HELDA
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

The (non) accuracy of mitochondrial genomes for family-level phylogenetics in Erebidae (Lepidoptera)

Ghanavi, Hamid Reza

2022-11

Ghanavi , H R , Twort , V , Hartman , T J , Zahiri , R & Wahlberg , N 2022 , ' The (non) accuracy of mitochondrial genomes for family-level phylogenetics in Erebidae (Lepidoptera) ' , Zoologica Scripta , vol. 51 , no. 6 , pp. 695-707 . https://doi.org/10.1111/zsc.12559

http://hdl.handle.net/10138/349891
https://doi.org/10.1111/zsc.12559

cc_by_nc
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Check for updates

Received: 10 March 2022 | Revised: 13 July 2022 | Accepted: 13 July 2022

DOI: 10.1111/zsc.12559

ORIGINAL ARTICLE

Zoologica Scripta | WILEY

The (non) accuracy of mitochondrial genomes for family-level phylogenetics in Erebidae (Lepidoptera)

Hamid Reza Ghanavi[1] | Victoria Twort[1,2] | Tobias Joannes Hartman[1] | Reza Zahiri[3,4] | Niklas Wahlberg[1]

[1]Systematic Biology Group, Biology Department, Lund University, Lund, Sweden

[2]Zoology Unit, The Finnish Museum of Natural History Luomus, University of Helsinki, Helsinki, Finland

[3]Center of Natural History, University of Hamburg, Hamburg, Germany

[4]Ottawa Plant Laboratory, Canadian Food Inspection Agency, Entomology Diagnostic Laboratory, Ottawa, Ontario, Canada

Correspondence
Hamid Reza Ghanavi, Systematic Biology Group, Biology Department, Lund University, Lund, Sweden.
Email: hamid.ghanavi@biol.lu.se

Funding information
European Union's Horizon 2020 research and innovation program, Marie Skldowska-Curie, Grant/Award Number: 6422141; Vetenskapsrådet, Grant/Award Number: 2015-04441

Abstract

The use of molecular data to study the evolutionary history of organisms has revolutionized the field of systematics. Now with the appearance of high throughput sequencing (HTS) technologies, more and more genetic sequence data are available. One of the important sources of genetic data for phylogenetic analyses has been mitochondrial DNA. The limitations of mitochondrial DNA for the study of phylogenetic relationships have been thoroughly explored in the age of single locus phylogenetic studies. Now with the appearance of genomic scale data, increasing number of mitochondrial genomes are available, leading to an increasing number of mitophylogenomic studies. Here, we assemble 47 mitochondrial genomes using whole genome Illumina short reads from representatives of the family Erebidae (Lepidoptera), in order to evaluate the accuracy of mitochondrial genome application in resolving deep phylogenetic relationships. We find that mitogenomes are inadequate for resolving subfamily-level relationships in Erebidae, but given good taxon sampling, we see its potential in resolving lower level phylogenetic relationships.

KEYWORDS

Erebidae, Lepidoptera, mitochondrial genome, old DNA extract, phylogenomics

1 | INTRODUCTION

The ability to study the evolutionary histories of organisms has been revolutionized by the appearance and broad applicability of molecular methods. This ability to infer phylogenetic relationships based on molecular data was a major step forward in our understanding compared to traditional morphological comparative methods. Mitochondrial genomes offered the first possibility to use genomic scale data to infer phylogenetic hypotheses early in the history of molecular systematics. The newly accessible mitogenomic approach saw a rise in its use for resolving deep phylogenetic relationships, in arthropods and in other groups (Nardi, 2003; Simon & Hadrys, 2013; Song et al., 2016). Since such methods became popular, some researchers have questioned the limitations of mitochondrial genetic data for resolving early divergence events or deep phylogenetic relationships (Cameron et al., 2004; Talavera & Vila, 2011; Zardoya & Meyer, 1996). Nevertheless, many studies have applied mitochondrial genomes as a source of information to resolve phylogenetic relationships of varied evolutionary

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Zoologica Scripta* published by John Wiley & Sons Ltd on behalf of Royal Swedish Academy of Sciences.

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

depth. Some studies focused on the relationships within a superorder (Cameron et al., 2009; Li et al., 2015; Talavera & Vila, 2011), an order (Cameron et al., 2007; Dowton et al., 2009; Fenn et al., 2008; Kim et al., 2011; López-López & Vogler, 2017; Timmermans et al., 2014; Yang et al., 2019), a family (Chen et al., 2014; Chen, Zheng, et al., 2020; Li et al., 2018, 2020; Xu et al., 2020; Yang et al., 2015; Zhang et al., 2020) or shallower relationships such as at the species or population-level.

The phylogenetic depth of a relationship can affect the amount of phylogenetic signal coded in molecular data. In general, markers with higher mutation rates are only informative for shallower evolutionary relationships or recent divergences. For clades splitting deeper in time, fast-evolving markers will accumulate too many saturated sites and, therefore, tend to not resolve their phylogenetic relationships accurately. On the other hand, for markers with a very low mutation rate, the marker may not accumulate enough changes and lack phylogenetic signal for shallow relationships. Compared to the nuclear genome, mitochondrial genomes have long been thought to contain relatively homogenous molecular markers in terms of mutation rate (Papadopoulou et al., 2010).

A peculiarity of the mitochondrial genome is the lack of recombination, which means that in practice mitochondrial DNA behaves as a single genetic marker with a unique evolutionary history. In addition, the mitogenome is only maternally inherited, meaning that it has an effective population size one fourth of the nuclear genome. The mitochondrial genome is also susceptible to selective sweeps (Rubinoff et al., 2006; Sperling, 2003). Other discordance between mitochondrial and nuclear genome phylogenies can be associated with introgression following hybridization or due to retained ancestral polymorphism (Sperling, 2003). Therefore, mitochondrial markers can be misleading in cases of hybridization and are more affected by demographic factors than nuclear markers.

The initial approaches to sequencing mitochondrial genomes used PCR to amplify long pieces of overlapping molecules, Sanger sequencing of the long molecules and manually assembling sequence data. The labour intensiveness and costly nature of these methods, has put mitochondrial genomes out of reach for many research groups. With the appearance of High Throughput Sequencing (HTS) methods, the price per bp of sequencing data is dropping considerably. Analytical advances for HTS data and the wide variety of easily accessible bioinformatic pipelines, have simplified their use for a large number of research groups around the world (Cameron, 2014). Therefore, it is currently easier and more economical to obtain a high number of mitochondrial genomes. Ease of sequencing mitochondrial genomes has resulted in the publication of single genomes often without addressing specific research questions. Some authors responded to these poor scientific practices by publishing a larger number of mitochondrial genomes to address a clear question at phylogenetic depths appropriate to these markers (Chen, Wahlberg, et al., 2020).

