Using big data to understand evolutionary patterns in Geometridae and Erebidae, the two most diverse families of Lepidoptera

Ghanavi, Hamid

2020

Using big data to understand evolutionary patterns in Geometridae and Erebidae, the two most diverse families of Lepidoptera

HAMID REZA GHANAVI
DEPARTMENT OF BIOLOGY | FACULTY OF SCIENCE | LUND UNIVERSITY

List of papers

I. Murillo-Ramos L, Brehm G, Sihvonen P, Hausmann A, Holm S, **Ghanavi HR**, Õunap E, Truuverk A, Staude H, Friedrich E, Tammaru T, Wahlberg N. 2019. A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies. PeerJ 7:e7386 DOI 10.7717/peerj.7386

II. **Ghanavi HR**, Chazot N, Sanmartín I, Murillo-Ramos L, Duchêne S, Sihvonen P, Brehm G, Wahlberg N. 2020. Biogeography and Diversification Dynamics of the Moth Family Geometridae (Lepidoptera). Manuscript.

III. **Ghanavi HR**, Twort V, Zahiri R, Wahlberg N. 2020. Phylogenomics of Erebidae (Lepidoptera): using old DNA extracts to resolve old phylogenetic questions with whole genome sequencing. Manuscript.

IV. **Ghanavi HR**, Twort V, Hartman T, Zahiri R, Wahlberg N. 2020. The accuracy of mitochondrial genomes for family level phylogenies, the case of erebid moths (Lepidoptera; Erebidae). Manuscript.

V. **Ghanavi HR**, Twort V, Duplouy A. 2020. Exploring nonspecific/bycatch diversity of organisms in whole genome sequencing of Erebidae moths (Lepidoptera). Manuscript.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Science
Department of Biology

ISBN 978-91-7895-642-5

Using big data to understand evolutionary patterns in Geometridae and Erebidae, the two most diverse families of Lepidoptera

Hamid Reza Ghanavi

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at Blå hallen, Ekologihuset, Sölvegatan 37, Lund. On Friday 9th October 2020, at 13h00.

Faculty opponent
Dr Dimitar Dimitrov

Uviversity Museum of Bergen, Norway

Organization
LUND UNIVERSITY

Department of Biology
Systematic Biology Group

Ecology Building
SE-22362, Lund.

Author: Hamid Reza Ghanavi

Document name Doctoral Dissertation

Date of issue October 2020

Sponsoring organization: This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Slodowska-Curie grant agreement No. 6422141

Title and subtitle Using big data to understand evolutionary patterns in Geometridae and Erebidae, the two most diverse families of Lepidoptera

Abstract

Lepidoptera (moths and butterflies) are one of the most diverse groups of organisms on earth. They have conquered all the continents apart from Antarctica. The reasons of such high diversity are still not clear. One of the first steps to study the causes of such evolutionary success is to have a clear idea of their phylogenetic relationships. In this thesis I focus on the diversity of two of the most diverse families of Lepidoptera, Geometridae with over 23,000 described species and Erebidae with over 24,000 species. In the case of Geometridae I focus in obtaining a robust phylogenetic hypothesis and then study the diversification dynamics and the biogeographical history which have shaped their actual diversity and distribution. In this project I used the most complete dataset of the family in order to study their evolutionary patterns. In the case of the Erebidae family, obtaining a robust phylogenetic hypothesis was more challenging. In the most complete study of the group up to date, using a multi locus Sanger based approach, it was not possible to resolve the deep phylogeny of the family. Therefore, I used high throughput sequencing (HTS) approaches to resolve the complex deep phylogenetic history of the group. In this project I used old DNA extracts of over 10 years old to explore the possibility of using this genetic resource for genomic studies. In addition, I evaluated the accuracy and range of resolution of the mitochondrial genomes in this family. And finally, I explored the alternative possibilities which the HTS approaches offer us to study the presence of symbiotic interactions using genomic data.

Key words
Lepidoptera, Megadiverse, Diversification, Biogeography, Phylogenomics, Geometridae, Erebidae

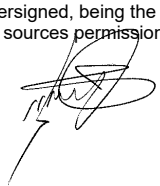
Classification system and/or index terms (if any)

Supplementary bibliographical information — **Language** English

ISSN and key title — **ISBN** 978-91-7895-642-5

Recipient's notes — **Number of pages** 174 — Price

Security classification

Signature 

Date 2020-09-01

Using big data to understand evolutionary patterns in Geometridae and Erebidae, the two most diverse families of Lepidoptera

Hamid Reza Ghanavi

LUND
UNIVERSITY

Cover design by Zahra Saberi

Faculty of Science
Department of Biology

ISBN 978-91-7895-642-5 (print)
ISBN 978-91-7895-643-2 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2020

To everyone who taught me something.

Table of Contents

List of papers

I. Murillo-Ramos L, Brehm G, Sihvonen P, Hausmann A, Holm S, **Ghanavi HR**, Õunap E, Truuverk A, Staude H, Friedrich E, Tammaru T, Wahlberg N. 2019. A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies. PeerJ 7:e7386 DOI 10.7717/peerj.7386

II. **Ghanavi HR**, Chazot N, Sanmartín I, Murillo-Ramos L, Duchêne S, Sihvonen P, Brehm G, Wahlberg N. 2020. Biogeography and Diversification Dynamics of the Moth Family Geometridae (Lepidoptera). Manuscript.

III. **Ghanavi HR**, Twort V, Zahiri R, Wahlberg N. 2020. Phylogenomics of Erebidae (Lepidoptera): using old DNA extracts to resolve old phylogenetic questions with whole genome sequencing. Manuscript.

IV. **Ghanavi HR**, Twort V, Hartman T, Zahiri R, Wahlberg N. 2020. The accuracy of mitochondrial genomes for family level phylogenies, the case of erebid moths (Lepidoptera; Erebidae). Manuscript.

V. **Ghanavi HR**, Twort V, Duplouy A. 2020. Exploring non-specific/bycatch diversity of organisms in whole genome sequencing of Erebidae moths (Lepidoptera). Manuscript.

Autor contribution

I. Contributed in the study conception. Contributed in the study design. Contributed to the data collection. Limited contribution to the analyses. Limited contribution to the manuscript preparation.

II. Contributed in the study conception. Contributed in the study design. Collected the data. Performed majority of the analyses. Took the lead in manuscript preparation.

III. Contributed in the study conception. Contributed in the study design. Collected the data. Performed all the analyses. Took the lead in manuscript preparation.

IV. Contributed in the study conception. Contributed in the study design. Collected the data. Performed all the analyses. Took the lead in manuscript preparation.

V. Contributed in the study conception. Contributed in the study design. Collected the data. Limited contribution to the analyses. Took the lead in manuscript preparation.

Introduction

Background

Many biologists throughout history have reflected over the diversity of organisms on Earth and the processes which have shaped the actual diversity (Hooker 1854; Darwin 1859; Wallace 1869). Some groups of organisms are very species rich, while others contain only a few representatives. Of course, the amount of time passed since the last most recent common ancestor of a group plays a big role in the diversity. Some smaller groups appeared only very recently. But there are also many cases where different groups of the same age have very different diversities in reality.

The process during which the diversity of a group changes is called diversification. Diversification comprises both the proliferation (i.e. speciation) and decline (i.e. extinction) of diversity. The actual observed diversity is a fruit of both processes. Different groups of organisms have been affected by these processes dissimilarly in different time periods. What is clear is that speciation and extinction are both dynamic processes which are affected by many biotic and abiotic factors.

Nowadays various methods exist for the study of diversification rates. Some more traditional methods focus on fossils and the occurrence of them over time, to try to infer the diversity of different groups based on observed fossils (Simpson 1944; Stanley 1980). The use of these methods is limited, especially for groups which do not fossilize well or their fossil data is scarce. This is usually the case for many of the most diverse groups of organisms on Earth: Insects. It happens that there is limited fossil work within the scientific literature for insects generally, and particularly for Lepidoptera. To be able to study the diversification dynamics within Lepidoptera, or any other fossil-lacking group, one can primarily rely on methods that use molecular data.

The alternative approach to study diversification is based generally on hypotheses of evolutionary history: Phylogenies (Hey 1992; Nee et al. 1992; Purvis 2008). Inferred phylogenetic trees reveal the relationships between different taxa where the tips usually stand for extant species and each branch is a different evolutionary entity. Therefore, each branching event in a tree can be interpreted as a speciation event. The increase in molecular data (i.e. genetic sequence information), not only allows us to obtain more and better phylogenetic hypotheses each day, but most importantly

enables us to add the time dimension. Thereby, enabling us to date speciation events. Time calibrated phylogenies allow us to also infer diversification rates.

Big data and phylogenomics

The appearance of the term big data in recent years is highly correlated with the advances in computer science and therefore generation of databases which are not easily accessible. Due to complexities of accessing and using such databases the so-called big data field emerged to extract and simplify the information out of such databases. In parallel, advances in sequencing technologies led to the appearance of high throughput sequencing (HTS) or -omics methods. These relatively recent technologies allow the researchers to generate huge amounts of genetic information, thereby creating the biological big data field, genomics. These huge genomic datasets are a great source of information to study evolutionary patterns within different organisms. But new approaches have their own challenges.

The data storage and computational power are probably the most evident challenges related to genomics. With the incomparably high amount of data created with the use of the HTS approaches, the data management and secure data storage solutions are more visible than ever before. Many research institutions invest in their own data storage servers while others use online cloud-based storage plans. Until not many years ago, the big majority of analyses one could do on genetic datasets were completely feasible on a normal private computer, or even a laptop. But in the genomic era, even very simple analyses need more potent computation nodes. In this case also many institutions opted for online or shared computation systems. The computation power needed for an analysis depends basically on the size of the dataset but also the complexity of the algorithms implementing the analyses. In phylogenetics or population genetics for example, some of the older programs to use with smaller datasets do not even manage to load the new datasets. The rise of HTS approaches (Figure 1) has pushed many computer scientists and bioinformaticians to develop newer more efficient ways of coding the old algorithms or developing completely new approaches.

On the other hand, the potential of using all the evolutionary information coded into the genome of organisms allowed study designs which were impossible using single or even multi locus datasets created with a Sanger sequencing approach. The high complexity of the phylogenetic relationships of many groups are higher than the resolution limits of the most complete multi locus datasets. This is especially true in groups which have experienced rapid radiations for example. In such cases the diversification processes which separated the different lineages occurred in a relatively short time. Such brief periods of time, are not long enough for many genetic markers to accumulate enough changes to be recovered with phylogenetic methods.

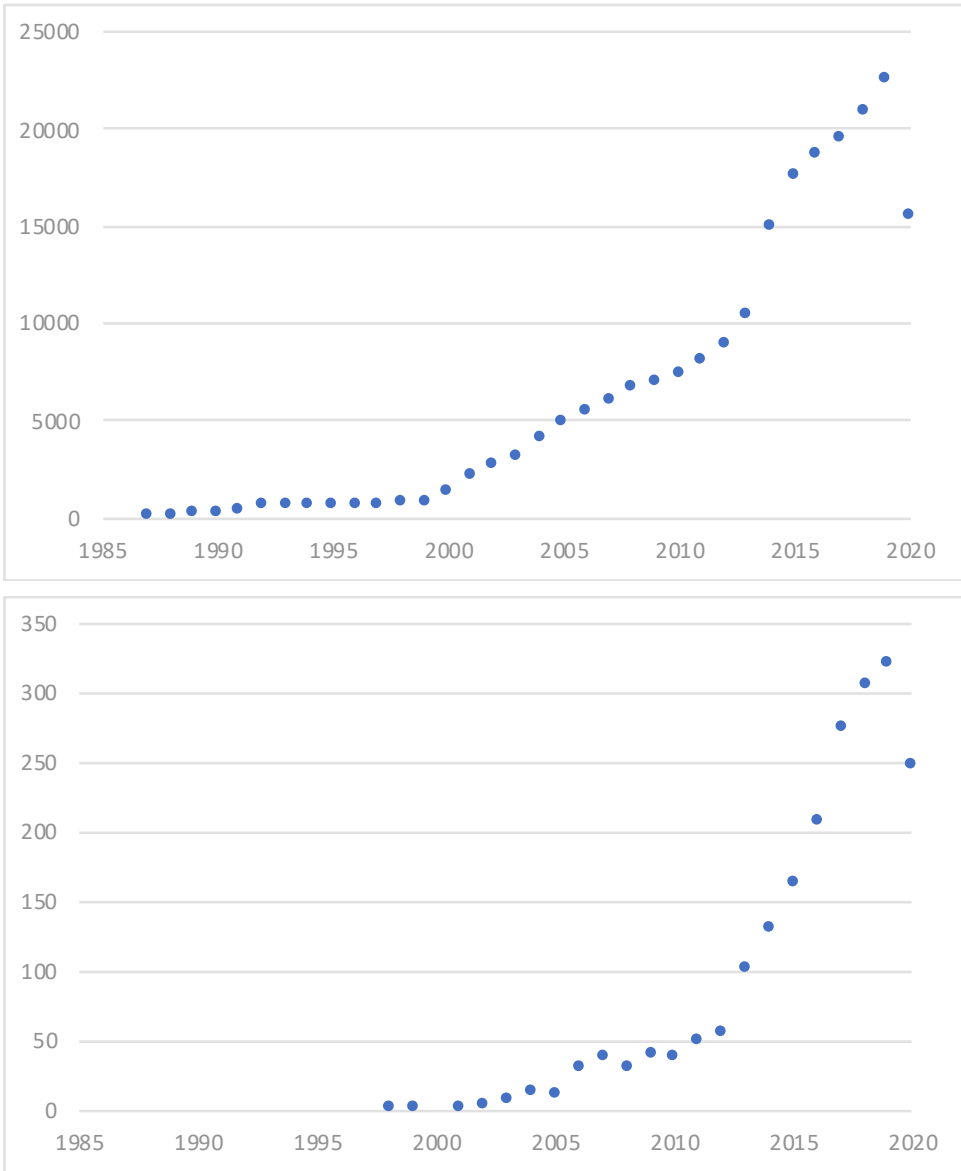


Figure 1. Number of published papers per year, as indexed in Pubmed, using the terms Genomics (top) and Phylogenomics (bottom).

The use of genomes is not limited to phylogenetic studies. Nowadays genomic approaches are used in numerous branches of life sciences. This is why the number of sequenced genomes is increasing on a fast track and does not appear to change any time soon (Figure 1). This creates a major publicly available resource for many different fields of study, phylogenetics included. In addition, with the advances in

the sequencing technologies and the radical drop in the sequencing prices, the generation of genomic data is becoming cheaper and more accessible for the scientific community every day. Therefore, the future challenge will be more of a big data nature, meaning that the mining, filtering and analysis of the data will be more important than the sequencing *per se*. In phylogenomics for example, using the full potential of available datasets is still an important challenge. With the available methods, some approaches to study evolutionary patterns are still not completely possible. Bayesian approaches to infer phylogenetic hypotheses based on genomic scale datasets, for example, are still very limited and unreliable. The most popular programs available in the actuality to study the phylogenetic relationships within a group are IQ-Tree 2 (Nguyen et al. 2015) and RAxML-NG (Kozlov et al. 2019) which are both maximum likelihood based. Another good example of inadequacy of traditional methods to use in phylogenomics is the question of evaluation of support values in inferred trees. The well-established bootstrap value in maximum likelihood phylogenetics for example, is of very little use in genomic scale phylogenetic trees. Alternative ways of measuring the support values have been suggested to deal with this problem. Gene and site concordance factors are an example of such new support values (Minh et al. 2020).

Finally, it is important to mention that the big data in evolutionary biology is not only limited to phylogenomics. In Bayesian methods, for timing a phylogeny or performing diversification analyses for example, where the probability space grows exponentially with number of parameters, the length of the dataset is not nearly as important as the number of individuals in the dataset. In these cases, basically a single tree with hundreds of final taxa, presents the same kind of challenges as in genomic analyses.

Diversification analyses

First works on diversification using phylogenetic methods date back to the 1990s (Hey 1992; Nee et al. 1992, 1994a, 1994b). In general, what we refer to as a "complete phylogenetic tree" is a "true" tree with all the extinct and extant species (Nee et al. 1994b). Now we could "know" the complete phylogenetic tree (in case of simulated trees) or not (usually the case in empirical studies!). The "reconstructed phylogenetic tree" in opposition, is the tree which includes "sampled tips", usually extant species, only (Nee et al. 1994b). The time since a clade is separated from its sister group is called the "root age" of that clade. The time of the first speciation event of a clade is called the "crown age" of that particular clade. In the example in Figure 2, looking at the tree on the right side (which is typically the only one we have access to), x0 will be the root age and x1 the crown age for the whole phylogeny. Note that x1 is also the root age for both clades "sp1, sp2" and "sp3, sp4, sp5".

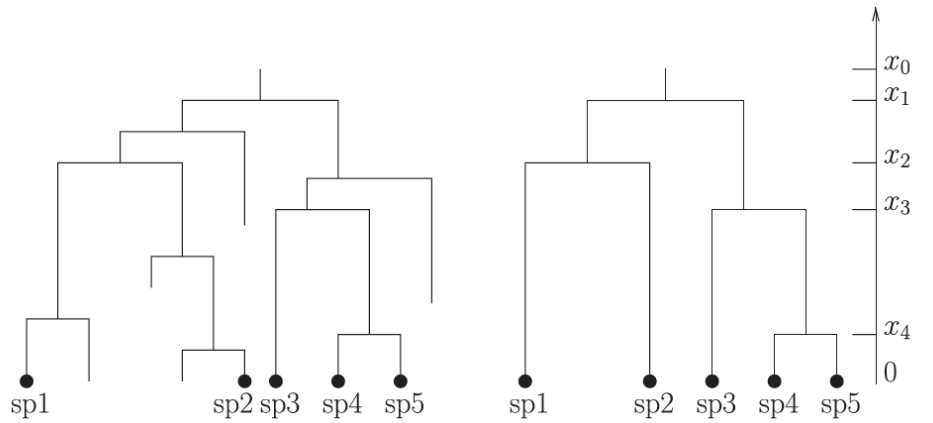


Figure 2. Left: complete phylogenetic tree with five sampled species (sp1–sp5). Right: reconstructed phylogenetic tree (modified from Stadler 2013).

In reconstructed phylogenetic trees, distances between species are in time units (relative or not). On the other hand, phylogenetic hypotheses obtained from empirical data are usually in units of evolutionary difference. When genetic data is used to obtain the phylogenetic trees, then the distance unit is genetic distance or amount of genetic changes. Therefore, these empirical trees are transformed into "time-calibrated trees" using a molecular clock, ideally using different sources of information (i.e. fossils, mutation rates and so on). A phylogenetic tree without any information about the branch lengths is called "tree shape" or "tree topology". A tree topology where the temporal order of divergence events is reflected in the order of nodes is called a "ranked tree" (Figure 3).

The diversification models can be classified from the simplest towards more complex recent ones. The simplest diversification model is one where the diversification rate is constant and it only includes a speciation (λ) process with no extinction (μ), also commonly known as a "Yule Process" (Yule 1925). The constant model could also allow for the extinction process to happen at a constant rate; this is called Time-constant model (Harvey et al. 1994; Nee et al. 1994b; Nee 2006). The next model can allow for the rates to vary in different ways. For example, we have Time-variable models where a different diversification rate is inferred for each time slice across the phylogeny (Rabosky and Lovette 2008a; Morlon et al. 2010, 2011). There is also the possibility of inferring clade-specific models where each clade could have a different diversification rate (Rabosky 2014). Some models reflect the effect of a trait or the evolution of a character on diversification (Maddison et al. 2007; Fitzjohn 2010; FitzJohn 2012; Magnuson-Ford and Otto 2012; Beaulieu and O'Meara 2016; Caetano et al. 2018) while others focus on the effect of the clade diversity itself (Walker and Valentine; Phillimore and Price 2008; Rabosky and Lovette 2008b; Burbrink and Pyron 2009; Etienne et al. 2012).

The simplest kind of diversification model is a constant rate model, where speciation and extinction rates do not vary during time and across the phylogeny (Raup et al. 1973). This type of model is usually called "constant rate birth–death" or "equal-rates" model and is considered as a null model (Nee et al. 1994b; Pybus and Harvey 2000; Paradis 2003; Stadler 2013; Morlon 2014). In this type of model, the speciation rate is greater than the extinction rate and, therefore, the number of lineages accumulates towards the present. A special case of this type of model is when no extinction occurs ($\mu = 0$) and it is a pure birth model (Yule 1925). In this case, the number of species increases linearly on a semi-logarithmic scale, where the slope is the rate of speciation (Morlon 2014). This increase in species number is usually visualised with the help of a lineage-through time (LTT) plot that is constructed from a dated phylogeny where one can count the number of species or lineages present at each time interval. Harvey et al (1994) observed that, when the extinction rate is higher than zero, an increase in the slope of species number accumulation is seen. This observed effect is called "pull of the present" and, in theory, is used to infer the extinction rate. The possibility of estimating extinction rates from phylogenies is highly debated (Kubo and Iwasa 1995; Paradis 2004; Quental and Marshall 2010; Rabosky 2010, 2016; Beaulieu and O'Meara 2015). The pull of the present can be understood as the effect of lineages closest to the present, with extinction not having had enough time to affect them. Remember that for a species to go extinct it should first speciate and "live" for a while.

Models based on constant rates of diversification may give some realistic results especially in small clades evolving over a relatively short period of time. However, constant rate models are usually not accurate for large datasets covering longer evolutionary times.

In time-variable models, extinction and speciation rates can be independent and vary through time. The variation in these rates usually depends on time linearly or exponentially. In other words, time is the main variable explaining the variation in diversification rates (Rabosky and Lovette 2008a; Morlon et al. 2010, 2011; Stadler 2011). Due to this, the time-variable models are also called "time-dependent models". These models are typically applied to a whole tree or a subtree. Even though, it is also possible to compare different diversification rates scenarios partitioning the tree in different subtrees. Usually the time-variable methods are used as an exploratory analysis to evaluate if any variation in speciation and extinction rates are observed. In such cases further analyses are needed to study the processes behind such variations.

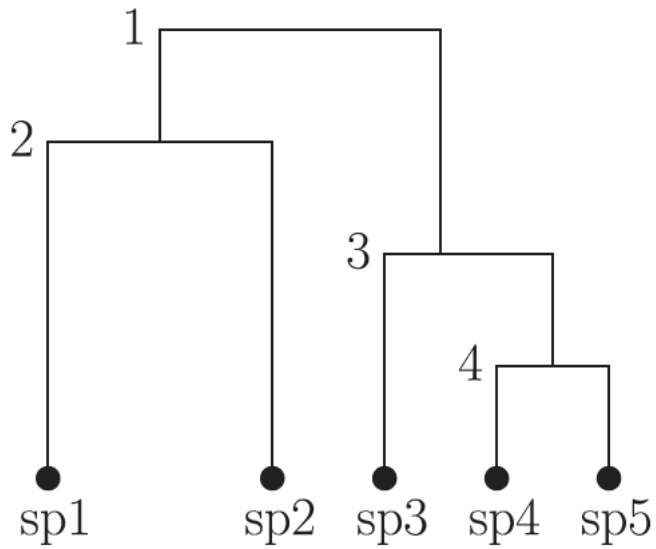


Figure 3. A ranked tree, where the order of divergence events is presented. Note that the branch lengths here are not informative (modified from Stadler 2013).

Sometimes not all lineages in a phylogeny share the same variation in diversification rates. This is especially important in deep and big phylogenies where distinct groups have been affected by very different evolutionary challenges. In such datasets, distinct parts of a tree are also expected to have different macroevolutionary dynamics. In the cases when you expect variation in diversification rates in different parts of the tree, clade-specific models offer a more realistic scenario for the evolutionary history of different groups.

In clade-specific models, different branches can vary in their diversification rates (Rabosky 2014). The popular and highly controversial BAMM (Rabosky and Huang 2016) software implements a clade-specific diversification model. A characteristic of this Bayesian software is that it does not allow the extinction rate to be higher than the speciation rate. Also, some of the mathematical implementations of the model in BAMM have been criticized. In a very recent critique of the BAMM method, Meyer and Wiens (2018), based on simulated data, studied its accuracy and concluded that BAMM could result in "biased and inaccurate estimates of diversification rate". This is likely due to the underestimation of the shift numbers in the phylogeny and the overestimation of diversification rates in small taxa which have low diversification rates (Meyer and Wiens 2018). In the response paper, Rabosky (2018) replies to these criticisms, by basically only justifying why the other method (Magallón and Sanderson 2001, Method-of-moments or MS estimator) is

not more accurate than BAMM and does not directly treat the problems in BAMM. In a response, Meyer et al. (2018) did further comparisons where they report inconsistencies in the results obtained from the BAMM analysis for the same clade, and finally concluded: “we strongly caution against using BAMM in empirical studies”.

Other Bayesian approaches exist where authors claim to have corrected for BAMM’s mathematical implementation errors (Höhna et al. 2019). In their implementation of the model, within RevBayes (Höhna et al. 2016), the authors relax the constraint of fixed extinction rates and allow it to vary independently from speciation rates. Furthermore, another issue in BAMM analysis is the taxon sampling fraction implementation in the model. It is not clear how it is implemented or the mathematics behind it, with no explanation in either the original paper or on the website. In the implementation of the branch-specific model in RevBayes different options are available for taking into account the taxon sampling fraction of the dataset. None of these methods allow the user to define a different taxon sampling fraction for different parts of the tree, which is what BAMM “allows” the user to do. However, as it is not clear how this is done in BAMM, the options available in RevBayes offer more transparency and perhaps reliability.

Historical Biogeography

The actual distribution ranges of organisms, given that their phylogenetic relationship is known, can inform us about their past distribution ranges. The geographical range where an organism occur affect in many evolutionary traits of that organisms. This is why numerous macro evolutionary processes of a lineage, as extinction and speciation, are directly affected by its past distribution ranges. Usually the historical changes in the distribution ranges are not directly observable. In order to study the historical variation in the distribution ranges biogeographic inferences can be applied. In the actual biogeographic models, usually the time calibrated phylogenetic history of the group is known, or inferred separately, and parameters as the actual distribution, historical variations in the dispersal rates, and changes of the biogeographical regions during time are introduced in the model.

The Dispersal-Extinction-Cladogenesis (DEC) process (Ree et al. 2005; Ree and Smith 2008) is the most popular biogeographical model nowadays. In this model is formed by three key components, the actual distribution, the anagenetic range changes and the cladogenetic range changes.

The actual distribution of the clades is coded into the model as a set of discrete traits of presence-absence type. In this case it is important to consider that the presence in more than a region is coded as an alternative trait and not the sum of the other traits. In other words, both presence and absence are coded together in a trait. For example,

if a taxon is present in one area (A) out of two possible areas (A and B), this is coded as presence in the area A and absence in the area B. therefore if a taxon is present in both areas, its distribution trait is an independent state from the distribution in each of the areas.

The anagenetic and cladogenetic range changes are in essence very similar. Both model dispersal and extinction events. In the anagenetic range change, both dispersal and extinction events occur along the same branch. Meaning that if a taxon is present in one area at time t, then at time t+1 it can disperse to another area or extinguish in that area. The cladogenetic range change explore the dispersal and extinction at a speciation even (a node) in a tree. Meaning that the daughter species at a node, do not necessarily inherit the original distribution of the root, and this variation on how the range is inherited is coded into the tree.

Study System

Some of the most diverse groups of organisms (i.e. very interesting in the point of view of diversification studies), are the ones that have very poor fossil records. One of these very interesting groups is the insect order Lepidoptera (i.e. butterflies and moths). Being one of the most diverse group of organisms (just after Coleoptera, and probably Diptera and Hymenoptera), with more than 157,000 described species (van Nieukerken et al. 2011), their diversity is often hypothesized to be related to the diversity of the plants they feed on. However, the general diversity of the group has been studied only in a few works. This is mainly due to the fact that until very recently the phylogenetic relationships between major groups was not well resolved and therefore, there was no information on the dating of major diversification events (Wahlberg et al. 2013).

To study diversification patterns within the order we decided to approach it by studying different families of Lepidoptera. Obtaining a well resolved and dated phylogeny for each family will allow us to study the diversification dynamics within each family. In the end comparing the diversification patterns which affect each family to each other, will allow us to look for general patterns which affected Lepidoptera as a whole. With this method in mind, the first family we chose to study the diversification dynamics of was the Geometridae moths.

Geometridae are one of the most diverse families within Lepidoptera (Figure 4), with more than 23,000 described species in more than 2,000 genera (van Nieukerken et al. 2011). Species of this family have a global distribution and some of them are important pests, having a big impact on human societies. The phylogeny of the family is still poorly known. The root of the family is around 82 My old (Wahlberg et al. 2013).

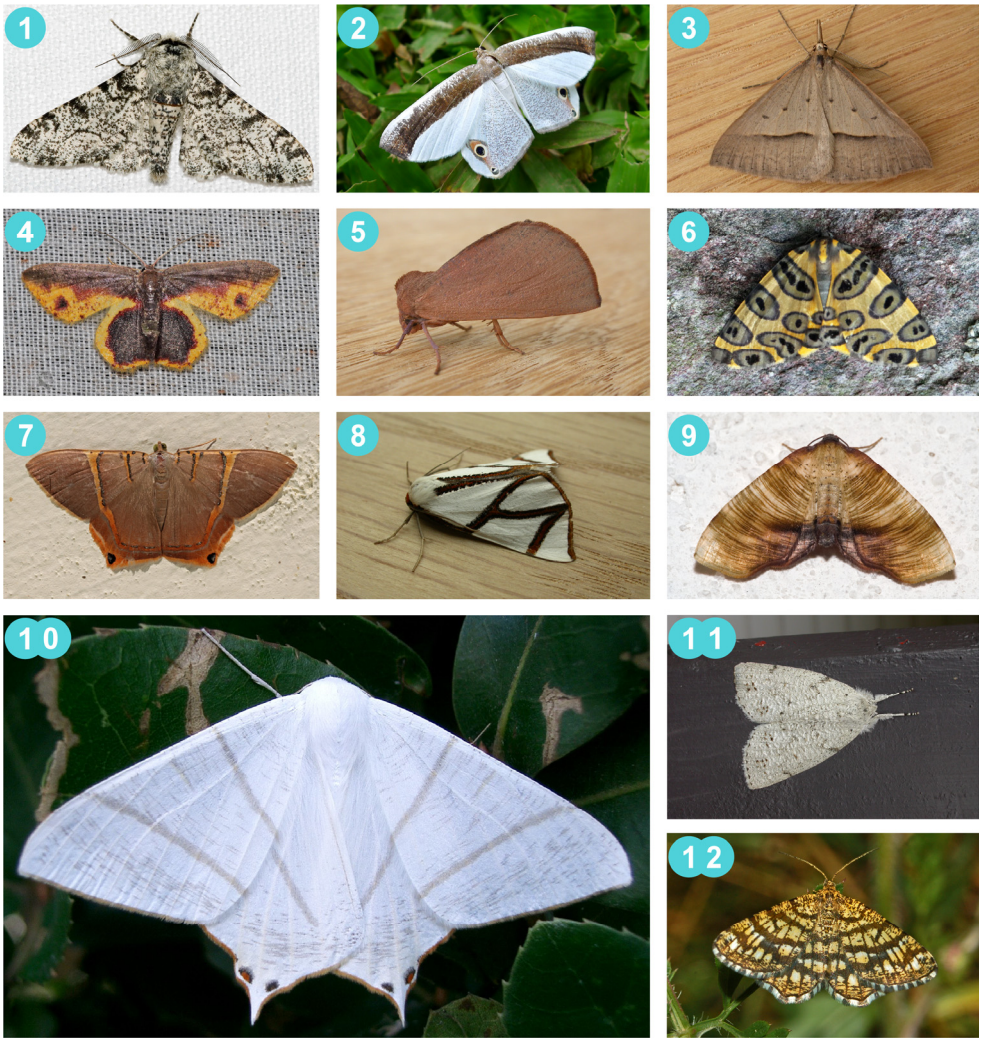


Figure 4. The diversity of the Geometridae. 1, *Biston betularia*; 2, *Opisthoxia amabilis*; 3, *Epidesmia hypenaria*; 4, *Chrysocraspeda mitigata*; 5, *Fisera perplexata*; 6, *Pantherodes pardalaria*; 7, *Phrygionis polita*; 8, *Thalaina clara*; 9, *Plagodis dolabraria*; 10; *Ourapteryx* sp.; 11, *Mochlotona phasmatias*; 12, *Chiasmia clathrata*.Pictures All the pictures are from Wikipedia by Chiswick Chap, Gail Hampshire, Donald Hobern, Alexey Yakovlev, Donald Hobern, Charles J Sharp, Charles J Sharp, Donald Hobern, Kulac, KENPEI, Donald Hobern and Hectonichus respectively.

Another megadiverse group of Lepidoptera is the moth family Erebidae (Figure 5). With more than 24,000 described species in over 1,700 genera (van Nieukerken et al. 2011). Like Geometridae this group is also distributed worldwide and its estimated age is in the same range (slightly younger at ~65 My, Wahlberg et al. 2013). The most complete phylogenetic hypothesis of the family is proposed by Zahiri et al. (2012) which relies on seven nuclear and one mitochondrial protein coding genes. Few subfamily level phylogenetic studies appeared in recent years

using a more modern high throughput sequencing (HTS) approach (Homziak et al. 2019; Dowdy et al. 2020).

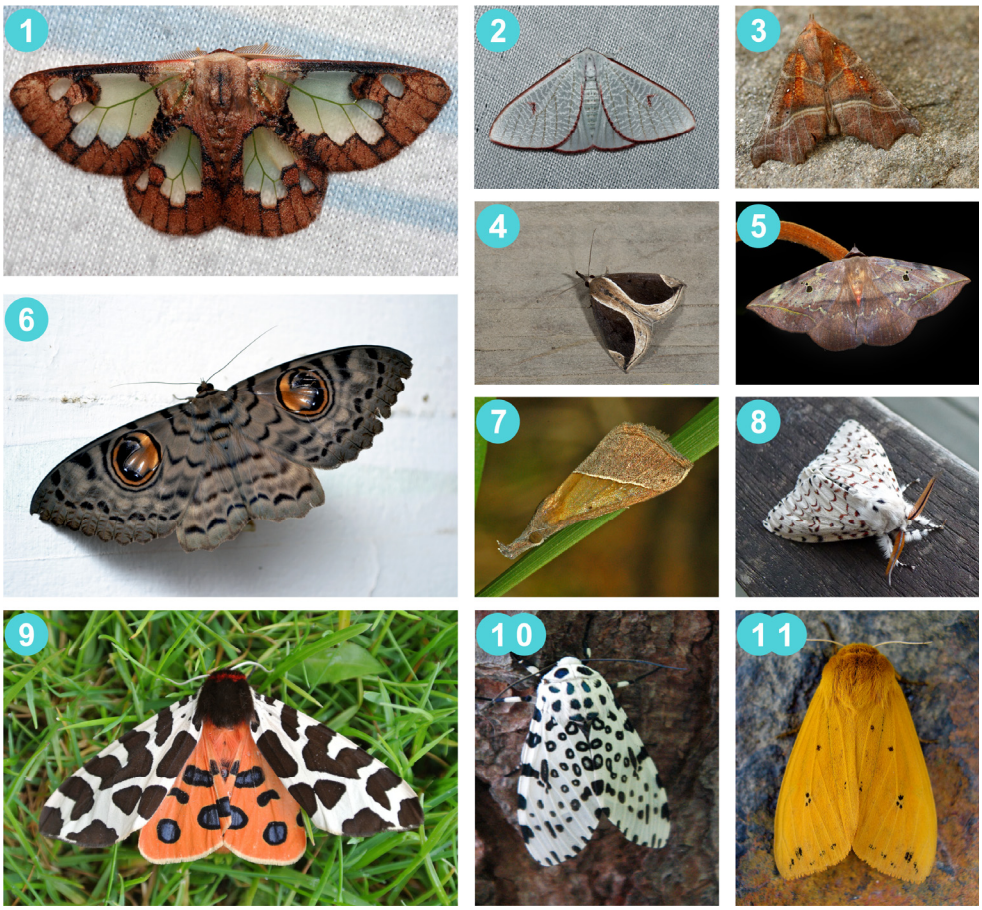


Figure 5. The diversity of the Erebidae. 1, *Carriola ecnomoda*; 2, *Arctornis egerina*; 3, *Scoliopteryx libatrix*; 4, *Hypena taiwana*; 5, *Hypopyra capensis*; 6, *Erebus macrops*; 7, *Hypena lividalis*; 8, *Lymantria nephrographa*; 9, *Arctia caja*; 10; *Hypercompe scribonia*; 11, *Pyrrharctia Isabella*. All the pictures are from Wikipedia by Alexey Yakovlev, Alexey Yakovlev, ©entomart, LiCheng Shih, Frank Vassen, Mullookkaaran, Hectonichus, Gail Hampshire, Temple of Mara, Jeremy Johnson and Steve Jurvetson respectively.

Aims of the Thesis

In this thesis I use molecular big data of two megadiverse family of moths to investigate their evolutionary patterns. That is why the first goal of the thesis is to obtain a robust and well resolved phylogenetic hypothesis for each of the families. I covered this first goal in **Paper I and III**. I used different approaches to resolve the phylogenetic relationships of the two families. In the case of phylogeny of Geometridae (**Paper I**) the approach focused on increasing the taxon sampling of the group, sequencing up to eleven genetic markers for over 1200 species. Then using the well supported phylogenetic tree of the Geometridae family, I investigated the diversification patterns and the biogeography of the family in the **Paper II**. The major challenges in these studies was dealing with the complexities of Bayesian analyses in dating, diversification and biogeography analyses of such a big dataset. In the **Paper III** the challenges were a little different. In this case the available phylogenetic hypothesis in the literature did recover very short unsupported internal branches, most probably due to rapid radiation events. To resolve the short internal branches and offer a supported phylogenetic hypothesis for the relationships between the different subfamilies of Erebidae we opted for a genomic approach. In this study (**Paper III**) I explored the possibility of using old genomic extracts which are considered not of good enough quality for genomic uses. In recent years, many mitochondrial genomes have been published and many phylogenenomic analyses have been done based on the mitochondrial genomes. In **Paper IV** I explore the utility of these datasets and their resolution resolving family level phylogenetic hypotheses in Lepidoptera. One of the appealing points of using whole genomes in phylogenetic study designs is the possibility of using the obtained genomic results for alternative, not phylogenetic related, questions. I explored one of these alternative applications of the genomic big data, studying the symbiotic diversity observed within the whole genomes of Erebidae (**Paper V**). To summarise, the major goals of my thesis were:

1. To study the evolutionary history of Geometridae and obtain a supported and well-resolved phylogenetic hypothesis resolving the relationships within each subfamily. **Paper I**
2. To explore for the first time the diversification patterns and the biogeography of Geometridae to understand its high diversity. **Paper II**

3. To explore the methodological challenges of sequencing whole genomes using available old genomic DNA extracts. **Paper III**
4. To resolve the backbone phylogenetic relationships within different subfamilies of Erebidae. **Paper III**
5. To explore and evaluate the phylogenetic power of mitochondrial genomes to resolve family level relationships within Lepidoptera. **Paper IV**
6. To investigate the exploratory power of screening genomic data obtained in our approach to look for symbiont diversity of Erebidae. **Paper V**

Brief Methodology

In this thesis I have explored the use of a varied set of methodologies. The first paper is focused on the phylogenetic relationships within the moth family Geometridae. In this study eleven genetic markers were amplified using the primers and PCR conditions described in Wahlberg and Wheat (2008) and Wahlberg et al. (2016). Briefly, sequences were aligned in Geneious 11.0.2 (Kearse et al. 2012) using the implemented MAFFT 7 algorithm (Katoh and Standley 2013). The dataset was constructed in VoSeq (Peña and Malm 2012), and the phylogenetic analyses were performed in IQ-Tree 1.6.10 (Nguyen et al. 2015).

In the second paper, I used the obtained phylogeny in the first chapter for the dating analysis in BEAST2 (Bouckaert et al. 2014). This dataset being the most complete of the family up to date, presented numerous methodological challenges for the analyses. The timing analysis for example was very time consuming. Two diversification analyses were performed on the time-calibrated tree, the Episodic Birth Death (EBD and the Branch Specific Diversification (BSD) models (Höhna et al. 2019) as implemented in RevBayes (Höhna et al. 2016). The Biogeography of the family was also studied using a Dispersal-Extinction-Cladogenesis (DEC) model (Ree et al. 2005; Ree and Smith 2008) also implemented a Bayesian framework in RevBayes (Landis et al. 2018). Then the result of both biogeography and BSD were combined to give a clearer image of the evolutionary patterns of the family Geometridae.

In the third paper, I wanted to resolve the deep phylogeny in the family Erebidae. The most recent published paper on the phylogeny of the group (Zahiri et al. 2012), using a multi locus Sanger approach did not resolve the relationships between different subfamilies. Thus, I used HTS methodology to sequence the whole genome of 47 species, sampled in order to recover all the deep nodes in Zahiri et al. (2012). For this paper I used the same DNA extracts available from that study and prepared libraries and cleaned the raw reads following the methodology in Twort et al. (2020). The raw reads were cleaned using Prinseq 0.20.4 (Schmieder and Edwards 2011) and Trimmomatic 0.38 (Bolger et al. 2014) and the cleaned reads were then de novo assembled with spAdes 3.13.0 (Nurk et al. 2013). Then using the MESPA protocol (Neethiraj et al. 2017), a set of genetic markers were extracted and uploaded into VoSeq (Peña and Malm 2012) database to generate the dataset. The final dataset included also all the available online genomes for Erebidae (7 in total) plus other 18 species from Euteliidae, Noctuidae and Notodontidae as outgroups.

Both nuclear and amino acid datasets were used in IQ-Tree2 (Nguyen et al. 2015) using ultrafast bootstrap approximations (UFBoot2) and SH-like approximate likelihood ratio test (Guindon et al. 2010; Hoang et al. 2018). To evaluate the support of the nodes we explored the gene concordance factor (gCF) and the site concordance factor (sCF) as implemented in IQ-Tree2 (Minh et al. 2020).

Using the genomes obtained in the paper three, I evaluated the accuracy of the phylogenetic studies using only mitochondrial genomes. For this the mitochondrial genomes of the 47 erebid moths were *de novo* assembled using 2 alternative approaches. We first used the Novoplasty (Dierckxsens et al. 2016) on all samples. For the samples which did not result in an acceptable circular genome we used the mirabait option in MIRA 4.0.2 (Chevreux et al. 1999, 2004) to find the reads corresponding to mitochondrial DNA. The mitochondrial reads were de novo assembled using three simultaneous approaches, the Geneious de novo assembler, SPAdes assembler 3.10.0 (Nurk et al. 2013) and plasmidSPAdes (Antipov et al. 2016), all of them implemented in Geneious 10.2.6 (Kearse et al. 2012). For each sample, all the contigs over 500 bp were aligned to a reference MtGenome of another species of Erebidae. Then the consensus sequence of the alignment was used as a reference to map the mitochondrial reads in Bowtie2 (Langmead and Salzberg 2012) as implemented in Geneious with default parameters. All the resulting assembled genomes were annotated using MITOS (Bernt et al. 2013). From the annotated genomes eleven protein coding genes (PCG) were extracted from all mitochondrial genomes (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 4L, *ND4L*). Each gene was aligned separately using MAFFT v7.450 (Katoh 2002; Katoh and Standley 2013) as implemented in Geneious with default options and uploaded to the VoSeq (Peña and Malm 2012) database. Then both amino acid and nucleotide datasets were created to perform maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses. The ML analyses were performed usin IQ-Tree2 and the BI analyses using MrBayes 3.2.7 (Ronquist et al. 2012).

For the last paper I explored the possibility and power of screening for symbionts using the low coverage whole genomes obtained in the paper 3. The raw reads were cleaned using Prinseq and Trimmomatic following the same parameters used in the paper 3. Cleaned reads were assigned taxonomic labels with Kraken2 (Wood and Salzberg 2014) and MetaPhlAn 2.0 (Segata et al. 2012). Kraken2 was run using a custom database, which contained the standard kraken database, the refseq viral, bacteria and plasmid databases and all available Lepidoptera genomes from genbank

Results and Discussion

In the first paper we obtained the most complete dataset for the family Geometridae covering up to half of the generic diversity of the family. We sampled 691 genera out of the 1961 described genera. In this study we offer a robust phylogenetic hypothesis resolving most of the evolutionary relationships within the family. The Sterrhinae subfamily was found as the sister group to the rest of the family. We confirmed the monophyly of Larentinae, as suggested by other authors. The subfamily Archiearinae was recovered as the sister group to the rest of the subfamilies. The subfamily Epidesmiinae was suggested as a new subfamily within Geometridae. In addition, many taxonomical issues have been addressed and this paper is already one of the main references for any work on the family.

Following on the results obtained in the first paper, I wanted to explore the diversification and the biogeography of the family. A major issue in this kind of studies is the taxon sampling of the dataset to be studied. The Geometridae dataset we have used is by far the best sampled dataset but it only covers around 5% of the whole diversity of the group. In our study this is a real problem but it mainly affects the branches toward the present. In other words, as we have the major lineages of the family sampled, going a few million years back in time, the majority of the lineages are sampled. A big challenge in this project was to analyse such big datasets. The actual methods for time calibration for example deal very poorly with large datasets. The complexity of a Bayesian analysis increases exponentially with each new parameter. A dataset with over 1200 species has many parameters and therefore is very complex to analyse. For example, each branch in a topology is a parameter. This complexity usually is translated into practical issues such as extremely long computation time, very large computation power needed or just the inaccuracy of the algorithms to deal with such multidimension probability spaces. In paper two I first time calibrate the phylogeny and then study in a Bayesian framework both the diversification and the historical biogeography of the family. In this study I present for the first time a date for the major divergence events within the family. Then using the time calibrated tree, I study the diversification dynamics within the family using two alternative approaches. The first approach measures a mean diversification rate for the whole family which vary through time. In this analysis I obtain two relatively short time periods around 35 and 10 million years ago when, on average, the family had higher net diversification rate. Then using a

second approach to study the variation in diversification rate for each branch in the tree, I find the lineages and the time frame where changes in the diversification rates are visible (Figure 6). Using the second diversification analysis approach I find numerous independent increases in diversification rates in independent lineages again around 35 and 10 million years ago. Then using the actual distribution of the family and the time calibrated tree I studied their historical biogeography. In this analysis the importance of different biogeographic areas was demonstrated and also the dispersal events which shaped the actual distribution of the family. As an example, we recovered that the family originated most probably in the Neotropics where extant diversity is high. By combining the diversification and historical biogeography analyses, I show the variation of the diversification rate of members of the family in the different biogeographical regions and through time. This analysis allowed me to understand better the role of different regions through time in creating the current diversity of the family.

