. (2010) Illumina sequencing library for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**(6), 5448.

Miller, M.A., Pfeiffer, W., & Schwartz, T. 2010. *Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees in Proceedings of the Gateway Computing Environments Workshop (GCE)*. 14 November, New Orleans, LA, pp. 1–8.

Mirarab, S., Reaz, R., Bayzid, S., Zimmermann, T., Swenson, M.S. & Warnow, T. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**, i541–i548.

Misof, B., Liu, S., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Insect Phylogenomics*, **346**, 763–767.

Mitchell, A., Mitter, C. & Regier, J.C. (2006) Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes. *Systematic Entomology*, **31**, 21–46.

van Nieukerken E.J., Kaila L., Kitching I.J., et al., 2011. Order Lepidoptera Linnaeus, 1758. *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness* (ed. by Zhang, Z.-Q). *Zootaxa*, **3148**, 212–221.

Ozsolak, F. & Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, **12**, 87–98.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Peters, R.S., Krogmann, L., Mayer, C. *et al.* (2017) Evolutionary history of the hymenoptera. *Current Biology*, **27**, 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027.

Pinheiro, L.R. & Duarte, M. (2016) Description of nine new species of *Heliura* Butler from South America (Lepidoptera, Erebidae, Arctiinae, Arctiini, Ctenuchina). *Annales de la Société entomologique de France (N.S.)*, **51**, 310–330.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. & Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rannala, B. & Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: the barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 1–10.

Rawlins J.E. 1984. Mycophagy in Lepidoptera. *Fungus-Insect Relationships: Perspectives in Ecology and Evolution* (ed. by

Wheeler Q., Blackwell M.). New York: Columbia University Press. pp. 382–423.

Rojas, B., Mappes, J. & Burdfield-Steel, E. (2019) Multiple modalities in insect warning displays have additive effects against wild avian predators. *Behavioral Ecology and Sociobiology*, **73**, 37.

Rokyta, D.R., Lemmon, A.R., Margres, M.J. & Aronow, K. (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*, **13**, 1–23.

Rönkä, K., Mappes, J., Kaila, L. & Wahlberg, N. (2016) Putting *Parasemia* in its phylogenetic place: a molecular analysis of the subtribe Arctiina (Lepidoptera). *Systematic Entomology*, **41**, 844–853.

Salichos, L. & Rokas, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, **497**, 327–331.

Sanderford, M.V. & Conner, W.E. (1995) Acoustic courtship communication in *Syntomeida epilais* Wlk. (Lepidoptera: Arctiidae, Ctenuchidae). *Journal of Insect Behavior*, **8**, 19–31.

Sanderford, M.V., Coro, F. & Conner, W.E. (1998) Courtship behavior in *Empyreuma affinis* Roths. (Lepidoptera, Arctiidae, Ctenuchinae): acoustic signals and tympanic organ response. *Naturwissenschaften*, **85**, 82–87.

Schmidt, B.C. & Sperling, F.A.H. (2008) Widespread decoupling of mtDNA variation and species integrity in *Grammia* tiger moths (Lepidoptera: Noctuidae). *Systematic Entomology*, **33**, 613–634.

Schmidt, B.C. & Sullivan, J.B. (2018) Three species in one: a revision of *Clemensia albata* Packard (Erebidae, Arctiinae, Lithosiini). *ZooKeys*, **788**, 39–55.

Scott, C.H. & Branham, M.A. (2012) A preliminary phylogeny of the lichen moth tribe Lithosiini (Lepidoptera: Erebidae: Arctiinae) based on morphological characters. *Insect Systematics and Evolution*, **43**, 321–369.

Scott Chialvo, C.H., Chialvo, P., Holland, J.D. *et al.* (2018) A phylogenetic analysis of lichen-feeding tiger moths uncovers evolutionary origins of host chemical sequestration. *Molecular Phylogenetics and Evolution*, **121**, 23–34.

Scott, C.H., Zaspel, J.M., Chialvo, P. & Weller, S.J. (2014) A preliminary molecular phylogenetic assessment of the lichen moths. *Systematic Entomology*, **39**, 286–303.

Seo, T.-K. (2008) Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution*, **25**, 960–971.

Shin, S., Clarke, D.J., Lemmon, A.R., Lemmon, E.M. *et al.* (2017) Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Molecular Biology and Evolution*, **35**, 823–836.

Simmons, R.B. & Conner, W.E. (1996) Ultrasonic signals in the defense and courtship of *Euchaetes egle* Drury and E. bolteri stretch (Lepidoptera, Arctiidae). *Journal of Insect Behavior*, **9**, 909–919.

Simmons, R.B. & Weller, S.J. (2001) Utility and evolution of cytochrome b in insects. *Molecular Phylogenetics and Evolution*, **20**, 196–210.

Simmons, R.B. & Weller, S.J. (2002) What kind of signals do mimetic tiger moths send? A phylogenetic test of wasp mimicry systems (Lepidoptera: Arctiidae: Euchromiini). *Proceedings of the Royal Society of London Series B-Biological Sciences*, **269**, 983–990.

Simmons, R.B., Weller, S.J. & Johnson, S.J. (2012) The evolution of androconia in mimetic tiger moths (Noctuoidea: Erebidae: Arctiinae: Ctenuchina and Euchromiina). *Annals of the Entomological Society of America*, **105**, 804–816.

Singer, M.S., Mace, K.C. & Bernays, E.A. (2009) Self-medication as adaptive plasticity: increased ingestion of plant toxins by parasitized caterpillars. *PLoS One*, **4**, 1–8.

Sohn, J., Labandeira, C., Davis, D. & Mitter, C. (2012) An annotated catalog of fossil and subfossil Lepioptera (Insecta: Holometabola) of the world. *Zootaxa*, **3286**, 1–132.