Considering the characteristics of the mitochondrial genome as a molecular marker, the question of phylogenetic utility arises. It is also unclear if this important genetic marker can reliably resolve phylogenetic relationships for groups, which have experienced rapid radiations. In case of rapid radiations, during a short period of time, numerous lineages arise. Resolving phylogenetic relationships from rapid radiation events is challenging due to the fact that the marker must evolve quickly enough to accumulate enough changes during the rapid radiation phase, but slow enough to not saturate afterwards. One such groups that present such phylogenetic challenges is the moth family Erebidae.

Erebidae is one of most diverse families of moths and butterflies (Lepidoptera) with over 24,500 species described (van Nieukerken et al., 2011). In the most complete phylogenetic study of the group to date (Zahiri et al., 2012), many short branches were recovered at deep levels, suggesting a possible rapid radiation event. The family is divided into 18 subfamilies at the moment, of which a few have received more attention from systematists (e.g. Lymantriinae, Erebinae and Arctiinae; see Dowdy et al., 2020; Homziak et al., 2019; Wang et al., 2015), while others have never been studied in detail. Relationships at the subfamily level within Erebidae are currently poorly resolved, likely due to the lack of phylogenetic signal in the markers previously used.

Here, we assemble new mitochondrial genomes of 47 species of Erebidae (Table 1) representing all known subfamilies and major lineages based on the most recent phylogenetic hypotheses in order to capture the deepest nodes within the family. In addition, we downloaded 37 publicly available mitochondrial genomes and mined five transcriptomes (Table 2) for 11 protein coding genes found in the mitochondrial genomes. We compare the obtained phylogenetic hypotheses with prior supported relationships recovered in other studies to evaluate the phylogenetic accuracy range of mitochondrial genomes as markers in Erebidae.

2 | MATERIAL AND METHODS

2.1 | Taxon sampling

We sequenced low coverage whole genomes from DNA extracts of 47 species of Erebidae (Table 1). DNA extracts were the same as those used by Zahiri et al. (2012). Taxon choice was made in order to recover the deepest nodes within recognized subfamilies and major lineages in the family Erebidae. We focus mainly on the short

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 List of species sequenced in this study

#	Codes	Subfamily	Tribe	Species	Circular	Length	#tRNA	GC%	AT%
1	RZ44	Aganainae	Aganaini	*Asota heliconia*	Yes	15,446	22	19.9	80.1
2	RZ268	Aganainae		*Mecodina praecipua*	Yes	15,501	22	19	81
3	RZ103	Anobinae		*Rema costimacula*	Yes	15,668	22	19.3	80.7
4	RZ332	Anobinae		*Anoba anguliplaga*	No	14,835	20	18.9	81.1
5	RZ404	Arctiinae	Amerilini	*Amerila astreus*	Yes	15,519	22	19.6	80.4
6	RZ30	Arctiinae	Arctiini	*Creatonotos transiens*	Yes	15,569	22	18.9	81.1
7	RZ28	Arctiinae	Lithosiini	*Brunia antica*	Yes	15,489	22	19.4	80.6
8	RZ8	Arctiinae	Syntomini	*Amata phegea*	Yes	15,534	22	18.9	81.1
9	RZ3	Boletobiinae	Aventiini	*Laspeyria flexula*	Yes	15,583	22	20.1	79.9
10	RZ104	Boletobiinae		*Saroba pustulifera*	Yes	15,731	22	19.5	80.5
11	RZ41	Boletobiinae	Aventiini	*Metaemene atrigutta*	Yes	15,629	22	20.5	79.5
12	RZ336	Calpinae	Calpini	*Calyptra hokkaida*	Yes	15,562	22	18.3	81.7
13	RZ337	Calpinae	Calpini	*Oraesia excavata*	Yes	15,769	22	18.6	81.4
14	RZ56	Calpinae	Phyllodini	*Phyllodes eyndhovii*	Yes	15,612	22	18.2	81.8
15	RZ248	Erebinae	Acantholipini	*Acantholipes circumdata*	Yes	16,224	25	20.7	79.3
16	RZ11	Erebinae	Erebini	*Erebus ephesperis*	Yes	15,688	22	18.6	81.4
17	RZ39	Erebinae	Hulodini	*Ericeia subcinerea*	Yes*	15,880	24**	19.7	80.3
18	RZ149	Erebinae	Hypopyrini	*Hypopyra capensis*	Yes	15,702	22	19.1	80.9
19	RZ58	Erebinae	Melipotini	*Melipotis jucunda*	Yes*	16,616	22	18.5	81.5
20	RZ21	Erebinae	Ophiusini	*Ophiusa coronata*	Yes	15,762	22	18.8	81.2
21	RZ313	Erebinae	Sypnini	*Sypnoides fumosa*	Yes	15,527	22	19.4	80.6
22	RZ48	Erebinae	Erebini	*Sympis rufibasis*	Yes	15,572	22	18.5	81.5
23	RZ59	Eulepidotinae	Eulepidotini	*Panopoda rufimargo*	Yes	15,986	22	18.8	81.2
24	RZ22	Eulepidotinae		*Azeta ceramina*	Yes	15,696	22	19	81
25	RZ180	Herminiinae		*Nodaria verticalis*	No	14,175	18	18.5	81.5
26	RZ271	Herminiinae		*Idia aemula*	No	15,464	22	18.8	81.2
27	RZ367	Hypeninae		*Hypena baltimoralis*	No	14,724	20	19.6	80.4
28	RZ138	Hypenodinae	Micronoctuini	*Micronoctua sp.*	Yes	15,466	22	19	81
29	RZ42	Hypenodinae		*Luceria striata*	Yes	15,383	22	20.1	79.9
30	RZ105	Hypocalinae		*Hypocala deflorata*	No	14,428	19	18.8	81.2
31	RZ89	Lymantriinae	Arctornithini	*Arctornis sp.*	Yes	15,506	22	21.4	78.6
32	RZ34	Lymantriinae	Nygmiini	*Nygmia plana*	No	14,479	19	19.1	80.9
33	RZ18	Pangraptinae		*Masca abactalis*	Yes	15,562	22	19.5	80.5
34	RZ40	Pangraptinae		*Pangrapta bicornuta*	Yes*	15,957	22	18.1	81.9
35	RZ159	Rivulinae		*Rivula ochrea*	No	14,510	19	18.2	81.8
36	RZ94	Rivulinae		*Alesua etialis*	Yes	15,198	19	17.7	82.3
37	RZ9	Scolecocampinae		*Scolecocampa liburna*	Yes	15,580	22	18.9	81.1

(Continues)

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

#	Codes	Subfamily	Tribe	Species	Circular	Length	#tRNA	GC%	AT%
38	RZ13	Scoliopteryginae	Anomini	*Gonitis involuta*	Yes	15,695	22	19.2	80.8
39	MM00407	Scoliopteryginae	Scoliopterygini	*Scoliopteryx libatrix*	Yes*	15,617	22	19.4	80.6
40	RZ331	Tinoliinae		*Tinolius eburneigutta*	No	15,026	21	19.1	80.9
41	RZ389	Tinoliinae		*Tamsia hieroglyphica*	Yes	15,598	22	20	80
42	RZ57	Toxocampinae		*Lygephila maxima*	Yes	15,591	22	19.3	80.7
43	RZ111	Unassigned		*Platyjionia mediorufa*	Yes	15,329	22	19.9	80.1
44	RZ119	Unassigned		*Schistorhynx argentistriga*	Yes*	16,660	27	19.5	80.5
45	RZ265	Unassigned		*Rhesala imparata*	Yes	15,583	22	18.4	81.6
46	RZ4	Boletobiinae	Phytometrini	*Colobochyla salicalis*	Yes	16,449	22	18.5	81.5
47	RZ93	Unassigned		*Epitausa dilina*	Yes	15,440	22	18.7	81.3

