Author/s:
Milla, L;Moussalli, A;Wilcox, SA;van Nieukerken, EJ;Young, DA;Halsey, M;McConville, T;Jones, TM;Kallies, A;Hilton, DJ

Title:
Phylotranscriptomics resolves phylogeny of the Heliozelidae (Adeloidea: Lepidoptera) and suggests a Late Cretaceous origin in Australia

Date:
2020-01-01

Citation:
Milla, L., Moussalli, A., Wilcox, S. A., van Nieukerken, E. J., Young, D. A., Halsey, M., McConville, T., Jones, T. M., Kallies, A. & Hilton, D. J. (2020). Phylotranscriptomics resolves phylogeny of the Heliozelidae (Adeloidea: Lepidoptera) and suggests a Late Cretaceous origin in Australia. Systematic Entomology, 45 (1), pp.128-143. https://doi.org/10.1111/syen.12383.

Persistent Link:
https://hdl.handle.net/11343/286225

Milla Liz (Orcid ID: 0000-0002-6139-4336)

van Nieukerken Erik (Orcid ID: 0000-0002-5721-1840)

Phylotranscriptomics resolves phylogeny of the Heliozelidae (Adeloidea: Lepidoptera) and suggests a Late Cretaceous origin in Australia

Liz Milla[1,6*], Adnan Moussalli[2], Stephen A. Wilcox[7], Erik J. van Nieukerken[3], David A. Young[4], Mike Halsey[5], Tom McConville[1], Therésa M. Jones[1], Axel Kallies[1] and Douglas J. Hilton[1]

[1]School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia

[2]Museum Victoria, 11 Nicholson St, Carlton, Victoria, Australia

[3]Naturalis Biodiversity Center, Leiden, The Netherlands

[4]D'Estrees Entomology & Science Services, Kingscote, Australia

[5]Faculty of Health and Life Sciences, Oxford Brookes University, England

[6]Centre for Australian National Biodiversity Research, CSIRO, Canberra, Australia

[7]Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia

[*]Correspondence: Liz Milla, Centre for Australian National Biodiversity Research, CSIRO, PO Box 1700, Canberra, ACT 2602, Australia. E-mail: liz.milla@csiro.au

Abstract

Heliozelidae are a cosmopolitan family of small, day-flying moths, and include some pest species of commercial crops. Overall, the family is poorly known and lacks a well-resolved phylogeny. Previous molecular and taxonomic work has revealed rich undescribed diversity within the family, particularly in Australia; however, the relationships amongst the major clades or genera were not resolved. We sequenced the transcriptomes of 39 taxa, representing all major genera of Heliozelidae, and seven outgroups representing most other Adeloidea families and the putative sister superfamily, Andesianoidea. The resulting phylogeny, based on the coding sequences of up to 1049 nuclear genes, provides a robust hypothesis for the generic relationships within Heliozelidae. On the basis of this analysis, the genus *Plesiozela*, previously proposed sister group of all other Heliozelidae, is excluded from the family and formally transferred to Incurvariidae. By incorporating fossil and secondary time calibrations into our phylogeny, we estimated that Heliozelidae ancestors first appeared at the beginning of the Late Cretaceous, approximately 95 Mya. We propose an ancestral biogeographical range hypothesis of the family, based on a combination of our transcriptome data and a previous multigene study including over 100 species. Our ancestral range modelling results suggest Heliozelidae are likely to have originated in the Australian region, with subsequent range expansions to the rest of the world.

Introduction

Advances in genomic sequencing have enabled rapid progress in producing robust molecular phylogenies using the sequences of hundreds to thousands of genes from a wide range of organisms (Kocot *et al.*, 2011; Lemmon *et al.*, 2012; Misof *et al.*, 2014; Crawford *et al.*, 2015; Garrison *et al.*, 2016; Léveillé-Bourret *et al.*, 2017; O'Hara *et al.*, 2017). In the past few years, these phylogenomic approaches have been used to generate phylogenies for several lepidopteran families and superfamilies (Bazinet *et al.*, 2013; Kawahara & Breinholt, 2014; Breinholt *et al.*, 2017; Toussaint *et al.*, 2017; Homziak *et al.*, 2019). Generally, these phylogenies have focused on groups within the derived Ditrysia, which contains 98% of described lepidopteran species (van Nieukerken *et al.*, 2011; Mitter *et al.*, 2017). Nonditrysia, a paraphyletic assemblage of superfamilies evolved from some of the oldest divergences in Lepidoptera (Regier *et al.*, 2015b; Mitter *et al.*, 2017), are morphologically very diverse but are considered species-poor compared to the mega-diverse ditrysians (Davis & Gentili, 2003; Kristensen *et al.*, 2015; Regier *et al.*, 2015a). The majority of nonditrysian moths are small (wingspans of less than 10 mm), rendering them difficult to detect and generally challenging to survey and study, a factor likely contributing to their under-collection and perceived lack of diversity. Consequently, comprehensive phylogenies of nonditrysian families are lacking.

Heliozelidae (or "shield-bearer" moths) are a nonditrysian family of small, predominantly day-flying moths with a worldwide distribution, found on all continents except Antarctica (Nielsen, 1980; Nielsen & Davis, 1985; van Nieukerken *et al.*, 2011; Regier *et al.*, 2015b). Several studies have placed Heliozelidae within the superfamily Adeloidea, which comprises Heliozelidae, Adelidae, Prodoxidae, Incurvariidae, Cecidosidae and the recently described family Tridentaformidae (Nielsen, 1980; Nielsen & Davis, 1985; van Nieukerken *et al.*, 2011; Regier *et al.*, 2015b). There are currently 125 described species of Heliozelidae contained in twelve genera (van Nieukerken *et al.*, 2011; van Nieukerken *et al.*, 2012; van Nieukerken & Geertsema, 2015), with most species restricted to four genera (*Antispila* Hübner, *Coptodisca* Walsingham, *Heliozela* Herrich-Schäffer and *Hoplophanes* Meyrick). Most of the described species with known life histories are leaf-miners, with some species recognized as important pests of commercial crops, including grape vines, cranberry and walnut (Maier, 1988; van Nieukerken *et al.*, 2012; Bernardo *et al.*, 2015; van Nieukerken & Geertsema, 2015). Fewer than twenty new species of Heliozelidae have been described in the last few decades, mostly from Japan and the Americas (Opler, 1971; Lafontaine, 1974; Kuroko, 1982; Karsholt & Kristensen, 2003; Lee *et al.*, 2006b, a; van Nieukerken *et al.*, 2012; Lee & Hirowatari, 2013; van Nieukerken & Geertsema, 2015; Liu & Wang, 2017; Wang *et al.*, 2018). Heliozelid diversity is thought to be highest in North America and Australia, with many putative species yet to be described, particularly in Australia (van Nieukerken *et al.*, 2011; van Nieukerken *et al.*, 2012; Milla *et al.*, 2018). However, some regions, including India, South America and Africa, still remain poorly surveyed for heliozelid diversity. A lack of current and detailed taxonomic knowledge has hindered the study and identification of heliozelid species, including those that are potentially invasive (Bernardo *et al.*, 2011; van Nieukerken & Geertsema, 2015).

A recent molecular phylogeny based on four genes (Milla *et al.*, 2018) revealed strong support for the monophyly of four of the described genera (*Antispilina*, *Coptodisca*, *Holocacista* Walsingham & Durrant and *Hoplophanes*), while recovering *Antispila* as polyphyletic and *Heliozela* as paraphyletic with the inclusion of the monotypic genus *Tyriozela* Meyrick. However, none of the relationships between the major clades and genera of Heliozelidae were resolved, thus no inferences could be made as to the pattern and timing of evolution of the family. *Plesiozela* Karsholt & Kristensen, a genus containing three species and endemic to South America, has been proposed as the sister group to the rest of the family based on several putative Heliozelidae apomorphies (Nielsen, 1980; Karsholt & Kristensen, 2003), but molecular evidence has not been available to support this placement. Additionally, the phylogeny of Adelidae ("fairy" or "longhorn" moths), considered the sister family of Heliozelidae, has not been resolved, with some studies suggesting Adelidae to be polyphyletic (Regier *et al.*, 2013; Regier *et al.*, 2015b; Milla *et al.*, 2018). Thus, a comprehensive and robust phylogenetic study was required to resolve some of the questions about the phylogenetic relationships among Adeloidea families, and more specifically, Heliozelidae.

The major goals of our study were to confirm the monophyly of Heliozelidae and to resolve the evolutionary relationships between the major heliozelid lineages. We have taken advantage of the power of transcriptome sequencing for phylogenetic inference (Kawahara & Breinholt, 2014; Wen *et al.*, 2015; Bazinet *et al.*, 2016; Irisarri *et al.*, 2017; Janouškovec *et al.*, 2017), and the availability of published Lepidoptera transcriptome data (Misof *et al.*, 2014; Bazinet *et al.*, 2016) to extend our dataset. We combined transcriptome sequences from specimens representing the major heliozelid genera (a total of 39 described and putative species) with seven representatives from three of the five remaining Adeloidea families and the sister superfamily Andesianoidea. Based on our data, we propose the first robust molecular phylogeny of Heliozelidae. We also time-calibrated our phylogeny, using fossil and secondary calibrations, allowing us to provide an estimate of the establishment of Heliozelidae around the world. Finally, we combined our comprehensive dataset with previously published sequences to broaden the range of heliozelid species in our study, and we present an ancestral range reconstruction of the Heliozelidae that suggests the biogeographical origins of the family. Thus, this study represents one of the most comprehensive generic-level studies of nonditrysian Lepidoptera families.

Materials and Methods

Selected taxa and sample processing

We collected 39 specimens that were representative of the main Heliozelidae groups defined in Milla *et al.* (2018): three *Antispila* groups (*Antispila* group I, *Antispila* group II, and *A. ampelopsifoliella* group), *Antispilina*, *Coptodisca*, *Heliozela+Tyriozela*, *Holocacista*, *Hoplophanes* and *Pseliastis* Meyrick, as well as *Plesiozela*, the endemic South American genus proposed by Nielsen (1980) and Karsholt & Kristensen (2003) as sister group to all other Heliozelidae. *Antispila* group I, represented here by *A. freemani* Lafontaine, is regarded as the "true" *Antispila*, and thus we have renamed this group *Antispila sensu stricto* following van Nieukerken *et al.* (2012). The group labelled "Undescribed group I", containing three undescribed species from Australia, Costa Rica and Indonesia (Milla *et al.*, 2018), is not represented here, as we could not obtain fresh material for any of these species. We were also unable to obtain fresh material for the monotypic genera *Tyriozela* Meyrick (contained within *Heliozela*), *Ischnocanaba* Bradley and *Phanerozela* Meyrick from Japan, Solomon Islands and Brazil respectively, and *Microplitica* Meyrick, which contains two species from India and Indonesia. We have been able to include specimens of several taxa representing other undescribed species and genera according to their distinct morphology and host associations (Hilton *et al.*, unpublished). Identifiers for putative species were formed by a combination of the genus name and an epithet formed by non-italicised hostplant genus (where known), with country and, in the case of Australian undescribed species, state of origin (e.g. *Heliozela* EucalyptusAusWA). Putative genera have been temporarily assigned the label "genus" followed by a number (e.g. "genus2") and will be formally described in future publications. In this study, we will focus on their phylogenetic placement and do not address their taxonomic rank. For the outgroups, we selected six taxa from Adeloidea: three Adelidae, two Incurvariidae and one Prodoxidae, and a specimen of the superfamily Andesianoidea, which molecular studies have proposed as the sister group to Adeloidea (Mutanen *et al.*, 2010; Bazinet *et al.*, 2016).

