Branching out: the role of host plants in the diversification of leaf-mining moths

Doorenweerd, C.

Publication date
2016
Document Version
Final published version

ISBN 978-3-16-148410-0

Branching out: the role of host plants in the diversification of leaf-mining moths

C. Doorenweerd 2016

Branching out: the role of host plants in the diversification of leaf-mining moths

Camiel Doorenweerd

Branching out: the role of host plants in the diversification of leaf-mining moths

Camiel Doorenweerd

Doorenweerd C, 2016.
Branching out: the role of host plants in the diversification of leaf-mining moths
PhD thesis, University of Amsterdam, The Netherlands

Cover design by Frank Stokvis, adult moth watercolours © Roland Johansson and The Encyclopedia of the Swedish Flora and Fauna, with permission from The Swedish Species Information Center.
Layout by Niko Korenhof, Naturalis Biodiversity Center
Printed by Amsterdam University Press, Amsterdam
ISBN: 978-94-91407-40-6

This research was carried out at Naturalis Biodiversity Center, Leiden and the Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam. The research was funded by the Infrastructure Reinforcement fund (FES) of the Dutch Ministry of Finance.

Branching out: the role of host plants in the diversification of leaf-mining moths

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 25 oktober 2016, te 14.00 uur

door Camiel Doorenweerd

geboren te Middelburg

Promotiecommissie

Promotor:	prof. dr. S.B.J. Menken	Universiteit van Amsterdam
Copromotor:	dr. E.J. van Nieukerken	Naturalis Biodiversity Center
Overige leden:	prof. dr. P.H. van Tienderen	Universiteit van Amsterdam
	prof. dr. J.C. Biesmeijer	Universiteit van Amsterdam
	prof. dr. N. Wahlberg	Lund University
	prof. dr. M. Schilthuizen	Universiteit Leiden
	dr. A.T. Groot	Universiteit van Amsterdam
	dr. J.A.J. Breeuwer	Universiteit van Amsterdam
	dr. V.S.F.T. Merckx	Naturalis Biodiversity Center

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Voor mijn ouders

Table of contents

Chapter 1	Introduction	9
Chapter 2	DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-alpha: two are better than one in recognising cryptic species *Contributions to Zoology 2012*	29
Chapter 3	A global phylogeny of leafmining *Ectoedemia* moths (Lepidoptera: Nepticulidae): exploring host plant family shifts and allopatry as drivers of speciation *PLoS One 2015*	63
Chapter 4	A revised checklist of Nepticulidae fossils (Lepidoptera) indicates an Early Cretaceous origin *Zootaxa 2015*	91
Chapter 5	Phylogeny, classification and divergence times of pygmy leafmining moths (Lepidoptera: Nepticulidae): the earliest lepidopteran radiation on Angiosperms? *Accepted for publication in Systematic Entomology*	147
Chapter 6	Increased diversification rates across lineages of leaf-mining moths after entering the new adaptive zone of the temperate region *To be submitted*	183
Chapter 7	Epilogue	207
Chapter 8	Summary / Samenvatting	213
Chapter 9	Word of thanks	223
CV and publications		227

1

Introduction

Macroevolutionary patterns and speciation

Macroevolutionary patterns

Darwin's work (Darwin, 1859) marked a revolution in our thinking on evolution. Despite over 150 years of evolutionary research, however, the factors that drive the key process of speciation, the splitting of one species into two, still largely remain elusive. There have been roughly two different approaches to understanding the process of speciation. The microevolutionary approach works from an ecological perspective and focuses on the interactions between populations of, commonly, a single species and its biotic and abiotic environment, how these interactions may result in restricted gene-flow and eventually lead to reproductively isolated species (Kozak et al., 2015; Brakefield et al., 2009; Day et al., 2008; Mayr, 1963). Potential microevolutionary drivers can be tested in natural situations or in a controlled laboratory environment (Matsubayashi et al., 2010; Ohshima, 2008). The macroevolutionary approach, on the other hand, studies the patterns that emerge when the relationships between large groups of species are examined, extending long back into evolutionary history, preferably as close to the origin of the group under study as possible. This field has its roots in Palaeontology, where such patterns first became available from studying the fossil record (Simpson, 1953). With the advent of molecular methods to reconstruct the Tree of Life (Letunic & Bork, 2007), the field has grown wider, and as fossil and molecular data are being integrated, slowly we are able to get to a more complete image of the evolutionary history of life on earth (Condamine et al., 2016; Heikkilä et al., 2015; Sohn et al., 2015; Hedges et al., 2006; Grimaldi & Engel, 2005).

The origin and diversification of Lepidoptera

Diversification is a loosely defined term which is increasingly used in macroevolutionary studies and commonly indicates that the study analysed the sum of the speciation and extinction events for a group of species over a period of time (Condamine et al., 2016; Tank et al., 2015; Toussaint et al., 2015; Wahlberg et al., 2013). Diversification studies often focus on understanding why some of the groups that exist today have many more species than others. More than half of the world's biodiversity consists of insects (Mitter et al., 1988), and the moths and butterflies, the order Lepidoptera, is one of the four largest insect orders. The Lepidoptera are the largest radiation that is almost entirely associated with flowering plants (Wiens et al., 2015). A staggering number of over 155,000 Lepidoptera species have been described, and likely a similar number awaits description or discovery (Nieukerken et al., 2011). There is no full consensus yet on their age of origin, but most estimates agree that they diverged from their sister-group, the caddisflies (Trichoptera), in the Late Triassic, over 200 million years ago (Misof et al., 2014; Wahlberg et al., 2013). For quite some time, the group was not very diverse, and it was not until the flowering plants, the Angiosperms, started to become a dominant part of terrestrial ecosystems that the Lepidoptera diversification began to accelerate, i.e. some 100 million years ago, during the Cretaceous (Condamine et al., 2016; Regier et al., 2015; Wahlberg et al., 2013).

Around that time, we also find the first evidence of leaf-mining moths in the fossil record (this thesis, chapter 4; Labandeira et al., 1994).

Leaf-mining moths

The leaf-mining guild

Leaf-miners form a guild, an ecologically defined group of insects, which are characterized by having larvae that feed within plant tissue and create channels that are shut off from the outside by the epidermal layers of the plant (Fig. 1) (Hering, 1951). Leaf mining is not limited to Lepidoptera, but can be found in all holometabolous insect orders and leaf-miners have in many lineages diversified into a large number of species; there are for example over 7,500 species of sawflies (Hymenoptera: Tenthredinidae) (Davis et al., 2010) and over 2,500 species of leaf-miner flies (Diptera: Agromizidae) (Spencer, 2012). Although the name 'leaf-miners' suggests that they only feed in plant leaves, there are variations to the theme, as well as intermediate forms. Leaf-miners may for instance also use petioles, flowers or fruits. Insects that feed internally in other parts of the plants, such as stems and roots, or those that create galls, are usually classified separately (Donovan et al., 2014; Knor et al., 2012; Hering, 1951). Feeding within plant tissue requires specific evolutionary adaptations. On the one hand the enclosed environment protects against desiccation and from certain predatory threats, on the other hand it imposes physical restrictions and individuals are fully exposed to the plant's defence mechanisms. As a result of this close association, leaf-miners have a high degree of host specialization (Menken et al., 2010).

Figure 1: Leaf-mines created by an undescribed species of *Ectoedemia* on beech (*Fagus crenata*) in Japan.

Host fidelity

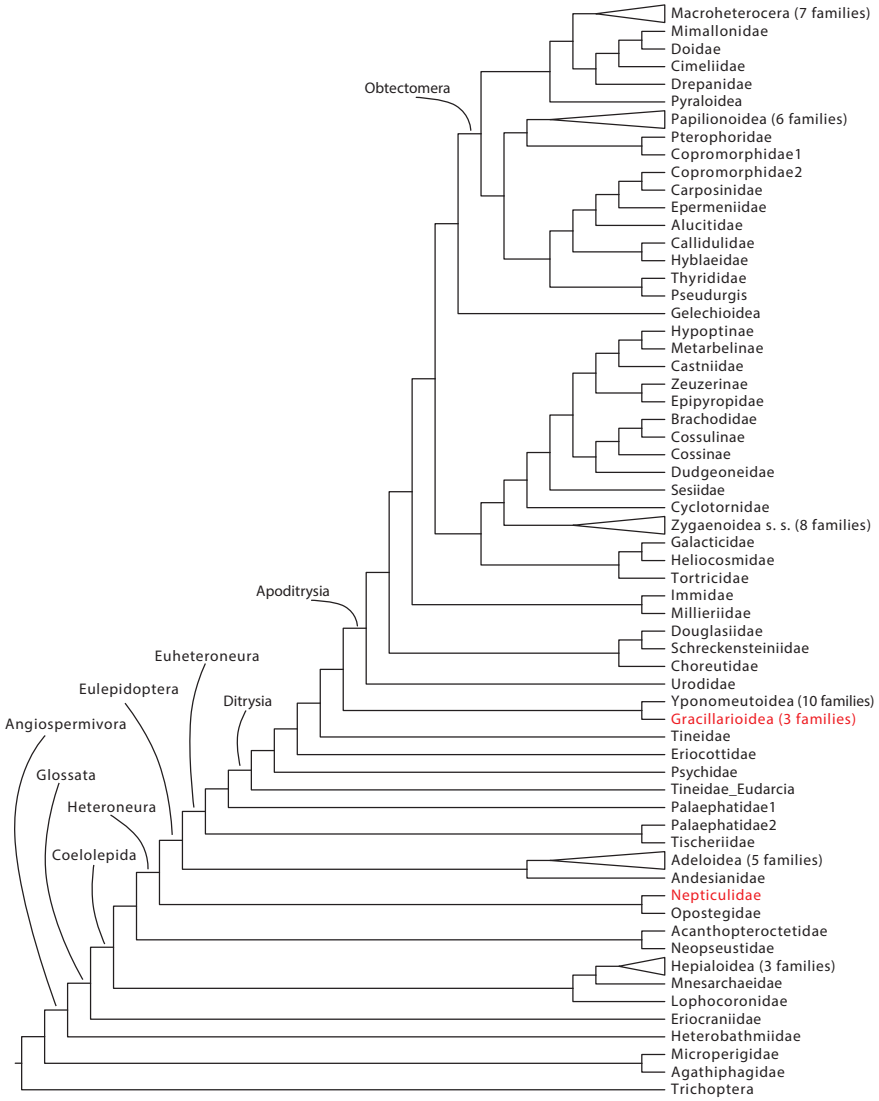
Typically, leaf-mining larvae cannot change their host, which makes the oviposition choice of their mother especially important. Most leaf-mining species are monophagous, they are restricted to host plants of a single genus, some are oligophagous, restricted to plants within a single family, and there are only exceptional cases of polyphaghy, where they can feed on plants from different families. Also, related leaf-miners tend to feed on related host plants. These patterns are found across all groups of leaf-miners (this thesis; Winkler & Mitter, 2008; Nyman et al., 2006). Within Lepidoptera, the degree of host specialization is generally higher in early diverging lineages, and always higher in leaf-mining groups (Menken et al., 2010). The pygmy leaf-mining moths, Nepticulidae, and the leaf blotch mining moths, Gracillariidae, are two families that are found among the earlier split-offs in the Lepidoptera tree of life, i.e. among the non-apodytrisian lineages (Fig. 2) (Regier et al., 2013). Both families are species rich, commonly found across all types of terrestrial ecosystems and almost all of the species are leaf miners. Within Gracillariidae, the subfamily Lithocolletinae in particular bears striking resemblance to Nepticulidae when it comes to host family preferences, suggesting that there have been common evolutionary drivers during their diversification.

Nepticulidae

The lepidopteran family Nepticulidae, commonly called pygmy leaf-mining moths because they have wingspans of less than 10 mm, represents probably the earliest radiation of moths on angiosperms (see Fig. 2) (this thesis; Regier et al., 2015; Kristensen & Skalski, 1998). Nepticulidae comprise 854 named species and an estimated total of 2,500 species worldwide (Nieukerken et al., 2016a). Nepticulidae are invariably specialists and only a few species are oligophagous, such as *Stigmella corylifoliella* or polyphagous, such as *Ectoedemia atricollis*. The majority of the hosts are woody plants belonging to the rosid I clade (Fig. 3) (Menken et al., 2010; APG III, 2009). Out of the 22 described genera (Nieukerken et al., 2016a; Nieukerken et al., 2016b; this thesis, chapter 5), the genera *Ectoedemia* and *Stigmella* together make up for approximately 60% of the species and are particularly rich on northern hemisphere hardwood trees. The most commonly used host plant families in these two genera are all ecologically dominant groups in the temperate region, viz. the beeches and oaks (Fagaceae), willows and poplars (Salicaceae), the rose family (Rosaceae), the legumes (Fabaceae) and the birch family (Betulaceae). There is a recurring pattern in which morphologically defined species groups are found on a single host plant family, but the species within such a group are specialised on separate host plant species (van Nieukerken & Johansson, 2003; Johansson et al., 1990; Kemperman et al., 1985).

Lithocolletinae

There are 611 species described in 11 genera within the Gracillariidae subfamily Lithocolletinae, but most of the genera have few species: *Phyllonorycter* and *Cameraria* together comprise approximately 90% of all the species (De Prins & De Prins, 2016). The moths are generally slightly larger than Nepticulidae, but still



small and their wingspans do not exceed 15 mm. The colours of the wings often include different shades of orange and dark and light bands (Fig. 4). Lithocolletinae share many of the host families with *Stigmella* and *Ectoedemia*, displaying again the pattern of species groups often feeding on a single host family (Lopez-Vaamonde et al., 2006; Lopez-Vaamonde et al., 2003), and individual species being highly host-specific. There are also differences, for example in the diversity distribution of *Phyllonorycter*, *Cameraria*, *Stigmella* and *Ectoedemia*. The distribution of *Cameraria* is less cosmopolitan than the other genera: the main diversity occurs in North America, less so in Asia and there is only a single European species (*Cameraria ohridella)*, which is currently a well-known pest on

◀ Figure 2: Lepidoptera tree of life based on 483 taxa and 19 genes, reworked from (Regier et al., 2015; van Nieukerken et al., 2011; Regier et al., 2013). Taxa studied in this thesis are indicated in red. Although there is now a broad consensus on the back-bone of the Lepidoptera phylogeny, several groups require taxonomic revision, evident from the non-monophyletic clades in this tree. Nepticulidae together with Opostegidae are joined in Nepticuloidea and the first clade to split-off in the heteroneuran clade. The term Heteroneura refers to the hindwing venation with Sc and R veins fused beyond a short distance from the wing base, and vein Rs is unbranched, and more morphological characters support this clade (Nielsen & Kristensen, 1996). Gracillarioidea together with Yponomeutoidea is part of the Ditrysia clade, which can easily be recognised by the separate copulatory orifice and ovipore. Gracillarioidea is the last clade to branch off before the Apoditrysia clade, which includes 93% of all described species (van Nieukerken et al., 2011), and is composed of the Roeslerstammiidae, Bucculatricidae and Gracillariidae. Lithocolletinae is one of eight subfamilies within Gracillariidae and includes ca. a third of the described gracillariid species (De Prins & De Prins, 2016; Kawahara et al., 2016).

horse-chestnut trees (*Aesculus hippocastanum*) (Lees et al., 2011; Gilbert et al., 2005). *Phyllonorycter* and *Stigmella* are both cosmopolitan, with the exclusion of Antarctica, and *Ectoedemia* has its main diversity in the West-Palaearctic. The molecular phylogenetics of Gracillariidae have been studied more extensively prior to this thesis than those of Nepticulidae. Although some of these studies have focussed on inter-generic relationships (Kawahara et al., 2011; Lopez-Vaamonde et al., 2006), there has also been an influential study with a time-calibrated species-level phylogeny of the genus *Phyllonorycter* (Lopez-Vaamonde et al., 2006). The main conclusion from this study was that the leaf-miners most likely diversified millions of years after their hosts.

Taxonomy and systematics

Species concepts

Studying speciation requires a clear concept that can be used to define where one species ends, and the next species begins. Mayr's biological species concept is generally accepted throughout science and adopted in most biology text books: "a species is a group of organisms that can produce fertile offspring together, and are reproductively isolated from other groups" (Mayr, 1963). Although seemingly simple and clear, there are many practical and some theoretical issues with this concept (Coyne & Orr, 2004; Mayr, 1963). In many cases there is little knowledge to what extent individuals can and do reproduce with each other, for example because we only know about a few specimens or they are geographically isolated, and things become increasingly complicated with asexually reproducing organisms. Also in Nepticulidae and Lithocolletinae, there are some parthenogenetically reproducing species, for example *Stigmella microtheriella*, *Ectoedemia argyropeza* and *Phyllonorycter emberizaepenella*. There are many alternative species concepts available, including ecological, evolutionary, phylogenetic, genotypic cluster and phenetic concepts [for an overview and definitions see (De Queiroz, 2007)]. In practice, taxonomists often combine properties of different concepts into a 'total evidence' approach that should be a

reliable proxy for the biological species concept. For leaf-mining moths, this involves combining morphological characters, interpreted in a Hennigian cladistic sense (Hennig, 1966), with molecular data, commonly DNA barcodes (this thesis, chapter 2; Hebert et al., 2003), distribution information and life history information,

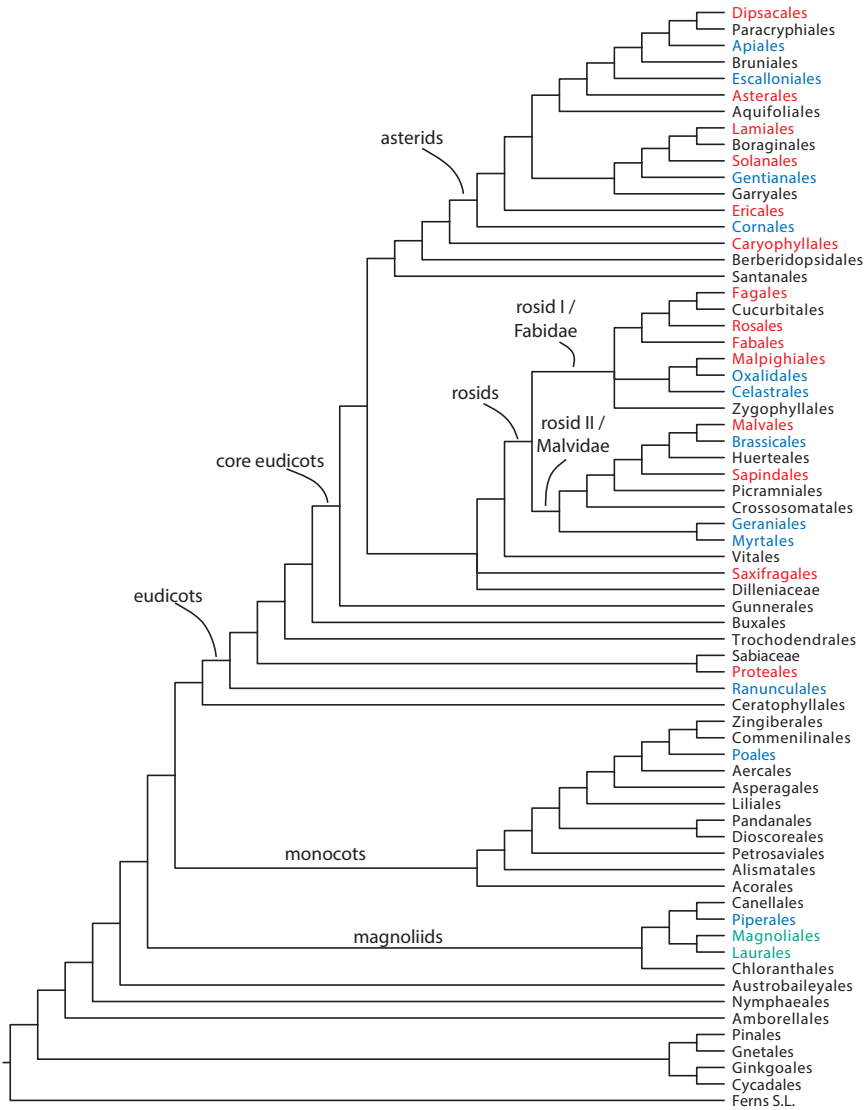


Figure 3: Angiosperm plants tree of life at the order level, reworked from the APG III classification (APG III, 2009). Nepticulidae and Lithocolletinae can be found feeding on many of these plant orders: red taxa are used by both, green taxa by Lithocolletinae only and blue taxa by Nepticulidae only. Although many orders are coloured, Nepticulidae and Lithocolletinae are predominantly specialised on plants in the rosid I clade, particularly the orders Fagales (on Fagaceae and Betulaceae), Malphigiales (on Salicaceae), Rosales (on Rosaceae) and Fabales (on Fabaceae).

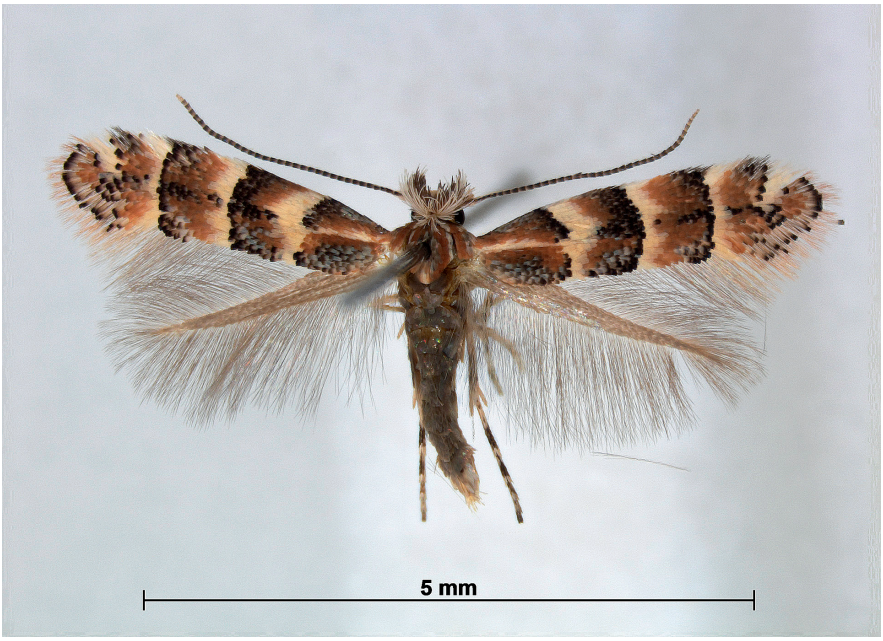


Figure 4: A pinned and spread male *Phyllonorycter pumilae* from South Korea. Leg. Erik J. van Nieukerken. RMNH.5007980.

which includes host plant associations. However, even when all this information is combined, the interpretation of the information by the taxonomist to define a species remains somewhat arbitrary, and should always be viewed as a hypothesis.