The column 'circular' states whether the result of Novoplasty was a circular genome (yes) or a linear one which we manually circularized (yes*) or not (no). Length is in base pair (bp). #tRNA is the number of tRNA recognized by MITOS. ** this genome was manually circularized, and bordering the overlapping region 2 tRNAs were repeated.

deep branches that form the unresolved part of the tree for this family in published phylogenetic hypotheses. We also downloaded all the available Erebidae mitochondrial genomes from GenBank (37 genomes, Table 2), as well as mined the mitochondrial protein coding genes from five publicly available transcriptomes (Table 2). As outgroups, we used 17 taxa, consisting of 10 Noctuidae, 3 Notodontidae, 3 Nolidae and 1 Euteliidae (Table 3).

2.2 | Library preparation and sequencing

In this study old DNA extracts, obtained over 10 years ago, were used to generate libraries following the protocol of Twort et al. (2021). DNA quality was checked using electrophoretic agarose gels and high molecular weight samples were sonicated to approx. 200–300 bp fragments using a Bioruptor®. Fragmented DNA was blunt-end repaired with T4 Polynucleotide Kinase (New England Biolabs), followed by a reaction clean up with the MinElute purification kit (Qiagen). This was followed by adapter ligation, reaction purification and adapter fill in. The resulting reactions were then indexed using unique dual indexes. Indexing PCR was carried out in six independent reactions to avoid amplification bias, with 15 cycles b for each reaction. Indexing PCR reactions were pooled prior to magnetic bead clean up with Sera-Mag SpeedBeads™ carboxylate-modified hydrophilic (Sigma-Aldrich). An initial bead concentration of 0.5× was used to remove long fragments that are likely to represent contamination from fresh DNA and libraries were selected with a bead concentration of 1.8× to size select the expected library range of ~300 bp. The resulting libraries were quantified and quality checked with Quanti-iT™ PicoGreen™ dsDNA assay and with Bioanalyzer 2100, respectively. The final indexed libraries were pooled together prior to sequencing on an Illumina Novaseq platform (one lane, 2x150 bp, S4 flow cell) at Swedish National Genomics Institute (NGI) in Stockholm.

2.3 | MtGenome assembly

In order to assemble the mitochondrial genomes (de novo) we have used Novoplasty (Dierckxsens et al., 2016) on the newly sequenced samples. As stated in the manual for Novoplasty, raw uncleaned read files were used with kmer = 21. This approach gave a clean circular genome in 34 samples (72%). For an additional 5 samples (11%) it was sufficient to manually circularize in Geneious 10.2.6 (Kearse et al., 2012). The remaining 8 samples (17%) did not result in an assembled mitogenome using this approach probably due to their lower depth of sequencing. For these remaining samples, we used Prinseq 0.20.4 (Schmieder & Edwards, 2011) to remove the reads with ambiguous bases. We then cleaned the reads to remove low quality bases from the beginning (LEADING: 3) and end (TRAILING: 3) and reads less than 30 bp in length in Trimmomatic 0.38 (Bolger et al., 2014). Quality was measured for sliding windows of 4 bp and had to be greater than

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 List of the Erebidae samples retrieved from other studies and their GenBank accession number (GB)

#	Subfamily	Tribe	Species	GB
1	Aganainae		*Asota plana*	KJ173908
2	Arctiinae	Arctiini	*Callimorpha dominula*	NC_027094
3	Arctiinae	Arctiini	*Hyphantria cunea*	NC_014058
4	Arctiinae	Arctiini	*Lemyra melli*	NC_026692
5	Arctiinae	Arctiini	*Nyctemera arctata*	KM244681
6	Arctiinae	Arctiini	*Vamuna virilis*	NC_026844
7	Arctiinae	Arctiini	*Spilarctia subcarnea*	KT258909
8	Arctiinae	Arctiini	*Spilarctia alba*	KX753670
9	Arctiinae	Arctiini	*Aglaomorpha histrio*	KY800518
10	Arctiinae	Arctiini	*Arctia plantaginis*	ERR1856313*
11	Arctiinae	Lithosiini	*Cyana sp*	KM244679
12	Arctiinae	Lithosiini	*Paraona staudingeri*	KY827330
13	Arctiinae	Lithosiini	*Eilema ussuricum*	MN696172
14	Arctiinae	Syntomini	*Amata formosae*	NC_021416
15	Calpinae	Calpini	*Oraesia emarginata*	SRR5128005*
16	Calpinae	Ophiderini	*Eudocima salaminia*	SRR1300148*
17	Calpinae		*Paragabara curvicornuta*	KT362742
18	Erebinae	Catocalini	*Catocala sp*	KJ432280
19	Herminiinae		*Hydrillodes lentalis*	MH013484
20	Lymantriinae	Lymantriini	*Lachana alpherakii*	KJ957168
21	Lymantriinae	Lymantriini	*Lymantria umbrosa*	KY923066
22	Lymantriinae	Lymantriini	*Lymantria dispar*	KY923067
23	Lymantriinae	Lymantriini	*Lymantria albescens*	MH388823
24	Lymantriinae	Lymantriini	*Lymantria mathura*	MH388824
25	Lymantriinae	Lymantriini	*Lymantria monacha*	MH388825
26	Lymantriinae	Lymantriini	*Lymantria postalba*	MH388826
27	Lymantriinae	Lymantriini	*Lymantria xylina*	MH388827
28	Lymantriinae	Lymantriini	*Lymantria sugii*	MT265380
29	Lymantriinae	Lymantriini	*Lymantria dispar*	SRR1021618*
30	Lymantriinae	Nygmiini	*Euproctis pseudoconspersa*	NC_027145
31	Lymantriinae	Nygmiini	*Euproctis similis*	KT258910
32	Lymantriinae	Nygmiini	*Euproctis cryptosticta*	KY996558
33	Lymantriinae	Nygmiini	*Somena scintillans*	MH051839
34	Lymantriinae	Nygmiini	*Euproctis seitzi*	MN916588
35	Lymantriinae	Nygmiini	*Euproctis chrysorrhoea*	SRR1040496*
36	Lymantriinae	Orgyiini	*Gynaephora menyuanensis*	NC_020342
37	Lymantriinae	Orgyiini	*Gynaephora aureata*	KJ507132
38	Lymantriinae	Orgyiini	*Gynaephora qinghaiensis*	KJ507133
39	Lymantriinae	Orgyiini	*Gynaephora qumalaiensis*	KJ507134
40	Lymantriinae	Orgyiini	*Gynaephora ruoergensis*	KY688083
41	Lymantriinae	Orgyiini	*Gynaephora jiuzhiensis*	KY688085
42	Lymantriinae	Orgyiini	*Gynaephora minora*	KY688086

Transcriptomic data used for mining of protein coding genes are marked with an asterisk (*).