All specimens included in this study were collected as either adults or larvae. They were either preserved in RNALater or killed by freezing at -80°C and later placed into RNALater. Table 1 lists all transcriptome specimens; full specimen details are in Table S1. RNA was extracted from specimens using the Nucleospin RNA XS (Macherey-Nagel) kit following the protocol for small samples, and RNA quality was checked using an Agilent 2200 TapeStation (Agilent Technologies Inc.). RNA libraries were prepared with the TruSeq mRNA library protocol for low inputs in half the standard reaction volume (25 μl) and quality checked using the TapeStation. Samples were multiplexed and sequenced on the Illumina NextSeq platform using 80 bp or 250 bp paired reads. Output statistics are in Table S2. Demultiplexed raw reads were quality trimmed using BIOVIA Pipeline Pilot 9.1 (Dassault Systemes) using the Trim by Quality component. The quality cutoff was set to Q20, except for samples MMP.003062, MMP.003067 and MMP.003042, where the cutoff was set to Q30, due to the longer 250 bp reads resulting in a sharp drop in quality after 150 bases. Each assembly was built using Trinity version 2.2, using the Trimmomatic option to remove the Illumina adapters. 'Raw' assemblies were merged by CD-HIT-EST (Li & Godzik, 2006; Fu *et al.*, 2012) using 100% identity to remove any identical fragments. The completeness of each merged assembly was then quantified by running it against the Endopterygota odb9 dataset using BUSCO 3.0.2 (Simão *et al.*, 2015). For two of our outgroups (*Nemophora scopolii* Kozlov *et al.*, SRA: SRR921621 and *Tegeticula yuccasella* (Riley), SRA: SRR3180626) published RNA-Seq data were downloaded from NCBI and assembled using the methods described above. The Trinity assembly and BUSCO results for all 46 samples used in this study are in Table S2. In most cases, whole specimens were used for RNA extraction; hence, to confirm species identification, we downloaded COI sequences of the same species from Genbank to use as vouchers and compared them to the transcriptome COI sequences. Where voucher COI sequences were not available, we sequenced COI from pinned conspecific specimens. Details of additional specimens sequenced as DNA vouchers are listed in Table S1, and p-distance comparisons between all voucher and transcriptome COI sequences are listed in Table S3.

Single-copy orthology search and alignment

We used the SCOPP pipeline (https://github.com/Victaphanta/SCOPP) to identify single copy, orthologous nuclear genes from the above transcriptomes. We used the genome of the silkworm (*Bombyx mori* (Linneaus)) as our reference, namely the CDS set as annotated in version ASM15162v1 (http://metazoa.ensembl.org/Bombyx_mori/Info/Index). Briefly, the SCOPP pipeline has three core modules to perform the following functions: a) search for homologs amongst multiple transcriptome assemblies using LAST (Kiełbasa *et al.*, 2011) (with an e-value cut-off of 1^{-10} in this study), b) align all matched sequences with MAFFT (Katoh & Standley, 2013), and c) merge redundant sequences based on an identity threshold (97% in this study).

From the homolog alignments generated by the SCOPP pipeline, we chose alignments where at least 70% of taxa were represented, and where only a single representative sequence remained per taxon. To extend this subset, we also examined all homolog alignments where only one or two taxa had two representative sequences remaining. All of these alignments were visually inspected to determine whether any sequences were the result of contamination, mis-indexing, paralogs or simply the retention of unrelated, low scoring matches. To assist with this assessment, we constructed individual nucleotide gene trees using IQ-TREE 1.6.1a (Nguyen *et al.*, 2014; Kalyaanamoorthy *et al.*, 2017; Hoang *et al.*, 2018) with the parameters: -m TESTNEW, -bb 1000, -nt 1 and *B. mori* as the outgroup (option -o). Sequences that were found to be either possible contamination, erroneous homolog matches or suspected out-paralogs were excluded from each alignment. We then used TreSpEx (Struck, 2014) to find potential 'hidden paralogs' (single copy, non-orthologous sequences). TreSpEx flags sequences found to be in conflict with well-supported, *a priori* clades within individual trees. In this study, we specified four large *a priori* clades strongly supported by Milla *et al.* (2018), namely: 1) *Hoplophanes*, 2) *Heliozela*, 3) *Antispila* sensu stricto + *A.* group II + *A. ampelopsifoliella* group + *Coptodisca* + *Holocacista* + *Antispilina*, and 4) *Pseliastis* + related taxa. The final set of single copy, orthologous gene alignments were trimmed using BMGE (Criscuolo & Gribaldo, 2010) with the options -t CODON, -m BLOSUM62 and -g 0.5 to remove poorly aligned regions and contigs less than 50% of the CDS length.

To address possible compositional heterogeneity, which can have an effect on tree estimation (Zwick *et al.*, 2012; Breinholt & Kawahara, 2013; Regier *et al.*, 2015a), we used BaCoCa (Kück & Struck, 2014) to run a Chi-square test on the homogeneity of base frequencies across taxa for each alignment. We also calculated average p-distances from a reference taxon for each nucleotide and amino acid alignment using a custom script. We chose a specimen, representing a species within the ingroup, *Pseliastis* PhebaliumAusWA, as the reference to calculate p-distances. A filtered gene subset (filtered_nt123) was created where we excluded nucleotides that did not meet the BaCoCa compositional heterogeneity test ($P<0.05$) or had average p-distances greater than 0.25. We also produced a degeneracy coded (full_degen) dataset from the full_nt123 matrix. This dataset was produced using Degen v1.4 (Zwick *et al.*, 2012) which degenerates nucleotides to IUPAC codes at positions that can undergo synonymous change to reduce the effect of compositional heterogeneity. All custom scripts are available on github (https://github.com/Victaphanta/SCOPP), and all specimen details have been uploaded to BOLD and COI data to Genbank (IDs in Table S1). Raw transcriptome FASTQ files have been submitted to NCBI SRA (PRJNA544864), and supplementary files, including syntax and files used in the SCOPP pipeline have been submitted to Dryad (doi:10.5061/dryad.c0pv774).

Phylogenetic analysis

Maximum likelihood phylogenies were estimated for each of the datasets using IQ-TREE 1.6.1a. The full_nt123 and filtered_nt123 datasets were partitioned by codon position. Evolutionary models were selected based on the best BIC score as calculated by ModelFinder (Kalyaanamoorthy *et al.*, 2017) within IQ-TREE. For both datasets, the best models were GTR+F+R4 for the first two codon positions, and GTR+F+R6 for the third codon position (GTR=General Time Reversible, F=empirically calculated base frequencies, R=FreeRate model with number of categories). For the full_degen dataset, the best model was GTR+F+R5, and for the full_aa dataset, LG4M+R6. Branch support was estimated with 1000 ultrafast (UFBoot) bootstraps (Hoang *et al.*, 2018) and 1000 SH-like approximate likelihood ratio (sh-aLRT) test replicates (Guindon *et al.*, 2010). We considered a branch to be strongly supported if the UFBoot values were >= 95% and the sh-aLRT values were >= 80%. We also used ASTRAL-III version 5.6.1 (Zhang *et al.*, 2017) to estimate a coalescent-based species tree, using individual gene trees estimated using IQ-TREE and the full_nt123 dataset. Prior to the ASTRAL-III analyses we collapsed all branches with UFBoot support less than 95% using the newick utilities library (Junier & Zdobnov, 2010).

Divergence time estimates

For our dating analyses we relied on two calibrations outside Heliozelidae. The first calibration was based on a fossil from the Lower Cretaceous, assigned by Whalley (1978) to the proposed genus *Incurvarites* Rebel, within the superfamily Incurvarioidea (now known as Adeloidea). Whalley (1978) examined wing margin scales from all families that were considered as Microlepidoptera at the time, as well as some specimens from Macrolepidoptera families. His preliminary assignment of these fossil wing margin scales to Incurvariidae was based on their highly toothed apex, which he found similar to other incurvariids, particularly the genus *Prodoxus* Riley (now in Prodoxidae, previously within Incurvariidae). Because this fossil identification is uncertain (Sohn *et al.*, 2015) and we cannot determine to which of the two families it should be assigned, we have used its approximate age to set the minimum age of the Incurvariidae and Prodoxidae split at 100 Ma. A secondary calibration was used for the root of the tree based on the estimates by Wahlberg *et al.* (2013), which calculated the split of Andesianidae and Prodoxidae at around 143±20 Mya (95% HPD). Accordingly, we used a soft-bounded flat prior for the root distribution with an upper bound of 163 Mya and a lower bound of 123 Ma. We used a relaxed clock model with uncorrelated rate as implemented in MCMCTree, part of the PAML suite (Yang, 2007), to estimate time calibrated trees based on both nucleotide and amino acid data. For the full_nt123 dataset, the dataset was then partitioned into the three codon positions and baseml was run to obtain an estimate of substitution rates to assist in setting appropriate corresponding priors in the MCMCTree analyses. To account for the uncertainty of the fossil and secondary calibrations, we selected a diffuse truncated Cauchy distribution (Inoue *et al.*, 2009) on the prior age distribution of these nodes, with an offset of 0.1 and a scale parameter of 0.5. We then ran two independent MCMCTree runs for 125M generations using the approximation method for large datasets as per Reis and Yang (2011), sampling every 5000 generations. We checked for convergence between the independent runs by generating a convergence plot of the posterior mean times. We used codeml on the full_aa dataset to obtain an estimate global substitution rate, and a Hessian and gradient approximation matrix, using the JTT+R5 model (second best model from IQ-TREE, as protein mixture models are not available in PAML). Two independent MCMCTree runs were performed with 100 million generations each, sampling every 5000 generations. Convergence in posterior nodal time estimates for the two amino acid runs was qualified by checking their convergence plot.

Ancestral range reconstruction

To estimate the ancestral range for the Heliozelidae family, we defined six major biogeographic areas where heliozelids are known to occur: Palearctic (Pa), Oriental (Or), Australian (Au), Neotropical (Nt), Nearctic (Na) and Afrotropical (Af). The extant range for each species was based on the biogeographical data from Milla *et al.* (2018), but here we combined East and West Palearctic areas into a single Palearctic region. Using the BioGeoBEARS v1.1.1 R package (Matzke, 2013), we ran the three available models (DEC, DIVALIKE and BAYAREALIKE) with and without founder events (specified by j parameter). Founder events allow individual lineages to disperse outside the ancestral range, for example, in the case of a long-distance dispersal event.

These events have been found to be important in explaining disjunct distributions, including those of small insects with poor flying ability such as micro-moths (Pyron, 2014; Rota *et al.*, 2016) . We also reran all six models to incorporate a time-restricted dispersal matrix ("stratified" models). This matrix imposes penalties for dispersal into non-adjacent areas during specified time periods. For this matrix we used the time slices and dispersal probabilities as per Rota *et al.* (2016), listed on Table S4. Given the global scale of the analysis and considering that no extant species are known to be naturally distributed beyond a single biogeographic region, we constrained the maximum number of biogeographic regions that a single species can occupy to two regions.

To maximize taxonomic sampling for the ancestral range estimation, we combined the COI, COII, and 28S sequences from our transcriptome assemblies with data from the recently published Heliozelidae molecular study (Milla *et al.*, 2018). We excluded H3 from our study as it is highly conserved and provided little phylogenetic information. This resulted in a final alignment of 116 Heliozelidae species, plus one outgroup (*Nemophora scopolii*). We used IQ-TREE to estimate an ML phylogeny from the concatenated COI+COII+28S alignment, imposing topological constraints consistent with the best ML tree derived from the transcriptome-based analysis, specifically the full_nt123 tree. Then, a time-calibrated tree was estimated for the resulting tree topology using MCMCTree, imposing secondary time calibration constraints specified by the 95% posterior intervals from the full_nt123 time-calibrated tree. We ran MCMCTree twice for just over 41 million generations using the COI+COII+28S dataset partitioned by first and second codon position for COI+COII, with the whole of 28S as the third partition. The third base position for the mitochondrial data was discarded to avoid issues with oversaturated signal. The MCMCTree runs were then combined to generate a consensus tree. The resulting tree (without the outgroup) was collapsed to a single representative per species and used as input into the ancestral range analysis described above.