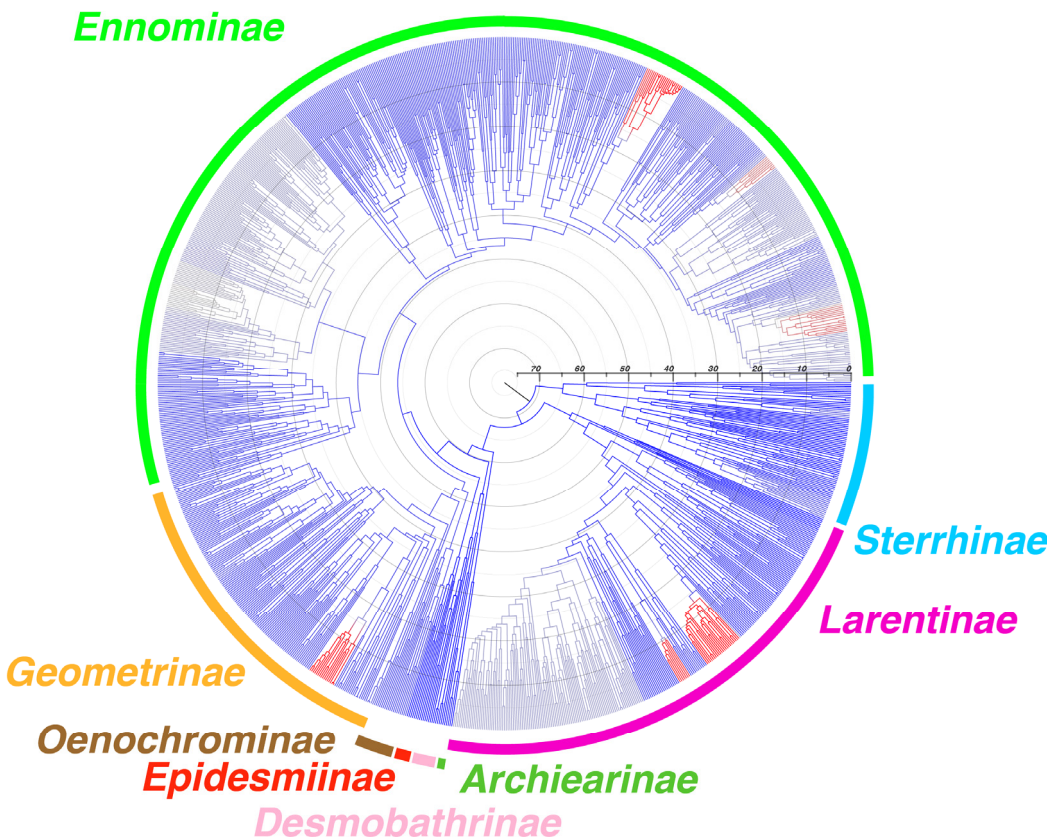


Figure 6. The result of BSD analysis in Geometridae.

In the third paper we had a similar question but a different problem. The third paper deals with the evolutionary history of the family Erebidae in Lepidoptera. This is also a megadiverse family of moth with over 24,000 species. In this case, a similar approach to the first paper had been already tested in the literature to try to resolve the phylogenetic relationships of the subfamilies (Zahiri et al. 2012). Zahiri et al (2012) did use a large dataset of Sanger based genetic markers, but none of the internal branches at the subfamily level was resolved and the relationships between different subfamilies were unclear or not supported. Zahiri et al. (2012) suggested that these numerous internal short branches are most probably due to a rapid radiation even early in the history of the Erebidae subfamily. In this case in order to resolve the internal short branches and find a robust phylogenetic hypothesis, I used a high throughput sequencing (HTS) approach. I chose 47 species based on Zahiri et al (2012) in order to recover the majority of the deep nodes and short branches. Using the same old DNA extracts from that paper, I made libraries and sequenced the whole genome of the 47 species. The first challenge in this study was to being able of sequencing whole genomes only based on old DNA extracts which do not have the high quality and high quantity of DNA needed for the majority of the protocols to prepare libraries. In paper three, for the first time I show the utility of these, often forgotten, DNA samples in order to obtain genomic scale sequences. This shows the value of old genomic extractions available to numerous research groups. Once I obtained the genomes, I assembled them and using a relatively new approach, I extracted a gene set of over 200 genetic markers. Using this large dataset, I obtained a phylogenetic hypothesis for the evolutionary relationships within the family. A typical problem with the use of genomic scale data for phylogenetic studies is the calculation of support values and uncertainty in the obtained tree. The traditionally used methods as bootstrapping and its variations tend to offer poor confidence and accuracy, or at least should not be interpreted the same way as the traditional single of multi locus genetic marker trees based on low number of markers. To overcome this issue and to be able to evaluate the support of my phylogenetic hypothesis I applied the concordance factor approach. This new method is not as easy to interpret as the traditional support values but it offers an alternative way of presenting the uncertainty designed for genomic datasets. Finally, using a genomic scale dataset and applying the most up to date methodology I obtained the most robust phylogenetic hypothesis for the family Erebidae, resolving the majority of the deep node evolutionary relationships.

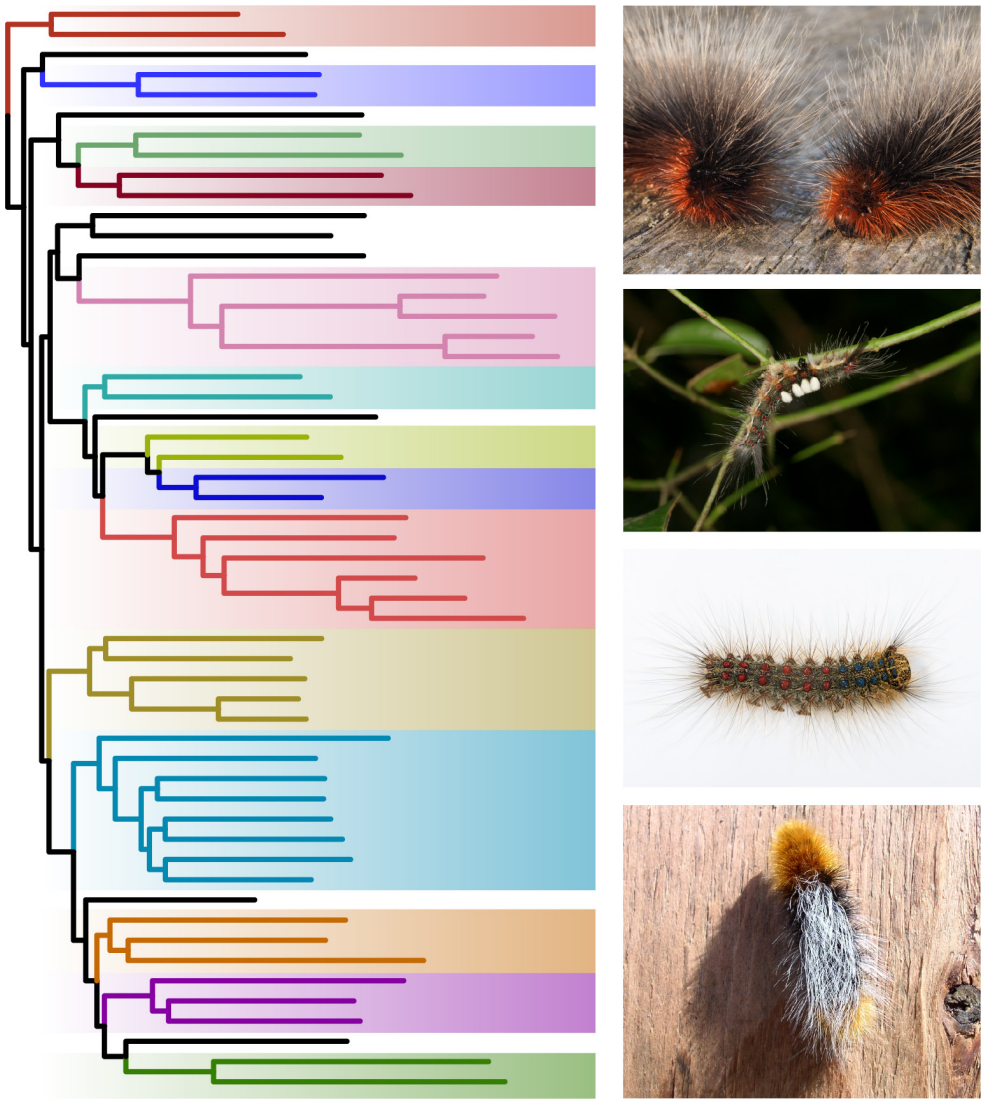


Figure 7. The phylogenomic tree of the family Erebidae. Each colour represents a subfamily. The caterpillar pictures on the right, from top to bottom are *Arctia caja*, *Orgyia* sp., *Lymantria dispar* and *Platyprepia virginalis*. All the pictures are from Wikipedia by BaykedeVries, © 2016 Jee & Rani Nature Photography, Didier Descouens and Beatriz Moisset respectively.

Mitochondrial genes have been used to study phylogenetic hypotheses since the beginning of molecular systematics. This is due to many factors such as the relative ease of sequencing of these markers. With the appearance of HTS methodologies and their standardisation there is a rise in the publication of mitochondrial genomes and often their use to respond different phylogenetic questions. In the fourth paper I wanted to evaluate the accuracy of using mitochondrial genome data in order to

resolve deep phylogenetic relationships. I first assembled the mitochondrial genomes out of the whole genome results of the paper 3, and using all the mitochondrial genomes available online I tried to resolve the complex deep node relationships and internal short branches within Erebidae. Comparing different phylogenetic approaches, maximum likelihood (ML) and Bayesian inference (BI), and different datasets, the amino acid or the nucleotide datasets, I showed the limitations of mitochondrial genomes as an information source for family level phylogenetic studies at least in Lepidoptera. In this study I showed that this approach is not accurate enough in order to resolve the relationships between the different subfamilies but, given a good taxon sampling, it might be a relatively good approach for the phylogenetically shallower scale questions, within a tribe or subfamily for example.

An increasing number of scientists sees organisms as communities of interacting species rather than independent entities. A growing number of studies focus on the symbiotic interactions between microorganisms and their hosts. In insects particularly, the majority of symbiotic studies focus in *Wolbachia* and other organisms are overlooked. In paper five I wanted to explore the exploratory power of whole genome techniques in order to look for the presence of possible symbionts in the genome data. Using the 47 newly sequenced whole genomes of Erebidae in the paper 3, I reported for the first time four new species of moths that have an infection by *Wolbachia*, in one other species I found *Burkholderia*, and *Sodalis* and *Arsenophonus* simultaneously in two other species. Interestingly, one species was infected by a bacterium that is described as a hemipteran organ symbiont. I also observed numerous cases of *bracovirus* reads related to a parasitoid braconid wasp which could inform us about the complex cycles of this virus in both the wasp and the lepidopteran larvae. At the end I discuss the high potential of using this technique as an exploratory tool using all the publicly available genomic information online.

Conclusions

In general, the thesis deal with challenges of working with big datasets at different levels to understand evolutionary patterns within two of the most diverse families of organisms. In the first paper a robust phylogenetic hypothesis is obtained resolving the complex relationships within the major lineages of the family Geometridae. After obtaining a robust hypothesis on the complex evolutionary history of the Geometridae, I used the results obtained in the fist paper to study the diversification processes affecting the history of the family. Again, here one of the main challenges were to use a very species reach phylogenetic tree and perform analyses which are not performing best with big datasets. In the paper 2 I studied for the first time the diversification and the biogeography of the family Geometridae. In this paper I show the variation in the diversification rate between the different lineages through time. I also studied their historical biogeography. The joint study of both diversification and biogeography allowed me to compare the importance of the different biogeographic areas during different time periods. This has shown for example the importance of the Palaearctic region as one of a land bridge allowing the dispersal of different group between major regions. Or the early importance of the Neotropics in hosting a high diversity and exporting it to other regions. The observation of two major rise in the mean diversification rates in the family around 30 million years ago (mya) and again around 10 mya using the EBD method, coincided with the increase in the diversification rates of numerous independent lineages recovered in the BSD method. These important events were most probably due to the major climatic changes at that period.

In the third paper I use a different approach to resolve a question similar to the paper one. Here the phylogenetic relationships of the Erebidae family is in the focus. The most complete phylogenetic study on this family recovered numerous deep node uncertainties and did not resolve the relationships within the different subfamilies. In addition, they showed how the internal branches were very short, probably due to a rapid radiation. Therefore, the approach we used was to sequence whole genomes of representatives of different subfamilies in order to resolve the deep node placements and subfamily level relationships. In addition, we explored the utility of the old genomic DNA extracts for such studies. The results of our study show clearly the high potential of these forgotten genetic resources to be used with the newer HTS approaches and in addition we obtained the most robust phylogenetic hypothesis for the family at the subfamily level up to date.

When thinking on the best approach to deal with the challenges in the chapter three, we observed the relatively high number of available mitochondrial genome publications. The high majority of these papers did not have a clear study design or research question, and nearly all of them performed a poor phylogenetic study together with their mitochondrial genome results. To explore the accuracy of such, usually deep, phylogenetic studies, we decided to assemble the mitochondrial genomes obtained in the paper three and evaluate their accuracy resolving deep family level phylogenetic questions, at least in Lepidoptera. Our result unsurprisingly demonstrated how poorly the mitochondrial genomes performed resolving relationships at subfamily level, but also show that given a good taxon sampling, they might be good tools to study shallower phylogenetic questions.

At the end, in order to explore the possibility of using whole genome data to study symbiont diversity, we screened the genomes obtained in the paper three. In this paper we show the exploratory power of such approaches using metagenomic pipelines to screen for clues on the possible symbiotic relationships. In addition, for the first time we recovered the occurrence of known *Wolbachia* in four species of Erebidae where they have not been reported previously. Also, we reported the presence of other interesting symbionts which are usually less studied in the symbiotic interaction studies. Our results clearly show the high exploratory potential of this approach for mining available genomic data through the online databases.

Acknowledgements

This thesis would not have been possible without the help and support of many people.

Niklas, I will never be able to thank you enough for offering me this opportunity, for your patience and for not giving up each time I came up with a new idea which never worked! You have always been there, to support any crazy ideas, to respond stupid questions, to tell me when I was wrong, to tell me when I was right, to bring me back on track when I was lost, to advise me whenever I needed it and for anything I could think of. You always gave me the freedom to think, to experiment, to analyse, to be wrong but to learn from it. It was really nice to be your PhD student, I might get the PhD soon, but I will still stay your student!
By the way… I am not sending this text to you to ask you to correct with my poor English, so it will surely have plenty of errors and colloquialisms! :)

I would like to thank my PhD committee, **Anders**, **Staffan** and especially **Honor** who have always been there for questions have been so flexible for the meetings and helped me through all the questions I had.

Leidys, thank you for… que no voy a escribir en inglés! Gracias por ser tu, por ser tan terca y por reganarme y decirme siempre que algo te parece mal. Gracias por los chismes, los cafés, las fiestas, las bromas pesadas, por todo lo que hemos vivido juntos estos años. Pero sobretodo, gracias por tu amistad :)

Elsa, on a été là depuis le tout début :) plus de 4 ans maintenant ! Des conversations qui fait peur aux suédois, dans un hot tube ou dans la cuisine ; de l'alcool pour nettoyer, pour le labo et pour les conférence ; Ce qui se passe a Groningen reste en Groningen, et en Tovetorp , et en Sofia, et en Podgora, et en… bref partout dans le monde ! Merci pour toujours être là. Et aussi merci pour rire avec mes blagues qui font pas rire :D

Victoria, thank you for always being so helpful. You have always helped me with everything, from the lab work, to writing, to complaining, to fishing to everything :) McDonalds breakfasts are not that bad after all! We still need to organise that roadtrip in NZ!!!! You just need to learn some proper English b….

Nicolas, **Jadranka**, **Andrea**, **Anne**, **Simeao**, **Eduardo**, **Irenka**, **Tobias**, **Amandine** and **Laure**, thank you for all the fikas, meetings, conversations, labworks, talk practices, cakes, parties and everything that we experienced together. You have been the closest people to me during these years.

Christoffer, thank you for being always up for identifying a weird bug, or going sampling to Iran, or organize other samplings, or help with a DIY project, or talk about cameras and photography or … We should definitely go sampling again! **Rune**, thank you for all the help with material, with identification, for being always smiling for being always helpful.

Johanna, **Dafne**, **Theo**, **Oscar**, **Katja**, **Carsten**, **Gróa**, **Homa**, **Chon**, **Ainara**, **Carlos**, **Annick**, **Juan Pablo**, **Micaela**, **Romain**, **La Mari**, **Tristan**, **Daniel**, **Ivette**, **Martin**, **Cecilia**, **Albert**, **Melanie**, **Milda**, **Fabian**, **Robin**, **Kajsa**, **Rachel**, **Monica Aivars**, **Yesbol**, **Raphael**, **Jesus**, **Margarida**, **David**, **Zach** and **Emma**, thank you all for all the conversations, laughs, collaborations, drinks (warm and cold), lunches, parties and everything which made my time here unforgettable.

Thank you to **Daniel**, aunque no lo parezca llevamos mucho tiempo conociéndonos! (aun sin contar la carrera digo… como sabes no me acuerdo mucho… :p) Eres un amigo único, gracias por estar allí siempre que lo necesité. Aunque vivimos en ciudades diferente, eso de saber que vivimos en el mismo país me hace sentir como tener un hermano cerca.

Marisa, gracias por ser tan buena amiga :) ya que acabamos los 3 en Suecia porque no nos vemos mas a menudo? Si en Madrid íbamos a la maquina de café siempre que podíamos!

Thank you to **Vanilla Cupcake**! You are an amazing friend, you are a great scientist and you have this amazing positive energy all the time (oh that sounded pseudoscientific!). **Emmanuel** gracias por ser tan buen amigo. En donde coincidimos me sentía como si tuviera un viejo amigo de toda la vida (será porque éramos los mas viejos?). **Victor**, thank you for all the not funny jokes we were the only ones laughing to :D **Miroslav**, **Trevor**, **Si-Pei**, **Erik**, **Matthias**, **Igor**, **Josh**, **Ashish**, **Anne Sarah** and all the BIG4 supervisors, thank you all.

Fredrik and **Isabel**, thank you for allowing me to be part of your teams for a short time. Both my stay in Stockholm and Madrid taught me a lot. I can never forget how challenging was for me the deep scientific questions you were asking me and how much these questions changed the way I used to think.

Saeed jaan, doctor vaaghean dastet dard nakone baabate hameye komakaayi ke bem kardi tu in moddat. Kaamelan mesle baraadaram budi tamaame in moddat.

Thank you **Chema**. I owe you a lot of what I am now. You were the first person who gave me the opportunity to enter a research environment and learn. I have really been lucky to know you early as a student. Also a huge thank you to everyone from el departamento de bioquimica y biologia molecular de la facultad de veterinaria, especially **Azar** and **Rafa**. Azar I owe you you also a lot, but I didnt listen to you! As you see I finished as a biologist unfortunately :p Rafa, la proxima vez que vayamos a matalascañas conduzco yo ida y vuelta ;)

Nacho, no se de verdad como agradecerte por todo, Eres como un padre para mi, siempre has sido muy generoso, siempre me has apoyado y siempre seré tu estudiante.

Manuel, hermano, no me puedo creer que en el poco de tiempo que nos hemos conocido, hayamos compartido tantas cosas y que te hayas vuelto, como siempre dices, mi hermano :)

Thank you to **Good Constellation** team! **Ana**, **Joana**, **Raúl**, **Nataly**, **Lenard** and **Matej**, you are such an amazingly crazy team! Let's organize our meetings regularly for many years to come.

Zahra merci baabate hameye in moddat. Rupushe aazmaayeshgaaham yaadete? Merci baabate un rupushe kaamelan khaas :p

At the end, but most importantly, **maman**, **baba**, **Hamed**, merci!

And finally, Thank you all.

PS: I hope for the next thesis I will come up with a shorter and funnier acknowledgement, this one resulted not that funny after all!

References

Antipov D., Hartwick N., Shen M., Raiko M., Lapidus A., Pevzner P.A. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics. 32:btw493.

Beaulieu J.M., O'Meara B.C. 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution (N. Y). 69:1036–1043.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst. Biol. 65:583–601.

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. Mol. Phylogenet. Evol. 69:313–319.

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

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:e1003537.

Burbrink F.T., Pyron R.A. 2009. How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in new world ratsnakes (tribe Lampropeltini)? Evolution (N. Y). 64:934–943.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution (N. Y). 72:2308–2324.

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 Res. 14:1147–1159.

Chevreux B., Wetter T., Suhai S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Comput. Sci. Biol. Proc. Ger. Conf. Bioinforma.:45–56.

Darwin C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: John Murray.

Dierckxsens N., Mardulyn P., Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:gkw955.

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). Syst. Entomol.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. B Biol. Sci. 279:1300–1309.
Fitzjohn R.G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.
FitzJohn R.G. 2012. Diversitree : comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.
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. Syst. Biol. 59:307–321.
Harvey P.H., May R.M., Nee S. 1994. Phylogenies Without Fossils. Evolution (N. Y). 48:523.
Hey J. 1992. Using Phylogenetic Trees to Study Speciation and Extinction. Source Evol. Evol.
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.
Höhna S., Freyman W.A., Nolen Z., Huelsenbeck J., May M.R., Moore B.R. 2019. A Bayesian Approach for Estimating Branch-Specific Speciation and Extinction Rates. bioRxiv.:555805.
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. Syst. Biol. 65:726–736.
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.
Hooker J.D. 1854. Himalayan journals; or, Notes of a naturalist in Bengal, the Sikkim and Nepal Himalayas, the Khasia Mountains, &c. London,: J. Murray,.
Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 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.
Kozlov A.M., Darriba D., Flouri T., Morel B., Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35:4453–4455.
Kubo T., Iwasa Y. 1995. Inferring the Rates of Branching and Extinction from Molecular Phylogenies. Evolution (N. Y). 49:694.
Landis M.J., Freyman W.A., Baldwin B.G. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. Evolution (N. Y). 72:2343–2359.
Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. Syst. Biol. 56:701–710.

Magallón S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution (N. Y). 55:1762–1780.
Magnuson-Ford K., Otto S.P. 2012. Linking the Investigations of Character Evolution and Species Diversification. Am. Nat. 180:225–245.
Meyer A.L.S., Román-Palacios C., Wiens J.J. 2018. BAMM gives misleading rate estimates in simulated and empirical datasets. Evolution (N. Y). 72:2257–2266.
Meyer A.L.S., Wiens J.J. 2018. Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. Evolution (N. Y). 72:39–53.
Minh B.Q., Hahn M.W., Lanfear R. 2020. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. Mol. Biol. Evol.
Morlon H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.
Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. 108:16327–16332.
Morlon H., Potts M.D., Plotkin J.B. 2010. Inferring the Dynamics of Diversification: A Coalescent Approach. PLoS Biol. 8:e1000493.
Nee S. 2006. Birth-Death Models in Macroevolution. Annu. Rev. Ecol. Evol. Syst. 37:1–17.
Nee S., Holmes E.C., May R.M., Harvey P.H. 1994a. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. - R. Soc. London, B. 344:77–82.
Nee S., May M.R., Harvey P.H. 1994b. The reconstructed evolutionary process. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 344:305–311.
Nee S., Mooers A.O., Harvey P.H. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci.
Neethiraj R., Hornett E.A., Hill J.A., Wheat C.W. 2017. Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. Mol. Ecol. 26:4990–5002.
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. Mol. Biol. Evol. 32:268–274.
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., Davis D.R., De Prins J., De Prins W., Epstein M.E., Gentili-Poole P., Gielis C., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S. (J C., Kozlov M. V, Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-T., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-C., Solis M.A., Tarmann G.M., Warren A.D., Weller S., Yakovlev R. V, Zolotuhin V. V, Andreas Z. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 1758:212–221.
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. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng M., Jiang R., Sun F., Zhang X., editors. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 158–170.

Paradis E. 2003. Analysis of diversification: Combining phylogenetic and taxonomic data. Proc. R. Soc. B Biol. Sci.
Paradis E. 2004. Can extinction rates be estimated without fossils? J. Theor. Biol.
Peña C., Malm T. 2012. VoSeq: A Voucher and DNA Sequence Web Application. PLoS One. 7:e39071.
Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. PLoS Biol. 6:0483–0489.
Purvis A. 2008. Phylogenetic Approaches to the Study of Extinction. Annu. Rev. Ecol. Evol. Syst. 39:301–319.
Pybus O.G., Harvey P.H. 2000. Testing macro–evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. London. Ser. B Biol. Sci. 267:2267–2272.
Quental T.B., Marshall C.R. 2010. Diversity dynamics: Molecular phylogenies need the fossil record. Trends Ecol. Evol. 25:435–441.
Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution (N. Y). 64:1816–1824.
Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLoS One. 9:e89543.
Rabosky D.L. 2016. Challenges in the estimation of extinction from molecular phylogenies: A response to Beaulieu and O'Meara. Evolution (N. Y). 70:218–228.
Rabosky D.L. 2018. BAMM at the court of false equivalency: A response to Meyer and Wiens. Evolution (N. Y). 72:2246–2256.
Rabosky D.L., Huang H. 2016. A Robust Semi-Parametric Test for Detecting Trait-Dependent Diversification. Syst. Biol. 65:181–193.
Rabosky D.L., Lovette I.J. 2008a. Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? Evolution (N. Y). 62:1866–1875.
Rabosky D.L., Lovette I.J. 2008b. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B Biol. Sci. 275:2363–2371.
Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic Models of Phylogeny and the Evolution of Diversity. J. Geol. 81:525–542.
Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution (N. Y). 59:2299–2311.
Ree R.H., Smith S.A. 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. Syst. Biol. 57:4–14.
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. Syst. Biol. 61:539–542.
Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics.
Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods. 9:811–814.
Simpson G.G. 1944. Tempo and Mode of Evolution. Columbia Classic in Evolution.
Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. 108:6187–6192.

Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.
Stanley S.M. 1980. Macroevolution, Pattern and process. W H Freeman & Co.
Twort V.G., Minet J., Wheat C.W., Wahlberg N. 2020. Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life. bioRxiv.:2020.08.18.255182.
Wahlberg N., Peña C., Ahola M., Wheat C.W., Rota J. 2016. PCR primers for 30 novel gene regions in the nuclear genomes of lepidoptera. Zookeys. 2016:129–141.
Wahlberg N., Wheat C.W. 2008. Genomic Outposts Serve the Phylogenomic Pioneers: Designing Novel Nuclear Markers for Genomic DNA Extractions of Lepidoptera. Syst. Biol. 57:231–242.
Wahlberg N., Wheat C.W., Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One. 8:1–8.
Walker T.D., Valentine J.W. Equilibrium Models of Evolutionary Species Diversity and the Number of Empty Niches. Am. Nat. 124:887–899.
Wallace A.R. 1869. The Malay Archipelago: The Land of the Orang-Utan and the Bird of Paradise. A Narrative of Travel, with Studies of Man and Nature. New York: Harper & Brothers.
Wood D.E., Salzberg S.L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15:R46.
Yule G.U. 1925. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. Philos. Trans. R. Soc. B Biol. Sci. 213:21–87.
Zahiri R., Holloway J.D., Kitching I.J., Lafontaine J.D., Mutanen M., Wahlberg N. 2012. Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). Syst. Entomol. 37:102–124.

Paper I

A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies

Leidys Murillo-Ramos[1,2], Gunnar Brehm[3], Pasi Sihvonen[4], Axel Hausmann[5], Sille Holm[6], Hamid Reza Ghanavi[2], Erki Õunap[6,7], Andro Truuverk[8], Hermann Staude[9], Egbert Friedrich[10], Toomas Tammaru[6] and Niklas Wahlberg[2]

[1] Grupo Biología Evolutiva, Department of Biology, Universidad de Sucre, Sincelejo, Sucre, Colombia
[2] Systematic Biology Group, Department of Biology, Lund University, Lund, Sweden
[3] Institut für Zoologie und Evolutionsbiologie, Phyletisches Museum, Jena, Germany
[4] Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland
[5] Staatliche Naturwissenschaftliche Sammlungen Bayerns, München, Germany
[6] Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
[7] Estonian University of Life Sciences, Institute of Agricultural and Environmental Sciences, Tartu, Estonia
[8] Natural History Museum, University of Tartu, Tartu, Estonia
[9] LepsocAfrica, Magaliesburg, South Africa
[10] Berghoffsweg 5, Jena, Germany

ABSTRACT

Our study aims to investigate the relationships of the major lineages within the moth family Geometridae, with a focus on the poorly studied Oenochrominae-Desmobathrinae complex, and to translate some of the results into a coherent subfamilial and tribal level classification for the family. We analyzed a molecular dataset of 1,206 Geometroidea terminal taxa from all biogeographical regions comprising up to 11 molecular markers that includes one mitochondrial (COI) and 10 protein-coding nuclear gene regions (wingless, ArgK, MDH, RpS5, GAPDH, IDH, Ca-ATPase, Nex9, EF-1alpha, CAD). The molecular data set was analyzed using maximum likelihood as implemented in IQ-TREE and RAxML. We found high support for the subfamilies Larentiinae, Geometrinae and Ennominae in their traditional scopes. Sterrhinae becomes monophyletic only if *Ergavia* Walker, *Ametris* Hübner and *Macrotes* Westwood, which are currently placed in Oenochrominae, are formally transferred to Sterrhinae. Desmobathrinae and Oenochrominae are found to be polyphyletic. The concepts of Oenochrominae and Desmobathrinae required major revision and, after appropriate rearrangements, these groups also form monophyletic subfamily-level entities. Oenochrominae *s.str.* as originally conceived by Guenée is phylogenetically distant from *Epidesmia* and its close relatives. The latter is hereby described as the subfamily Epidesmiinae Murillo-Ramos, Sihvonen & Brehm, **subfam. nov.** Epidesmiinae are a lineage of "slender-bodied Oenochrominae" that include the genera *Ecphyas* Turner, *Systatica* Turner, *Adeixis* Warren, *Dichromodes* Guenée, *Phrixocomes* Turner, *Abraxaphantes* Warren, *Epidesmia* Duncan & Westwood and *Phrataria* Walker. Archiearinae are monophyletic when *Dirce* and *Acalyphes* are formally transferred to Ennominae. We

Submitted 6 February 2019
Accepted 1 July 2019
Published 27 August 2019

Corresponding authors
Leidys Murillo-Ramos,
leidys.murillo@unisucre.edu.co
Niklas Wahlberg,
niklas.wahlberg@biol.lu.se

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 32

DOI **10.7717/peerj.7386**

How to cite this article Murillo-Ramos L, Brehm G, Sihvonen P, Hausmann A, Holm S, Reza Ghanavi H, Õunap E, Truuverk A, Staude H, Friedrich E, Tammaru T, Wahlberg N. 2019. A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies. **PeerJ 7:e7386 DOI 10.7717/peerj.7386**

also found that many tribes were para- or polyphyletic and therefore propose tens of taxonomic changes at the tribe and subfamily levels. Archaeobalbini **stat. rev.** Viidalepp (Geometrinae) is raised from synonymy with Pseudoterpnini Warren to tribal rank. Chlorodontoperini Murillo-Ramos, Sihvonen & Brehm, **trib. nov.** and Drepanogynini Murillo-Ramos, Sihvonen & Brehm, **trib. nov.** are described as new tribes in Geometrinae and Ennominae, respectively.

Subjects Entomology, Evolutionary Studies, Taxonomy, Zoology
Keywords New subfamily, Phylogeny, Moths, Epidesmiinae, Taxonomy, Loopers

INTRODUCTION

Geometridae are the second most species-rich family of Lepidoptera, with approximately 24,000 described species (number from *Van Nieukerken et al. (2011)* updated by the authors) found in all regions except Antarctica. The monophyly of Geometridae is well supported based on distinctive morphological characters (*Cook & Scoble, 1992*; *Scoble, 1992*; *Minet & Scoble, 1999*). In particular, adult members of the family possess paired tympanal organs at the base of the abdomen, while in larvae the prolegs are reduced to two pairs in almost all species, which causes the larvae to move in a looping manner (*Minet & Scoble, 1999*).

The phylogenetic relationships of the major subdivisions of Geometridae have been studied based on molecular data, which have contributed to the understanding of the evolutionary relationships within the family (*Abraham et al., 2001*; *Yamamoto & Sota, 2007*; *Sihvonen et al., 2011*). Eight subfamilies are currently recognized in Geometridae (*Sihvonen et al., 2011*). Several recent molecular and morphological studies have attempted to confirm the monophyly or clarify the taxonomy of most of these groups, for instance: Sterrhinae (*Holloway, 1997*; *Hausmann, 2004*; *Sihvonen & Kaila, 2004*; *Õunap, Viidalepp & Saarma, 2008*), Larentiinae (*Holloway, 1997*; *Mironov, 2003*; *Viidalepp, 2006*, *2011*; *Hausmann & Viidalepp, 2012*; *Õunap, Viidalepp & Truuverk, 2016*), Desmobathrinae (*Holloway, 1996*; *Hausmann, 2001*), Archiearinae (*Hausmann, 2001*; *Young, 2006*), Oenochrominae (*Holloway, 1996*; *Scoble & Edwards, 1990*; *Cook & Scoble, 1992*; *Hausmann, 2001*; *Young, 2006*), Geometrinae (*Cook et al., 1994*; *Pitkin, 1996*; *Hausmann, 2001*; *Ban et al., 2018*), Orthostixinae (*Holloway, 1997*) and Ennominae (*Holloway, 1994*; *Pitkin, 2002*; *Beljaev, 2006*; *Young, 2006*; *Wahlberg et al., 2010*; *Õunap et al., 2011*; *Skou & Sihvonen, 2015*; *Sihvonen, Staude & Mutanen, 2015*), but questions remain. An important shortcoming is that our understanding of geometrid systematics is biased towards the long-studied European fauna, whereas the highest diversity of this family is in the tropics, which are still largely unexplored (*Brehm et al., 2016*). Many species remain undescribed and there are many uncertainties in the classification of tropical taxa.

One of the most comprehensive phylogenetic studies on Geometridae to date was published by *Sihvonen et al. (2011)*. They analyzed a data set of 164 taxa and up to eight genetic markers, and the most species-rich subfamilies were confirmed as monophyletic. However, the systematic positions of Oenochrominae and Desmobathrinae remained uncertain due to low taxon sampling and genetic markers, and both subfamilies were

suspected to be polyphyletic. Moreover, because of taxonomic uncertainty, many geometrid genera, especially among tropical taxa, remained unassigned to any tribe.

This study is the first in a series of papers that investigate the phylogenetic relationships of Geometridae on the basis of global sampling. Our dataset comprises 1,192 terminal taxa of Geometridae and 14 outgroup taxa, with samples from all major biomes, using up to 11 molecular markers. Our paper includes an overview of the relationships of the major lineages within the family, with the particular aim of defining the limits and finding the phylogenetic affinities of the subfamilies, with a focus on Oenochrominae and Desmobathrinae. Further papers in the series will focus on particular subfamilies and regions, and will build upon the taxonomic changes proposed in the present article: e.g., relationships in Sterrhinae (P. Sihvonen et al., 2019, unpublished data), New World taxa (G. Brehm et al., 2019, unpublished data), Larentiinae (E. Õunap et al., 2019, unpublished data) and the ennomine tribe Boarmiini (L. Murillo-Ramos et al., 2019, unpublished data).

Oenochrominae and Desmobathrinae are considered the most controversial subfamilies in Geometridae. A close relationship of these subfamilies has been proposed both in morphological (*Meyrick, 1889*; *Cook & Scoble, 1992*; *Holloway, 1996*) and in molecular studies (*Sihvonen et al., 2011*; *Ban et al., 2018*). In early classifications, species of Desmobathrinae and Oenochrominae were classified in the family Monocteniadae (*Meyrick, 1889*), which is currently considered a junior synonym of Oenochrominae Guenée. Meyrick diagnosed them on the basis of the position of the R veins in the hindwing and Sc+R1 in the forewing (*Scoble & Edwards, 1990*). However, the classification proposed by Meyrick was not fully supported by subsequent taxonomic work (*Scoble & Edwards, 1990*; *Cook & Scoble, 1992*; *Holloway, 1996*). Too often, Oenochrominae was used for geometrids that could not be placed in other subfamilies, and at some point, even included Hedylidae, the moth-butterflies (*Scoble, 1992*). Unsurprisingly, many taxa formerly classified in Oenochrominae have recently been shown to be misplaced (*Holloway, 1997*; *Staude, 2001*; *Sihvonen & Staude, 2011*; *Staude & Sihvonen, 2014*). In *Scoble & Edwards (1990)*, the family concept of Oenochrominae was restricted to the robust-bodied Australian genera, with one representative from the Oriental region. *Scoble & Edwards (1990)* were not able to find synapomorphies to define Monocteniadae *sensu* Meyrick, and referred back to the original grouping proposed by *Guenée (1858)*. They restricted Oenochrominae to a core clade based on male genitalia: the diaphragm dorsal to the anellus is fused with the transtilla to form a rigid plate. Additionally, *Cook & Scoble (1992)* suggested that the circular form of the lacinia and its orientation parallel to the tympanum was apomorphic for these robust-bodied Oenochrominae.

In an extensive morphological study, *Holloway (1996)* delimited the subfamily Desmobathrinae to include species with slender appendages and bodies previously assigned to Oenochrominae. According to *Holloway (1996)*, Desmobathrinae comprises two tribes: Eumeleini and Desmobathrini. However, no synapomorphies were found to link the two tribes. *Holloway (1996)* noted that the modification of the tegumen of the male genitalia was variable in both groups but that the reduction of cremastral spines in the pupa from eight to four in *Ozola* Walker, 1861 and *Eumelea* Duncan & Westwood, 1841 provided evidence of a close relationship between Eumeleini and Desmobathrini.

Currently, 328 species (76 genera) are included in Oenochrominae, and 248 species (19 genera) are assigned to Desmobathrinae (*Beccaloni et al., 2003*; *Sihvonen et al., 2011*; *Sihvonen, Staude & Mutanen, 2015*).

Most recent molecular phylogenies have shown Oenochrominae and Desmobathrinae to be intermingled (*Sihvonen et al., 2011*; *Ban et al., 2018*), but previous taxon sampling was limited to eight and four species, respectively. The poor taxon sampling and unresolved relationships around the oenochromine and desmobathrine complex called for additional phylogenetic studies to clarify the relationships of these poorly known taxa within Geometridae. We hypothesize that both Oenochrominae and Desmobathrinae are para- or polyphyletic assemblages, and we address this hypothesis with studying 29 terminal taxa of Oenochrominae and 11 representatives of Desmobathrinae, mostly from the Australian and Oriental Regions.

MATERIALS AND METHODS

Material acquisition, taxon sampling and species identification

In addition to 461 terminal taxa with published sequences (see Data S1), we included sequences from 745 terminal taxa in our study (Data S1). Representative taxa of all subfamilies recognized in Geometridae were included, except for the small subfamily Orthostixinae for which most molecular markers could not be amplified successfully. A total of 93 tribes are represented in this study following recent phylogenetic hypotheses and classifications (*Sihvonen et al., 2011*; *Wahlberg et al., 2010*; *Sihvonen, Staude & Mutanen, 2015*; *Õunap, Viidalepp & Truuverk, 2016*; *Ban et al., 2018*). In addition, 14 non-geometrid species belonging to other families of Geometroidea were included as outgroups based on the hypothesis proposed by *Regier et al. (2009*, *2013)*. Where possible, two or more samples were included per tribe and genus, especially for species-rich groups that are widely distributed and in cases where genera were suspected to be poly- or paraphyletic. We emphasized type species or species similar to type species, judged by morphological characters and/or genetic similarity of DNA barcodes in order to better inform subsequent taxonomic work, to favor nomenclatorial stability and to establish the phylogenetic positions of genera unassigned to tribes.

Sampled individuals were identified by the authors using appropriate literature, by comparing them with type material from different collections, museums and DNA barcode sequences. Moreover, we compiled an illustrated catalog of all Archiearinae, Desmobathrinae and Oenochrominae taxa included in this study, to demonstrate their morphological diversity and to facilitate subsequent verification of our identifications. This catalog contains images of all analyzed specimens of the above-mentioned taxa as well as photographs of the respective type material (Data S2). Further taxa from other subfamilies will be illustrated in other papers (G. Brehm et al., 2019, unpublished data, P. Sihvonen et al., 2019, unpublished data, E. Õunap et al., 2019, unpublished data). Some of the studied specimens could not yet be assigned to species, and their identifications are preliminary, particularly for (potentially undescribed) tropical species. Taxonomic data, voucher IDs, number of genes, current systematic placement and references to relevant literature with regard to tribal assignment, are shown in Data S1.

Molecular techniques

DNA was extracted from one to three legs of specimens either preserved in ethanol or dry. In a few cases, other sources of tissue were used, such as parts of larvae. The remaining parts of specimens were preserved as vouchers deposited in the collections of origin, both public and private (eventually private material will be deposited in public museum collections). Genomic DNA was extracted and purified using a NucleoSpin® Tissue Kit (MACHEREY-NAGEL, Düren, Germany), following the manufacturer's protocol. DNA amplification and sequencing were carried out following protocols proposed by *Wahlberg & Wheat (2008)* and *Wahlberg et al. (2016)*. PCR products were visualized on agarose gels. PCR products were cleaned enzymatically with Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher Scientific, Waltham, MA, USA) and sent to Macrogen Europe (Amsterdam, Netherlands) for Sanger sequencing. One mitochondrial (*cytochrome oxidase subunit I*, COI) and 10 protein-coding nuclear gene regions, *carbamoylphosphate synthetase* (CAD), *Ribosomal Protein S5* (RpS5), *wingless* (wgl), *cytosolic malate dehydrogenase* (MDH), *glyceraldehydes-3-phosphate dehydrogenase* (GAPDH), *Elongation factor 1 alpha* (EF-1alpha), *Arginine Kinase* (ArgK), *Isocitrate dehydrogenase* (IDH), *sorting nexin-9-like* (Nex9) and *sarco/endoplasmic reticulum calcium ATPase* (Ca-ATPase), were sequenced. To check for potential misidentifications, DNA barcode sequences were compared to those in BOLD (*Ratnasingham & Hebert, 2007*) where references of more than 21,000 geometrid species are available, some 10,000 of them being reliably identified to Linnean species names (*Ratnasingham & Hebert, 2007*). GenBank accession numbers for sequences used in this study are provided in Data S1.

Alignment and cleaning sequences

Multiple sequence alignments were carried out in MAFFT as implemented in Geneious v.11.0.2 (Biomatters, http://www.geneious.com/) for each gene based on a reference sequence of Geometridae downloaded from the database VoSeq (*Peña & Malm, 2012*). The alignment of each gene was carefully checked by eye relative to the reference sequence, taking into account the respective genetic codes and reading frames. Heterozygous positions were coded with IUPAC codes. Sequences with bad quality were removed from the alignments. Aligned sequences were uploaded to VoSeq (*Peña & Malm, 2012*) and then assembled into a dataset comprising 1,206 taxa. The final dataset had a concatenated length of 7665 bp including gaps. To check for possible errors in alignments, potentially contaminated or identical sequences and misidentifications, we constructed maximum-likelihood trees for each gene. These preliminary analyses were conducted using RAxML-HPC2 V.8.2.10 (*Stamatakis, 2014*) on the web-server CIPRES Science Gateway (*Miller, Pfeiffer & Schwartz, 2010*). The final data set included at least three genes per taxon except for *Oenochroma vinaria* (Guenée, 1858), *Acalyphes philorites* Turner, 1925, *Dirce lunaris* (Meyrick, 1890), *D. aesiodora* Turner, 1922, *Furcatrox australis* (Rosenstock, 1885), *Chlorodontopera mandarinata* (Leech, 1889), *Chlorozancla falcatus* (Hampson, 1895), *Pamphlebia rubrolimbraria* (Guenée, 1858) and *Thetidia albocostaria* (Bremer, 1864). For these taxa, included in studies by *Young (2006)* and *Ban et al. (2018)*, only two markers were available. The final data matrix included 32% missing data.

Tree search strategies and model selection

We ran maximum likelihood analyses with a data set partitioned by gene and codon position using IQ-TREE V1.6.10 (*Nguyen et al., 2015*) and data partitioned by codon in RAxML (*Stamatakis, 2014*). Best-fitting substitution models were selected by ModelFinder, which is a model-selection method that incorporates a model of flexible rate heterogeneity across sites (*Kalyaanamoorthy et al., 2017*). ModelFinder implements a greedy strategy as implemented in PartitionFinder that starts with the full partitioned model and consequentially merges partitions (MFP+MERGE option) until the model fit does not increase (*Lanfear et al., 2012*). After the best model has been found, IQ-TREE starts the tree reconstruction under the best model scheme. The phylogenetic analyses were carried out with the *-spp* option that allowed each partition to have its own evolutionary rate. The RAxML-HPC2 V.8.2.10 analysis was carried out on CIPRES using the GTR+CAT option.

Support for nodes was evaluated with 1,000 ultrafast bootstrap (UFBoot2) approximations (*Hoang et al., 2018*) in IQ-TREE, and SH-like approximate likelihood ratio test (*Guindon et al., 2010*). Additionally, we implemented rapid bootstrap (RBS) in RAxML (*Stamatakis, Hoover & Rougemont, 2008*). To reduce the risk of overestimating branch supports in UFBoot2 test, we implemented *-bnni* option, which optimizes each bootstrap tree using a hill-climbing nearest neighbor interchange search. Trees were visualized and edited in FigTree v1.4.3 software (*Rambaut, 2012*). The final trees were rooted with species of the families Sematuridae, Epicopeiidae, Pseudobistonidae and Uraniidae following previous hypotheses proposed in *Regier et al. (2009, 2013)*, *Rajaei et al. (2015)* and *Heikkilä et al. (2015)*.

Taxonomic decisions

The electronic version of this article in Portable Document Format (PDF) will represent a published work according to the International Commission on Zoological Nomenclature (*International Commission on Zoological Nomenclature, 2012*), and hence the new names contained in the electronic version are effectively published under that Code from the electronic edition alone. This published work and the nomenclatural acts it contains have been registered in ZooBank. The ZooBank LSIDs (Life Science Identifiers) can be resolved and the associated information viewed through any standard web browser by appending the LSID to the prefix http://zoobank.org/. For this publication: LSIDurn:lsid:zoobank.org:pub:662A9A18-B620-45AA-B4B1-326086853316. The online version of this work is archived and available from the following digital repositories: PeerJ, PubMed Central and CLOCKSS.