PHRED 25 on average. We then used the mirabait option in MIRA 4.0.2 (Chevreux et al., 1999, 2004) to find reads that corresponded to mitochondrial DNA. Mitochondrial reads were de novo assembled using three approaches, the Geneious de novo assembler, SPAdes assembler 3.10.0 (Nurk et al., 2013) and plasmidSPAdes (Antipov et al., 2016), all of which are implemented in Geneious. For each sample, all contigs over 500 bp were aligned

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 List of the outgroups used in this study and their GenBank accession number (GB)

#	Family	Species	GB
1	Euteliidae	*Anigraea rubida*	SRR1299755*
2	Noctuidae	*Mythimna separata*	NC_023118
3	Noctuidae	*Sesamia inferens*	NC_015835
4	Noctuidae	*Helicoverpa zea*	SRX371342*
5	Noctuidae	*Agrotis segetum*	SRR1231960*
6	Noctuidae	*Athetis lepigone*	SRR796575*
7	Noctuidae	*Trichoplusia ni*	NC_045936
8	Noctuidae	*Helicoverpa armigera*	SRR1565435*
9	Noctuidae	*Chrysodeixis includens*	SRR2049082*
10	Noctuidae	*Heliothis subflexa*	ERR738599*
11	Noctuidae	*Mythimna separata*	SRR5115697*
12	Nolidae	*Gabala argentata*	NC_026842
13	Nolidae	*Risoba prominens*	NC_026841
14	Nolidae	*Manoba major*	SRR1300145*
15	Notodontidae	*Ochrogaster lunifer*	NC_011128
16	Notodontidae	*Phalera flavescens*	NC_016067
17	Notodontidae	*Notoplusia minuta*	SRR1299746*

Transcriptomic data used for mining of protein coding genes are marked with an asterisk (*).

to a reference mitochondrial genome of *Lymantria dispar* (Erebidae). A consensus sequence of aligned contigs was then used to reference map mitochondrial reads in Bowtie2 (Langmead & Salzberg, 2012) as implemented in Geneious with default parameters. All the resulting assembled genomes were annotated using MITOS (Bernt et al., 2013). The COI gene was extracted from all the assembled genomes to compare with the sequences obtained with Sanger sequencing as an extra quality control.

2.4 | Phylogenetic analyses

Eleven protein coding genes (PCG) were extracted from all mitochondrial genomes. This data set includes the genes coding for ATP synthase membrane subunit 6 (*ATP6*), cytochrome *c* oxidase subunit I to III (*COI-III*), cytochrome b (*Cytb*), NADH dehydrogenase 1 to 5 (*ND1* - *ND5*) and the NADH–ubiquinone oxidoreductase chain 4 L (*ND4L*). We excluded two genes (*ATP8* and *ND6*) from our data set as their alignments were ambiguous with many indels. Each gene was aligned separately using MAFFT v7.450 (Katoh, 2002; Katoh & Standley, 2013) as implemented in Geneious with default options. Sequences were curated in VoSeq (Peña & Malm, 2012), after revision and manual correction of the alignments. Using the VoSeq database application, we created a concatenated nucleotide data set (nt) with a total length of 10,245 bp and an amino acid data set (aa) of 3415 characters.

We ran maximum likelihood (ML) analyses with both nt (partitioned by gene and codon position) and aa (partitioned by gene) data sets using IQ-TREE 2.0.6 (Nguyen et al., 2015). In both analyses the best substitution model and partitioning scheme was selected by ModelFinder (Kalyaanamoorthy et al., 2017) with '-m MFP + MERGE' option. We evaluated branch supports with ultrafast bootstrap approximations (UFBoot2, 5000 reps) and SH-like approximate likelihood ratio test (1000 reps) (Guindon et al., 2010; Hoang et al., 2018) using the '-B 5000 -alrt 1000' option. We used the '-bnni' option to reduce the risk of overestimating branch supports in ultrafast bootstrap approximation. In addition, we tested the effect of the third codon position by removing it from the data set and repeating the same analysis in IQ-TREE. Additionally, we tested partitioning scheme for the nucleotide data set partitioned by gene only, in PartitionFinder2 (Lanfear et al., 2017). In this analysis we limited the tested models with the option 'models = mrbayes'. The obtained partitioning scheme was used to perform a Bayesian phylogenetic analysis in MrBayes 3.2.7 (Ronquist et al., 2012). This analysis ran for two independent runs of 10^7 generations sampling every 10^3 steps. This analysis was repeated five times. Convergence of the runs was checked in Tracer 1.7.1 (Rambaut et al., 2018). Resulting trees were visualized in FigTree v1.4.3 (Rambaut, 2018). In order to evaluate the effect of possible rogue samples in the phylogenetic analysis, RogueNaRok (Aberer et al., 2013) was used with a threshold of 90. After the detection of the rogue samples, they were deleted from the database and new trees were constructed in IQ-TREE to compare with the original tree.

Mira and Novoplasty were run using the resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2018-8-347. The software PartitionFinder2 and MrBayes were run using the CIPRES Science Gateway infrastructures (Miller et al., 2010). The raw whole genome data is deposited in GenBank under the BioProject number PRJNA702831. All data in the supplementary materials, including the alignment, annotated genomes and tree results, can be found on the GitHub repository: github.com/Hamidhrg/ErebidMtGenome.

3 | RESULTS

From the total number of 47 obtained genomes, 34 were fully assembled as circularized genomes. For base

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

frequency and basic genome compositions we only focus on the 34 good quality genomes. They varied in length from 15,198 bp in *Alesua etialis* (Rivulinae) to 16,449 bp in *Colobochyla salicalis* (Boletobiinae, Phytometrini). AT base frequency ranged between 78.6% in *Arctornis* sp. (Lymantriinae) and 82.3% in *Alesua*. Their tRNA number was between 19 in *Alesua* to 25 in *Acantholipes circumdata* (Erebinae; Table 1). The annotated genomes are available through our online GitHub repository.

ModelFinder in IQ-TREE merged the 33 possible partitions of the nucleotide data set into 13 and found corresponding best substitution models (Table 4). Partition sizes ranged between 96 and 1411 bp (788 bp mean partition size). The data set included 4789 parsimony informative sites.

Maximum likelihood analysis of the nt data set resulted in the best resolved tree (Figure 1). The family Erebidae was a well-supported, monophyletic group. All other families, used as outgroups, were also recovered as monophyletic with relatively high support. Within Erebidae, most of subfamilies with more than one representative were found to be monophyletic, including the Lymantriinae, Arctiinae and Erebinae. Several subfamilies did not form monophyletic groups, including the Pangraptinae and Aganainae. The result of RogueNaRok run found four rogue samples: RZ268 (*Mecodina praecipua*), RZ93 (*Epitausa dilina*), RZ119 (*Schistorhynx argentistriga*) and RZ105 (*Hypocala deflorate*). The deletion of mentioned samples from the data set improved slightly the support values mainly at the more terminal branches, but did not have any effect on the shape of the tree or the (lack of) support in the internal deep branches. The result of analysing the data set without the third codon position did not improve significantly any of the branch support values or relationships (Supplementary Material).