St. Laurent, R.A., Hamilton, C.A. & Kawahara, A.Y. (2018) Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae). *Systematic Entomology*, **43**, 1–33.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Toussaint, E.F.A., Condamine, F.L., Kergoat, G.J., Capdevielle-Dulac, C., Barbut, J., Silvain, J. & Le Ru, B.P. (2012) Palaeoenvironmental shifts drove the adaptive radiation of a noctuid stemborer tribe Lepidoptera, (Noctuidae, Apameini) in the Miocene. *PLoS One*, **7**, e41377.

Toussaint, E.F.A., Breinholt, J.W., Earl, C. *et al.* (2018) Anchored phylogenomics illuminates the skipper butterfly tree of life. *BMC Evolutionary Biology*, **18**, 101.

Vincent, B. & Laguerre, M. (2014) Catalogue of the Neotropical Arctiini leach, [1815] (except Ctenuchina Kirby, 1837 and Euchromiina Butler, 1876) (Insecta, Lepidoptera Erebidae, Arctiinae). *Zoosystema*, **36**, 137–533.

Vincent, B., Laguerre, M. & Rougerie, R. (2009) Contribution à la connaissance du genre *Opharus* Walker avec description de deux nouvelles espèces. Apport des codes barres ADN (Lepidoptera, Arctiidae). *Bulletin de la Société entomologique de France*, **114**, 69–78.

Vincent, B., Hajibabaei, M. & Rougerie, R. (2014) A striking new genus and species of tiger-moth (Lepidoptera: Erebidae, Arctiinae, Arctiini) from the Caribbean, with molecular and morphological analysis of its systematic placement. *Zootaxa*, **3760**, 289–300.

Volynkin, A.V., Dubatolov, V.V. & Kishida, Y. (2018) *Miltochrista wangmini*, a new species from China (Lepidoptera, Erebidae, Arctiinae). *Zootaxa*, **4394**, 147–150.

Wahlberg, N. & Wheat, C.W. (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**, 231–242.

Watson, A. & Goodger, D.T. (1986) Catalogue of the Neotropical tiger moths. *Occasional Papers on Systematic Entomology*, **1**, 1–57.

Weller, S.J., Simmons, R.B. & Carlson, A. (2004) *Empyreuma* species and species limits: evidence from morphology and molecules (Arctiidae: Arctiinae: Ctenuchini). *The Journal of the Lepidopterists' Society*, **58**, 21–32.

Weller S.J., DaCosta M., Simmons R., Dittmar K., & Whiting M. 2009. Evolution and taxonomic confusion in Arctiidae. Tiger Moths and Woolly Bears: Behavior, Ecology, and Evolution of the Arctiidae (ed. by Conner W.E.). New York: Oxford University Press. pp. 11–30.

Wickett, N.J., Mirarab, S., Nguyen, N. *et al.* (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, **111**, E4859–E4868.

Winterton, S.L., Lemmon, A.R., Gillung, J.P. *et al.* (2018) Evolution of lacewings and allied orders using anchored phylogenomics (Neuroptera, Megaloptera, Raphidioptera). *Systematic Entomology*, **43**, 330–354.

Young, A.D., Lemmon, A.R., Skevington, J.H. *et al.* (2016) Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC Evolutionary Biology*, **16**, 1–13.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.Y. (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36.

Zahiri, R., Kitching, I.J., Lafontaine, J.D., Mutanen, M., Kaila, L., Holloway, J.D. & Wahlberg, N. (2011) A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). *Zoologica Scripta*, **40**, 158–173.

Zahiri, R., Holloway, J.D., Kitching, I.J., Lafontaine, J.D., Mutanen, M. & Wahlberg, N. (2012) Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). *Systematic Entomology*, **37**, 102–124.

Zahiri, R., Lafontaine, J.D., Schmidt, B.C., Zakharov, E.V. & Hebert, P.D.N. (2014) A transcontinental challenge – a test of DNA barcode performance for 1,541 species of Canadian Noctuoidea (Lepidoptera). *PLoS One*, **9**, e92797.

Zahiri, R., Lafontaine, J.D., Schmidt, B.C., Zakharov, E.V. & Hebert, P.D.N. (2017) Probing planetary biodiversity with DNA barcodes: the Noctuoidea of North America. *PLoS One*, **12**, e0178548.

Zaspel, J.M. & Weller, S.J. (2006) Review of generic limits of the tiger moth genera *Virbia* Walker and *Holomelina* Herrich-Schäffer (Lepidoptera: Arctiidae: Arctiinae) and their biogeography. *Zootaxa*, **1159**, 1–68.

Zaspel, J.M., Weller, S.J., Wardwell, C.T., Zahiri, R. & Wahlberg, N. (2014) Phylogeny and evolution of pharmacophagy in tiger moths (Lepidoptera: Erebidae: Arctiinae). *PLoS One*, **9**, e101975.

Zenker, M.M., Wahlberg, N., Brehm, G., Teston, J.A., Przybylowicz, L., Pie, M.R. & Freitas, A.V.L. (2016) Systematics and origin of moths in the subfamily Arctiinae (Lepidoptera, Erebidae) in the Neotropical region. *Zoologica Scripta*, **46**, 348–362.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018a) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, **19**(S6), 153.

Zhang, S.Q., Che, L.H., Li, Y., Liang, D., Pang, H., Ślipiński, A. & Zhang, P. (2018b) Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nature Communications*, **9**, 1–11.

Zhou, X., Lutteropp, S., Czech, L., Stamatakis, A., von Looz, M. & Rokas, A. (2019) Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. *Systematic Biology*, syz058. https://doi.org/10.1093/sysbio/syz058.

Accepted 28 March 2020