RESULTS

Searching strategies and model selection

The ModelFinder analysis resulted in 26 partitions with associated best-fit models (Table 1). IQ-TREE and RAxML analyses resulted in trees with nearly identical topology. Also, the different methods of evaluating robustness tended to agree in supporting the same nodes. However, in most of the cases UFBoot2 from IQ-TREE showed higher support values compared to RBS in RAxML (RAxML tree with support values is shown in

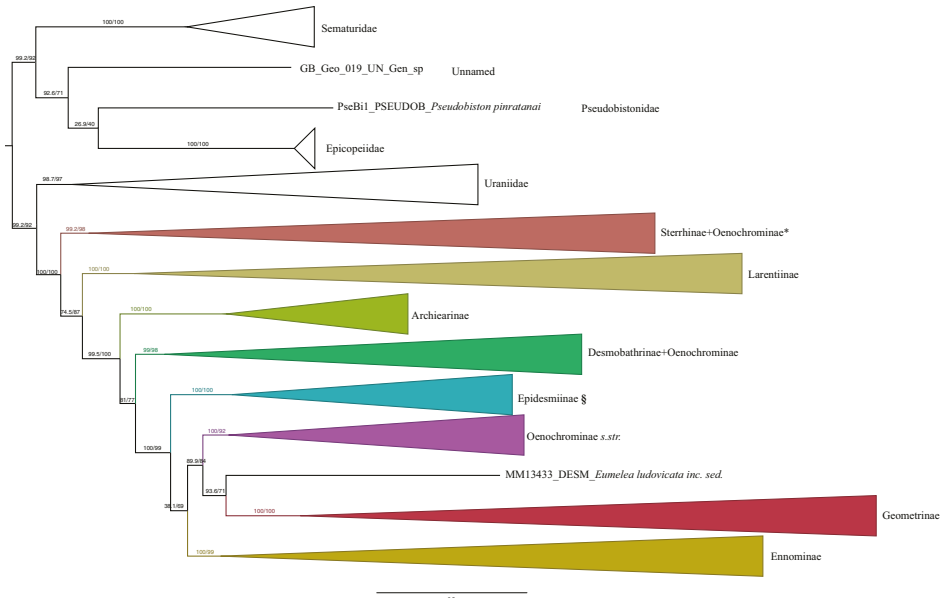


Figure 1 Evolutionary relationships of major groups of the family Geometridae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). *Formal taxonomic treatment will be dealt with in P. Sihvonen et al., 2019, unpublished data. § Epidesmiinae **subfam. nov.** See Oenochrominae section for more details. Full-size DOI: 10.7717/peerj.7386/fig-1

Data S3). SH-like and UFBoot2 performed similarly, with UFBoot2 showing slightly higher values, and both tended to show high support for the same nodes (Fig. 1). As noted by the authors of IQ-TREE, values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*).

General patterns in the phylogeny of Geometridae

Analyses of the dataset of 1,206 terminal taxa, comprising up to 11 markers and an alignment length of 7,665 bp recovered topologies with many well-supported clades. About 20 terminal taxa are recovered as very similar genetically and they are likely to represent closely related species, subspecies or specimens of a single species. The examination of their taxonomic status is not the focus of this study, so the number of unique species in the analysis is slightly less than 1,200. Our findings confirm the monophyly of Geometridae (values of SH-like, UFBoot2= 100) (Fig. 1). The general patterns in our phylogenetic hypotheses suggest that Sterrhinae are the sister group to the rest of Geometridae. This subfamily is recovered as monophyletic when three genera traditionally included in Oenochrominae are considered to belong to Sterrhinae

Table 1 Evolutionary models recovered in ModelFinder.

Evolutionary models	Codon position	Data type
SYM+R5	ArgK_pos1	Nuclear
SYM+R4	ArgK_pos2_Ca-ATPase_pos2	Nuclear
GTR+F+R6	ArgK_pos3	Nuclear
GTR+F+R5	Ca-ATPase_pos1_IDH_pos1	Nuclear
SYM+I+G4	Ca-ATPase_pos3	Nuclear
SYM+I+G4	CAD_pos1	Nuclear
K3P+I+G4	CAD_pos2	Nuclear
GTR+F+R7	CAD_pos3	Nuclear
TIM2+F+I+G4	COI_pos1	Mitochondrial
K2P+R8	COI_pos2_MDH_pos2_RpS5_pos2_WntGeo_pos2	Mitochondrial/Nuclear
GTR+F+ASC+R10	COI_pos3	Mitochondrial
TIM2e+R10	EF1a_pos1	Nuclear
TIM+F+I+G4	EF1a_pos2	Nuclear
SYM+R10	EF1a_pos3_GAPDH_pos3_RpS5_pos3	Nuclear
TVM+F+I+G4	GAPDH_pos1	Nuclear
SYM+I+G4	GAPDH_pos2	Nuclear
GTR+F+R4	IDH_pos2	Nuclear
SYM+R6	IDH_pos3	Nuclear
GTR+F+I+G4	MDH_pos1	Nuclear
SYM+I+G4	MDH_pos3	Nuclear
SYM+I+G4	Nex9_pos1	Nuclear
K3P+I+G4	Nex9_pos2	Nuclear
GTR+F+R6	Nex9_pos3	Nuclear
SYM+I+G4	RpS5_pos1	Nuclear
GTR+F+I+G4	WntGeo_pos1	Nuclear
SYM+R7	WntGeo_pos3	Nuclear

(see details below). Tribes in Sterrhinae, such as Timandriini, Rhodometrini, Lythriini, Rhodostrophiini and Cyllopodini, are not recovered as monophyletic (Fig. 2). A detailed analysis, including formal changes to the classification of Sterrhinae, will be provided by P. Sihvonen et al., 2019, unpublished data.

The monophyly of Larentiinae is established in previous studies (*Sihvonen et al., 2011*; *Õunap, Viidalepp & Truuverk, 2016*) and our results are largely in agreement with their hypotheses. However, our results do not support the sister relationship between Sterrhinae and Larentiinae found in previous studies. Rather, we find that Sterrhinae are the sister to the rest of Geometridae. Within Larentiinae, in concordance with recent findings (*Sihvonen et al., 2011*; *Õunap, Viidalepp & Truuverk, 2016*; *Strutzenberger et al., 2017*), we find Dyspteridini as the sister group to the remaining Larentiinae (Fig. 3). Phylogenetic relationships within Larentiinae were treated in detail by *Õunap, Viidalepp & Truuverk (2016)*. Further details of the analyses and changes to the classification of Larentiinae will be discussed by G. Brehm et al., 2019, unpublished data and E. Õunap et al., 2019, unpublished data.

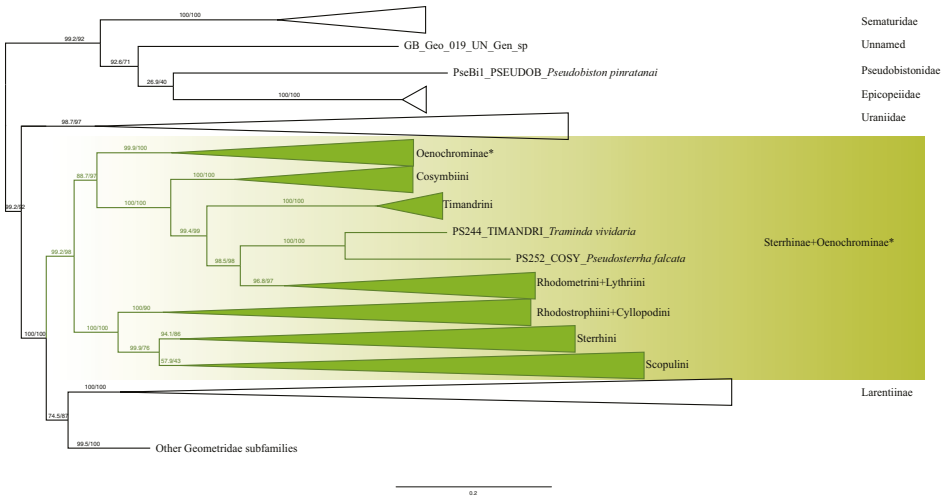


Figure 2 Evolutionary relationships of the subfamily Sterrhinae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). *Formal taxonomic treatment will be dealt with in P. Sihvonen et al., 2019, unpublished data. Full-size DOI: 10.7717/peerj.7386/fig-2

Archiearinae are represented by more taxa than in a previous study (*Sihvonen et al., 2011*). Archiearinae grouped as sister to Oenochrominae + Desmobathrinae complex + *Eumelea* + Geometrinae and Ennominae (Fig. 4). The monophyly of this subfamily is well supported (values of SH-like, UFBoot2 = 100). However, as in the previous study (*Sihvonen et al., 2011*), the Australian genera *Dirce* Prout, 1910 and *Acalyphes* Turner, 1926 are not part of Archiearinae but can clearly be assigned to Ennominae. Unlike previously assumed (e.g., *McQuillan & Edwards, 1996*), the subfamily Archiearinae probably does not occur in Australia, despite superficial similarities of *Dirce*, *Acalyphes* and Archiearinae.

Desmobathrinae were shown to be paraphyletic by *Sihvonen et al. (2011)*. In our analysis, the monophyly of this subfamily is not recovered either, as we find two genera traditionally placed in Oenochrominae (i.e. *Zanclopteryx* Herrich-Schäffer, (1855) and *Racasta* Walker, 1861) nested within Desmobathrinae (Fig. 4). We formally transfer these genera to Desmobathrinae. In the revised sense, Desmobathrinae form a well-supported group with two main lineages. One of them comprises *Ozola* Walker, 1861, *Derambila* Walker, 1863 and *Zanclopteryx*. This lineage is sister to a well-supported clade comprising *Conolophia* Warren, 1894, *Noreia* Walker, 1861, *Leptoctenopsis* Warren, 1897, *Racasta*, *Ophiogramma* Hübner, 1831, *Pycnoneura* Warren, 1894 and *Dolichoneura* Warren, 1894.

Oenochrominae in the broad sense are not a monophyletic group. However, Oenochrominae *sensu stricto* (*Scoble & Edwards, 1990*) form a well-supported lineage

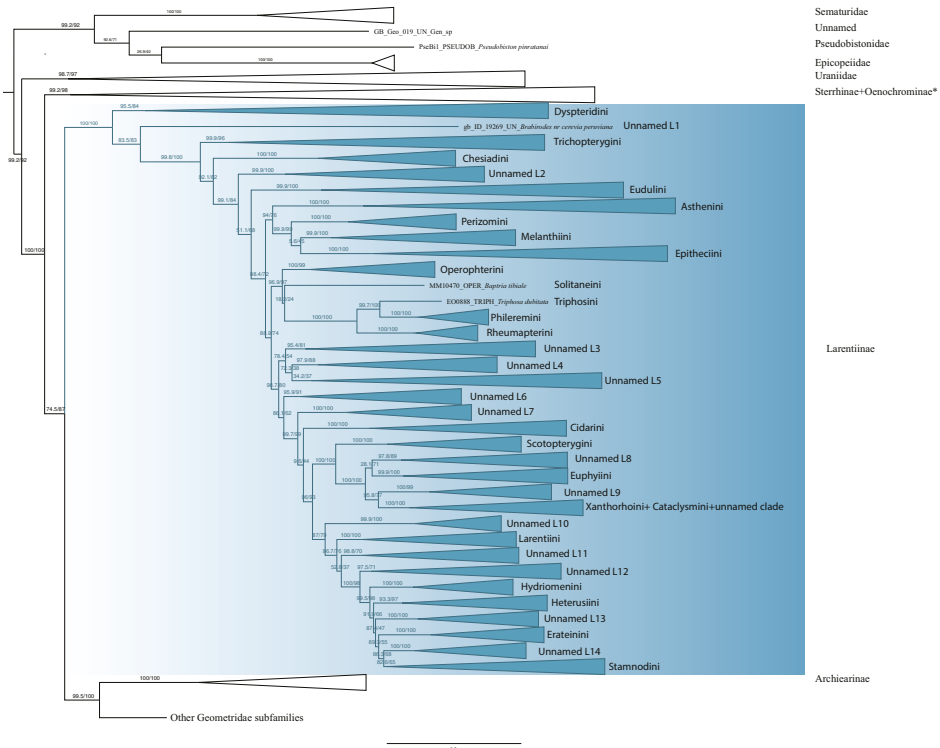


Figure 3 Evolutionary relationships of the subfamily Larentiinae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). *Formal taxonomic treatment will be dealt with in P. Sihvonen et al., 2019, unpublished data. Full-size DOI: 10.7717/peerj.7386/fig-3

comprising two clades. One of them contains a polyphyletic *Oenochroma* with *Oenochroma infantilis* Prout, 1910 being sister to *Dinophalus* Prout, 1910, *Hypographa* Guenée, 1858, *Lissomma* Warren, 1905, *Sarcinodes* Guenée, 1858 and two further species of *Oenochroma*, including the type species *Oenochroma vinaria* Guenée, 1858. The other clade comprises *Monoctenia* Guenée, 1858, *Onycodes* Guenée, 1858, *Parepisparis* Bethune-Baker, 1906, *Anticteniа* Prout, 1910, *Arthodia* Guenée, 1858, *Gastrophora* Guenée, 1858 and *Homospora* Turner, 1904 (Fig. 4). Most of the remaining genera traditionally placed in Oenochrominae, including e.g. *Epidesmia* Duncan & Westwood, 1841, form a well-supported monophyletic clade that is sister to Oenochrominae *s.str.* + *Eumelea ludovicata* + Geometrinae + Ennominae assemblage.

The genus *Eumelea* Duncan & Westwood, 1841 has an unclear phylogenetic position in our analyses. The IQ-TREE result suggests *Eumelea* to be sister to the subfamily

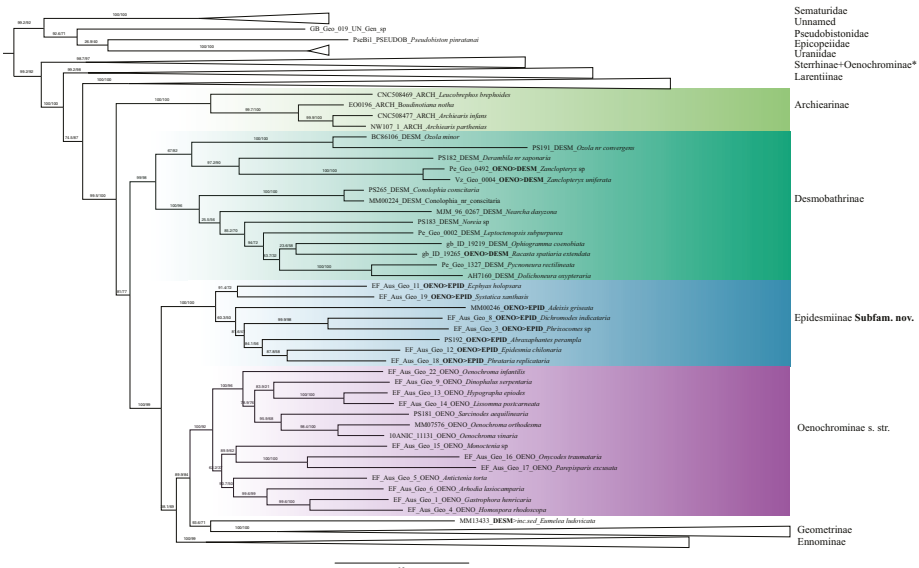


Figure 4 Phylogenetic relationships of the subfamilies Archierinae, Desmobathrinae, Epidesmiinae subfam. nov., Oenochrominae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). Taxonomic changes are indicated by a symobolized arrow >. *Formal taxonomic treatment will be dealt with in P. Sihvonen et al., 2019, unpublished data. Full-size DOI: 10.7717/peerj.7386/fig-4

Geometrinae (SH-like = 93.6, UFBoot2 = 71), whereas RAxML recovered *Eumelea* in Ennominae as sister of *Plutodes* Guenée, 1858 (RBS = 60).

The monophyly of Geometrinae is well supported (Fig. 5) and in IQ-TREE results Geometrinae are recovered as the sister-taxon of *Eumelea*. The *Eumelea* + Geometrinae clade is sister to Oenochrominae *s.str*. Although a recent phylogenetic study proposed several taxonomic changes (*Ban et al., 2018*), the tribal composition in Geometrinae is still problematic. Many tribes are recovered as paraphyletic. Our results suggest that *Ornithospila* Warren, 1894 and *Agathia* Guenée, 1858 form a lineage sister to the rest of Geometrinae. *Chlorodontopera* is placed as an isolated lineage sister to Aracimini, Neohipparchini, Timandromorphini, Geometrini and Comibaenini, which are recovered as monophyletic groups, respectively. Synchlorini are nested within Nemoriini in a well-supported clade (support branch SH-like = 98.3, UFBoot2 = 91, RBS = 93).
The monophyly of Pseudoterpnini could not be recovered, instead this tribe splits up into three well-defined groups. Several genera currently placed in Pseudoterpnini *s.l.* are recovered as an independent lineage clearly separate from Pseudoterpnini *s.str*. (SH-like, UFBoot2 = 100). *Xenozancla* Warren, 1893 is sister to a clade comprising Dysphaniini and Pseudoterpnini *s.str*. Hemitheini *sensu Ban et al. (2018)* are recovered as a well-supported

Figure 5 Evolutionary relationships of the subfamily Geometrinae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). Taxonomic changes are indicated by a symbolized arrow >. § New subfamily. Full-size DOI: 10.7717/peerj.7386/fig-5

clade. *Crypsiphona ocultaria* Meyrick, 1888 was resolved as a single lineage, close to *Lophostola* + Hemitheini .

Ennominae are strongly supported as monophyletic in IQ-TREE analyses (SH-like = 100, UFBoot2 = 99) whereas in RAxML the monophyly is weakly supported (RBS = 63). Detailed results concerning the classification, especially for the Neotropical taxa, will be presented by G. Brehm et al. (2019, unpublished data), but the main results are summarized here (Fig. 6). Very few tribes are monophyletic according to the results of the present study. One group of Neotropical taxa currently assigned to Gonodontini (unnamed E1), *Idialcis* Warren, 1906 (unnamed clade E2), Gonodontini *s.str.*, Gnophini, Odontoperini, unnamed clade E3, Nacophorini and Ennomini (*sensu* Beljaev, 2008) group together (SH-like = 90.3, UFBoot2 = 87). Ennomini were sister to this entire group. Campaeini is recovered as sister of Alsophilini + Wilemaniini and Colotoini. In turn they are sister to a clade comprising a number of taxa. These include the New Zealand genus *Declana* Walker, 1858 (unnamed E4) which appear as sister to a large complex including *Acalyphes* Turner, 1926 + *Dirce* Prout, 1910, Lithinini, intermixed with some genera currently placed in Nacophorini and Diptychini.

Neobapta Warren, 1904 and *Oenoptila* Warren, 1895 form an independent lineage (unnamed E5) sister to Theriini, which in turn form a supported clade with *Lomographa* (Baptini) (SH-like, UFBoot2 = 100). Likewise, we recovered *Erastria* Hübner, 1813 + *Metarranthis* Warren, 1894 (both as unnamed E5) as sister to Plutodini + Palyadini. The IQ-TREE analyses show Palyadini as a well-defined lineage, sister to *Plutodes*. However, in RAxML analyses, *Eumelea* and *Plutodes* group together and Palyadini cluster with a group of Caberini species. In the IQ-TREE analysis Apeirini formed a lineage with Hypochrosini, Epionini, *Sericosema* Warren, 1895 and *Ithysia* Hübner, 1825. This lineage is in turn sister of African *Drepanogynis* Guenée, 1858 which groups together with *Sphingomima* Warren, 1899, *Thenopa* Walker, 1855 and *Hebdomophruda* Warren, 1897. Caberini are sister to an unnamed clade composed of *Trotogonia* Warren, 1905, *Acrotomodes* Warren, 1895, *Acrotomia* Herrich-Schäffer 1855 and *Pyrinia* Hübner, 1818. Finally, our analyses recover a very large, well-supported clade comprising the tribes Macariini, Cassymini, Abraxini, Eutoeini and Boarmiini (SH-like = 100, UFBoot2 = 99). This large clade has previously been referred to informally as the "boarmiines" by *Forbes (1948)* and *Wahlberg et al. (2010)*. The tribe Cassymini is clearly paraphyletic: genera such as *Cirrhosoma* Warren, 1905, *Berberodes* Guenée, 1858, *Hemiphricta* Warren, 1906 and *Ballantiophora* Butler, 1881 currently included in Cassymini, cluster in their own clade together with *Dorsifulcrum* Herbulot, 1979 and *Odontognophos* Wehrli, 1951. We were unable to include Orthostixinae in the analyses, so we could not clarify the taxonomic position of this subfamily with regard to its possible synonymy with Ennominae (*Sihvonen et al., 2011*).

DISCUSSION

Optimal partitioning scheme and support values

The greedy algorithm implemented in ModelFinder to select the best-fitting partitioning scheme combined the codon partitions into 26 subsets (Table 2). These results are not different from previous studies that tested the performance of different data partitioning

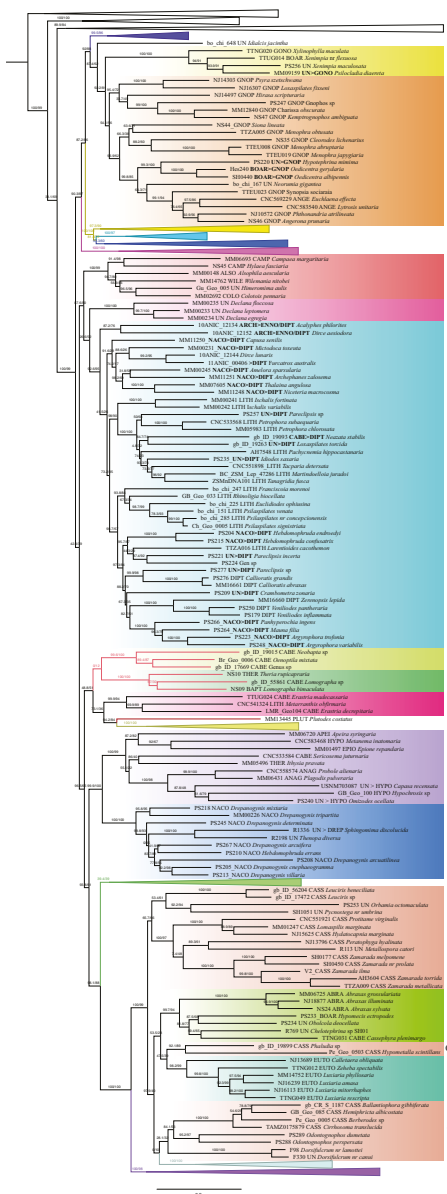

Epidesmiinae
Oenochrominae
Geometrinae
Unnamed E1*
Unnamed E2
Gonodontini
Gnophini
Odontoperini+Azelinini*
Unnamed E3*
Nacophorini*
Ennomini
Campaeini+ Alsophilini+Wilemaniini
+Prosopolophini
Unnamed E4
Diptychini + Lithinini syn. nov.
Unnamed E5*
Theriini+ Baptini
Unnamed E6
Plutodini
Palyadini
Apeirini+Epionini
+Anagogini+Hypochrosini
Drepanogynini trib. nov.
Caberini
Cassymini I
Abraxini
Cassymini II
Eutoeini
Cassymini III
Macariini
Boarmiini

Figure 6 Evolutionary relationships of the subfamily Ennominae. Numbers above branches are SH-aLRT support (%)/ultrafast bootstrap support, UFBoot2(%), for nodes to the right of the numbers. Values of SH ≥ 80 and UFBoot2 ≥ 95 indicate well-supported clades (*Trifinopoulos & Minh, 2018*). Taxonomic changes are indicated by a symobolized arrow >. *Formal taxonomic treatment will be dealt with in G. Brehm et al., 2019, unpublished data. Full-size DOI: 10.7717/peerj.7386/fig-6

schemes and found that in some cases partitioning by gene can result in suboptimal partitioning schemes and may limit the accuracy of phylogenetic analyses (*Rota, 2011*; *Lanfear et al., 2012*). However, we note that although the AIC and BIC values were lower when the data were partitioned by gene, the tree topology recovered was nevertheless almost the same as when data were partitioned by codon position, suggesting that much of the phylogenetic signal in the data is robust to partitioning schemes. As would be expected, the analyses resulted in some disagreements between the different measures of node support. Ultrafast bootstrap gave the highest support values, followed by SH-like and finally standard bootstrap as implemented in RAxML gave the lowest. Although support indices obtained by these methods are not directly comparable, differences in node support of some clades can be attributed to the small number of markers, insufficient phylogenetic signal or saturated divergence levels (*Guindon et al., 2010*).

Current understanding of Geometridae phylogeny and taxonomic implications

Geometridae Leach, 1815

The phylogenetic hypothesis presented in this study is by far the most comprehensive to date in terms of the number of markers, sampled taxa and geographical coverage. In total, our sample includes 814 genera, thus representing 41% of the currently recognized Geometridae genera (*Scoble & Hausmann, 2007*). Previous phylogenetic hypotheses were based mainly on the European fauna and many clades were ambiguously supported due to low taxon sampling. The general patterns of the phylogenetic relationships among the subfamilies recovered in our study largely agrees with previous hypotheses based on morphological characters and different sets of molecular markers (*Holloway, 1997*; *Abraham et al., 2001*; *Yamamoto & Sota, 2007*; *Sihvonen et al., 2011*). However, the results of our larger dataset differ in many details and shed light on the phylogenetic relationships of several, poorly resolved, small subfamilies.

Sterrhinae are recovered as the sister subfamily to the remaining Geometridae. This result is not in concordance with *Sihvonen et al. (2011)*, *Yamamoto & Sota (2007)* and *Regier et al. (2009)*, who found a sister group relationship between Sterrhinae and Larentiinae which in turn were sister to the rest of Geometridae. *Sihvonen et al. (2011)* showed the Sterrhinae + Larentiinae sister relationship with low support, while *Yamamoto & Sota (2007)* and *Regier et al. (2009)* included only a few samples in their analyses. Our analyses include representatives from almost all known tribes currently included in Sterrhinae and Larentiinae. The higher number of markers, improved methods of analysis, the broader taxon sampling as well as the stability of our results suggests that Sterrhinae are indeed the sister group to the remaining Geometridae. Sterrhinae (after transfer of *Ergavia*, *Ametris* and *Macrotes*, see details below), Larentiinae, Archiearinae,

Table 2 Summary of formally proposed taxonomic changes.

Transfer from Archiearinae to Ennominae

Acalyphes Turner, 1926, to Ennominae: Diptychini
Dirce Prout, 1910, to Ennominae: Diptychini

Transfer from Oenochrominae to Desmobathrinae (Desmobathrini):

Nearcha Guest, 1887
Racasta Walker, 1861
Zanclopteryx Herrich-Schäffer, 1855

Transfer from Oenochrominae to Epidesmiinae:

Abraxaphantes Warren, 1894
Adeixis Warren 1987
Dichromodes Guenée 1858
Ecphyas Turner, 1929
Epidesmia Duncan & Westwood, 1841
Phrixocomes Turner, 1930
Phrataria Walker, 1863
Systatica Turner, 1904

New tribe combinations in Ennominae

Psilocladia Warren, 1898, from unassigned to Gonodontini
Oedicentra Warren, 1902, from Boarmiini to Gnophini
Hypotephrina Janse, 1932, from unassigned to Gnophini
Capusa Walker, 1857, from Nacophorini to Diptychini
Mictodoca Meyrick, 1892, from Nacophorini to Diptychini
Furcatrox McQuillan, 1996, from Nacophorini to Diptychini
Amelora Guest, 1897, from Nacophorini to Diptychini
Archephanes Turner, 1926, from Nacophorini to Diptychini
Thalaina Walker, 1855, from Nacophorini to Diptychini
Niceteria Turner, 1929, from Nacophorini to Diptychini
Neazata Warren, 1906 from Caberini to Diptychini
Idiodes Guenée, 1858 from unassigned to Diptychini
Panhyperochia Krüger, 2013, from Nacophorini to Diptychini
Mauna Walker, 1865, from Nacophorini to Diptychini
Pareclipsis Warren, 1894, from unassigned to Diptychini
Crambometra Prout, 1915, from unassigned to Diptychini
Hebdomophruda Warren, 1897, from Nacophorini to Diptychini
Pareclipsis Warren, 1894, from unassigned to Diptychini
Capasa Walker 1866, from unassigned to Hypochrosini
Omizodes Warren, 1894, from unassigned to Hypochrosini
Metallospora Warren, 1905, from unassigned to Cassymini
Obolcola Walker, 1862, from unassigned to Abraxini
Chelotephrina Fletcher, 1958 from unassigned to Abraxini
Cassephyra Holloway, 1994 from Cassymini to Abraxini
Thenopa Walker, 1855 from unassigned to Drepanogynini
Drepanogynis Guenée, 1858 from Nacophorini to Drepanogynini

Table 2 (continued).

New tribe combinations in Geometrinae	
Agathiopsis Warren 1896, from unassigned to Hemitheini	
Albinospila Holloway, 1996, from Geometrini to Hemitheini	
Antharmostes Warren 1899, from unassigned to Hemitheini	
Ctenoberta Prout 1915, from unassigned to Hemitheini	
Comostolopsis Warren 1902, from unassigned to Hemitheini	
Oenospila Swinhoe 1892, from Geometrini to Hemitheini	
New and upgraded tribes in Geometrinae	**Included taxa**
Archaeobalbini, **stat. rev.**	Type genus: *Herochroma* Swinhoe, 1893 (syn. *Archaeobalbis* Prout, 1912). Other included genera: *Pachyodes* Guenée, 1858; *Metallolophia* Warren, 1895; *Actenochroma* Warren, 1893; *Absala* Swinhoe 1893; *Metaterpna* Yazaki, 1992; *Limbatochlamys* Rothschild, 1894; *Psilotagma* Warren, 1894; *Dindica* Warren, 1893; *Dindicodes* Prout, 1912; *Lophophelma* Prout, 1912.
Chlorodontoperini, Murillo-Ramos, Sihvonen & Brehm, **trib. nov.**	Type genus: *Chlorodontopera* Warren, 1893. Species included: *C. discospilata* (Moore, 1867); *C. mandarinata* (Leech, 1889); *C. chalybeata* (Moore, 1872); *C. taiwana* (Wileman, 1911).
New tribe in Ennominae	**Included taxa**
Drepanogynini, Murillo-Ramos, Sihvonen & Brehm, **trib. nov.**	Type genus: *Drepanogynis* Guenée, 1858. Other included genera: *Thenopa* Walker, 1855. Species included, genus combination uncertain (*incertae sedis*): *"Sphingomima" discolucida* Herbulot, 1995 (transferred from unassigned to Drepanogynini); *"Hebdomophruda" errans* Prout, 1917 (transferred from Nacophorini to Drepanogynini).
Synonymized tribes	**Valid tribe**
Lithinini Forbes, 1948, **syn. nov.**	Diptychini Janse, 1933 (Ennominae)
Synchlorini Ferguson, 1969 **syn. nov.**	Nemoriini Gumppenberg, 1887 (Geometrinae)
Incertae sedis	
Eumelea Duncan & Westwood, 1841	

Geometrinae and Ennominae were highly supported as monophyletic. Oenochrominae and Desmobathrinae formed polyphyletic and paraphyletic assemblages, respectively. The monophylies of Oenochrominae and Desmobathrinae have long been questioned. Morphological studies addressing Oenochrominae or Desmobathrinae have been limited and the majority of genera have never been examined in depth. In addition, it has been very difficult to establish the boundaries of these subfamilies on the basis of morphological structures (*Scoble & Edwards, 1990*). *Sihvonen et al. (2011)* showed that neither Oenochrominae nor Desmobathrinae were monophyletic, but these results were considered preliminary due to the limited number of sampled taxa, and as a consequence no formal transfers of taxa were proposed.

The systematic status of Orthostixinae remains uncertain because it was not included in our study. *Sihvonen et al. (2011)* included the genus *Naxa* Walker, 1856, formally placed in Orthostixinae, and found it to be nested within Ennominae. However, only three genes were successfully sequenced from this taxon, and its position in the phylogenetic tree turned out to be highly unstable in our analyses. It was thus excluded from our dataset.

Orthostixis Hübner, 1823, the type genus of the subfamily, needs to be included in future analyses.

Sterrhinae Meyrick, 1892

We included 74 Sterrhinae taxa in our analyses, with all tribes recognized in *Forum Herbulot (2007)* being represented. The recovered patterns generally agree with previous phylogenetic hypotheses of the subfamily (*Sihvonen & Kaila, 2004*, *Sihvonen et al., 2011*). The genera *Ergavia* Walker, 1866, *Ametris* Guenée, 1858 and *Macrotes* Westwood, 1841, which currently are placed in Oenochrominae were found to form a well-defined lineage within Sterrhinae with strong support (SH-Like = 99 UFBoot2 = 100). These genera are distributed in the New World, whereas the range of true Oenochrominae is restricted to the Australian and Oriental Regions. *Sihvonen et al. (2011)* already found that *Ergavia* and *Afrophyla* Warren, 1895 belong to Sterrhinae and suggested more extensive analyses to clarify the position of these genera, which we did. *Afrophyla* was transferred to Sterrhinae by *Sihvonen & Staude (2011)* and *Ergavia*, *Ametris* and *Macrotes* (plus *Almodes* Guenée, (1858)) will be transferred by P. Sihvonen et al. (2019, unpublished data).

Cosymbiini, Timandrini, Rhodometrini and Lythriini are closely related as shown previously (*Sihvonen & Kaila, 2004*; *Õunap, Viidalepp & Saarma, 2008*; *Sihvonen et al., 2011*). Cosymbiini appear as sister to the Timandrini + *Traminda* Saalmüller, 1891 + *Pseudosterrha* Warren, 1888 and Rhodometrini + Lythriini clade. Lythriini are closely related to Rhodometrini as shown by *Õunap, Viidalepp & Saarma (2008)* with both molecular and morphological data. *Traminda* (Timandrini) and *Pseudosterrha* (Cosymbiini) grouped together forming a lineage that is sister to the Rhodometrini + Lythriini clade (Fig. 2).

Rhodostrophiini and Cyllopodini were recovered as polyphyletic with species of Cyllopodini clustering within Rhodostrophiini. Similar results were recovered previously (*Sihvonen & Kaila, 2004*; *Sihvonen et al., 2011*), suggesting that additional work is needed to be done to clarify the status and systematic positions of these tribes. Sterrhini and Scopulini were recovered as sister taxa as proposed by *Sihvonen & Kaila (2004)*, *Hausmann (2004)*, *Õunap, Viidalepp & Saarma (2008)* and *Sihvonen et al. (2011)*. Our new phylogenetic hypothesis constitutes a large step towards understanding the evolutionary relationships of the major lineages of Sterrhinae. Further taxonomic changes and more detailed interpretation of the clades will be dealt with by P. Sihvonen et al. (2019, unpublished data).

Larentiinae Duponchel, 1845

Larentiinae are a monophyletic entity (Fig. 3). In concordance with the results of *Sihvonen et al. (2011)*, *Viidalepp (2011)*, *Õunap, Viidalepp & Truuverk (2016)* and *Strutzenberger et al. (2017)*, Dyspteridini are supported as sister to all other larentiines. Remarkably, *Brabirodes* Warren, 1904 forms an independent lineage. Chesiadini are monophyletic and sister to all larentiines except Dyspteridini, *Brabirodes* and Trichopterygini. These results do not support the suggestion by *Viidalepp (2006)* and *Sihvonen et al. (2011)* that Chesiadini are sister to Trichopterygini.

In our phylogenetic hypothesis, Asthenini are sister to the Perizomini + Melanthiini + Eupitheciini clade. These results do not fully agree with *Õunap, Viidalepp & Truuverk (2016)* who found Asthenini to be sister to all Larentiinae except Dyspteridini, Chesiadini, Trichopterygini and Eudulini. However, our results do support the Melanthiini + Eupitheciini complex as a sister lineage to Perizomini. *Sihvonen et al. (2011)* recovered Phileremini and Rheumapterini as well-supported sister taxa. Our results suggest *Triphosa dubitata* Linnaeus 1758 (Triphosini) is sister to Phileremini, with Rheumapterini sister to this clade. Cidariini were recovered as paraphyletic, as the genera *Coenotephria* Prout, 1914 and *Lampropteryx* Stephens, 1831 cluster in a different clade (unnamed clade L7) apart from the lineage comprising the type genus of the tribe, *Cidaria* Treitschke, 1825. *Ceratodalia* Packard, 1876, currently placed in Hydriomenini and *Trichodezia* Warren, 1895 are nested within Cidariini. These results are not in concordance with *Õunap, Viidalepp & Truuverk (2016)*, who regarded this tribe to be monophyletic. Scotopterygini are sister to a lineage comprising *Ptychorrhoe blosyrata* Guenée (1858), *Disclisioprocta natalata* (Walker, 1862) (placed in the unnamed clade L8), Euphyiini, an unnamed clade L9 comprising the genera *Pterocypha*, *Archirhoe* and *Obila*, Xanthorhoini and Cataclysmini. Euphyiini are monophyletic, but Xanthorhoini are recovered as mixed with Cataclysmini. The same findings were shown by *Õunap, Viidalepp & Truuverk (2016)*, but no taxonomic rearrangements were proposed. Larentiini are monophyletic and sister of Hydriomenini, Heterusiini, Erateinini, Stamnodini and some unnamed clades (L11–14). Although with some differences, our results support the major phylogenetic patterns of *Õunap, Viidalepp & Truuverk (2016)*.

Despite substantial progress, the tribal classification and phylogenetic relationships of Larentiinae are far from being resolved (*Õunap, Viidalepp & Truuverk, 2016*). *Forbes (1948)* proposed eight tribes based on morphological information, *Viidalepp (2011)* raised the number to 23 and *Õunap, Viidalepp & Truuverk (2016)* recovered 25 tribes studying 58 genera. Our study includes 23 of the currently recognized tribes and 125 genera (with an emphasis on Neotropical taxa). However, the phylogenetic position of many taxa remains unclear, and some tropical genera have not yet been formally assigned to any tribe. Formal descriptions of these groups will be treated in detail by G. Brehm et al. (2019, unpublished data) and E. Õunap et al. (2019, unpublished data).

Archiearinae Fletcher, 1953

The hypothesis presented in this study recovered Archiearinae as a monophyletic entity after some taxonomic rearrangements are performed. This subfamily was previously considered as sister to Geometrinae + Ennominae (*Abraham et al., 2001*), whereas *Yamamoto & Sota (2007)* proposed them to be the sister-taxon to Orthostixinae + Desmobathrinae. Our findings agree with *Sihvonen et al. (2011)* who recovered Archiearinae as the sister-taxon to the rest of Geometridae excluding Sterrhinae and Larentiinae, although only one species was included in their study. *Archiearis* Hübner, (1823) is sister to *Boudinotiana* Esper, 1787 and these taxa in turn are sister to *Leucobrephos* Grote, 1874 (Fig. 4). The southern hemisphere Archiearinae require more attention. *Young (2006)* suggested that two Australian Archiearinae genera, *Dirce* and

Acalyphes, actually belong to Ennominae. Our analyses clearly support this view and we therefore propose to formally transfer *Dirce* and *Acalyphes* to Ennominae (all formal taxonomic changes are provided in Table 2). Unfortunately, the South American Archiearinae genera *Archiearides* Fletcher, 1953 and *Lachnocephala* Fletcher, 1953, and Mexican *Caenosynteles* Dyar, 1912 (*Pitkin & Jenkins, 2004*), could not be included in our analyses. These presumably diurnal taxa may only be superficially similar to northern hemisphere Archiearinae as was the case with Australian *Dirce* and *Acalyphes*.

Desmobathrinae Meyrick, 1886

Taxa placed in Desmobathrinae were formerly recognized as Oenochrominae genera with slender appendages. *Holloway (1996)* revived Desmobathrinae from synonymy with Oenochrominae and divided it into the tribes Eumeleini and Desmobathrini. Desmobathrinae species have a pantropical distribution and they apparently (still) lack recognized morphological apomorphies (*Holloway, 1996*). Our phylogenetic analysis has questioned the monophyly of Desmobathrinae *sensu* Holloway because some species currently placed in Oenochrominae were embedded within the group (see also *Sihvonen et al., 2011*), and also the phylogenetic position of the tribe Eumeleini is unstable (see below). Desmobathrinae can be regarded as a monophyletic group after the transfer of *Zanclopteryx, Nearcha* and *Racasta* from Oenochrominae to Desmobathrinae, and the removal of Eumeleini (Table 2). Desmobathrinae as circumscribed here are an independent lineage that is sister to all Geometridae except Sterrhinae, Larentiinae and Archiearinae.

The monobasic Eumeleini has had a dynamic taxonomic history: *Eumelea* was transferred from Oenochrominae *s.l.* to Desmobathrinae based on the pupal cremaster (*Holloway, 1996*), whereas *Beljaev (2008)* pointed out that *Eumelea* could be a member of Geometrinae based on the skeleto-muscular structure of the male genitalia. Molecular studies (*Sihvonen et al., 2011*, *Ban et al., 2018*) suggested that *Eumelea* was part of Oenochrominae *s.str*., but these findings were not well-supported and no formal taxonomic changes were proposed. Our analyses with IQTREE and RAxML recovered Eumeleini in two very different positions, either as sister to Geometrinae (SH-like = 93.6, UFBoot2 = 71) (Figs. 4 and 5), or as sister of *Plutodes* in Ennominae (RBS = 60) (Data S3). The examination of morphological details suggests that the position as sister to Geometrinae is more plausible: hindwing vein M2 is present and tubular; anal margin of the hindwing is elongated; and large coremata originate from the saccus (*Holloway, 1994*, our observations). The morphology of *Eumelea* is partly unusual, and for that reason we illustrate selected structures (Data S4), which include for instance the following: antennae and legs of both sexes are very long; forewing vein Sc (homology unclear) reaches wing margin; in male genitalia coremata are extremely large and branched; uncus is cross-shaped (cruciform); tegumen is narrow and it extends ventrally beyond the point of articulation with vinculum; saccus arms are extremely long, looped; and vesica is with lateral rows of cornuti. However, the green geoverdin pigment concentration of *Eumelea* is low in comparison to Geometrinae (*Cook et al., 1994*). We tentatively conclude that *Eumelea* is probably indeed associated with Geometrinae. However, since eleven genetic

markers were not sufficient to clarify the phylogenetic affinities of *Eumelea*, we provisionally place the genus as *incertae sedis* (Table 2).

Oenochrominae Guenée, 1858

Oenochrominae has obviously been the group comprising taxa that could not easily be assigned to other subfamilies. Out of the 76 genera currently assigned to Oenochrominae, our study includes 25 genera (28 species). Three of these genera will be formally transferred to Sterrhinae (P. Sihvonen et al., 2019, unpublished data), three are here transferred to Desmobathrinae (see above, Table 2), and eight are transferred to Epidesmiinae (see below). In agreement with *Sihvonen et al. (2011)*, Oenochrominae *s.str.* grouped together in a well-supported lineage. Genera of this clade can be characterized as having robust bodies, and their male genitalia have a well-developed uncus and gnathos, broad valvae and a well-developed anellus (*Scoble & Edwards, 1990*). Common host plants are members of Proteaceae and Myrtaceae (*Holloway, 1996*). Our results strongly suggest that the genus *Oenochroma* is polyphyletic: *Oenochroma infantilis* is sister to a clade including *Dinophalus*, *Hypographa*, *Lissomma*, *Sarcinodes* and (at least) two species of *Oenochroma*. To date, 20 species have been assigned to *Oenochroma* by *Scoble (1999)*, and one additional species was described by *Hausmann et al. (2009)*, who suggested that *Oenochroma vinaria* is a species complex. We agree with *Hausmann et al. (2009)*, who pointed out the need for a major revision of *Oenochroma*.

In our phylogenetic hypothesis, *Sarcinodes* is sister to *O. orthodesma* and *O. vinaria*, the type species of *Oenochroma*. Although *Sarcinodes* and *Oenochroma* resemble each other in external morphology, a sister-group relationship between these genera has not been hypothesized before. The inclusion of *Sarcinodes* in Oenochrominae is mainly based on shared tympanal characters (*Scoble & Edwards, 1990*). However, the circular form of the lacinia, which is an apomorphy of Oenochrominae *s.str.* is missing or not apparent in *Sarcinodes* (*Holloway, 1996*). In addition, *Sarcinodes* is found in the Oriental rather than in the Australian region, where all *Oenochroma* species are distributed. A second clade of Oenochrominae *s.str.* comprises the genera *Monoctenia*, *Onycodes*, *Parepisparis*, *Antictenia*, *Arhodia*, *Gastrophora* and *Homospora*, which clustered together as the sister of *Oenochroma* and its relatives. These genera are widely recognized in sharing similar structure of the male genitalia (*Scoble & Edwards, 1990*), yet their phylogenetic relationships have never been tested. *Young (2006)* suggested the monophyly of Oenochrominae *s.str.*, however, with a poorly resolved topology and low branch support. In her study, *Parepisparis*, *Phallaria* and *Monoctenia* shared a bifid head, while in *Parepisparis* and *Onychodes*, the aedeagus was lacking caecum and cornuti. Our analysis supports these morphological similarities. *Monoctenia*, *Onycodes* and *Parepisparis* clustered together. However, a close relationship of the genera *Antictenia*, *Arhodia*, *Gastrophora* and *Homospora* has not been suggested before. Our analysis thus strongly supports the earliest definition of Oenochrominae proposed by *Guenée (1858)*, and reinforced by *Cook & Scoble (1992)*. Oenochrominae should be restricted to *Oenochroma* and related genera such as *Dinophalus*, *Hypographa*, *Lissomma*, *Sarcinodes*, *Monoctenia*, *Onycodes*, *Parepisparis*, *Antictenia*, *Arhodia*, *Gastrophora*, *Homospora*, *Phallaria* and

Palaeodoxa. We consider that genera included in Oenochrominae by *Scoble & Edwards (1990)*, but recovered in a lineage separate from *Oenochroma* and its close relatives in our study, belong to a hitherto unknown subfamily, which is described below.

Epidesmiinae Murillo-Ramos, Brehm & Sihvonen **new subfamily**

LSIDurn:lsid:zoobank.org:act:34D1E8F7-99F1-4914-8E12-0110459C2040

Type genus: *Epidesmia* Duncan & Westwood, 1841.

Material examined: Taxa included in the molecular phylogeny: *Ecphyas holopsara* Turner, 1929, *Systatica xanthastis* Lower, 1894, *Adeixis griseata* Hudson, 1903, *Dichromodes indicataria* Walker, 1866, *Phrixocomes* sp. Turner, 1930, *Abraxaphantes perampla* Swinhoe, 1890, *Epidesmia chilonaria* (Herrich-Schäffer, 1855), *Phrataria replicataria* Walker, 1866.

Most of the slender-bodied Oenochrominae, excluded from Oenochrominae *s.str*. by *Holloway (1996)*, were recovered as an independent lineage (Fig. 4) that consists of two clades: *Ec. holopsara* + *S. xanthastis* and *Ep. chilonaria* + five other genera. Branch support values from IQ-TREE strongly support the monophyly of this clade (SH-like and UFBoot2 = 100), while in RAxML the clade is moderately supported (RBS = 89). These genera have earlier been assigned to Oenochrominae *s.l.* (*Scoble & Edwards, 1990*). However, we recovered the group as a well-supported lineage independent from Oenochrominae *s.str*. and transfer them to Epidesmiinae, subfam. n. (Table 2).

Phylogenetic position: Epidesmiinae is sister to Oenochrominae *s.str*. + *Eumelea* + Geometrinae + Ennominae.

Short description of Epidesmiinae: Antennae in males unipectinate (exception: *Adeixis*), shorter towards the apex. Pectination moderate or long. Thorax and abdomen slender (unlike in Oenochrominae). Forewings with sinuous postmedial line and areole present. Forewings planiform (with wings lying flat on the substrate) in resting position, held like a triangle and cover the hindwings.