Results

Sequence generation and matrix construction

We processed our 46 transcriptomes through the SCOPP pipeline and identified 1049 candidate homolog alignments with a minimum of 70% taxa completeness for any given reference gene. Within the 1049 selected alignments, 541 contained one or two taxa with two matched sequences. Through our visual inspection of the alignments and associated gene trees we identified and removed one or two erroneous sequences (contaminants, out-paralogs or unrelated matches) from 462 of the 541 alignments with multiple matches. In the other 79 alignments we found one taxon with two sequence fragments that were placed together in the gene tree but had not been merged by the assembler because they did not overlap or because their small overlap was just above the 97% similarity threshold. In these cases, the two fragments were manually merged into one sequence. Guided by the TreSpEx results, we identified a further five alignments where a sequence was in strong conflict with the four major *a priori* Heliozelidae clades, and the conflicting sequence was removed. The results from the transcriptome assemblies are found in Table S2, and the original and edited alignments are available on Dryad (doi:10.5061/dryad.c0pv774). The final nucleotide (full_nt123) and degen (full_degen) matrices of concatenated alignments contained 849,279 nucleotides with 14.7% gaps, and the translated amino acid (full_aa) matrix of concatenated alignments contained 283,093 amino acids. After filtering on BaCoCa homogeneity test results and p-distances, we obtained a subset of 581 genes, which were combined into a matrix of 437,709 nucleotides with 13.6% gaps (filtered_nt123).

Phylogeny of Heliozelidae and its relationship with sister groups

The phylogenetic analyses provided strong support for the relationships between all major Heliozelidae genera, contained in three major clades (described below). The majority of branches (maximum of 39 out of 42) received full or strong bootstrap support from both UFBoot and sh-aLRT bootstrap methods in all four ML analyses (Fig. 1). Similarly, there was a high degree of topological congruence between all ML analyses and there were no instances of strongly supported topological conflict (Figs S1-4). The topology of the ASTRAL-III coalescent gene tree (Fig. S5) was very similar to the topology of the ML phylogenies, with only three branches that were well-supported by ML (specifically filtered_nt123) but not recovered by ASTRAL-III (indicated as dashed lines in Fig. S5). Quartet support along the ASTRAL-III branches was variable, with little or no support for the branches that were different to the ML topology. Three of the ASTRAL-III branches had low local posterior probability (<0.95), a different measure of branch support for coalescent gene trees (Sayyari & Mirarab, 2016). Overall, the ML phylogenies provided better resolution than the ASTRAL-III tree. One possible reason for this difference is that the individual gene trees used as input to ASTRAL-III had many branches collapsed due to low support, decreasing the amount of information available for the quartet support calculation. Because the full_nt123 and filtered_nt123 ML phylogenies produced the same topology and had the highest number of resolved branches, we will primarily focus on the results from the filtered_nt123 phylogeny.

In all of the phylogenetic analyses, Heliozelidae was recovered as monophyletic by excluding *Plesiozela* (Figs S1-4), and it is in this sense we will consider the family in the following discussion. The first major clade is represented by Australian endemic *Hoplophanes* group (Clade 1, Fig. 1), containing many *Hoplophanes* and closely related species. The rest of the family is divided into two large clades, one clade containing the Australian endemic *Pseliastis* group (Clade 2, Fig. 1) and several undescribed closely related taxa, and a cosmopolitan clade containing *Heliozela* + *Antispila sensu stricto* + *Holocacista* + *Antispilina* + *Antispila* group II + *Coptodisca* (Clade 3, Fig. 1). Within the *Pseliastis* clade we recovered a number of very shallow branches, three of which received less than full bootstrap support. Within the cosmopolitan clade, the *Heliozela* group contained two lineages, one containing all *Heliozela* species and another containing undescribed Australian species associated with Dilleniaceae (labeled "genus8"). The remaining Heliozelidae within the cosmopolitan clade formed three major groups. One contained *Antispila sensu stricto* (as defined in Milla *et al.*, 2018), and represented here by *A. freemani*, a second group contained the cosmopolitan genus *Holocacista* and the genus

Antispilina, and third group contained only North and South American species, namely the endemic North American genus *Coptodisca*, North American species of *Antispila*, and a lineage represented by an undescribed species from South America, associated with *Nothofagus* sp. The representatives of *Antispila* group II (*A. argentifera*) and *A. ampelopsifoliella* group (*A. oinophylla*) as defined by Milla *et al.* (2018), joined together here with full support; thus, we have collapsed them into one group, labelled *Antispila* group II. The placement of *Plesiozela*, previously considered a heliozelid genus and sister to all other Heliozelidae, was fully supported within Incurvariidae, represented in this study by *Perthida* EucalyptusAusACT and *Incurvaria pectinea*. All ML phylogenies fully supported *Nemophora* Hoffmannsegg (subfamily Adelinae, Adelidae) as sister group to Heliozelidae, with *Ceromitia* Zeller + *Nematogopon* Zeller (subfamily Nematopogoninae, Adelidae) as sister to Heliozelidae + Adelinae, and Incurvariidae (including *Plesiozela*) as sister to Heliozelidae + Adelinae + Nematopogoninae. Details of all phylogeny results are in Figs S1-5.

Divergence time estimates

The independent MCMC runs of both nucleotide and amino acid analyses were found to have reached convergence (see plots in Fig. S6). The mean time estimates obtained from both the nucleotide and amino acid MCMCTree analyses (Fig. S7 and Fig. S8) were also very close, with each estimate from one type of analysis within the respective confidence range of the other. We therefore focused on the results from the nucleotide estimate, summarized in Fig. 2. The analysis placed the origin of Heliozelidae at the start of the Late Cretaceous (95.4 Ma, 95% CI 81.6-116.2), when the family split from Adelinae, represented here by *Nemophora scopolii*. A major split occurred during this period, with the *Hoplophanes* group separating from the rest of the family at around 84 Mya (95% CI 72-102). The next major split occurred between the *Pseliastis* clade and the cosmopolitan clade, at around 78.7 Mya (95% CI 66.9-96), followed by the *Heliozela* group from the rest of the cosmopolitan clade, at around 73.1 Mya (95% CI 62-89.2). Subsequently, there were several splits within the remaining cosmopolitan clade during the early Eocene giving rise to the *Antispila* group II + *Coptodisca* and *Holocacista* + *Antispilina* groups. There appears to be have been another burst of diversification during the Miocene, giving rise to several lineages within the *Pseliastis* clade. *Plesiozela*, is shown to have split from its sister genus in this analysis, *Perthida* (Incurvariidae), during the Early Eocene (53.5 Ma, 95% CI 37.7-72.9).

Ancestral range reconstruction

The time-calibrated ML tree generated with the combined COI, COII and 28S data and used in the ancestral range reconstruction is shown in Fig. S9. Recently, it has been suggested that direct comparisons of models that include jump dispersal parameters (such as DEC+J) and those without them (DEC) should be avoided (Ree & Sanmartín, 2018). Thus, we separately discuss the top performing models from both categories – with and without jump dispersal parameter (Table 2).

All models tested within BioGeoBEARS proposed the Australian region as one of the biogeographical regions where Heliozelidae originated (see Table 2). Eight of the models proposed Australia as the most likely ancestral region, while four models proposed two regions, Australia and either Oriental (DEC+j stratified) or Palearctic (DEC stratified, BAYAREALIKE and BAYAREALIKE stratified), as the most likely ancestral range. The top two performing models in the jump dispersal category ranked by AICc values (AIC corrected by sample size) returned very similar likelihood values, differing by less than 0.25 AICc units, and were non-stratified models. The best model in this category was the DEC+j model (LnL = -102.8493, AICc = 211.9110), followed by the DIVALIKE+j model (LnL = -103.0827, AICc = 212.3778). In the category without the jump dispersal parameter, the top two results differed by 2.39 AICc units. The best performing models in this category were DIVALIKE (LnL = -142.5435, AICc = 289.1923), followed by DEC (LnL = -143.74, AICc = 291.5854). One of the main differences between the top models in each category was that the DIVALIKE model estimated some ancestral ranges encompassing more than one biogeographic area, while the ancestral range for each node in the DEC+J model encompassed a single area. In the DIVALIKE scenario, some expansions into new areas could have occurred earlier than estimated by the DEC+J model. For example, the *Heliozela* ancestor was estimated by the DIVALIKE model to occupy both Australian and Neotropical regions earlier than 50 Mya, as estimated by the DEC+J model, while the rest of the Cosmopolitan clade was estimated to occupy the Australian and Palearctic regions earlier than 65 Mya, compared to the DEC+J, which estimated the incursion into the Palearctic region at just over 40 Mya (Fig. S10 and Fig. S12). Probabilities at each node also differed slightly between the two models (Fig. S11 and Fig. S13), with the DIVALIKE model returning higher confidence at the oldest nodes (e.g. root node probability of Australian origin was 0.996 with DIVALIKE vs. 0.782 with DEC+J).

One of the advantages of the DEC models is that they allow for the possibility of subset sympatry, where a daughter species inherits a subset of a large ancestral area while remaining within the ancestral area. As we have defined very large biogeographical areas, this type of speciation event is possible. Alternatively, DIVALIKE allows for widespread vicariance, where two daughter species inherit a large range. Because Heliozelidae are very small and tend to be closely associated with a particular host, we concluded subset sympatry would be a more likely event than widespread vicariance, hence we considered the DEC models to be more relevant to this study. The BAYAREALIKE models do not allow for any type of vicariance events to occur, a potentially significant speciation process. Therefore, for the discussion we have chosen to focus on the best-fitting DEC model (DEC+J), and we present the combined time-calibrated tree with the major ancestral range estimation results in Fig. 3.

Discussion

Phylogeny of Heliozelidae and relationships with sister groups

The use of large-scale, phylogenomic-level data has enabled us to generate a robust hypothesis regarding the monophyly of the Heliozelidae family and the evolutionary relationships between the major genera within it. One of the main findings concerns the placement of *Plesiozela*, a genus tentatively assigned to Heliozelidae and considered sister to the rest of Heliozelidae (Karsholt & Kristensen, 2003). *Plesiozela* is endemic to South America and contains only two described species, *P. nielseni* Karsholt & Kristensen and *P. patagonica* Karsholt & Kristensen. The genus was represented in this study by a specimen of *P. nielseni. Plesiozela* was recovered, invariably and with full support, within Incurvariidae, while the rest of Heliozelidae were consistently recovered as a monophyletic group. Both Nielsen (1980) and Karsholt & Kristensen (2003) suggested *Plesiozela* as the sister group to all other Heliozelidae based on four putative Heliozelidae apomorphies: a) dorsoventrally curved anterior tentorial arms (part of the internal head structures; parallel or converging in the rest of Adeloidea); b) M-CuA crossvein absent in hindwing (generally present in other adeloid genera); c) pilifers on the labrum absent (present in other Adeloidea); and d) reduced mandibles, present as small sclerotized protuberances in *Plesiozela* and *Hoplophanes*, but absent in other genera. However, Nielsen (1980) also suggested the assignment of *Plesiozela* into Incurvariidae as an alternative hypothesis, which is strongly supported by our data. This reassignment requires the reassessment of the apomorphic characters within the Adeloidea superfamily. The four original characters cannot be considered as solely Heliozelidae apomorphies, as they are present in at least some genera of Incurvariidae. Rather, based on Nielsen's analyses (Nielsen, 1980), the following characters could be considered as Heliozelidae apomorphies: a) "smooth" head, covered by flattened lamellar scales (all other Adeloidea, including *Plesiozela*, have a "rough" head covered by raised piliform scales); b) epicranial suture absent (presence considered part of the Adeloidea ground plan, present in *Plesiozela*); c) maxillary palps with four or less segments (plesiomorphic state is 5-segmented maxillary palps, as found in *Plesiozela*); and d) veins 1A+2A in forewing joined from base (separate at the base in *Plesiozela*). Thus, the molecular results presented here, together with the reassessment of the morphological characters discussed above, are compelling enough for us to justify the inclusion of *Plesiozela* within Incurvariidae. Nielsen (1980) also considered the length of the proboscis an important character for Heliozelidae. It is much longer than the labial palps in most heliozelids, while it is similar size or shorter than the labial palps in most other adeloids. However, it should be noted that a longer proboscis is present in *Nemophora* (Adelinae), the putative sister group to Heliozelidae, while some Heliozelidae genera have short proboscises.