In contrast, the ML analysis of the aa data set resulted in very anomalous trees (Supplementary Material). First, it appeared very sensitive to missing data; therefore, three samples with the highest amount of missing data (*Lymantria monacha* from the ingroup, and *Thaumetopoea pityocampa* and *Agrotis segetum* from the outgroup) were removed and a new analysis was run. The resulting tree improved slightly; however, it was still very anomalous. For example, almost no subfamily was monophyletic, including such well-defined subfamilies as Arctiinae and Lymantriinae. Bayesian inference (BI) using the nt data partitioned according to the PartitionFinder2 analysis, all 10 chains (five runs of 2 independent chains) reached stationary phase, but none of the runs converged with each other. The analysis was repeated for a longer (up to 10^8) generation number and with a higher swap temperature (up to temp = 0.7) resulting in the same issue.

4 | DISCUSSION

The most comprehensive study focused on the phylogenetic relationships of Erebidae up to date was published by Zahiri et al. (2012). Using seven nuclear and one mitochondrial markers (for a total of 6407 bp) it inferred a phylogenetic hypothesis with numerous unsupported, short branches that did not resolve relationships among the different subfamilies and tribes. Similarly, we find that mitochondrial genomic data were not able to resolve the

TABLE 4 ModelFinder best partitioning scheme

Partition	Markers	Length (bp)	Infor	Invar	Model
1	ATP6_pos1, COII_pos1, COIII_pos1, CytB_pos1	1083	397	592	GTR + F + R7
2	ATP6_pos2, COI_pos2, COII_pos2	965	139	758	GTR + F + R7
3	ATP6_pos3	227	197	15	TPM2 + F + I + G4
4	COI_pos1	510	130	345	GTR + F + I + G4
5	COI_pos3, COII_pos3, ND3_pos3	847	580	227	TIM + F + R5
6	COIII_pos2, CytB_pos2, ND2_pos2, ND3_pos2	1039	218	705	TVM + F + R4
7	COIII_pos3, CytB_pos3	628	572	29	TPM3 + F + R7
8	ND1_pos1, ND4_pos1, ND4L_pos1, ND5_pos1	1411	599	635	GTR + F + R5
9	ND1_pos2, ND4_pos2, ND4L_pos2, ND5_pos2	1411	310	950	GTR + F + R5
10	ND1_pos3, ND4_pos3, ND5_pos3	1315	1090	93	TIM + F + R7
11	ND2_pos1, ND3_pos1	411	210	133	GTR + F + R5
12	ND2_pos3	302	269	13	GTR + F + I + G4
13	ND4L_pos3	96	78	4	GTR + F + I + G4
Total		10,245	4789	4499	

Length column corresponds to the total number of base pairs forming the partition. ‘Infor’ and ‘invar’ stand for number of informative and invariable sites, respectively.

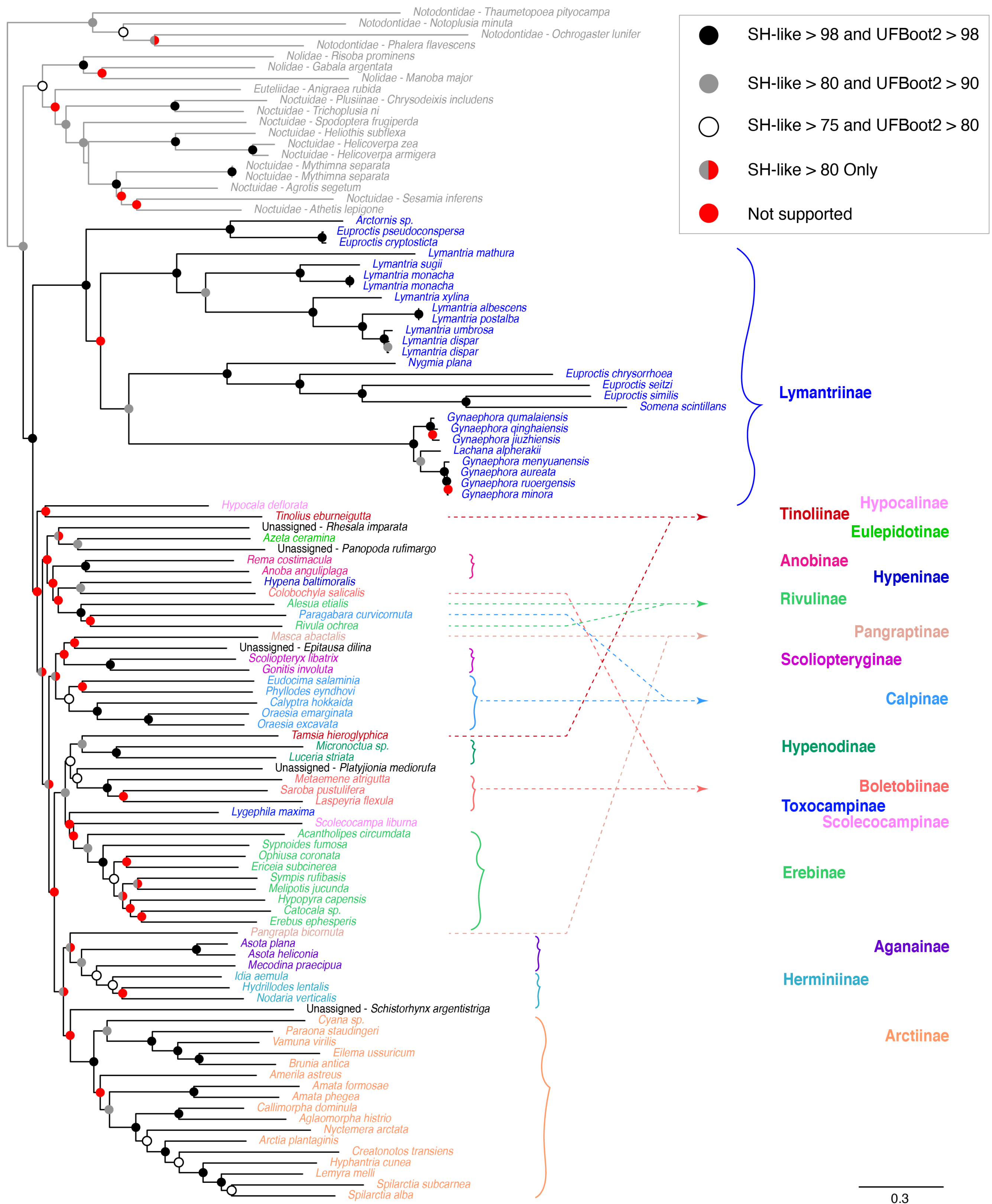

FIGURE 1 The ML tree obtained using the nt data set in IQ-TREE. The species names are coloured based on corresponding subfamilies. The subfamily names are placed in front of the corresponding clade. The cases where members of a subfamily were placed in different parts of the tree, an arrow was drawn to point it out. Black circles represent highly supported nodes, grey supported nodes, white low support and red not supported nodes. The outgroup lineages are coloured in grey.