Diagnosis of Epidesmiinae: The genera included in this subfamily form a strongly supported clade with DNA sequence data from the following gene regions (exemplar *Epidesmia chilonaria* (Herrich-Schäffer, 1855)) ArgK (MK738299), Ca-ATPase (MK738690), CAD (MK738960), COI (MK739187), EF1a (MK740168), GAPDH (MK740402), MDH (MK740974) and Nex9 (MK741433). A thorough morphological investigation of the subfamily, including diagnostic characters, is under preparation.

Distribution: Most genera are distributed in the Australian region, with some species ranging into the Oriental region. *Abraxaphantes* occurs exclusively in the Oriental region.

Geometrinae Stephens, 1829

The monophyly of Geometrinae is strongly supported, but the number of tribes included in this subfamily is still unclear. *Sihvonen et al. (2011)* analyzed 27 species assigned to 11 tribes, followed by *Ban et al. (2018)* with 116 species in 12 tribes. *Ban et al. (2018)* synonymized nine tribes, and validated the monophyly of 12 tribes, with two new tribes Ornithospilini and Agathiini being the first two clades branching off the main lineage of

Geometrinae. Our study (168 species) validates the monophyly of 13 tribes, eleven of which were defined in previous studies: Hemitheini, Dysphaniini, Pseudoterpnini *s.str.*, Ornithospilini, Agathiini, Aracimini, Neohipparchini, Timandromorphini, Geometrini, Comibaeini, Nemoriini. One synonymization is proposed: Synchlorini Ferguson, 1969 **syn. nov.** is synonymized with Nemoriini Gumppenberg, 1887. One tribe is proposed as new: Chlorodontoperini **trib. nov.**, and one tribe (Archaeobalbini Viidalepp, 1981, **stat. rev.**) is raised from synonymy with Pseudoterpnini.

Ban et al. (2018) found that *Ornithospila* Warren, 1894 is sister to the rest of Geometrinae, and *Agathia* Guenée, 1858 is sister to the rest of Geometrinae minus *Ornithospila*. Although weakly supported, our results (with more species of *Agathia* sampled) placed Ornisthospilini+Agathiini together and these tribes are the sister to the rest of Geometrinae. *Chlorodontopera* is placed as an isolated lineage as shown by *Ban et al. (2018)*. Given that *Chlorodontopera* clearly forms an independent and well-supported lineage we propose the description of a new tribe Chlorodontoperini.

Chlorodontoperini Murillo-Ramos, Sihvonen & Brehm, **new tribe**

LSIDurn:lsid:zoobank.org:act:0833860E-A092-43D6-B2A1-FB57D9F7988D

Type genus: *Chlorodontopera* Warren, 1893

Material examined: Taxa in the molecular phylogeny: *Chlorodontopera discospilata* (Moore, 1867) and *Chlorodontopera mandarinata* (Leech, 1889).

Some studies (*Inoue, 1961*; *Holloway, 1996*) suggested the morphological similarities of *Chlorodontopera* Warren, 1893 with members of Aracimini. Moreover, *Holloway (1996)* considered this genus as part of Aracimini. Our results suggest a sister relationship of *Chlorodontopera* with a large clade comprising Aracimini, Neohipparchini, Timandromorphini, Geometrini, Nemoriini and Comibaenini. Considering that our analysis strongly supports *Chlorodontopera* as an independent lineage (branch support SH-like = 99 UFBoot2 = 100, RBS = 99), we introduce the monobasic tribe Chlorodontoperini. This tribe can be diagnosed by the combination of DNA data from six genetic markers (exemplar *Chlorodontopera discospilata*) CAD (MG015448), COI (MG014735), EF1a (MG015329), GAPDH (MG014862), MDH (MG014980) and RpS5 (MG015562). *Ban et al. (2018)* did not introduce a new tribe because the relationship between *Chlorodontopera* and *Euxena* Warren, 1896 was not clear in their study. This relationship was also been proposed by *Holloway (1996)* based on similar wing patterns. Further analyses are needed to clarify the affinities between *Chlorodontopera* and *Euxena*.

The tribe Chlorodontoperini is diagnosed by distinct discal spots with pale margins on the wings, which are larger on the hindwing; a dull reddish-brown patch is present between the discal spot and the costa on the hindwing, and veins M3 and CuA1 are not stalked on the hindwing (*Ban et al., 2018*). In the male genitalia, the socii are stout and setose and the lateral arms of the gnathos are developed, not joined. Sternite 3 of the male has setal patches (see *Holloway, 1996* for illustrations). Formal taxonomic changes are listed in Table 2.

Aracimini, Neohipparchini, Timandromorphini, Geometrini and Comibaenini were recovered as monophyletic groups. These results are in full agreement with

Ban et al. (2018). However, the phylogenetic position of *Eucyclodes* Warren, 1894 is uncertain (unnamed G2). The monophyly of Nemoriini and Synchlorini is not supported. Instead, Synchlorini are nested within Nemoriini (support branch SH-like = 98.3, UFBoot2 = 91, RBS = 93). Our findings are in concordance with Sihvonen et al. (2011) and Ban et al. (2018), but our analyses included a larger number of markers and a much higher number of taxa. Thus, we formally synonymize Synchlorini **syn. nov.** with Nemoriini (Table 2).

The monophyly of Pseudoterpnini *sensu* Pitkin, Han & James (2007) could not be recovered. Similar results were shown by Ban et al. (2018) who recovered Pseudoterpnini *s.l.* including all the genera previously studied by Pitkin, Han & James (2007), forming a separate clade from *Pseudoterpna* Hübner, 1823 + *Pingasa* Moore, 1887. Our results showed African *Mictoschema* Prout, 1922 falling within Pseudoterpnini *s.str*., and it is sister to *Pseudoterpna* and *Pingasa*. A second group of Pseudoterpnini *s.l.* was recovered as an independent lineage clearly separate from Pseudoterpnini *s.str.* (SH-like = 88.3, UFBoot2 = 64). Ban et al. (2018) did not introduce a new tribe due to the morphological similarities and difficulty in finding apomorphies of Pseudoterpnini *s.str*. In addition, their results were weakly supported. Considering that two independent studies have demonstrated the paraphyly of Pseudoterpnini *sensu* Pitkin et al. (2007), we see no reason for retaining the wide concept of this tribe. Instead, we propose the revival of the tribe status of Archaeobalbini.

Archaeobalbini Viidalepp, 1981, **status revised**

(original spelling: Archeobalbini, justified emendation in Hausmann (1996))

Type genus: *Archaeobalbis* Prout, 1912 (synonymized with *Herochroma* Swinhoe, 1893 in Holloway (1996))

Material examined: *Herochroma curvata* Han & Xue, 2003, *H. baba* Swinhoe 1893, *Metallolophia inanularia* Han & Xue, 2004, *M. cuneataria* Han & Xue, 2004, *Actenochroma muscicoloraria* (Walker, 1862), *Absala dorcada* Swinhoe, 1893, *Metaterpna batangensis* Hang & Stüning, 2016, *M. thyatiraria* (Oberthür, 1913), *Limbatochlamys rosthorni* Rothschild, 1894, *Psilotagma pictaria* (Moore, 1888), *Dindica para* Swinhoe, 1893, *Dindicodes crocina* (Butler, 1880), *Lophophelma erionoma* (Swinhoe, 1893), *L. varicoloraria* (Moore, 1868), *L. iterans* (Prout, 1926) and *Pachyodes amplificata* (Walker, 1862).

This lineage splits into four groups: *Herochroma* Swinhoe, 1893 + *Absala* Swinhoe, 1893 + *Actenochroma* Warren, 1893 is the sister lineage of the rest of Archaeobalbini that were recovered as three clades with unresolved relationships comprising the genera *Limbatochlamys* Rothschild, 1894, *Psilotagma* Warren, 1894, *Metallolophia* Warren, 1895, *Metaterpna* Yazaki, 1992, *Dindica* Warren, 1893, *Dindicodes* Prout, 1912, *Lophophelma* Prout, 1912 and *Pachyodes* Guenée, 1858. This tribe can be diagnosed by the combination of DNA data from six genetic markers, see for instance *Pachyodes amplificata* CAD (MG015522), COI (MG014818), EF1a (MG015409), GAPDH (MG014941), MDH (MG015057) and RpS5 (MG015638). Branch support values in IQ-TREE confirm the

monophyly of this clade (SH-like = 88.3, UFBoot2 = 64). GenBank accession numbers are shown in Supplementary Material. A morphological diagnosis requires further research.

Xenozancla Warren, 1893 (unnamed G3) is sister to the clade comprising Dysphaniini and Pseudoterpnini *s.str*. *Sihvonen et al. (2011)* did not include *Xenozancla* in their analyses and suggested a sister relationship of Dysphaniini and Pseudoterpnini, but with low support. According to *Ban et al. (2018)*, *Xenozancla* is more closely related to Pseudoterpnini *s.str*. than to Dysphaniini. However, due to low support, *Ban et al. (2018)* did not propose a taxonomic assignment for *Xenozancla*, which is currently not assigned to a tribe. Although our IQ-TREE results show that *Xenozancla* is sister to a clade comprising Dysphaniini and Pseudoterpnini *s.str*., the RAxML analysis did not recover the same phylogenetic relationships. Instead, Dysphaniini + Pseudoterpnini *s.str*. are found to be sister taxa, but *Xenozancla* is placed close to *Rhomborista monosticta* (Wehrli, 1924). As in *Ban et al. (2018)*, our results do not allow us to reach a conclusion about the phylogenetic affinities of these tribes, due to low support of nodes.

The Australian genus *Crypsiphona* Meyrick, 1888 (unnamed G4) was placed close to Hemitheini. *Crypsiphona* has been assigned to Pseudoterpnini (e. g. *Pitkin, Han & James, 2007*, *Õunap & Viidalepp, 2009*), but is recovered as a separate lineage in our tree. Given the isolated position of *Crypsiphona*, the designation of a new tribe could be considered, but due to low support of nodes in our analyses, further information (including morphology) is needed to confirm the phylogenetic position of this genus. In our phylogenetic hypothesis, a large clade including the former tribes Lophochoristini, Heliotheini, Microloxiini, Thalerini, Rhomboristini, Hemistolini, Comostolini, Jodini and Thalassodini is recovered as sister to the rest of Geometrinae. These results are in full agreement with *Ban et al. (2018)*, who synonymized all of these tribes with Hemitheini. Although the monophyly of Hemitheini is strongly supported, our findings recovered only a few monophyletic subtribes. For example, genera placed in Hemitheina were intermixed with those belonging to Microloxiina, Thalassodina and Jodina. Moreover, many genera which were unassigned to tribe, were recovered as belonging to Hemitheini. Our findings recovered *Lophostola* Prout, 1912 as sister to all Hemitheini. These results are quite different from those found by *Ban et al. (2018)* who suggested Rhomboristina as being sister to the rest of Hemitheini. In contrast, our results recovered Rhomboristina mingled with Hemistolina. These different results are probably influenced by the presence of African and Madagascan *Lophostola* in our analysis. In our opinion the subtribe concept, as applied in Hemitheini earlier, is not practical and we do not advocate its use in geometrid classification.

Ennominae Duponchel, 1845

Ennominae are the most species-rich subfamily of geometrids. The loss of vein M2 on the hindwing is probably the best apomorphy (*Holloway, 1994*), although vein M2 is present as tubular in a few ennomine taxa (*Staude, 2001*; *Skou & Sihvonen, 2015*). Ennominae are a morphologically highly diverse subfamily, and attempts to find further synapomorphies shared by all major tribal groups have failed.

The number of tribes as well as phylogenetic relationships among tribes are still debated (see *Skou & Sihvonen, 2015* for an overview). Moreover, the taxonomic knowledge of this

subfamily in tropical regions is still poor. *Holloway (1994)* recognized 21 tribes, *Beljaev (2006)* 24 tribes, and *Forum Herbulot (2007)* 27 tribes. To date, four molecular studies have corroborated the monophyly of Ennominae (*Yamamoto & Sota, 2007*; *Wahlberg et al., 2010*; *Õunap et al., 2011*, *Sihvonen et al., 2011*), with *Young (2006)* being the only exception who found Ennominae paraphyletic. Moreover, four large-scale taxonomic revisions (without a phylogenetic hypothesis) were published by *Pitkin (2002)* for the Neotropical region, *Skou & Sihvonen (2015)*, *Müller et al. (2019)* for the Western Palaearctic region, and *Holloway (1994)* for Borneo. More detailed descriptions of taxonomic changes in Ennominae will be given by G. Brehm et al. (2019, unpublished data) and L. Murillo-Ramos et al. (2019, unpublished data). We here discuss general patterns and give details for taxonomic acts not covered in the other two papers.

Our findings recover Ennominae as a monophyletic entity, but results were not highly supported in RAxML (RBS = 67) compared to IQ-TREE (SH-Like =100, UFBoot2 = 99). The lineage comprising Geometrinae and Oenochrominae is recovered as the sister clade of Ennominae. In previous studies, *Wahlberg et al. (2010)* sampled 49 species of Ennominae, *Õunap et al. (2011)* sampled 33 species, and *Sihvonen et al. (2011)* 70 species including up to eight markers per species. All these studies supported the division of Ennominae into "boarmiine" and "ennomine" moths (*Holloway, 1994*). This grouping was proposed by *Forbes (1948)* and *Holloway (1994)*, who suggested close relationships between the tribes Boarmiini, Macariini, Cassymini and Eutoeini based on the bifid pupal cremaster and the possession of a fovea in the male forewing. The remaining tribes were defined as "ennomines" based on the loss of a setal comb on male sternum A3 and the presence of a strong furca in male genitalia. Both *Wahlberg et al. (2010)* and *Sihvonen et al. (2011)* found these two informal groupings to be reciprocally monophyletic.

In our analyses, 653 species with up to 11 markers were sampled, with an emphasis on Neotropical taxa, which so far had been poorly represented in the molecular phylogenetic analyses. Our results recovered the division into two major subclades (Fig. 6), a core set of ennomines in a well-supported clade, and a poorly supported larger clade that includes the "boarmiines" among four other lineages usually thought of as "ennomines". The traditional "ennomines" are thus not found to be monophyletic in our analyses, questioning the utility of such an informal name. Our phylogenetic hypothesis supports the validation of numerous tribes proposed previously, in addition to several unnamed clades. We validate 23 tribes (*Forum Herbulot, 2007*; *Skou & Sihvonen, 2015*): Gonodontini, Gnophini, Odontoperini, Nacophorini, Ennomini, Campaeini, Alsophilini, Wilemaniini, Prosopolophini, Diptychini, Theriini, Plutodini, Palyadini, Hypochrosini, Apeirini, Epionini, Caberini, Macariini, Cassymini, Abraxini, Eutoeini and Boarmiini. We hereby propose one new tribe: Drepanogynini **trib. nov.** (Table 2). Except for the new tribe, most of the groups recovered in this study are in concordance with previous morphological classifications (*Holloway, 1994*; *Beljaev, 2006*, *2016*; *Forum Herbulot, 2007*; *Skou & Sihvonen, 2015*; *Müller et al., 2019*).

Five known tribes and two further unnamed lineages (E1, E2 in Fig. 6) form the core Ennominae: Gonodontini, Gnophini, Odontoperini, Nacophorini and Ennomini. Several Neotropical clades that conflict with the current tribal classification of Ennominae

will be described as new tribes by G. Brehm et al. (2019, unpublished data). Gonodontini and Gnophini are recovered as sister taxa. Gonodontini was defined by *Forbes (1948)* and studied by *Holloway (1994)*, who showed synapomorphies shared by *Gonodontis* Hübner, (1823), *Xylinophylla* Warren, 1898 and *Xenimpia* Warren, 1895. Our results recovered the genus *Xylinophylla* as sister of *Xenimpia* and *Psilocladia* Warren, 1898. *Psilocladia* is an African genus currently unassigned to tribe (see *Sihvonen, Staude & Mutanen, 2015* for details). Considering the strong support and that the facies and morphology are somewhat similar to other analyzed taxa in Gonodontini, we formally include *Psilocladia* in Gonodontini (Table 2). Gnophini are monophyletic and we formally transfer the African genera *Oedicentra* Warren, 1902 and *Hypotephrina* Janse, 1932, from unassigned to Gnophini (Table 2). The total number of species, and number of included genera in Gnophini are still uncertain (*Skou & Sihvonen, 2015*; *Müller et al., 2019*). Based on morphological examination, *Beljaev (2016)* treated Angeronini as a synonym of Gnophini. The costal projection on male valva bearing a spine or group of spines was considered as a synapomorphy of the group. Using molecular data, *Yamamoto & Sota (2007)* showed a close phylogenetic relationship between *Angerona* Duponchel, 1829 (Angeronini) and *Chariaspilates* Wehrli, 1953 (Gnophini). Similar results were shown by *Sihvonen et al. (2011)* who recovered *Angerona* and *Charissa* Curtis, 1826 as sister taxa, and our results also strongly support treating Angeronini as synonym of Gnophini.

Holloway (1994) suggested close affinities among Nacophorini, Azelinini and Odontoperini on the basis of larval characters. In a morphology-based phylogenetic study, *Skou & Sihvonen (2015)* suggested multiple setae on the proleg on A6 of the larvae as a synapomorphy of the group. Our results also support a close relationship of Nacophorini, Azelinini and Odontoperini. These clades will be treated in more detail by G. Brehm et al. (2019, unpublished data).

Following the ideas of *Pitkin (2002)*, *Beljaev (2008)* synonymized the tribes Ourapterygini and Nephodiini with Ennomini. He considered the divided vinculum in male genitalia and the attachment of muscles *m3* as apomorphies of the Ennomini, but did not provide a phylogenetic analysis. *Sihvonen et al. (2011)* supported Beljaev's assumptions and recovered *Ennomos* Treitschke, 1825 (Ennomini), *Ourapteryx* Leach, 1814 (Ourapterygini) and *Nephodia* Hübner, 1823 (Nephodiini) as belonging to the same clade. Our comprehensive analysis confirms those previous findings and we agree with Ennomini as the valid tribal name for this large clade. This clade will be treated in more detail by G. Brehm et al. (2019, unpublished data).

Campaeini, Alsophilini, Wilemaniini and Prosopolophini grouped together in a well-supported clade (SH-like = 100, UFBoot2 = 99). Previous molecular analyses have shown an association of Colotoini [= Prosopolophini] and Wilemaniini (*Yamamoto & Sota, 2007*; *Sihvonen et al., 2011*), although no synapomorphies are known to support synonymization (*Skou & Sihvonen, 2015*). The Palaearctic genera *Compsoptera* Blanchard, 1845, *Apochima* Agassiz, 1847, *Dasycorsa* Prout, 1915, *Chondrosoma* Anker, 1854 and *Dorsispina* Nupponen & Sihvonen, 2013, are potentially part of the same complex (*Skou & Sihvonen, 2015*, Sihvonen pers. obs.), but they were not included in the current study. Campaeini is a small group including four genera with Oriental, Palaearctic and Nearctic

distribution, apparently closely related to Alsophilini and Prosopolophini, but currently accepted as a tribe (*Forum Herbulot, 2007*; *Skou & Sihvonen, 2015*). Our results support the close phylogenetic affinities among these tribes, but due to the limited number of sampled taxa, we do not propose any formal changes.

The genus *Declana* Walker, 1858 is recovered as an isolated clade sister to Diptychini. This genus is endemic to New Zealand, but to date has not been assigned to tribe. According to our results, *Declana* could well be defined as its own tribe. However, the delimitation of this tribe is beyond the scope of our paper and more genera from Australia and New Zealand should first be examined. A close relationship between Nacophorini and Lithinini was suggested by *Pitkin (2002)*, based on the similar pair of processes of the anellus in the male genitalia. Pitkin also noted a morphological similarity in the male genitalia (processes of the juxta) shared by Nacophorini and Diptychini. In a study of the Australasian fauna, *Young (2008)* suggested the synonymization of Nacophorini and Lithinini. This was further corroborated by *Sihvonen, Staude & Mutanen (2015)* who found that Diptychini were nested within some Nacophorini and Lithinini. However, none of the studies proposed formal taxonomic changes because of limited taxon sampling. In contrast, samples in our analyses cover all biogeographic regions and the results suggest that true Nacophorini is a clade which comprises almost exclusively New World species. This clade is clearly separate from Old World "nacophorines" (cf. *Young, 2003*) that are intermixed with Lithinini and Diptychini. We here formally transfer Old World nacophorines to Diptychini and synonymize Lithinini **syn. nov.** with Diptychini (Table 2). Further formal taxonomic changes in the Nacophorini complex are provided by G. Brehm et al. (2019, unpublished data).

Theria Hübner 1825, the only representative of Theriini in this study, clustered together with *Lomographa* Hübner, 1825 (Baptini in *Skou & Sihvonen, 2015*), in a well-supported clade, agreeing with the molecular results of *Sihvonen et al. (2011)*. The placement of *Lomographa* in Caberini (*Rindge, 1979*; *Pitkin, 2002*) is not supported by our study nor by that of *Sihvonen et al. (2011)*. The monophyly of *Lomographa* has not been tested before, but we show that one Neotropical and one Palaearctic *Lomographa* species indeed group together. Our results show that Caberini are not closely related to the Theriini + Baptini clade, unlike in earlier morphology-based hypotheses (*Rindge, 1979*; *Pitkin, 2002*). Morphologically, Theriini and Baptini are dissimilar, therefore we recognize them as valid tribes (see description and illustrations in *Skou & Sihvonen, 2015*).

According to our results, 11 molecular markers were not enough to infer phylogenetic affinities of Plutodini (represented by one species of *Plutodes*). Similar results were found by *Sihvonen et al. (2011)*, who in some analyses recovered *Plutodes* as sister of *Eumelea*. Our analyses are congruent with those findings. IQ-TREE results suggest that *Plutodes* is sister to Palyadini, but RAxML analyses recovered *Eumelea* as the most probable sister of *Plutodes*. Given that our analyses are not in agreement on the sister-group affinities of *Plutodes*, we do not make any assumptions about its phylogenetic position. Instead, we emphasize that further work needs to be done to clarify the phylogenetic positions of *Plutodes* and related groups.

Hypochrosini is only recovered in a well-defined lineage if the genera *Apeira* Gistl, 1848 (Apeirini), *Epione* Duponchel, 1829 (Epionini), *Sericosema* (Caberini), *Ithysia* (Theriini),

Capasa Walker, 1866 (unassigned) and *Omizodes* Warren, 1894 (unassigned) were transferred to Hypochrosini. *Skou & Sihvonen (2015)* already suggested a close association of Epionini, Apeirini and Hypochrosini. We think that synonymizing these tribes is desirable. However, due to the limited number of sampled taxa we do not propose any formal changes until more data becomes available. We do suggest, however, formal taxonomic changes for the genera *Capasa* and *Omizodes* from unassigned to Hypochrosini (Table 2).

The southern African genus *Drepanogynis* is paraphyletic and has earlier been classified as belonging in Ennomini, and later in Nacophorini (*Krüger, 2002*). In our phylogeny, it is intermixed with the genera *Sphingomima* Warren, 1899, and *Thenopa* Walker, 1855. *Hebdomophruda errans* Prout, 1917 also clusters together with these taxa, apart from other *Hebdomophruda* Warren, 1897 species, which suggests that this genus is polyphyletic. These genera form a clade sister to the lineage that comprises several Hypochrosini species. Considering that our analysis strongly supports this clade, we place *Thenopa*, *Sphingomima* and *Drepanogynis* in a tribe of their own.

Drepanogynini Murillo-Ramos, Sihvonen & Brehm **new tribe**

LSIDurn:lsid:zoobank.org:act:AA384988-009F-4175-B98C-6209C8868B93

Type genus: *Drepanogynis* Guenée, (1858)

The African genera *Thenopa*, *Sphingomima* and *Drepanogynis* appear as a strongly supported lineage (SH-like, UFBoot2 and RBS = 100). *Krüger (1997*, p. 259) proposed "Boarmiini and related tribes as the most likely sister group" for *Drepanogynis*, whereas more recently *Drepanogynis* was classified in the putative southern hemisphere Nacophorini (*Krüger, 2014*; *Sihvonen, Staude & Mutanen, 2015*). In the current phylogeny, *Drepanogynis* is isolated from Nacophorini *sensu stricto* and from other southern African genera that have earlier been considered to be closely related to it (*Krüger, 2014* and references therein). The other southern African genera appeared to belong to Diptychini in our study. The systematic position of *Drepanogynis tripartita* (Warren, 1898) has earlier been analyzed in a molecular study (*Sihvonen, Staude & Mutanen, 2015*). The taxon grouped together with the Palaearctic species of the tribes Apeirini, Theriini, Epionini and putative Hypochrosini. *Sihvonen, Staude & Mutanen (2015)* noted that *Argyrophora trofonia* (Cramer, 1779) (representing *Drepanogynis* group III *sensu Krüger, 1999*) and *Drepanogynis tripartita* (representing *Drepanogynis* group IV *sensu Krüger, 2002*) did not group together, but no formal changes were proposed. Considering that the current analysis strongly supports the placement of *Drepanogynis* and related genera in an independent lineage, and the aforementioned taxa in the sister lineage (Apeirini, Theriini, Epionini and putative Hypochrosini) have been validated at tribe-level, we place *Drepanogynis* and related genera in a tribe of their own.

Material examined and taxa included: *Drepanogynis mixtaria* (Guenée, 1858), *D. tripartita*, *D. determinata* (Walker, 1860), *D. arcuifera* Prout, 1934, *D. arcuatilinea Krüger, 2002*, *D. cnephaeogramma* (Prout, 1938), *D. villaria* (Felder & Rogenhofer, 1875), "*Sphingomima*" *discolucida* Herbulot, 1995 (genus combination uncertain, see taxonomic

notes below), *Thenopa diversa* Walker, 1855, *"Hebdomophruda" errans* Prout, 1917 (genus combination uncertain, see taxonomic notes below).

Taxonomic notes: We choose *Drepanogynis* Guenée, 1858 as the type genus for Drepanogynini, although it is not the oldest valid name (ICZN Article 64), because extensive literature has been published on *Drepanogynis* (Krüger, 1997, 1998, 1999, 2014), but virtually nothing exists on *Thenopa*, Walker, 1855, except the original descriptions of its constituent species. Current results show the urgent need for more extensive phylogenetic studies within Drepanogynini. *Thenopa* and *Sphingomima* are embedded within *Drepanogynis*, rendering it paraphyletic, but our taxon coverage is too limited to propose formal changes in this species-rich group. Drepanogynini, as defined here, are distributed in sub-Saharan Africa. *Drepanogynis sensu* Krüger (1997, 1998, 1999, 2014) includes over 150 species and it ranges from southern Africa to Ethiopia (Krüger, 2002, Vári, Kroon & Krüger, 2002), whereas the genera *Sphingomima* (10 species) and *Thenopa* (four species) occur in Central and West Africa (Scoble, 1999). *Sphingomima* and *Thenopa* are externally similar, so the recovered sister-group relationship in the current phylogeny analysis was anticipated. In the current analysis, *Hebdomophruda errans* Prout, 1917 is isolated from other analyzed *Hebdomophruda* species (the others are included in Diptychini), highlighting the need for additional research. Krüger (1997, 1998) classified the genus *Hebdomophruda* into seven species groups on the basis of morphological characters, and *H. errans* group is one of them (Krüger, 1998). We do not describe a new genus for the taxon *errans*, nor do we combine it with any genus in the Drepanogynini, highlighting its uncertain taxonomic position (*incertae sedis*) pending more research. In the current analysis, *Sphingomima discolucida* Herbulot, 1995 is transferred from unassigned tribus combination to Drepanogynini, but as the type species of *Sphingomima* (*S. heterodoxa* Warren, 1899) was not analyzed, we do not transfer the entire genus *Sphingomima* into Drepanogynini. We highlight the uncertain taxonomic position of the taxon *discolucida*, acknowledging that it may eventually be included again in *Sphingomima* if the entire genus should be transferred to Drepanogynini.

Diagnosis: Drepanogynini can be diagnosed by the combination of DNA data with up to 11 genetic markers (exemplar *Drepanogynis mixtaria* (Guenée, 1858)) ArgK (MK738841), COI (MK739615), EF1a (MK739960), IDH (MK740862), MDH (MK741181), Nex9 (MK741630), RpS5 (MK741991) and Wingless (MK742540). In the light of our phylogenetic results, the *Drepanogynis* group of genera, as classified earlier (Krüger, 2014), is split between two unrelated tribes (Drepanogynini and Diptychini). More research is needed to understand how other *Drepanogynis* species and the *Drepanogynis* group of genera *sensu* Krüger (1997, 1998, 1999, 2014) (at least 11 genera), should be classified.

Boarmiini are the sister group to a clade that comprises Macariini, Cassymini, Abraxini and Eutoeini. We found that many species currently classified as Boarmiini are scattered throughout Ennominae. Boarmiini *s.str*. are strongly supported but are technically not monophyletic because of a large number of genera which need to be formally transferred from other tribes to Boarmiini (G. Brehm et al., 2019, unpublished data for Neotropical taxa and L. Murillo-Ramos et al., 2019, unpublished data for other taxa). The results are

principally in concordance with *Jiang et al. (2017)*, who supported the monophyly of Boarmiini but with a smaller number of taxa.

The divided valva in male genitalia was suggested as a synapomorphy of Macariini + Cassymini + Eutoeini by *Holloway (1994)*. In addition, he proposed the inclusion of Abraxini in Cassymini. Although our findings support a close relationship, this group requires more study and a more extensive sampling effort. Similar findings were provided by *Jiang et al. (2017)* who suggested more extensive sampling to study the evolutionary relationships of these tribes.

Orthostixinae Meyrick, 1892

Orthostixinae were not included in our study. *Sihvonen et al. (2011)* showed this subfamily as deeply embedded within Ennominae, but unfortunately it was not represented by the type genus of the subfamily. These results agree with *Holloway (1996)* who examined *Orthostixis* Hübner, (1823) and suggested the inclusion in Ennominae despite the full development of hindwing vein M2, the presence of a forewing areole and the very broad base of the tympanal ansa. We sampled the species *Naxa textilis* (Walker, 1856) and *Orthostixis cribraria* (Hübner, 1799), but only three and one marker were successfully sequenced for these samples, respectively. We included these species in the preliminary analyses but results were so unstable that we excluded them from the final analysis. Further research including fresh material and more genetic markers are needed to investigate the position of Orthostixinae conclusively.

CONCLUSIONS

This study elucidated important evolutionary relationships among major groups within Geometridae. The monophyly of the subfamilies and the most widely accepted tribes were tested. We found strong support for the traditional concepts of Larentiinae, Geometrinae and Ennominae. Sterrhinae also becomes monophyletic when *Ergavia*, *Ametris* and *Macrotes*, currently placed in Oenochrominae, are formally transferred to Sterrhinae. The concepts of Oenochrominae and Desmobathrinae required major revision and, after appropriate rearrangements, these groups also form monophyletic subfamily-level entities. Archiearinae are monophyletic with the transfer of *Dirce* and *Acalyphes* to Ennominae. We treat Epidesmiinae as a new subfamily.

This study proposes the recognition of eight monophyletic geometrid subfamilies. Many geometrid tribes were recovered para- or polyphyletic. We attempted to address the needed taxonomic changes, in order to favor taxonomic stability of the subfamilies and many tribes, even if in an interim way, to allow other researchers to use an updated higher-taxonomic structure that better reflects our current understanding of geometrid phylogeny. Although we included a large number of new taxa, in our study, many clades remain poorly represented. This is particularly true for taxa from tropical Africa and Asia. Tribes in special need of reassessment include Eumeleini, Plutodini, Eutoeini, Cassymini and Abraxini. We hope the phylogenetic hypotheses shared here will open new paths of inquiry across Geometridae. Morphological synapomorphies have not yet been identified for many of the re- and newly defined higher taxa circumscribed by our 11-gene

data set. Likewise, there is great need, across the family, to begin the work of mapping behavioral and life history attributes to the clades identified in this work.

ACKNOWLEDGEMENTS

Harri Sihvonen (Finland) is thanked for preparing extensive African materials for the study. We are grateful to Cathy Byrne (Hobart, Australia), B.C. Schmidt (Canada, Ottawa), Alfred Moser (Curitiba, Brazil), Rolf Mörtter (Karlsruhe, Germany), Daniel Bolt (Domat/Ems, Switzerland), Florian Bodner (Vienna, Austria), Dominik Rabl (Vienna, Austria), Aare Lindt (Tallinn, Estonia), Luis Parra (Concepción, Chile), Andreas Kopp (St. Margarethen, Switzerland), Stefan Naumann (Berlin, Germany), Jaan Viidalepp (Tartu, Estonia) for providing samples for this study. We thank John Chainey, Geoff Martin and Linda Pitkin at the NHM (London) for providing access to the collections and photographs of Neotropical Ennominae moths. We thank David Wagner, Andreas Zwick and Kevin Keegan for constructive comments on the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Niklas Wahlberg received funding from the Academy of Finland (Grant No. 265511) and the Swedish Research Council (Grant No. 2015-04441). Leidys Murillo-Ramos received funding from Colciencias, 756-2016 and Universidad de Sucre, Colombia. Hamid Reza Ghanavi was funded from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no. 642241 (BIG4). Sille Holm, Erki Õunap, Andro Truuverk and Toomas Tammaru were supported by institutional research funding IUT (IUT20-33) of the Estonian Ministry of Education and Research. Gunnar Brehm received funding for fieldwork in Peru (DFG grant Br 2280/6-1) and for visits to the NHM (SYNTHESYS grant GB TAF1048 and 6817). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Academy of Finland: 265511.
Swedish Research Council: 2015-04441.
Colciencias: 756-2016.
Universidad de Sucre, Colombia.
European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no. 642241 (BIG4).
Estonian Ministry of Education and Research: IUT20-33.
Funding for fieldwork in Peru: DFG grant Br 2280/6-1) and for visits to the NHM: SYNTHESYS grant GB TAF1048 and 6817.

Competing Interests

All the authors declare that they have no competing interests.

Author Contributions

- Leidys Murillo-Ramos conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Gunnar Brehm conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Pasi Sihvonen conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Axel Hausmann contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Sille Holm performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Hamid Reza Ghanavi performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Erki Õunap performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Andro Truuverk performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Hermann Staude contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Egbert Friedrich contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Toomas Tammaru contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Niklas Wahlberg conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequences described here are accessible via GenBank with the following accession numbers:

MK738162–MK738576; MK738577–MK738903; MK738904–MK739052; MK739053–MK739692; MK739693–MK740306; MK740307–MK740765; MK740766–MK740930; MK740931–MK741338; MK741339–MK741692; MK741693–MK742127; MK742128–MK742716.

Data Availability

The following information was supplied regarding data availability:

The raw analyses are available in the Supplemental Files. The RAxML tree was used to compare some clades recovered in IQTREE.

New Species Registration

The following information was supplied regarding the registration of a newly described species:

Publication LSID: urn:lsid:zoobank.org:pub:662A9A18-B620-45AA-B4B1-326086853316

Epidesmiinae LSID: urn:lsid:zoobank.org:act:34D1E8F7-99F1-4914-8E12-0110459C2040

Tribe Chlorodontoperini LSID: urn:lsid:zoobank.org:act:0833860E-A092-43D6-B2A1-FB57D9F7988D

Tribe Drepanogynini LSID: urn:lsid:zoobank.org:act:AA384988-009F-4175-B98C-6209C8868B93.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7386#supplemental-information.

REFERENCES

Abraham D, Ryrholm N, Wittzell H, Jeremy DH, Scoble MJ, Löfstedt C. 2001. Molecular phylogeny of the subfamilies in Geometridae (Geometroidea: Lepidoptera). *Molecular Phylogenetics and Evolution* **20(1)**:65–77 DOI 10.1006/mpev.2001.0949.

Ban X, Jiang N, Cheng R, Xue D, Han H. 2018. Tribal classification and phylogeny of Geometrinae (Lepidoptera: Geometridae) inferred from seven gene regions. *Zoological Journal of the Linnean Society* **184(3)**:653–672 DOI 10.1093/zoolinnean/zly013.

Beccaloni G, Scoble M, Kitching I, Simonsen T, Robinson G, Pitkin B, Hine A, Lyal C. 2003. "Subfamily Oenochrominae" The Global Lepidoptera Names Index. Natural History Museum, London. *Available at http://www.nhm.ac.uk/our-science/data/lepindex/* (accessed 3 August 2018).

Beljaev EA. 2006. A morphological approach to the Ennominae phylogeny (Lepidoptera, Geometridae). *Spixiana* **29**:215–216.

Beljaev EA. 2008. A new concept of the generic composition of the geometrid moth tribe Ennomini (Lepidoptera, Geometridae) based on functional morphology of the male genitalia. *Entomological Review* **88(1)**:50–60 DOI 10.1134/S0013873808010089.

Beljaev E. 2016. *Annotated catalogue of the insects of Russian Far East*. Volume II. Lepidoptera. Vladivostok: Dalnauka, 812.

Brehm G, Hebert PDN, Colwell RK, Adams M-O, Bodner F, Friedemann K, Möckel L, Fiedler K. 2016. Turning up the heat at a hotspot: DNA barcodes reveal 80% more species of geometrid moths along an Andean elevational gradient. *PLOS ONE* **11(3)**:e0150327 DOI 10.1371/journal.pone.0150327.

Cook MA, Harwood LM, Scoble MJ, McGavin GC. 1994. The chemistry and systematic importance of the green wing pigment in emerald moths (Lepidoptera: Geometridae,

Geometrinae). *Biochemical systematics and ecology* **22(1)**:43–51 DOI 10.1016/0305-1978(94)90113-9.

Cook MA, Scoble MJ. 1992. Tympanal organs of geometrid moths: a review of their morphology, function, and systematic importance. *Systematic Entomology* **17(3)**:219–232 DOI 10.1111/j.1365-3113.1992.tb00334.x.

Forbes WTM. 1948. Lepidoptera of New York and neighboring states. II. *Memoirs of the Cornell University Agricultural Experiment Station* **274**:1–263.

Forum Herbulot. 2007. World list of family-group names in Geometridae. *Available at http://www.herbulot.de/famgroup.htm* (accessed 3 August 2018).

Guenée A. 1858. Histoire naturelle des insectes (Lepidoptera), Species Général des Lépidoptères. *Tom IX. X. Uranides et Phalenites I. II*. Paris: Roret, 304.

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(3)**:307–321 DOI 10.1093/sysbio/syq010.

Hausmann A. 2001. *Geometrid moths of Europe. Vol. 1: introduction to the series. Archiearinae, Oenochrominae, Geometrinae*. Stenstrup: Apollo books.

Hausmann A. 2004. *Geometrid moths of Europe. Vol. 2: Sterrhinae*. Stenstrup: Apollo books.

Hausmann A, Hebert PDN, Mitchell A, Rougerie R, Sommerer M, Edwards T, Young K. 2009. Revision of the Australian *Oenochroma vinaria* Guenée, 1858 species-complex (Lepidoptera: Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. *Zootaxa* **2239**:1–21 DOI 10.5281/zenodo.190505.

Hausmann A, Viidalepp J. 2012. *Geometrid moths of Europe. Vol. 3: Larentiinae I*. Stenstrup: Apollo books.

Heikkilä M, Mutanen M, Wahlberg N, Sihvonen P, Kaila L. 2015. Elusive ditrysian phylogeny: an account of combining systematized morphology with molecular data (Lepidoptera). *BMC Evolutionary Biology* **15(1)**:1–27 DOI 10.1186/s12862-015-0520-0.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35(2)**:518–522 DOI 10.1093/molbev/msx281.

Holloway J. 1994. The moths of Borneo, part 11: family Geometridae, subfamily Ennominae. *Malayan Nature Journal* **47**:1–309.

Holloway J. 1996. The moths of Borneo, part 9: Geometridae (incl. Orthostixini), Oenochrominae, Desmobathrinae, Geometrinae. *Ennominae Malayan Nature Journal* **49**:147–326.

Holloway J. 1997. The moths of Borneo, part 10: family Geometridae, subfamilies Sterrhinae and Larentiinae. *Malayan Nature Journal* **51**:1–242.

Inoue H. 1961. Lepidoptera: Geometridae. *Insecta Japonica* **4**:1–106.

International Commission on Zoological Nomenclature. 2012. International Code of Zoological Nomenclature. Fourth Edition. *Available at https://www.iczn.org/the-code/the-international-code-of-zoological-nomenclature/the-code-online/*.

Jiang N, Li X, Hausmann A, Cheng R, Xue DY, Han HX. 2017. A molecular phylogeny of the Palaearctic and Oriental members of the tribe Boarmiini (Lepidoptera: Geometridae: Ennominae). *Invertebrate Systematics* **31(4)**:427–441 DOI 10.1071/IS17005.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14(6)**:587–589 DOI 10.1038/nmeth.4285.

Krüger M. 1997. Revision of Afrotropical Ennominae of the *Drepanogynis* group I: the genus *Hebdomophruda* Warren, Part 1. *Annals of the Transvaal Museum* **36**:257–291.

Krüger M. 1998. Revision of Afrotropical Ennominae of the *Drepanogynis* group II: the genus *Hebdomophruda* Warren, Part 2. *Annals of the Transvaal Museum* **36**:333–349.

Krüger M. 1999. Revision of Afrotropical Ennominae of the *Drepanogynis* group III: the genera *Argyrophora* Guenée, *Pseudomaenas* Prout and *Microligia* Warren. *Annals of the Transvaal Museum* **36**:427–496.

Krüger M. 2002. Revision of Afrotropical Ennominae of the *Drepanogynis* group IV: the genus Drepanogynis Guenée (Lepidoptera: Geometridae). *Transvaal Museum Monograph* **13**:1–220 incl. 442 figs.

Krüger M. 2014. A revision of the *Mauna* Walker, 1865 and *Illa* Warren, 1914 group of genera (Lepidoptera: Geometridae: Ennominae: Nacophorini). *Annals of the Ditsong National Museum of Natural History* **4**:77–173.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29(6)**:1695–1701 DOI 10.1093/molbev/mss020.

McQuillan PB, Edwards ED. 1996. Geometroidea. In: Nielsen ES, Edwards TE, Rangsi TV, eds. *Checklist of the Lepidoptera of Australia*. Clayton: CSIRO Publishing.

Meyrick E. 1889. Revision of Australian Lepidoptera. *Proceedings of the Linnean Society of New South Wales* **41**:1117–1216 DOI 10.5962/bhl.part.15082.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8. *Available at http://www.phylo.org* (accessed June 2018).

Minet J, Scoble MJ. 1999. The Drepanoid/Geometroid assemblage. In: Kristensen NP, ed. *Handbook of Zoology, part 35, Lepidoptera, Moths and Butterflies, Vol. 1, Evolution, Systematics, and Biogeography*. Berlin: De Gruyter, 301–320.

Mironov V. 2003. Larentiinae II (Perizomini and Eupitheciini). In: Hausmann A, ed. *The Geometrid Moths of Europe* 4. Stenstrup: Apollo Books, 1–463.

Müller B, Erlacher S, Hausmann A, Rajaei H, Sihvonen P, Skou P. 2019. Ennominae II. In: Hausmann A, Sihvonen P, Rajaei H, Skou P, eds. *Geometrid Moths of Europe.* Vol. 6. Leiden: Brill, 906.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32(1)**:268–274 DOI 10.1093/molbev/msu300.

Õunap E, Javoiš J, Viidalepp J, Tammaru T. 2011. Phylogenetic relationships of selected European Ennominae (Lepidoptera: Geometridae). *European Journal of Entomology* **108(2)**:267–273 DOI 10.14411/eje.2011.036.

Õunap E, Viidalepp J. 2009. Description of *Crypsiphona tasmanica* sp. nov. (Lepidoptera: Geometridae: Geometrinae), with notes on limitations in using DNA barcodes for delimiting species. *Australian Journal of Entomology* **48(2)**:113–124 DOI 10.1111/j.1440-6055.2009.00695.x.

Õunap E, Viidalepp J, Saarma U. 2008. Systematic position of Lythriini revised: transferred from Larentiinae to Sterrhinae (Lepidoptera, Geometridae). *Zoologica Scripta* **37(4)**:405–413 DOI 10.1111/j.1463-6409.2008.00327.x.

Õunap E, Viidalepp J, Truuverk A. 2016. Phylogeny of the subfamily Larentiinae (Lepidoptera: Geometridae): integrating molecular data and traditional classifications. *Systematic Entomology* **21(4)**:824–843 DOI 10.1111/syen.12195.

Peña C, Malm T. 2012. VoSeq: a voucher and DNA sequence web application. *PLOS ONE* **7(6)**: e39071 DOI 10.1371/journal.pone.0039071.

Pitkin LM. 1996. Neotropical emerald moths: a review of the genera (Lepidoptera: Geometridae, Geometrinae). *Zoological Journal of the Linnean Society* **118(4)**:309–440 DOI 10.1111/j.1096-3642.1996.tb01268.x.

Pitkin L. 2002. Neotropical Ennomine moths: a review of the genera (Lepidoptera: Geometridae). *Zoological Journal of the Linnean Society* **135(2–3)**:121–401 DOI 10.1046/j.1096-3642.2002.01200.x.

Pitkin LM, Han HX, James S. 2007. Moths of the tribe Pseudoterpnini (Geometridae: Geometrinae): a review of the genera. *Zoological Journal of the Linnean Society* **150(2)**:343–412 DOI 10.1111/j.1096-3642.2007.00287.x.

Pitkin B, Jenkins P. 2004. Butterflies and moths of the world, generic names and their type-species. *Available at http://www.nhm.ac.uk/our-science/data/butmoth/* (accessed 29 August 2018).

Rajaei H, Greve C, Letsch H, Stüning D, Wahlberg N, Minet J, Misof B. 2015. Advances in Geometroidea phylogeny, with characterization of a new family based on *Pseudobiston pinratanai* (Lepidoptera, Glossata). *Zoologica Scripta* **44(4)**:418–436 DOI 10.1111/zsc.12108.

Rambaut A. 2012. Figtree 1.4.0. *Available at http://tree.bio.ed.ac.uk/software/figtree/* (accessed 3 August 2018).

Ratnasingham S, Hebert PDN. 2007. BOLD: the barcode of life data systems. *Molecular Ecology Notes* **7(3)**:355–364 DOI 10.1111/j.1471-8286.2007.01678.x.

Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ, Cho S, Davis DR, Baixeras J, Brown J, Parr C, Weller S, Lees DC, Mitter KT. 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLOS ONE* **8(3)**:e58568 DOI 10.1371/journal.pone.0058568.

Regier JC, Zwick A, Cummings MP, Kawahara AY, Cho S, Weller S, Roe A, Baixeras J, Brown JW, Parr C, Davis DR, Epstein M, Hallwachs W, Hausmann A, Janzen DH, Kitching IJ, Solis MA, Yen SH, Bazinet AL, Mitter C. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evolutionary Biology* **9(1)**:280 DOI 10.1186/1471-2148-9-280.