Adelidae have been proposed as a monophyletic group, and sister family to Heliozelidae (Mutanen *et al.*, 2010; Wahlberg *et al.*, 2013). However, the monophyly of Adelidae is currently in question, as previous analyses suggest that Adelidae may be polyphyletic, with the Adelinae sub-family as the sister group to Heliozelidae (Regier *et al.*, 2013; Regier *et al.*, 2015b). The results obtained here support the notion of a polyphyletic Adelidae, with *Nemophora* (Adelinae) as sister to Heliozelidae, while *Ceromitia* and *Nematopogon* (Nematopogoninae) formed a sister clade to Heliozelidae + *Nemophora.* To robustly resolve the phylogeny of Adelidae, a phylogenomic approach combined with expanded taxon sampling may be necessary, as recent studies, including this one, are based on a relatively small number of genes and taxa, and have encountered conflicting results.

The phylogeny of the Heliozelidae presented in this study is largely consistent with previous studies (van Nieukerken *et al.*, 2012; Bernardo *et al.*, 2015; Milla *et al.*, 2018); however, we have provided robust support for relationships between major clades, which were previously unresolved. One of the main differences between this study and the most recent phylogeny (Milla *et al.*, 2018) is the placement of *Antispila sensu stricto* within the clade containing the majority of cosmopolitan genera (Clade 3, Fig. 1), and the clade containing *Pseliastis* (Clade 2), as sister clade to the cosmopolitan clade. There are other lines of evidence supporting the molecular relationships between genera recovered here, namely morphological and ecological similarities. The sister groups *Coptodisca* and *Antispila* group II contain species endemic to North America, and these species produce similar, distinctive leaf mines (van Nieukerken *et al.*, 2012). Closely related to these two groups, is a South American lineage represented in this study by a single species associated with *Nothofagus* Blume (labelled 'spNothofagusArg'), for which the name '*Neospila*' was proposed by Nielsen (1980). These are very small heliozelid moths from the Patagonian regions of Chile and Argentina, which exhibit shiny black forewings with white marks. *Holocacista* and *Antispilina* formed a monophyletic group, with the females of these two genera sharing a flattened, deeply dented ovipositor (Nielsen, 1980). *Antispilina* appears to be the only heliozelid genus of which all species feed on the leaves of herbaceous plants (*Persicaria* Mill. and *Bistorta* L. (Scop.), Polygonaceae), and is not as geographically widespread as the polyphagous *Holocacista* (Regier *et al.*, 2015a; Milla *et al.*, 2018). *Heliozela* species are associated with a large range of host plants, in particular from the family Myrtaceae. They generally exhibit 'typical' dark brown forewings with white markings and traverse-linear scale arrangements (Simonsen, 2001) and are petiole and leaf-miners (Nielsen, 1980). They are sister to a group of outwardly similar *Hibbertia*-associated species (labelled "genus8 HibbertiaAusWA" and "genus8 HibbertiaAusSA") that are seed and flower feeders.

The sister clade to the cosmopolitan clade, containing all of the aforementioned taxa, is the *Pseliastis* group (Clade 2), where we find several undescribed species, distinguished by their host associations and morphology. In this group, there were several shallow branches, some of which could not be resolved. This branching pattern is suggestive of a rapid radiation within this group, which may be able to be resolved with more concentrated sampling. Most species in this group are broadly similar in appearance to *Pseliastis sensu stricto*, namely small and silver or dark, sometimes with transverse or oblique white wing fasciae, feeding on either flowers or seeds during their larval stage. The vast majority of species in the *Pseliastis* group are strongly associated with hostplants of the Rutaceae family. Several species (labelled with "B." followed by *Boronia* Sm. species name and locality, e.g. "B.heterophyllaAusWA") have a very strong association with the endemic Australian Rutaceae genus *Boronia* (Milla *et al.*, 2018). Our recent morphological examination of a syntype of *Prophylactis argochalca* Meyrick, the type species of *Prophylactis* Meyrick, suggested that specimens labelled as "genus1" here may belong to that genus. Furthermore, both morphological and preliminary molecular evidence indicate that *Prophylactis* is a separate genus and not a synonym of *Hoplophanes*, as suggested by Nielsen *et al.* (1996). *Hoplophanes* are morphologically distinct from other Heliozelidae, exhibiting the largest wingspan of the family and a long, pointed ovipositor in females, a possible apomorphy for this genus (Nielsen, 1980).

Divergence times and the ancestral range of Heliozelidae

Within the relatively poor Lepidoptera fossil record there are a number of adeloid fossils, but very few that can be confidently assigned to Heliozelidae (Doorenweerd *et al.*, 2015; Sohn *et al.*, 2015). Labandeira (2002) identified possible Heliozelidae fossil mines in North America dating to the early Lutetian, middle Eocene (Sohn *et al.*, 2012). These mines were found in fossilized leaves of *Comptonia* L'Hér. ex Ait. (Myricaceae) and are similar to those formed by *Antispila* in Corneaceae and Vitaceae. Labandeira (2002) noted that Myricaceae was not known as a host family of North American *Antispila*. However, it is now known that *A. argentifera* Braun feeds on Myricaceae (including *Comptonia*) and that an unnamed species of *Coptodisca* feeds also on Myricaceae (Eiseman, 2019; van Nieukerken, unpublished). Consequently, we could not determine whether this fossil calibration should be placed at the split of *Coptodisca* and *Antispila* group II (which contains *A. argentifera*), or a higher level, specifically at the split between *Antispila sensu stricto* and *Holocacista* + *Antispilina* + *Antispila* group II + *Coptodisca*, and we excluded these fossils from our dating calculations. Considering the estimated age of the fossil mines and our results, the mines are more likely to belong to an ancestor of *Antispila* group II + *Coptodisca*, as ancestors of *Antispila sensu stricto* are estimated to have appeared in North America later, during the late Oligocene or early Miocene (Fig. 3).

A previous study estimated the superfamily Adeloidea split from its putative sister superfamily Andesianoidea in the late Jurassic/early Cretaceous, with the crown age of Adeloidea estimated at 115.8 Mya (Wahlberg *et al.*, 2013). Our estimate for the split between Andesianoidea and Adeloidea was 135.6 Mya (95% CI 122-161 Mya, Fig. 2), around 20 million years earlier than previous estimates. Likewise, our study has pushed back other estimates obtained by Wahlberg *et al.* (2013). For example, we estimated the split between Prodoxidae and the rest of the Adeloidea at 129 Mya (95% CI 112-156 Ma) compared to the previous estimate of 115 Ma, and between Incurvariidae and Adelinae + Nematopogoninae + Heliozelidae at around 111 Mya (95% CI 96-135), compared to 96 Mya. In particular, our estimate for the split between Adelidae (Adelinae) and Heliozelidae is much earlier at 95.4 Mya (95% CI 82-116 Mya) compared to 77 Mya in the Wahlberg *et al.* (2013) study, which may be due to our expanded taxon sampling and more precise branch length estimates from our phylogenomic scale dataset. We acknowledge that our time estimates are based on a single fossil calibration and a secondary calibration, both of which are placed outside of the family. Additional calibration points using the ages of confirmed Heliozelidae fossils would be required to improve the confidence in these estimates.

The ancestral range reconstruction, based on the three-gene dataset calibrated by the transcriptome-based time estimates (Fig. 3), provides a working hypothesis for the timing of Heliozelidae establishment around the world. All model reconstructions proposed a likely Australian origin of the family, although some suggested the Heliozelidae ancestor may have occupied an additional region, either Palearctic or Oriental. The crown of the Heliozelidae family was dated to the Late Cretaceous. During this time, the Australian continent is thought to have begun its separation from East Antarctica, the only Gondwanan landmass to which it remained connected (McLoughlin, 2001). An Australian ancestral origin was estimated (with high to medium probability) for the three major clades: *Hoplophanes* (1.00), *Pseliastis* (1.00) and the cosmopolitan clade (0.64). Two of the earliest diverging clades of Heliozelidae, represented today by species of *Hoplophanes*, *Pseliastis*, and related undescribed taxa, are currently known only from Australia. The most recent common ancestor of the *Pseliastis* group dates to the late Eocene, suggesting either limited cladogenesis or a high extinction rate during this period. Karyotype analyses in the tribe Boroniae, which includes the genus *Boronia* (Rutaceae), have suggested a range expansion and radiation around the Late Eocene (Smith-White, 1954; Shan *et al*., 2003). This would coincide with the estimated crown origin and observed radiation of the *Pseliastis* group, many species of which have a strong association with *Boronia*.

The first range expansion outside of Australia may have been the lineage that gave rise to *Antispila sensu stricto*, which appeared in the Palearctic region sometime between the Late Cretaceous and Middle Paleogene. Little ancestral diversity has been retained in this lineage, similar to the *Pseliastis* group. However, *Antispila sensu stricto* appears to have better ability to disperse and colonise new hosts than *Pseliastis* and *Hoplophanes*, as it expanded its Palearctic range during the Oligocene and early Miocene, with several independent incursions into the Nearctic region. Today, *Antispila sensu stricto* is associated with at least five different host plant families (van Nieukerken *et al*., 2012, Milla *et al*., 2018), and contains native North American *Antispila* species, including *A. isabella* Clemens, *A. cornifoliella* Clemens, and *A. freemani*.

The lineage represented by extant *Heliozela* species appears to have originated from Australia and spread to the Neotropics during the Eocene, and to the Oriental region during the Oligocene, followed by establishment in the Palearctic region by the early Miocene. The expansion from the Neotropics into the Palearctic region is estimated to have occurred during a period when South America was isolated from other continents except Antarctica, so an incursion into this area would have required long-distance dispersal events. This pattern also suggests good dispersal and colonisation ability, as today species of *Heliozela* are associated with at least four host plant families. *Holocacista* is another genus that also spread widely, colonising four different biogeographical regions (Australian, Afrotropical, Oriental and Palearctic), and is known to be associated with multiple host plant families. *Holocacista* appears to be the only genus present in Australia that has not originated there, having expanded into Australia through the Oriental region from its Palearctic origins. It is also the most diverse genus present in Africa, although the African heliozelid fauna has not been extensively investigated (van Nieukerken & Geertsema, 2015).

A lineage that originated in the Neotropics sometime during the Eocene gave rise to the rich North American fauna of today, including species of *Coptodisca* and *Antispila* group II. These two groups appear to now be restricted to the Nearctic regions, with their closest relative, represented by a *Nothofagus*-associated undescribed species, found in South America. This expansion from the Neotropics into the Nearctic region would have required a long-distance dispersal event, as at the time, South America was not connected to North America. The subsequent radiation of *Coptodisca* during the Mid-Late Miocene occurred during a time of great plant diversification, marked by the appearance of open grasslands and forests in many parts of the world, including North America (Kürschner *et al*., 2008).

In the transcriptome dataset there were few specimens from geographic regions that are currently under-surveyed, such as South America and Africa. Effects of sparse taxon sampling on ancestral range estimation can be mitigated when taxa representing geographically important areas are included, and the underlying phylogenetic tree is robust (Wang *et al.*, 2017). In our ancestral range analysis, we included at least one representative from each major genus identified in Milla *et al.* (2018), and the phylogenomic size of the dataset contributed to the robustness of the underlying phylogeny. We extended our dataset by incorporating results from a study with extensive representation of species from most biogeographical regions (Milla *et al.*, 2018), thus, we consider our ancestral range hypothesis to be plausible based on our current knowledge of Heliozelidae.