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relationships of subfamilies with any confidence. We do find that the family itself is a strongly supported monophyletic group. As earlier studies have found (Regier et al., 2017; Zahiri et al., 2011), our results also showed that the relationships among the other four lineages (Notodontidae, Nolidae, Euteliidae and Noctuidae) are not clear, although they do form a monophyletic assemblage with good support. Within the quadrifid noctuoids (Noctuidae, Nolidae, Euteliidae and Erebidae), the sister group of Erebidae remains unresolved. Our phylogeny found Erebidae to be sister to the other quadrifids (Nolidae, Euteliidae and Noctuidae), however, (Zahiri et al., 2012) found Euteliidae+Noctuidae in this position in their ML analysis and Noctuidae as sister to Erebidae in the parsimony analysis.

We find that the subfamily Lymantriinae is sister to the rest of Erebidae; however, this position has no support. Zahiri et al. (2012) did not find Lymantriinae in the same position but also in that study its position is not supported. Fibiger & Lafontaine (2005) placed Lymantriinae sister to Arctiinae to reflect the close association found by Mitchell et al. (1997, 2000). They also noted that arctiines and lymantriines are not basal clades but appear to be highly specialized lineages derived from within Erebidae. Lymantriinae like Herminiinae, Aganainae and Arctiinae share a unique apomorphic character—a prespiracular counter-tympanal hood—that had been interpreted as the plesiomorphic condition in quadrifid Noctuoidea. Our results do not support such a relationship. Branch lengths within the Lymantriinae clade appear to be longer than in the rest of the tree (Figure 1). This pattern of exceptionally long branch lengths among Lymantriinae was also observed by Zahiri et al. (2012). One reason for this could be a higher rate of molecular evolution within Lymantriinae, although the reasons for a higher rate are not known at the moment. Support values in this clade appear high, but it is clear that the high support values correspond to relationships within genera and not between different genera. Wang et al. (2015) studied this subfamily using eight molecular markers, and they found that relationships between tribes are similarly poorly supported. We did not include the tribe Daplasini which Wang et al. (2015) found to be sister to the rest of the subfamily. With our taxon sampling, we found the tribe Arctornithini to be sister to the rest of the included Lymantriinae taxa, a result, which is in concordance with Wang et al. (2015).

We find Pangraptinae to be polyphyletic with *Pangrapta bicornuta* as sister to Aganainae and Herminiinae and *Masca abactalis* in a clade with Scoliopteryginae and Calpinae (Figure 1). In Zahiri et al. (2012) *Masca* is placed as sister to the rest of Pangraptinae, which, in turn, is sister to Aganainae + Herminiinae + Arctiinae. In our study, the clade (*Pangrapta* + Aganainae + Herminiinae) is weakly associated with an unsupported clade consisting of the enigmatic genus *Schistorhynx* Hampson, and Arctiinae. Zahiri et al. (2012) placed Pangraptinae within a group of subfamilies with prespiracular counter-tympanal hoods, although morphological examinations of various pangraptine genera revealed that they have a typical erebine post-spiracular hood. Zahiri et al. (2012) concluded that the prespiracular feature is either the result of convergent evolution between Lymantriinae and the clade comprising Herminiinae, Aganainae and Arctiinae, or a feature of the larger clade that encompasses all these groups, with subsequent reversal in Pangraptinae. Our results suggest independent convergent evolution of prespiracular counter-tympanal hoods in Lymantriinae and a clade of four subfamilies Pangraptinae, Herminiinae, Aganainae, Arctiinae with subsequent reversal in Pangraptinae. We find *Mecodina* to be sister to Herminiinae, instead of Aganainae as in Zahiri et al. (2012). The position of *Mecodina* is not supported in either our study or Zahiri et al. (2012), although the taxon does seem to be part of the Aganainae+Herminiinae clade with good support.

The relationships within Arctiinae are better supported. The clade composed of *Cyana* sp., *Paraona staudingeri*, *Vamuna virilis*, *Eilema ussuricum* and *Brunia antica*, representing the tribe Lithosiini, is sister to the rest of the subfamily. This position of Lithosiini is in concordance with previously published studies (Dowdy et al., 2020; Rönkä et al., 2016; Zahiri et al., 2012; Zaspel et al., 2014). Also, the position of *Amerila astreus* (Amerilini), even though it is not significantly supported, and the relationships of the Callimorphina and Arctiinae subtribes are similar to prior studies.

Erebinae was recovered as monophyletic, however, within the subfamily there is a lack of support for the resolution of the relationships among different tribes and genera. The placement of *Acantholipes circumdata* (Acantholipini) as sister to the rest of the subfamily was also recovered by Zahiri et al. (2012) and Homziak et al. (2019). Homziak et al. (2019) used anchored hybrid enrichment (AHE) phylogenomics to resolve deep relationships within this subfamily. The position of *Sypnoides fumosa* (Sypnini) in Erebinae is in concordance with Zahiri et al. (2012) and Homziak et al. (2019). The other relationships within the subfamily are poorly resolved and do not agree with the prior studies.

The calpine genus *Paragabara* is found within Rivulinae with relatively good support, making Calpinae polyphyletic and Rivulinae paraphyletic (Figure 1). *Paragabara* has not been included previously in a phylogenetic analysis and has no nuclear gene sequences available. It is possible that it has been misplaced previously in Calpinae and should thus be moved to Rivulinae, rendering both subfamilies

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

monophyletic. Within Calpinae (with *Paragabara* removed), our analysis supported the monophyly of Calpini, containing the type genus *Calyptra* Ochsenheimer and *Oraesia* Guenée. This clade is placed sister to a weakly supported clade consisting of genera assigned to Ophiderini (*Eudocima* Billberg) and Phyllodini (*Phyllodes* Boisduval), corroborating the results of Zaspel et al. (2012).

We find the subfamily Tinoliinae to be polyphyletic, with *Tinolius* being sister to Hypocalinae and *Tamsia* belonging to a clade with Hypenodinae and Boletobiinae (Figure 1). Zahiri et al. (2012) found a monophyletic Tinoliinae, but with no support.

We included a number of taxa that are currently unassigned to subfamily (Figure 1). In general, mitochondrial genome data do not help us resolve the phylogenetic positions of these taxa. One case was similar to the results of Zahiri et al. (2012), *Colobochyla* Hübner (represented by the type species) grouped as sister to Hypeninae with good support.