Molecular taxonomy

Since the early 1980's, molecular techniques have increasingly been applied in systematic studies to supplement morphological data. Early studies used gel-electrophoreses on allozymes, variant forms of enzymes, to distinguish between species and indicate relatedness between species and higher systematic ranks. Such methods have also been applied to Nepticulidae and brought new insight in the relations between species that were difficult to recognise morphologically (e.g. van Driel & Menken, 1988). Although still sometimes used today (Richardson et al., 2012), most scientists prefer to study DNA directly to prevent the unpredictable potential effects that, among others, regulation and translation steps can have on the outcome of allozyme studies. There is now a large range of different molecular methods available and the selection of the one that is appropriate involves matching the research question with the resolution that is required, the age, quality and number of available samples and the available time and budget. Since its official launch in 2003, DNA barcoding has proven its worth as a molecular method for species recognition and delimitation, provided that it is used with care (Mutanen et al., 2016; this thesis, chapter 2). It uses a universally agreed upon relatively short

sequence of DNA, for animals a 658 base-pair section of the mitochondrial cytochrome c oxidase I (COI) gene. The sequence can be determined relatively cheaply with different methods, including Sanger sequencing and several Next-Generation-Sequencing methods (e.g. Prosser et al., 2016). The DNA barcode sequences are made available to the global community through the online database BOLD (Ratnasingham & Hebert, 2007) and are linked to voucher specimens that are stored in Natural History collections to allow for future reinterpretations. As such, the DNA barcodes that are available in BOLD provide an invaluable way to digitize taxonomic information that is present in Natural History collections globally and provides ways for scientists to greatly enhance their sampling in taxonomic, as well as phylogenetic studies, which is exemplified by the chapters in this thesis.

Molecular phylogenetics

In order to study diversification patterns, the relationships between taxa should be studied from present time (i.e. species delimitation) to the origin of the group, which may go back hundreds of millions of years (Misof et al., 2014; Grimaldi & Engel, 2005). Reconstructing reliable phylogenetic trees where the evolutionary relationships between species, genera and families are inferred requires more data than COI sequences alone can provide (Wahlberg & Wheat, 2008). In general, more data is better and there seems to be no upper limit to the amount of sequence information that can be used, other than computational and budgetary limitations (Kawahara et al. 2011; Zwick et al. 2011; Regier et al., 2009). In practice, the taxon sampling and number of genetic markers used should be balanced to include as many representatives as possible and at the same time resolve all phylogenetic relationships with high statistical support. The selection of taxa within a given group depends on the research question. For systematic purposes the aim is often to include representatives for all rankings above species level and disregard multiple specimens per species and closely related species. For diversification studies, however, a random taxon sampling is often required across different systematic levels and the aim should be to include as many species as possible. The 'wet-work' for molecular biology has seen much automation in recent years and as the production of large amounts of sequence data has become cheaper and faster, the most challenging aspects of molecular biology now lie with bioinformatics and computational resources to calculate phylogenies from the wealth of sequence data. Ideally, all possible tree configurations are compared and weighed - the heuristic approach - but the number of possible combinations increases exponentially with increasing data and taxa, and likewise the computation time. Two commonly applied algorithms to circumvent this are maximum likelihood and Bayesian statistics. They each apply different methodologies to sample from the possible combinations and calculate which is the most likely phylogenetic tree. Neither method is perfect, but the combination provides a good indication of the robustness of the inferred phylogeny (Douady et al., 2003). Cloud computing, such as the OpenStack platform implemented at Naturalis, further offers ways to efficiently allocate computational resources, and recent developments in phylogenetic software focus on using multiple cores (Aberer et al., 2014; Stamatakis, 2006). Combining the latest software developments with the latest

hardware developments and high taxon coverage from the natural history collections enables the inference of large phylogenies, suitable for detailed diversification analyses.

Divergence times
Once a phylogeny has been reconstructed, the divergence times between different taxa can be estimated by molecular dating, where geological ages are assigned to particular nodes. These ages can be derived from fossil evidence that provides a minimum age, or geological events (such as the break-up of Gondwana). As more calibration points are included, the reliability of the divergence time estimations increases (Magallón et al., 2013; Warnock et al., 2012). Lepidoptera have a poor fossil record compared to other insects, owing to the delicate structure of the wings and body that easily decomposes. The two most important types of fossilized material for Lepidoptera are therefore Amber entombments, in which insects have been stuck in plant-produced resin which hardened over time (Fig. 5), and leaf mine impressions, where the plant leaves that included mines have fossilized or left an impression in rock (this thesis, chapter 3; Sohn et al., 2015). Especially for Nepticulidae, there is a relatively rich record of fossil leaf-mines, although the variation of leaf-mine morphologies makes it difficult to reliably assign fossil leaf-mines to particular clades. The fossil leaf-mine record for Gracillariidae has been summarized in an overview (Sohn et al., 2012) but requires reinterpretation following the latest phylogenetic insights (Kawahara et al., 2016; this thesis, chapter 6). Alternatively to calibrating nodes with fossil evidence or geological events directly, external calibrations can be applied, where the estimated divergence from a deeper-level phylogeny with overlapping taxa is used (this thesis, chapter 6).

Candidate drivers of speciation

Theories on speciation
There are many theories on speciation, which each predict different evolutionary drivers. It is universally accepted that allopatric speciation is one of the most important processes for animals, where populations become genetically isolated due to physical barriers (e.g. rivers, mountains, oceans) and diverge over time (Mayr, 1963; 1947). To what extent there may be sympatric speciation, where populations are not physically separated, and which drivers may be involved, has been widely debated (for reviews see Tilmon, 2008; Bolnick & Fitzpatrick, 2007). The adaptive radiation theory (Simpson, 1953) is one of the more general theories, not limited to plant-herbivore relationships, and suggests that in any situation where a lineage is able to colonize and adapt to a new resource-space this will be followed by increased rates of speciation and adaptation to all the different ecological niches that have become within reach. In the plant-herbivore framework, a new resource space could translate to shifts to a novel host plant family (this thesis, chapter 3; Schluter, 2000). The specific interactions and repeated patterns of host use in herbivorous insects have led to multiple theories that propose co-evolution between the plants and insects. The classic ‘escape and radiate’ theory

(Ehrlich & Raven, 1964) suggested, based on patterns of host use in butterflies, that co-evolution between the hosts and the insects had been the driving force of evolution in a reciprocal co-evolutionary process: the plants would diversify their chemical defences, releasing them from herbivore pressure, and the insects would subsequently adapt to all the new forms and specialise. It is important to note that co-evolution can be interpreted in different ways. With strict co-evolution, the plants and insects exert selective pressures on each other, with diffuse co-evolution or community evolution the plants influence the evolution of the insects, but not vice versa (Nyman et al. 2012). Co-evolution in either sense can lead to cospeciation, sometimes referred to as co-cladogenesis or co-diversification, where the evolution of host plants and insects has been contemporaneous and as a result the phylogeny of the plants will mirror that of the insects. However, cospeciation can also occur without any co-evolution; as a result from shared allopatric situations. In herbivorous insects, cospeciation patterns are most evident in groups that have obligate mutualisms that involve pollination, for example in fig wasps (Agaonidae) and figs (*Ficus*), yucca moths (Prodoxidae) and yucca trees (*Yucca*), and *Epicephala* moths and Phyllantheae trees, but even there the mirror images are not perfect and patterns are (partly) better explained by host-shift speciation than cospeciation (Kawakita 2010; Page, 2003; Pelmyr et al., 1996). Another explanation for mirroring phylogenies of plant hosts and insects comes from the 'sequential evolution' theory, where insects track the evolution of their hosts by following a pattern of resource-similarity, but this evolution occurred much later and independently of the evolution of the hosts. However, resource-similarity may be different in the perception of the insects than phylogenetic similarity of the plants (Jermy, 1993). The recognition of host plants has been suggested as a major factor in driving insect diet breadth and specialisation and speciation, with examples supporting this scenario in ermine moths (Yponomeutidae): the host shifts of *Yponomeuta rorrella* (Lepidoptera: Yponomeutidae) and *Yponomeuta malinellus* appear related to the ability to recognize particular chemical compounds in the hosts (Menken et al., 1992). Theories suggesting yet different drivers and mechanisms for the evolution of herbivorous insects include the 'oscillation theory' (Janz and Nylin, 2008), the 'phylogenetic constraint theory' (Price et al., 1990) and the 'intermediate resource theory' (Nyman et al., 2010). For each there is some evidence from the groups for which they have been tested, but they make overlapping predictions and none have been tested comparatively across different groups. If, and to what extent, the evolution of the insects and herbivorous insects has been contemporaneous is a central issue in many of the co-evolutionary theories. The patterns that have been uncovered by time-calibrated phylogenies in herbivorous insects are complex and possibly involved different drivers and mechanisms for different groups or through time (this thesis, chapters 5 and 6; Winkler & Mitter, 2009; McKenna, 2009; Lopez-Vaamonde et al., 2006; Labandeira et al., 1994).

Common drivers?

The most diverse lineages within Nepticulidae and Lithocolletinae are found in temperate regions, on woody host plants that are commonly present in these

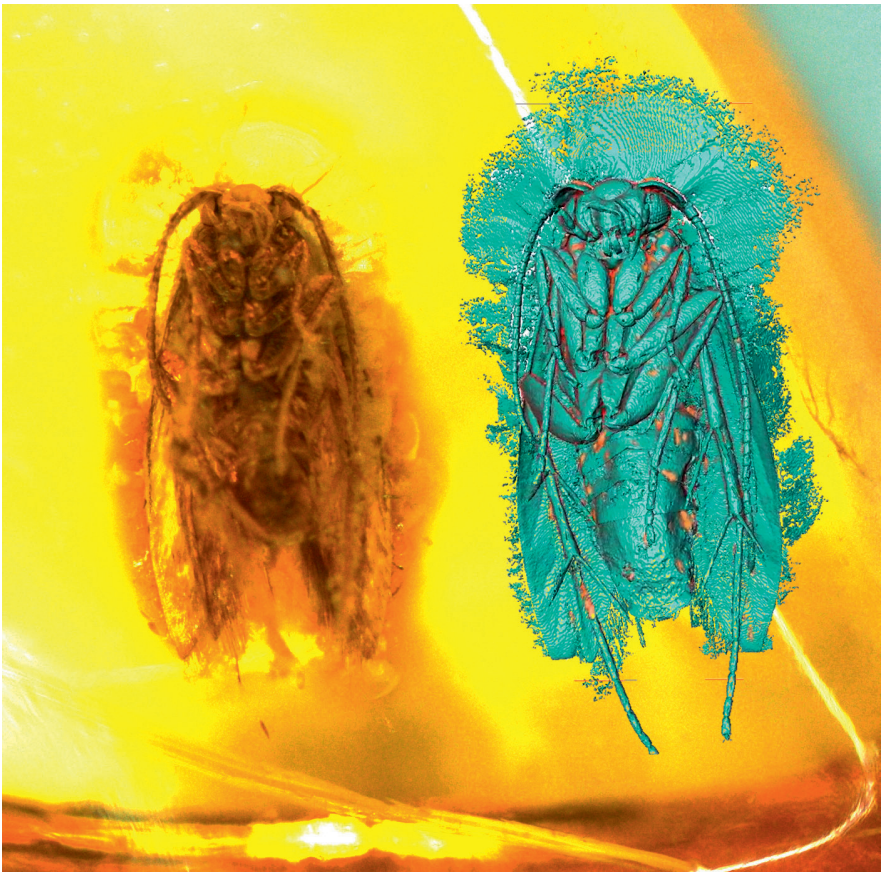


Figure 5: A nepticulid moth, described as *Bohemannia butzmanni*, that has been trapped in Amber resin 43–45 million years ago (Fischer, 2013). The metallic blue insert on the right is a 2D rendering of the preliminary result of a 3D micro-CT scan of the specimen that visualises more characters than observable by light microscopy (Photograph of the Amber stone by Erik van Nieukerken, CT scan by Dirk van der Marel, 2D and 3D rendering with help from Martin Rücklin). The size of the moth from head to tip of the abdomen is about 1.5 mm.

ecosystems. The diversity of leaf miners in combination with clear host relationships makes them a suitable group for studying potentially host driven evolution (Feder et al., 2005; Leppänen et al., 2012; Winkler et al., 2009). The green world in which these insects live, the colonized or candidate host plants, has also evolved and diversified throughout the past 100 million years. Deciduousness in broad-leaved trees has evolved in disturbed environments during the late Cretaceous and was favoured by conditions after the K-Pg mass extinction event, 65 million years ago (Donovan et al., 2014). Widespread deciduous forests in the northern hemisphere spread especially during the gradual cooling and drying of the climate in the Paleogene (Wolfe, 1975). How leafmining moths have adapted to the changing situations of host plants and climate over millions of years, by

gradual adaptation to deciduousness within the same host group or by massive host shifts, and establishing whether host plants have been the driving factor for leaf miner diversification, whether these have been a common factor for all groups or whether entirely different factors have been the most important are the main aims of this thesis.

Thesis outline

In this thesis, I studied the diversification patterns of two groups of leafmining Lepidoptera: Nepticulidae and Lithocolletinae (Gracillariidae). Both are found on all continents except Antarctica and are species-rich, but to different extents in different lineages and their centres of diversity are in different biogeographic regions, even though they predominantly feed on similar host plant families. The research I performed focussed on comparing their species-level phylogenetic diversification patterns, based on a dataset that included the majority of their global diversity, to understand common factors that have driven their evolution.

The first step in any study involving species-level phylogenies is delimiting the species and establishing reliable species-concepts. In part the diversity of our focal groups has been described in a Linnean sense, but a significant part of the diversity remains unlinked to described species, unnamed or undiscovered. The molecular approach of DNA barcoding uses a section of 658 base-pairs of the mitochondrial COI gene to efficiently gather information on genetic differentiation, which can then be combined with data on morphology, distribution and life history for a 'total evidence' taxonomy approach. We tested such an approach in **Chapter 2** where we include a detailed analysis of about 90% of all described species of *Ectoedemia* (Nepticulidae). A further 20% of the dataset consisted of undescribed species. We paid particular attention to complexes of cryptic species, which are often not only difficult to recognize morphologically, but also molecularly. Next to using the official DNA barcoding gene COI, we tested the applicability of a second, nuclear, gene for species delimitation and recognition, Elongation Factor 1 alpha. At the time of publication of this chapter, *Ectoedemia* was still defined in a broader sense, so we focussed on the subgenus *Ectoedemia*, which was later raised to full genus based on the results in **Chapter 5.** The approach outlined in chapter 2 was applied to all other groups that were included in later chapters of the thesis.

Moving from datasets with DNA barcodes to robust phylogenies required sequencing more genes. To this end, primer sets for Lepidoptera for eight genes (Wahlberg & Wheat, 2008) were tested in Turku in the lab of Prof. Wahlberg. Several of these primer combinations worked well, and with several other primer combinations from literature I had the methods available to sequence up to eight genes. Because the molecular facilities of Naturalis Biodiversity Center had been adapted to high-throughput sequencing, the molecular work could quickly progress. The first group for which we were able to reconstruct a detailed molecular phylogeny was again the subgenus *Ectoedemia*, on which we mapped host-plant

and biogeographic data using ancestral state reconstructions, and explored to what extent they could explain the observed diversification pattern in **Chapter 3**.

As multiple generic phylogenies grew with species, we began searching for published fossil records of leafminers to calibrate the phylogenies in time using molecular dating. Realising that it would be impossible to find sufficient fossils that could be assigned within genera, we started working on phylogenies for the larger groups: Nepticulidae, and Lithocolletinae. Starting with Nepticulidae, including fossil data proved less straightforward than anticipated, as we discovered many discrepancies in the published work and different approaches to describing fossils made it difficult to assign fossils to extant taxa. In **Chapter 4** we review all Nepticulidae fossils, over 70 fossil records with usually multiple exemplars, in a fruitful collaboration with colleagues at the Smithsonian Institution.

Chapter 5 presents a phylogeny, classification and divergence time estimates for Nepticulidae, including revised systematic ranks and molecular dating to determine the age of the different genera. This work includes 339 species from 22 nepticulid genera and constitutes a major advance in our understanding of the diversification of the family. Parallel to this chapter, in separate papers not included in this thesis (Nieukerken et al., 2016a; Nieukerken et al., 2016b), three new Nepticulidae genera were described and a new catalogue for Nepticulidae was submitted for publication.

A synthesis of my PhD work follows in **Chapter 6**, where we comparatively analyse shifts in the rate of speciation and their correspondence with host plant family shifts and other candidate drivers of speciation in the reconstructed species-level phylogenies of Nepticulidae and Lithocolletinae. At this point we included 609 species of Nepticulidae from 20 genera, and 345 species of Lithocolletinae from 11 genera. We used Bayesian inference to test along which lineages there have been shifts in the rate of diversification and inferred overall speciation rate patterns. Using our data on host plant family shifts allowed us to test if such shifts are followed by shifts in speciation rates, which would follow from the adaptive radiation hypothesis (Simpson 1953) or if other explanations may be preferred. Also, the results tell us if speciation occurs at a steady pace, as several studies have recently indicated (Hedges et al., 2015; Venditti et al., 2010), or that there are bursts of speciation, which are only revealed when data with the right resolution is available. In **Chapter 7**, I reflect on the findings of this thesis as a whole and discuss them from a broader perspective.

References

Aberer, A.J., Kobert, K., & Stamatakis, A. (2014) ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

APG III: Bremer, B., Bremer, K., Chase, M.W., Fay, M.F., Reveal, J.L., Soltis, D.E., Soltis, P.S., Stevens, P.F., Anderberg, A.A., Moore, M.J., Olmstead, R.G., Rudall, P.J., Sytsma, K.J., Tank, D.C., Wurdack, K., Xiang, J.Q.Y., & Zmarzty, S. (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, **161**, 105–121.

Bolnick, D.I. & Fitzpatrick, B.M. (2007) Sympatric Speciation: Models and Empirical Evidence. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 459–487.

Brakefield, P.M., Beldade, P., & Zwaan, B.J. (2009) The African butterfly *Bicyclus anynana*: a model for evolutionary genetics and evolutionary developmental biology. *Cold Spring Harbor Protocols*, **4**, 1–9.

Condamine, F.L., Clapham, M.E., & Kergoat, G.J. (2016) Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Nature Scientific Reports*, **6**, 19208.

Coyne, J.A. & Orr, H.A. (2004) Speciation, pp 545. Sinauer Associates, Inc.

Darwin, C. (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *Nature*, **5**, pp. 502.

Davis, R.B., Baldauf, S.L. & Mayhew, P.J. (2010) The origins of species richness in the Hymenoptera: insights from a family-level supertree. *BMC Evolutionary Biology*, **10**, 1–16.

Day, J.J., Cotton, J.A., & Barraclough, T.G. (2008) Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PLoS ONE*, **3**, e1730.

De Prins, J. & De Prins, W. (2016) *Global Taxonomic Database of Gracillariidae (Lepidoptera)*. Belgian Biodiversity Platform. URL: http://www.gracillariidae.net [accessed on: 10 July 2016].

De Queiroz, K. (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879–886.

Donovan, M.P., Wilf, P., Labandeira, C.C., Johnson, K.R. & Peppe, D.J. (2014) Novel insect leaf-mining after the end-Cretaceous extinction and the demise of Cretaceous leaf miners, Great Plains, USA. *PLoS ONE*, **9**, e103542.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F. & Douzery, E.J.P. (2003) Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. *Molecular Biology and Evolution*, **20**, 248–254.

van Driel, J.W. & Menken, S.B.J. (1988) Biochemical systematics of the leaf-mining moth family Nepticulidae (Lepidoptera). I. Allozyme variability in the *Stigmella betulicola*, *S. lapponica* and *S. marginicolella* groups. *Entomologica Scandinavica*, **19**, 131–142.

Feder, J.L., Xie, X., Rull, J., Velez, S., Forbes, A., Leung, B., Dambroski, H., Filchak, K.E. & Aluja, M. (2005) Mayr, Dobzhansky, and Bush and the Complexities of Sympatric Speciation in Rhagoletis. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6573–6580.

Fischer, T.C. (2013) Pygmy moths (Lepidoptera, Nepticulidae) from Baltic Amber (Eocene). *Zitteliana*, **53**, 85–92.

Gilbert, M., Guichard, S., Freise, J., Grégoire, J.C., Heitland, W., Straw, N., Tilbury, C. & Augustin, S. (2005) Forecasting Cameraria ohridella invasion dynamics in recently invaded countries: from validation to prediction. *Journal of Applied Ecology*, **42**, 805–813.

Grimaldi, D.A. & Engel, M.S. (2005) Evolution of the insects, pp 755. Cambridge University Press, New York.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Hedges, S.B., Dudley, J. & Kumar, S. (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, **22**, 2971–2972.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, **32**, 835–845.

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**, 260.

Hennig, W. (1966) Phylogenetic systematics, English translation., pp 263. University of Illinois Press, Urbana [etc.].

Hering, E.M. (1951) Biology of the Leaf Miners. Dr W. Junk, 's-Gravenhage.

ICZN (1999) International Code of Zoological Nomenclature. Fourth edition. Adopted by the International Union of Biological Sciences, pp 306. The International Trust for Zoological Nomenclature, London.

Johansson, R., Nielsen, E.S., van Nieukerken, E.J. & Gustafson, B. (1990) The Nepticulidae and Opostegidae (Lepidoptera) of North West Europe. In Fauna Entomologica Scandinavica (ed. by R. Johansson), Vol. 23, pp. 1–739. Brill.

Kawahara, A.Y., Ohshima, I., Kawakita, A., Regier, J.C., Mitter, C., Cummings, M.P., Davis, D.R., Wagner, D.L., De Prins, J. & Lopez-Vaamonde, C. (2011) Increased gene sampling strengthens support for higher-level groups within leaf-mining moths and relatives (Lepidoptera: Gracillariidae). *BMC Evolutionary Biology*, **11**, 182.

Kawahara, A.Y., Plotkin, D., Ohshima, I., Lopez-Vaamonde, C., Houlihan, P.R., Breinholt, J.W., Kawahara, A.Y., Xiao, L., Regier, J.C., Davis, D.R., Kumata, T., Sohn, J.C., De Prins, J. & Mitter, C. (2016) A molecular phylogeny and revised higher-level classification for the leaf-mining moth family Gracillariidae and its implications for larval host use evolution. *Systematic Entomology*, **in press.**

Kemperman, T.C.M., Wilkinson, C., Kuroko, H., & Kumata, T. (1985) Japanese species of the genus *Stigmella* (Nepticulidae: Lepidoptera). *Insecta Matsumurana*, **32**, 1–107.

Knor, S., Prokop, J., Kvaček, Z., Janovský, Z. & Wappler, T. (2012) Plant-arthropod associations from the Early Miocene of the Most Basin in North Bohemia – Palaeoecological and palaeoclimatological implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **321**, 102–112.

Kozak, K.M., Wahlberg, N., Neild, A.F.E., Dasmahapatra, K.K., Mallet, J. & Jiggins, C.D. (2015) Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic Heliconius Butterflies. *Systematic Biology*, **64**, 505–524.

Kristensen, N.P. & Skalski, A.W. (1998). Phylogeny and palaeontology. In Handbuch der Zoologie, Volume IV Arthropoda: Insecta, Part 35, Lepidoptera, Moths and Butterflies, volume 1. Evolution, systematics and biogeography (ed. by N.P. Kristensen), Handbuch der Zoologie/ Handbook of Zoology Vol. 4, pp. 7–25. De Gruyter, Berlin & New York.

Labandeira, C.C., Dilcher, D.L., Davis, D.R. & Wagner, D.L. (1994) 97-Million years of Angiosperm-Insect association - Paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 12278–12282.

Lees, D.C., Lack, H.W., Rougerie, R., Hernandez-Lopez, A., Raus, T., Avtzis, N.D., Augustin, S. & Lopez-Vaamonde, C. (2011) Tracking origins of invasive herbivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. *Frontiers in Ecology and the Environment*, **9**, 322–328.

Leppänen, S.A., Altenhofer, E., Liston, A.D. & Nyman, T. (2012) Phylogenetics and evolution of host-plant use in leaf-mining sawflies (Hymenoptera: Tenthredinidae: Heterarthrinae). *Molecular Phylogenetics and Evolution*, **64**, 331–341.

Letunic, I. & Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.