Conclusions

With this study, we have addressed three major questions regarding the evolution of the Heliozelidae family: the relationships between the major genera, the timing of divergence and the ancestral origin of the family. We have shown that a phylogenomic-level dataset is suitable for resolving the relationships between the major genera of a poorly known family. The phylogeny presented here provides a well-supported, robust framework for future studies into this diverse and intriguing moth family. We have also contributed a comprehensive transcriptome dataset that can be incorporated into other phylogenetic analyses, particularly those encompassing the most ancient lineages of Lepidoptera.

Acknowledgements

The authors acknowledge the support of many people and organisations during this study. For collecting and sending material: Rudolf Bryner (Biel, Switzerland), Leigh Steyn (Stellenbosch, South Africa), Steve Wullaert (Paal, Belgium). For molecular lab support: Camiel Doorenweerd (Leiden, Netherlands, now Honolulu, USA), Alex Gorman and Jordan Wilcox. For expert bioinformatics advice: Andreas Zwick (ANIC, CSIRO). For generous hospitality and support while in the field: Terry and Jen Haddon (Nannup, Australia). For use of computational resources: Melbourne Bioinformatics at the University of Melbourne (UOM0024). We would like to thank the staff of the following government departments and institutes for their assistance obtaining research permits and for their generous advice and help they have provided in the field: Department of Parks and Wildlife (Western Australia), especially Sarah Barrett, Danny Stefoni, Alan Wills, Jo Shalders, Mike Paxman and Barbara Beliers; Parks Victoria (Victoria, Australia), Department of Primary Industries, Parks, Water and Environment (Tasmania, Australia) and Office of Environment and Heritage (NSW, Australia), Museo Argentino de Ciencias Naturales "Bernadino Rivadiavia" (Buenos Aires, Argentina), particularly Dr. Luis Compagnucci; Paola Favre (Administration de Parques Nacionales, Argentina), Nicolás Pacheco (Parque Nacional Puyehue, Chile), Dr. Luis Parra (Universidad de Concepción, Chile) and Juan Gamin Muñoz (Corporación Nacional Forestal, Chile). For additional photographs we'd like to thank R. Bryner, Charley

Eiseman (Northfield, MA, USA) and Darren Carman (Australia). The authors would also like to thank the two external reviewers of this manuscript for their insightful comments and helpful suggestions. This work was supported by the following grants: Hermon Slade Foundation (HSF15/6), The Holsworth Wildlife Research Endowment - Equity Trustees Charitable Foundation (HOLSW2015-1-F067 and HOLSW2016-R1-F009), Australian Lepidoptera Research Endowment. L. Milla received support from a Research Training Program Scholarship, provided by the Australian government. All authors read and approved the final manuscript. The authors declare no conflict of interest.

References

Bazinet, A.L., Cummings, M.P., Mitter, K.T., & Mitter, C.W. (2013) Can RNA-Seq Resolve 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. (2016) Phylotranscriptomics resolves ancient divergences in the Lepidoptera. *Systematic Entomology*, **42**, 305-316.

Bernardo, U., Sasso, R., Gebiola, M., & Viggiani, G. (2011) First record of a walnut shield bearer Coptodisca (Lepidoptera: Heliozelidae) in Europe. *Journal of applied entomology*, **136**, 638-640.

Bernardo, U., van Nieukerken, E., Sasso, R., Gebiola, M., Gualtieri, L., & Viggiani, G. (2015) Characterization, distribution, biology and impact on Italian walnut orchards of the invasive North-American leafminer Coptodisca lucifluella (Lepidoptera: Heliozelidae). *Bull Entomol Res*, **105**, 210-224.

Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L., & Kawahara, A.Y. (2017) Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for Anchored Phylogenomics. *Systematic Biology*, syx048.

Breinholt, J.W. & Kawahara, A.Y. (2013) Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome biology and evolution*, **5**, 2082-2092.

Crawford, N.G., Parham, J.F., Sellas, A.B., Faircloth, B.C., Glenn, T.C., Papenfuss, T.J., Henderson, J.B., Hansen, M.H., & Simison, W.B. (2015) A phylogenomic analysis of turtles. *Mol Phylogenet Evol*, **83**, 250-257.

Criscuolo, A. & Gribaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*, **10**, 210.

Davis, D.R. & Gentili, P. (2003) Andesianidae, a new family of monotrysian moths (Lepidoptera: Andesianoidea) from austral South America. *Invertebrate Systematics*, **17**, 15-26.

Doorenweerd, C., Nieukerken, E.J.V., Sohn, J.-C., & Labandeira, C.C. (2015) A revised checklist of Nepticulidae fossils (Lepidoptera) indicates an Early Cretaceous origin. *Zootaxa*, **3963**, 295-334.

Eiseman, C. (2019) Leafminers of North America, Vol. 2019. Electronic book, privately published. http://charleyeiseman.com/leafminers/.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150-3152.

Garrison, N.L., Rodriguez, J., Agnarsson, I., Coddington, J.A., Griswold, C.E., Hamilton, C.A., Hedin, M., Kocot, K.M., Ledford, J.M., & Bond, J.E. (2016) Spider phylogenomics: untangling the Spider Tree of Life. *PeerJ*, **4**, e1719.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., & Vinh, L.S. (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*, **35**, 518-522.

Homziak, N.T., Breinholt, J.W., Branham, M.A., Storer, C.G., & Kawahara, A.Y. (2019) Anchored hybrid enrichment phylogenomics resolves the backbone of erebine moths. *Mol Phylogenet Evol*, **131**, 99-105.

Inoue, J., Donoghue, P.C., & Yang, Z. (2009) The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology*, **59**, 74-89.

Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J.-Y., Kupfer, A., Petersen, J., Jarek, M., Meyer, A., & Vences, M. (2017) Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nature Ecology & Evolution*, **1**, 1370.

Janouškovec, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S.G., Bright, K.J., Imanian, B., Strom, S.L., & Delwiche, C.F. (2017) Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences*, **114**, E171-E180.

Junier, T. & Zdobnov, E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, **26**, 1669-1670.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*, **14**, 587.

Karsholt, O. & Kristensen, N.P. (2003) Plesiozela, gen. nov. from temperate South America: apparent sister-group of the previously known Heliozelidae (Lepidoptera: Incurvarioidea: Heliozelidae). *Invertebrate Systematics*, **17**, 39-46.

Katoh, K. & Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*, **30**, 772-780.

Kawahara, A.Y. & Breinholt, J.W. (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. In Proc. R. Soc. B, Vol. 281, pp. 20140970. The Royal Society.

Kieł basa, S.M., Wan, R., Sato, K., Horton, P., & Frith, M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Research*, **21**, 487-493.

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., & Lieb, B. (2011) Phylogenomics reveals deep molluscan relationships. *Nature*, **477**, 452.

Kristensen, N.P., Hilton, D.J., Kallies, A., Milla, L., Rota, J., Wahlberg, N., Wilcox, S.A., Glatz, R.V., Young, D.A., & Cocking, G. (2015) A new extant family of primitive moths from Kangaroo Island, Australia, and its significance for understanding early Lepidoptera evolution. *Systematic Entomology*, **40**, 5-16.

Kück, P. & Struck, T.H. (2014) BaCoCa–A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Mol Phylogenet Evol*, **70**, 94-98.

Kuroko, H. (1982) Moths of Japan.Heliozelidae Vol. Kodansha, Tokyo, Japan.

Kürschner, W.M., Kvaček, Z., & Dilcher, D.L. (2008) The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. *Proceedings of the National Academy of Sciences*, **105**, 449-453.

Labandeira, C.C. (2002) Paleobiology of middle Eocene plant-insect associations from the Pacific Northwest: a preliminary report. *Rocky Mountain Geology*, **37**, 31-59.

Lafontaine, J. (1974) new species of Coptodisca (Heliozelidae) from Mississippi on farkleberry (Vaccinium arboreum). *Journal of the Lepidopterists' Society*, **28**, 126-130.

Lee, B.-W. & Hirowatari, T. (2013). Heliozelidae. In The standard of moths in Japan (ed. by N.Y. Hirowatari T, Sakamaki Y, Kishida Y), Vol. 3, pp. 99–102. Gakken Kyoiku Shuppan, Tokyo.

Lee, B.-W., Hirowatari, T., & Kuroko, H. (2006a) Five new species of the genus Heliozela Herrich-Schaffer (Lepidoptera, Heliozelidae) from Japan. *Transactions of the Lepidopterological Society of Japan*, **57**, 81-91.

Lee, B.-W., Hirowatari, T., & Kuroko, H. (2006b) A new species of Antispila (Lepidoptera: Heliozelidae) from the Ryukyus, Japan, with descriptions of immature stages. *Annals of the Entomological Society of America*, **99**, 1019-1027.

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

Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E., & Lemmon, A.R. (2017) Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology*, **67**, 94-112.

Li, W. & Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

Liu, T. & Wang, S. (2017) First report of the leaf-mining genus Antispila Hübner,[1825] from mainland China, with the description of a new species feeding on Cornus (Lepidoptera, Heliozelidae). *ZooKeys*, 95.

Maier, C.T. (1988) Life cycle of Coptodisca negligens (Lepidoptera: Heliozelidae) on cranberry. *J Econ Entomol*, **81**, 497-500.

Matzke, N.J. (2013) BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 0.2*, **1**, 2013.

McLoughlin, S. (2001) The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian journal of botany*, **49**, 271-300.

Milla, L., van Nieukerken, E.J., Vijverberg, R., Doorenweerd, C., Wilcox, S.A., Halsey, M., Young, D.A., Jones, T.M., Kallies, A., & Hilton, D.J. (2018) A preliminary molecular phylogeny of shield-bearer moths (Lepidoptera: Adeloidea: Heliozelidae) highlights rich undescribed diversity. *Mol Phylogenet Evol*, **120**, 129-143.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., & Beutel, R.G. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763-767.

Mitter, C., Davis, D.R., & Cummings, M.P. (2017) Phylogeny and Evolution of Lepidoptera. *Annu Rev Entomol*, **62**, 265-283.

Mutanen, M., Wahlberg, N., & Kaila, L. (2010) Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 2839-2848.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, **32**, 268-274.

Nielsen, E.S. (1980) The Incurvarioid genera of the world (Lepidoptera: Incurvarioidea) [unpublished thesis, typescript]. University of Copenhagen.

Nielsen, E.S. & Davis, D.R. (1985) The first southern hemisphere prodoxid and the phylogeny of the Incurvarioidea (Lepidoptera). *Systematic Entomology*, **10**, 307-322.

Nielsen, E.S., Edwards, E.D., Rangsi, T.V., & CSIRO (1996) Checklist of the Lepidoptera of Australia / E.S. Nielsen, E.D. Edwards, and T.V. Rangsi (eds.).Monographs on Australian Lepidoptera Vol., pp xiv, 529 p. :. CSIRO Australia, Collingwood, Vic., Australia.

O'Hara, T.D., Hugall, A.F., Thuy, B., Stöhr, S., & Martynov, A.V. (2017) Restructuring higher taxonomy using broad-scale phylogenomics: the living Ophiuroidea. *Mol Phylogenet Evol*, **107**, 415-430.

Opler, P.A. (1971) Seven new lepidoterous leaf-miners associated with Quercus agrifolia (Heliozelidae, Gracilariidae). *Lepidopterists Soc J*, 194-211.

Pyron, R.A. (2014) Biogeographic Analysis Reveals Ancient Continental Vicariance and Recent Oceanic Dispersal in Amphibians. *Systematic Biology*, **63**, 779-797.

Ree, R.H. & Sanmartín, I. (2018) Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography*, **45**, 741-749.