Rindge FH. 1979. A revision of the North American moths of the genus *Lomographa* (Lepidoptera, Geometridae). *American Museum Novitates* **2673**:1–18.

Rota J. 2011. Data partitioning in Bayesian analysis: molecular phylogenetics of metalmark moths (Lepidoptera: Choreutidae). *Systematic Entomology* **36(2)**:317–329 DOI 10.1111/j.1365-3113.2010.00563.x.

Scoble MJ. 1992. *Lepidoptera: form function and diversity*. Oxford: Oxford University Press.

Scoble MJ. 1999. *Geometrid Moths of theWorld: a catalogue (Lepidoptera, Geometridae)* 1, 2. Collingwood: CSIRO.

Scoble MJ, Edwards ED. 1990. *Parepisparis* Bethune-Baker and the composition of the Oenochrominae (Lepidoptera: Geometridae). *Entomologica Scandinavica* **20(4)**:371–399 DOI 10.1163/187631289X00375.

Scoble MJ, Hausmann A. 2007. *Online list of valid and available names of the Geometridae of the world. Available at http://www.lepbarcoding.org/geometridae/species_checklists.php*.

Sihvonen P, Kaila L. 2004. Phylogeny and tribal classification of Sterrhinae with emphasis on delimiting Scopulini (Lepidoptera: Geometridae). *Systematic Entomology* **29(3)**:324–358 DOI 10.1111/j.0307-6970.2004.00248.x.

Sihvonen P, Mutanen M, Kaila L, Brehm G, Hausmann A, Staude HS. 2011. Comprehensive molecular sampling yields a robust phylogeny for geometrid moths (Lepidoptera: Geometridae). *PLOS ONE* **6(6)**:e20356 DOI 10.1371/journal.pone.0020356.

Sihvonen P, Staude H. 2011. Geometrid moth *Afrophyla vethi* (Snellen, 1886) transferred from Oenochrominae to Sterrhinae (Lepidoptera: Geometridae). *Metamorphosis* **22**:102–113.

Sihvonen P, Staude HS, Mutanen M. 2015. Systematic position of the enigmatic African cycad moths: an integrative approach to a nearly century old problem (Lepidoptera: Geometridae, Diptychini). *Systematic Entomology* **40(3)**:606–627 DOI 10.1111/syen.12125.

Skou P, Sihvonen P. 2015. *The Geometrid moths of Europe. Vol. 5: Ennominae I*. Stenstrup: Apollo Books.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30(9)**:1312–1313 DOI 10.1093/bioinformatics/btu033.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* **57(5)**:758–771 DOI 10.1080/10635150802429642.

Staude HS. 2001. A revision of the genus *Callioratis* Felder (Lepidoptera: Geometridae: Diptychinae). *Metamorphosis* **12**:125–156.

Staude H, Sihvonen P. 2014. Revision of the African geometrid genus *Zerenopsis* C. &. R. Felder-moths with peculiar life histories and mating behaviors (Geometridae: Ennominae: Diptychini). *Metamorphosis* **25**:11–55.

Strutzenberger P, Brehm G, Gottsberger B, Bodner F, Seifert CL, Fiedler K. 2017. Diversification rates, host plant shifts and an updated molecular phylogeny of Andean *Eois* moths (Lepidoptera: Geometridae). *PLOS ONE* **12(12)**:e018843 DOI 10.1371/journal.pone.0188430.

Trifinopoulos J, Minh B. 2018. IQ-TREE manual: frequently asked questions. *Available at http://www.iqtree.org/doc/Frequently-Asked-Questions* (accessed 13 August 2018).

Van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen S, Zahiri R, Adamski D, Baixeras J, Bartsch D, Bengtsson BÅ, Brown JW, Bucheli SR, Davis DR, Prins J de, De Prins W, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara AY, Koster J, Kozlov M, Lafontaine JD, Lamas G, Landry J, Lee S, Nuss M, Park K, Penz C, Rota J, Schintlmeister A, Schmidt BC, Sohn J, Solis MA, Tarmann GM, Warren AD, Weller S, Yakovlev RV, Zolotuhin VV, Zwick A. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.), Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148**:212–221.

Viidalepp J. 2006. Cladistic analysis of the subfamily Larentiinae. In: Hausmann A, McQuillan P, eds. Proceedings of the Forum Herbulot 2006. Integration of molecular, ecological and morphological data: recent progress towards the higher classification of the Geometridae (Hobart, 19–20 January 2006). *Spixiana* **29**:202–203.

Viidalepp J. 2011. A morphological review of tribes in Larentiinae (Lepidoptera: Geometridae). *Zootaxa* **3136(1)**:1–44 DOI 10.11646/zootaxa.3136.1.1.

Vári L, Kroon DM, Krüger M. 2002. *Classification and checklist of the species of Lepidoptera recorded in Southern Africa*. Chatswood: Simple Solutions.

Wahlberg N, Peña C, Ahola M, Wheat CW, Rota J. 2016. PCR primers for 30 novel gene regions in the nuclear genomes of Lepidoptera. *ZooKeys* **596**:129–141 DOI 10.3897/zookeys.596.8399.

Wahlberg N, Snäll N, Viidalepp J, Ruohomäki K, Tammaru T. 2010. The evolution of female flightlessness among Ennominae of the Holarctic forest zone (Lepidoptera, Geometridae). *Molecular Phylogenetics and Evolution* **55(3)**:929–938 DOI 10.1016/j.ympev.2010.01.025.

Wahlberg N, Wheat CW. 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology* **57(2)**:231–242 DOI 10.1080/10635150802033006.

Yamamoto S, Sota T. 2007. Phylogeny of the Geometridae and the evolution of winter moths inferred from a simultaneous analysis of mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution* **44(2)**:711–723 DOI 10.1016/j.ympev.2006.12.027.

Young CJ. 2003. The place of the Australian Nacophorini in the Geometridae. *Spixiana* **26**:199–200.

Young CJ. 2006. Molecular relationships of the Australian Ennominae (Lepidoptera: Geometridae) and implications for the phylogeny of the Geometridae from molecular and morphological data. *Zootaxa* **1264(1)**:1–147 DOI 10.11646/zootaxa.1264.1.1.

Young CJ. 2008. Characterization of the Australian Nacophorini using adult morphology, and phylogeny of the Geometridae based on morphological characters. *Zootaxa* **1736(1)**:1–141 DOI 10.11646/zootaxa.1736.1.1.

Paper II

Biogeography and Diversification Dynamics of the Moth Family Geometridae (Lepidoptera)

Hamid Reza Ghanavi[1], Nicolas Chazot[2], Isabel Sanmartín[3], Leidys Murillo-Ramos[1,4], Sebastián Duchêne[5], Pasi Sihvonen[6], Gunnar Brehm[7] and Niklas Wahlberg[1]

[1] Systematic Biology Group, Biology department, Lund University, Lund, Sweden.

[2] Department of Biological & Environmental Sciences Gothenburg University, Gothenburg, Sweden.

[3] Real Jardín Botánico (RJB), CSIC, Madrid, Spain.

[4] Grupo Biología Evolutiva, Department of Biology, Universidad de Sucre, Sincelejo, Sucre, Colombia.

[5] Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia.

[6] Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland.

[7] Institut für Zoologie und Evolutionsbiologie, Phyletisches Museum, Jena, Germany.

Abstract

Highly diverse groups offer a great opportunity to study the processes and mechanisms which have shaped their evolutionary success. Within Lepidoptera, moths of the family Geometridae, with 23 thousand described species and a worldwide distribution excepting Antarctica, are one such group. Here, we present the first study on the diversification dynamics and biogeographic processes that

shaped the current diversity patterns and distribution ranges in this family. In the absence of reliable fossils for the group, we used a (published) multi-locus data set representing XX generic diversity in the family, secondary calibration points from the literature and relaxed molecular clocks to generate a time-calibrated phylogeny using the software BEAST2. This time tree was used to reconstruct the biogeographic evolution of Geometridae, implementing in RevBayes a Bayesian approach to the Dispersal-Extinction-Cladogenesis (DEC) model that incorporates palaeographic-based dispersal graphs with uncertainty in geological ages. We also implemented a Bayesian time-variable, episodic birth-death model and a model that allows branch-specific speciation rates, to reconstruct the diversification dynamics in the family. Our results suggest that the most recent common ancestor of Geometridae was distributed in the New World, with the Neotropics the most likely ancestral area. An increase in diversification rates occurred at circa 30-40 mya, at a time of a major global climate cooling. There were also shifts in speciation rates that were clade-specific at around 25-35 mya, coincident with a period of major climate change. These results point out to different biogeographical and evolutionary histories per area, to show the differences of the diversification rates in different biogeographical areas through time, showing the relative importance of each region in the diversification history of the family.

Key words: Geometridae, diversification, biogeography, Lepidoptera

Introduction

Since the inception of natural sciences as a field at the end of the XVIII century, the heterogeneity in the spatial distribution of biodiversity on Earth has captivated the interest of scientists (Hooker 1854; Darwin 1859; Wallace 1869, 1876). For example, the Latitudinal Diversity Gradient (LDG) describes a general pattern in which the number of species of organisms increases from the poles to the equator, with tropical regions being the most species-rich (Fischer 1960; Buzas et al. 2002; Jablonski et al. 2006; Brown 2014; Kinlock et al. 2017). The LDG has been partly explained by present-day differences in environmental factors, such as temperature and precipitation, favouring higher productivity in these regions. However, there is increasing consensus that historical processes such as the climatic and geological history of a landmass, driving lineage speciation and extinction rates, played a prominent role in shaping biodiversity patterns (Linder 2001; Antonelli et al. 2009; Buerki et al. 2011; Condamine et al. 2013a).

Extant diversity also differs greatly among groups of organisms. In insects, for example, the order Strepsiptera currently includes very few representatives (about 600 species), whereas related groups with a similar time of origin (e.g., Coleoptera), exhibit species numbers that are orders of magnitude higher (more than 300,000

described species). Also, present-day diversity is only a narrow snapshot of the course of evolution. A great percentage of species that originated on Earth have now become extinct, and many major branches in the Tree of Life have not left any extant descendants, such as the ammonites and trilobites (Payne and Clapham 2012). Understanding the origin of present-day extant diversity is a challenging task, but reconstructing the extinct and unobserved diversity is even more complex (Purvis 2008; Meseguer et al. 2015; Sanmartín and Meseguer 2016). Initially, fossils and their associated stratigraphic age were used for exploring hidden patterns of diversity in the past and the role of extinction in shaping these patterns (e.g. Simpson 1944; Stanley 1980). In recent decades, statistical methods have emerged that make use of molecular rates and phylogenetic information to elucidate the dynamics of diversification within a group of organisms (Hey 1992; Nee et al. 1992, 1994; Purvis 2008; Morlon et al. 2011a; Sanmartín and Meseguer 2016). These methods require information about the times of divergence of extant lineages, often calibrated with fossil information or rates of molecular evolution from related groups. The power of such methods to estimate past diversification rates has been recently fostered by high-throughput sequencing (HTS) techniques, allowing an exponential increase in the number of genes analysed, the development of more realistic models, permitting rates of diversification to change over time and across lineages, and the addition of Bayesian approaches to account for error in parameter estimation (Höhna et al. 2011; Morlon et al. 2011b; Stadler 2013; May et al. 2016; Sanmartín and Meseguer 2016). Fossils remain important to provide a temporal calibration point or to inform on the magnitude or even direction of change, and have demonstrated that their inclusion can dramatically alter conclusions on the evolutionary history of a group, especially in deep time (Mao et al. 2012; Meseguer et al, 2015, 2018; Landis et al. 2020). However, the majority of organisms have highly incomplete fossil records, or even absent, often due to poor fossilization rates.

One such group that lacks a reliable fossil record but still exhibits remarkable levels of diversity are Geometridae moths. With more than 23,000 described species (van Nieukerken et al. 2011), the family represents one of the major radiations within Lepidoptera, and insects in general. Yet, fossil information is extremely scarce: thus far, only one fossil has been unambiguously assigned to the family (Fischer et al. 2019). Recently, Geometridae has been the focus of a series of phylogenetic studies, thanks to the publication of new extensive molecular datasets (Brehm et al. 2019; Murillo-Ramos et al. 2019; Sihvonen et al. 2020). The increase in taxon sampling, especially for higher-level lineages, and the generation of well-supported multi-locus phylogenetic hypotheses, enables the investigation of diversification dynamics in the group.

Geometrid moths have a worldwide distribution and also exhibit decreasing diversity levels from tropical to temperate areas, in accordance with the LDG pattern (Beck et al. 2017). Western Palaearctic species are by far the best studied, due to the work of many entomologists who have been observing and describing species

since the Age of Exploration. Nowadays, relatively few new species are described from this region compared with tropical regions such as South East Asia (e.g. Holloway 1993, 1996, 1997), Africa (e.g. Tujuba et al. 2020) and the Neotropics (e.g. Brehm 2018).

Despite the family's cosmopolitan distribution, most species of Geometridae are endemic to a single continent or biogeographic region, and often limited to a single locality. The age of origin of the family has been placed in the Cretaceous in previous studies focusing on the order Lepidoptera (Wahlberg et al. 2013; Kawahara et al. 2019) . This relatively ancient age and the endemic distribution patterns of extant species makes this group especially attractive for studying the historical processes that drive patterns of diversity over space and time. To date, there is no biogeographic study on the origin of geometrid moths, when and how they reached their current cosmopolitan distribution and excess levels of diversity. The large majority of works in Lepidoptera have focused on the butterfly families Nymphalidae (Wahlberg and Freitas 2007; Kodandaramaiah and Wahlberg 2007, 2009; Aduse-Poku et al. 2009, 2015; Müller et al. 2010; Matos-Maraví et al. 2013, 2014; Toussaint and Balke 2016; Kodandaramaiah et al. 2018; Toussaint et al. 2020), Papilionidae (Condamine et al. 2013b), Pieridae (Müller et al. 2013) Lycaenidae (Vila et al. 2011), and Hesperiidae (Toussaint et al. 2019). Only one published study exists on the biogeography of a moth family (Choreutidae, Rota et al. 2016). Also, few works have studied patterns of speciation and extinction using statistical methods in butterflies (Condamine et al. 2012, 2018; Toussaint and Balke 2016; Chazot et al. 2019, 2020). Many of these works highlight the major role played by global changes in climate and past geography in the evolution of Lepidoptera extant diversity.

In this study, we used a published multi-locus phylogeny of Geometridae as a robust phylogenetic framework for inferring biogeographic and diversification patterns in the family. Specifically, we employed Bayesian statistical approaches to molecular dating, biogeographic reconstruction and diversification analyses to infer lineage divergence times, ancestral geographic ranges, changes in rates of migration, extirpation (local extinction), and lineage speciation and extinction over time. Our aim was to identify key time periods, with major climatic or geological changes, as well as biogeographic regions (landmasses), that played a significant role in the origin and evolution of Geometridae extant diversity.

Material and Methods

Taxon sampling and phylogeny

Our study is based on the most recent and comprehensive molecular dataset of family Geometridae by Murillo-Ramos et al. (2019). This dataset includes 93 tribes

and 1192 species sequenced for 10 low-copy nuclear markers (*ArgK*, *Ca-ATPase*, *CAD*, *EF-1alpha*, *GAPDH*, *IDH*, *MDH*, *Nex9*, *RpS5* and *wingless*) and one mitochondrial gene (*COI*), with a total length of 7665 bp. Most nodes in this phylogeny received high clade support (Bayesian posterior probabilities), and thus constitute a sound phylogenetic template to explore biogeographic and diversification patterns. The phylogeny was rooted using taxa from the families Uraniidae, Pseudobistonidae, Epicopeiidae and Sematuridae, which are the closest taxa to Geometridae (Regier et al. 2009, 2013).

Time Calibration

To generate a time-calibrated topology, we used relaxed molecular clocks implemented in the Bayesian software BEAST2 (Bouckaert et al. 2014). The molecular dataset above was partitioned by gene, with independent molecular substitution models (Table 1) inferred by ModelFinder (Kalyaanamoorthy et al. 2017) implemented in the maximum likelihood inference software IQ-TREE 1.6.5 (Nguyen et al. 2015). Initial test runs failed to converge, probably due to the size and complexity of the dataset, which included XX percentage of missing data. To make the analysis computationally possible, we enforced the tree topology obtained by Murillo-Ramos et al. (2019) by modifying the xml file manually and disabling operators for MCMC tree moves. A birth-death tree prior with incomplete sampling (Stadler 2013) and a lognormal relaxed clock model (Drummond et al. 2006) were used. The clock and tree priors were linked across all partitions. Four calibration points based on secondary age estimates were used to infer absolute divergence times for the following nodes: the most recent common ancestor (MRCA) of Geometridae; the MRCA of *Scopula* and *Cyclophora*; the MRCA of *Archiearis* and *Alsophila*; and the MRCA of *Biston* and *Alsophila*. Secondary age estimates to calibrate these nodes were obtained from the Lepidoptera time tree of Wahlberg et al. (2013). Normal distribution priors were used, with a standard deviation spanning the 95% high-posterior-density (HPD) credibility intervals estimated by Wahlberg and collaborators (2013); see Table 2 for details on the value of parameters for these priors. A pre-run analysis of one MCMC chain for over $2x10^8$ generations was performed as burnin to ensure the fine-tuning of priors. Four independent chains (with different seed numbers) were then run for over 10^8 generations using the pre-run as a starting point. Each run was checked for adequate convergence and mixing using TRACER 1.6 (Rambaut et al. 2018), first independently and then together. All parameter ESS values scored higher than 200. The software LogCombiner and TreeAnnotator (both available in the BEAST2 package) were used to merge the tree files from the independent runs, and generate a maximum clade credibility (MCC) tree, representing the mean and 95% HPD interval for all nodal ages. Outgroup taxa were then removed from this tree, using the R statistical framework (R Core Team 2017) with package *ape* v. 5.2 (Paradis and Schliep 2019), before running the biogeographic and diversification analyses to avoid incomplete taxon sampling

biases, since outgroups were represented by a few species. The BEAST MCC tree is provided in the Supplementary Material.

Table 1: Gene partitions and the substitution model as inferred by ModelFinder. "Infor" stands for the number of parsimony-informative sites and "Invar" for the number of invariant sites.

Partition	Markers	Length (bp)	Infor	Invar	Model
1	*ArgK*	388	192	152	GTR+F+I+G4
2	*Ca-ATPase*	444	176	221	SYM+I+G4
3	*CAD*	865	482	318	GTR+F+I+G4
4	*COI*	1476	802	502	GTR+F+I+G4
5	*EF1a*	1240	516	577	SYM+I+G4
6	*GAPDH*	691	324	290	SYM+I+G4
7	*IDH*	722	363	294	GTR+F+I+G4
8	*MDH*	407	209	161	SYM+I+G4
9	*Nex9*	420	241	138	GTR+F+I+G4
10	*RpS5*	603	259	265	SYM+I+G4
11	*WntGeo*	409	269	98	SYM+I+G4
Total		**7665**	**3833**	**3016**	

Table 2: Calibration points used to date the phylogeny of geometrid moths. The age unit is million years.

Dated nodes	Age (mean)	95% CI	Prior distribution	Prior sigma
Crown	71±5.3	(61–82)	Normal	1
Scopula and *Cyclophora* ancestor	67±5.8	(56–79)	Normal	7
Archiearis and *Alsophila* ancestor	60±5.6	(49–71)	Normal	6.7
Biston and *Alsophila* ancestor	41±5.1	(31–51)	Normal	6.1

Diversification analyses

To study major changes or shifts in extinction and speciation rates over time that affected all lineages in the family simultaneously, we used time-variable, episodic birth death models (EBD) implemented in the software RevBayes (Höhna et al. 2016). The EBD model is a Bayesian approach to the maximum likelihood discrete,

episodic birth-death model implemented in TreePar (Stadler 2011a, 2011b). In this model, time is divided into discrete time bins. The rates of speciation and extinction remain constant within each time slice but can vary between time slices according to a Compound Poisson Process (Condamine et al. 2018). In our analysis, we evaluated three different values for the width of time bins: ten, five and two million-year (My) intervals. To account for incomplete taxon sampling in our phylogeny, we assumed that our taxon sampling was random and introduced a global sampling fraction parameter (ρ), which was set to a fixed value of 0.05. This was estimated as the number of tips in the phylogeny (1192) divided by the total number of described species in Geometridae (24000 species, updated from van Nieukerken et al. 2011), which is a conservative approach for the total number of species. Priors for speciation and extinction rates were modelled as lognormal distributions with the mean centred in the extant diversity (ln (Nº extant species/2.0 / Root age), and the standard deviation set to 0.587405, which places 95% uncertainty of one order of magnitude around the mean. After 10^4 generations as a burnin or pre-run, the analysis was run for another $5x10^4$ generations for each time interval width. The results were then visualized in R (R Core Team 2017) using the package RevGadget (Figure 2).

To infer shifts in rates of diversification over time that affect only a given clade in the phylogeny, we used a branch-specific diversification (BSD) model implemented in RevBayes (Höhna et al. 2019). The model is a discretized approximation to the continuous-time clade-diversification approach implemented in BAMM (Rabosky 2014). As in BAMM, a stick break-point process, Compound Poisson Process (CPP) is used to detect points in time where there is a significant change in diversification rates and to discriminate between different diversification regimes/scenarios; the model can account for incomplete taxon sampling. However, unlike BAMM, the BSD model properly accounts for the possibility of changes or shifts in speciation rates in unobserved, extinct lineages by using Maddison et al. (2007) numerical integration approximation accounting for all possible of event change types in discrete, infinitesimal time bins (Höhna et al. 2019). Failing to do so has been shown to bias posterior estimates for shifts in diversification rates in BAMM (Meyer and Wiens 2018). As in BAMM, extinction rates are modelled as constant in BSD and only the magnitude and direction of shifts in speciation rates are inferred; accounting for both shifts in extinction and speciation rates can introduce unidentifiability of parameters and diversification regimes (Höhna et al. 2019). The rho parameter and all other priors were set as in the EBD analysis. We ran the analysis for $2x10^4$ generations. Scripts to run these analyses are provided in the Supplementary Material.

Biogeography

Biogeographic evolution was inferred using the Dispersal-Extinction-Cladogenesis model (Ree et al. 2005; Ree & Smith, 2008) implemented in a Bayesian framework in RevBayes (Landis et al. 2018). An epoch, time-stratified DEC analysis was

performed following the settings in the website tutorial (https://revbayes.github.io/tutorials/biogeo/biogeo_epoch.html), with modifications as in Theode et al. (2019). Following Ree and Sanmartín (2018), cladogenetic events were limited to narrow sympatry, peripheral-isolate speciation (subset sympatry) and vicariance (allopatry). We used seven biogeographic regions as units for the analysis, corresponding to cratons or persistent landmasses since the Late Mesozoic, and which harbour endemic species: the Afrotropics, Australia, Nearctic, Neotropics, New Zealand, Oriental and Palaearctic. An eighth area, Antarctica, was included in the analysis, even if currently no geometrid species are known to occur there, because it potentially played an important role as a land bridge for dispersal in the southern continents over the Late Mesozoic and Cenozoic. The rate of biogeographic change was modelled as a uniform distribution between 10E-4 and 10E-1/Myr^{-1}. The rate of extirpation was modelled as a loguniform distribution with an expectation of one event per million year. The phylogeny was sliced into five consecutive time intervals or bins; each was assigned a different dispersal rate matrix, where the baseline migration rate (1.0) was multiplied by a relative "scaler" value according to paleogeographical connectivity through time (Table 3). To define paleogeographical connectivity between our biogeographic units (cratonic landmasses), we used the EarthViewer application (www.hhmi.org/biointeractive/earthviewer), which is based on the Paleomap Project by Christopher R. Scotese (http://www.scotese.com). We employed our own R script to extract and plot marginal probabilities for each geographic state as a pie chart on the nodes of the MCC tree. See Supplementary Material for more details on area definition, dispersal rate scalers for the epoch model and the script used in the analysis.

To summarize the frequency of dispersal events between pairs of areas or regions over the reconstructed history of Geometridae, we used the R package qgraph (Epskamp et al. 2012). We made several assumptions to estimate the number of dispersal events. The source and sink regions for dispersal events were identified by comparing the most probable ranges at the upper and lower node of a branch. If at least one area was gained along the branch, we assumed at least one dispersal event took place. In cases where two or more areas could have been the source of the dispersal event, we used the time-stratified dispersal matrices implemented in the epoch model to choose the most probable source area. If two or more potential source areas had the same probability, we randomly picked one. Finally, a timing for the dispersal event was then randomly sampled along the branch. This procedure was repeated 1000 times and the sum of events between regions was calculated at each repetition and finally averaged.

Biogeography and diversification

To assess differences among biogeographic regions in lineage accumulation over time, we performed two different analyses. First, we estimated the relative frequency of lineages occupying a given region through time. To do so, we

performed an estimation of the frequency of dispersal events and their timing, using a similar approach to the analysis above for counting the total number of events between regions. Then, we divided time into 1 my time bins, and for each time bin we calculated the relative frequency of lineages in each region during that time period.

Second, we estimated the average rate of diversification in each region through time by combining the biogeographic ancestral state estimate and the branch-specific diversification analysis, using a similar approach to Chazot et al. (2020). From the DEC biogeographic analysis, we identified the dispersal events along each branch and randomly sampled its timing, following a similar procedure to the analysis above for counting the total number of dispersal events between regions. In addition, for each branch, we recovered the net diversification rate estimated by the branch-specific diversification analysis. Hence, for each branch we obtained the most probable biogeographic state (geographic range) and the net diversification rate. Finally, we divided time (the phylogeny time scale) into 1 my time bins, and for each time bin we calculated the mean diversification rate for all branches inferred as occupying a given a biogeographic state during that time interval. We repeated this procedure 1000 times. This allowed us to estimate the mean diversification rate per biogeographical area over time. To perform this analysis we used dendextend (Galili 2015), phyloch (Heibl 2008), phylotate (Beer and Beer 2019) and TreePar (Stadler 2015) R packages, and a custom script modified from the approach described in Chazot et al. (2020).

Table 3: Biogeographical areas (A) and time slices (B) used in the biogeography analysis. "mya" stands for million years ago.

A)

Areas	Code
Neotropic	N
Afrotropic	F
Nearctic	A
Palearctic	P
Oriental	O
Australia	S
New Zealand	Z
Antarctica	T

B)

Number	Time-frame
1	0 - 10 mya
2	10 - 30 mya
3	30 - 40 mya
4	40 - 50 mya
5	> 50 mya

All data in the supplementary material, the alignment, the script files and the results can be found and downloaded from the GitHub repository: github.com/Hamidhrg/Geometridae2020.

Results

Time-calibration

Divergence time estimation in BEAST reveals that Geometridae diverged from its sister-family Uraniidae (i.e., the stem age of geometrids) in the Late Cretaceous, at approximately 76.8 (95% HPD: 72.9–80.7) million years ago (mya, Figure 1). The first split within the extant radiation of Geometridae (i.e., the crown age of the family), occurred shortly after, at 70.8 mya (68.9–72.7), separating subfamily Sterrhinae from the rest of the family. The other subfamilies diverged during the Early Cenozoic, with ages ranging between 68 and 50 million years (Figure 1). The start of lineage diversification within each subfamily was also diverse, with Sterrhinae as the oldest (65 mya) and Epidesmiinae as the youngest (32 Mya, Figure 1).

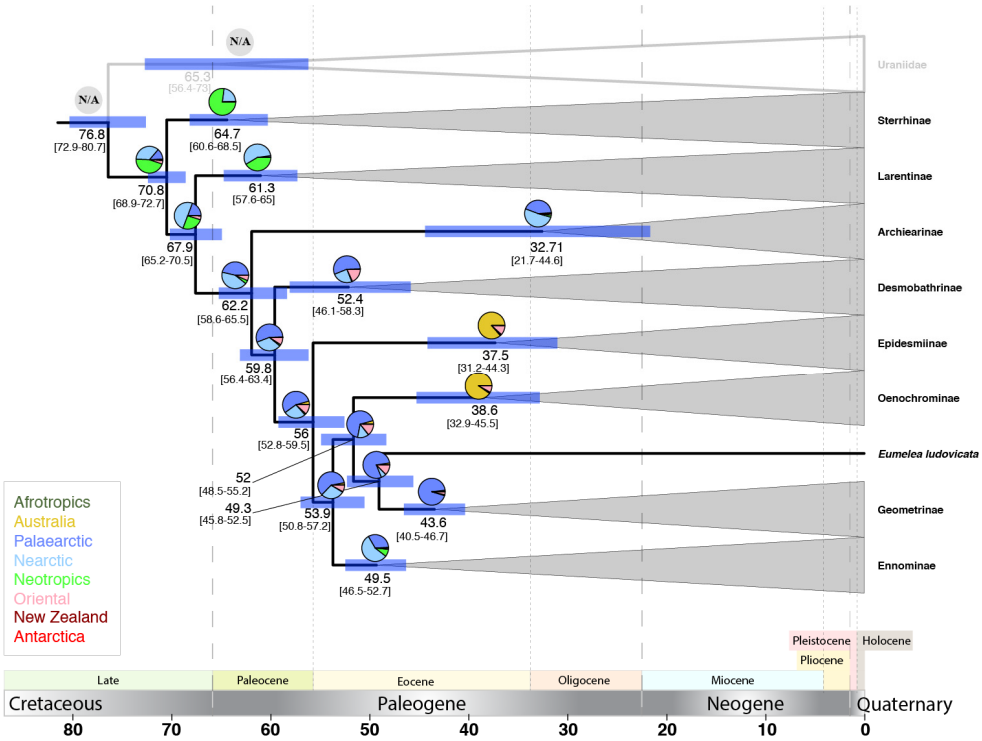


Figure 1: Times of divergence of the major lineages (subfamilies) within Geometridae and their inferred distributions. The age of each node is shown in million years with its 95% interval. Pie charts show the sum of the marginal probability of each inferred biogeographical state.

Diversification

The EBD analysis (Figure 2) supported a gradual increase in the net rate of diversification over time in the family, punctuated by with two upward shifts in diversification rates at 35 and 10 mya.

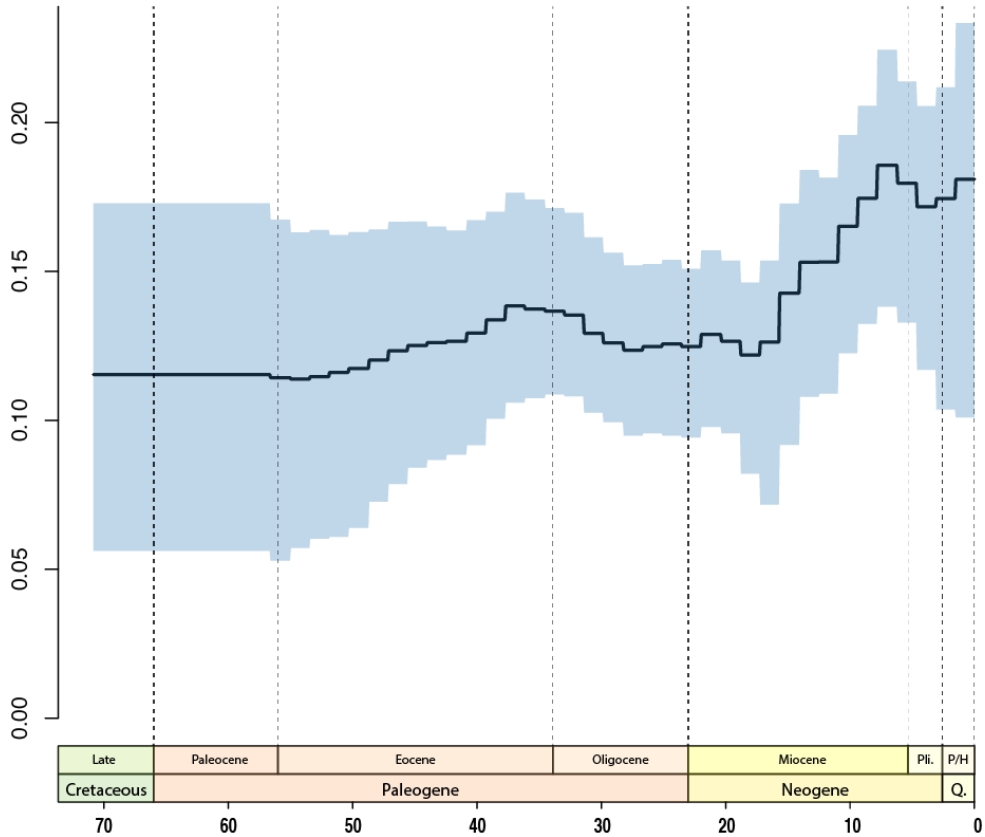


Figure 2: Net diversification rate over time in Geometridae inferred with the episodic birth death model (EBD) in RevBayes, showing the upward rate shifts at 35 and 10 mya. “Pli” stands for Pliocene, “P/H” for Pleistocene/Holocene, and “Q” for Quaternary. The scale on the X axis is on millions of years (mya), and that of the Y axis indicates the net diversification rate.

The BSD analysis (Figure 3) reveals a homogeneous net diversification rate for the major part of the tree, with independent increases in the rate of speciation for several clades within subfamilies. In Larentiinae, a nearly threefold increase in diversification rates is observed at 35 mya on the clade comprising Scotopterygini, the Euphyiini-Xanthorrhoini complex, and the Larentiini complex (4.2.18–4.2.20 in Brehm et al. 2019) (Figure 3 A). Two further shifts towards higher diversification rates are detected at around 15 and 10 mya in the node leading to Triphosini and a subclade within genus Eupithecia (Figure 3 B and C). In subfamily Geometrinae, tribe Geometrini also presents a pronounced increase in diversification rates dated circa 15 mya (Figure 3 D).

Subfamily Ennominae harbours approximately half of the total known diversity of Geometridae. Increases in net diversification rates are detected in two independent

lineages at around 35 mya: one clade comprises mainly tribe Ennomini, and the other one is located within tribe Boarmiini (Figure 3 E and F). A more recent upward shift in speciation rates can be observed within this subfamily, at 15 mya, affecting species within genus Cleora (Figure 3). The diversification dynamics of the tribe Boarmiini are investigated in more detail by Murillo-Ramos et al. (submitted).

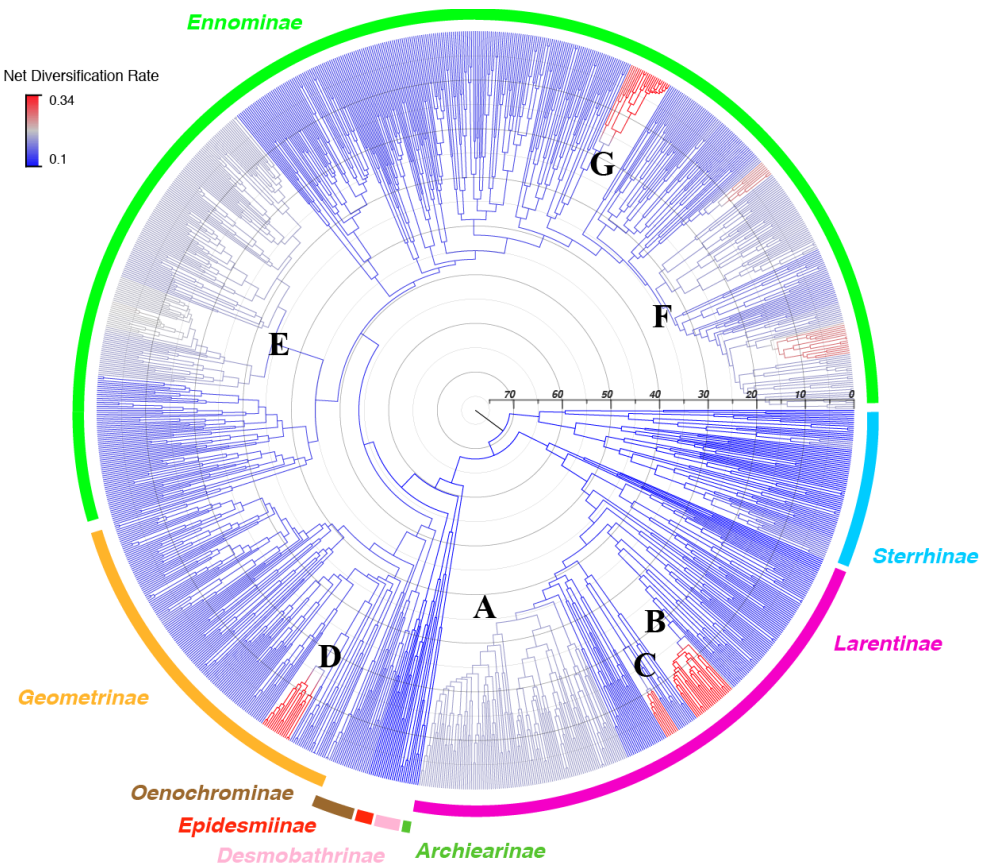


Figure 3: Diversification rate variation across Geometridae subfamilies identified with the branch-specific diversification analysis (BDS) in RevBayes. The numbers on the time scale are in million years before present. The letters A-G represent lineages discussed in the text.

Biogeography

Our results suggest that the MRCA of Geometridae (Figure 1) was distributed on the American continents, most probably in the Neotropics (marginal posterior probability, pp = 0.43). The MRCA of subfamily Sterrhinae (64.7 Ma) share the

same Neotropical ancestral distribution (pp = 0.74). The ancestor of Larentiinae and the other subfamilies (64.9 Ma) is reconstructed as Nearctic (pp = 0.51), although the Neotropics is another possibility. The MRCA of Larentiinae (61.3 Ma) is inferred to have originated in the Nearctic region with high marginal probability (pp = 0.55). The MRCA of the remaining subfamilies (62.2 Ma) was probably Holarctic distributed, with the Palaearctic and Nearctic regions receiving similar marginal posterior probabilities (pp = 0.4 or 0.37, Figure 1). The MRCA of Archiearinae (32.72 Ma) is inferred as distributed in the Nearctic (pp = 0.51) or the Palaearctic region (pp = 0.37). The subfamily Desmobathrinae likely originated in the Palaearctic region (pp = 0.52), with the Nearctic and Oriental regions as alternative ancestral ranges. The MRCAs of Epidesmiinae and Oenochrominae are reconstructed as distributed in the Australian region with high probability (pp = 0.86 and pp = 0.88 respectively); their extant distribution is inferred to have resulted from independent dispersal events from the Palaearctic region (possibly via the Oriental region; see discussion below). Finally, the MRCAs of subfamilies Geometrinae and Ennominae are reconstructed as Palaearctic (pp = 0.75) and Nearctic (pp = 0.51), respectively.

Estimates of the frequency of dispersal events between pairs of regions (Figure 4) show that the large majority of dispersal events occurred between geographically adjacent regions, as expected from the model (277 adjacent dispersal events, Figure 4; Supplementary Material). Long distance dispersal events between regions that were not connected by landmass are present (Figure 4) but are less common (46 long distance events, Supplementary Material). Another observation is the dominance of migration events between the Palaearctic, Nearctic, Neotropical and Afrotropical regions, and the relative isolation of the Australian and Oriental regions. In particular, the Palaearctic region appears as the source area with the highest number of dispersal events, mostly towards the Afrotropical and Oriental regions (Figure 4). A high rate of dispersal towards the Nearctic is also inferred; however, dispersal in the opposite direction, from the Nearctic to the Palaearctic, appears more frequent in our reconstruction. The Nearctic, indeed, is inferred as the second most important hub of dispersal events, mainly towards the Nearctic and Neotropical regions (Figure 4). New Zealand is reconstructed as the sink of some dispersal events, with no "outwards" migration, in agreement with our DEC ancestral range reconstruction (Figure 1).

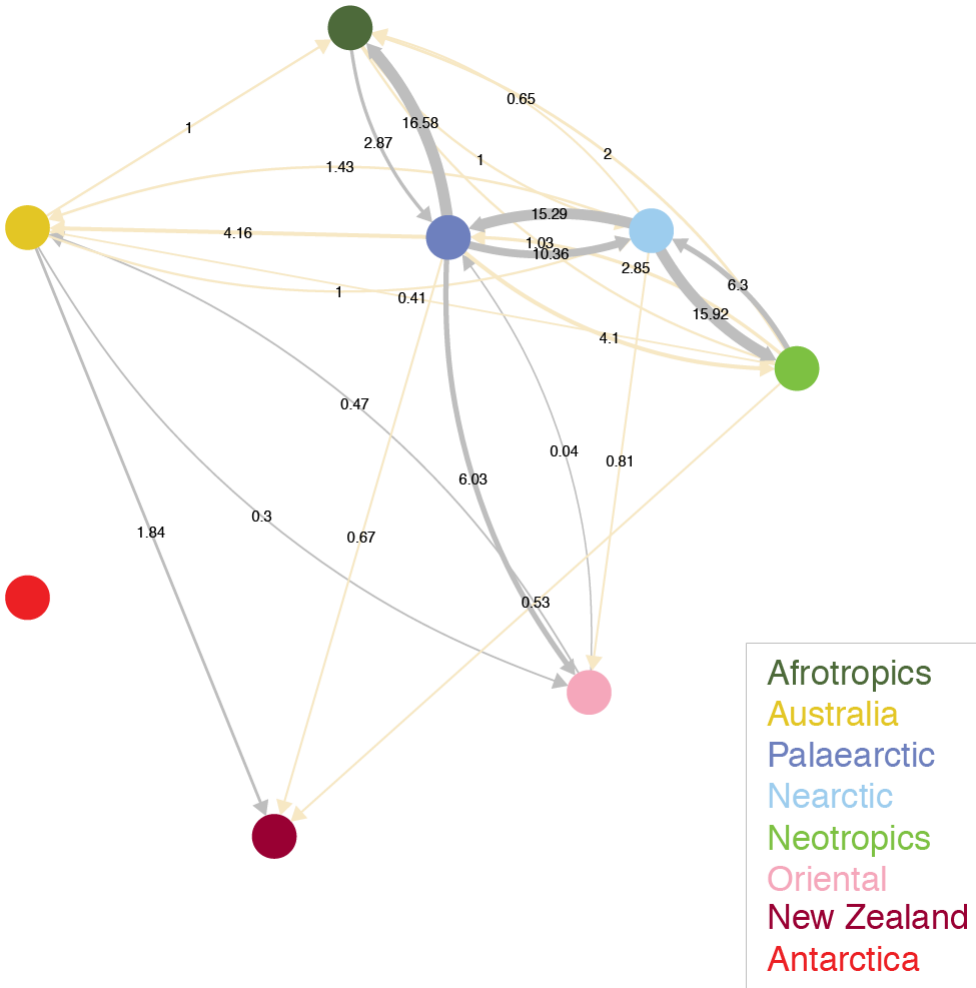


Figure 4: Number of dispersal events between the different biogeographical areas without considering the last 20 Myr. Adjacent dispersals between landmasses or continents that showed paleogeographical connectivity are shown with grey arrows, and long-distance dispersals with yellow arrows.

Figure 5 shows estimates of the relative frequency of lineages occurring in each biogeographic region through time (Figure 5a), and the variation in the net diversification rate per region over time (Figure 5b). Both estimates show a similar pattern, in accordance with the ancestral range reconstruction (Figure 1). The Neotropics and Nearctic regions show the highest frequency and net diversification rate at the root of the tree, being gradually replaced by the Palearctic region, once this landmass was colonized in the late Cretaceous. From c. 45 mya, the Palaearctic region becomes the main cradle of diversification, with a rise in the frequency of lineages, which stabilises around 25 mya but and increases again towards the present

(Figure 5a,b). Though values are lower than in the Palearctic, the Afrotropical region and Australia show a similar pattern, of increasing relative frequency of lineages towards the present, especially from the Mid Cenozoic onwards (Late Oligocene, 25 mya; Figure 5a). The net diversification rate in Africa also increases sharply after 25 mya, but this pattern cannot be observed in Australia (Figure 5b). The Oriental region and New Zealand exhibit a nearly flat pattern of diversification after their colonization circa 25 mya (Figures 1, 5).

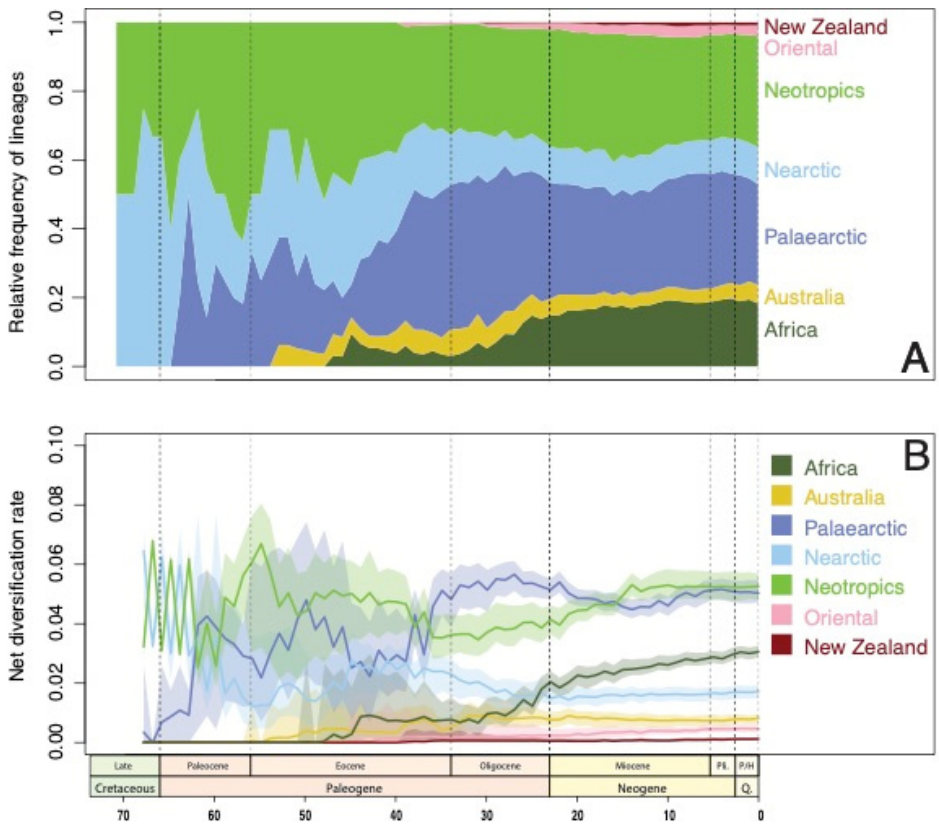


Figure 5: A) Relative frequency of lineages in each biogeographic region through time. B) Net diversification rate variation through time per biogeographic region. Shaded polygons are the distributions of 1000 mean net diversification rates in each biogeographic region. The bold line is the mean of that distribution.

Discussion

Given our calibration of the age of the crown group of Geometridae in the Maastrichtian of the Late Cretaceous, we find that early divergence of subfamily lineages (Figure 1) took place in the Paleocene and Eocene, much like in the butterflies (Heikkilä et al. 2012; Espeland et al. 2018). Many family lineages in Lepidoptera are inferred to have diverged from each other shortly before or after the KT event, while diversification within families has happened in the Cenozoic (Wahlberg et al. 2013). It is likely that the diversification of the major Lepidoptera lineages is connected to the diversification of their host plants, the angiosperms (Davis et al. 2005; Wahlberg et al. 2013).