The remaining subfamilies with more than one representative, that is Boletobiinae, Scoliopteryginae, Hypenodinae and Anobinae are recovered as monophyletic entities. However, the interrelationships of subfamilies are either weakly or not supported in our analysis. Relationships among subfamilies have not been resolved in any phylogenetic work up to now. Zahiri et al. (2012) suggested that the short internal branches connecting different subfamilies and some tribes are potentially due to a rapid radiation. Therefore, more data and a more comprehensive taxon sampling are needed in order to resolve these relationships. The results of this study show very low support values for these internal nodes suggesting that the amount of information coded in the mitochondrial genome is not sufficient to deal with such rapid radiations of similar or older ages. AHE approaches have been successful within subfamilies (Dowdy et al., 2020; Homziak et al., 2019) and would likely be sufficient to resolve the relationships of the subfamilies. On the other hand, whole genome shotgun approaches are becoming more feasible and affordable (e.g., Twort et al., 2021 for Lepidoptera), suggesting an alternative avenue of gathering large amounts of data. We are currently investigating such a route by extracting nuclear genes of interest from the raw data associated with the current study (Ghanavi et al., in prep.).

One of the caveats of our study is the sporadic taxon sampling in our data set. Although our data set has low taxon sampling, it is still comparable to most multi-locus phylogenetic studies in species diversity and definitely larger than most phylogenomic data sets. Hence, we believe that expanding the taxon sampling will improve phylogenetic resolution. Nevertheless, it is probable, that divergence events are resolvable at species level and perhaps in some cases at the genus level, as demonstrated in the better sampled clades in our study (e.g., Arctiinae and Lymantriinae).

In the data set studied here, the amount of phylogenetic signal coded in the mitochondrial genome was not sufficient to resolve satisfactorily the relationships among the representatives of the Erebidae family. This is especially visible at deeper nodes (Figure 1). The lack of resolution for deep, short branches suggest that on one hand, this data set has relatively high mutation rates, which cause saturation issues for deep relationships, and the amount of signal in mitochondrial genome, seems inadequate to recover deep, short branches.

5 | CONCLUSION

Advances in sequencing technologies and bioinformatics support, have revolutionized molecular systematics, evolutionary biology and phylogenomics. Sequencing a large number of mitochondrial genomes is relatively cheap and does not need extra infrastructure beyond traditional PCR labs. This has allowed a rise in the number of new mitochondrial genomes being published practically on a weekly basis in the last few years. These short publications usually publish a single new mitochondrial genome together with a very brief and rudimentary phylogenetic analysis. Here, we show that it is relatively easy to increase the number of taxa sequenced for a single study by sequencing 47 specimens using whole genome shotgun methods from DNA extractions used in Sanger sequencing studies previously.

However, in this study, we question the utility of mitochondrial genome data to resolve deep phylogenetic relationships accurately or to resolve relationships of rapid radiation events in Erebidae. Based on our findings, mitochondrial genomes are not sufficient to resolve erebid relationships within and between subfamilies. Relationships between close tribes could potentially be studied with a dense enough taxon sampling in Erebidae. We also show that it is clear that amino acid data sets based on mitochondrial protein coding genes are not useful to study phylogenetic relationships at this level.

ACKNOWLEDGEMENTS

HG received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skldowska-Curie grant agreement No. 6422141. NW acknowledges funding from the Swedish Research Council (Grant No. 2015-04441). We thank Marko Mutanen for sending us the DNA extract for *Scoliopteryx libatrix*. Voucher specimens of many species are deposited in the Biological Museum, Lund University (10.15468/dahk2a).

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ORCID

Hamid Reza Ghanavi https://orcid.org/0000-0003-1029-4236

REFERENCES

Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Systematic Biology*, *62*, 162–166.

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., & Pevzner, P. A. (2016). plasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics*, *32*(22):3380–3387. https://doi.org/10.1093/bioinformatics/btw493

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution.*, *69*, 313–319.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Cameron, S. L. (2014). How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology.*, *39*, 400–411.

Cameron, S. L., Lambkin, C., Barker, S. C., & Whiting, M. F. (2007). A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology.*, *32*, 40–59.

Cameron, S. L., Miller, K. B., D'Haese, C. A., Whiting, M. F., & Barker, S. C. (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). *Cladistics*, *20*, 534–557.

Cameron, S. L., Sullivan, J., Song, H., Miller, K. B., & Whiting, M. F. (2009). A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zoologica Scripta*, *38*, 575–590.

Chen, L., Wahlberg, N., Liao, C. Q., Bin, W. C., Ma, F. Z., & Huang, G. H. (2020). Fourteen complete mitochondrial genomes of butterflies from the genus Lethe (lepidoptera, Nymphalidae, Satyrinae) with mitogenome-based phylogenetic analysis. *Genomics*, *112*, 4435–4441.

Chen, L. P., Zheng, F. Y., Bai, J., Wang, J. M., Lv, C. Y., Li, X., Zhi, Y. C., & Li, X. J. (2020). Comparative analysis of mitogenomes among six species of grasshoppers (orthoptera: Acridoidea: Catantopidae) and their phylogenetic implications in wing-type evolution. *International Journal of Biological Macromolecules*, *159*, 1062–1072.

Chen, M. M., Li, Y., Chen, M., Wang, H., Li, Q., Xia, R. X., Zeng, C. Y., Li, Y. P., Liu, Y. Q., & Qin, L. (2014). Complete mitochondrial genome of the atlas moth, Attacus atlas (lepidoptera: Saturniidae) and the phylogenetic relationship of Saturniidae species. *Gene*, *545*, 95–101.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E., Wetter, T., & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research*, *14*, 1147–1159.

Chevreux, B., Wetter, T., & Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology*, 45–56.

Dierckxsens, N., Mardulyn, P., & Smits, G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*, e18.

Dowdy, N. J., Keating, S., Lemmon, A. R., Lemmon, E. M., Conner, W. E., Scott Chialvo, C. H., Weller, S. J., Simmons, R. B., Sisson, M. S., & Zaspel, J. M. (2020). 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). *Systematic Entomology*, *45*, 874–893.

Dowton, M., Cameron, S. L., Austin, A. D., & Whiting, M. F. (2009). Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – A lineage with both rapidly and slowly evolving mitochondrial genomes. *Molecular Phylogenetics and Evolution.*, *52*, 512–519.

Fenn, J. D., Song, H., Cameron, S. L., & Whiting, M. F. (2008). A preliminary mitochondrial genome phylogeny of orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Molecular Phylogenetics and Evolution*, *49*, 59–68.

Fibiger, M., & Lafontaine, J. D. (2005). A review of the higher classification of the Noctuoidea (Lepidoptera) with special reference to the Holarctic fauna. *Esperiana*, *11*, 7–92.

Ghanavi, H. R., Twort, V., Yappar, E., Zahiri, R., Wahlberg, N. Phylogenomics of Erebidae (Lepidoptera): using old DNA extracts to resolve old phylogenetic questions with whole genome sequencing. Manuscript in preparation.

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. *Molecular Biology and Evolution*, *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. *Molecular Phylogenetics and Evolution*, *131*, 99–105.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589.

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*, 3059–3066.

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.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649.