Lopez-Vaamonde, C., Charles, H., Godfray, J. & Cook, J.M. (2003) Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution*, **57**, 1804–1821.

Lopez-Vaamonde, C., Wikström, N., Labandeira, C., Godfray, H.C.J., Goodman, S.J. & Cook, J.M. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, **19**, 1314–1326.

Magallón, S., Hilu, K.W. & Quandt, D. (2013) Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, **100**, 556–573.

Matsubayashi, K.W., Ohshima, I. & Nosil, P. (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis Et Applicata*, **134**, 1–27.

Mayr, E. (1947) Ecological Factors in Speciation. *Evolution*, **1**, 263–288.

Mayr, E. (1963) Animal species and evolution, pp 797. The Belknap Press of Harvard University Press, Cambridge, MA.

Menken, S.B.J., Boomsma, J.J. & van Nieukerken, E.J. (2010) Large-scale evolutionary patterns of host plant associations in the Lepidoptera. *Evolution*, **64**, 1098–1119.

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., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, 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., Schütte, 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*, **346**, 763–767.

Mitter, C., Farrell, B. & Wiegmann, B. (1988) The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist*, 107–128.

Mutanen, M., Kivela, S.M., Vos, R.A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., Huemer, P., Dinca, V., van Nieukerken, E.J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaens, T., Efetov, K.A., Hebert, P.D., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., Segerer, A., Tarmann, G., Zahiri, R. & Godfray, H.C. (2016) Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. *Systematic Biology*. **Early online**

Nielsen, E.S. & Kristensen, N.P. (1996) The Australian moth family Lophocoronidae and the basal phylogeny of the Lepidoptera-Glossata. *Invertebrate Taxonomy*, **10**, 1199–1302.

van Nieukerken, E.J.. & Johansson, R. (2003) The Quercus feeding Stigmella species of the West Palaearctic: new species, key and distribution (Lepidoptera: Nepticulidae). *Tijdschrift Voor Entomologie*, **146**, 307–370.

van Nieukerken, E.J., Doorenweerd, C. & Hoare, R.J.B. (2016a) Revised classification and catalogue of global Nepticulidae and Opostegidae (Lepidoptera: Nepticuloidea). *ZooKeys*, **submitted**.

van Nieukerken, E.J., Doorenweerd, C., Nishida, K. & Snyers, C. (2016b) New taxa, including two new genera show uniqueness of Neotropic Nepticulidae (Lepidoptera). *ZooKeys*, **accepted**.

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, J.C., Kozlov, M., 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. & 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.

Nyman, T., Farrell, B.-D., Zinovjev, A.-G. & Vikberg, V. (2006) Larval habits, host-plant associations, and speciation in nematine sawflies (Hymenoptera: Tenthredinidae). *Evolution*, **60**, 1622–1637.

Ohshima, I. (2008) Host race formation in the leaf-mining moth *Acrocercops transecta* (Lepidoptera: Gracillariidae). *Biological Journal of the Linnean Society*, **93**, 135–145.

Prosser, S.W., deWaard, J.R., Miller, S.E. & Hebert, P.D. (2016) DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources*, **16**, 487–497.

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

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

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS ONE*, **8**, e58568.

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 Evolutionary Biology*, **9**, 280.

Richardson, B.J., Baverstock, P.R., & Adams, M. (2012) Allozyme electrophoresis: a handbook for animal systematics and population studies, pp 420. Academic Press.

Schluter, D. (2000) The Ecology of Adaptive Radiation, pp 288. Oxford University Press, Oxford.

Simpson, G.G. (1953) The Major Features of Evolution, pp 434. Columbia University Press.

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

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

Spencer, K.A. (2012) Host specialization in the world Agromyzidae (Diptera). Series Entomologica Vol. 45, pp 444. Springer Science & Business Media, Dordrecht.

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B. & Harmon, L.J. (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist*, **207**, 454–67.

Tilmon, K.J. (2008) The evolutionary biology of herbivorous insects: specialization, speciation and radiation, pp 341. University of California Press.

Toussaint, E.F., Condamine, F.L., Hawlitschek, O., Watts, C.H., Porch, N., Hendrich, L. & Balke, M. (2015) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, **64**, 3–24.

Venditti, C., Meade, A. & Pagel, M. (2010) Phylogenies reveal new interpretation of speciation and the Red Queen. *Nature*, **463**, 349–52.

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

Wahlberg, N., Wheat, C.W., & Pena, C. (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS ONE*, **8**, e80875.

Warnock, R.C.M., Yang, Z. & Donoghue, P.C.J. (2012) Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, **8**, 156–159.

Wiens, J.J., Lapoint, R.T. & Whiteman, N.K. (2015) Herbivory increases diversification across insect clades. *Nature Communications*, **6**, 8370.

Winkler, I.S. & Mitter, C. (2008). The phylogenetic dimension of insect-plant interactions: a review of recent evidence. In The evolutionary biology of herbivorous insects: specialization, speciation and radiation. (ed. by K.J. Tilmon), pp. 240–263. University of California Press.

Winkler, I.S., Mitter, C. & Scheffer, S.J. (2009) Repeated climate-linked host shifts have promoted diversification in a temperate clade of leaf-mining flies. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18103–18108.

Wolfe, J.A. (1975) Some aspects of plant geography of Northern Hemisphere during late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden*, **62**, 264–279.

Zwick, A., Regier, J.C., Mitter, C. & Cummings, M.P. (2011) Increased gene sampling yields robust support for higher-level clades within Bombycoidea (Lepidoptera). *Systematic Entomology*, **36**, 31–43.

2

DNA barcoding of *Ectoedemia* s. str. with COI and EF1-α

DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-α: two are better than one in recognising cryptic species

Erik J. van Nieukerken[1, 2], Camiel Doorenweerd[1], Frank R. Stokvis[1], Dick S.J. Groenenberg[1]

[1] *Netherlands Centre for Biodiversity Naturalis, PO Box 9517, 2300 RA Leiden, The Netherlands*
[2] *E-mail: nieukerken@naturalis.nl*

Key words: pairwise difference, Palearctic

Abstract

We sequenced 665bp of the Cytochrome C Oxidase I (COI) barcoding marker for 257 specimens and 482bp of Elongation Factor 1-α (EF1-α) for 237 specimens belonging to the leaf-mining subgenus *Ectoedemia* (*Ectoedemia*) in the basal Lepidopteran family Nepticulidae. The dataset includes 45 out of 48 West Palearctic *Ectoedemia* s. str. species and several species from Africa, North America and Asia. Both COI and EF1-α proved reliable as an alternative to conventional species identification for the majority of species and the combination of both markers can aid in species validation. A clear barcode gap is not present, and in some species large K2P intraspecific pairwise differences are found, up to 6.85% in COI and 2.9% in EF1-α. In the *Ectoedemia rubivora* species complex, the species *E. rubivora, E. arcuatella* and *E. atricollis* share COI barcodes and could only be distinguished by EF1-α. Diagnostic base positions, usually third codon positions, are in this and other cases a useful addition to species delimitation, in addition to distance methods. *Ectoedemia albifasciella* COI barcodes fall into two distinct clusters not related to other characters, whereas these clusters are absent in EF1-α, possibly caused by mtDNA anomalies or hybridisation. In the *Ectoedemia subbimaculella* complex, both sequences fail to unequivocally distinguish the species *E. heringi, E. liechtensteini, E. phyllotomella* and one population of *E. subbimaculella*. DNA barcodes confirm that North American *Ectoedemia argyropeza* are derived from a European introduction. We strongly advocate the use of a nuclear marker in addition to the universal COI barcode marker for better identifying species, including cryptic ones.

Introduction

The Lepidoptera are one of the megadiverse groups of organisms, with currently more than 157,000 described species (van Nieukerken *et al*., 2011), but also a group for which there are many specialists and ongoing taxonomic studies (Kristensen *et al*., 2007). It is therefore no surprise that Lepidoptera were considered particularly useful for DNA barcoding studies, and they have figured importantly in studies on DNA barcoding since the idea was launched in 2003 (Hebert *et al*., 2003a, b, 2004a; Janzen *et al*., 2005; Hajibabaei *et al*., 2006a). The All Lepidoptera barcoding campaign (http://www.lepbarcoding.org/) has resulted in an increasing database of Lepidoptera barcodes (559,920 on October 21, 2011), particularly derived from geographic campaigns in North America, Australia and Europe, and from global campaigns covering the families Sphingidae, Saturniidae and Geometridae (Hebert *et al*., 2010; Wilson *et al*. 2011). The method has proven successful for identifying most morphologically recognised species and has many interesting applications. The most frequently cited application is the recognition of cryptic species (Hebert *et al*., 2004a; Hausmann *et al*., 2009; Janzen *et al*., 2009), but on the contrary barcodes can also confirm that a polyphagous species is indeed one species (Hulcr *et al*., 2007). DNA barcoding also makes matching of unknown immatures with adults possible (Miller *et al*., 2006; Janzen *et al*., 2009), or matching the sexes of dimorphic species (Janzen *et al*., 2009). Further it allows identification of food remains from gut contents (Matheson *et al*., 2007) and even can be used to identify specimens in collections that have lost important characters. Despite this success story, there has been criticism on the use of a mitochondrial marker and on this particular part of COI. Roe and Sperling (2007) concluded that the barcoding region is not discriminating species better than other parts of the COI-COII genes and suggested the use of longer sequences. In Australian Elachistidae several recently diverged species could not be recognised by another 700 bp part of COI (Kaila and Ståhls, 2006) and in some Hesperiidae the differences between species were just three nucleotides (Burns *et al*., 2007). In general, mitochondrial DNA has particular issues related to the nature of mitochondrial biology: reduced effective population size and introgression, maternal inheritance, inconsistent mutation rate, heteroplasmy, compounding evolutionary processes and nuclear pseudogenes are some of the cited causes for problems in species discrimination (Rubinoff *et al*., 2006). In such cases mtDNA based clusters can be composed of specimens belonging to different species through introgression (Ballard and Whitlock, 2004; Ballard and Rand, 2005; Stone *et al*., 2007) and variability within a species can be far larger than between species, thus incorrectly suggesting the presence of cryptic species (Stone *et al*., 2007).

Various authors have suggested the use of additional genes for DNA barcoding, particularly nuclear genes (Sonnenberg *et al*., 2007; Zakharov *et al*., 2009). For discriminating two cryptic species of *Crypsiphona* (Geometridae), Õunap and Viidalepp (2009) used both COI and EF1-α, but did not comment on differences between the two genes. Dasmahapatra *et al*. (2009) found that the COI DNA

barcode recognised more haplogroups than they could recover with techniques such as AFLP.

We are not aware of any study that tried to compare species discrimination throughout any animal taxonomic group with two or more barcoding markers. We realised that the dataset that we obtained during a phylogenetic study of the subgenus *Ectoedemia* Busck, 1907 s. str. (Lepidoptera: Nepticulidae) provides an ideal set for just this type of comparison. When using a mitochondrial marker as well as a nuclear marker the concerns with both markers can hopefully be reduced. We studied the COI barcode and a part of the nuclear gene Elongation Factor 1-α, a frequently studied gene in Lepidoptera, providing much phylogenetic signal (Caterino *et al*., 2000). The Nepticulidae are one of the most speciose, basal, non-Ditrysian, Lepidopteran families, with currently slightly over 800 named species (van Nieukerken *et al*., 2011). They comprise tiny moths of only 3-10 mm wingspan of which the larvae are plant-miners, the majority feeding in leaves. The species are almost invariably monophagous or at most oligophagous, and feed particularly on woody plants in the Rosid clade of the Eudicots, with the notion that related species often use related plants (Menken *et al*., 2010). The genus *Ectoedemia* is one of the larger genera, divided into a number of subgenera (van Nieukerken, 1986). The subgenus *Ectoedemia* is mainly Holarctic, with around 90 known species, feeding on a small number of tree families. The 48 Western Palearctic species (including one that is still unnamed) have been fully revised (van Nieukerken, 1985; van Nieukerken *et al*., 2010), 20 Eastern Palearctic species have been described (Puplesis, 1994), and 18 species were recorded from North America (Wilkinson, 1981; Wilkinson and Newton, 1981), including the European *E. argyropeza*. Outside the Holarctic region, five species are known from southern Africa (Scoble, 1978, 1979; Mey, 2004) and two from Central America (Puplesis and Robinson, 2000), while some unnamed species from the Oriental region are recorded here. Nepticulidae are an ideal group for barcode studies since larvae are easily collected within their leafmines, simultaneously providing information on their host plants. Reared adults provide further tests of species identity, but in many cases identification of larvae and leafmines is possible. The subgenus *Ectoedemia* provides an interesting mix of species that are straightforward to identify in all stages and sexes, species only identifiable by genitalia, and a few species complexes in Europe of which the members are hard to identify at all.

The *Ectoedemia angulifasciella* complex comprises four species feeding on Rosaceae: *E. angulifasciella* mainly on *Rosa*, *E. rubivora* on *Rubus*, *E. arcuatella* on *Fragaria* and *E. atricollis* on several Rosaceous trees (Wilkinson *et al*., 1983; van Nieukerken, 1985). Since these differ in only few morphological characters, these species are most easily identified by their hostplant. Only *E. angulifasciella* can be safely identified by their male genitalia. In the Fagaceae-feeding *Ectoedemia subbimaculella* group, two complexes occur: the *E. albifasciella* complex with four nominal species and the *E. subbimaculella* complex with between two to five species (van Nieukerken, 1985; van Nieukerken *et al*., 2010). The species of the *Ectoedemia albifasciella* complex can only be identified easily by their female genitalia, males

can only be identified from a combination of hostplant, larva and leafmine data when reared, and two species then still cannot be separated. The species of the *subbimaculella* group are almost inseparable as adults, with a slight difference in the female genitalia that distinguishes *E. subbimaculella* from the other species (van Nieukerken, 1985). *Ectoedemia subbimaculella* can also be distinguished from the others by the conspicuous behavioural character in the leafmine, in which the larva makes a slit to prevent waterlogging, but otherwise a combination of hostplant, larva and adult is usually the only way to get an acceptable identification. It would thus be interesting to test the ability of DNA barcoding to separate species in these complexes and to assess whether the taxonomic decisions on the species level are also supported by molecular data. In fact, the identity of at least one of the species, *E. liechtensteini,* has been questioned (van Nieukerken, 1985; van Nieukerken *et al*., 2010). Allozyme studies could not separate three of the studied species of the *subbimaculella* complex (Menken, 1990).

Our phylogenetic studies (Doorenweerd and van Nieukerken, in prep.) have shown that most of the species groups recognised in Europe (van Nieukerken *et al*., 2010) are recovered as monophyletic, when the small *occultella* group is included in the *angulifasciella* group. Here we will therefore use the European group names with the exception of the *occultella* group. All Western Palaearctic species, except three, and in addition three from southern Africa, four from North America, eight from the East Palearctic (including two trans-palearctic species) and approximately five species from Southeast Asia were analysed.

Material and methods

Material

Material was either collected for this project, present in the collections of NCB Naturalis or received from third parties. Typically larvae were collected by searching occupied leafmines, after which individual larvae were immediately conserved in 96 or 100% ethanol (occasionally without ethanol) and later kept in a freezer at minus 80°C. Additional specimens were reared to the adult stage in the laboratory. Cocoons with hibernating larvae of European species were kept in polystyrene rearing containers in an unheated shed and taken indoors at ambient temperatures from March onwards, where the adults finally emerged. Tropical species were reared in temperatures around 25°C and high humidity in a climate cabinet. From small samples often all larvae were preserved directly. The leafmines from which larvae were taken were dried and kept as vouchers. Larvae were identified as far as possible on the basis of external larval characters (such as body colour, head colour, presence of plates) and the larval feeding pattern combined with host-species identity. After rearing, the samples were re-identified on the basis of the adults, by dissection of genitalia if needed. In many cases when larvae could not be identified with certainty, sequences were used for a final identification check. In such species, sequences from correctly identified adults serve as barcoding standard.

The dataset includes 45 out of the 48 known West Palearctic species of *Ectoedemia* s. str. (van Nieukerken *et al*., 2010). All names and full authorities are given in Appendix 1. We were unable to get amplifiable DNA from only three species: *E. hexapetaleae* (Szőcs, 1957), *E. similigena* Puplesis, 1994 and *E. albida* Puplesis, 1994, for which we only had relatively old specimens. In addition to the European species we used material of several species from Asia, Africa or North America, including some undescribed ones and six species of the subgenus *Ectoedemia* (*Zimmermannia*) Hering, 1940 as outgroup. All species analysed are listed with full nomenclature in Appendix 1. Some unnamed species are indicated with tentative names based on hostplants and/or distribution.

For the species of the species complexes, we sampled up to 18 specimens each during focussed collecting. For the remaining species we aimed to use at least two, but preferably more specimens, with the largest possible geographic distances between them to observe representative intraspecific variation (Lukhtanov *et al*., 2009). The majority of adults studied were at the same time examined for a taxonomic revision (van Nieukerken *et al*., 2010), and the data are included in the supplementary data of that paper, using the same registry numbers, and also available on GBIF (http://data.gbif.org/welcome.htm). The material includes type material of several species that were described in the cited paper. All specimens received a registry number for our collection, whether extracted destructively or not. Also specimens not belonging to the RMNH collection (NCB Naturalis collection, former Leiden Zoology collections), received such a number as well for practical reason: the number represents in that case the DNA extract, of which the remaining aliquots are kept in our DNA collection. In our laboratory an extra sequence tracking number was added to each extract. Tissue samples of larvae were usually kept in 96% ethanol in a minus 80 freezer, used adults are kept as mounted specimens in the dry collection, with separate permanent genitalia slides.

All sequences are publicly available on the Barcoding of Life Database (BOLD – www.barcodinglife.com) under the project Nepticulidae of the World – *Ectoedemia Public Records*, with full collection data and images when available. In online Table S3 we provide for each sequenced specimen the identification, sample ID's, Process ID's, GenBank accession numbers and the GBIF data portal URL plus some data on occurrence. Further details can be seen on the BOLD site.

Extraction, amplification and sequencing

Genomic DNA was extracted with the Qiagen DNEasy Blood & Tissue kit. Different types of tissue were used for extraction, depending whether the number of the available specimens allowed destructive extraction or not. Hindleg(s) were cut in small pieces with a scalpel prior to digestion with proteinase K, larvae were homogenised with a disposable pestle. Non-destructive extractions from the abdomen (Knölke *et al*., 2005 slightly modified) were used to combine genitalia preparations with DNA extractions; some larvae were treated in a similar way in order to be able to mount the larval cuticle on a slide.

Primers

For the list of primers see Table 1. We used part of mitochondrial Cytochrome C Oxidase I (COI) – the selected barcoding marker for animals (Hebert *et al*., 2003a), and amplified a part of 665 bp in length with the Lep primers (Hebert *et al*., 2004a). We also sequenced a section of 482 bp of the nuclear Elongation Factor 1-α (EF1-α) marker for most of the specimens. Initial attempts to amplify a 1240 bp fragment of this gene by using the primers (five sets) of Cho *et al*. (1995) largely failed. Only primer M44-1 with rcM52.6 (Cho *et al*., 1995) amplified a 701 bp fragment consistently for at least five different genera of Nepticulidae (*Ectoedemia, Enteucha, Parafomoria, Trifurcula* and *Stigmella*). Based on these results, Nepticulidae-specific primers, EF-NepF and EF-NepR (Table 1) that amplified a 482 bp fragment of this gene, were designed and used throughout this study.

For many specimens we used T7 promotor and T3 tailed primers for both COI and EF1-α, as this speeds up the work-flow and may improve results (Regier and Shi, 2005; Wahlberg and Wheat, 2008). For some older museum specimens, the DNA was too degraded for amplifying sections over 400-bp long. For these we used internal primers for COI (Hajibabaei *et al*., 2006a, b) and EF1-α (specially designed for Nepticulidae). As a consequence, for some specimens there is only a shorter sequence available, denoted with (p) for partial. These shorter sections are, respectively, for COI a 310-bp amplicon and for EF1-α a 251-bp amplicon.

Table 1. Primers used. The names are those that are used on the BOLD site. T-primers are tailed primers, in forward direction tailed with T7 promotor, in reverse with a T3 tail (in bold). The first two published primers are denoted short, because the version most used on BOLD has three more bases in either primer than these.

Primer name	marker	Direction		Reference
LepF1-short	COI	F	ATTCAACCAATCATAAAGATAT	Hebert *et al*. 2004a
LepR1-short	COI	R	TAAACTTCTGGATGTCCAAAAA	Hebert *et al*. 2004a
T-LepF1-short	COI	F	**TAATACGACTCACTATAGGG** ATTCAACCAATCATAAAGATAT	new
T-LepR1-short	COI	R	**ATTAACCCTCACTAAAG** TAAACTTCTGGATGTCCAAAAA	new
MLepF1* (MF1)	COI	F	GCTTTCCCACGAATAAATAATA	Hajibabaei *et al*. 2006a
MLepR1* (MH-MR1)	COI	R	CCTGTTCCAGCTCCATTTTC	Hajibabaei *et al*. 2006a
EF-NepF	EF1-α	F	GCCCCCGGACACAGAGATTTCA	new
EF-NepR	EF1-α	R	CACGACCTACTGGCACTGTTCC	new
T-EF-NepF	EF1-α	F	**TAATACGACTCACTATAGGG** GCCCCCGGACACAGAGATTTCA	new
T-EF-NepR	EF1-α	R	**ATTAACCCTCACTAAAG** CACGACCTACTGGCACTGTTCC	new
T-EFcdwF2*	EF1-α	F	**TAATACGACTCACTATAGGG** CCCAGATTYGARGAAATYAAR	new
T-EFcdwR*	EF1-α	R	**ATTAACCCTCACTAAAG** GCAACDGCAGCTGGRTTRTA	new

PCR
The PCR cycle consisted of 3 minutes initial denaturation at 94°C, 15 seconds cycle denaturation at 94°C, 30 seconds cycle at annealing temperature, 40 seconds cycle extension at 57°C for 40 cycles. A final extension at 57°C for 5 minutes occurred after all cycles had finished. The annealing temperature for COI was 50°C, for EF1-α 57°C. PCR was performed in volumes of 25 μl. For many samples the product was purified using the Promega Wizard Genomic Purification kit using the manufacturers 'spin column protocol', for others the purification was done by Macrogen with a Montage purification kit (Millipore). All samples were sequenced in both directions on an ABI 3730 by Macrogen.

Alignment
Sequencher 4.2 software was used to align the forward and reverse sequences, to manually check for ambiguities in the chromatograms and to export contigs. In EF1-α heterozygous bases are scored with ambiguity codes. The sequences of both markers contain no gaps or stopcodons and were aligned by eye in BioEdit 7.0.9.0 (Hall, 2004).

Neighbor joining trees and distances
Neighbor joining trees were created in Paup* 4.0b10 (Swofford, 2003) using Kimura 2 Parameter distance, the algorithm also used for species identification in the BOLD datasystems. Ten thousand bootstrap replicates were performed with Paup*, and bootstrap and distance values are shown on the respective branches present in a Neighbor joining tree (Figs S1-S2). For the trees of different clusters we ran separate analyses. Distances were also calculated using Kimura 2 Parameter distance, either by BOLD tools or with MEGA5 (Tamura *et al*., 2011). As outgroup we used species of the subgenus *Ectoedemia* (*Zimmermannia*), which is the sistergroup of *Ectoedemia* s.s. in both ongoing unpublished family level phylogenetic studies (Hoare and van Nieukerken, in prep; van Nieukerken *et al*., in prep.) and based on morphological characters (van Nieukerken, 1985; van Nieukerken *et al*., 2010).