Regier, J.C., Mitter, C., Davis, D.R., Harrison, T.L., Sohn, J.C., Cummings, M.P., Zwick, A., & Mitter, K.T. (2015a) A molecular phylogeny and revised classification for the oldest ditrysian moth lineages (Lepidoptera: Tineoidea), with implications for ancestral feeding habits of the mega-diverse Ditrysia. *Systematic Entomology*, **40**, 409-432.

Regier, J.C., Mitter, C., Kristensen, N.P., Davis, D.R., van Nieukerken, E.J., Rota, J., Simonsen, T.J., Mitter, K.T., Kawahara, A.Y., & Yen, S.H. (2015b) A molecular phylogeny for the oldest (nonditrysian) lineages of extant Lepidoptera, with implications for classification, comparative morphology and life-history evolution. *Systematic Entomology*, **40**, 671-704.

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.-C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C., & Mitter, K.T. (2013) A Large-Scale, Higher-Level, Molecular Phylogenetic Study of the Insect Order Lepidoptera (Moths and Butterflies). *PLoS ONE*, **8**, e58568.

Reis, M.d. & Yang, Z. (2011) Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol*, **28**, 2161-2172.

Rota, J., Peña, C., & Miller, S.E. (2016) The importance of long-distance dispersal and establishment events in small insects: historical biogeography of metalmark moths (Lepidoptera, Choreutidae). *Journal of Biogeography*, **43**, 1254-1265.

Sayyari, E. & Mirarab, S. (2016) Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol Biol Evol*, **33**, 1654-1668.

Shan, F., Yan, G., & Plummer, J.A. (2003) Karyotype evolution in the genus Boronia (Rutaceae). *Botanical Journal of the Linnean Society*, **142**, 309-320.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212.

Simonsen, T.J. (2001) The wing vestiture of the non-ditrysian Lepidoptera (Insecta). Comparative morphology and phylogenetic implications. *Acta Zoologica*, **82**, 275-298.

Smith-White, S. (1954) Chromosome numbers in the Boronieae (Rutaceae) and their bearing on the evolutionary development of the tribe in the Australian flora. *Australian journal of botany*, **2**, 287-303.

Sohn, J.-C., Labandeira, C.C., & Davis, D.R. (2015) The fossil record and taphonomy of butterflies and moths (Insecta, Lepidoptera): implications for evolutionary diversity and divergence-time estimates. *BMC Evol Biol*, **15**, 12.

Sohn, J.-C., Labandeira, C.C., Davis, D.R., & Mitter, C. (2012) An annotated catalog of fossil and subfossil Lepidoptera (Insecta: Holometabola) of the world.

Struck, T.H. (2014) TreSpEx—Detection of Misleading Signal in Phylogenetic Reconstructions Based on Tree Information. *Evolutionary Bioinformatics*, **10**, EBO. S14239.

Toussaint, E.F., Hendrich, L., Hájek, J., Michat, M.C., Panjaitan, R., Short, A.E., & Balke, M. (2017) Evolution of Pacific Rim diving beetles sheds light on Amphi-Pacific biogeography. *Ecography*, **40**, 500-510.

van Nieukerken, E., Kaila, L., Kitching, I., Kristensen, N.P., Lees, D., Minet, J., Mitter, J., Mutanen, M., Regier, J., Simonsen, T., Wahlberg, N., Yen, S., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B., Brown, J., Bucheli, R., Davis, D., Prins, J.D., Prins, W.D., Epstein, M., Gentili-Poole, P., Gielis, C., Hättenschwiler, P., Hausmann, A., Holloway, J., Kallies, A., Karsholt, O., Kawahara, A., Koster, S., Kozlov, M., Lafontaine, J., Lamas, G., Landry, J., Lee, S., Nuss, M., Park, K.-T., Penz, C., Rota, J., Schmidt, B., Schintlmeister, A., Sohn, J., Solis, M., Tarmann, G., Warren, A., Weller, S., Yakovlev, Y., Zolotuhin, V., & Zwick, A. (2011) Order Lepidoptera Linnaeus, 1758. In Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (ed. by Z.-Q. Zhang), Vol. 3148, pp. 212-221. Zootaxa, 3148.

van Nieukerken, E.J. & Geertsema, H. (2015) A new leafminer on grapevine and Rhoicissus (Vitaceae) in South Africa within an expanded generic concept of Holocacista (Insecta, Lepidoptera, Heliozelidae). *ZooKeys*, **507**, 41-97.

van Nieukerken, E.J., Wagner, D., Baldessari, M., Mazzon, L., Angeli, G., Girolami, V., Duso, C., & Doorenweerd, C. (2012) Antispila oinophylla new species (Lepidoptera, Heliozelidae), a new North American grapevine leafminer invading Italian vineyards: taxonomy, DNA barcodes and life cycle. *ZooKeys*, **170**, 29-77.

Wahlberg, N., Wheat, C.W., & Peña, C. (2013) Timing and Patterns in the Taxonomic Diversification of Lepidoptera (Butterflies and Moths). *PLoS ONE*, **8**, e80875.

Wang, N., Kimball, R.T., Braun, E.L., Liang, B., & Zhang, Z. (2017) Ancestral range reconstruction of Galliformes: the effects of topology and taxon sampling. *Journal of Biogeography*, **44**, 122-135.

Wang, N., Liu, T., Xu, J., & Jiang, B. (2018) The leaf-mining genus Antispila Hübner, 1825 feeding on Vitaceae in Shandong Peninsula, China with one new species (Lepidoptera, Heliozelidae). *ZooKeys*, 49.

Wen, J., Egan, A.N., Dikow, R.B., & Zimmer, E.A. (2015) Utility of transcriptome sequencing for phylogenetic inference and character evolution. *Next-generation sequencing in plant systematics*.

Whalley, P. (1978) New taxa of fossil and recent Micropterigidae with a discussion of their evolution and a comment on the evolution of Lepidoptera (Insecta). *Annals of the Transvaal Museum*, **31**, 71-86.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, **24**, 1586-1591.

Zhang, C., Sayyari, E., & Mirarab, S. (2017) ASTRAL-III: increased scalability and impacts of contracting low support branches. In RECOMB International Workshop on Comparative Genomics, pp. 53-75. Springer.

Zwick, A., Regier, J.C., & Zwickl, D.J. (2012) Resolving discrepancy between nucleotides and amino acids in deep-level Arthropod phylogenomics: differentiating serine codons in 21-amino-acid models. *PLoS ONE*, **7**, e47450.

Table 1. Transcriptome specimen details. *The resulting FASTQ files for these two specimens were merged into one and are subsequently referred to by ID MMP.003414.

Specimen ID	Family	Genus	Species	Host family	Host genus	Country
MMP.003062	Heliozelidae	*Antispila*	*argentifera*	Myricaceae	*Myrica*	USA
MMP.003070	Heliozelidae	*Antispila*	*freemani*	Cornaceae	*Cornus*	USA
MMP.003061	Heliozelidae	*Antispila*	*oinophylla*	Vitaceae	*Vitis*	Canada
MMP.004868	Heliozelidae	*Antispilina*	*ludwigi*	Polygonaceae	*Bistorta*	Belgium
MMP.003068	Heliozelidae	*Coptodisca*	*lucifluella*	Juglandaceae	*Carya*	Canada
MMP.003067	Heliozelidae	*Coptodisca*	*splendoriferella*	Rosaceae	*Prunus*	USA
MMP.003360	Heliozelidae	*Heliozela*	ChamelociumAusWA	Myrtaceae	*Chamelaucium*	Australia
MMP.003398	Heliozelidae	*Heliozela*	EucalyptusAusWA	Myrtaceae	*Eucalyptus*	Australia
MMP.003005	Heliozelidae	*Heliozela*	UnknownAusVIC			Australia
MMP.003428	Heliozelidae	*Heliozela*	UnknownChile			Chile
MMP.004870	Heliozelidae	*Holocacista*	*capensis*	Vitaceae	*Vitis*	South Africa
MMP.003035	Heliozelidae	*Holocacista*	CoprosmaAusNSW	Rubiaceae	*Coprosma*	Australia
MMP.003374	Heliozelidae	*Holocacista*	LoganiaAusWA	Loganiaceae	*Logania*	Australia
MMP.003007	Heliozelidae	*Hoplophanes*	AcrotricheAusSA	Ericaceae	*Acrotriche*	Australia
MMP.003122	Heliozelidae	*Hoplophanes*	Leucopogon1AusWA	Ericaceae	*Leucopogon*	Australia
MMP.003130	Heliozelidae	*Hoplophanes*	Leucopogon2AusWA	Ericaceae	*Leucopogon*	Australia
MMP.004899	Heliozelidae	*Hoplophanes*	UnknownAusWA	Unknown		Australia
MMP.003309	Heliozelidae	*Pseliastis*	B.citriodoraAusTas	Rutaceae	*Boronia*	Australia
MMP.003358	Heliozelidae	*Pseliastis*	B.ramosaAusWA	Rutaceae	*Boronia*	Australia
MMP.003151	Heliozelidae	*Pseliastis*	PhebaliumAusWA	Rutaceae	*Phebalium*	Australia
MMP.003315	Heliozelidae	*Pseliastis*	*xanthodisca*	Rutaceae	*Nematolepis*	Australia
MMP.003394	Heliozelidae	genus1	B.clavataAusWA	Rutaceae	*Boronia*	Australia
MMP.003376	Heliozelidae	genus1	B.tetrandraAusWA	Rutaceae	*Boronia*	Australia
MMP.003124	Heliozelidae	genus2	B.heterophyllaAusWA	Rutaceae	*Boronia*	Australia
MMP.003123	Heliozelidae	genus2	B.megastigmaAusWA	Rutaceae	*Boronia*	Australia
MMP.003030	Heliozelidae	genus3	DodonaeaAusSA	Sapindaceae	*Dodonaea*	Australia
MMP.003319	Heliozelidae	genus4	PomaderrisAusVIC	Rhamnaceae	*Pomaderris*	Australia
MMP.003343	Heliozelidae	genus5	GeleznowiaAusWA	Rutaceae	*Geleznowia*	Australia
MMP.003042	Heliozelidae	genus6	B.algidaAusVIC	Rutaceae	*Boronia*	Australia
MMP.003086	Heliozelidae	genus7	ZieriaAusVIC	Rutaceae	*Zieria*	Australia
MMP.003020	Heliozelidae	genus8	HibbertiaAusSA	Dilleniaceae	*Hibbertia*	Australia
MMP.003372	Heliozelidae	genus8	HibbertiaAusWA	Dilleniaceae	*Hibbertia*	Australia
MMP.003040	Heliozelidae	genus9	GrevilleaAusNSW	Proteaceae	*Grevillea*	Australia
MMP.003153	Heliozelidae	genus10	RhadinothamnusAusWA	Rutaceae	*Rhadinothamnus*	Australia
MMP.003125	Heliozelidae	genus11	B.spathulataAusWA	Rutaceae	*Boronia*	Australia
MMP.003027	Heliozelidae	genus13	MelaleucaAusSA	Myrtaceae	*Melaleuca*	Australia
MMP.003423	Heliozelidae	genus14	NothofagusArg	Nothofagaceae	*Nothofagus*	Argentina
MMP.003347	Heliozelidae	genus15	B.corulescensAusWA	Rutaceae	*Boronia*	Australia
MMP.003331	Adelidae	*Ceromitia*	*iolampra*	Fabaceae	*Acacia*	Australia
MMP.004875	Adelidae	*Nematopogon*	*adansoniella*	Fagaceae		Switzerland
SRR921621	Adelidae	*Nemophora*	*scopolii*			Austria
MMP.003304	Incurvariidae	*Incurvaria*	*pectinea*	Betulaceae	*Betula*	United Kingdom
MMP.003037	Incurvariidae	*Perthida*	EucalyptusAusACT	Myrtaceae	*Eucalyptus*	Australia
MMP.003414*	Incurvariidae	*Plesiozela*	*nielseni*	Nothofagaceae	*Nothofagus*	Argentina
MMP.003415*	Incurvariidae	*Plesiozela*	*nielseni*	Nothofagaceae	*Nothofagus*	Argentina
SRR3180626	Prodoxidae	*Tegeticula*	*yuccasella*	Asparagaceae		USA
MMP.003425	Andesianoidea	*Andesiana*	*lamellata*	Nothofagaceae	*Nothofagus*	Argentina

Table 2. BioGeoBEARS analysis results ordered number of parameters and AICc score. Stratified models refer to models that incorporated a time-restricted dispersal matrix.