Our diversification rate analyses inferred two periods of significant increase in the net rate of speciation around 30 and 15 Ma (Figure 2). The timing of these rate increases across the family agrees well with the clade-specific diversification shifts estimated by the BSD approach (Figure 3). These upward shifts in the rate of speciation are located in seven non-closely related clades (tribes or genera) within different families of Geometridae. They are dated around 40-30 mya (three lineages) and between 10 and 15 mya (four lineages; Figure 3). This concentration of independent events of increase in the rate of diversification at around 30 mya and 15 mya agrees well with the pattern found by Chazot et al. (2020). Their study on the diversification dynamics of Nymphalidae differs from ours in a better taxon sampling, especially for the more recent ages (i.e. the last 15 million years), and their use of a different method to evaluate branch-specific rates. However, it is noteable that Chazot et al. (2020) also inferred increases in the rate of diversification in several independent lineages at the end of the Eocene period and around the Mid Miocene.

The global cooling event at the end of the Eocene, known as the Terminal Eocene Event (TEE, Tiffney 1985; Sanmartín et al. 2001; Meseguer et al. 2015) or the Late Eocene-Early Oligocene cooling event (LEEOC, Crisp and Cook 2007; Zachos et al. 2008), dated between 35-32 Ma, is considered as one of the most influential climatic events of the Cenozoic. Temperatures dropped nearly 10º C worldwide – accompanied by a major drop in CO2 concentrations (Beerling et al. 2009) –, followed by major changes in vegetation and associated fauna (Tiffney 1985; Morley 2000; Sanmartín et al. 2001; Plana 2004). The event was caused, among others, by the opening of the Drake Passage between South America and Antarctica, and the onset of the Circum-Antarctic Current. In the Holarctic, the closing of the Turgai Strait between Europe and Asia (c. 30 Ma), brought about a more continental climate into the Western Palearctic (Sanmartín et al. 2001). The TEE has been associated to widespread extinction in flowering plant (angiosperm) families (Plana 2004; Pan et al. 2006; Crisp and Cook 2007; Antonelli and Sanmartín 2011). A warm-adapted boreotropical flora was replaced across the Holarctic continents, Eurasia and North America, by the temperate mixed-mesophytic forest (Tiffney

1985; Meseguer et al. 2015); in Africa, humid tropical forests were replaced by more xeric vegetation (Plana 2004; Pan et al. 2006). A long-term cooling trend, punctuated by warming events, started with the TEE, which saw the expansion of C4 grasses in temperate and subequatorial regions, and the rise to dominance of coniferous forests in the boreal Holarctic regions (Meseguer et al. 2015; Kergoat et al. 2018). These changes in vegetation composition were concomitant with changes in the associated fauna (Sanmartín et al. 2001; Kergoat et al. 2018), with expansion and adaptation of insect families to the new grassland biomes, boreal and tundra forests, etc., but also events of widespread extinction.

The second period of major change in diversification rates in Geometridae, at c. 15 mya (Figure 3), coincides with another event of rapid global change and biotic extinction. The Mid Miocene Climate Optimum (MMCO) at 17-14 Ma (Zachos et al. 2008) was a dramatic warming event, which saw increases in temperature close to 5º C worldwide (Bohaty and Zachos 2003; Steffen et al. 2018). In the Holarctic and temperate areas, the MMCO was accompanied by range expansion of dryland floras worldwide (Edwards et al. 2010; Spriggs et al. 2014), and the colonization of tropical mountain regions by temperate lineages (Meseguer et al. 2015). In tropical regions like Africa, warmer and drier climates led to appearance of xeric-adapted floras and the extinction of humid-adapted, subtropical lineages (Morley 2000; Plana 2004; Pan et al. 2006; Pokorny et al. 2015). This warming trend was brought about by regional and global tectonic events, such as the closing of the eastern arm of the Tethys Seaway that connected the Indian and Atlantic Oceans, around 15 mya; the rapid uplift of the Eastern African Plateau (~17 mya, Sepulchre et al. 2006) and the Tibetan Plateau (Yin and Harrison 2000); and the continental collision of the Australian and Eurasian plates, which led to the aridification of central Australia (~ 15 mya, Crisp and Cook 2007).

The expansion of grassland biomes after the MMCO has been argued as a diversification driver in herbivore lineages, such as mammals (Stebbins 1981; MacFadden 2005), beetles (Micó et al. 2009) and Satyrinae butterflies (Peña and Wahlberg 2008); in contrast, Kergoat et al. (2018) did not find evidence of synchronous diversification in stemborer moths. Chazot et al. (2020) and our results here support an increase in net diversification rates in Nymphalidae butterflies and Geometridae moths around the MMCO; yet, this shift was not general, affecting some clades within different subfamilies, and we did not test for a causal connection. In the family Nymphalidae, the pattern is most probably driven by Satyrinae which are mainly grass feeder specialists. On the other hand, members of Geometridae are not grass feeders; they are specialized in forest habitats. The appearance of grasslands, which is the most probable reason for the diversification shift in Nymphalidae, meant habitat fragmentation and appearance of gene flow barriers to Geometridae which probably could explain the observation of the same pattern in two different specialist groups.

Though events of rapid climate change like the TEE and MMCO were global, they probably did not affect equally to all biogeographic regions. This can be seen in our biogeographic reconstruction (Figure 1), which shows that certain regions became hubs of diversification and range expansion events (Figure 5a,b) around the TEE and the MMCO. Until approximately 40 mya, the Neotropics exhibits the highest diversification rate among regions followed by the Palaearctic, but this order is reverted after 45-35 mya, when the Palearctic shows the highest diversification rate until 20 mya (Figure 5a,b). This pattern of geographic-dependent diversification rates agrees remarkably well with the results obtained by Chazot et al. (2020) for Nymphalidae, which show an increase in the net diversification rate in the Palaearctic region between 40 and 20 mya. As these authors, we hypothesize that the observed increase around 40 mya responded to the Eocene-Oligocene transition (LEEOC), and the appearance of higher seasonality and colder climates, which favoured higher diversification rates for lineages adapting to these new conditions.

The Palearctic pattern is somewhat mirrored by the Nearctic (Figure 5a,b), though in our reconstruction the Nearctic is also part of the root ancestral range (Figure 1). Like the Palearctic, the Nearctic is a relatively well studied region, which probably played an important role in the divergence of many of the deepest lineages of Geometridae, as a bridge between the Neotropics and the Palaearctic (Figure 1; see also Brehm et al. 2019). Interestingly, though the Nearctic mirrors the Palearctic in the relative frequency of lineages (Figure 5a), the net diversification rate of this region over time decreases in the last 40 mya compared to other regions (Figure 5b). Chazot et al. (2020) also found a similar pattern in the family Nymphalidae, with a low relative diversity in the Nearctic region since the Eocene. The reason for this could be related to the differential effect of Pleistocene glaciations, which hit harder the Eastern Nearctic region, where most geometrid moth species are endemic, compared to the Palearctic region, especially the eastern (Asian) parts (Sanmartín et al. 2001).

Compared to Nymphalidae, the colonization of the Afrotropics by geometrid moths occurred later (c. 50 mya, Figure 1), and the region had a relatively low diversification rate during the first 20 mya compared to other regions (Figure 5a,b). At around 25 mya, at the Oligocene–Miocene boundary, the Afrotropical lineages of geometrids show a sharp increase in their net diversification rate (Figure 5b), probably associated to the expansion of grasslands and the retraction of the tropical mesic forests (Plana 2004; Paun et al. 2005).

A general increase in the diversification rate towards the present is seen in all biogeographic regions, excepting Australia (Figure 5b). In Nymphalidae the Australian region exhibits a relatively low diversification rate compared to other regions, which becomes the lowest towards the present (Chazot et al. 2020). In Geometridae (Figure 5a,b), the Australian region shows a constantly low relative diversity compared to the Afrotropics since its colonization in the Early-Mid Cenozoic (Figs. 1, 5a), and the net diversification rate slightly decreases in the last

20 mya (Figure 5b). However, we treated New Zealand as a different region from Australia, but it has been considered as part of the same region in other studies. Thus, it might be that the Australian region harbours a relatively higher diversity than represented in our study.

We need to consider here, however, that some of these patterns might be a consequence of the different level of sampling of these regions in our dataset. The Neotropics and the Palearctic regions have been historically better studied, and/or are better sampled in our study, then the Oriental region, Australia or the Afrotropics. Towards the present, those regions that are historically better studied and sampled will have comparatively a better representation of younger clades (i.e., genera, species); we therefore expect them to exhibit higher diversification rates towards the present compared to the other regions. Similarly, the fact that the net diversification rates in most regions stabilizes towards the present (Figure 5a,b) is most likely due to the effect of incomplete taxon sampling (i.e., missing taxa), which affects more younger clades than clades with deeper divergences (i.e., subfamilies, tribes).

Even taking these deficiencies into consideration, the role of the Neotropics as a region that has historically harboured a higher diversity than other regions is apparent in Geometridae. This agrees with the higher relative diversity observed in other Lepidoptera families (Vila et al. 2011; Chazot et al. 2020). However, unlike Nymphalidae, where the Neotropics shows a higher relative diversity only after the Eocene, in Geometridae this region had a high relative diversity since the crown diversification in the Late Cretaceous (Figs. 1, 5).

In general, our biogeographic analysis supports very few ancestral range states including more than one region. This is in line with our observation of extant species being endemic to a single region, probably due to their limited dispersal capacities. Most dispersal events are inferred to have occurred between geographically adjacent regions (Figure 4; Supplementary Material). Moreover, a large number of the inferred long-distance dispersal events (Supplementary Material) could be the result of under sampling in the Asian and Afrortropical regions. This is probably the case for the unexpected long-distance jumps between Australia and the Palaearctic regions (Figure 4). The Oriental region – which serves as a historical bridge between these two regions – is severely underrepresented in our study. On the other hand, the high frequency of dispersal events between the Neotropics and the Nearctic region, or between the Nearctic and Palaearctic regions, and between the Palaearctic and Afrotropical regions (which are very common in both directions) are in agreement with paleogeographic scenarios, supporting the presence of current or past connections among these regions, for example the trans-Atlantic and Beringian land bridges between both halves of the Holarctic (Sanmartín et al. 2001; Peña et al. 2010; Vila et al. 2011).

Conclusion

Using the most complete and up to date phylogeny of Geometridae, we studied for the first time their diversification patterns and biogeographical history. We reveal a high number of dispersal events between Afrotropic, Nearctic, Neotropic and Palaearctic regions which have shaped the actual dispersion ranges of the family. The Neotropical region has the highest recent diversity of the family. This appears to be a result of higher relative diversification rates and the early presence of the family in this region. The Palaearctic region also presents a relatively high diversification rate and similarly high diversity. The majority of the dispersal events are recorded between adjacent regions. Doubtlessly, a better and more even taxon sampling across all regions, in particular the African and Oriental region is desired in future studies.

Aknowledgement

HG received funding from the European Union's Horizon 2020 research and innovation program under the Marie Slodowska-Curie grant agreement No. 6422141. NW acknowledges funding from the Swedish Research Council (Grant No. 2015-04441). LM-R acknowledges funding from Colciencias, 756-2016 and Universidad de Sucre, Colombia. IS was funded by the Spanish government (MINECO/MICIU/AEI) and the European Regional Development Fund (FEDER) through projects CGL2015-67849-P and PGC2019-XXXX-P.

References

Aduse-Poku K., Brattström O., Kodandaramaiah U., Lees D.C., Brakefield P.M., Wahlberg N. 2015. Systematics and historical biogeography of the old world butterfly subtribe Mycalesina (Lepidoptera: Nymphalidae: Satyrinae). BMC Evol. Biol. 15:1–14.

Aduse-Poku K., Vingerhoedt E., Wahlberg N. 2009. Out-of-Africa again: A phylogenetic hypothesis of the genus Charaxes (Lepidoptera: Nymphalidae) based on five gene regions. Mol. Phylogenet. Evol. 53:463–478.

Antonelli A., Nylander J.A.A., Persson C., Sanmartín I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. Proc. Natl. Acad. Sci. 106:9749–9754.

Antonelli A., Sanmartín I. 2011. Why are there so many plant species in the Neotropics? Taxon. 60:403–414.

Beck J., McCain C.M., Axmacher J.C., Ashton L.A., Bärtschi F., Brehm G., Choi S.-W., Cizek O., Colwell R.K., Fiedler K., Francois C.L., Highland S., Holloway J.D., Intachat J., Kadlec T., Kitching R.L., Maunsell S.C., Merckx T., Nakamura A., Odell E., Sang W., Toko P.S., Zamecnik J., Zou Y., Novotny V. 2017. Elevational species richness gradients in a hyperdiverse insect taxon: a global meta-study on geometrid moths. Glob. Ecol. Biogeogr. 26:412–424.

Beer D., Beer A. 2019. phylotate: Phylogenies with Annotations. .

Bohaty S.M., Zachos J.C. 2003. Significant Southern Ocean warming event in the late middle Eocene. Geology. 31:1017.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:e1003537.

Brehm G. 2018. Revision of the genus Callipia Guenée, 1858 (Lepidoptera, Geometridae), with the description of 15 new taxa. Eur. J. Taxon. 2018:1–52.

Brehm G., Murillo-Ramos L., Sihvonen P., Hausmann A., Schmidt B.C., Õunap E., Moser A., Mörtter R., Bolt D., Bodner F., Lindt A., Parra L.E., Wahlberg N. 2019. New World geometrid moths (Lepidoptera: Geometridae): Molecular phylogeny, biogeography, taxonomic updates and description of 11 new tribes. Arthropod Syst. Phylogeny. 77:457–486.

Brown J.H. 2014. Why are there so many species in the tropics? J. Biogeogr. 41:8–22.

Buerki S., Forest F., Alvarez N., Nylander J.A.A., Arrigo N., Sanmartín I. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. J. Biogeogr. 38:531–550.

Buzas M.A., Collins L.S., Culver S.J. 2002. Latitudinal difference in biodiversity caused by higher tropical rate of increase. Proc. Natl. Acad. Sci. 99:7841–7843.

Chazot N., Condamine F.L., Dudas G., Peña C., Matos-Maraví P., Freitas A.V.L., Willmott K.R., Elias M., Warren A., Aduse-Poku K., Lohman D.J., Penz C.M., DeVries P., Kodandaramaiah U., Fric Z.F., Nylin S., Müller C., Wheat C., Kawahara A.Y., Silva-Brandão K.L., Lamas G., Zubek A., Ortiz-Acevedo E., Vila R., Vane-Wright R.I., Mullen S.P., Jiggins C.D., Slamova I., Wahlberg N. 2020. The latitudinal diversity gradient in brush-footed butterflies (Nymphalidae): conserved ancestral tropical niche but different continental histories. bioRxiv.:2020.04.16.045575.

Chazot N., Willmott K.R., Lamas G., Freitas A.V.L., Piron-Prunier F., Arias C.F., Mallet J., De-Silva D.L., Elias M. 2019. Renewed diversification following Miocene landscape turnover in a Neotropical butterfly radiation. Glob. Ecol. Biogeogr. 28:1118–1132.

Condamine F.L., Rolland J., Höhna S., Sperling F.A.H., Sanmartín I. 2018. Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies. Syst. Biol. 67:940–964.

Condamine F.L., Rolland J., Morlon H. 2013a. Macroevolutionary perspectives to environmental change. Ecol. Lett. 16:72–85.

Condamine F.L., Sperling F.A.H., Kergoat G.J. 2013b. Global biogeographical pattern of swallowtail diversification demonstrates alternative colonization routes in the Northern and Southern hemispheres. J. Biogeogr. 40:9–23.

Condamine F.L., Sperling F.A.H., Wahlberg N., Rasplus J.-Y., Kergoat G.J. 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. Ecol. Lett. 15:267–277.

Crisp M.D., Cook L.G. 2007. A congruent molecular signature of vicariance across multiple plant lineages. Mol. Phylogenet. Evol. 43:1106–1117.

Darwin C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: John Murray.

Davis C.C., Webb C.O., Wurdack K.J., Jaramillo C.A., Donoghue M.J. 2005. Explosive Radiation of Malpighiales Supports a Mid-Cretaceous Origin of Modern Tropical Rain Forests. Am. Nat. 165:E36–E65.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699–710.

Edwards E.J., Osborne C.P., Stromberg C.A.E., Smith S.A., Bond W.J., Christin P.A., Cousins A.B., Duvall M.R., Fox D.L., Freckleton R.P., Ghannoum O., Hartwell J., Huang Y., Janis C.M., Keeley J.E., Kellogg E.A., Knapp A.K., Leakey A.D.B., Nelson D.M., Saarela J.M., Sage R.F., Sala O.E., Salamin N., Still C.J., Tipple B. 2010. The Origins of C4 Grasslands: Integrating Evolutionary and Ecosystem Science. Science (80-.). 328:587–591.

Epskamp S., Cramer A.O.J., Waldorp L.J., Schmittmann V.D., Borsboom D. 2012. qgraph: Network Visualizations of Relationships in Psychometric Data. J. Stat. Software; Vol 1, Issue 4 .

Espeland M., Breinholt J., Willmott K.R., Warren A.D., Vila R., Toussaint E.F.A., Maunsell S.C., Aduse-Poku K., Talavera G., Eastwood R., Jarzyna M.A., Guralnick R., Lohman D.J., Pierce N.E., Kawahara A.Y. 2018. A Comprehensive and Dated Phylogenomic Analysis of Butterflies. Curr. Biol. 28:770-778.e5.

Fischer A.G. 1960. LATITUDINAL VARIATIONS IN ORGANIC DIVERSITY. Evolution (N. Y). 14:64–81.

Fischer T.C., Michalski A., Hausmann A. 2019. Geometrid caterpillar in Eocene Baltic amber (Lepidoptera, Geometridae). Sci. Rep. 9:2–5.
Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 31:3718–3720.
Heibl C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. .
Heikkilä M., Kaila L., Mutanen M., Peña C., Wahlberg N. 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proc. R. Soc. B Biol. Sci. 279:1093–1099.
Hey J. 1992. Using Phylogenetic Trees to Study Speciation and Extinction. Source Evol. Evol.
Höhna S., Freyman W.A., Nolen Z., Huelsenbeck J., May M.R., Moore B.R. 2019. A Bayesian Approach for Estimating Branch-Specific Speciation and Extinction Rates. bioRxiv.:555805.
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. Syst. Biol. 65:726–736.
Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. Mol. Biol. Evol. 28:2577–2589.
Holloway J.D. 1993. The moths of Borneo : Pt.11.: Family Geometridae, subfamily Ennominae. Kuala Lumpur: Malaysian Nature Society.
Holloway J.D. 1996. The moths of Borneo : Pt.9.: Family Geometridae, subfamilies Oenochrominae, Desmobathrinae and Geometrinae. Kuala Lumpur: Malaysian Nature Society.
Holloway J.D. 1997. The moths of Borneo : Pt.10.: Family Geometridae, subfamilies Sterrhinae and Larentiinae. Kuala Lumpur: Malayan Nature Society.
Hooker J.D. 1854. Himalayan journals; or, Notes of a naturalist in Bengal, the Sikkim and Nepal Himalayas, the Khasia Mountains, &c. London,: J. Murray,.
Jablonski D., Roy K., Valentine J.W. 2006. Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient. Science (80-.). 314:102–106.
Kalyaanamoorthy S., Minh B.Q., Wong T.K., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods.
Kawahara A.Y., Plotkin D., Espeland M., Meusemann K., Toussaint E.F.A., Donath A., Gimnich F., Frandsen P.B., Zwick A., dos Reis M., Barber J.R., Peters R.S., Liu S., Zhou X., Mayer C., Podsiadlowski L., Storer C., Yack J.E., Misof B., Breinholt J.W. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc. Natl. Acad. Sci. 116:22657–22663.
Kergoat G.J., Condamine F.L., Toussaint E.F., Capdevielle-Dulac C., Clamens A.-L., Barbut J., Goldstein P.Z., Le Ru B. 2018. Opposite macroevolutionary responses to environmental changes in grasses and insects during the Neogene grassland expansion. Nat. Commun. 9:5089.
Kinlock N.L., Prowant L., Herstoff E.M., Foley C.M., Akin-Fajiye M., Bender N., Umarani M., Ryu H.Y., Şen B., Gurevitch J. 2017. Explaining global variation in the latitudinal diversity gradient: Meta-analysis confirms known patterns and uncovers new ones. Glob. Ecol. Biogeogr. 27:125–141.

Kodandaramaiah U., Braby M.F., Grund R., Müller C.J., Wahlberg N. 2018. Phylogenetic relationships, biogeography and diversification of Coenonymphina butterflies (Nymphalidae: Satyrinae): intercontinental dispersal of a southern Gondwanan group? Syst. Entomol. 43:798–809.

Kodandaramaiah U., Wahlberg N. 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus Junonia (Nymphalidae: Nymphalinae). J. Evol. Biol.

Kodandaramaiah U., Wahlberg N. 2009. Phylogeny and biogeography of Coenonympha butterflies (Nymphalidae: Satyrinae) - Patterns of colonization in the Holarctic. Syst. Entomol. 34:315–323.

Linder H.P. 2001. Plant diversity and endemism in sub-Saharan tropical Africa. J. Biogeogr. 28:169–182.

MacFadden B.J. 2005. EVOLUTION: Fossil Horses--Evidence for Evolution. Science (80-.). 307:1728–1730.

Matos-Maraví P., Núñez Águila R., Peña C., Miller J.Y., Sourakov A., Wahlberg N. 2014. Causes of endemic radiation in the Caribbean: Evidence from the historical biogeography and diversification of the butterfly genus Calisto (Nymphalidae: Satyrinae: Satyrini). BMC Evol. Biol. 14.

Matos-Maraví P.F., Peña C., Willmott K.R., Freitas A.V.L., Wahlberg N. 2013. Systematics and evolutionary history of butterflies in the “Taygetis clade” (Nymphalidae: Satyrinae: Euptychiina): Towards a better understanding of Neotropical biogeography. Mol. Phylogenet. Evol. 66:54–68.

May M.R., Höhna S., Moore B.R. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods Ecol. Evol. 7:947–959.

Meseguer A.S., Lobo J.M., Ree R., Beerling D.J., Sanmartín I. 2015. Integrating Fossils, Phylogenies, and Niche Models into Biogeography to Reveal Ancient Evolutionary History: The Case of Hypericum (Hypericaceae). Syst. Biol. 64:215–232.

Meyer A.L.S., Wiens J.J. 2018. Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. Evolution (N. Y). 72:39–53.

Micó E., Sanmartín I., Galante E. 2009. Mediterranean diversification of the grass-feeding Anisopliina beetles (Scarabaeidae, Rutelinae, Anomalini) as inferred by bootstrap-averaged dispersal-vicariance analysis. J. Biogeogr. 36:546–560.

Morley R.J. 2000. Origin and evolution of tropical rain forests. Chichester: John Wiley & Sons.

Morlon H., Parsons T.L., Plotkin J.B. 2011a. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. 108:16327–16332.

Morlon H., Schwilk D.W., Bryant J.A., Marquet P.A., Rebelo A.G., Tauss C., Bohannan B.J.M., Green J.L. 2011b. Spatial patterns of phylogenetic diversity. Ecol. Lett. 14:141–149.

Müller C.J., Matos-Maraví P.F., Beheregaray L.B. 2013. Delving into Delias Hübner (Lepidoptera: Pieridae): Fine-scale biogeography, phylogenetics and systematics of the world’s largest butterfly genus. J. Biogeogr. 40:881–893.

Müller C.J., Wahlberg N., Beheregaray L.B. 2010. “After Africa”: The evolutionary history and systematics of the genus Charaxes Ochsenheimer (Lepidoptera: Nymphalidae) in the Indo-Pacific region. Biol. J. Linn. Soc. 100:457–481.

Murillo-Ramos L., Brehm G., Sihvonen P., Hausmann A., Holm S., Ghanavi H.R., Õunap E., Truuverk A., Staude H., Friedrich E., Tammaru T., Wahlberg N. 2019. A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies. .

Nee S., May M.R., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 344:305–311.

Nee S., Mooers A.O., Harvey P.H. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci.

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. Mol. Biol. Evol. 32:268–274.

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., Davis D.R., De Prins J., De Prins W., Epstein M.E., Gentili-Poole P., Gielis C., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S. (J C., Kozlov M. V, Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-T., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-C., Solis M.A., Tarmann G.M., Warren A.D., Weller S., Yakovlev R. V, Zolotuhin V. V, Andreas Z. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 1758:212–221.

Pan A.D., Jacons B.F., Dransfield J., Baker W.J. 2006. The fossil history of palms (Arecaceae) in Africa and new records from the Late Oligocene (28–27 Mya) of north-western Ethiopia. Bot. J. Linn. Soc. 151:69–81.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Paun O., Lehnebach C., Johansson J.T., Lockhart P., Hörandl E. 2005. Phylogenetic relationships and biogeography of Ranunculus and allied genera (Ranunculaceae) in the Mediterranean region and in the European Alpine System. Taxon. 54:911–932.

Payne J.L., Clapham M.E. 2012. End-permian mass extinction in the oceans: An ancient analog for the twenty-first century? Annu. Rev. Earth Planet. Sci. 40:89–111.

Peña C., Nylin S., Freitas A.V.L., Wahlberg N. 2010. Biogeographic history of the butterfly subtribe Euptychiina (Lepidoptera, Nymphalidae, Satyrinae). Zool. Scr. 39:243–258.

Peña C., Wahlberg N. 2008. Prehistorical climate change increased diversification of a group of butterflies. Biol. Lett. 4:274–278.

Plana V. 2004. Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 359:1585–1594.

Pokorny L., Riina R., Mairal M., Meseguer A.S., Culshaw V., Cendoya J., Serrano M., Carbajal R., Ortiz S., Heuertz M., Sanmartín I. 2015. Living on the edge: Timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Front. Genet. 6:1–15.

Purvis A. 2008. Phylogenetic Approaches to the Study of Extinction. Annu. Rev. Ecol. Evol. Syst. 39:301–319.

R Core Team. 2017. R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria.

Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLoS One. 9:e89543.
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67:901–904.
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. J. Biogeogr.
Regier J.C., Mitter C., Zwick A., Bazinet A.L., Cummings M.P., Kawahara A.Y., Sohn J.-C.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.
Regier J.C., Zwick A., Cummings M.P., Kawahara A.Y., Cho S., Weller S., Roe A., Baixeras J., Brown J.W., Parr C., Davis D.R., Epstein M., Hallwachs W., Hausmann A., Janzen D.H., Kitching I.J., Solis M.A., Yen S.-H., Bazinet A.L., Mitter C. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. BMC Evol. Biol. 9:280.
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). J. Biogeogr. 43:1254–1265.
Sanmartín I., Enghoff H., Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73:345–390.
Sanmartín I., Meseguer A.S. 2016. Extinction in phylogenetics and biogeography: From timetrees to patterns of biotic assemblage. Front. Genet. 7:1–17.
Sepulchre P., Ramstein G., Fluteau F., Schuster M., Tiercelin J.-J., Brunet M. 2006. Tectonic Uplift and Eastern Africa Aridification. Science (80-.). 313:1419–1423.
Sihvonen P., Murillo-Ramos L., Brehm G., Staude H., Wahlberg N. 2020. Molecular phylogeny of Sterrhinae moths (Lepidoptera: Geometridae): towards a global classification. Syst. Entomol. 45:606–634.
Simpson G.G. 1944. Tempo and Mode of Evolution. Columbia Classic in Evolution.
Spriggs E.L., Christin P.-A., Edwards E.J. 2014. C4 Photosynthesis Promoted Species Diversification during the Miocene Grassland Expansion. PLoS One. 9:e97722.
Stadler T. 2011a. Inferring speciation and extinction processes from extant species data. Proc. Natl. Acad. Sci. 108:16145–16146.
Stadler T. 2011b. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. 108:6187–6192.
Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.
Stadler T. 2015. TreePar: Estimating birth and death rates based on phylogenies. .
Stanley S.M. 1980. Macroevolution, Pattern and process. W H Freeman & Co.
Stebbins G.L. 1981. Coevolution of Grasses and Herbivores. Ann. Missouri Bot. Gard. 68:75.
Steffen W., Rockström J., Richardson K., Lenton T.M., Folke C., Liverman D., Summerhayes C.P., Barnosky A.D., Cornell S.E., Crucifix M., Donges J.F., Fetzer I., Lade S.J., Scheffer M., Winkelmann R., Schellnhuber H.J. 2018. Trajectories of the Earth System in the Anthropocene. Proc. Natl. Acad. Sci. 115:8252–8259.
Tiffney B.H. 1985. The eocene north atlantic land bridge: its importance in tertiary and modern phytogeography of the northern hemisphere. J. Arnold Arbor. 66:243–273.

Toussaint E.F.A., Balke M. 2016. Historical biogeography of Polyura butterflies in the oriental Palaeotropics: trans-archipelagic routes and South Pacific island hopping. J. Biogeogr. 43:1560–1572.
Toussaint E.F.A., Müller C.J., Morinière J., Tänzler R., Balke M. 2020. A glide over the Indo-Australian geological maze: repeated transgressions of Lydekker's and Wallace's Lines in archdukes, barons and dukes (Nymphalidae: Limenitidinae: Adoliadini). Biol. J. Linn. Soc. 129:810–821.
Toussaint E.F.A., Vila R., Yago M., Chiba H., Warren A.D., Aduse-Poku K., Storer C., Dexter K.M., Maruyama K., Lohman D.J., Kawahara A.Y. 2019. Out of the Orient: Post-Tethyan transoceanic and trans-Arabian routes fostered the spread of Baorini skippers in the Afrotropics. Syst. Entomol. 44:926–938.
Tujuba T.F., Hausmann A., Sciarretta A. 2020. Revision of the Orbamia Herbulot, 1966 group of genera with description of two new genera, ten new species, and two new subspecies (Lepidoptera, Geometridae, Ennominae, Cassymini). Zookeys. 929:53–77.
Vila R., Bell C.D., Macniven R., Goldman-Huertas B., Ree R.H., Marshall C.R., Bálint Z., Johnson K., Benyamini D., Pierce N.E. 2011. Phylogeny and palaeoecology of polyommatus blue butterflies show beringia was a climate-regulated gateway to the new world. Proc. R. Soc. B Biol. Sci. 278:2737–2744.
Wahlberg N., Freitas A.V.L. 2007. Colonization of and radiation in South America by butterflies in the subtribe Phyciodina (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. 44:1257–1272.
Wahlberg N., Wheat C.W., Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One. 8:1–8.
Wallace A.R. 1869. The Malay Archipelago: The Land of the Orang-Utan and the Bird of Paradise. A Narrative of Travel, with Studies of Man and Nature. New York: Harper & Brothers.
Wallace A.R. 1876. The Geographical Distribution Of Animlas. with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface. New York: .
Yin A., Harrison T.M. 2000. Geologic evolution of the Himalayan-Tibetan orogen. Annu. Rev. Earth Planet. Sci. 28:211–280.
Zachos J.C., Dickens G.R., Zeebe R.E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature. 451:279–283.

Paper III

Phylogenomics of Erebidae (Lepidoptera): using old DNA extracts to resolve old phylogenetic questions with whole genome sequencing

Hamid Reza Ghanavi[1], Victoria Twort[1,2], Reza Zahiri[3] and Niklas Wahlberg[1]

[1] Systematic Biology Group, Biology department, Lund University, Lund, Sweden.

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

[3] Centrum für Naturkunde, University of Hamburg, Martin-Luther-King Platz 3, 20146 Hamburg, Germany.

Abstract

The appearance of high throughput sequencing (HTS) is a major advance in numerous fields within life sciences, comparable to the discovery of polymerase chain reaction (PCR). Although Evolutionary Biology and Phylogenetics are within the fields which are adapting to a more common use of such approaches, the full potential of these methods is still to be explored. One of the limiting factors in the more common use of such methods is the lack of expertise in designing studies using such approaches and lack of expertise in the bioinformatic skills needed to deal with genome scale data sizes. Even if the sequencing costs have dropped continuously and radically during the last years, still for many phylogenetic studies based on a handful of genes, the traditional Sanger sequencing methods are cheaper. Therefore, it is important to explore the type of questions which are best investigated using HTS approaches and where these approaches are cheaper and less labour intensive. In addition, a general belief exists about the need for high quality and high quantity of DNA to be used in HTS approaches, which limits their use in many cases. Here we explore the utility of these methods to study the complex phylogenetic relationships within Erebidae moths due to a potential rapid radiation. We use >10 year old genomic DNA extracts, used previously for Sanger sequencing studies, to prepare libraries and obtain whole genomes. We apply our approach to 47 samples

of Erebidae moths in order to sequence their whole genome and extract a protein coding gene set, and to study the deep relationships within the family. Using this approach, we obtained a well-resolved deep phylogenetic hypothesis for the family and demonstrated the utility of these low quality and often forgotten genetic resources.

Key words: Erebidae, phylogenomics, whole genome, old DNA extract, Lepidoptera

Introduction

Lepidoptera (commonly known as moths and butterflies), together with Coleoptera, Hymenoptera and Diptera, present exceptional high numbers of species which have appeared during the last 200 million years (Misof et al. 2014). With close to 160,000 described species (van Nieukerken et al. 2011) they have colonised every continent apart from Antarctica and are adapted to a wide diversity of terrestrial habitats. Within the order Lepidoptera, Erebidae is the most diverse family with over 27,000 described species around the globe (van Nieukerken et al. 2011). The phylogenetic relationships within this mega diverse family of moths are not fully understood. This is especially true for the deep relationships between different subfamilies (Zahiri et al. 2012). One of the big challenges in the study of deep nodes within Erebidae is the probable rapid radiation that they experienced in the early divergences of the family. This rapid radiation is translated today into very short internal branches.

To resolve phylogenetic questions involving rapid radiations in the past, the challenge is to find markers which have the correct phylogenetic resolution. If a marker has a relatively high mutation rate, it accumulates too many changes and becomes saturated towards the deeper parts of the tree. On the other hand, if a marker is slowly evolving, it needs longer periods of time to accumulate enough changes to be informative. In short, the challenge is to find the right set of markers, evolving slowly enough to not get saturated resolving deep relationships, and fast enough to accumulate enough changes in the short period of time corresponding to the rapid radiation events. We use a gene set used in other phylogenetic studies on Lepidoptera, which have been shown to be useful for resolving deep phylogenetic relationships (Twort et al. 2020), and investigate their utility for family-level studies. We adopted a High Throughput Sequencing (HTS) approach for obtaining these markers.

With the advances in genomic methods, HTS approaches become more accessible to use. One of the earliest limitations to use such methods was its inhibiting price. But especially in recent years the price of sequencing genomes has had such a

significant drop that it has become cheaper than the traditional Sanger sequencing (cost per bp of sequenced data). Another limitation to use such methods has been the low efficiency of molecular techniques to prepare libraries using low amounts or poor-quality DNA. This has changed greatly due to the development and standardisation of methods to deal with historical DNA (Sproul and Maddison 2017; Li et al. 2019; Allio et al. 2020; Call 2020; Lopez et al. 2020; Twort et al. 2020).

On the other hand, the development of computer technologies and computing power, facilitate the use of such big datasets. In addition to advances in the hardware, also the algorithms and software dealing with these datasets is evolving on a daily basis. These methodological advances helped the rise of phylogenomics. Different approaches have been used in arthropod phylogenomics. Maybe one of the most popular approaches among entomologists these days is the target enrichment (TE) method (Blaimer et al. 2016; Young et al. 2016; Espeland et al. 2018; Godwin et al. 2018; St Laurent et al. 2018; Kieran et al. 2019). In this method, small probes are designed to preferentially attach to conserved parts of the genome and this way only a smaller proportion of the genome is sequenced. This is why this is one of the so-called genome reduction methods. In general, these genome reduction methods need a high amount of extracted DNA. Methods not reducing the input DNA, such as shotgun whole genome sequencing, are theoretically less affected by the lower quality and quantity of the input DNA.

In this study we opted for an unconventional source of DNA to prepare our HTS libraries. Here we explored the possibility of using low concentration and old genomic DNA extracts from over 10 years ago to create our libraries for whole genome sequencing. In this study, we selected 47 species covering all major subfamilies to: i) resolve the phylogenetic deep relationships within Erebidae ii) evaluate the possibility of using low-quality/low-concentration DNA extracts to sequence whole genomes as they are widely available in numerous research groups around the world.

Material and Methods

Taxon sampling and DNA quality control

We based our study design and taxon sampling on the most complete phylogenetic hypothesis available by Zahiri et al. (2012). Using the relationships published in that study, we selected 47 species (Table 1) to cover a maximum number of deep nodes and therefore be able to focus on the internal short branches. The same DNA extracts that were used in Zahiri et al. (2012) were used in this study. These have been stored at +4°C for 10-12 years in their original tubes. We added up to 50 ul MilliQ water to microtubes where the water content had evaporated. We ran 5 ul of the extract in

a 2% agarose gel to roughly evaluate the size range and concentration of the DNA in the samples.

Table 1: The list of Erebidae samples used in whole genome sequencing. # refers to the number of raw reads in millions of reads.

ID	Species	Subfamily	COUNTRY	#
RZ44	Asota heliconia	Aganainae	HONG KONG	39,9
RZ268	Mecodina praecipua	Aganainae	HONG KONG	26,4
RZ332	Anoba anguliplaga	Anobinae	GHANA	42,2
RZ103	Rema costimacula	Anobinae	HONG KONG	22,1
RZ404	Amerila astreus	Arctiinae	MALAYSIA	45,3
RZ28	Brunia antica	Arctiinae	HONG KONG	77
RZ30	Creatonotos transiens	Arctiinae	HONG KONG	30,4
RZ8	Syntomis phegea	Arctiinae	HUNGARY	21,7
RZ3	Laspeyria flexula	Boletobiinae	HUNGARY	54,2
RZ41	Metaemene atrigutta	Boletobiinae	HONG KONG	16,9
RZ104	Saroba pustulifera	Boletobiinae	HONG KONG	20,7
RZ336	Calyptra hokkaida	Calpinae	JAPAN	34,5
RZ337	Oraesia excavata	Calpinae	HONG KONG	38,2
RZ56	Phyllodes eyndhovii	Calpinae	TAIWAN	63,7
RZ248	Acantholipes circumdata	Erebinae	UAE	27,8
RZ11	Erebus ephesperis	Erebinae	TAIWAN	105,9
RZ39	Ericeia subcinerea	Erebinae	HONG KONG	80
RZ149	Hypopyra capensis	Erebinae	GHANA	52,6
RZ58	Melipotis jucunda	Erebinae	USA	57,1
RZ21	Ophiusa coronata	Erebinae	MALAYSIA	42,3
RZ48	Sympis rufibasis	Erebinae	HONG KONG	52,4
RZ313	Sypnoides fumosa	Erebinae	JAPAN	86,7
RZ22	Azeta ceramina	Eulepidotinae	COSTA RICA	55,2
RZ59	Panopoda rufimargo	Eulepidotinae	USA	41,6
RZ271	Idia aemula	Herminiinae	USA	52,3
RZ180	Nodaria verticalis	Herminiinae	GHANA	38,4
RZ367	Hypena baltimoralis	Hypeninae	USA	34,9
RZ42	Luceria striata	Hypenodinae	HONG KONG	27,4
RZ138	Micronoctua sp.	Hypenodinae	INDONESIA	106,7

RZ105	Hypocala deflorata	Hypocalinae	HONG KONG	47,9
RZ89	Arctornis sp.	Lymantriinae	JAPAN	33,1
RZ34	Nygmia plana	Lymantriinae	HONG KONG	19
RZ18	Masca abactalis	Pangraptinae	INDONESIA	44,8
RZ40	Pangrapta bicornuta	Pangraptinae	HONG KONG	63,2
RZ94	Alesua etialis	Rivulinae	COSTA RICA	16
RZ159	Rivula ochrea	Rivulinae	GHANA	59,4
RZ9	Scolecocampa liburna	Scolecocampinae	USA	52
RZ13	Gonitis involuta	Scoliopteryginae	TANZANIA	17,5
MM00407	Scoliopteryx libatrix	Scoliopteryginae	FINLAND	38,4
RZ111	Platyjionia mediorufa	Tinoliinae	HONG KONG	25,5
RZ389	Tamsia hieroglyphica	Tinoliinae	MALAYSIA	26,4
RZ331	Tinolius eburneigutta	Tinoliinae	THAILAND	33,4
RZ57	Lygephila maxima	Toxocampinae	JAPAN	41,4
RZ4	Colobochyla salicalis	Unassigned	HUNGARY	44,3
RZ93	Epitausa dilina	Unassigned	COSTA RICA	40,8
RZ265	Rhesala imparata	Unassigned	HONG KONG	37,9
RZ119	Schistorhynx argentistriga	Unassigned	HONG KONG	55,9

Library preparation

In the cases where high molecular weight DNA was observed, the DNA was sonicated to approximatively 200 – 300 bp fragments using a Bioryptor® with the following settings: (M) medium power output, 30 sec ON/ 90 sec OFF pulses for 45 minutes in a 4oC water bath, followed by vacuum centrifugation and resuspension in 50 µl of MilliQ water. Libraries were prepared for sequencing using a modified protocol of Meyer and Kircher (2010). Full details of the protocol are given in Twort et al. (2020). This protocol is derived from the protocol created by Rohland and Reich (2012). Briefly, DNA was blunt-end repaired with T4 Polynucleotide Kinase (BioLabs), followed by a reaction clean up with the MinElute purification kit (Qiagen). 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 to ten independent reactions to avoid amplification bias, with up to 15 cycles being used for each reaction. Indexing PCR reactions for each library were pooled together prior to the final magnetic bead clean up with Sera-Mag™ SpeedBeads™ (Thermo Fisher Scientific). An initial bead concentration of 0.5X was used to remove long fragments. Then libraries were selected with a bead concentration of 1.8X to size select the expected library range of ~300 bp. The

resulting libraries were quantified and quality checked with Quanti-iTTM PicoGreenTM dsDNA assay and with a DNA chip on a Bioanalyzer 2100, respectively.

Sequencing and Genome Assembly

Based on the observed library size and its measured nucleic acid concentration on PicoGreen, the final concentration of libraries was calculated. Using the final concentration of the libraries, they have been uniformly pooled together. The pooled library was sent to the Swedish National Genomic Institute (NGI) to be sequenced on a single lane of Illumina NovaSeq 6000 (PE 150bp). At this stage, the raw sequences for five genomes within Noctuoidea (Table 2) were downloaded from GenBank and added to our data. Raw reads were quality checked with FASTQC v0.11.8 (Andrews 2010). Reads containing ambiguous bases were removed from the dataset using Prinseq 0.20.4 (Schmieder and Edwards 2011). Reads were cleaned to remove low quality bases from the beginning (LEADING: 3) and end (TRAILING: 3), reads less than 30 bp in length, and evaluation of read quality with a sliding window approach was done with Trimmomatic 0.38 (Bolger et al. 2014). Quality was measured for sliding windows of 4 bp and had to be greater than PHRED 25 on average. The cleaned reads were then de novo assembled with spAdes 3.13.0 (Nurk et al. 2013) using kmer values of 21, 55 and 77.

Table 2: The list of short reads downloaded from GenBank.

Species	Family	GenBank
Busseola fusca	Noctuidae	ASM784487
Mamestra configurata	Noctuidae	ASM219265
Agrotis ipsilon	Noctuidae	ASM419385
Spodoptera litura	Noctuidae	ASM270686
Hyphantria cunea	Erebidae	ASM370950

Dataset generation

The orthologous markers were extracted from the assembled genomes using the MESPA (Neethiraj et al. 2017) pipeline. This method uses a set of reference markers to identify and pull them out from assembled genomes. The reference used in this study is a manually curated gene set with a total of 330 genes (11 mitochondrial and 319 nuclear) which has proven to be phylogenetically informative at deeper nodes of the Lepidoptera order (Twort et al. 2020). The MESPA pipeline output both

nucleotide and amino acid sequence of the references used. Sometimes a single sequence of a marker is divided into various pieces, usually related to the quality of the genome assembly. The amino acid sequences were aligned to reference sequences using MAFFT v7.450 (Katoh and Standley 2013) using the default options. Both amino acid alignments and nucleotide sequences were screened in Geneious 10.2.6 (Kearse et al. 2012) to correct for possible reading frame shifts, pseudogenes, mis-alignments or any other methodological errors. Then the amino acid alignments were reverse translated using the nucleotide sequences in Pal2Nal v14 (Suyama et al. 2006) perl code. The obtained aligned nucleotide sequences were uploaded to the VoSeq (Peña and Malm 2012) database. In addition, the data from 24 Noctuoidea transcriptomes (Table 3) of interest available from online databases, were retrieved and added to VoSeq. We thus included all the publicly available Erebidae genomes (6 in total).

Table 3: The list of available genomic information retrieved from public online repositories.

Species	Family	Reference
Arctia plantaginis	Erebidae	ERR1856313
Eudocima salaminia	Erebidae	SRX553954
Euproctis chrysorrhoea	Erebidae	SRR1040496
Lymantria dispar	Erebidae	SRX371346
Lymantria monacha	Erebidae	SRR1055268
Oraesia emarginata	Erebidae	SRR5128005
Anigraea rubida	Euteliidae	SRR1299755
Agrotis segetum	Noctuidae	SRR1231960
Athetis lepigone	Noctuidae	SRR796575
Chrysodeixis includens	Noctuidae	SRR2049082
Helicoverpa armigera	Noctuidae	SRR1565435
Helicoverpa zea	Noctuidae	SRX371342
Heliothis subflexa	Noctuidae	ERR738599
Heliothis virescens	Noctuidae	SRX371341
Mythimna separata	Noctuidae	SRR5115697
Sesamia inferens	Noctuidae	SRR867201
Sesamia nonagrioides	Noctuidae	ERR424922

Spodoptera exigua	Noctuidae	SRR525279
Spodoptera frugiperda	Noctuidae	SRR3406055
Striacosta albicosta	Noctuidae	SRX017236
Trichoplusia ni	Noctuidae	SRR544891
Manoba major	Nolidae	SRR1300145
Notoplusia minuta	Notodontidae	SRR1299746
Thaumetopoea pityocampa	Notodontidae	SRR1284701

Phylogenetic analyses

Both nucleotide and amino acid datasets (partitioned by gene) were used to infer Maximum Likelihood (ML) based phylogenetic tree. The ML trees were obtained using IQ-TREE 2.0.6 (Nguyen et al. 2015). In both analyses the best substitution model and partitioning scheme was determined using ModelFinder (Kalyaanamoorthy et al. 2017) with "-m MFP+MERGE" option. We evaluated the node supports with ultrafast bootstrap approximations (UFBoot2) and SH-like approximate likelihood ratio test (Guindon et al. 2010; Hoang et al. 2018) using the "-B 1000 -alrt 1000" option. We used the "-bnni" option to reduce the risk of overestimating branch supports in ultrafast bootstrap approximation analysis. The use of genomic scale datasets is known to increase the bootstrap-based support values in phylogenetic analyses. In order to further evaluate the support values of the nodes, the gene concordance factor (gCF) and the site concordance factor (sCF) were calculated for each node as implemented in IQ-Tree2 (Minh et al. 2020) for the nucleotide dataset (Figure 1). In addition, this method calculates a discordance factor value which is useful in the interpretation of the concordance factor results. The gDF (or sDF) measures the proportion of genes (or sites) which support a topology with one Nearest Neighbor Interchange (NNI) distance away. A variation of gCF was also calculated where, in place of gene trees, partition trees were used (here we call it partition concordance factor, pCF). The same way gCF and sCF measure the proportion of genes and sites, respectively, which support a split in the tree, the pCF measure the proportion of the partitions supporting a split. The resulting trees were visualized and rooted in FigTree v1.4.3 (Rambaut 2016) using the outgroups. The rooting point were chosen as the branch leading to the Notodontidae samples based on other studies (Regier et al. 2009, 2013; Zahiri et al. 2011).