Diagnostic positions
In cases of closely related species, where sufficient sequences were available, we also analysed the sequences for mutually exclusive diagnostic base positions. They were also defined as 'simple character attributes' in the Character Attribute Organisation System (DeSalle *et al*., 2005; DeSalle, 2006; Rach *et al*., 2008). Results are depicted in table form (Tables 4-6), we indicated whether these positions are third codon positions or not.

Results

In total we obtained 262 COI sequences (ten partial) belonging to *ca*. 64 species of *Ectoedemia* sensu stricto, including five sequences (one partial) of four species in the subgenus *Zimmermannia*. Further we obtained 240 EF1-α sequences (25 partial)

of *ca*. 62 *Ectoedemia* s. str. and three sequences of three *Zimmermannia* species (one partial).

Quality material
Material collected for molecular studies, usually larvae, kept in 96% or 100% ethanol in a -80 freezer, was almost always successful. In total 85.8% of the successfully extracted material yielded the full COI barcode and 80.4% the EF1-α sequence (see Table 2). Dried collection material was also successful when only a few years old, with a progressive decline for older material, but still some full barcodes were obtained from 19 year old specimens, in all cases extracting DNA from abdomens. Shorter barcodes (335 bp) and a shorter part of EF1-α (251 bp) were obtained from 3-25 year old material. Older larvae kept in 70% ethanol (collected for morphological studies), were partly successful: from 54% we obtained at least a short barcode for material up to 28 years after collecting; this material was less successful for EF1-α: out of 10 larvae we got just one full and one partial sequence.

COI versus EF1-α
In COI 148 out of 658 basepairs are variable in our dataset (22.3%) and in EF1-α 152 out of 482 (31.5%). The effectiveness of COI and EF1-α for barcoding this group was compared by plotting the maximum intraspecific distance against the minimum

Table 2. Quality of sequenced material of *Ectoedemia* s. str. by tissue type, results of COI and EF1-α compared.

	Age material	COI (#)	%	Age material	EF1-α (#)	%
Larvae						
directly in ethanol 96% in minus 80						
full sequence	0-9 yrs	161	97.6%	0-9 yrs	156	96.3%
in pieces or short fragment only				7 yrs	1	0.6%
failed	0-1 yrs	4	2.4%		5	3.1%
older ethanol 70% material						
full sequence	17 yrs	1	9.1%	17 yrs	1	10.0%
in pieces or short fragment only	17-25 yrs	5	45.5%	25 yrs	1	10.0%
failed	25 yrs	5	45.5%	25 yrs	8	80.0%
Adults						
Abdomens, non destructive						
full sequence	0-19 yrs	77	74.8%	0-11 yrs	59	61.5%
in pieces or short fragment only	3-19 yrs	14	13.6%	0-19 yrs	23	24.0%
failed	0-28 yrs	12	11.7%	0-16 yrs	14	14.6%
Hindlegs						
full sequence	6/7 yrs	3	75.0%	7 yrs	2	66.7%
failed	6 yrs	1	25.0%	7 yrs	1	33.3%
Overall results						
full sequence		242	85.8%		218	80.4%
in pieces or short fragment only		19	6.7%		25	9.2%
failed		21	7.4%		28	10.3%

interspecific distance for 39 species pairs of which more than one specimen was available for both markers (Fig. 1a). If the minimum distance between species is larger than the maximum distance within species, they can reliably be assigned to a species and it can be said that there is a 'barcoding gap' present. The vast majority of data points is well above the 1:1 barcoding gap line, indicating that the COI and EF1-α sections we used are reliable barcoding markers. The data points below the 1:1 line are from species pairs belonging to the complexes treated in detail below. The graph also shows that the maximum intraspecific distance of COI for these species can be as high as 3.5%, whereas EF1-α values remain below 2.0%. The effectiveness was further examined by plotting the pairwise distances of COI and EF1-α between specimens in a scattergram (Fig. 1b). If both markers would evolve at exactly the same rate, all data points would be expected to be on the 1:1 diagonal. If the rate with which mutation accumulate would differ constantly between COI and EF1-α, all datapoints would fall either below or above the diagonal. The latter clearly is not the case with our data: EF1-α and COI evolve at roughly the same rate.

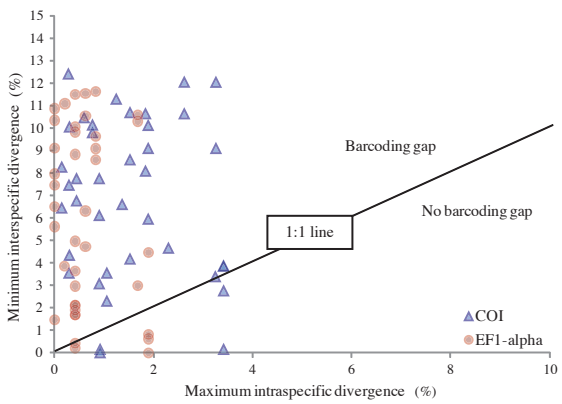


Fig. 1a. Comparison of maximum interspecific divergence versus maximum intraspecific divergence percentages between 39 *Ectoedemia* s. str. species pairs for which multiple sequences for both markers were available. EF1-α reaches the same minimum interspecific distances as COI, but the maximum intraspecific divergence is much lower. Species pairs below the barcoding gap line involve species belonging to the complexes.

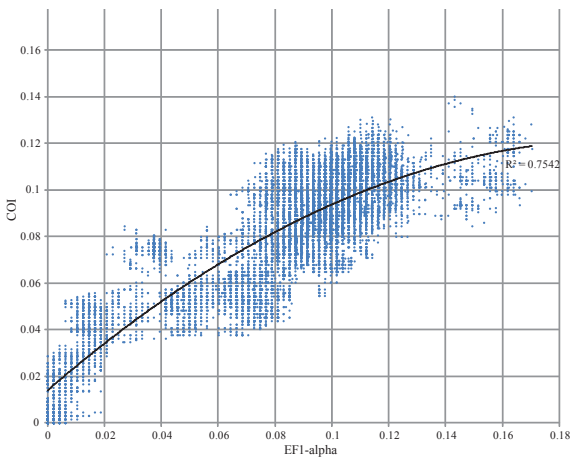


Fig. 1b. Scattergram containing the 21.945 pairwise distance values of COI and EF1-α between all specimens. The polynomous trend line gave the best fit for the data, with an R^2 of 0.75. The data does not show that either marker evolves at a higher rate in general, but closely related specimens show larger distances in COI, where more distantly related specimens are the exact opposite of this and show larger distances in EF1-α.

However, the most fitting trend line, polynomous with an R^2 of 0.75, indicates that pairwise distances between closely related specimens (*i.e.* within species) are higher in COI than in EF1-α, and pairwise distances between more distantly related specimens (*i.e.* between species groups or subgenera) are higher in EF1-α than in COI. So, even though there is no linear difference between the mutation rate of both markers, this indicates there is an evolutionary difference between both markers nonetheless.

Species recognition

All specimens of a single species, except the species in the *Ectoedemia rubivora* complex (*E. arcuatella, atricollis* and *rubivora*), the *E. albifasciella* complex and part of the *E. subbimaculella* complex form distinct clusters in both the COI and the EF1-α neighbour joining trees with all taxa included (Figs S1-S2). They are unambiguously distinguishable from other species by using distance methods for either marker. BOLD highlights intraspecific pairwise differences that exceed 2% as potentially containing cryptic species (Ratnasingham and Hebert, 2007). Several species in our dataset exceed this threshold, their maximum distances as well as their respective values in EF1-α and maximum geographic distance between the samples are shown in Table 3. Below we will discuss species for which we have sequenced more than one specimen from more than one locality.

The Ectoedemia angulifasciella *group*

Species in this group largely feed on Rosaceae, with the exception of the Betulaceae feeders *E. occultella* and *E. minimella*. The taxonomic status of three specimens found feeding on *Rubus* in Vietnam and Borneo, and a fourth specimen caught at light in Vietnam, is uncertain (Fig. 2). The external morphology, geographic region, host species and feeding pattern of the *Rubus* miners suggest that they might belong to a single species, but insufficient material is available to conduct a conclusive morphological study. Both COI and EF1-α results indicate that they likely represent several species, with pairwise differences between 5.6% and 6.2% in COI and 2.1% and 4.0% in EF1-α. The results also suggest that the specimen collected at light is a different species closely related to the aforementioned three, and likely to be found mining *Rubus* sp.

Table 3. Maximum K2P intraspecific pairwise distance of *Ectoedemia* sequences exceeding 2%

Species	COI (%)	EF1-α (%)	Geographic distance (km)
E. alnifoliae	2.62	0.21	12
E. pseudoilicis	3.26	0.21	770
E. haraldi	4.26	2.53	2455
E. erythrogenella	3.26	0.42	3200
E. spiraeae	6.85	2.90	7388
E. intimella	6.50	1.20	9458

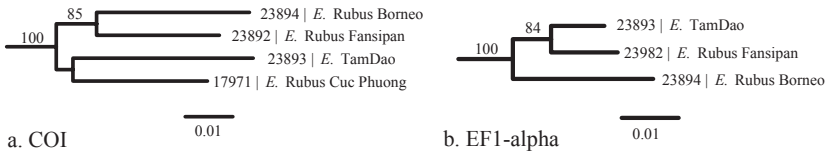


Fig. 2. K2P Neighbor joining trees containing possibly four species in the *Ectoedemia angulifasciella* group. Specimen numbers are RMNH.INS registry numbers. All specimens were collected in Vietnam or Borneo, at light as adult or as larva on *Rubus* spp. 2a: The COI tree. The distances between specimens are large, indicating that they are likely to represent several species. 2b: The EF1-α tree. One specimen less was included here, but the tree also shows relatively long distances between specimens.

In another *Rubus* feeder, *E. erythrogenella*, specimens from Spain, France, Sardinia and Greece hardly show differences, whereas the single specimen from Turkey shows a pairwise difference in COI with the others greater than 2%, but less than 0.5% in EF1-α (Table 3). In *Ectoedemia spinosella*, the single specimen from Greece and from another hostplant (*Prunus webbii*) differs considerably in both genes from the completely identical sequences from specimens from France, Italy and the Netherlands (all feeding on sloe, *Prunus spinosa*). In the closely related *E. mahalebella*, specimens from southwest France, Italy and Croatia hardly show any variation.

The two specimens of *E. spiraeae* studied, one from Europe (Slovakia) and one from China, show a distance of 6.47% in COI and 2.9% in EF1-α, which can be correlated with a very large geographic distance. The *E. spiraeae* species cluster is the only that did not get a bootstrap support over 60 in both markers (online Figs S1-S2). *Ectoedemia spiraeae* has a scattered distribution from eastern Europe through Siberia to China and Japan, with a relative large gap between Europe and Asia (van Nieukerken *et al*., 2010). These results suggest that possibly different species are involved, a possibility to investigate by a morphological and molecular study of more material from a wider range of populations.

Specimen RMNH.INS.23741 was discovered in Norway and provisionally described as *Ectoedemia* sp. n. (Bengtsson *et al*., 2008). Barcodes show that it is almost identical to larvae and reared adults from *Rosa* from France. This species has now been described as *E. rosae* van Nieukerken and Berggren, 2011.

The Ectoedemia angulifasciella *complex* (Fig. 3).
Although this complex was originally established as a complex containing four cryptic species based on morphological characters (Wilkinson *et al*., 1983; van Nieukerken, 1985), *E. angulifasciella* can easily be discriminated from the other three by 23 diagnostic basepairs: 3.5% of the entire sequence (Fig. 3a); this is also the only species that can be separated by at least one character in the male genitalia. By contrast, there are no diagnostic base positions in COI at all that discriminate between the three remaining species *E. arcuatella, E. rubivora* and *E. atricollis*. Since the distance from *E. angulifasciella* to the other species is also large in EF1-α

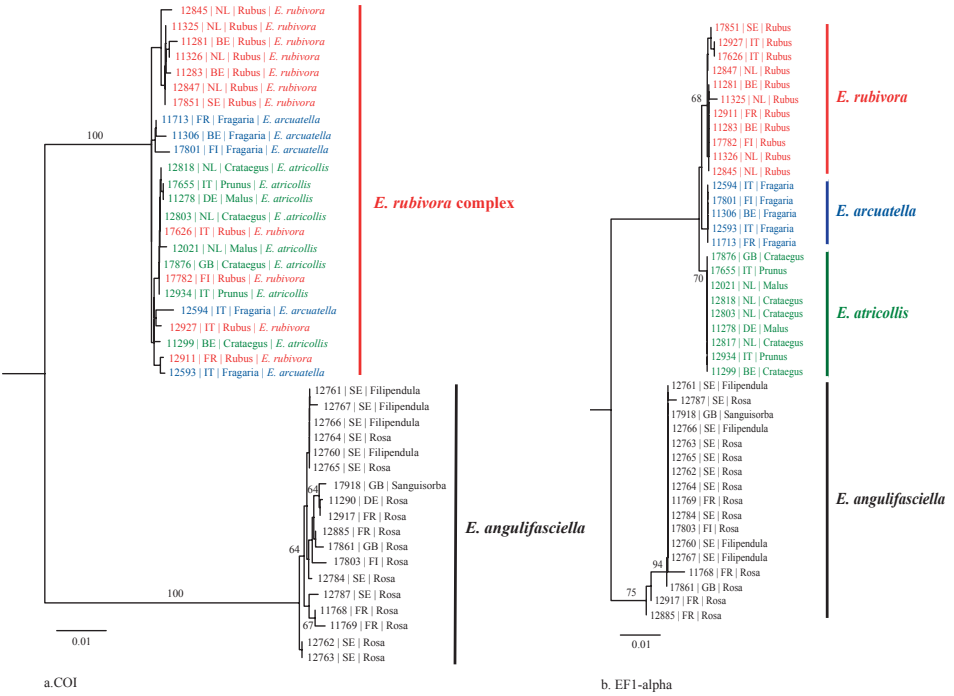


Fig. 3. K2P Neighbor joining trees of *Ectoedemia angulifasciella* and the *E. rubivora* complex with bootstrap values. The colours denote the different species. The annotation starts with the RMNH.INS registry number, followed by ISO coded country of origin and host. Outgroup for these trees is *E. terebinthivora*, bootstrap values represent 10,000 replicates.
3a: The COI tree. There is a large distance between *E. angulifasciella* and the *rubivora* complex, but within *E. angulifasciella* there is little variation and none that can be correlated with different hostplants. *E. rubivora*, *E. arcuatella* and *E. atricollis* cannot be distinguished in COI.
3b: The EF1-α tree. As in COI *E. angulifasciella* seems only distantly related to the others. *E. rubivora*, *E. atricollis* and *E. arcuatella* group on species clusters and can thus be distinguished using EF1-α, albeit based on only two positions (see also Table 4).

(Fig. 3b), this strongly suggests that *E. angulifasciella* should not be regarded as part of this complex; we therefore suggest renaming this the *E. rubivora* complex. The four species can be distinguished by the host plant they are found on and some minor morphological characters. They have been lumped or split in the past depending on the emphasis on biological data versus morphological data (for a review see Wilkinson *et al*., 1983). Five sequences of COI, belonging to specimens of *E. atricollis* and *E. rubivora*, were completely identical (RMNH.INS #'s 11278, 17626, 17782, 12818 and 12803); a few haplotypes in this complex do not coincide with species boundaries at all. Where COI fails to distinguish species, we found that *E. atricollis*, *E. rubivora* and *E. arcuatella* can be distinguished molecularly based on two synonymous mutations (diagnostic) at third codon positions in EF1-α (Table 4, Fig. 3b). For all three species we have included material originating from a large part of their European range.

Eighteen specimens of *E. angulifasciella* were sequenced, in order to test whether populations on different hosts can be differentiated by their barcodes. This species feeds mainly on *Rosa* species, but also locally on *Filipendula vulgaris* and *Sanguisorba* species. The fact that the species in Öland (Sweden) can be abundant on *Filipendula* and completely absent from *Rosa* in the same locality, and vice versa on other localities, suggests that there might be different, morphologically cryptic species specialising on these different hosts (see Bengtsson *et al*., 2008). *Sanguisorba* feeders have also been described as several different species in the past (synonymised by van Nieukerken, 1985). However, the molecular results do not show any difference for material from various hosts, but show a rather invariable *E. angulifasciella* throughout Europe with a maximum intraspecific pairwise difference of 0.77% in COI and 0.84% in EF1-apha, thus confirming the morphological findings.

The Ectoedemia suberis *group*

All species in this oak mining group show little or no intraspecific variation. The species *E. hendrikseni*, *E. phaeolepis* and *E. heckfordi* have recently been discovered and belong to a morphologically difficult complex that also includes *E. andalusiae* and *E. suberis* (van Nieukerken *et al*., 2010). All are found in West and Southwest Europe with partly overlapping geographic ranges. COI and EF1-α support their full species status, with interspecific genetic distances varying between 2.6% and 6.8% in COI and distances between 2.7% and 4.0% in EF1-α, comparable to the distances between other species. All species form monophyletic clusters with high bootstrap support (Fig. 4). The branching pattern between these species differs significantly in both markers, but they always group together as four. Besides confirming the species status of these five species, these results show that COI and EF1-α can readily be used to distinguish these species.

Table 4. Simple character attribute positions within EF1-α to distinguish three closely related species of the *Ectoedemia rubivora* complex. Both positions are 3rd codon positions, the substitutions are synonymous.

Species\position	80	230
E. atricollis	T	C
E. rubivora	C	T
E. arcuatella	C	C

The Ectoedemia populella *group*

All species feed on Salicaceae. A few specimens from North America are included., The pairwise distances within *E. intimella* are very large, correlated with a very large geographic distance. There was one *E. intimella* specimen included from Japan, with a distance of 6.74% from the others. This female specimen is morphologically indistinguishable from European specimens (van Nieukerken *et al*., 2010), but since we have not seen other Japanese material, nor any intermediate populations, no taxonomic conclusions can be based on this finding.

From *Ectoedemia argyropeza* a North American subspecies, *E. argyropeza downesi* Wilkinson & Scoble, 1979, has been described on slight morphological differences. Wilkinson and Scoble (1979) did not consider the possibility that the North American

populations are introduced from Europe. Later Menken and Wiebosch Steeman (1988) concluded on the basis of allozymes that this is most likely the case. Five of the European COI sequences of *E. argyropeza* are also 100% identical to several North American specimens registered on BOLD when using the BOLD identification engine, corroborating the earlier findings.

The Ectoedemia subbimaculella *group*

The *E. subbimaculella* group is the second group specialised on Fagaceae (*Quercus*), although probably some species feeding on other hosts belong here as well. It includes the two species complexes discussed separately below. For *E. alnifoliae* and *E. pseudoilicis* we found intraspecific pairwise differences in COI greater than 2%, but less than 0.5% in EF1-α (Table 3). In the case of *E. pseudoilicis* it is the Turkish specimen that differs from the Greek ones, but in *E. alnifoliae* the sampled populations that show these differences (in Turkey) are just 12 km apart. In *E. haraldi* we found a maximum pairwise difference of 4.26% in COI, and 2.53% in EF1-α (Table 3). Four specimens of this species were included, the two western European ones (from France and Spain) and two eastern specimens (Greece and Turkey) form separate clades. Also all four specimens have large pairwise differences between them, with a minimum of 1.2% in COI and an average of 3.16% (Table 3). Superficially the eastern and western populations are similar in morphology and biology, a detailed morphological analysis should be carried out to see if the molecular differences are paired with morphological differences. *Ectoedemia heringella* shows two clusters in COI, both including specimens from Italy and Great-Britain. This may indicate that the introduced British population (van

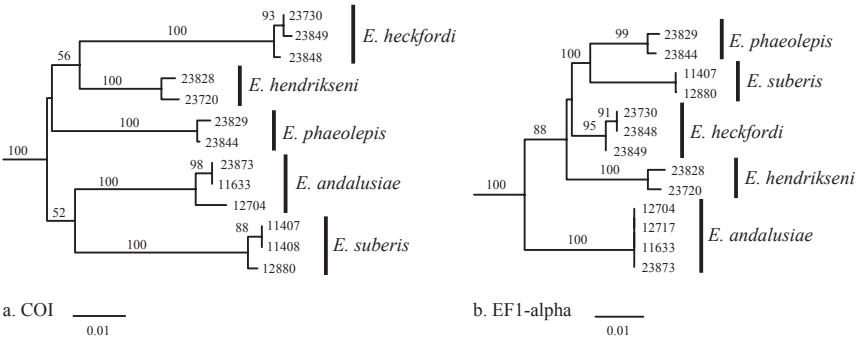


Fig. 4. K2P Neighbor joining trees containing species of the *Ectoedemia suberis* complex. Specimen numbers are RMNH.INS registry numbers. All hosts are *Quercus* spp. and all species are only found in South-West Europe. Outgroup for this analysis was *E. terebinthivora*, bootstrap values represent 10,000 replicates.
4a: The COI tree. The species are on their own clusters with high support and distances comparable to those between other species.
4b: The EF1-α tree. Much the same as the COI tree, with high support for separating the species and good distances. The positions of the clusters here are completely different from those in COI.

Nieukerken *et al*., 2010) is genetically structured, and not the result of the introduction of a single specimen. The COI clusters are not paralleled in EF1-α.

The Ectoedemia albifasciella *complex*

This complex comprises four *Quercus* (Fagaceae) mining species: *E. albifasciella*, *E. pubescivora*, *E. contorta* and *E. cerris* (van Nieukerken, 1985; van Nieukerken *et al*., 2010). These species can be distinguished using COI, but there are also two distinct clusters (2.17% difference) in *E. albifasciella*, making it polyphyletic (Fig. 5a). There are no indications that NUMT's (nuclear mitochondrial inserts, Zhang and Hewitt, 1996) are the cause of this, since the chromatograms contain no double signals, and a translation into amino acids showed no difference between both haplotypes nor the presence of stop codons within these sequences. In EF1-α these two haplotypes are not recovered, we believe this different haplotype is the result of a mitochondrial anomaly (Ballard and Whitlock, 2004; Rubinoff *et al*., 2006). There is no biological or morphological data to support the split of the two COI haplotypes, but the second haplotype has up to now only been recovered from immature specimens. Only the geographic data suggests that the rare haplotype might have a limited distribution in the Netherlands and adjacent West Germany, but more specimens will have to be included to confirm this. The distribution of the more common haplotype also includes this area, and some samples from a single locality contain both haplotypes. By looking at mutually exclusive diagnostic base positions or simple character attributes, specimens can be identified, even though *E. albifasciella* appears polyphyletic. Apart from one, all these differences are synonymous (Table 5). It should be noted that for the other three species relatively few specimens are sequenced. It is thus possible that simple character attributes might disappear when more specimens are examined and show intraspecific variation. In EF1-α *E. albifasciella* is paraphyletic relative to the other three species. *E. contorta*, *E. cerris* and *E. pubescivora* are represented as a clade in the Neighbor joining tree (Fig. 5b), but there are no simple character attributes to distinguish them. A single studied male *E. albifasciella* from Morocco (not in Fig. 5) is identical in EF1-α, but possibly belongs to another COI haplotype with 1.4 % difference, not grouping with other *E. albifasciella* in the NJ trees; it has 13 out of the 15 diagnostic basepairs of *E. albifasciella*.

Table 5. Simple character attribute positions in COI to distinguish species within the *Ectoedemia albifasciella* complex. Most positions are 3rd, apart from 389 and 584. All substitutions, except at 389, are synonymous.