Model	LnL	Number of parameters	d	e	j	AICc	AICc weighted	Proposed Heliozelidae ancestral range
With jump dispersal								
DEC+j	-102.8493	3	1.00E-12	1.00E-12	0.0243	211.9110	3.89E-01	Australia
DIVALIKE+j	-103.0827	3	1.00E-12	1.00E-12	0.0243	212.3778	3.08E-01	Australia
BAYAREALIKE+J	-103.0963	3	1.00E-07	1.00E-07	0.0240	212.4051	3.04E-01	Australia
DEC+j Stratified	-117.0417	3	1.00E-12	1.00E-12	0.1784	240.2958	2.67E-07	Australia+Oriental
DIVALIKE+j Stratified	-120.6133	3	1.00E-12	1.00E-12	0.1872	247.4390	7.50E-09	Australia
BAYAREALIKE+J Stratified	-120.6154	3	1.00E-07	1.00E-07	0.1711	247.4433	7.48E-09	Australia
Without jump dispersal								
DIVALIKE	-142.5435	2	1.74E-03	1.00E-12	-	289.1923	6.43E-18	Australia
DEC	-143.7400	2	1.45E-03	1.76E-03	-	291.5854	1.94E-18	Australia
DEC Stratified	-158.2622	2	8.42E-03	3.33E-03	-	320.6297	9.59E-25	Australia+Palearctic
DIVALIKE Stratified	-164.6689	2	1.05E-02	2.51E-03	-	333.4430	1.58E-27	Australia
BAYAREALIKE	-173.1765	2	1.86E-03	1.39E-02	-	350.4583	3.19E-31	Australia+Palearctic
BAYAREALIKE Stratified	-180.9694	2	1.02E-02	1.50E-02	-	366.0440	1.32E-34	Australia+Palearctic

Supporting Information

Fig. S1. Nucleotide ML phylogeny generated using IQ-TREE based on 1049 genes (full_nt123). Numbers above the branches are support based on 1000 ultrafast bootstraps (UFBoot)/1000 SH-like approximate likelihood ratio replicates (sh-aLRT).

Fig. S2. Amino-acid ML phylogeny generated using IQ-TREE based on 1049 genes (full_aa). Numbers above the branches are support based on 1000 ultrafast bootstraps (UFBoot)/1000 SH-like approximate likelihood ratio replicates (sh-aLRT).

Fig. S3. Degeneracy-coded nucleotide phylogeny generated using IQ-TREE based on 1049 genes (full_degen). Numbers above the branches are support based on 1000 ultrafast bootstraps (UFBoot)/1000 SH-like approximate likelihood ratio replicates (sh-aLRT).

Fig. S4. Nucleotide ML phylogeny generated using IQ-TREE based on 581 genes filtered on compositional heterogeneity test result and p-distance (filtered_nt123). Numbers above the branches are support based on 1000 ultrafast bootstraps (UFBoot)/1000 SH-like approximate likelihood ratio replicates (sh-aLRT).

Fig. S5. Coalescent gene tree of 1049 genes generated with ASTRAL-III. Numbers above branches indicate quartet support/local posterior branch support (Sayyari & Mirarab, 2016). Dashed lines indicate branches not supported by ML phylogenies.

Fig. S6. Convergence plots of mean posterior times for independent runs of MCMCTree tree. Top: Run 1 vs Run 2 for full_nt123 dataset. Bottom: Run 1 vs Run 2 for full_aa dataset.

Fig. S7. Time-calibrated phylogeny from 1049 genes based on nucleotide alignment and generated using MCMCTree.

Fig. S8. Time-calibrated phylogeny from 1049 genes based on amino-acid alignment and generated using MCMCTree.

Fig. S9. MCMCTree tree of three genes (COI+COII+28S), partitioned by first and second codon position (COI+COII) and all positions (28S), with topology constrained by transcriptome-generated ML phylogeny result.

Fig. S10. BioGeoBEARS ancestral biogeographical range estimation for Heliozelidae based on top model with jump dispersal, DEC+J (LnL = -102.8493, AICc = 211.911).

Fig. S11. BioGeoBEARS ancestral biogeographical range reconstruction of Heliozelidae. Output from top model with jump dispersal, DEC+J (LnL = -102.8493, AICc = 211.911).

Fig. S12. BioGeoBEARS ancestral biogeographical range estimation of Heliozelidae based on top model without jump dispersal, DIVALIKE (LnL = -142.5435, AICc = 289.1923).

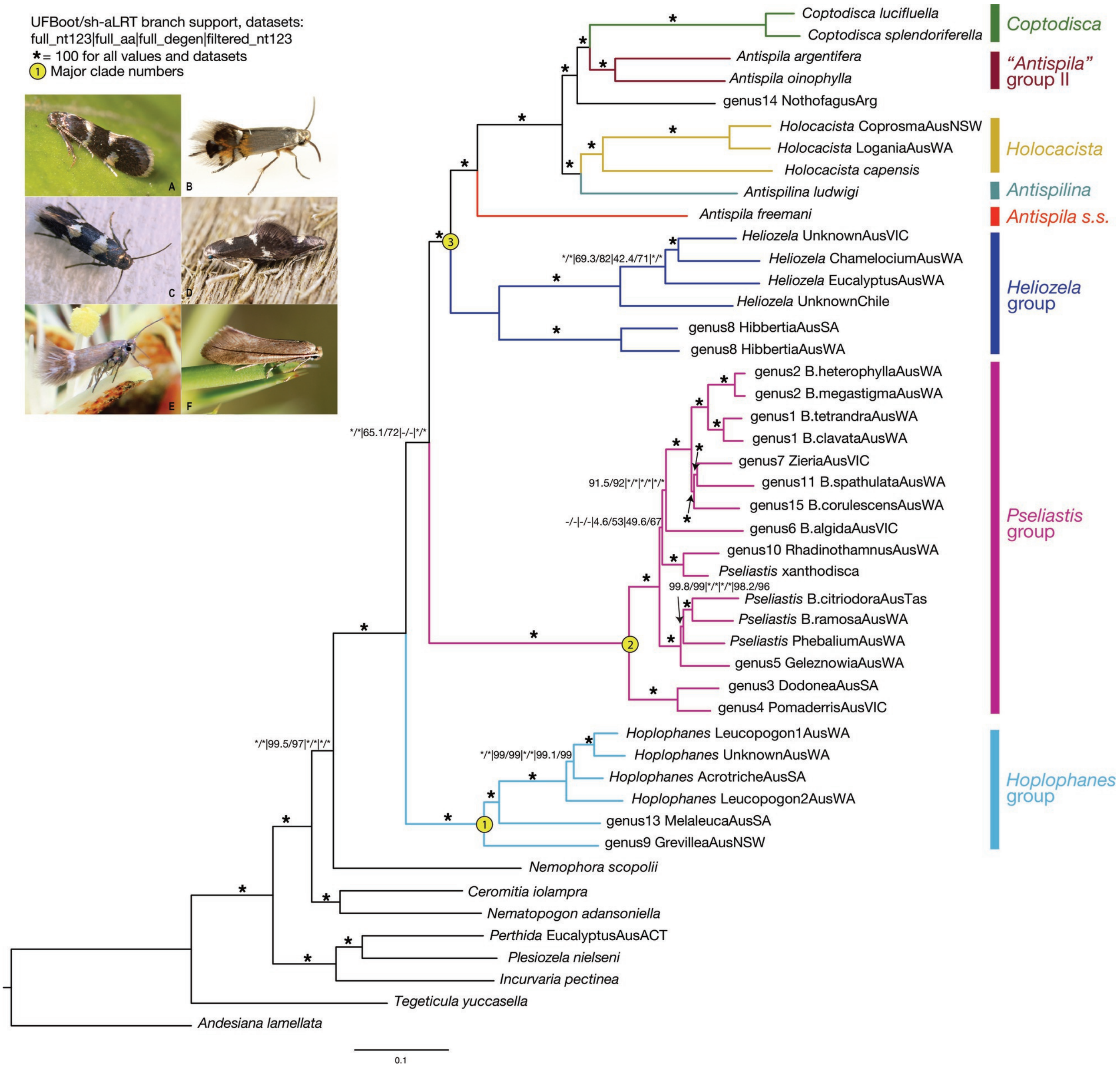
Fig. S13. BioGeoBEARS ancestral biogeographical range probabilities of Heliozelidae based on top model without jump dispersal, DIVALIKE (LnL = -142.5435, AICc = 289.1923).

Table S1. Full specimen details for transcriptomes and DNA vouchers.

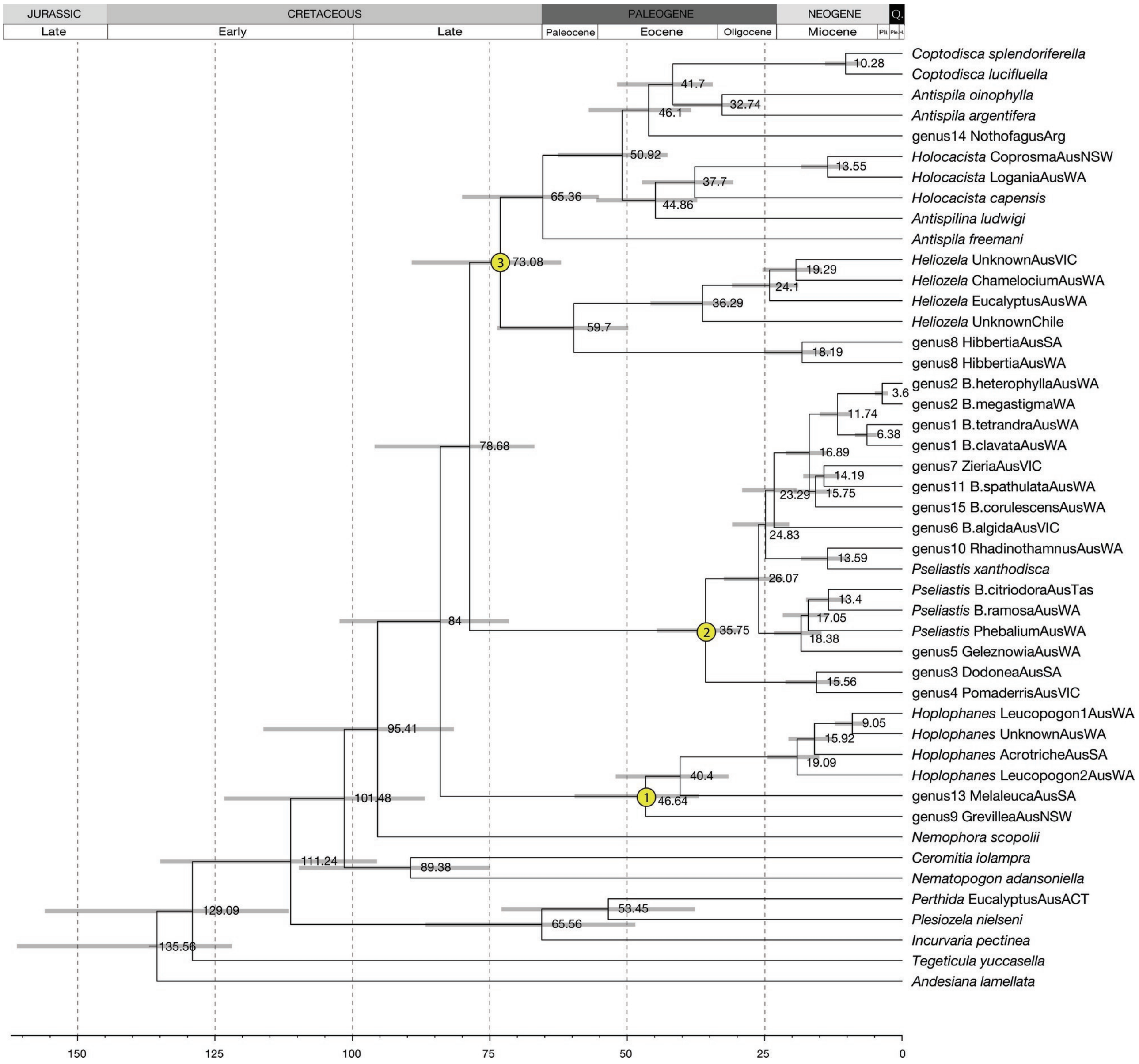
Table S2. Transcriptome raw read counts, Trinity 2.2 assembly statistics and BUSCO results. ^Indicates library was sequenced over two runs, statistics are given based on combined output files. Pair read length for both sequencing runs shown (if different between runs). *Due to poor quality, reads for two samples of *Plesiozela nielseni* (MMP.003414 and MMP.003415) were merged, identified subsequently as MMP.003414.