Most of the bioinformatic pipelines were run using the resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2018-8-347. All data in the supplementary

material, the assembled genomes, the alignments and the results can be found and downloaded from the GitHub repository: github.com/Hamidhrg/Erebidae.

Results

Genome assembly

We successfully prepared libraries and whole genome sequenced 47 old genomic DNA extracts from all the major subfamilies of Erebidae. The mean number of obtained raw reads per library was 45 Million reads ranging from 16 to 107 (Table 1). The assembled genomes had a mean N50 of 1,568 bp ranging from 284 to 4,579 and a mean N90 of 394 bp, ranging from 58 to 1,015. The GC content of the assembled genomes varied between 33.5% and 40.3% with a mean of 36%. The average contig length was between 159 bp to 1,931 bp (mean of 790 bp). The total assembly size varied between 202 Mbp and 1,622 Mbp (Suplementary Table 1). Using the MESPA pipeline, we successfully recovered and aligned the 330 gene set. After manually checking and filtering, the final dataset consisted of 76 taxa and 208 genes (11 mitochondrial and 197 nuclear) with a total length of 159,546 sites.

Phylogenetic analyses

The result of ModelFinder merged the 208 possible gene-partitions of the nucleotide dataset into 44 partitions and into 45 partitions for the amino acid dataset. The best partitioning scheme and substitution models is shown in the Supplementary materials. SH-like and UFBoot2 support values performed similarly in the majority of the cases. As mentioned in IQ-TREE2 manual, values of SH-like $\geq$ 80 and UFBoot2 $\geq$ 90 are considered as well-supported clades. The gCF and sCF values are however more complex to interpret. In general, both concordance factor values were recovered as relatively low in this analysis. The pCF value was generally higher than the other two concordance factor values.

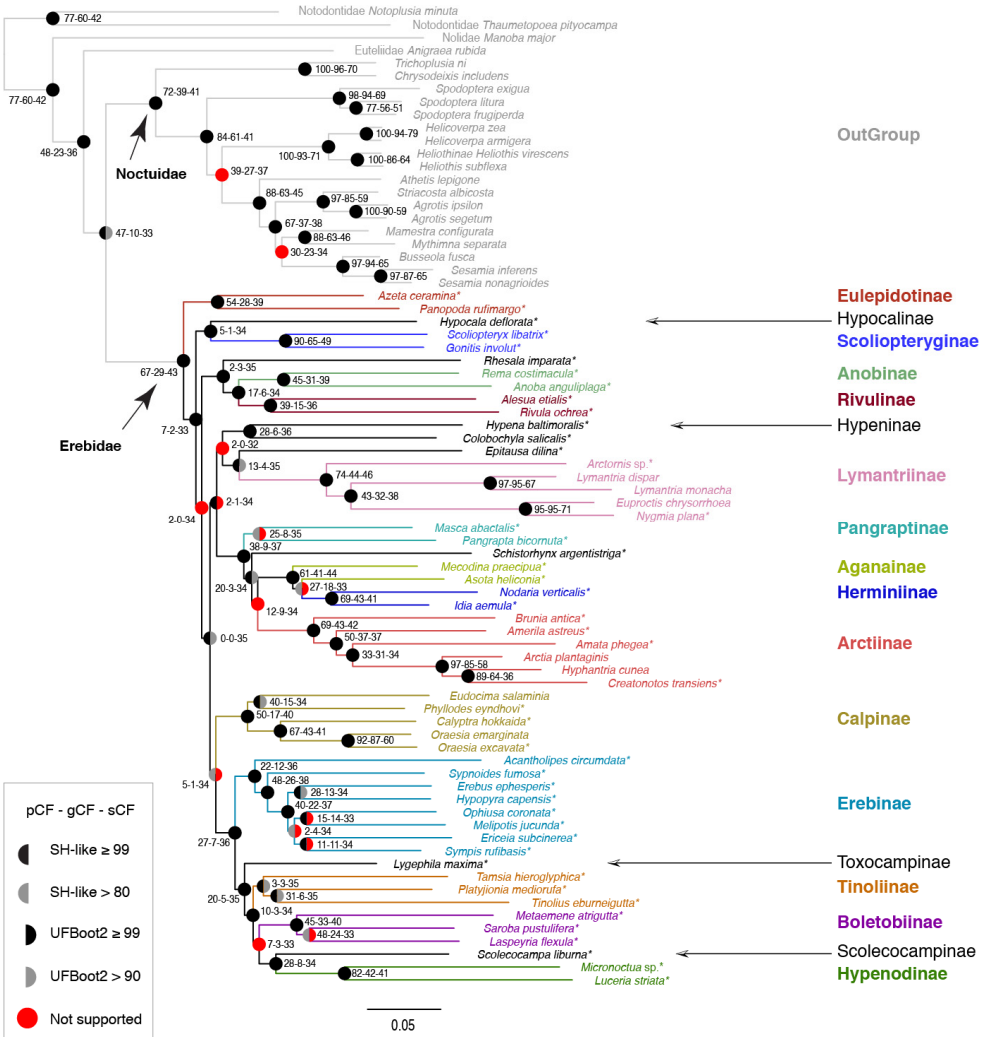


Figure 1: The ML tree of Erebidae. Terminal taxa are coloured based on the subfamilies they belong to. Species names marked by a star* are sequenced in this study. The pCF, gCF and sCF values are placed by the corresponding nodes. Species *Rhesalia imparata*, *Colobochyla salicalis*, *Epitausa dilina* and *Schistorhynx argentistriga* are not placed into any subfamily.

Discussion

Old DNA extract phylogenomics

In this study we successfully managed to sequence whole genomes of Erebidae moth samples using >10 year old genomic extracts. The DNA solutions were stored in a normal refrigerator in the lab at 4°C. The samples were stored in the original elution buffer or MilliQ water without any special treatment prior to storing them. Many of the samples had completely dried out. We resuspended the DNA molecules by adding 50 ul of MilliQ water to the apparently empty microtubes. Some of these samples did not show a clear smear of DNA when run on an electrophoretic gel, probably due to very low DNA concentrations. Despite the apparent lack of DNA, library preparation was successful, and the sequenced DNA clearly belonged to the species in question.

Using existing DNA extracts for genomic-level work is still not a common practice, although current restrictions in many biodiverse countries make sampling fresh specimens difficult. Apart from the costs of a new sampling campaign and DNA extraction, in many cases it is inhibitively complicated to obtain the necessary permits for such sampling. This is especially important since the implementation of Nagoya Protocol (NP) by many countries. Without entering the debate about the effects of such measures (See Neumann et al. 2018), it is clear that NP affects many researchers who have encountered very time-consuming practical issues in order to obtain the proper permits. A practical solution to reduce the unnecessary costs of sampling and getting the permits for it, could be to use specimens collected with the necessary permits in the past, which are now stored in refrigerators and freezers around the world. This solution has been proposed already by some researchers showing the importance of natural history museums and public collections (Sproul and Maddison 2017; Allio et al. 2020; Call 2020; Twort et al. 2020). Here we show that it is even possible to bring one step forward the approach proposed by these researchers making use of the extracted DNA available in numerous laboratories around the world. The costs of our approach are reasonably low when considering the amount of data created, especially considering that the cost of sampling and DNA extraction is zero. In this work we recovered slightly over 200 genes due to time constraints. However, it is possible to extract many more markers from the same assemblies, e.g. all protein-coding genes. The genomic results of this project have also been used in a parallel study to explore the diversity of symbionts within these species (Chapter 5).

This approach has some limitations also. The quality of our assemblies is relatively low and are not suitable for studies which need chromosome quality assemblies. Another important observation of our results is that even if we recover a relatively high number of the markers of interest, a significant number of them are incomplete or they are in short pieces which need to be manually assembled. This might be due

to many factors. Degraded DNA, contamination, repetitive sequences and skewed base frequencies are among the main factors affecting the assembly and therefore can cause the incomplete data observed here. Although here we based our library preparation only on short insert sizes (~300 bp) we had to sonicate the original DNA to homogenize our sequencing run. The mean amount of raw data obtained for each library was 13.5 Gb. Considering an average genome size of 0.5 Gb for Lepidoptera (Hanrahan and Johnston 2011; Triant et al. 2018) and a library size of 300bp (PE 150bp), in theory, we could achieve a mean coverage of 27x for each genome. In this rough calculation many factors reduce the efficiency as mentioned before. In practice we expect the coverage to be two to three times lower than that. This is usually translated into a lower quality assembly. Our approach was sufficient to extract the protein coding markers to study the phylogenetic relationships of the group. We did observe quite high molecular weight fragments in the DNA extracts, and one possibility left to be explored is using long read sequencing platforms (e.g. PacBio) to increase the quality of assemblies from old DNA extracts.

Many recent phylogenetic studies have used HTS approaches. These methods use different sources of -omic data. Some studies use transcriptomes (e.g. Borner et al. 2014; Bazinet et al. 2017; Naumenko et al. 2017; Price and Bhattacharya 2017; Leduc-Robert and Maddison 2018; Foley et al. 2019; Kawahara et al. 2019), others use whole-genomes (e.g. Sims et al. 2009; Bonaventura et al. 2010; Karnkowska et al. 2016; Cloutier et al. 2019) or TE methods (Young et al. 2016; Dietrich et al. 2017; Dornburg et al. 2017; Espeland et al. 2018; Godwin et al. 2018; Haddad et al. 2018). Another interesting point in favour of our approach is that we can potentially mine the obtained genomes for the markers used in all of these alternative approaches which allow us to have access to a broader range of genetic resources.

Phylogeny

The obtained phylogenetic tree (Figure 1) is by far the best resolved phylogenetic hypothesis for the family with respect to the deep nodes. In our result the subfamily Eulepidotinae is recovered with high support as the sister group to the rest of the family. This result is different from the Zahiri et al. (2012) placing the subfamily Scoliopteryginae as the sister group to the rest of the family (not supported). The next group in diverging is a well-supported clade comprising subfamilies Hypocalinae and Scoliopteryginae. Although the placement of these two subfamilies as the sister group of each other is well supported, their phylogenetical position as being the sister group to the rest of Erebidae (apart from Eulepidotinae) is not supported.

We sequenced four unplaced taxa, Rhesala imparata, Colobochyla salicalis, Epitausa dilina, and Schistorhynx argentistriga. Rhesala imparata forms a well-supported sister lineage to the subfamilies Anobinae and Rivulinae, a result also found by Zahiri et al. (2012). Despite high support for the placement of R. imparata, the pCF value suggests that an alternative placement one NNI move away (for a

more detailed discussion see Minh et al. 2020) is equally supported (Figure 2). This result suggest that Rhesala represents an undescribed subfamily within Erebidae. The placement of the species Colobochyla salicalis as the sister group to the subfamily Hypeninae represents a similar situation. This well-supported relationship is also observed in Zahiri et al. (2012). In contrast, the relatively well-supported placement of the species Epitausa dilina as the sister group to the subfamily Lymantriinae is very different to the result in Zahiri et al. (2012), who found it to be close to the subfamily Calpinae with low support. The subfamilies Pangraptinae, Aganainae, Herminiinae, Arctinae and the species Schistorhynx argentistriga form a well-supported clade which is also observed in Zahiri et al. (2012). The subfamily Pangraptinae is recovered as monophyletic with relatively low UFBoot2 support value, although all the other support measures support it, but its position as the sister group to the other members of this clade is well supported. The species S. argentistriga is recovered as the sister group to the remaining three subfamilies Arctiinae, Herminiinae and Aganainae with poor support. The only subfamily which is not recovered monophyletic is Aganainae which appeared paraphyletic but poorly supported. We believe to improve the support values and to clarify the relationships within this part of the three probably a better taxon sampling within Aganainae and Pangraptinae will have a major effect as Arctiinae is relatively well sampled. Within Arctiinae The well-supported positioning of the tribes Lithosiini, Amerilini, Syntomini and Arctiini is in concordance with other studies of this subfamily (Zahiri et al. 2012; Zaspel et al. 2014; Rönkä et al. 2016; Dowdy et al. 2020).

The placement of Calpinae as the sister group to the rest of the tree is similar to Zahiri et al. (2012) and similarly not well supported. Further up the tree, Erebinae is placed as the sister group to Toxocampinae, Tinoliinae, Boletobiinae, Scolecocampinae and Hypenodinae with high support. Within the subfamily Erebinae, the position of the species Acantholipes circumdata (tribe Acantholipini) as the sister group to the rest of the subfamily, is in concordance with Homziak et al. (2019) using an anchored hybrid enrichment (AHE) approach. The well-supported placement of the species Sypnoides fumosa (tribe Sypnini) is also reported in the literature (Zahiri et al. 2012; Homziak et al. 2019). It is interesting to observe low support values in UFBoot2 (Figure 1) and concordance factors (Figure 2) in the rest of the subfamily although it is a a more recent part of the tree. This might be a sign of low taxon sampling in this part as apparently in Homziak et al. (2019) this part is well resolved with more terminal taxa.

The subfamily Toxocampinae is placed with high support as the sister group to Tinoliinae, Boletobiinae, Scolecocampinae and Hypenodinae. It is interesting to observe that, even if not supported at all, Zahiri et al. (2012) recovered the same pattern with extremely short internal branches. In the rest of the tree Scolecocampinae and Hypenodinae form a well-supported clade but its relationship with Boletobiinae and Tinoliinae is not well supported (Figure 1 and 2).

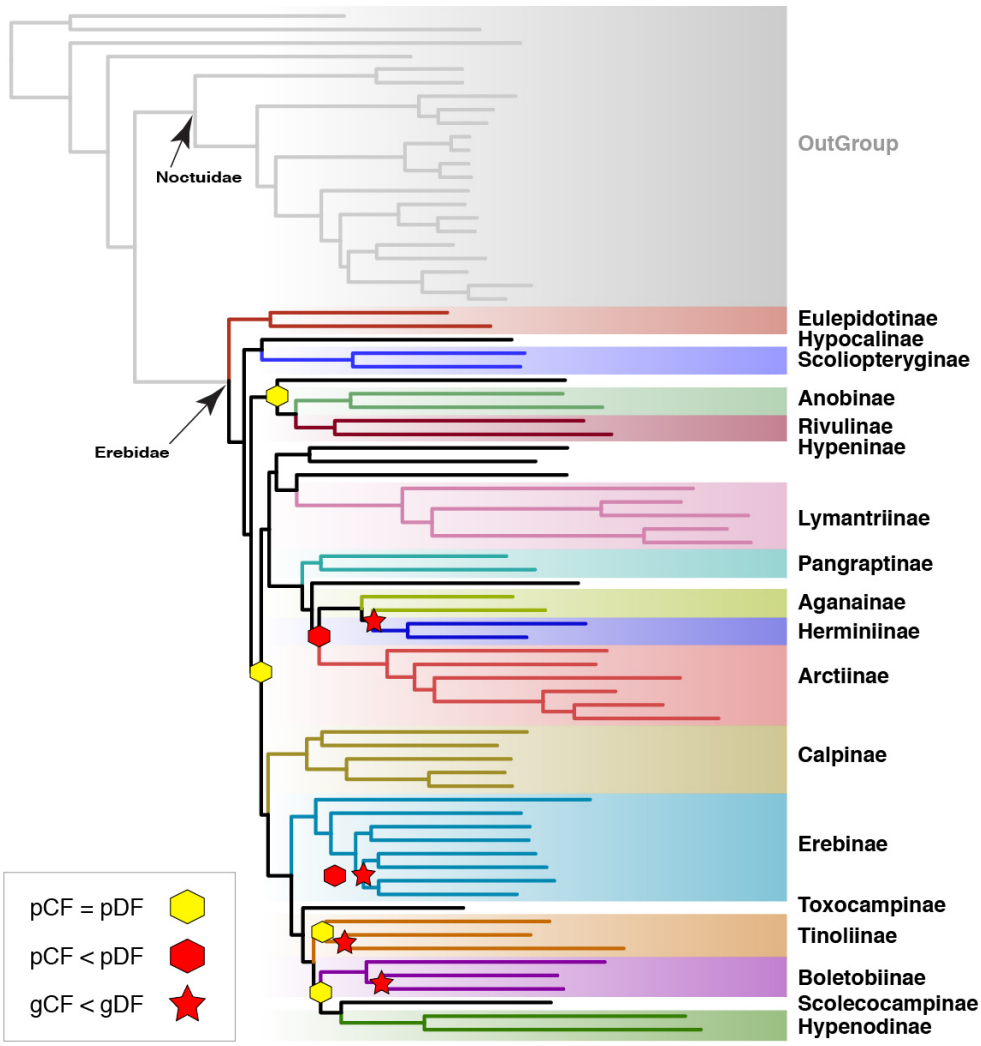


Figure 2: Concordance factors within Erebidae phylogenetic tree as obtained from IQ-Tree2. pCF is the partition concordance factor and gCF the gene concordance factor. pDF is any of the two partition discordance factors and gDF, similarly, any of the two gene discordance factors.

The interpretation of the concordance factor support values is complex. Here we applied both sCF and gCF to our analysis in addition to a variation of gCF considering each partition as a gene (figure 1 and 2). The sCF method was quite poor in informing us how sites supported deep nodes, especially because in the majority of the cases the differences between the sCF and sDF values were very small. We believe that sCF is very difficult to use in a comparative way. Most of the sCF values in our study scored less than 50% (Figure 3). Considering that a 33%

sCF corresponds to a random distribution of nucleotide site support for nodes, our result suggests that a large majority of the nodes have no support from individual sites, which is clearly not in line with our other measures of support. The gCF seems a better option and slightly easier to interpret. Here, as our genetic markers vary radically in size with some of them being relatively short, we used a modified gCF where we combined the genes into bigger clusters using the ModelFinder's best partitioning scheme. This partition based concordance factor (pCF) seemed to perform slightly better than the gCF (Figure 2).

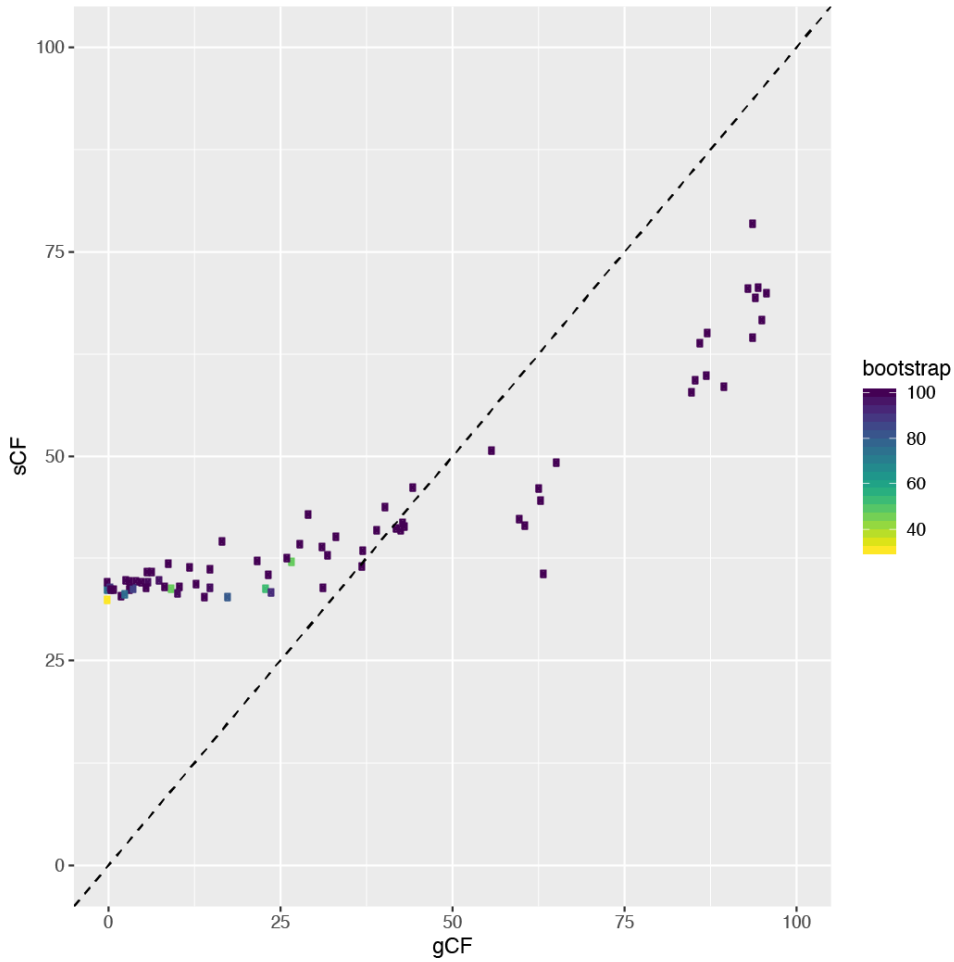


Figure 3: The interaction of site concordance factor (sCF) and gene concordance factor (gCF) with UFBoot2 values.

Conclusion

In this study we successfully sequenced whole genomes of Erebidae moths from 47 DNA extracts that were over 10 years old, demonstrating the high value of this resource for HTS approaches. This has a high importance in many cases where fresh sampling is not possible due to legislative or economic reasons. In addition, we show the power of a manually curated gene set alignment to resolve a complex phylogenetic relationship due to short internal branches.

The number and diversity of the sequenced genomes of this moth family increased substantially passing from only 7 species from 3 subfamilies to 54 species from 18 subfamilies. We offer the best resolved deep phylogenetic hypothesis for Erebidae moths up to date. We found that the subfamily Eulepidotinae is the sister group to all the other members of the family. We resolved the relationship between numerous subfamilies for the first time. For further studies we believe a more complete taxon sampling is needed specially in Aganainae, Pangraptinae, Erebinae, Tinoliinae and Boletobiinae. And finally, exploring the concordance factors, all of them being very complex to interpret, we believe the gCF or its variation, pCF, is more useful than sCF.

Aknowledgement

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). The authors acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We thank Marko Mutanen for providing the DNA extract for *Scoliopteryx libatrix*.

References

Allio R., Scornavacca C., Nabholz B., Clamens A.-L., Sperling F.A., Condamine F.L. 2020. Whole Genome Shotgun Phylogenomics Resolves the Pattern and Timing of Swallowtail Butterfly Evolution. Syst. Biol. 69:38–60.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bazinet A.L., Mitter K.T., Davis D.R., van Nieukerken E.J., Cummings M.P., Mitter C. 2017. Phylotranscriptomics resolves ancient divergences in the Lepidoptera. Syst. Entomol. 42:305–316.

Blaimer B.B., Lloyd M.W., Guillory W.X., Brady S.G. 2016. Sequence Capture and Phylogenetic Utility of Genomic Ultraconserved Elements Obtained from Pinned Insect Specimens. PLoS One. 11:e0161531.

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

Bonaventura M.P. Di, Lee E.K., DeSalle R., Planet P.J. 2010. A whole-genome phylogeny of the family Pasteurellaceae. Mol. Phylogenet. Evol. 54:950–956.

Borner J., Rehm P., Schill R.O., Ebersberger I., Burmester T. 2014. A transcriptome approach to ecdysozoan phylogeny. Mol. Phylogenet. Evol. 80:79–87.

Call E. 2020. The age of museomics - How to get genomic information from museum specimens of Lepidoptera. .

Cloutier A., Sackton T.B., Grayson P., Clamp M., Baker A.J., Edwards S. V. 2019. Whole-Genome Analyses Resolve the Phylogeny of Flightless Birds (Palaeognathae) in the Presence of an Empirical Anomaly Zone. Syst. Biol. 68:937–955.

Dietrich C.H., Allen J.M., Lemmon A.R., Lemmon E.M., Takiya D.M., Evangelista O., Walden K.K.O., Grady P.G.S., Johnson K.P. 2017. Anchored Hybrid Enrichment-Based Phylogenomics of Leafhoppers and Treehoppers (Hemiptera: Cicadomorpha: Membracoidea). Insect Syst. Divers. 1:57–72.

Dornburg A., Townsend J.P., Brooks W., Spriggs E., Eytan R.I., Moore J.A., Wainwright P.C., Lemmon A.R., Lemmon E.M., Near T.J. 2017. New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. Mol. Phylogenet. Evol. 110:27–38.

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). Syst. Entomol.

Espeland M., Breinholt J., Willmott K.R., Warren A.D., Vila R., Toussaint E.F.A., Maunsell S.C., Aduse-Poku K., Talavera G., Eastwood R., Jarzyna M.A., Guralnick R., Lohman D.J., Pierce N.E., Kawahara A.Y. 2018. A Comprehensive and Dated Phylogenomic Analysis of Butterflies. Curr. Biol. 28:770-778.e5.

Foley S., Lüddecke T., Cheng D.-Q., Krehenwinkel H., Künzel S., Longhorn S.J., Wendt I., von Wirth V., Tänzler R., Vences M., Piel W.H. 2019. Tarantula phylogenomics: A robust phylogeny of deep theraphosid clades inferred from transcriptome data sheds light on the prickly issue of urticating setae evolution. Mol. Phylogenet. Evol. 140:106573.

Godwin R.L., Opatova V., Garrison N.L., Hamilton C.A., Bond J.E. 2018. Phylogeny of a cosmopolitan family of morphologically conserved trapdoor spiders (Mygalomorphae, Ctenizidae) using Anchored Hybrid Enrichment, with a description of the family, Halonoproctidae Pocock 1901. Mol. Phylogenet. Evol. 126:303–313.

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. Syst. Biol. 59:307–321.

Haddad S., Shin S., Lemmon A.R., Lemmon E.M., Svacha P., Farrell B., Ślipinski A., Windsor D., McKenna D.D. 2018. Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). Syst. Entomol. 43:68–89.

Hanrahan S.J., Johnston J.S. 2011. New genome size estimates of 134 species of arthropods. Chromosom. Res. 19:809–823.

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.

Kalyaanamoorthy S., Minh B.Q., Wong T.K., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods.

Karnkowska A., Vacek V., Zubáčová Z., Treitli S.C., Petrželková R., Eme L., Novák L., Žárský V., Barlow L.D., Herman E.K., Soukal P., Hroudová M., Doležal P., Stairs C.W., Roger A.J., Eliáš M., Dacks J.B., Vlček Č., Hampl V. 2016. A Eukaryote without a Mitochondrial Organelle. Curr. Biol. 26:1274–1284.

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., Plotkin D., Espeland M., Meusemann K., Toussaint E.F.A., Donath A., Gimnich F., Frandsen P.B., Zwick A., dos Reis M., Barber J.R., Peters R.S., Liu S., Zhou X., Mayer C., Podsiadlowski L., Storer C., Yack J.E., Misof B., Breinholt J.W. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc. Natl. Acad. Sci. 116:22657–22663.

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.

Kieran T.J., Gordon E.R., Forthman M., Hoey-Chamberlain R., Kimball R.T., Faircloth B.C., Weirauch C., Glenn T.C. 2019. Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. Mol. Phylogenet. Evol. 130:297–303.

Leduc-Robert G., Maddison W.P. 2018. Phylogeny with introgression in Habronattus jumping spiders (Araneae: Salticidae). BMC Evol. Biol. 18:24.

Li W., Cong Q., Shen J., Zhang J., Hallwachs W., Janzen D.H., Grishin N. V. 2019. Genomes of skipper butterflies reveal extensive convergence of wing patterns. Proc. Natl. Acad. Sci. 116:6232–6237.

Lopez L., Turner K.G., Bellis E.S., Lasky J.R. 2020. Genomics of Natural History Collections for Understanding Evolution in the Wild. Mol. Ecol. Resour. n/a:1755–0998.13245.

Meyer M., Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb. Protoc. 2010:pdb.prot5448-pdb.prot5448.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. Mol. Biol. Evol.

Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer A.J., Aspock U., Aspock H., Bartel D., Blanke A., Berger S., Bohm A., Buckley T.R., Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y., Jermiin L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z., Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G., Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y., Pass G., Podsiadlowski L., Pohl H., von Reumont B.M., Schutte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X., Szucsich N.U., Tan M., Tan X., Tang M., Tang J., Timelthaler G., Tomizuka S., Trautwein M., Tong X., Uchifune T., Walzl M.G., Wiegmann B.M., Wilbrandt J., Wipfler B., Wong T.K.F., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates D.K., Yoshizawa K., Zhang Q., Zhang R., Zhang W., Zhang Y., Zhao J., Zhou C., Zhou L., Ziesmann T., Zou S., Li Y., Xu X., Zhang Y., Yang H., Wang J., Wang J., Kjer K.M., Zhou X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science (80-.). 346:763–767.

Naumenko S.A., Logacheva M.D., Popova N. V, Klepikova A. V, Penin A.A., Bazykin G.A., Etingova A.E., Mugue N.S., Kondrashov A.S., Yampolsky L.Y. 2017. Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection. Mol. Ecol. 26:536–553.

Neethiraj R., Hornett E.A., Hill J.A., Wheat C.W. 2017. Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. Mol. Ecol. 26:4990–5002.

Neumann D., Borisenko A. V, Coddington J.A., Häuser C.L., Butler C.R., Casino A., Vogel J.C., Haszprunar G., Giere P. 2018. Global biodiversity research tied up by juridical interpretations of access and benefit sharing. Org. Divers. Evol. 18:1–12.

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. Mol. Biol. Evol. 32:268–274.

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., Davis D.R., De Prins J., De Prins W., Epstein M.E., Gentili-Poole P., Gielis C., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S. (J C., Kozlov M. V, Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-T., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-C., Solis M.A., Tarmann G.M., Warren A.D., Weller S., Yakovlev R. V, Zolotuhin V. V, Andreas Z. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 1758:212–221.

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. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng M., Jiang R., Sun F., Zhang X., editors. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 158–170.

Peña C., Malm T. 2012. VoSeq: A Voucher and DNA Sequence Web Application. PLoS One. 7:e39071.

Price D.C., Bhattacharya D. 2017. Robust Dinoflagellata phylogeny inferred from public transcriptome databases. J. Phycol. 53:725–729.

Rambaut A. 2016. FigTree v1.4.3. .

Regier J.C., Mitter C., Zwick A., Bazinet A.L., Cummings M.P., Kawahara A.Y., Sohn J.-C.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.

Regier J.C., Zwick A., Cummings M.P., Kawahara A.Y., Cho S., Weller S., Roe A., Baixeras J., Brown J.W., Parr C., Davis D.R., Epstein M., Hallwachs W., Hausmann A., Janzen D.H., Kitching I.J., Solis M.A., Yen S.-H., Bazinet A.L., Mitter C. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. BMC Evol. Biol. 9:280.

Rohland N., Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22:939–946.

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

Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics.

Sims G.E., Jun S.-R., Wu G.A., Kim S.-H. 2009. Whole-genome phylogeny of mammals: Evolutionary information in genic and nongenic regions. Proc. Natl. Acad. Sci. 106:17077–17082.

Sproul J.S., Maddison D.R. 2017. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. Mol. Ecol. Resour. 17:1183–1201.

St Laurent R.A., Hamilton C.A., Kawahara A.Y. 2018. Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae). Syst. Entomol. 43:729–761.

Suyama M., Torrents D., Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.
Triant D.A., Cinel S.D., Kawahara A.Y. 2018. Lepidoptera genomes: current knowledge, gaps and future directions. Curr. Opin. Insect Sci. 25:99–105.
Twort V.G., Minet J., Wheat C.W., Wahlberg N. 2020. Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life. bioRxiv.:2020.08.18.255182.
Young A.D., Lemmon A.R., Skevington J.H., Mengual X., Ståhls G., Reemer M., Jordaens K., Kelso S., Lemmon E.M., Hauser M., De Meyer M., Misof B., Wiegmann B.M. 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol. Biol. 16:143.
Zahiri R., Holloway J.D., Kitching I.J., Lafontaine J.D., Mutanen M., Wahlberg N. 2012. Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). Syst. Entomol. 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). Zool. Scr. 40:158–173.
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.

Paper IV

The accuracy of mitochondrial genomes for family level phylogenies, the case of erebid moths (Lepidoptera; Erebidae)

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

[1] Systematic Biology Group, Biology department, Lund University, Lund, Sweden.

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

[3] Centrum für Naturkunde, University of Hamburg, Martin-Luther-King Platz 3, 20146 Hamburg, Germany.

Abstract

Many technological advances have affected the life sciences, but very few of them as much as the discovery of polymerase chain reaction or PCR. The ease and reliability of PCR allowed many researchers an invaluable access to molecular genetic data. The use of molecular data in order to study evolutionary history of different organisms, revolutionised the field of systematics. Now with the appearance of high throughput sequencing (HTS) technologies more and more genetic sequence data is available. One of the important sources of genetic data for phylogenetic analyses has always been mitochondrial DNA. The limitations of mitochondrial DNA for phylogenetic studies has been thoroughly studied in the age of single locus phylogenetic studies. Now with the appearance of genomic scale data, more and more mitochondrial genomes are available. Here we assemble 47 mitochondrial genomes using whole genome Illumina short reads of samples of family Erebidae (Lepidoptera), in order to evaluate the accuracy of use of mitochondrial genome in resolving deep phylogenetic relationships. We find the inadequacy of mitochondrial genomes resolving subfamily level relationships in Erebidae, but given good taxon sampling, we see its potential in resolving lower level phylogenetic relationships.

Key words: Erebidae, phylogenomics, mitochondrial genome, Lepidoptera

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 and Hadrys 2013; Song et al. 2016). Since the beginning of the popularization of such methods, some groups questioned the limitations of mitochondrial genetic data for resolving deep phylogenetic relationships (Zardoya and Meyer 1996; Cameron et al. 2004; Talavera and Vila 2011). Nevertheless, many studies used this source of information to resolve phylogenetic relationships of varied evolutionary depth. Some studies focused on the relationships within a superorder (Cameron et al. 2009; Talavera and Vila 2011; Li et al. 2015), an order (Cameron et al. 2007; Fenn et al. 2008; Dowton et al. 2009; Kim et al. 2011; Timmermans et al. 2014; López-López and Vogler 2017; Yang et al. 2019), a family (Chen et al. 2014, 2020b; Yang et al. 2015; Li et al. 2018, 2020; Xu et al. 2020; Zhang et al. 2020) or shallower relationships.

The phylogenetic depth of a relationship affects the amount of phylogenetic signal coded in the molecular data. In general, markers with higher mutation rates are only informative for the shallower evolutionary relationships or recent divergences. For clades splitting deeper in time, these fast-evolving markers tend to accumulate too many saturated sites and therefore tend to not resolve their phylogenetic relationship accurately. On the other hand, for markers having a very low mutation rate, a phylogenetic relationship can be too shallow for the marker to accumulate enough changes and have enough phylogenetic signal. Mitochondrial genomes usually contain relatively homogenous molecular markers in term of mutation rate (Brower 1994). Also, the lack of recombination in mitochondrial genomes means that in practice mitochondrial DNA behaves as a single genetic marker with a unique evolutionary history. In addition, the mitochondrial genome is only maternally inherited, meaning that it has an effective population size one fourth of the nuclear genome. Therefore, mitochondrial markers can be misleading in cases of hybridization and are more affected by demographic factors than nuclear markers.

The first approaches to sequence the mitochondrial genomes used PCR to amplify long pieces of overlapping molecules, Sanger sequencing the long molecules and manually assembling the sequence data. The labour intensiveness and pricy nature of these methods, made the 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. Therefore, it is currently

easier and more economical to obtain an even higher number of mitochondrial genomes. The modern-day ease of sequencing mitochondrial genomes has caused a rise in the publication of single genomes practically without addressing any research question. Some authors have already started to respond to these poor scientific practices by publishing a larger number of mitochondrial genomes with a clear question at the phylogenetic depths which the markers are proven to have enough phylogenetic signal (e.g. Chen et al. 2020a).

Considering the characteristics of a mitochondrial genome as a phylogenetic marker, the question of the phylogenetic depths where this marker is useful arises. Also important is the question if this important genetic marker can reliably resolve phylogenetic relationships in groups which have experienced rapid radiations. In case of rapid radiations, during a short period of time, numerous lineages arise. Resolving phylogenetic relationships in past rapid radiation events is challenging due to the fact that the marker should be fast evolving enough to accumulate enough changes during the rapid radiation, but slow enough to not saturate the signal afterwards. One of the groups which present such challenging phylogenetic dilemmas 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 very short branches are recovered at the deeper levels which suggested a rapid radiation event. The relationships at the subfamily level within Erebidae are poorly resolved probably due to the lack of phylogenetic signal in the markers used.

Here we assemble mitochondrial genome of 47 species of Erebidae representing all the subfamilies and major lineages based on the most recent phylogenetic hypotheses in order to capture all the deepest nodes within the family. We compare the obtained phylogenetic hypotheses with known and supported relationships recovered in other studies to evaluate the phylogenetic range of accuracy of mitochondrial genome as a marker.

Material and Methods

Taxon sampling

In this study we use the reads obtained from low coverage whole genome sequencing of 47 Erebidae moths (Table 1) obtained in the chapter 3. The taxon choice has been made in order to recover all the deepest nodes within the subfamilies and the major lineages in the Erebidae family. This allows us to focus mainly on the deep short branches which are usually the unresolved part of the tree for this family in published phylogenetic hypotheses. We also downloaded all the available

Erebidae mitochondrial genomes from GenBank (43 genomes). A total of 19 mitochondrial genomes were also used as outgroups which consisted of 11 Noctuidae, 4 Notodontidae, 3 Nolidae and one Euteliidae (Table 2).

Table 1: List of species used in this study. In column "circular" it is stated 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.

#	Codes	Species	circular	Length	#tRNA	GC%	AT%
1	MM00407	*Scoliopteryx libatrix*	Yes*	15617	22	19.4	80.6
2	RZ103	*Rema costimacula*	Yes	15668	22	19.3	80.7
3	RZ104	*Saroba pustulifera*	Yes	15731	22	19.5	80.5
4	RZ105	*Hypocala deflorata*	No	14428	19	18.8	81.2
5	RZ11	*Erebus ephesperis*	Yes	15688	22	18.6	81.4
6	RZ111	*Platyjionia mediorufa*	Yes	15329	22	19.9	80.1
7	RZ119	*Schistorhynx argentistriga*	Yes*	16660	27	19.5	80.5
8	RZ13	*Gonitis involuta*	Yes	15695	22	19.2	80.8
9	RZ138	*Micronoctua* sp.	Yes	15466	22	19	81
10	RZ149	*Hypopyra capensis*	Yes	15702	22	19.1	80.9
11	RZ159	*Rivula ochrea*	No	14510	19	18.2	81.8
12	RZ18	*Masca abactalis*	Yes	15562	22	19.5	80.5
13	RZ180	*Nodaria verticalis*	No	14175	18	18.5	81.5
14	RZ21	*Ophiusa coronata*	Yes	15762	22	18.8	81.2
15	RZ22	*Azeta ceramina*	Yes	15696	22	19	81
16	RZ248	*Acantholipes circumdata*	Yes	16224	25	20.7	79.3
17	RZ265	*Rhesala imparata*	Yes	15583	22	18.4	81.6
18	RZ268	*Mecodina praecipua*	Yes	15501	22	19	81
19	RZ271	*Idia aemula*	No	15464	22	18.8	81.2
20	RZ28	*Brunia antica*	Yes	15489	22	19.4	80.6
21	RZ3	*Laspeyria flexula*	Yes	15583	22	20.1	79.9
22	RZ30	*Creatonotos transiens*	Yes	15569	22	18.9	81.1
23	RZ313	*Sypnoides fumosa*	Yes	15527	22	19.4	80.6

24	RZ331	*Tinolius eburneigutta*	No	15026	21	19.1	80.9
25	RZ332	*Anoba anguliplaga*	No	14835	20	18.9	81.1
26	RZ336	*Calyptra hokkaida*	Yes	15562	22	18.3	81.7
27	RZ337	*Oraesia excavata*	Yes	15769	22	18.6	81.4
28	RZ34	*Nygmia plana*	No	14479	19	19.1	80.9
29	RZ367	*Hypena baltimoralis*	No	14724	20	19.6	80.4
30	RZ389	*Tamsia hieroglyphica*	Yes	15598	22	20	80
31	RZ39	*Ericeia subcinerea*	Yes*	15880	24**	19.7	80.3
32	RZ4	*Colobochyla salicalis*	Yes	16449	22	18.5	81.5
33	RZ40	*Pangrapta bicornuta*	Yes*	15957	22	18.1	81.9
34	RZ404	*Amerila astreus*	Yes	15519	22	19.6	80.4
35	RZ41	*Metaemene atrigutta*	Yes	15629	22	20.5	79.5
36	RZ42	*Luceria striata*	Yes	15383	22	20.1	79.9
37	RZ44	*Asota heliconia*	Yes	15446	22	19.9	80.1
38	RZ48	*Sympis rufibasis*	Yes	15572	22	18.5	81.5
39	RZ56	*Phyllodes eyndhovii*	Yes	15612	22	18.2	81.8
40	RZ57	*Lygephila maxima*	Yes	15591	22	19.3	80.7
41	RZ58	*Melipotis jucunda*	Yes*	16616	22	18.5	81.5
42	RZ59	*Panopoda rufimargo*	Yes	15986	22	18.8	81.2
43	RZ8	*Syntomis phegea*	Yes	15534	22	18.9	81.1
44	RZ89	*Arctornis* sp.	Yes	15506	22	21.4	78.6
45	RZ9	*Scolecocampa liburna*	Yes	15580	22	18.9	81.1
46	RZ93	*Epitausa dilina*	Yes	15440	22	18.7	81.3
47	RZ94	*Alesua etialis*	Yes	15198	19	17.7	82.3

MtGenome assembly

In order to assemble the mitochondrial genomes (de novo) we have used Novoplasty (Dierckxsens et al. 2016) on all samples. For this analysis the raw forward and reverse read files were used with a kmer of 21. This approach gave a clean circular genome in 34 samples (72%). In an additional 5 samples (11%) the result was

sufficient to manually circularize them in Geneious 10.2.6 (Kearse et al. 2012). The remaining 8 samples (17%) did not result with 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 and Edwards 2011) to remove the reads containing 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 PHRED 25 on average. Then on the cleaned reads, we used the mirabait option in MIRA 4.0.2 (Chevreux et al. 1999, 2004) to find the reads corresponding to mitochondrial DNA. The mitochondrial reads were de novo assembled using three simultaneous approaches, the Geneious de novo assembler, SPAdes assembler 3.10.0 (Nurk et al. 2013) and plasmidSPAdes (Antipov et al. 2016), all of them implemented in Geneious. For each sample, all the contigs over 500 bp were aligned to a reference MtGenome of another species of Erebidae. Then the consensus sequence of the alignment was used as a reference to map the mitochondrial reads in Bowtie2 (Langmead and Salzberg 2012) as implemented in Geneious with default parameters. All the resulting assembled genomes were annotated using MITOS (Bernt et al. 2013).

Phylogenetic analyses

Eleven protein coding genes (PCG) were extracted from all mitochondrial genomes. This dataset 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 4L (ND4L). We excluded two genes (ATP8 and ND6) from our dataset as they did not align properly. Each gene was aligned separately using MAFFT v7.450 (Katoh 2002; Katoh and Standley 2013) as implemented in Geneious with default options. After the revision and manually correction of the alignments, they have been uploaded to a private database VoSeq (Peña and Malm 2012). Using VoSeq database application, we created a nucleotide concatenated dataset (nc) with a total length of 10,245 bp and an amino acid dataset (aa) of 3,415 characters.

We ran maximum likelihood (ML) analyses with both nc (partitioned by gene and codon position) and aa (partitioned by gene) datasets 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 the node supports with 5000 ultrafast bootstrap approximations (UFBoot2) and 1000 SH-like approximate likelihood ratio test (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 analysis. Additionally we tested the best partitioning scheme for the nucleotide dataset 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 107 generations sampling every 103 steps. This analysis was repeated five times. The convergence of the runs were checked in Tracer 1.7.1 (Rambaut et al. 2018). The resulting trees were visualized and rooted in FigTree v1.4.3 (Rambaut 2016) using the outgroups. The COI gene was extracted from all the assembled genome to compare with the sequences obtained with Sanger sequencing as an extra quality control.

Table 2: 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	Spodoptera frugiperda	SRR3406055
5	Noctuidae	Helicoverpa zea	SRX371342
6	Noctuidae	Agrotis segetum	SRR1231960
7	Noctuidae	Athetis lepigone	SRR796575
8	Noctuidae	Trichoplusia ni	???
9	Noctuidae	Helicoverpa armigera	SRR1565435
10	Noctuidae	Chrysodeixis includens	SRR2049082
11	Noctuidae	Heliothis subflexa	ERR738599
12	Noctuidae	Mythimna separata	SRR5115697
13	Nolidae	Gabala argentata	NC_026842
14	Nolidae	Risoba prominens	NC_026841
15	Nolidae	Manoba major	SRR1300145
16	Notodontidae	Ochrogaster lunifer	NC_011128
17	Notodontidae	Phalera flavescens	NC_016067
18	Notodontidae	Notoplusia minuta	SRR1299746
19	Notodontidae	Thaumetopoea pityocampa	SRR1284701

The software Mira and Novoplasty were run using the resources provided by SNIC through 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). All data in the supplementary material, the alignment, the annotated genomes and the results can be found and downloaded from the GitHub repository: github.com/Hamidhrg/ErebidMtGenome.

Results

From the total number of 47 obtained genomes, 34 were fully assembled as circularized genomes. For the base frequency and basic genome composition result we only focus on the 34 good quality genomes. They varied in length from 15,198 bp in the sample RZ94 to 16,449 bp in the sample RZ4. Their AT base frequency ranged between 78.6% in RZ89 to 82.3% in RZ94. Their tRNA number was between 19 in RZ94 to 25 in RZ248 (Table 1). The annotated genomes are available through our online GitHub repository.

The ModelFinder in IQ-Tree2 merged the 33 possible partitions of the nucleotide dataset into 13 and found their corresponding best substitution models (Table 3). The partition sizes ranged between 96 to 1,411 bp (788 bp mean partition size). In total the dataset included 4,789 phylogenetically informative sites.