Species\position	46	121	187	223	235	238	346	389	394	463	568	577	584	604	619
E. albifasciella	A	T	A	C	C	C	C	T	A	T	A	T	T	C	T
E. pubescivora	A	C	A	C	C	C	C	T	A	T	A	T	T	T	C
E. contorta	G	T	A	T	C	C	T	C	A	C	A	T	T	C	T
E. cerris	A	T	G	C	T	T	C	T	G	T	G	C	C	C	T

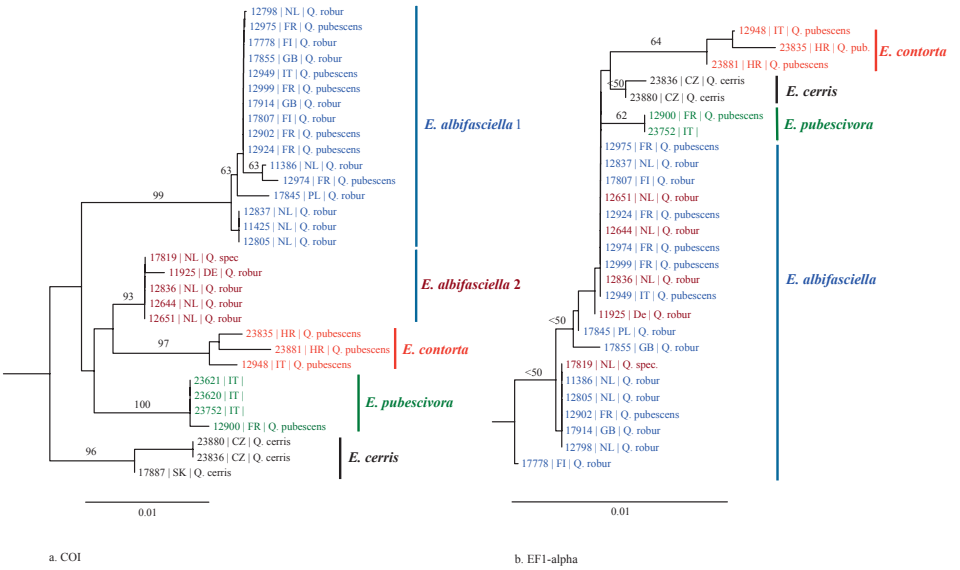


Fig. 5. K2P Neighbor joining trees of the *Ectoedemia albifasciella* complex with bootstrap values. The annotation starts with the RMNH.INS registry number, followed by ISO coded country of origin and host, when collected as larva. All hosts are *Quercus* spp. The colours denote the different species and the two haplotype clusters of *E. albifasciella*. Outgroup for this analysis was *E. rufifrontella*, bootstrap values represent 10,000 replicates.
5a: The COI tree. The species have their own clusters with high support and distances over 1.2%. There are haplotype clusters in *E. albifasciella*, with a 2.17% distance.
5b: The EF1-α tree. Only *E. contorta* and *E. pubescivora* can be distinguished in EF1-α. The two *E. albifasciella* haplotypes found in COI are not found here. The specimen from Morocco (see text) is not included in this analysis.

The Ectoedemia subbimaculella *complex*

This is the second complex found in this group (van Nieukerken, 1985; van Nieukerken *et al*., 2010). It contains the widespread *E. subbimaculella*, as well as *E. heringi*, *E. phyllotomella* and *E. liechtensteini* which all three are restricted to southern or eastern Europe, the last two specialising on *Quercus cerris*. *Ectoedemia subbimaculella* is placed on a well supported cluster in the COI tree, but with a bootstrap value of 52 unsupported in EF1-α. Also, RMNH.INS.23514, identified as *E. heringi* (by COI!) falls inside the *E. subbimaculella* clade in EF1-α; this may be a case of introgression in mtDNA. The other species are more problematic. Our results for *E. liechtensteini* are inconclusive, we therefore cannot assess the usability of either marker for this putative species. *Ectoedemia phyllotomella* cannot be satisfactorily distinguished in either marker. Especially interesting is the clade consisting of two specimens identified as *E. subbimaculella* from Hungary (feeding on *Quercus cerris*) and two specimens identified as *E. heringi* from Greece (found on *Quercus ithaburensis*), further complicating this complex. They share five third codon positions in COI where they differ from all other complex members (Table 6). In EF1-α these two supposed *E. subbimaculella* specimens are placed basally to the

Table 6. Simple character attribute positions in COI to distinguish some species or unidentified clusters within the *Ectoedemia subbimaculella* complex. *E. subbimaculella* multistate positions: 568, T:C = 13:1; 628, A:G = 11:3. All positions are 3rd codon positions, the substitutions are synonymous.

Species\position	76	103	226	325	352	376	547	565	568	628
E. subbimaculella	A	T	T	A	A	G	A	T	T/C	A/G
E. heringi	A	T	T	G	A	A	A	T	A	A
E. liechtensteini	G	T	T	G	A	A	A	T	A	A
E. phyllotomella	A	T	T	G	A	A	A	T	A	A
E. 'subbimaculella' from *Q. cerris*	A	G	G	G	G	A	G	C	A	T
E. 'heringi' from *Q. ithaburensis*	A	G	G	G	G	A	G	C	A	T

whole group, although without support. The general picture for this complex is that *E. heringi* and *E. phyllotomella* cannot be distinguished based by COI or EF1-α barcodes, but that *E. subbimaculella* can, when the *Quercus cerris* feeding form is excluded. *E. subbimaculella* differs by three simple character attributes in COI from the others and one position in COI distinguishes *E. liechtensteini* (taking into consideration that only three specimens have been sequenced). In EF1-α there is just a single character attribute: *E. subbimaculella* (including the misidentified RMNH.INS.23514) has an A in position 221, where the others have a G. All of these are synonymous substitutions at third codon positions. By barcoding we could also identify a peculiar colour aberration reared from *Quercus* as *E. subbimaculella*: RMNH.INS.23671, see photograph on BOLD.

Discussion

One or two genes

Ever since the original studies on DNA barcoding in animals (Hebert *et al.,* 2003a, b, 2004b), the COI section amplified by Folmer primers and derivates has become the standard barcoding marker for animals, at least for insects and vertebrates (CBOL Database Working Group, 2009). The use of a single marker is in contrast with other groups of organisms, where also the selection of barcoding has been, or still is a lengthy process. In land plants the choice for two markers, the chloroplast genes rbcL and matK has taken several years (Chase *et al.,* 2005; CBOL Plant Working Group, 2009). Mycologists seem finally to settle for a single gene, ITS, after many years of discussion (Schoch *et al.,* 2011, Santamaria *et al.,* 2009). The choice for COI as single standard marker has enormous advantages, but even in the animal kingdom is no longer the only standard, since LSU and SSU are used frequently for *e.g.* nematodes (Blaxter, 2004; Blaxter *et al.,* 2005). Despite scientific objections (Will and Rubinoff, 2004; Rubinoff *et al.,* 2006; Roe and Sperling, 2007; Song *et al.,* 2008), the choice for COI has been accepted from the onset. With the present size of the barcode database with 1.4 million records, changing the barcode marker is obviously no option, and not something we would like to advocate, but using additional markers may well be the way forward.

Our study of *Ectoedemia* barcodes shows that COI is able to identify the majority of *Ectoedemia* species, but the three species in the *rubivora* complex share barcodes and show variation independent of species boundaries. In other species complexes, species can be distinguished, but often show interspecific distances far below the 'ideal' threshold value of 2%. In contrast, in several species the intraspecific variation is high, much larger than the 2% threshold, and in *E. albifasciella* occurs even a deep split. In other words, a clear barcoding gap does not exist in parts of the genus, whereas it may be present in other parts. The partial sequence of Elongation Factor 1-α is also able to identify the majority of species, in this case including the *rubivora* complex species, but not most species in the *subbimaculella* complex. In contrast to COI, the intraspecific variation is much smaller (up to 2.5%) and the extra haplotype of *E. albifasciella* is not present in EF1-α. Because the intraspecific variation is much larger in COI (up to 6.85%) than in EF1-α (up to 2.53%), while the interspecific variation is rather similar in both genes, the latter in fact might be considered more suitable as 'barcoding marker' for Nepticulidae. In COI there is more intraspecific variation than in EF1-α, which could be mitochondrial anomalies (Ballard and Whitlock, 2004; Rubinoff *et al*., 2006) since most of these differences are not observed in EF1-α. The increase in pairwise distances between more distantly related specimens in EF1-α compared to COI could be explained by a higher level of saturation in COI, and thus loss of information in this marker at this taxonomic level. However, both genes have their limitations and cannot identify all species on their own. Taken together the resolution becomes much better and almost all species are straightforward to identify. Our results support the idea that two barcoding markers are better than one, and we will routinely continue to use these two genes, and in addition usually also the D2–D3 region of the nuclear ribosomal marker 28S. The latter and EF1-α provide better phylogenetic resolution and therefore can more easily place unknown taxa in their correct phylogenetic position. In practice we find that COI sequences often lead to taxonomic mismatches, in contrast to a recent study on sphingid moths (Wilson *et al*., 2011). The nuclear EF1-α gene was amplified almost as easily as COI from museum material (80 versus 85%), indicating it can be a useful alternative marker. However, we do not advocate that EF1-α should be a second universal barcoding marker throughout the Metazoa or even the insects, in too many cases introns and multiple copies make this gene less suitable (Caterino *et al*., 2000; Djernæs and Damgaard, 2006). For Lepidoptera this might be the ideal addition, since introns and multiple copies are as yet unknown and the gene is routinely sequenced for phylogenetic studies (Caterino *et al*., 2000; Wahlberg and Wheat, 2008). However, even in Lepidoptera specific primers for several subgroups are needed (Cho *et al*., 1995; Yamamoto and Sota, 2007).

Barcoding gap

The fact that a barcoding gap exists is an interesting biological phenomenon. However, as also has been observed earlier, when larger geographic areas are sampled, the gap may disappear (Lukhtanov *et al*., 2009) and this is what we see in some of our examples. In the cases of the widespread species *E. intimella* and *E.*

spiraeae, where we have a gap in observations, we are unable to decide on the basis of the barcoding results whether these are widespread species or should be split in two taxa. In the first case we have not enough material to check it by morphology, the single female from Japan is inseparable from European ones, although the barcode would clearly suggest two different species, but we suspect that this species has a continuous range from Europe to Japan. In the case of *E. spiraeae* it might be possible that more species are present, because of the large gap between European and Siberian populations (although this needs to be checked), and also here more morphological study is needed as well as molecular analyses from intermediate populations. In other cases where both markers show a deep split, there may be cryptic species present (the oriental *Rubus* feeders, *E. haraldi*) that need to be scrutinised. In some species we found the largest distances in COI between specimens from Turkey and Europe, a phenomenon known in various other animal groups and thought to be originating from different glacial refuges (*e.g.* oak galls, Stone *et al*., 2007). However, the amount of Turkish material we have seen is too limited for further conclusions.

Species complexes

In the species complexes previously defined by morphology (van Nieukerken, 1985), several species could not be distinguished by DNA barcoding, and when they could, the intraspecific distances are far below the 2% threshold. In the *E. rubivora* complex, three species share the same barcode variation, but they differ in EF1-α by two simple character attribute positions. The species in this complex have been lumped or split in the past depending on the emphasis on biological data or morphological data (for a review see Wilkinson *et al*., 1983). This complex has been of molecular interest before and Wilkinson *et al*. (1983) used allozyme analysis to clarify their specific status. They found evidence for two pairs of sibling species, with *E. angulifasciella* closest to *E. atricollis* and *E. arcuatella* closest to *E. rubivora*. We, however, found in both COI and EF1-α large pairwise differences between *E. angulifasciella* specimens and the other species and thus the complex is reduced to just the other three species (the *rubivora* complex). Whether it is due to recent speciation and incomplete lineage sorting that COI cannot be used to separate these species, or because there has been hybridisation remains an open question. The latter possibility is suggested by the fact that part of the COI haplotypes seem to group by species.

In the *E. albifasciella* complex we found species clusters for all four species in COI, but also two distinct haplotypes for *E. albifasciella*. Aside from this issue, the species can be distinguished using both distance and simple character attribute methods in COI. In EF1-α however three species have only poorly supported species clusters, and *E. albifasciella* is paraphyletic with regard to the other three. This could be because EF1-α is more conservative and has not yet accumulated the differences needed to distinguish species. Both markers combined suggest that there has been recent allopatric speciation and possibly secondary sympatry after post glacial dispersion. We consider the possibility that the second haplotype of *E. albifasciella* represents a separate species as rather

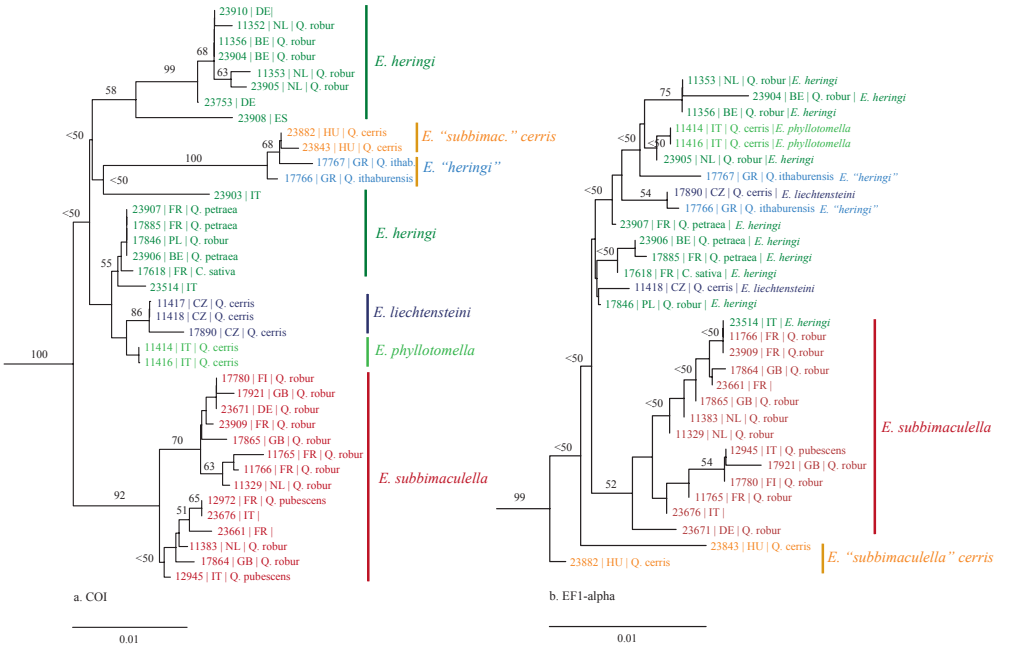


Fig. 6. K2P Neighbor joining trees of the *Ectoedemia subbimaculella* complex with bootstrap values. The annotation starts with the RMNH.INS registry number, followed by ISO coded country of origin and host, when collected as larva. All hosts are *Quercus* spp., except for specimen 17618 which was collected on *Castanea sativa*. Sequences of *E. rufifrontella* were used as outgroup for both trees, bootstrap values represent 10,000 replicates. The colours denote the different species and the aberrant forms of *E. heringi* on *Quercus ithaburensis* and *E. subbimaculella* on *Q. cerris*.
6a: The tree based on COI data. A large *E. subbimaculella* cluster; the '*E. subbimaculella*' specimens on *Q. cerris* group with the '*E. heringi*' specimens on *Quercus ithaburensis*.
6b: The tree based on EF1-α data. Although there are many clusters, they have low bootstrap values.

unlikely, but admittedly, we only got this sequence from larvae, and thus have no information on the adult morphology.

In the *E. subbimaculella* complex, distance methods cannot be used to confidently distinguish species, despite the overall genetic distances in the complex. There is some clustering for species, but because we were only able to include few specimens of *E. liechtensteini* and *E. phyllotomella* we cannot yet conclude if they can be distinguished using these markers or not. There seem to be large clusters for *E. heringi* and *E. subbimaculella* but there are also specimens that are placed at the base of the tree instead in those clusters. However, from the limited data there appear at least to be some simple character attributes for parts of this complex. The molecular results have not brought us much closer to the understanding of this complex than morphological methods did 25 years ago (van Nieukerken, 1985), but

at least these results provide a basis for a detailed genetic analysis of this intriguing complex of two or more species, that possibly show extensive hybridisation after secondary sympatry. The specimen 23514 shows the danger of barcode identification: the nuclear gene identifies this specimen as *E. subbimaculella*, whereas COI recognizes it as *heringi*. Because of the widespread occurrence of introgression in mtDNA (Ballard and Whitlock, 2004), the first identification seems to be the more likely one.

Introduced species

Another useful application of DNA barcoding is the possibility to identify potential introduced species, since we expect that barcode haplotypes of the introduced population will form a subset of the source haplotypes. We confirmed that COI sequences of European *E. argyropeza* are identical to North American sequences already on BOLD, corroborating earlier findings with allozymes (Menken and Wiebosch Steeman, 1988). Already many cases of overlooked introductions from Europe to North America and vice versa have been found by this method (V. Nazari, pers. comm.).

Conclusion

In conclusion, COI and EF1-α are both, and preferably in combination, useful as barcoding markers for most *Ectoedemia* s. str. species, including some cryptic species with small genetic distances. In addition diagnostic base pair positions are helpful for identification in both genes. In species complexes a combination of both markers will usually identify the species, by distance methods or diagnostic basepairs. Whereas in the past thresholds for species delimitation were often proposed and used (Hebert *et al*., 2003a; DeSalle *et al*., 2005), this method has been highly criticised and is now only used as a first indication of the possible presence of cryptic species. For future studies we suggest using two different (independent) markers with comparable resolution, preferably from different genomes, in concord with the advice by Zakharov *et al*. (2009). This will make it possible to rule out artefacts and anomalies caused by one marker, and strengthen patterns when both markers show the same topology. It will also provide additional phylogenetic information when the correct methods are used. In Nepticulidae and even in Lepidoptera in general, EF1-α is a good candidate as a second marker.

Acknowledgements

The authors thank Kees van den Berg (NCB Naturalis) for assistance with rearing, fieldwork and preparation of many of the adults.Many thanks go out to the following people for donating or lending samples for analysis: Giorgio Baldizzone (Asti, Italy), Bengt Å. Bengtsson (Färjestaden, Sweden), Kaj Berggren (Kristiansand, Norway), Willy Biesenbaum (Velbert-Langenberg, Germany), Rob Edmunds (Downham Market, United Kingdom), Willem Ellis (Amsterdam, Netherlands), Bob (R.J.) Heckford (Plymouth, United Kingdom), Henk ten Holt (Nijmegen, Netherlands), Martin R. Honey (London, Great Britain), Ronald Johansson (Växjö, Sweden), Jari

Junnilainen (Vantaa, Finland), Lauri Kaila (Helsinki, Finland), Timm Karisch (Dessau, Germany), Ole Karsholt (Kopenhagen, Denmark), Susumi Kawahara (Koshimizu, Japan), Sjaak Koster (Losser, Netherlands), Hiroshi Kuroko (Osaka, Japan), John Langmaid (Southsea, United Kingdom), Aleš Laštůvka (Prostejov, Czech Republic), Zdeňek Laštůvka (Brno, Czech Republic), Anna Mazurkiewicz (Kurów, Poland), Wolfram Mey (Berlin, Germany), Marko Mutanen (Oulu, Finland), Greg R. Pohl (Edmonton, Alberta), Yositaka Sakamaki (Kagoshima, Japan), Arnold Schreurs (Kerkrade, Netherlands), Rudi Seliger (Schwalmtal, Germany), Philip H. Sterling (Dorchester, United Kingdom), Andreas Stübner (Jänschwalde/Ost, Germany), Paolo Triberti (Verona, Italy), David L. Wagner (Storrs, USA), Andreas Werno (Landsweiler-Reden, Germany), Hugo van der Wolf (Nuenen, Netherlands), and Kazuo Yamazaki (Osaka, Japan). Many data in France and Italy were obtained during the EU funded EDIT WP 7 project 'All Taxa Biodiversity Inventories in the Mercantour/Alpi Marittime natural parks', we thank Marta de Biaggi (Valdieri, Italy) and Marie-France Leggia (Nice, France) for arranging permits. We thank the following persons and institutes for allowing collecting material in Vietnam: the director professor Le Xuan Canh and further professors May Phu Quy, Dr. Khuat Dang Long and Dr. Dzuong Huyen (Institute of Ecology and Biological Resources, Hanoi), in China: professor Liu You Qiao (Zoological Institute, Academia Sinica, Beijing) and in Indonesia, Kalimantan: Tropenbos International Indonesia (Balikpapan, Indonesia). René Glas and Bastian Reijnen (NCB Naturalis, Leiden) assisted with work in the molecular lab. The BOLD team at the Canadian Centre for DNA Barcoding (Guelph), particularly Rodolphe Rougerie, Megan Milton, Kara Layton and Claudia Steinke, are acknowledged for their assistance with the BOLD database and advice. Gerard van der Velde (Nijmegen, Netherlands) jointly supervised with EJvN the MSc thesis by CD that serves as basis for this paper. We are grateful to Barbara Gravendeel, Jan van Tol (both NCB Naturalis, Leiden), Felix Sperling (Edmonton, Canada) and two anonymous referees for comments on earlier versions of the manuscript.

References

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* 13: 729–744. doi: 10.1046/j.1365-294X.2003.02063.x.

Ballard JWO, Rand DM. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics* 36: 621–642. doi: 10.1146/annurev.ecolsys.36.091704.175513.

Bengtsson BÅ, Johansson R, Palmqvist G. 2008. *Fjärilar: Käkmalar - säckspinnare. Lepidoptera: Micropterigidae - Psychidae*. Uppsala: ArtDatabanken, SLU.

Blaxter ML. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B* 359: 669–679. doi: 10.1098/rstb.2003.1447.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London B* 360: 1935. doi: 10.1098/rstb.2005.1725.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *Journal of the Lepidopterists' Society* 61: 138–153.

Caterino MS, Cho S, Sperling FAH. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology* 45: 1–54. doi: 10.1146/annurev.ento.45.1.1.

CBOL Database Working Group. 2009. Data Standards for BARCODE Records in INSDC (BRIs), v.2.3. *Barcode of Life, identifying species with DNA barcoding*. [available at http://www.barcodeoflife.org/content/resources/standards-and-guidelines].

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106: 12794–12797. doi: 10.1073/pnas.0905 845106.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society of London B* 360: 1889. doi: 10.1098/rstb.2005.1720.

Cho S, Mitchell A, Regier JC, Mitter C, Poole RW, Friedlander TP, Zhao S. 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation factor-1 α recovers morphology-based tree for heliothine moths. *Molecular Biology and Evolution* 12: 650–656.

Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J. 2009. Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources* 10: 264–273. doi: 10.1111/j.1755-0998.2009.02763.x.

DeSalle R, Egan MG, Siddall M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London B* 360: 1905–1916. doi: 10.1098/rstb.2005.1722.

DeSalle R. 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. *Conservation Biology* 20: 1545–1547. doi: 10.1111/j.1523-1739.2006.00543.x.

Djernæs M, Damgaard J. 2006. Exon-intron structure, paralogy and sequenced regions of elongation factor-1 alpha in Hexapoda. *Arthropod Systematics & Phylogeny* 64: 45–52.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. 2006a. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* 103: 968–971. doi: 10.1073/pnas.0510466103.

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. 2006b. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6: 959–964. doi: 10.1111/j.1471-8286.2006.01470.x.

Hall TA. 2004. *BioEdit sequence alignment editor* 7.0.1. Isis Pharmaceuticals. [available at http://www.mbio.ncsu.edu/bioedit/bioedit.html].