Kim, M. J., Kang, A. R., Jeong, H. C., Kim, K.-G., & Kim, I. (2011). Reconstructing intraordinal relationships in lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, Spindasis takanonis and Protantigius superans (lepidoptera: Lycaenidae). *Molecular Phylogenetics and Evolution.*, *61*, 436–445.

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*, 772–773.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, *9*, 357–359.

Li, H., Shao, R., Song, N., Song, F., Jiang, P., Li, Z., & Cai, W. (2015). Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Scientific Reports*, *5*, 8527.

Li, Q., Wang, X., Chen, X., & Han, B. (2018). Complete mitochondrial genome of the tea looper caterpillar, Ectropis obliqua (lepidoptera: Geometridae) with a phylogenetic analysis of Geometridae. *International Journal of Biological Macromolecules*, *114*, 491–496.

Li, X., Yan, L., Pape, T., Gao, Y., & Zhang, D. (2020). Evolutionary insights into bot flies (Insecta: Diptera: Oestridae) from comparative analysis of the mitochondrial genomes. *International Journal of Biological Macromolecules.*, *149*, 371–380.

López-López, A., & Vogler, A. P. (2017). The mitogenome phylogeny of Adephaga (coleoptera). *Molecular Phylogenetics and Evolution*, *114*, 166–174.

Miller M.A., Pfeiffer W., Schwartz T. (2010). *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*. 2010 Gateway Computing Environments Workshop, GCE 2010.

Mitchell, A., Cho, S., Regier, J. C., Mitter, C., Poole, R. W., & Matthews, M. (1997). Phylogenetic utility of elongation factor-1″ in Noctuoidea (Insecta: Lepidoptera): the limits of synonymous substitution. *Molecular Biology and Evolution*, *14*, 381–390.

Mitchell, A., Mitter, C., & Regier, J. C. (2000). More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Systematic Biology*, *49*, 202–224.

Nardi, F. (2003). Hexapod origins: Monophyletic or paraphyletic? *Science*, *299*, 1887–1889.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel, S. R., Woyke, T., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2013). In M. Deng, R. Jiang, F. Sun, & X. Zhang (Eds.), *Assembling genomes and mini-metagenomes from highly chimeric reads* (pp. 158–170). Springer Berlin Heidelberg.

Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: The mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27*, 1659–1672.

Peña, C., & Malm, T. (2012). VoSeq: A voucher and DNA sequence web application. *PLoS One*, *7*, e39071.

Rambaut A. (2018). FigTree v1.4.4. Accessed July 13, 2019. https://github.com/rambaut/figtree/releases

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, *67*, 901–904.

Regier, J. C., Mitter, C., Mitter, K. T., Cummings, M. P., Bazinet, A. L., Hallwachs, W., Janzen, D. H., & Zwick, A. (2017). Further progress on the phylogeny of Noctuoidea (Insecta: Lepidoptera) using an expanded gene sample. *Systematic Entomology*, *42*, 82–93.

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.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542.

Rubinoff, D., Cameron, S., & Will, K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity*, *97*, 581–594.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*, 863–864.

Simon, S., & Hadrys, H. (2013). A comparative analysis of complete mitochondrial genomes among Hexapoda. *Molecular Phylogenetics and Evolution*, *69*, 393–403.

Song, F., Li, H., Jiang, P., Zhou, X., Liu, J., Sun, C., Vogler, A. P., & Cai, W. (2016). Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. *Genome Biology and Evolution*, *8*, 1411–1426.

Sperling, F. (2019). CHAPTER 20. Butterfly Molecular Systematics: From Species Definitions to Higher-Level Phylogenies. In C. L. Boggs, W. B. Watt, & P. R. Ehrlich (Eds.), *Butterflies: Ecology and evolution taking flight* (pp. 431–458). University of Chicago Press. https://doi.org/10.7208/9780226063195-023

Talavera, G., & Vila, R. (2011). What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evolutionary Biology*, *11*, 315.

Timmermans, M. J. T. N., Lees, D. C., & Simonsen, T. J. (2014). Towards a mitogenomic phylogeny of lepidoptera. *Molecular Phylogenetics and Evolution*, *79*, 169–178.

Twort, V. G., Minet, J., Wheat, C. W., & Wahlberg, N. (2021). Museomics of a rare taxon: Placing Whalleyanidae in the lepidoptera tree of life. *Systematic Entomology*, *46*, 926–937.

van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., Mitter, C., Mutanen, M., Regier, J. C., Simonsen, T. J., Wahlberg, N., Yen, S.-H., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B. Å., Brown, J. W., Bucheli, S. R., … Andreas, Z. (2011). Order Lepidoptera Linnaeus, 1758. In Z.-Q. Zhang (Ed.), *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness* (pp. 212–221). Zootaxa.

Wang, H., Wahlberg, N., Holloway, J. D., Bergsten, J., Fan, X., Janzen, D. H., Hallwachs, W., Wen, L., Wang, M., & Nylin, S. (2015). Molecular phylogeny of Lymantriinae (lepidoptera, Noctuoidea, Erebidae) inferred from eight gene regions. *Cladistics*, *31*, 579–592.

Xu, H., Wu, Y., Wang, Y., & Liu, Z. (2020). Comparative analysis of five mitogenomes of Osmylinae (Neuroptera: Osmylidae) and their phylogenetic implications. *International Journal of Biological Macromolecules*, *164*, 447–455.

Yang, M., Song, L., Shi, Y., Li, J., Zhang, Y., & Song, N. (2019). The first mitochondrial genome of the family Epicopeiidae and higher-level phylogeny of Macroheterocera (lepidoptera: Ditrysia). *International Journal of Biological Macromolecules*, *136*, 123–132.

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Yang, X., Cameron, S. L., Lees, D. C., Xue, D., & Han, H. (2015). A mitochondrial genome phylogeny of owlet moths (lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. *Molecular Phylogenetics and Evolution*, *85*, 230–237.

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., 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.

Zardoya, R., & Meyer, A. (1996). Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, *13*, 933–942.

Zaspel, J., Zahiri, R., Hoy, M., Janzen, D., Weller, S., & Wahlberg, N. (2012). A molecular phylogenetic analysis of the vampire moths and their fruit-piercing relatives (lepidoptera: Erebidae: Calpinae). *Molecular Phylogenetics and Evolution*, *65*, 786–791.

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*, 1–10.

Zhang, Z., Xing, Y., Cheng, J., Pan, D., Lv, L., Cumberlidge, N., & Sun, H. (2020). Phylogenetic implications of mitogenome rearrangements in east Asian potamiscine freshwater crabs (Brachyura: Potamidae). *Molecular Phylogenetics and Evolution.*, *143*, 106669.

SUPPORTING INFORMATION

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

How to cite this article: Ghanavi, H. R., Twort, V., Hartman, T. J., Zahiri, R., & Wahlberg, N. (2022). The (non) accuracy of mitochondrial genomes for family-level phylogenetics in Erebidae (Lepidoptera). *Zoologica Scripta*, *51*, 695–707. https://doi.org/10.1111/zsc.12559

14636409, 2022, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/zsc.12559 by University Of Helsinki, Wiley Online Library on [16/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License