The ML analysis of the nc dataset resulted in the best resolved tree (Figure 1). Our samples of Erebidae in comparison to the selected outgroups resulted in a well-supported monophyletic group. All the other families used as outgroups were also recovered as monophyletic with more or less high support. Within Erebidae, the subfamily Lymantriinae was recovered as monophyletic and as the sister group to all other species of the family (not supported). The subfamilies Arctiinae and Erebinae were each recovered as monophyletic respectively with high support.

In contrast the ML analysis of aa dataset resulted in very anomalous trees. First of all, it appeared very sensitive to missing data. Therefore, 3 samples with the highest amount of missing data were deleted from the dataset and a new analysis was run. The resulting tree improved very slightly, however it was still very anomalous. In the case of the Bayesian inference (BI) in MrBayes, all of the ten chains (five runs of 2 independent chains) reached the stationary phase but none of the runs converged with each other. The analysis was repeated for a longer (up to 108) generation number and with a higher temperature (up to temp = 0.7) resulting in the same issue.

Table 3: The list of available genomic information retrieved from public online repositories.

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		**10245**	**4789**	**4499**	

Discussion

The most complete study focused on Erebidae up to date is the phylogeny of the family published by Zahiri et al. (2012). Using seven nuclear and one mitochondrial marker (for a total of 6,407 bp) they inferred a phylogenetic hypothesis with numerous unsupported short branches which did not resolve the relationship between different subfamilies and tribes. In chapter 3 we use some of the same specimens as used in Zahiri et al. (2012) to perform a whole genome sequencing approach to resolve the mentioned unsupported deep nodes. Here we will use the result of both Zahiri et al. (2012) and chapter 3 of this thesis, to evaluate the resolution of our dataset and its power in answering unresolved phylogenetic relationships within Erebidae.

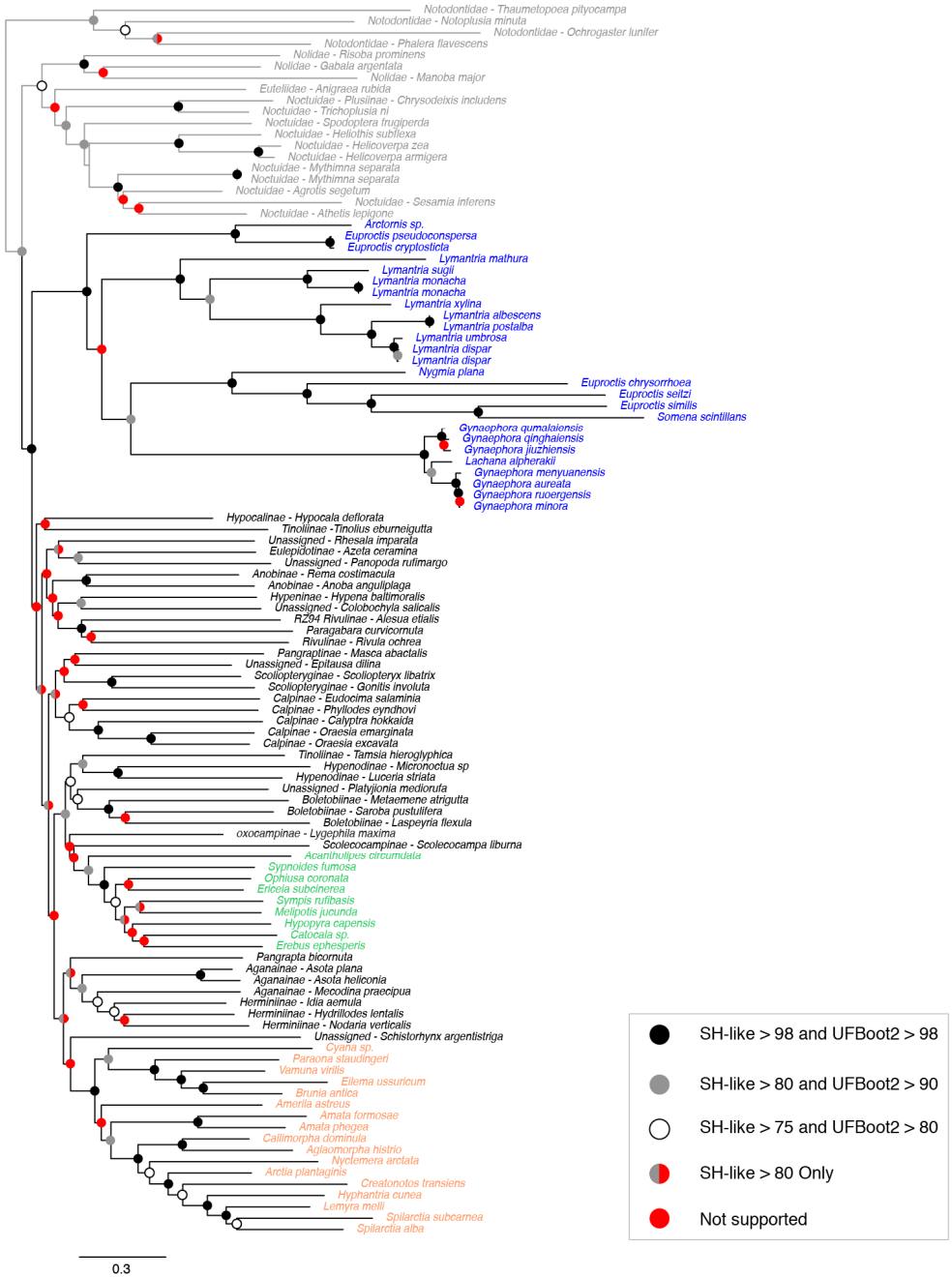


Figure 1: The ML tree obtained using the nc dataset in IQ-Tree2. The clade coloured in blue is the one corresponding to Lymantriinae subfamily, green Erebinae and orange Arctiinae. Black circles represent highly supported nodes, grey supported nodes, white low support and red not supported nodes. The outgroup clade is coloured in grey.

In Zahiri et al. (2012) the outgroup was formed by the same families chosen in this study. Therefore, we rooted the tree the same way to place Notodontidae as the sister group to the rest of the families (Nolidae, Eutelidae, Noctuidae and Erebidae). Within Erebidae we find that mitochondrial genomic data were not able to resolve the relationships of subfamilies with any confidence. The first group diverging from the rest of Erebidae in our study corresponds to the subfamily Lymantriinae. The position of this species as the sister group to the rest of the family is not supported in our study. Zahiri et al. (2012) do not find Lymantriinae in the same position but also in that study its position is not supported. In the chapter 3 on the other hand the family Eulepidotinae was recovered as the sister group to all other Erebidae subfamilies. Branch lengths within the Lymantriinae clade appear to be longer than in the rest of the tree. This could be due to the difference in taxon sampling within this clade compared to the rest of the tree (Fig. 1). The support values of the nodes in this clade could appear high at first, but after further attention it is clear that the high support values only correspond to the relationships within the same genus and not between the different genera. Wang et al. (2015) studied this subfamily using eight molecular markers. In the mentioned study the relationship between different tribes are poorly supported. The most basal clade in that study is the tribe Daplasini which we did not include in our dataset. The second basal clade in their study corresponds to the tribe Arctornithini which is in concordance with our results.

Within the subfamily Erebinae there is a lack of support for the resolution of the relationships between different genera. The placement of Acantholipes circumdata (Acantholipini) as the sister group of the rest of the subfamily is also recovered in Homziak et al. (2019) and in the chapter three of this thesis. Homziak et al. (2019) used anchored hybrid enrichment (AHE) phylogenomics to resolve the deep node relationships within this subfamily. The position of the species Sypnoides fumosa (Sypnini) in Erebinae clade is in concordance with Zahiri et al. (2012), Homziak et al. (2019) and the chapter three. The rest of the relationships within the subfamily are poorly resolved and do not agree with the mentioned studies.

The relationships within Arctiinae are better supported and appear to be better resolved. The clade composed by Cyana sp., Paraona staudingeri, Vamuna virilis, Eilema ussuricum and Brunia antica, representing the tribe Lithosiini, is placed as the sister group to the rest of the subfamily. This position of the Lithosiini tribe is in concordance with other studies (Zahiri et al. 2012; Zaspel et al. 2014; Rönkä et al. 2016; Dowdy et al. 2020). Also, the position of Amerila astreus (Amerilini), even though it is not supported, and the relationship of Callimorphina and Arctiina tribes are similar to the afore mentioned studies.

The relationships between the subfamilies have not been resolved in any published phylogenetic work up to date. 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 more comprehensive taxon sampling are needed in order to resolve these relationships which is the approach we have used

in the chapter three of this thesis. The results of our study show very low support values for these internal nodes suggesting that the amount of information coded in the mitochondrial genome is not enough to deal with such rapid radiations of similar or older ages. One of the caveats of our study is the taxon sampling in our dataset. Although our dataset has low taxon sampling, it is still comparable to most multi-marker phylogenetic studies in number of species and definitely larger than most phylogenomic datasets. Hence, we believe that improving the taxon sampling will definitely improve the phylogenetic resolution. Nevertheless, most probably, it will only affect the more recent divergence events as is visible in the better sampled clades in our study (e.g. Arctiinae).

Conclusion

The advances in sequencing technologies and the bioinformatics supporting it have revolutionized the molecular systematics, evolutionary biology and phylogenomics, among other fields. Especially with the advances in HTS, nowadays, sequencing a big number of mitochondrial genomes is relatively cheap and does not need much more infrastructure than the traditional PCR lab. This has allowed a rise in the number of new mitochondrial genomes being published practically on a weekly base in the last few years. These short publications usually publish a single new mitochondrial genome together with a very brief and rudimentary phylogenetic analysis.

In this study we question the utility of mitochondrial genome data to resolve deeper phylogenetic relationships accurately, or to resolve relationships of groups involving rapid radiation events. Based on our findings, at least for the erebid moths, mitochondrial genomes are not a good enough source of information per se, to resolve the relationships within and between subfamilies. The relationships between different close tribes could potentially be studied with a high enough taxon sampling in Erebidae. We also show that it is clear that amino acid datasets based on mitochondrial protein coding genes are not useful to study phylogenetic relationships at this level.

Aknowledgement

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 providing the DNA extract for *Scoliopteryx libatrix*.

References

Antipov D., Hartwick N., Shen M., Raiko M., Lapidus A., Pevzner P.A. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics. 32: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. Mol. Phylogenet. Evol. 69:313–319.

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

Brower A. V. 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. 91:6491–6495.

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. Syst. Entomol. 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. Zool. Scr. 38:575–590.

Chen L., Wahlberg N., Liao C.Q., Wang C. Bin, Ma F.Z., Huang G.H. 2020a. 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. 2020b. Comparative analysis of mitogenomes among six species of grasshoppers (Orthoptera: Acridoidea: Catantopidae) and their phylogenetic implications in wing-type evolution. Int. J. Biol. Macromol. 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 Res. 14:1147–1159.

Chevreux B., Wetter T., Suhai S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Comput. Sci. Biol. Proc. Ger. Conf. Bioinforma.:45–56.

Dierckxsens N., Mardulyn P., Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:gkw955.
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). Syst. Entomol.
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. Mol. Phylogenet. Evol. 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. Mol. Phylogenet. Evol. 49:59–68.
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. Syst. Biol. 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.
Kalyaanamoorthy S., Minh B.Q., Wong T.K., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods.
Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 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). Mol. Phylogenet. Evol. 61:436–445.
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. Mol. Biol. Evol. 34:772–773.
Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. 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. Sci. Rep. 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. Int. J. Biol. Macromol. 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. Int. J. Biol. Macromol. 149:371–380.
López-López A., Vogler A.P. 2017. The mitogenome phylogeny of Adephaga (Coleoptera). Mol. Phylogenet. Evol. 114:166–174.
Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010.
Nardi F. 2003. Hexapod Origins: Monophyletic or Paraphyletic? Science (80-.). 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. Mol. Biol. Evol. 32:268–274.
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., Davis D.R., De Prins J., De Prins W., Epstein M.E., Gentili-Poole P., Gielis C., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S. (J C., Kozlov M. V, Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-T., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-C., Solis M.A., Tarmann G.M., Warren A.D., Weller S., Yakovlev R. V, Zolotuhin V. V, Andreas Z. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 1758:212–221.
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. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng M., Jiang R., Sun F., Zhang X., editors. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 158–170.
Peña C., Malm T. 2012. VoSeq: A Voucher and DNA Sequence Web Application. PLoS One. 7:e39071.
Rambaut A. 2016. FigTree v1.4.3. .
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67:901–904.
Rönkä K., Mappes J., Kaila L., Wahlberg N. 2016. Putting Parasemia in its phylogenetic place: a molecular analysis of the subtribe Arctiina (Lepidoptera). Syst. Entomol. 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. Syst. Biol. 61:539–542.
Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics.

Simon S., Hadrys H. 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda. Mol. Phylogenet. Evol. 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 Biol. Evol. 8:1411–1426.
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 Evol. Biol. 11:315.
Timmermans M.J.T.N., Lees D.C., Simonsen T.J. 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79:169–178.
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. Int. J. Biol. Macromol. 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). Int. J. Biol. Macromol. 136:123–132.
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. Mol. Phylogenet. Evol. 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). Syst. Entomol. 37:102–124.
Zardoya R., Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13:933–942.
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). Mol. Phylogenet. Evol. 143:106669.

Paper V

Exploring non-specific/bycatch diversity of organisms in whole genome sequencing of Erebidae moths (Lepidoptera)

Hamid Reza Ghanavi[1], Victoria Twort[1,2] and Anne Duplouy[1,3]

[1] Systematic Biology Group, Biology department, Lund University, Lund, Sweden.

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

[3] Organismal and Evolutionary Biology Research Program, The University of Helsinki, Helsinki, Finland.

Abstract

Models estimate that about 80% of all butterfly and moth species host vertically transmitted endosymbiotic microorganisms, which can affect the host fitness, metabolism, reproduction, population dynamics, and genetic diversity, among others. The supporting empirical data are however currently highly biased towards the generally more colourful butterflies, and include less information about moths. Additionally, studies of symbiotic partners of Lepidoptera predominantly focus on the common bacterium *Wolbachia pipientis*, while infections by other inherited microbial partners have more rarely been investigated. Here, we mine the genomes of 47 species of Erebidae moths, with the aims to both inform on the diversity of symbionts potentially associated to this Lepidoptera group, and discuss the potential of metagenomic approaches to inform on such diversity. Based on the result of Kraken2 and Methaphlan2 analyses, we found clear evidence of the presence of *Wolbachia* in four species. Our result also suggests the presence of three other bacterial symbionts (*Burkholderia* spp., *Sodalis* spp. and *Arsenophonus* spp.), in three other moth species. Additionally, we recovered genomic material from bracovirus in about half of our samples. The detection of the latter, usually found in mutualistic association to braconid parasitoid wasps, may inform on host-parasite interactions that take place in the natural habitat of the Erebidae moths.

Key words: Erebidae, phylogenomics, whole genome, old DNA extract, Lepidoptera

Introduction

A growing scientific community now sees each organism as a community of interacting species rather than as an independent entity. Insects are no exception, and are now well-known for hosting a variety of microbial symbionts sitting both inside and outside the host cells. These microorganisms are at least as numerous as the number of host cells, and may constitute up to 10% of the host total mass (Douglas 2015). The effect of the symbionts on the insect hosts is as diverse as their taxonomy, ranging from pathogenic to obligate mutualists, and all the intermediate possible relationships (Dillon and Dillon 2004). This diversity has recently attracted the growing interest of the scientific community, but studies on insect symbiosis still see many biases. For example, in Lepidoptera, research in symbiosis has mostly focused on the most charismatic groups of colourful diurnal butterflies (Altizer et al. 2000; Jiggins et al. 2000; Duplouy et al. 2010) or pest species to the human society (Xu et al. 2014; Chen et al. 2017; Bapatla et al. 2018). In contrast, although the rest of Lepidoptera (commonly called moths) are highly diverse, encompassing no less than 130,000 species (van Nieukerken et al. 2011), the symbiont diversity in this group has only been investigated in a few species (Duplouy and Hornett 2018).

High throughput sequencing technologies (HTS) now provide a relatively easy and cheap way to obtain large amounts of genetic data. These technologies used to generate genomic data are varied and broadly applicable to the widest range of organisms. Thereby they are currently revolutionising the accessibility to genetic resources for addressing questions in the natural sciences that could have been previously out of reach. Consequently, we currently see a fast growth in the number of genomes and genomic data being produced, and made freely accessible through online repositories.

When generating genomic data, DNA is extracted from the study organism, either entirely or a specific part of it. Sequencing such material results in a mix of primary host specific DNA, but also DNA from other sources. These other sources potentially include ectosymbionts, endosymbionts, food, and/or parasites among others. Such genomic data open up the genomic analyses towards broader targets, including investigating the diversity of symbionts that might be associated to particular targeted hosts.

In this study we mine the genomic data from 47 individually sequenced moth species from the family Erebidae to i) explore the potential diversity of symbionts associated to this megadiverse Lepidoptera family; ii) and to evaluate the exploratory power of recovering information on natural host-symbiont associations from the low coverage genome sequencing approaches.

Material and Methods

Genomes

For this study, we used 47 of the existing whole genomes of Erebidae produced for Chapter 3 of this thesis. The sampling information is shown in Table 1. In general, these genomes represent all the described subfamilies and major lineages within the family. The genomes where sequenced using DNA extracts that are over one-decade old (taken from a study by Zahiri et al. 2012), and originated from one or two legs. It is important to keep in mind that the genome sequencing approach generating this dataset is not optimized to recover the symbiont diversity of these organisms, therefore the diversity is likely to be systematically underestimated.

Metagenomic analysis

The raw reads were quality checked with FASTQC v0.11.8 (Andrews 2010). Reads containing ambiguous bases were removed from the dataset using Prinseq 0.20.4 (Schmieder and Edwards 2011). Reads were cleaned to remove low quality bases from the beginning (LEADING: 3) and end (TRAILING: 3) and reads less than 30 bp in length. The evaluation of read quality with a sliding window approach was done in Trimmomatic 0.38 (Bolger et al. 2014). Quality was measured for sliding windows of 4 bp and had to be greater than PHRED 25 on average. Cleaned reads were assigned taxonomic labels with Kraken2 (Wood and Salzberg 2014) and MetaPhlAn 2.0 (Segata et al. 2012). Kraken2 was run using a custom database, which contained the standard kraken database, the refseq viral, bacteria and plasmid databases and all available Lepidoptera genomes from genbank (Supplementary Table 1 contains a full list of taxa included), confidence threshold of 0.05, and a mpa style output. MetaPhIAn was run using the analysis type rel_ab_w_read_stats, which provides the relative abundance and an estimate of read numbers originating from each clade. We visually screened the result for each sample, focusing on seven genera of vertically transmitted bacterial symbionts (i.e. *Arsenophonus* sp., *Cardinium* sp., *Hamiltonella* sp., *Rickettsia* sp., *Sodalis* sp., *Spiroplasma* sp. and *Wolbachia* sp.), one group of fungal symbionts (Microsporidia), and three types of viral symbionts (i.e. *Wolbachia*-phage *WO*, ichnovirus and bracovirus). This represents a non-exhaustive list of the maternally inherited symbionts found in diverse insect hosts, but covers all of those that have already been characterized within Lepidoptera (Duplouy and Hornett 2018). We also checked on the presence of the gut bacteria *Burkholderia* sp., which are known to confer pesticide resistance to their host in the pest bean bug *Riportus pedestris* (e.g. ‘can degrade an organophosphate pesticide, fenitrothion) (Kikuchi and Yumoto 2013).

All data in the supplementary material, the tables and the results can be found and downloaded from the GitHub repository: github.com/Hamidhrg/ErebidSymbionts.

Results

Metagenomic analysis

We identified the species Idia aemula, Luceria striata, Acantholipes circumdata and Oraesia excavata (RZ271, RZ42, RZ248, and RZ337) as infected by Wolbachia, and Wolbachia-associated phage WO (Table 1). Additionally, the Illumina reads from the genome sequencing of sample RZ13 was also found to include 954 Wolbachia reads, which is more than any of the clearly uninfected specimens but considerably less than any of the four clearly infected specimens listed above. Specimens RZ103 and RZ111 also included considerably more reads from Sodalis bacteria (9,108 and 4,395, respectively), and from Arsenophonus bacteria (1,336 and 662, respectively), than any other samples (maximum of 50 reads in any other genome). A closer look at the kraken outputs from the latter two samples also revealed a possible infection with a Plautia stali symbiont (gammaproteobacteria; 3,856 and 1,914 reads, respectively), which was not detected in any of the other 45 samples. Looking for reads mapping to Burkholderia bacteria, all the samples presented a low number but the sample RZ30 which showed relatively higher number (1,995). Our Kraken and MetaPhlan analyses showed no to very few reads mapping to Cardinium, Hamiltonella or Spiroplasma bacteria, or to Microsporidian fungi, in any of the 47 metagenomes screened. Finally, we identified a considerable amount of reads from viruses of the polydnaviridae family, and especially of the Bracovirus genus, in the samples LM55 (1,288 reads), RZ18 (1,381 reads), and RZ44 (1,384 reads). All other samples only include less than 750 reads, and more often no reads, for these viruses.

All details of the screen for the common symbionts can be found in Table 1, while all results from the Kraken analyses can be found in the supplementary material and GitHub repository.

Table 1: The number of reads classified as originating from the host and various microorganisms. Underlined values highlight the values mentioned in the text, - represent samples with either zero or less than 20 reads classified. # stands for number of raw reads in million. Spi stands for *Spiroplasma*, Bur for *Burkholderia*, Sod for *Sodalis*, Ars for *Arsenophonus*, Ric for *Rickettsia*, Wol for *Wolbachia*, WO for *Wolbachia* Phage *WO*, Ich for *Ichnovirus*, Bra for *Bracovirus* and Mic for *Microsporidia*.

REF	Code	species	Country	#	Kraken2											Metaphlan2
					Lepidoptera	Spi	Bur	Sod	Ars	Ric	Wol	WO	Ich	Bra	Mic	Wolbachia
1	MM00407	Scoliopteryx libatrix	FINLAND	38	2,266,973	-	289	38	-	-	-	-	-	-	-	-
2	RZ103	Rema costimacula	HONG KONG	22	907,037	-	-	9,108	1,336	-	24	-	-	266	-	-
3	RZ104	Saroba pustulifera	HONG KONG	21	1,649,430	20	33	-	-	-	26	-	-	-	-	-
4	RZ105	Hypocala deflorata	HONG KONG	48	3,231,681	42	59	-	-	20	33	-	-	-	-	-
5	RZ11	Erebus ephesperis	TAIWAN	106	8,550,697	64	298	-	-	-	32	-	-	1,288	-	-
6	RZ111	Platyjionia mediorufa	HONG KONG	26	995,385	30	48	4,395	662	-	43	-	-	-	-	-
7	RZ119	Schistorhynx argentistriga	HONG KONG	56	5,928,236	56	99	-	-	29	-	-	-	-	-	-
8	RZ13	Gonitis involuta	TANZANIA	17	1,254,304	-	83	-	-	-	954	-	317	102	-	2,005
9	RZ138	Micronoctua sp.	INDONESIA	107	11,736,010	100	126	50	-	-	24	-	-	-	30	-
10	RZ149	Hypopyra capensis	GHANA	53	4,808,838	47	107	-	-	-	-	-	-	-	-	-
11	RZ159	Rivula ochrea	GHANA	59	6,499,556	71	216	-	-	26	-	-	-	-	-	-
12	RZ18	Masca abactalis	INDONESIA	45	4,175,988	46	67	-	-	-	22	-	-	1,381	-	-
13	RZ180	Nodaria verticalis	GHANA	38	4,198,076	44	116	-	-	26	20	-	-	1,731	-	-
14	RZ21	Ophiusa coronata	MALAYSIA	42	2,653,381	37	76	-	-	-	26	-	-	28	-	-
15	RZ22	Azeta ceramina	COSTA RICA	55	4,926,573	64	85	-	-	-	30	-	-	-	26	-
16	RZ248	Acantholipes circumdata	UAE	28	3,085,527	27	-	-	-	21	29,454	410	-	-	-	220,309
17	RZ265	Rhesala imparata	HONG KONG	38	6,206,848	34	67	-	-	29	-	-	-	-	-	-
18	RZ268	Mecodina praecipua	HONG KONG	26	2,200,296	20	36	-	-	-	-	-	-	790	-	-
19	RZ271	Idia aemula	USA	52	6,897,287	35	112	-	-	-	144,331	1,038	-	771	-	168,228
20	RZ28	Brunia antica	HONG KONG	77	7,118,395	59	242	-	-	50	30	-	-	-	-	-
21	RZ3	Laspeyria flexula	HUNGARY	54	7,583,217	46	82	-	-	22	161	-	20	-	-	-
22	RZ30	Creatonotos transiens	HONG KONG	30	6,196,702	28	1,995	-	-	-	-	-	-	198	-	-
23	RZ313	Sypnoides fumosa	JAPAN	87	10,986,269	48	505	-	-	33	47	-	-	104	26	576
24	RZ331	Tinolius eburneigutta	THAILAND	33	3,112,193	-	85	-	-	-	-	-	-	159	-	-
25	RZ332	Anoba anguliplaga	GHANA	42	1,874,468	-	79	-	-	-	42	-	-	-	-	-
26	RZ336	Calyptra hokkaida	JAPAN	34	5,835,726	-	122	-	-	-	-	-	-	341	-	-
27	RZ337	Oraesia excavata	HONG KONG	38	3,147,679	43	65	-	-	-	66,978	182	-	581	-	208,044

28	RZ34	Nygmia plana	HONG KONG	19	1,026,248	30	43	-	-	-	27	-	-	-	-	-
29	RZ367	Hypena baltimoralis	USA	35	3,005,435	46	54	-	-	-	24	-	-	27	-	-
30	RZ389	Tamsia hieroglyphica	MALAYSIA	26	1,285,828	23	63	-	-	-	30	-	-	572	-	-
31	RZ39	Ericeia subcinerea	HONG KONG	80	7,549,078	46	133	-	-	-	65	-	-	-	-	-
32	RZ4	Colobochyla salicalis	HUNGARY	44	5,510,176	33	57	-	-	-	35	-	-	72	-	-
33	RZ40	Pangrapta bicornuta	HONG KONG	63	7,415,193	43	316	-	-	25	60	-	-	107	22	-
34	RZ404	Amerila astreus	MALAYSIA	45	4,649,942	35	72	-	-	-	25	-	-	130	-	-
35	RZ41	Metaemene atrigutta	HONG KONG	17	1,079,839	-	29	-	-	-	-	-	-	-	-	-
36	RZ42	Luceria striata	HONG KONG	27	3,065,608	23	75	-	-	-	67,176	494	-	-	-	181,728
37	RZ44	Asota heliconia	HONG KONG	40	3,763,381	36	54	-	-	-	-	-	-	1,384	-	-
38	RZ48	Sympis rufibasis	HONG KONG	52	5,491,409	48	208	-	-	-	44	-	-	-	-	-
39	RZ56	Phyllodes eyndhovii	TAIWAN	64	4,058,586	44	118	-	36	26	45	-	-	21	-	-
40	RZ57	Lygephila maxima	JAPAN	41	3,832,732	35	117	-	-	-	-	-	-	-	-	-
41	RZ58	Melipotis jucunda	USA	57	5,764,266	40	101	-	-	-	-	-	-	101	44	-
42	RZ59	Panopoda rufimargo	USA	42	4,715,473	45	237	-	-	-	-	-	-	31	-	-
43	RZ8	Syntomis phegea	HUNGARY	22	1,675,147	20	44	-	-	-	-	-	-	309	177	-
44	RZ89	Arctornis sp.	JAPAN	33	3,256,478	-	50	-	-	-	-	-	-	94	-	-
45	RZ9	Scolecocampa liburna	USA	52	3,132,323	45	137	-	-	-	-	-	-	-	23	-
46	RZ93	Epitausa dilina	COSTA RICA	41	4,202,446	27	42	-	-	-	23	-	-	-	-	-
47	RZ94	Alesua etialis	COSTA RICA	16	1,605,058	-	76	-	-	-	-	-	-	-	-	-

Discussion

We confidently add four moth species (I.e. *Idia aemula*, *Luceria striata*, *Acantholipes circumdata* and *Oraesia excavata*) to the list of species hosting the intracellular alpha-proteobacterial symbiont *Wolbachia* (Hornett and Duplouy 2018) confirmed through two screening methods (i.e. Kraken2 and MetaPhlAn). With only 4/47 species (8%) found infected, this represents a lower infection rate than expected, as the literature suggest values between 16-79% of the studied lepidopteran groups infected with *Wolbachia* (Werren et al. 1995; Jiggins et al. 2001; Tagami and Miura 2004; Salunke et al. 2012; Ilinsky and Kosterin 2017; Duplouy and Brattström 2018). The general penetrance of *Wolbachia* can however be low in different species, thus our results are most likely underestimating of the true infection rate within the Erebidae moths. Future broader screenings of different populations will provide more accurate natural infection rates for these species. Noticeably, we observe the presence of *Wolbachia* phage *WO* in all the samples where *Wolbachia* presence is strongly supported. The interaction of this bacteriophage with *Wolbachia* has been the focus of many eco-evolutionary studies in the recent years (Gavotte et al. 2006; Tanaka et al. 2009; Wang et al. 2016b, 2016a; Kaushik et al. 2019). It is suggested that phage WO are associated with horizontal gene transfer in *Wolbachia*, and may affect the fitness of the bacterium. These bacteriophages have been observed in practically all the studied genomes of *Wolbachia* up to date, with very few obligate mutualistic exceptions (Gavotte et al. 2006; Kent and Bordenstein 2010; Bordenstein and Bordenstein 2016). In the sample RZ13, species *Gonitis involuta*, a relatively high number of reads mapped to *Wolbachia*, significantly lower than in the other 4 species, and no reads were mapped to phage-*WO*. This can be due to many non-excluding factors such as contamination from other genomic material, the integration of *Wolbachia* genomic material (partially or entirely) in the host genome, random errors in the Identification of the reads as *Wolbachia*, low quality genomic material or a combination of mentioned reasons. In this particular case, the sequencing produced a significantly lower number of reads than other processed samples, which may support the idea that the sample quality was low prior to sequencing. We however cannot rule out any of the other possibilities, and more studies are needed to fully confirm or reject the presence of *Wolbachia* in this species.

The two moth samples, *Rema costimacula* and *Platyjionia mediorufa*, were of particular interests. Both the Kraken and the MetaPhlan analyses suggest the presence of the gammaproteobacteria endosymbionts, *Sodalis*, *Arsenophonus* and *Plautia stali* symbiont in both samples. *Sodalis* has been characterized from different insects, including tsetse flies (Dale et al. 2001), seal louse (Boyd et al. 2016), pigeon louse (Fukatsu et al. 2007), loose flies (Šochová et al. 2017), aphids (Burke et al. 2009), seed bug (Santos-Garcia et al. 2017), weevils (Conord et al. 2008; Toju and Fukatsu 2011), stinkbugs (Kaiwa et al. 2011), bees (Rubin et al.

2018), and ants (Sameshima et al. 1999), among others. To our best knowledge however, this is the first time the three symbionts are found in Lepidoptera (Duplouy and Hornett 2018). This suggest that *Sodalis* bacteria might affect a more diverse group of organisms than is currently known. We are however cautious with the interpretation of this result. The simple discovery of these bacteria in the genomic data does not inform us about the nature of the interaction, consequently this result should not be interpreted as a symbiotic relationship between *Sodalis* bacteria and the moths. Contamination of those two samples prior to DNA extraction is always possible. However, the sequenced host genetic material did not include significant amount of hemipteran DNA (or any other non-lepidopteran insect order): comparable low numbers of reads (>1,500) mapped to Hemipterans in all the sequenced genomes, showing a negative result. This rule out DNA contamination by material from the confirmed hemipteran hosts of these three symbionts. It is shown that the female brown-winged green bug, *P. stali*, during oviposition, smears excrement over the egg surface. The nymphs right after hatching, ingest the excrement to acquire the symbionts (Oishi et al. 2019). Therefore, a possible contamination source could be any contact with such excrement/egg clusters. Once again, studies of the symbionts in natural populations of these moth species are needed to fully resolve the true infection state of these species and the relationship with their symbionts.

The moth species *Creatonotos transiens*, show the possible presence of the proteobacteria *Burkholderia* sp. Similarly to the other symbionts presented above, these bacteria are found in very diverse groups of organisms, from Amoebas to Orthoptera passing by humans, and plants (Itoh et al. 2014, 2019; Khojandi et al. 2019; Ohbayashi et al. 2019). It was also described in the microbiota associated to the moth *Lymantria dispar* (Mason and Raffa 2014). In the bean bug, *Riptortus pedestris*, studies have suggested that the bacteria can benefit their host by providing resistance to pesticides (Kikuchi et al. 2012). Although never tested, the presence of such Proteobacteria in moths could similarly enhance the host ability to resist pesticides, which could partially explain the global success of many pest moth species despite the development of various targeted control strategies.

Six genomes included significantly high amounts of *bracovirus* reads (RZ11, RZ18, RZ180, RZ268, RZ271 and RZ44). Bracoviruses are a known genus of mutualistic viruses with a complex life cycle. Integrated in the genome of a braconid parasitic wasp, the *bracovirus* is transcribed during oviposition in a lepidopteran larvae (Louis et al. 2013). The presence of this viral genetic material in adult moths might suggest an unsuccessful infection by the parasitoid, and the survival of the larvae carrying the parasitic viral particles. Another potential explanation could be that the viral DNA is integrated into the lepidopteran genome, as it is usually found in its common Hymenoptera host. Only studies simultaneously investigating parasitism success rate and tissue tropism of the bracoviruses in the Lepidoptera and Hymenoptera hosts, will be able to inform on the nature of these interactions.

From a methodological point of view, the present study shows the successful exploratory approach to mine for potentially hidden associated microbial diversity in genomes. The present study was performed on shallow genome short reads obtained using Illumina platforms for the purpose of studying the phylogenomics of the hosts species, but similar approach can be implemented to any publicly available genomic datasets. The popularity of genomic scale sequence data methods, such as Illumina short read approach, created a wide publicly open genomic resource for the research community to study questions which were not directly into the focus of the studies generating them. On the other hand, in such studies it is very important to take into consideration the limitations of such approaches. One of the important limitations that we observed in this study is the reference datasets needed for programs like Kraken2. The quality and completeness of reference affects highly the results. Another important point is the nature and the quality of the genomes. Incomplete and shallow genomes are an exceptional resource for many fields as phylogenomics or population genomics, but they tend to present false negatives when mined for many symbionts. Also, the origin of the DNA used for the genome sequencing has an important effect in designing such studies. In this study all the used genomes came from DNA extracted from legs, therefore there is a methodical hard bias against gut fauna for example. Another important limitation of this method is to inform about the nature of the interaction between the organisms found in the genomic mix. In the majority of cases this method does not allow us to inform about the origin of the found organisms either. This is especially important as sample contamination is a known problem for molecular techniques since the appearance of these methods. In addition, this method is not suitable for quantification of the present organisms neither. The sum of all these points is the main reason we insist in the exploratory nature of such approaches.

Conclusion

As we expected, our method detects *Wolbachia* and the bacteriophage *WO* in four moth species, *Burkholderia* in one other species, and *Sodalis* and *Arsenophonus* simultaneously in two species. Although symbiotic associations of Lepidoptera with *Wolbachia* is likely, similar long-term associations between the three other symbionts and Lepidoptera have yet to be described. Similarly, we detect DNA material from bracoviruses that are currently only described as mutualistic symbionts of Hymenoptera. The true nature of these associations requires further experimental investigation in order to be confirmed as true symbiotic infections and not simple contamination. For example, the detection here of *bracovirus* DNA could suggest ecological interactions between moths and parasitoids, and the ability of the former to naturally resist parasitoid attack strategies. Altogether our study presents a method and produces material supporting testable hypotheses about the diversity

and nature of symbiotic interactions in those particular Lepidoptera species. With the availability of open access metagenomics databases, this field promises extensive and exciting opportunities to explore potentially hidden symbiotic diversity.

Aknowledgement

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). The authors acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We thank Marko Mutanen for providing the DNA extract for *Scoliopteryx libatrix*.

References

Altizer S.M., Oberhauser K.S., Brower L.P. 2000. Associations between host migration and the prevalence of a protozoan parasite in natural populations of adult monarch butterflies. Ecol. Entomol. 25:125–139.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bapatla K.G., Singh A., Yeddula S., Patil R.H. 2018. Annotation of gut bacterial taxonomic and functional diversity in Spodoptera litura and Spilosoma obliqua. J. Basic Microbiol. 58:217–226.

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

Bordenstein S.R., Bordenstein S.R. 2016. Eukaryotic association module in phage WO genomes from Wolbachia. Nat. Commun. 7:13155.

Boyd B.M., Allen J.M., Koga R., Fukatsu T., Sweet A.D., Johnson K.P., Reed D.L. 2016. Two Bacterial Genera, Sodalis and Rickettsia, Associated with the Seal Louse Proechinophthirus fluctus (Phthiraptera: Anoplura). Appl. Environ. Microbiol. 82:3185–3197.

Burke G.R., Normark B.B., Favret C., Moran N.A. 2009. Evolution and Diversity of Facultative Symbionts from the Aphid Subfamily Lachninae. Appl. Environ. Microbiol. 75:5328–5335.

Chen F., Coates B., He K.-L., Bai S.-X., Zhang T.-T., Wang Z.-Y. 2017. Effects of Wolbachia on mitochondrial DNA variation in populations of Athetis lepigone (Lepidoptera: Noctuidae) in China. Mitochondrial DNA Part A. 28:826–834.

Conord C., Despres L., Vallier A., Balmand S., Miquel C., Zundel S., Lemperiere G., Heddi A. 2008. Long-Term Evolutionary Stability of Bacterial Endosymbiosis in Curculionoidea: Additional Evidence of Symbiont Replacement in the Dryophthoridae Family. Mol. Biol. Evol. 25:859–868.

Dale C., Young S.A., Haydon D.T., Welburn S.C. 2001. The insect endosymbiont Sodalis glossinidius utilizes a type III secretion system for cell invasion. Proc. Natl. Acad. Sci. 98:1883–1888.

Dillon R.J., Dillon V.M. 2004. The gut bacteria of insects: Nonpathogenic Interactions. Annu. Rev. Entomol. 49:71–92.

Douglas A.E. 2015. Multiorganismal Insects: Diversity and Function of Resident Microorganisms. Annu. Rev. Entomol. 60:17–34.

Duplouy A., Brattström O. 2018. Wolbachia in the Genus Bicyclus: a Forgotten Player. Microb. Ecol. 75:255–263.

Duplouy A., Hornett E.A. 2018. Uncovering the hidden players in Lepidoptera biology: the heritable microbial endosymbionts. PeerJ. 6:e4629.

Duplouy A., Hursts G.D.D., O'neill S.L., Charlat S. 2010. Rapid spread of male-killing Wolbachia in the butterfly Hypolimnas bolina. J. Evol. Biol. 23:231–235.
Fukatsu T., Koga R., Smith W.A., Tanaka K., Nikoh N., Sasaki-Fukatsu K., Yoshizawa K., Dale C., Clayton D.H. 2007. Bacterial Endosymbiont of the Slender Pigeon Louse, Columbicola columbae, Allied to Endosymbionts of Grain Weevils and Tsetse Flies. Appl. Environ. Microbiol. 73:6660–6668.
Gavotte L., Henri H., Stouthamer R., Charif D., Charlat S., Bouletreau M., Vavre F. 2006. A Survey of the Bacteriophage WO in the Endosymbiotic Bacteria Wolbachia. Mol. Biol. Evol. 24:427–435.
Ilinsky Y., Kosterin O.E. 2017. Molecular diversity of Wolbachia in Lepidoptera: Prevalent allelic content and high recombination of MLST genes. Mol. Phylogenet. Evol. 109:164–179.
Itoh H., Aita M., Nagayama A., Meng X., Kamagata Y., Navarro R., Hori T., Ohgiya S., Kikuchi Y. 2014. Evidence of environmental and vertical transmission of Burkholderia symbionts in the oriental chinch bug, Cavelerius saccharivorus (Heteroptera: Blissidae). Appl. Environ. Microbiol. 80:5974–5983.
Itoh H., Jang S., Takeshita K., Ohbayashi T., Ohnishi N., Meng X.Y., Mitani Y., Kikuchi Y. 2019. Host–symbiont specificity determined by microbe–microbe competition in an insect gut. Proc. Natl. Acad. Sci. U. S. A. 116:22673–22682.
Jiggins F.M., Bentley J.K., Majerus M.E., Hurst G.D. 2001. How many species are infected with Wolbachia ? Cryptic sex ratio distorters revealed to be common by intensive sampling. Proc. R. Soc. London. Ser. B Biol. Sci. 268:1123–1126.
Jiggins, Hurst, Dolman, Majerus. 2000. High-prevalence male-killing Wolbachia in the butterfly Acraea encedana. J. Evol. Biol. 13:495–501.
Kaiwa N., Hosokawa T., Kikuchi Y., Nikoh N., Meng X.Y., Kimura N., Ito M., Fukatsu T. 2011. Bacterial Symbionts of the Giant Jewel Stinkbug Eucorysses grandis (Hemiptera: Scutelleridae). Zoolog. Sci. 28:169–174.
Kaushik S., Sharma K.K., Ramani R., Lakhanpaul S. 2019. Detection of Wolbachia Phage (WO) in Indian Lac Insect [Kerria lacca (Kerr.)] and Its Implications. Indian J. Microbiol. 59:237–240.
Kent B.N., Bordenstein S.R. 2010. Phage WO of Wolbachia: lambda of the endosymbiont world. Trends Microbiol. 18:173–181.
Khojandi N., Haselkorn T.S., Eschbach M.N., Naser R.A., DiSalvo S. 2019. Intracellular Burkholderia Symbionts induce extracellular secondary infections; driving diverse host outcomes that vary by genotype and environment. ISME J. 13:2068–2081.
Kikuchi Y., Hayatsu M., Hosokawa T., Nagayama A., Tago K., Fukatsu T. 2012. Symbiont-mediated insecticide resistance. Proc. Natl. Acad. Sci. U. S. A. 109:8618–8622.
Kikuchi Y., Yumoto I. 2013. Efficient Colonization of the Bean Bug Riptortus pedestris by an Environmentally Transmitted Burkholderia Symbiont. Appl. Environ. Microbiol. 79:2088–2091.
Louis F., Bezier A., Periquet G., Ferras C., Drezen J.-M., Dupuy C. 2013. The Bracovirus Genome of the Parasitoid Wasp Cotesia congregata Is Amplified within 13 Replication Units, Including Sequences Not Packaged in the Particles. J. Virol. 87:9649–9660.

Mason C.J., Raffa K.F. 2014. Acquisition and Structuring of Midgut Bacterial Communities in Gypsy Moth (Lepidoptera: Erebidae) Larvae. Environ. Entomol. 43:595–604.

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., Davis D.R., De Prins J., De Prins W., Epstein M.E., Gentili-Poole P., Gielis C., Hättenschwiler P., Hausmann A., Holloway J.D., Kallies A., Karsholt O., Kawahara A.Y., Koster S. (J C., Kozlov M. V, Lafontaine J.D., Lamas G., Landry J.-F., Lee S., Nuss M., Park K.-T., Penz C., Rota J., Schintlmeister A., Schmidt B.C., Sohn J.-C., Solis M.A., Tarmann G.M., Warren A.D., Weller S., Yakovlev R. V, Zolotuhin V. V, Andreas Z. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 1758:212–221.

Ohbayashi T., Itoh H., Lachat J., Kikuchi Y., Mergaert P. 2019. Burkholderia gut symbionts associated with European and Japanese populations of the dock bug Coreus marginatus (Coreoidea: Coreidae). Microbes Environ. 34:219–222.

Oishi S., Moriyama M., Koga R., Fukatsu T. 2019. Morphogenesis and development of midgut symbiotic organ of the stinkbug Plautia stali (Hemiptera: Pentatomidae). Zool. Lett. 5:16.

Rubin B.E.R., Sanders J.G., Turner K.M., Pierce N.E., Kocher S.D. 2018. Social behaviour in bees influences the abundance of Sodalis (Enterobacteriaceae) symbionts. R. Soc. Open Sci. 5:180369.

Salunke B.K., Salunkhe R.C., Dhotre D.P., Walujkar S.A., Khandagale A.B., Chaudhari R., Chandode R.K., Ghate H. V, Patole M.S., Werren J.H., Shouche Y.S. 2012. Determination of Wolbachia Diversity in Butterflies from Western Ghats, India, by a Multigene Approach. Appl. Environ. Microbiol. 78:4458–4467.

Sameshima S., Hasegawa E., Kitade O., Minaka N., Matsumoto T. 1999. Phylogenetic Comparison of Endosymbionts with Their Host Ants Based on Molecular Evidence. Zoolog. Sci. 16:993–1000.

Santos-Garcia D., Silva F.J., Morin S., Dettner K., Kuechler S.M. 2017. The All-Rounder Sodalis: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug Henestaris halophilus (Heteroptera: Henestarinae) and a Critical Examination of Its Evolution. Genome Biol. Evol. 9:2893–2910.

Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics.

Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods. 9:811–814.

Šochová E., Husník F., Nováková E., Halajian A., Hypša V. 2017. Arsenophonus and Sodalis replacements shape evolution of symbiosis in louse flies. PeerJ. 5:e4099.

Tagami Y., Miura K. 2004. Distribution and prevalence of Wolbachia in Japanese populations of Lepidoptera. Insect Mol. Biol. 13:359–364.

Tanaka K., Furukawa S., Nikoh N., Sasaki T., Fukatsu T. 2009. Complete WO Phage Sequences Reveal Their Dynamic Evolutionary Trajectories and Putative Functional Elements Required for Integration into the Wolbachia Genome. Appl. Environ. Microbiol. 75:5676–5686.

Toju H., Fukatsu T. 2011. Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. Mol. Ecol. 20:853–868.
Wang G.H., Sun B.F., Xiong T.L., Wang Y.K., Murfin K.E., Xiao J.H., Huang D.W. 2016a. Bacteriophage WO Can Mediate Horizontal Gene Transfer in Endosymbiotic Wolbachia Genomes. Front. Microbiol. 7:1–16.
Wang N., Jia S., Xu H., Liu Y., Huang D. 2016b. Multiple Horizontal Transfers of Bacteriophage WO and Host Wolbachia in Fig Wasps in a Closed Community. Front. Microbiol. 7:1–10.
Werren J.H., Windsor D., Guo L. 1995. Distribution of Wolbachia among neotropical arthropods. Proc. R. Soc. London. Ser. B Biol. Sci. 262:197–204.
Wood D.E., Salzberg S.L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15:R46.
Xu P., Liu Y., Graham R.I., Wilson K., Wu K. 2014. Densovirus Is a Mutualistic Symbiont of a Global Crop Pest (Helicoverpa armigera) and Protects against a Baculovirus and Bt Biopesticide. PLoS Pathog. 10:e1004490.
Zahiri R., Holloway J.D., Kitching I.J., Lafontaine J.D., Mutanen M., Wahlberg N. 2012. Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). Syst. Entomol. 37:102–124.