Hausmann A, Hebert PDN, Mitchell A, Rougerie R, Manfred S, Edwards T, Young CJ. 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.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society Biological Sciences Series B* 270, 313–321. doi: 10.1098/rspb.2002.2218.

Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society Biological Sciences Series B* 270 (Supplement), S96–S99. doi: 10.1098/rsbl.2003.0025.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14812–14817. doi: 10.1073pnas.0406166101.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004b. Identification of Birds through DNA Barcodes. *PLoS Biology* 2: e312. doi: 10.1371/journal.pbio.0020312.

Hebert PDN, deWaard JR, Landry J-Fo. 2010. DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters* 6: 359–362. doi: 10.1098/rsbl.2009.0848.

Hulcr J, Miller SE, Setliff GP, Darrow K, Mueller ND, Hebert PDN, Weiblen GD. 2007. DNA barcoding confirms polyphagy in a generalist moth, *Homona mermerodes* (Lepidoptera: Tortricidae). *Molecular Ecology Notes* 7: 549–557. doi: 10.1111/j.1471-8286.2007.01786.x.

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society of London B* 360: 1835–1845. doi: 10.1098/rstb.2005.1715.

Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J-F, Lemaire

C, Miller JY, Miller JS, Miller L, E. Miller S, Montero J, Munroe E, Green SR, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources* 9: 1–26. doi: 10.1111/j.1755-0998.2009.02628.x.

Kaila L, Ståhls G. 2006. DNA barcodes: Evaluating the potential of COI to diffentiate closely related species of *Elachista* (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa* 1170: 1–26.

Knölke S, Erlacher S, Hausmann A, Miller MA, Segerer AH. 2005. A procedure for combined genitalia dissection and DNA extraction in Lepidoptera. *Insect Systematics and Evolution* 35: 401–409. doi: 10.1163/187631204788912463.

Kristensen NP, Scoble MJ, Karsholt O. 2007. Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. *Zootaxa* 1668: 699–747.

Lukhtanov VA, Sourakov A, Zakharov EV, Hebert PDN. 2009. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. *Molecular Ecology Resources* 9: 1302–1310. doi: 10.1111/j.1755-0998.2009.02577.x.

Matheson CD, Muller GC, Junnila A, Vernon K, Hausmann A, Miller MA, Greenblatt C, Schlein Y. 2007. A PCR method for detection of plant meals from the guts of insects. *Organisms Diversity and Evolution* 7: 294–303. doi: 10.1016/j.ode.2006.09.002.

Menken SBJ, Wiebosch Steeman M. 1988. Clonal diversity, population structure, and dispersal in the parthenogenetic moth *Ectoedemia argyropeza*. *Entomologia Experimentalis et Applicata* 49: 141–152. doi: 10.1111/j.1570-7458.1988.tb02485.x.

Menken SBJ. 1990. Biochemical systematics of the leaf mining moth family Nepticulidae (Lepidoptera). 3. Allozyme variation patterns in the *Ectoedemia subbimaculella* group. *Bijdragen tot de Dierkunde* 60: 189–197.

Menken SBJ, Boomsma JJ, van Nieukerken EJ. 2010. Large-scale evolutionary patterns of host plant associations in the Lepidoptera. *Evolution* 64: 1098–1119. doi: 10.1111/j.1558-5646.2009.00889.x.

Mey W. 2004. Nepticulidae (Lepidoptera, Nepticuloidea). Pp. 27–38 in: Mey W, ed., *The Lepidoptera of the Brandberg Massif in Namibia Part 1. Esperiana Memoir 1.*

Miller MA, Muller GC, Kravchenko VD, Junnila A, Vernon KK, Matheson CD, Hausmann A. 2006. DNA-based identification of Lepidoptera larvae and plant meals from their gut contents. *Russian Entomological Journal* 15: 427–432.

Nieukerken EJ van. 1985. A taxonomic revision of the western Palaearctic species of the subgenera *Zimmermannia* Hering and *Ectoedemia* Busck s. str. (Lepidoptera, Nepticulidae), with notes on their phylogeny. *Tijdschrift voor Entomologie* 128: 1–164. http://biodiversitylibrary.org/page/30334755.

Nieukerken EJ van. 1986. Systematics and phylogeny of Holarctic genera of Nepticulidae (Lepidoptera, Heteroneura: Monotrysia). *Zoologische Verhandelingen* 236: 1–93. http://www.repository.naturalis.nl/record/317599.

Nieukerken EJ van, Laštůvka A, Laštůvka Z. 2010. Western Palaearctic *Ectoedemia* (*Zimmermannia*) Hering and *Ectoedemia* Busck s. str. (Lepidoptera: Nepticulidae): five new species and new data on distribution, hostplants and recognition. *ZooKeys* 32: 1–82. doi: 10.3897/zookeys.32.282.

Nieukerken EJ van, Berggren K. 2011. *Ectoedemia rosae*, a new species from the French Alps and Norway (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie* 154: 181–191.

Nieukerken EJ van, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen S-H, Zahiri R, Adamski D, Baixeras J, Bartsch D, Bengtsson BÅ, Brown JW, Bucheli SR, Davis DR, De Prins J, De Prins W, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara A, Koster SJC, Kozlov M, Lafontaine JD, Lamas G, Landry J-F, Lee S, Nuss M, Park KT, Penz C, Rota J, Schmidt BC, Schintlmeister A, Sohn JC, 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.

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: 113–124. doi: 10.1111/j.1440-6055.2009. 00695.x.

Puplesis R. 1994. *The Nepticulidae of eastern Europe and Asia. Western, central and eastern parts*. Leiden: Backhuys Publishers.

Puplesis R, Robinson GS. 2000. A review of the Central and South American Nepticulidae (Lepidoptera) with special reference to Belize. *Bulletin of the Natural History Museum London. Entomology* 69: 3–114.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B-Biological Sciences* 275: 237–247. doi: 10.1098/rspb.2007.1290.

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364. doi: 10.1111/j.1471-8286.2006.01678.x.

Regier JC, Shi D. 2005. Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5′ tails. *Biotechniques* 38: 34–38.

Roe AD, Sperling FAH. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* 44: 325–345. doi: 10.1016/j.ympev.2006.12.005.

Rubinoff D, Cameron S, Will K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity* 97: 581–594. doi: 10.1093/jhered/esl036.

Santamaria M, Vicario S, Pappada G, Scioscia G, Scazzocchio C, Saccone C. 2009. Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. *BMC Bioinformatics* 10: S15. doi: 10.1186/1471-2105-10-S6-S15.

Schoch C, Seifert K, Crous P. 2011. Progress on DNA barcoding of Fungi. *IMA Fungus* 2: v–vi.

Scoble MJ. 1978. Nepticulidae (Lepidoptera) of southern Africa: the genus *Ectoedemia* Busck. *Journal of the Entomological Society of Southern Africa* 41: 81–86.

Scoble MJ. 1979. A new species of *Ectoedemia* Busck from Botswana with observations on its imaginal skeletal anatomy (Lepidoptera: Nepticulidae). *Annals of the Transvaal Museum* 32: 35–54. [available at: http://hdl.handle.net/10499/AJ1828].

Song H, Buhay JE, Whiting MF, Crandall KA. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America* 105: 13486–13491. doi: 10.1073/pnas.0803076105.

Sonnenberg R, Nolte A, Tautz D. 2007. An evaluation of LSU rDNA D1–D2 sequences for their use in species identification. *Frontiers in Zoology* 4: 6. doi: 10.1186/1742-9994-4-6.

Stone GN, Challis RJ, Atkinson RJ, Csóka G, Hayward A, Melika G, Mutun S, Preuss S, Rokas A, Sadeghi E, Schönrogge K. 2007. The phylogeographical clade trade: tracing the impact of human-mediated dispersal on the colonization of northern Europe by the oak gallwasp *Andricus kollari*. *Molecular Ecology* 16: 2768–2781. doi: 10.1111/j.1365-294X. 2007.03348.x.

Swofford DL. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4b10. Sunderland MA: Sinauer Associates.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739. doi: 10.1093/molbev/msr121.

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

Wilkinson C. 1981. A supplement to the genus *Ectoedemia* Busck (Nepticulidae: Lepidoptera) in North America, dealing with some difficult species and also some new ones. *Tijdschrift voor Entomologie* 124: 93–110. http://biodiversitylibrary.org/page/28244986.

Wilkinson C, Scoble MJ. 1979. The Nepticulidae (Lepidoptera) of Canada. *Memoirs of the Entomological Society of Canada* 111: 1–129. doi: 10.4039/entm111107fv.

Wilkinson C, Newton PJ. 1981. The micro-lepidopteran genus *Ectoedemia* Busck (Nepticulidae) in North America. *Tijdschrift voor Entomologie* 124: 27–92. http://biodiversitylibrary.org/page/28244920.

Wilkinson C, Bryan G, Menken SBJ, Nieukerken EJ van. 1983. A clarification of the status of four taxa in the *Ectoedemia angulifasciella* group (Nepticulidae: Lepidoptera). *Netherlands Journal of Zoology* 33: 211–224. doi: 10.1163/002829683X00093.

Will KW, Rubinoff D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47–55. doi: 10.1111/j.1096-0031.2003.00008.x.

Wilson JJ, Rougerie R, Schonfeld J, Janzen DH, Hallwachs W, Hajibabaei M, Kitching IJ, Haxaire J, Hebert PDN. 2011. When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology* 11: 18. doi: 10.1186/ 1472-6785-11-18.

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: 711–723. doi: 10.1016/j.ympev.2006.12.027.

Zakharov EV, Lobo NF, Nowak C, Hellmann JJ. 2009. Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae). *Heredity* 102: 590–599. doi: 10.1038/hdy.2009.26.

Zhang D-X, Hewitt GM. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution* 11: 247–251. doi: 10.1016/0169-5347(96)10031-8.

Online supplementary information (SI)

S1. K2P Neighbor joining tree of COI barcode sequences for all treated *Ectoedemia* species, except the species complexes *E. rubivora, E. albifasciella* and *E. subbimaculella*, which are collapsed and presented in Figs 3, 5-6. Specimen numbers are RMNH.INS registry numbers. Sequences of *Ectoedemia (Zimmermannia)* species were used as outgroup for the tree, bootstrap values represent 10,000 replicates.
http://www.contributionstozoology.nl/c/ctz/supp/8101a01S1.pdf

S2. K2P Neighbor joining tree of partial EF1-α sequences for all treated *Ectoedemia* species, except the species complexes *E. rubivora, E. albifasciella* and *E. subbimaculella*, which are collapsed. See Figs 3, 5-6 for those. Specimen numbers are RMNH.INS registry numbers. Sequences of *Ectoedemia (Zimmermannia)* species were used as outgroup for the tree, bootstrap values represent 10,000 replicates.
http://www.contributionstozoology.nl/c/ctz/supp/8101a01S2.pdf

S3. All sequenced samples used with identification, Sample ID (= voucher registry), BOLD processed, GenBank codes, GBIF portal URL and details on Stage, Country, province, collection, note and number of traces and images present. Further images will be posted to this site at a later date. All details can be consulted on the BOLD site (http://www.barcodinglife.com/) under the public project 'Nepticulidae - Ectoedemia – Public records [NEPEC]'.
http://www.contributionstozoology.nl/c/ctz/supp/8101a01S3.xlsx

Appendix

List of *Ectoedemia* species studied, with complete nomenclature, hostplants and distribution data. For both COI and EF1-α, mean and maximum intraspecific K2P distances are provided as well as the distance between nearest neighbours (K2P model) and the name of the nearest

			COI
Species	**Hostplant**	**Region**	**# sequences**
West Palearctic species			
Subgenus *Zimmermannia* Hering, 1940			
Ectoedemia longicaudella (Klimesch, 1953)	*Quercus*	West Palearctic	1
Ectoedemia liebwerdella Zimmermann, 1940	*Fagus*	West Palearctic	1
Ectoedemia amani Svensson, 1966	*Ulmus*	West and East Palearctic	2
Ectoedemia reichli Laštůvka, Z. & A., 1998		West Palearctic	1
Subgenus *Ectoedemia* Busck, 1907			
terebinthivora group			
Ectoedemia terebinthivora (Klimesch, 1975)	*Pistacia terebinthus*	West Palearctic	2
angulifasciella group			
Ectoedemia erythrogenella (Joannis, 1908)	*Rubus*	West Palearctic	5
Ectoedemia spiraeae Gregor & Povolny, 1983	*Spiraea*	West and East Palearctic	2
Ectoedemia agrimoniae (Frey, 1858)	*Agrimonia*	West Palearctic	3
Ectoedemia rosae Van Nieukerken & Berggren, 2011	*Rosa*	West Palearctic	4
Ectoedemia angulifasciella (Stainton, 1849)	*Rosa*	West Palearctic	18
Ectoedemia atricollis (Stainton, 1857)	*Rosaceae*: various trees	West Palearctic	9
Ectoedemia arcuatella (Herrich-Schäffer, 1855)	*Fragaria, Potentilla*	West Palearctic	6
Ectoedemia rubivora (Wocke, 1860)	*Rubus*	West Palearctic	11
Ectoedemia spinosella (Joannis, 1908)	*Prunus*	West Palearctic	5
Ectoedemia mahalebella (Klimesch, 1936)	*Prunus*	West Palearctic	6
Ectoedemia occultella (Linnaeus, 1767)	*Betula*	West and East Palearctic, North America	5
Ectoedemia minimella (Zetterstedt, 1839)	*Betula*	West Palearctic	3
suberis group			
Ectoedemia aegilopidella (Klimesch, 1978)	*Quercus ithaburensis*	West Palearctic	1
Ectoedemia caradjai (Groschke, 1944)	*Quercus* deciduous	West Palearctic	3
Ectoedemia suberis (Stainton, 1869)	*Quercus* evergreen	West Palearctic	3
Ectoedemia hendrikseni A. Laštůvka, Z. Laštůvka & Van Nieukerken, 2010	*Quercus* evergreen	West Palearctic	2
Ectoedemia andalusiae Van Nieukerken, 1985	*Quercus* evergreen	West Palearctic	4
Ectoedemia heckfordi Van Nieukerken, A. Laštůvka & Z. Laštůvka, 2010	*Quercus* deciduous	West Palearctic	3
Ectoedemia phaeolepis Van Nieukerken, A. Laštůvka & Z. Laštůvka, 2010	*Quercus* evergreen	West Palearctic	2
populella group			
Ectoedemia intimella (Zeller, 1848)	*Salix*	West and East Palearctic	6
Ectoedemia hannoverella (Glitz, 1872)	*Populus nigra, P. canadensis*	West and East Palearctic	4
Ectoedemia turbidella (Zeller, 1848)	*Populus alba, P. canescens*	West Palearctic	4
Ectoedemia klimeschi (Skala, 1933)	*Populus alba*	West Palearctic	2
Ectoedemia argyropeza (Zeller, 1839)	*Populus tremula*	West and East Palearctic, North America	6
preisseckeri group			
Ectoedemia preisseckeri (Klimesch, 1941)	*Ulmus*	West and East Palearctic	2
subbimaculella group			
Ectoedemia quinquella (Bedell, 1848)	*Quercus* deciduous	West Palearctic	3
Ectoedemia algeriensis Van Nieukerken, 1985	*Quercus* evergreen	West Palearctic	2
Ectoedemia coscoja Van Nieukerken, A. Laštůvka & Z. Laštůvka, 2010	*Quercus* evergreen	West Palearctic	1
Ectoedemia gilvipennella (Klimesch, 1946)	*Quercus* deciduous	West Palearctic	2
Ectoedemia leucothorax Van Nieukerken, 1985		West Palearctic	3

neighbour (species). Non applicable values are given as a dash. Darker colours indicate values that are below (Max Intra-sp) or above (Distance to NN) the barcode threshold of 2%.

COI	COI	COI	COI	COI	EF1-α	EF1-α	EF1-α	EF1-α	EF1-α	EF1-α
of which short	Mean intra-sp	Max intra-sp	Nearest species	Distance to NN	# sequences	of which short	Mean intra-sp	Max intra-sp	Nearest species	Distance to NN
	–	–	*E. liebwerdella*	4.85	0					
	–	–	*E. longicaudella*	4.85	1		–	–	*E. admiranda*	1.21
	0.15	0.15	*E. liebwerdella*	8.6	1		–	–	*E. liebwerdella*	8.12
	–	–	*E. longicaudella*	6.77	0					
	0	0	*E. caradjai*	6.27	2		0	0	*E. preisseckeri*	6.71
	1.6	3.25	*E. arcuatella*	8.09	5		0.17	0.42	*E. aegilopidella*	7.13
	6.82	6.82	*E. agrimoniae*	5.79	2	2	2.47	2.47	*E. agrimoniae*	5.01
	0.82	1.25	*E. spiraeae*	5.79	2		0	0	*E. spiraeae*	5.01
	0.99	1.52	*E. ivinskisi*	7.46	3		0.42	0.42	*E. aegilopidella*	6.69
	0.37	0.77	*E. arcuatella*	7.27	17		0.14	0.84	*E. rubivora*	2.99
	0.07	0.31	*E. rubivora*	0	9		0	0	*E. rubivora*	0.42
	0.58	0.91	*E. atricollis*	0.15	6		0	0	*E. rubivora*	0.21
	0.35	0.93	*E. atricollis*	0	11		0.09	0.42	*E. arcuatella*	0.21
	0.74	1.89	*E. mahalebella*	5.95	5		0.67	1.68	*E. mahalebella*	2.98
	0.15	0.3	*E. spinosella*	5.95	5		0.08	0.21	*E. spinosella*	2.98
	0.15	0.3	*E. caradjai*	5.64	5		0.08	0.21	*E. minimella*	4.97
	0.1	0.15	*E. caradjai*	6.44	3		0.28	0.42	*E. occultella*	4.97
1	–	–	*E. andalusiae*	5.36	1	1	–	–	*E. heckfordi*	3.69
	1.01	1.37	*E. chasanella*	4.65	3		0.28	0.42	*E. chasanella*	3.19
	0.3	0.45	*E. caradjai*	5.3	2		0	0	*E. heckfordi*	2.76
	0.6	0.6	*E. andalusiae*	4.98	2		0.63	0.63	*E. phaeolepis*	3.62
	0.61	0.91	*E. hendrikseni*	4.98	4		0	0	*E. heckfordi*	3.86
	0.3	0.45	*E. hendrikseni*	5.63	3		0.14	0.21	*E. phaeolepis*	2.33
	0.3	0.3	*E. hendrikseni*	5.14	2		0.42	0.42	*E. heckfordi*	2.33
	2.47	7.1	*E. klimeschi*	6.11	6	1	0.4	1.2	*E. canutus*	4.57
	0.91	1.53	*E. klimeschi*	4.18	5		0.25	0.63	*E. turbidella*	4.29
	0.15	0.3	*E. argyropeza*	3.55	4		0	0	*E. argyropeza*	1.47
	0.15	0.15	*E. argyropeza*	2.3	3		0.28	0.42	*E. turbidella*	1.9
	0.35	1.06	*E. klimeschi*	2.3	7		0	0	*E. turbidella*	1.47
	0	0	*E. phyllotomella*	5.94	3	2	0	0	*E. alnifoliae*	3.69
	0.1	0.15	*E. coscoja*	3.64	3		0	0	*E. algeriensis*	2.97
	0.15	0.15	*E. gilvipennella*	3.71	3	1	0.54	0.8	*E. coscoja*	2.75
	–	–	*E. quinquella*	3.64	2		0	0	*E. algeriensis*	2.75
	0.3	0.3	*E. algeriensis*	3.71	2		0.63	0.63	*E. algeriensis*	3.85
	0.1	0.15	*E. subbimaculella*	4.49	3		0.42	0.63	*E. haraldi*	5.41

List of *Ectoedemia* species (continued)

			COI
Species	**Hostplant**	**Region**	**# sequences**
Ectoedemia haraldi (Soffner, 1942)	*Quercus* evergreen	West Palearctic	4
Ectoedemia ilicis (Mendes, 1910)	*Quercus* evergreen	West Palearctic	4
Ectoedemia pseudoilicis Laštůvka, Z. & A., 1998	*Quercus* evergreen	West Palearctic	3
Ectoedemia heringella (Mariani, 1939)	*Quercus* evergreen	West Palearctic	8
Ectoedemia alnifoliae Van Nieukerken, 1985	*Quercus* evergreen	West Palearctic	3
Ectoedemia rufifrontella (Caradja, 1920)	*Quercus* deciduous	West Palearctic	4
Ectoedemia albifasciella (Heinemann, 1871)	*Quercus* deciduous	West Palearctic	22
Ectoedemia cerris (Zimmermann, 1944)	*Quercus cerris*	West Palearctic	3
Ectoedemia pubescivora (Weber, 1937)	*Quercus pubescens*	West Palearctic	4
Ectoedemia contorta Van Nieukerken, 1985	*Quercus* deciduous	West Palearctic	3
Ectoedemia subbimaculella (Haworth, 1828)	*Quercus* deciduous	West Palearctic	16
Ectoedemia heringi (Toll, 1934)	*Quercus* deciduous, Castanea	West Palearctic	17
Ectoedemia liechtensteini (Zimmermann, 1944)	*Quercus cerris*	West Palearctic	3
Ectoedemia phyllotomella (Klimesch, 1946)	*Quercus cerris*	West Palearctic	2
Species from other regions			
Subgenus *Zimmermannia* Hering, 1940			
Ectoedemia admiranda Puplesis, 1984		East Palearctic	
Subgenus *Ectoedemia* Busck, 1907			
angulifasciella group			
Ectoedemia Rubus Borneo	*Rubus*	Oriental	1
Ectoedemia Rubus Cuc Phuong	*Rubus*	Oriental	1
Ectoedemia Rubus Fansipan	*Rubus*	Oriental	1
Ectoedemia TamDao		Oriental	1
Ectoedemia picturata Puplesis, 1985	*Rosa*	East Palearctic	4
suberis group			
Ectoedemia ivinskisi Puplesis, 1984	*Quercus deciduous*	East Palearctic	2
Ectoedemia ornatella Puplesis, 1984	*Quercus deciduous*	East Palearctic	3
Ectoedemia chasanella (Puplesis, 1984)		East Palearctic	1
populella group			
Ectoedemia populella (Busck, 1907)	*Populus*	Nearctic	2
Ectoedemia canutus Wilkinson & Scoble, 1979	*Populus*	Nearctic	1
Ectoedemia California		Nearctic	1
subbimaculella group			
Ectoedemia Quercus gilva	*Quercus gilva*	East Palearctic	3
not assigned to group			
Ectoedemia quadrinotata (Braun, 1917)	Betulaceae	Nearctic	2
Ectoedemia arisi Puplesis, 1984		East Palearctic	1
Ectoedemia christophori Puplesis, 1985		East Palearctic	
Ectoedemia Annomocarya Vietnam	*Annomocarya*	Oriental	1
Ectoedemia Carpinus Vietnam	*Carpinus*	Oriental	1
Ectoedemia expeditionis Mey, 2004		Afrotropical	1
Ectoedemia tersiusi Mey, 2004		Afrotropical	1
Ectoedemia Namibia		Afrotropical	1
Total sequences			262