Table S3. COI voucher p-distances to transcriptome specimens. Specimen details for DNA vouchers are found in Table S1.

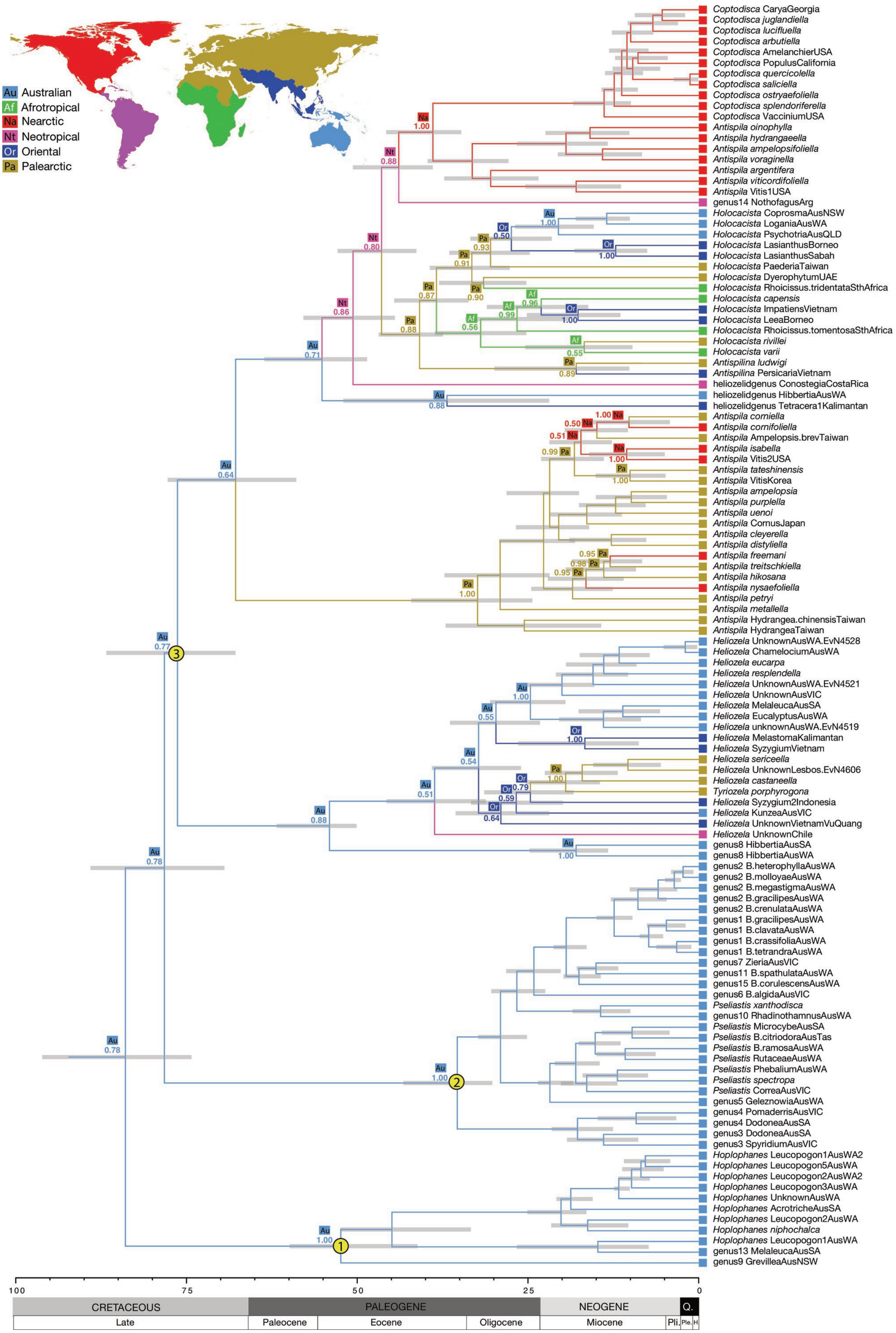
Table S4. Probability matrices used in BioGeoBEARS time-stratified analysis, following Rota *et al.* 2016. Each cell represents the probability of dispersal between intersecting regions at the given time periods.



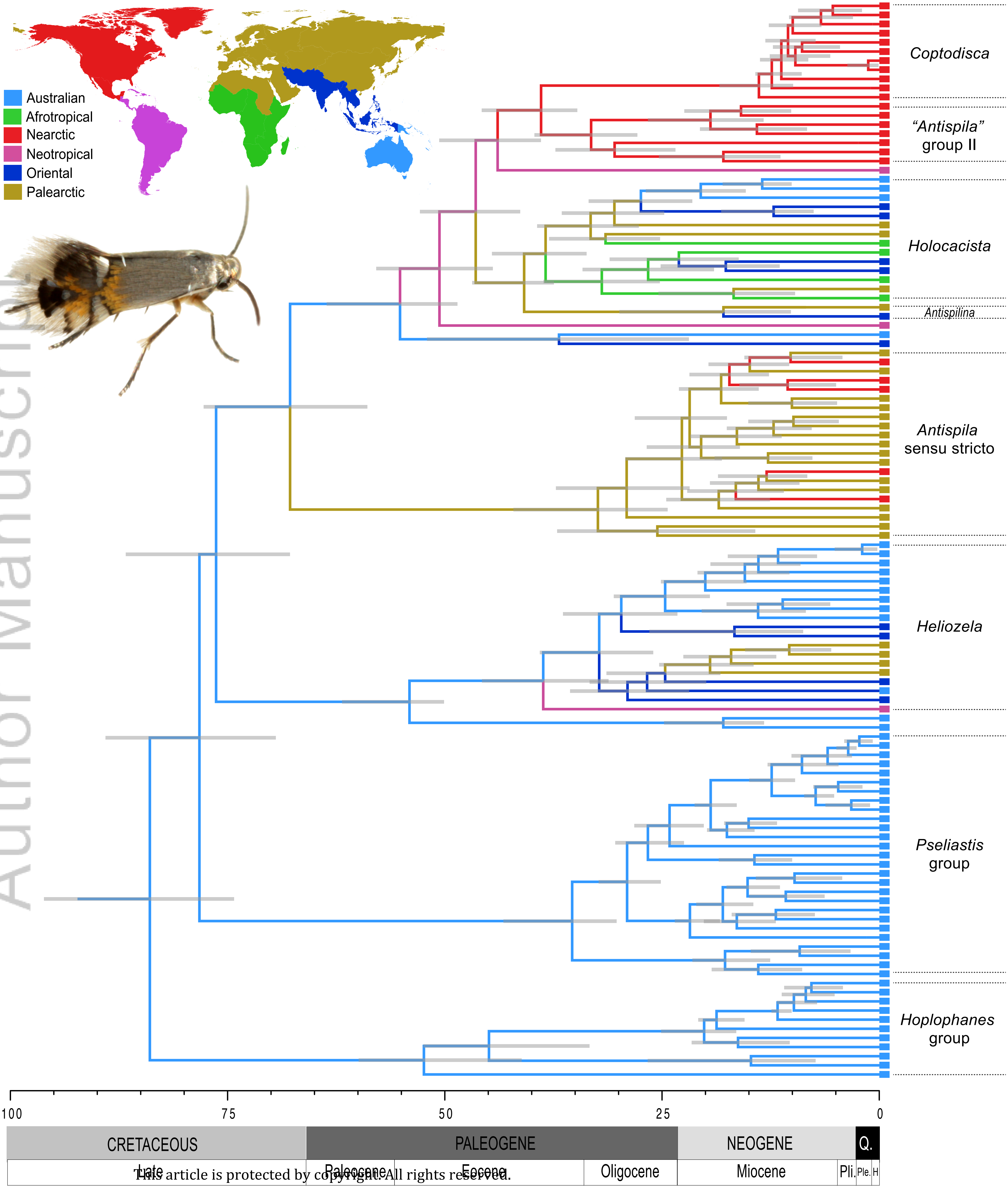
syen_12383_figure_1.eps



syen_12383_figure_2.eps



syen_12383_figure_3.eps


🟦 Australian
🟩 Afrotropical
🟥 Nearctic
🟪 Neotropical
🟦 Oriental
🟨 Palearctic
Coptodisca
"Antispila" group II
Holocacista
Antispilina
Antispila sensu stricto
Heliozela
Pseliastis group
Hoplophanes group
100
75
50
25
0
CRETACEOUS
PALEOGENE
NEOGENE
Q.
Late
Paleocene
Eocene
Oligocene
Miocene
Pli.
Ple.
H

Phylotranscriptomics resolves phylogeny of the Heliozelidae (Adeloidea: Lepidoptera) and suggests a Late Cretaceous origin in Australia

Liz Milla[1,6*], Adnan Moussalli[2], Stephen A. Wilcox[7], Erik J. van Nieukerken[3], David A. Young[4], Mike Halsey[5], Tom McConville[1], Therésa M. Jones[1], Axel Kallies[1] and Douglas J. Hilton[1]

[1]School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia

[2]Museum Victoria, 11 Nicholson St, Carlton, Victoria, Australia

[3]Naturalis Biodiversity Center, Leiden, The Netherlands

[4]D'Estrees Entomology & Science Services, Kingscote, Australia

[5]Faculty of Health and Life Sciences, Oxford Brookes University, England

[6]Centre for Australian National Biodiversity Research, CSIRO, Canberra, Australia

[7]Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia

[*]Correspondence: Liz Milla, Centre for Australian National Biodiversity Research, CSIRO, PO Box 1700, Canberra, ACT 2602, Australia. E-mail: liz.milla@csiro.au

Graphical abstract

Ancestral state reconstruction of Heliozelidae overlaid on time-calibrated MCMCTree phylogeny estimated from three genes and constrained by transcriptome-derived topology. Inset image: *Coptodisca splendoriferella* Clemens (credit: C. Eiseman).

Highlights

1. A robust, well-resolved molecular phylogeny of the Heliozelidae (Adeloidea) family based on a phylogenomic-level dataset derived from the transcriptomes of representative species.
2. The relationships between the major Heliozelidae clades are resolved, including the placement of *Plesiozela*, considered sister to all other Heliozelidae, now formally transferred to Incurvariidae.
3. The origin of the Heliozelidae family is estimated to the Late Cretaceous, pushing back earlier estimates. The ancestral range of the Heliozelidae is likely to have included the Australian region, with subsequent incursions into the rest of the world.