COI	COI	COI	COI	COI	EF1-α	EF1-α	EF1-α	EF1-α	EF1-α	EF1-α
of which short	Mean intra-sp	Max intra-sp	Nearest species	Distance to NN	# sequences	of which short	Mean intra-sp	Max intra-sp	Nearest species	Distance to NN
	3.16	4.22	*E. pseudoilicis*	3.54	4	2	0.69	2.53	*E. pseudoilicis*	2.02
	0.3	0.61	*E. heringella*	3.07	4		0.17	0.42	*E. heringella*	0.8
	2.21	3.24	*E. heringella*	2.77	2		0.21	0.21	*E. heringella*	1.68
	0.5	0.91	*E. pseudoilicis*	2.77	6	1	0.16	0.4	*E. ilicis*	0.8
	2.05	2.62	*E. pseudoilicis*	3.86	3		0.14	0.21	*E. haraldi*	3.28
	1.02	1.84	*E. phyllotomella*	3.64	3	1	0	0	*E. albifasciella*	1.61
	0.8	2.3	*E. pubescivora*	1.22	22		0.12	0.42	*E. cerris*	0
1	0	0	*E. albifasciella*	1.84	2	1	0	0	*E. albifasciella*	0
	0.08	0.15	*E. albifasciella*	1.22	2		0	0	*E. albifasciella*	0.21
	0.61	0.76	*E. albifasciella*	1.37	3	1	0.2	0.4	*E. cerris*	0.4
	1.2	3.4	*E. heringi*	0.3	15	1	0.62	1.62	*E. heringi*	0.63 (0)
	1.45	3.41	*E. phyllotomella*	0.15	12	1	0.66	1.89	*E. liechtensteini/phyllotomella*	0
	0.2	0.31	*E. phyllotomella*	0.3	2	1	0.4	0.4	*E. heringi*	0
	0	0	*E. heringi*	0.15	2		0	0	*E. heringi*	0
					1	1	–	–	*E. liebwerdella*	1.21
			E. Rubus Fansipan	5.78	1		–	–	*E. Rubus* Fansipan	4.08
			E. Rubus Fansipan	6.43						
			E. Rubus Borneo	6.25	1		–	–	*E.* TamDao	2.11
			E. Rubus Cuc Phuong	7.23	1		–	–	*E. Rubus* Fansipan	2.11
4	0.38	0.65	*E. spiraeae*	7.46	2	2	0	0	*E. aegilopidella*	5.83
1	–	–	*E. caradjai*	5.78	1		–	–	*E. aegilopidella*	7.61
1	0	0	*E. chasanella*	7.26	2	2	0	0	*E. caradjai*	4.99
	–	–	*E. caradjai*	4.65	1		–	–	*E. caradjai*	3.19
1	–	–	*E. canutus*	7.11	1	1	–	–	*E. canutus*	4.54
	–	–	*E. klimeschi*	4.66	1		–	–	*E. populella*	4.54
	–	–	*E. klimeschi*	4.08	1		–	–	*E. turbidella*	2.11
	0.2	0.3	*E. heringi*	4.42	3		0	0	*E. albifasciella*	4.07
	0.15	0.15	*E. heringi*	6.6	2	2	1.21	1.21	*E. christophori*	6.71
	–	–	*E. heringi*	5.65	1	1	–	–	*E. christophori*	3.28
			-	-	1				*E. arisi*	3.28
	–	–	*E. cerris*	5.42	1		–	–	*E. preisseckeri*	5.01
	–	–	*E. cerris*	4.18						
	–	–	*E. agrimoniae*	7.85	1		–	–	*E. tersiusi*	7.92
	–	–	*E. agrimoniae*	7.85	1		–	–	*E. expeditionis*	7.92
1	–	–	*E. spiraeae*	7						
10					240	25				

3

A Global Phylogeny of Leafmining *Ectoedemia* Moths

A Global Phylogeny of Leafmining *Ectoedemia* Moths (Lepidoptera: Nepticulidae): Exploring Host Plant Family Shifts and Allopatry as Drivers of Speciation

Camiel Doorenweerd[1],[2*], Erik J. van Nieukerken[1], Steph B. J. Menken[2]

[1] *Department of Terrestrial Zoology, Naturalis Biodiversity Center, Leiden, The Netherlands*
[2] *Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands*
[*] *camiel.doorenweerd@naturalis.nl*

Abstract

Background
Host association patterns in *Ectoedemia* (Lepidoptera: Nepticulidae) are also encountered in other insect groups with intimate plant relationships, including a high degree of monophagy, a preference for ecologically dominant plant families (e.g. Fagaceae, Rosaceae, Salicaceae, and Betulaceae) and a tendency for related insect species to feed on related host plant species. The evolutionary processes underlying these patterns are only partly understood, we therefore assessed the role of allopatry and host plant family shifts in speciation within *Ectoedemia*.

Methodology
Six nuclear and mitochondrial DNA markers with a total aligned length of 3692 base pairs were used to infer phylogenetic relationships among 92 species belonging to the subgenus *Ectoedemia* of the genus *Ectoedemia*, representing a thorough taxon sampling with a global coverage. The results support monophyletic species groups that are congruent with published findings based on morphology. We used the obtained phylogeny to explore host plant family association and geographical distribution to investigate if host shifts and allopatry have been instrumental in the speciation of these leafmining insects.

Significance
We found that, even though most species within species groups commonly feed on plants from one family, shifts to a distantly related host family have occasionally occurred throughout the phylogeny and such shifts are most commonly observed towards Betulaceae. The largest radiations have occurred within species groups that feed on Fagaceae, Rosaceae, and Salicaceae. Most species are restricted to one of the seven global biogeographic regions, but within species groups representatives are commonly found in different biogeographic regions. Although we find general patterns with regard to host use and biogeography, there are differences between clades that suggest that different drivers of speciation, and perhaps drivers that we did not examine, have shaped diversity patterns in different clades.

Introduction

Insect herbivores constitute the most species-rich group of insects and more than half of the world's known species are insects [1]. However, how evolutionary processes in insect-plant networks have shaped radiations in time and space is yet little understood [2–5]. As phylogenetic studies of insects as well as plants are becoming more complete in terms of taxon sampling, more robust by including more genetic markers and better dated with molecular clock methods, a complete picture of species interaction patterns as well as a more complete understanding of the relative importance of the various evolutionary factors that drive herbivorous insect radiations is coming within reach [5–7]. The two most salient characteristics of plant-insect associations were unveiled long before the phylogenetics era, but were recently corroborated by molecular studies, viz. 1) related insects tend to feed on related plants and 2) insect diet breadth decreases as interactions become more intimate, such as with internal feeding [3, 8]. Yet, strict co-cladogenesis between insects and their plant hosts is rare, and occasional host shifts towards distantly related plants occur throughout most groups [2, 9].

Multiple phylogenetic studies that incorporated molecular clock analyses have shown that contemporaneous parallel cladogenesis has not been common in the evolution of plants and their herbivorous insects [10–15]. Also little evidence has been found for 'asynchronous parallel cladogenesis', where plants diversified first and herbivores later mirrored the speciation events of their hosts through resource tracking [16]. Instead a 'resource archipelago' scenario seems to be realistic [17]. In this scenario, plants have radiated independently from herbivores. Herbivore radiation on the plant hosts occurred later, at least partly through ecological speciation and, as a result, the herbivore phylogeny mirrors the plant phylogeny only in part. During ecological speciation new species arise after diverging in resource use, which for herbivores is commonly a dietary change, and this may happen in sympatry or allopatry [18–20]. It has been argued quite a few times in the past century, however, that ecological factors driving speciation are often overestimated, whereas geographic factors are underestimated [16, 21, 22].

Phylogenies can be used to differentiate between the two speciation scenarios: in cases of ecological speciation, there will be resource-shifts between sister-species, but if allopatric speciation is the rule, the resources will remain the same [17]. Food resources for phytophagous insects constitute many different resource dimensions. Differences occur for example within a single plant between tissue types and developmental stages, between individual plants, between different parts of a plant population, within a single plant species in different seasons or between species [17]. Research on ecological speciation has long focussed on plant secondary metabolites, and the ability to digest and/or detoxify these by insects. For some groups in which both the herbivore and plant evolution has been studied in detail, plant chemistry has indeed been shown to drive speciation, at least partly, in both the insect and the plant [23, 24]. However, it is also becoming increasingly clear that for many groups of herbivores a wider range of resource dimensions may drive speciation [17, 18].

Changes in feeding mode or changes in host use are some of the evolutionary events of an herbivorous insect species that involve different resource dimensions. Feeding mode is the term loosely defining the combination of the plant part that is used (e.g. leaf, stem, fruit) and the method of residing in, or consuming, the plant tissue (e.g. external feeding, galling, mining). The feeding mode is constrained by morphological traits of the insect such as the shape and functionality of the mouthparts, but also for example the shape of the ovipositor, of which the dimensions determine the plant tissue that can be reached (e.g. a long and thin ovipositor can oviposit between long and dense hairs). We can divide changes in host use in close, intermediate and distant host changes. For monophagous insects a close change in host use generally involves a plant species within the same plant genus, an intermediate change a different plant genus in the same family and a distant change involves another plant family. As the distance to a different resource increases, utilizing that resource will require an increasing number of simultaneous adaptations to the divergent phytochemistry, different epidermal characteristics and different pattern of interactions with other taxa of the new host [25–29]. Although the odds of a successful distant host change are slim due to the many adaptive requirements, they may be enabled by increased rates of evolutionary change, processes which we are only beginning to understand [30]. According to the "enemy-free space" hypothesis, such changes in the evolutionary rate of change can be linked to the release of predation pressures, or they might be linked to reduced competition for resources, which may counteract the initially lower fitness on a new host [27, 31]. Diversification may follow from a distant host shift when this new resource creates access to many resources at intermediate distances. The intermediate resource hypothesis [17] predicts that speciation rates will be highest in situations in which herbivorous insects with limited resource usage (i.e. monophages) are in potential contact with a suite of resources at intermediate resource distances. However, one of the few cases where this was actually tested, a study on the diversification of the subgenus Drosophila, showed that speciation rates were not affected by changes in resource use [32].

Although changes in resource use and geographic distribution are two known drivers of diversification in phytophagous insect evolution [5], other candidate drivers should not be discarded beforehand. Increased mortality rates, such as those caused by parasitoid wasps or spiders, can be a strong selection pressure, as was shown by a study on jumping spiders predating on choreutid moths where species that mimicked the appearance and behaviour of the spiders had a higher survival rate [33]. If selective pressures are not continuously present throughout the distribution area of a species, they may lead to disruptive selection. Climatic events are known to have had a strong impact on ecological communities and may have caused bursts of taxonomic diversification in herbivorous insects as well [34]. Furthermore, the factors that drive speciation may influence each other directly or indirectly. Climatic conditions could have created ecological opportunities for host shifts during periods in which different host ranges overlapped. Following from the "enemy-free space" hypothesis, predation pressures may have been a strong selective pressure that favoured individuals that changed their host and were thereby temporarily released from the predation pressure [31, 35].

Because many of the aforementioned possible drivers of speciation potentially interacted and may have had a different effect for different groups or at different time periods, disentangling causality from mere correlation among the factors that potentially structure insect-plant networks is a challenging endeavour. This study addresses the phylogenetic relationships within *Ectoedemia* s. str., a subgenus of leafmining moths, and explores the importance of host shifts and allopatry as drivers for speciation. Larvae of leafmining insects are found within plant tissues and are not able to change plant during the larval stage. As a consequence, the host is selected only by the ovipositing female on the basis of visual, olfactory and/or contact chemosensory cues. Her offspring have the choice between surviving on the plant chosen by their mother or perish. There is a significantly higher degree of monophagy in leafminers compared to externally feeding insects, likely because the host not only represents a food resource, but also forms their complete larval habitat. This requires a combination of adapted traits to the various dimensions of the resource [3, 8, 36]. Ecological factors therefore likely play an important role in speciation in *Ectoedemia*, through changes in host use or one of the dimensions thereof.

With 91 described species, *Ectoedemia* is the largest subgenus within the nominate genus *Ectoedemia* in the leafmining moth family Nepticulidae. All *Ectoedemia* s. str. are leafminers, with one exception of a gall former, whereas some other subgenera display different feeding modes (viz. bark or fruit mining). *Ectoedemia* s. str. is best known from the West Palearctic with 48 described species. From the East Palearctic and Nearctic regions each about 20 species have been recorded. A smaller number of species is known from Africa, Central America and the Oriental region [37–40]. *Ectoedemia* s. str. seems absent from Australia, the Pacific and most of South America, with only a single species known from Belize and Ecuador. In tropical Asia many undescribed species have been collected recently, and will be described in the near future. In North America, only a few more unnamed species are known from collections and recent collecting efforts by the authors and colleagues, suggesting that sampling in that area is now fairly complete. Most *Ectoedemia* s. str. feed on rosid plant orders [41], with the Nearctic as the only exception where some species feed on *Platanus* (Proteales) and *Nyssa* (Cornales). The vast majority of rosid-feeding species is found on just four plant families: Fagaceae, Rosaceae, Salicaceae and Betulaceae. These plant families are ecologically dominant and widespread, particularly in the deciduous forests of the northern hemisphere. *Ectoedemia* s. str. thereby follows the host association patterns of many insect herbivores [3, 42]. The resemblance to the gracillariid leafmining genus *Phyllonorycter* and the sawfly subfamily Nematinae (Tenthredinidae, Hymenoptera) is striking, because both these groups also have Fagaceae, Rosaceae, Salicaceae and Betulaceae represented in its top five of most used host plant families [43, 44]. Most *Ectoedemia* are monophagous, commonly restricted to a single host genus or even a single host species. Some species within these clades with mostly monophagous species, however, display remarkable diet breadth, e.g. the European *Ectoedemia atricollis*, which not only feeds on a number of rosaceous tree genera, but also on the unrelated genus *Staphylea* (Staphyleaceae) [40].

A phylogenetic study based on morphology has revealed a number of monophyletic species groups in European *Ectoedemia*, of which all species utilise the same host family [40]. Also in North America and the East Palearctic, species have been grouped following morphological similarity, but these groups have been created in regional revisions without evaluating relationships with species from other regions, and have not been analysed following cladistic principles [45, 46]. Here, we use a dataset resulting from a largely complete global taxon sampling that includes 92 species from the entire distribution area of the subgenus. Six nuclear and mitochondrial molecular markers are analysed to reconstruct the phylogeny and to test the monophyly of species groups. Thereafter, we superimposed host association and geographic distribution on the obtained tree to evaluate hypotheses on processes of speciation, such as geographic isolation and resource shifts.

Materials and Methods

Taxon sampling and marker selection

Taxon sampling of West Palearctic *Ectoedemia* is almost complete, with 44 out of the 47 European species present in our phylogenetic data set, only lacking *Ectoedemia hexapetalae*, *E. similigena* and *E. aegilopidella*. From the Nearctic 18 species are analysed, including four undescribed ones, only missing two of the named species, viz. *E. marmaropa* and *E.canadensis* [46, 47]. From the East Palearctic we include 12 species and from tropical Asia 26, of which the majority is undescribed [the separation between these biogeographic regions is somewhat arbitrary, here we include all of Taiwan in tropical Asia, although some species are shared with Japan and mainland China]. Furthermore, three species from Africa are included, one unnamed and two named, whereas five named species are known from this area [39, 48]. We did not manage to include the only named Neotropical species, *E. fuscovittata* [49], the generic assignment of which is somewhat uncertain. Specimens included in this study are all registered on the Barcoding of Life Database (www.boldsystems.org/) under the project "*Ectoedemia* of the World" with detailed information on collecting localities and identification (also available in S1 Table).

The following institutes gave or organised permission for fieldwork and export of specimens in their respective areas: the Mercantour and Alpi Marittime parks in France and Italy (ATBI Mercantour / Alpi Marittime project), the Great Smoky Mountains National Park in the USA, the Institute of Ecology and Biological Resources, Hanoi in Vietnam, the Zoological Institute, Academia Sinica, Beijing in China, the Tropenbos International Indonesia, Balikpapan, Indonesia in Kalimantan, the Korean National Herbarium, Pocheon in South Korea and the National Sun Yat-Sen University, Kaohshiung in Taiwan.

For species that are undescribed we use provisional names, often indicating the distribution area and/or host plant, shown in text between quotation marks; these names also figure in the BOLD database and S1 Table and S5 Table without

quotation marks. From the *Ectoedemia* subgenus *Zimmermannia* we include six species as outgroup. In total we used 182 exemplars (174 ingroup and 8 outgroup) in the analyses. Part of this dataset has been analysed in a recent DNA barcoding study [37], which compared the utility of the mitochondrial cytochrome c oxidase subunit I (COI) barcoding gene [50] with the nuclear elongation factor 1 alpha (EF1-alpha) gene for species recognition. Recently obtained material was sequenced at these two loci. These data, in combination with morphology, life history and biogeographic data, were used to assign species boundaries to the undescribed material. The resulting selection contained 92, partly putative, species to be included in the phylogenetic study. Where possible, several representatives per species were included to be able to identify contamination during the molecular work and to detect intraspecific variability. Besides their utility for delimiting species, the COI barcode and EF1-alpha genes are also phylogenetically informative and were therefore included in the present phylogenetic dataset. To resolve the deeper parts of the phylogeny and to increase support in general, four more gene regions were also sequenced, viz. the D2 and D3 domains of the nuclear 28S ribosomal DNA gene, part of the nuclear isocitrate dehydrogenase (IDH) gene, the nuclear histone 3 gene and part of the mitochondrial cytochrome c oxidase II (COII) gene.

DNA extraction

DNA extractions were performed with a Qiagen DNeasy blood & tissue kit or with a Macherey Nagel magnetic bead tissue kit on an automated KingFisher flex system. Different types of tissue were used for extraction, depending on the abundance and significance of the specimens available. Hind legs were carefully removed from adults and cut into pieces with a scalpel. Non-destructive abdomen extractions were used so that genitalia preparations could be combined with DNA extraction [51]. Larvae were ground with a disposable pestle, or incisions were made in the cuticle to perform a non-destructive extraction of tissue material, after which the larval pelt was embedded in euparal on a microscopic slide.

PCR

Details on primers, including references, are listed in S2 Table. We sequenced the mtDNA cytochrome c oxidase I (COI) barcode region, which was 658 bp in length, sometimes in two parts if the material was degraded. Primer sets LepF1 and LCO1490 and LepR1 and HCO2198 attach to the same region and were often combined in a single PCR to obtain a higher DNA yield. A fragment of nuclear elongation factor 1-α (EF1-α) was 482 bp in length, and was sometimes also amplified in two parts. The D2 and D3 regions of the 28S ribosomal DNA gene had an aligned length of 873 bp. For the cytochrome oxidase II (COII) fragment, primer "eva" attaches to the tRNA-Lysine region at the 3' end of COII and this fragment was pruned for a stable aligned COII length of 638 bp; histone 3 sequences were 328 bp in length and isocitrate dehydrogenase (IDH) sequences 723 bp. A PCR cycle consisted of 3 minutes initial denaturation at 94°C, 15 seconds denaturation at 94°C, 30 seconds at the optimal annealing temperature, and 40 seconds extension at 57°C. A final extension at 72°C for 5 minutes completed the reactions. The

optimized annealing temperature for COI and COII was 50°C, for histone 3 and IDH 55°C, and for EF1-alpha and 28S rDNA 57°C. Universal tails were attached to the primers to facilitate higher throughput and to increase yield [52, 53]. PCR reactions were performed in volumes of 25 μl. Purification and bidirectional sequencing were outsourced to Macrogen Europe or Baseclear Leiden. Sequencher 4.2 (Gene Codes corporation) or Geneious R6 (www.geneious.com/) software was used to align the forward and reverse sequences, to manually check for ambiguities in the chromatograms and to create contigs. The COI, IDH, histone 3 and EF1-alpha alignments contained no gaps and were aligned by eye. COII contained a 3-bp insertion in *Ectoedemia quadrinotata* only. We used secondary structure to align the 28S rRNA sequences, despite the fact that within *Ectoedemia* there are only few ambiguous areas, making the alignment straightforward. Sequences were stored using the open-source package VoSeq 1.6.0 [54].

Phylogenetic analyses

Neighbor-joining trees were created using PAUP 4b10 [55] for each gene to assess congruence between markers. Alignments of different markers were then concatenated using VoSeq 1.6.0 and manually checked in Geneious R6, resulting in the preliminary dataset 'EctZimm2' with 187 exemplars representing 98 species. The most appropriate model for the phylogenetic analyses was determined with the online Findmodel service [56]. Single-gene as well as concatenated datasets were tested, and for all datasets the GTR+Gamma model was calculated to be most likely and used in subsequent analyses. Bayesian analyses were performed with MrBayes v3.2.1 [57, 58] on the Oslo Bioportal [59] and Cipres Portal [60]. Maximum likelihood analyses were done using the PhyML 2.2.0 plugin in Geneious R6 and RaxML HPC2 7.7.4 and Garli 1.0 on the Cipres portal. RaxML analyses were done both without partitions and with mtDNA, nuDNA and rDNA sequences separated in three partitions; these all resulted in highly similar tree topologies and branch support values. The resulting RaxML bootstrap trees from the initial analysis with all 187 exemplars were analysed with Roguenarok [61] to search for exemplars negatively influencing support values. The Roguenarok majority-rule consensus (MR) and extended majority rule consensus (MRE) algorithms flagged *Ectoedemia suberis* (RMNH.INS.12705), *E. "Namibia"* (RMNH.INS.23989), *E. aegilopidella* (RMNH.INS.23875) and both *E. picturata* specimens (RMNH.INS.23888 & RMNH.INS.23891), all exemplars that were missing sequence information for several markers. These taxa were subsequently removed in different steps, finally resulting in the dataset 'EctZimm5' with 182 exemplars representing 92 species. The analyses were then repeated on this culled dataset, after which there were no more exemplars flagged by Roguenarok. The increase in support values for key clades following the removal of 'rogue' examplars is shown in S3 Table. Garli maximum likelihood bootstrap (BS) and MrBayes Bayesian posterior probability (PP) support values were plotted on the respective branches of the PhyML best tree. We considered bootstrap values between 60 and 80 as moderate support, values above 80 as high support. For the posterior probabilities, we considered values of 0.85 to 0.95 as moderate support, and values above 0.95 as high support. We indicated higher clades with abbreviations of the species groups found within these


E. angulifasciella group
E. ornatella group
E. platanella group
E. suberis group
APOS
POS
bayesian E. angulifasciella group

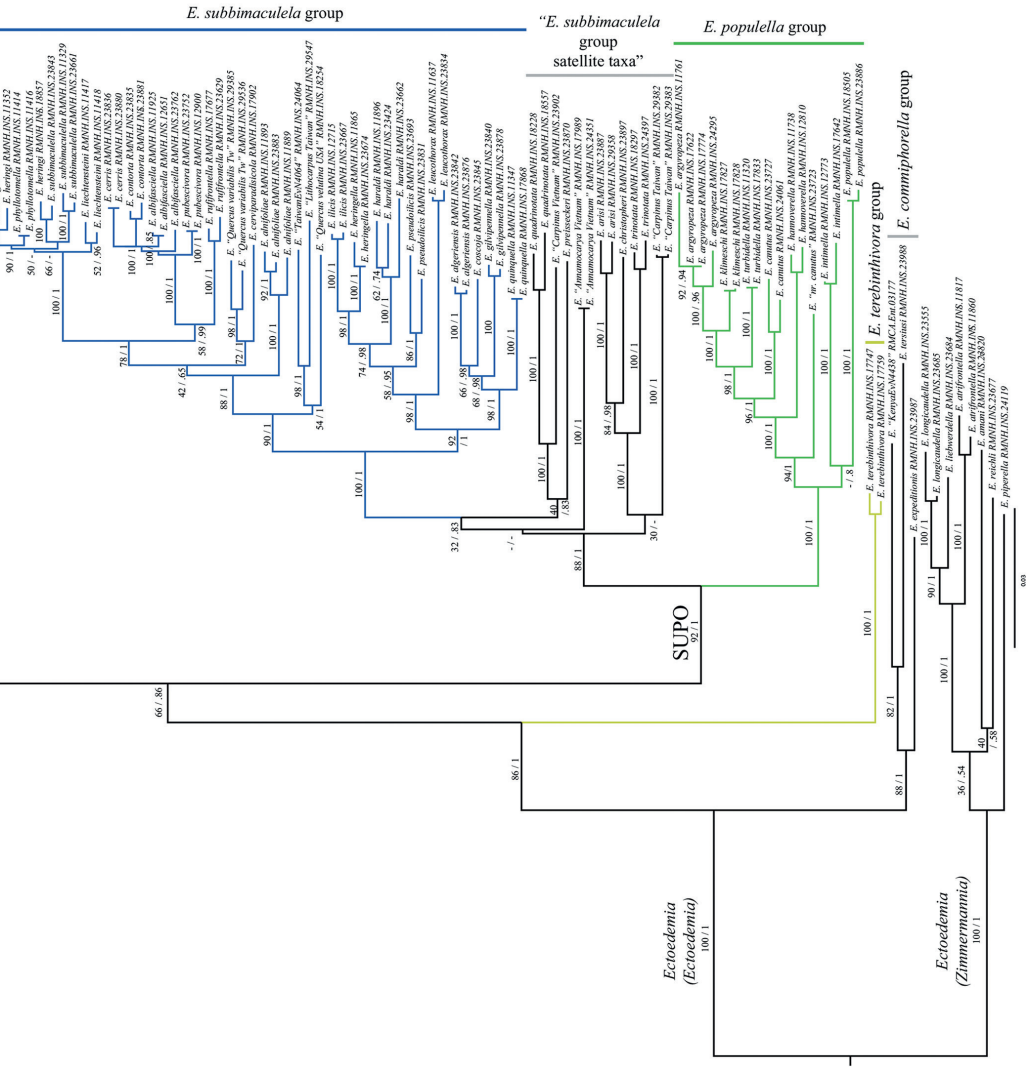


Fig. 1. Phylogeny of *Ectoedemia* s. str. PhyML best tree resulting from an analysis of 182 exemplars and six DNA markers, totalling 3692 bp. The outgroup contains six taxa of the subgenus *Ectoedemia* (*Zimmermannia*). In total 92 species of *Ectoedemia* s. str. are included. Species groups are colour coded and the support values of the Maximum Likelihood and Bayesian analyses are plotted on the branches. Because the topology of the *E. angulifasciella* group differed substantially, both the ML and Bayesian results are shown (see also S3 Table).

clades, for example SUPO for the combination of the *Ectoedemia subbimaculella* and *Ectoedemia populella* groups.

Character mapping
Mesquite version 2.75 (build 566) was used to map characters on the phylogenetic tree. The complete PhyML best tree with 182 exemplars was pruned so that a single terminal branch per species remained. Host plant family and geographic distribution were then plotted on the tree using the Trace Character History option under the Maximum Parsimony reconstruction with unordered states. This criterion calculates the fewest state changes required without assuming a mode of evolution or inferring information from the branch lengths. The results therefore often include equally likely states for many of the basal nodes, but indicate the strongest patterns in the data. The two resulting trees were arbitrarily ultrametricized and mirrored to allow for visual recognition of patterns between biogeography and host usage. Host plant associations at plant family level are mapped onto the phylogeny to identify distant changes in host use. The feeding mode, leafmining, is the same throughout the subgenus, with only one exception: the petiole galling *Ectoedemia populella*. Feeding mode is therefore not further analysed. Some species are specialized on a single host plant species, but most utilize several species within the same genus. A small number is oligophagous within a single plant family or disjunct oligophagous. In cases where more than one host plant species is used there often is a preference hierarchy of the ovipositing females that may vary throughout their distribution [62]. In this study we focus on host changes at the plant family level and evaluate how these correlate with shifts between major biogeographic regions. In cases in which a species can be found on multiple host families, only the dominant host has been indicated. To be able to identify shifts in distribution area, i.e. allopatry, throughout the phylogeny, the respective biogeographic region(s) of species has been mapped as well. We followed the general definitions of biogeographic regions [63], except that the Palearctic has further been divided into a West and East part with the "Turgai Strait", roughly following the 64–65 meridian as the divide [38], to allow for a more precise delineation of species' distributions. It should be noted though that this is still a rough estimate of species distributions, the actual distribution of most species only includes a smaller section of a biogeographic region and shifts within a single biogeographic region remain undetected.

Results

A global scale phylogeny of the subgenus Ectoedemia
The complete dataset contains 3692 bp from six genetic markers for 174 ingroup and eight outgroup exemplars, representing a largely complete global taxon sampling with 92 ingroup species. See S4 Table for Genbank accession numbers. Data from at least three DNA markers, encompassing a minimum of 1472 bp, was available for each taxon. The phylogeny based on the 3692 bp of the six markers combined is generally well resolved and supported (Fig. 1). The Maximum

Likelihood and Bayesian analyses provided similar topologies, allowing the support values to be plotted on the respective branches of the best Maximum Likelihood tree. The delimitation for species groups that we propose is indicated with coloured clades in the phylogeny (Fig. 1). The resulting new classification with all described and undescribed, but included in the phylogeny, *Ectoedemia* species is provided in S5 Table and is outlined as following:

Genus *Ectoedemia* Busck, 1907
Subgenus *Zimmermannia* Hering, 1940
Subgenus *Ectoedemia* s. s.
E. commiphorella group
E. terebinthivora group
SUPO—clade
    E. populella group
    E. subbimaculella group—satellite taxa
        [includes former *E. preisseckeri* group]
        E. subbimaculella group
APOS—clade
    POS-clade
        E. platanella group
        E. ornatella group
        E. suberis group
    E. angulifasciella group
        [includes former *E. rubifoliella* and *occultella* groups]

Only in the clade with the *angulifasciella* group there are substantial differences between the Bayesian and Maximum Likelihood analyses (Fig. 1). This is mainly due to the different position of *Ectoedemia "Prunus_Korea"*, which is sister to *E. "Pourthiaea_Taiwan"* in the Bayesian analysis but found basal to a clade with predominantly *Rubus*-feeding species in the Maximum Likelihood analysis, resulting in a different basal structuring of the *angulifasciella* group. The group as a whole is recovered as monophyletic in both phylogenetic analyses be it with medium statistical support.

Not all species could be assigned to a monophyletic species group. The "satellite taxa" basal to the *subbimaculella* group are all found in the same region of the molecular phylogeny, but cannot be placed in well-supported monophyletic species groups that are congruent with those based on morphology. Among these taxa are the former monotypic *E. preisseckeri* group, as well as the species *Ectoedemia christopheri* and *E. trinotata*, which have previously been regarded as part of the *populella* group on the basis of their genitalia [45, 46]. Besides the species groups, several other higher clades receive high support. Disregarding the basal section of the tree with the *E. commiphorella* and *E. terebinthivora* groups, there is a split into two large clades, named here as SUPO and APOS. The *subbimaculella* group with its basal satellite taxa and the *populella* group are placed together in the SUPO clade (BS = 92, PP = 1). The second clade, APOS, is recovered with high support

(BS 92, PP 1) and contains the*angulifasciella* group as well as a less well-supported clade named POS (BS 60, PP 0.86), containing the *platanella*, *ornatella* and *suberis* groups. Below we treat each species group in detail, the detailed new classification is presented in S5 Table.

The E. commiphorella *and* E. terebinthivora *groups*
The African taxa form the most basal split-off from the outgroup. Two of the three taxa included in the molecular phylogeny have been described, but they have previously not been attributed to any species group. Although the host plants of the African species that we sequenced are unknown, the hosts for two described closely related *Ectoedemia* are *Commiphora* species (Burseraceae) [64, 65], and it is not unlikely that other African species also feed on this widespread and diverse plant genus. We here erect the *Ectoedemia commiphorella* group for the five named (including *E. commiphorella*) and the two unnamed African *Ectoedemia* species, which are all morphologically similar to each other [48, 64, 65], and constitute a monophyletic group in the molecular phylogeny. The *E. terebinthivora* group contains only *E. terebinthivora*and represents a split-off between the *E. commiphorella* group and the rest of the ingroup. The species is found in the eastern Mediterranean area and is unique in that it feeds on Anacardiaceae (it is monophagous on *Pistacia terebinthus*). The unique position as a stand-alone species in the molecular analysis is concordant with its position in the morphological tree [40].

The E. subbimaculella *group*
Within the SUPO clade, the *Ectoedemia subbimaculella* group is recovered with high support. All species in this group feed on oaks and relatives (Fagaceae, primarily the genus *Quercus*). The group is well defined by morphological characters and is monophyletic in the molecular phylogeny with high support. Comparing the molecular results with published results on a subset of the European species that was previously assigned to this group based on morphology [40], displays remarkable similarities. In both phylogenies there is a clade around*E. quinquella* (best recognised by the characteristic forewing colour pattern and presence of a hair pencil at the base of the male hindwing), a southern European clade composed of species that feed on evergreen oaks (*E. leucothorax*, *E. pseudoilicis*, *E. haraldi*, *E. ilicis* and *E. heringella*) and a clade comprising the *subbimaculella* complex, *E. rufifrontella* and the *albifasciella* complex (sensu [40]). The latter clade is expanded in the molecular phylogeny and has at its basis *E. alnifoliae*, several Asian species and the single North American representative of the *subbimaculella* group, a single larva found on *Quercus velutina* in the Great Smoky Mountains (USA: North Carolina).

The E. populella *group*
Also part of the SUPO clade, the *populella* group is recovered with high support in all analyses. It consists of a variety of Palearctic and Nearctic species, including the type species of the genus, the Nearctic *Ectoedemia populella*. The group is well-defined morphologically [40], amongst others by the presence of two pairs of

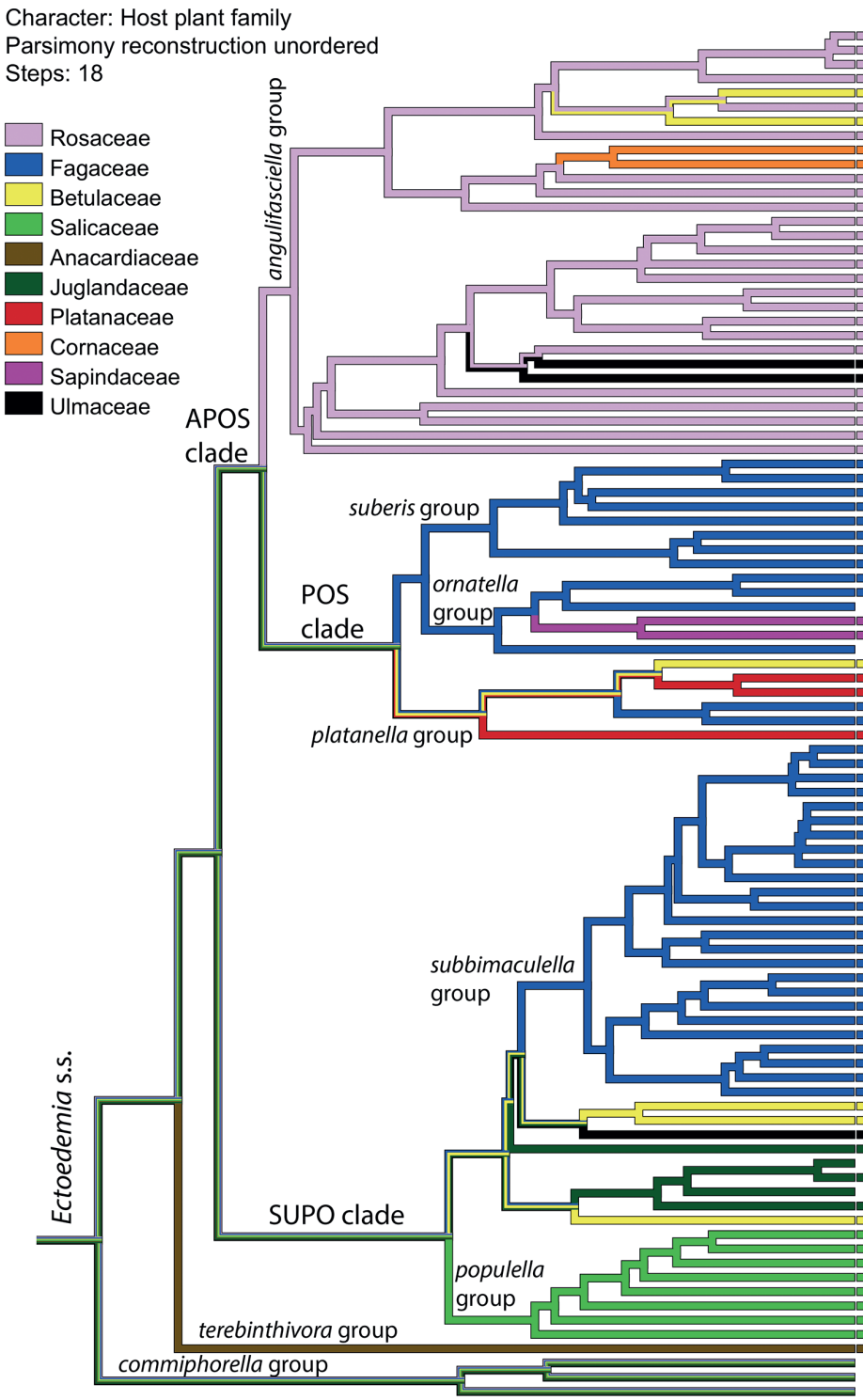
strongly developed carinae in the male genitalia. The group contains all Salicaceae-feeding *Ectoedemia*: *Ectoedemia intimella* feeds on *Salix* spp., whereas all other species feed on *Populus* spp. All first instars of these species start feeding in the petiole or midrib, and only later produce a leafmine, except *E. populella* which makes a gall in the petiole. The undescribed *E. "nr canutus"* also belongs to this group.

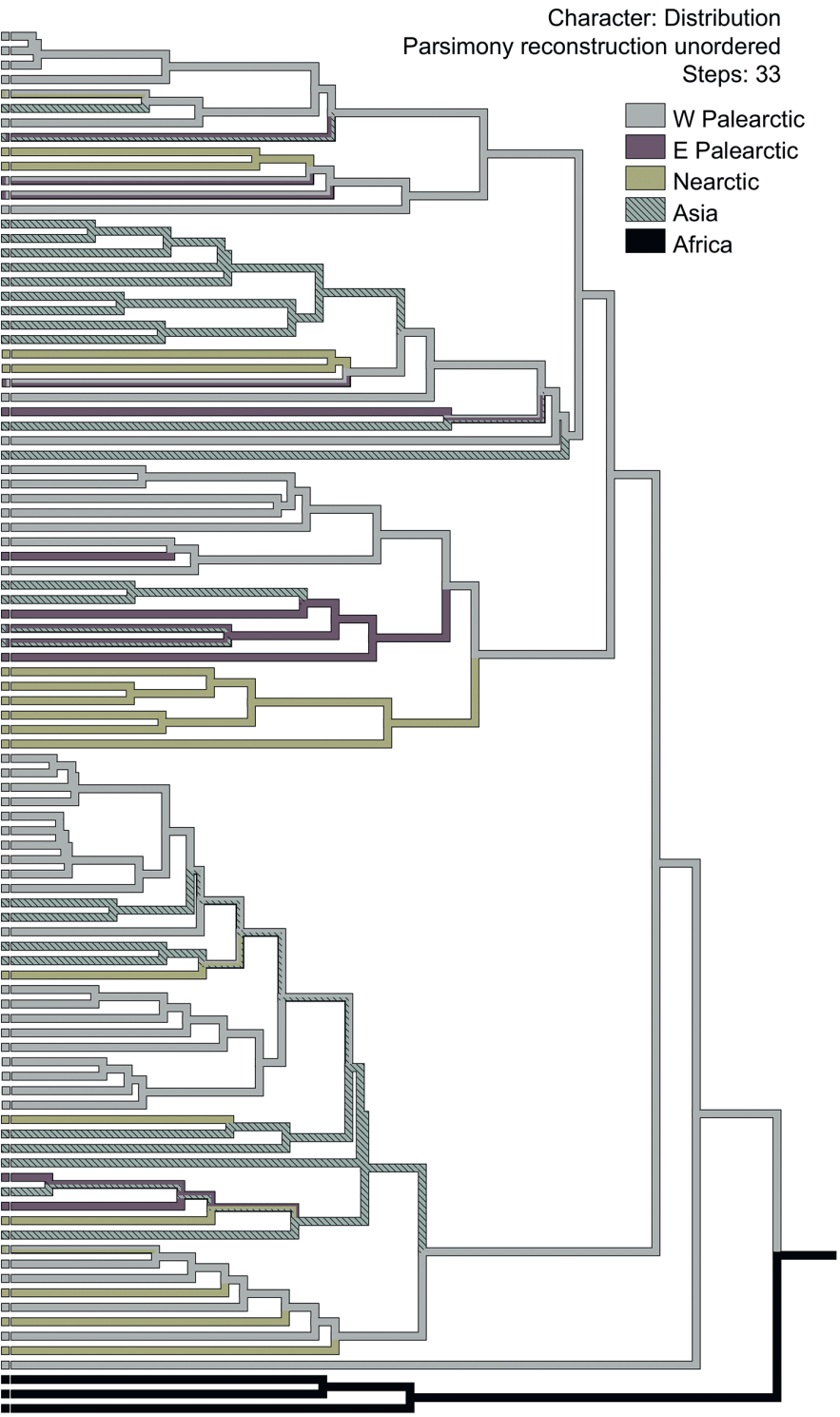
The E. angulifasciella *group*

The *angulifasciella* group is the largest species group of the APOS clade and is recovered with medium statistical support (BS 68, PP 0.86). When the group was erected based on morphology there was some uncertainty about its monophyly, as there were no clear common apomorphies to distinguish the group as a whole [40]. Nonetheless, the delimitation based on morphology and DNA is congruent and furthermore, it comprises all Rosaceae-feeding *Ectoedemia* species. Only a few species in this group feed on other plant families, viz. Betulaceae (the former *occultella* group), Ulmaceae (two species) and Cornaceae (two North American *Nyssa*-feeding species). The Maximum Likelihood and Bayesian analyses differed in the structuring of the *angulifasciella* group, due to a different position of the undescribed *E."Prunus_Korea"*; both alternatives are shown in Fig. 1. Disregarding the position of *E."Prunus_Korea"*, the *angulifasciella* group is divided into three larger subclades in both the Bayesian and Maximum Likelihood tree. The first group consists of the *E. rubivora* cryptic species complex + *E. angulifasciella* and *E. occultella*, *E. minimella* and E. "*Sorbus_Korea*", together with four East Palearctic species (PP 0.85). The second large clade (PP 1) contains almost all *Rubus*-feeding *Ectoedemia* and two *Ulmus*-feeding species, with the West Palearctic species pair *E. spinosella* and *E. mahalebella* as sister clade. The *Rubus*-feeding taxa from Southeast and East Asia show a remarkable molecular diversity: specimens that look morphologically almost the same seem to form a complex of sibling species, with at least several clearly differentiated taxa in Taiwan and Vietnam. The third clade (PP 1) groups the North American *Nyssa*-feeding clade with three West Palearctic Rosaceae feeders.

The POS clade: E. platanella, E. ornatella *and* E. suberis *groups*

The POS clade spans three species groups that are each found in a different biogeographic region. The *E. platanella* group is Nearctic, the *ornatella* group is East Palearctic/Oriental and the *suberis* group is West Palearctic. The *E. platanella* group is morphologically characterised by very similar male genitalia with multi-branched setae on the inner surface of the valves as a unique apomorphy [46]. It was initially only known from eastern North America. In the molecular phylogeny the group is recovered with high support, but support is much lower if the undescribed western North American species, collected as larva from *Platanus wrightii*, is included (RMNH.INS.29651). The Arizonan larvae of this host may belong to the same species that has been reared from *Platanus racemosa* in California; we have tentatively included it in the *platanella* group. Although specimens found on *Carya* were initially considered to be a different species, genitalia and DNA proved them to be conspecific with *Ectoedemia virgulae*. This


Character: Host plant family
Parsimony reconstruction unordered
Steps: 18
Rosaceae
Fagaceae
Betulaceae
Salicaceae
Anacardiaceae
Juglandaceae
Platanaceae
Cornaceae
Sapindaceae
Ulmaceae
angulifasciella group
APOS clade
suberis group
ornatella group
POS clade
platanella group
subbimaculella group
Ectoedemia s.s.
SUPO clade
populella group
terebinthivora group
commiphorella group


Character: Distribution
Parsimony reconstruction unordered
Steps: 33
W Palearctic
E Palearctic
Nearctic
Asia
Africa

◀ *Fig 2*. Ancestral state reconstruction of biogeography and host family. Ancestral state reconstruction of biogeography (right) and host plant family (left) in two mirrored trees with topology from the Bayesian analysis. Squares in the centre indicate for a taxon whether the information was known (square present) or reconstructed (square absent). Only the dominant family is indicated in case a species uses more than one host family. Multiple colours on a branch indicate ambiguous reconstructed states, except for the terminal branches where they indicate multiple states for that taxon.

means that *E. virgulae* is disjunct oligophagous, with larvae commonly found feeding on*Corylus* (Betulaceae) as well as *Carya* (Juglandaceae). In fact, this entire species group is characterised by distant host plant shifts: three species feed on Platanaceae, one on Fagaceae and one on Betulaceae and Juglandaceae. In contrast to the distant host shifts in the *platanella*group, all species in the *ornatella* and *suberis* groups feed on plant species within a single host genus (*Quercus* spp.) (see [37, 38], with as only exceptions *E. olvina* and its undescribed sister species *E. "Acer_Taiwan"* which both feed on *Acer* (Sapindaceae). The *ornatella* and *suberis*group delimitations in the molecular phylogeny are congruent with the delimitation based on morphology.

Host plant associations

The Maximum Parsimony reconstruction of biogeographic ancestral states (left tree) juxtaposed to those of host family (right tree) is shown in Fig. 2. Taxa with squares indicate that both host and biogeography are known, for those that do not have squares for the host, this information is reconstructed. Eighteen changes are required to explain the observed pattern of host use at family level. For all species groups except the *platanella* group, the ancestral host can unambiguously be assigned. The ancestral host of the *angulifasciella* group most likely fed on Rosaceae, and represents the only shift to Rosaceae in the subgenus *Ectoedemia*. Fagaceae are the most likely ancestral host for three species groups: the *suberis* group, the *ornatella*group and the *subbimaculella* group. Similar to Rosaceae, Salicaceae have only been colonised once, by the ancestor of the *populella* group. The ancestral host for the *platanella*group remains ambiguous and may have belonged to the Platanaceae, Fagaceae or Betulaceae. More basal in the tree, however, the ancestral host remains ambiguous. The ancestors of the APOS, POS and SUPO clades fed equally parsimoniously on Rosaceae, Fagaceae, Betulaceae, Salicaceae, Anacardiaceae, Juglandaceae and Platanaceae. The same seven plant families are the equally parsimonious ancestral hosts further up the tree towards the outgroup. The hosts of the outgroup taxa have purposely not been indicated to avoid bias in the analysis. The host plant association results corroborate earlier morphological studies that showed that most species groups are found on one host plant family only.

Biogeography

There are more changes in biogeography throughout the phylogeny than there are changes in host plant family. A minimum of 33 changes is required to explain the

observed pattern in biogeography, whereas only 18 host family changes are needed to explain the host plant family pattern. These shifts in biogeography are not evenly distributed over the tree, instead they are most common in the *Ectoedemia angulifasciella* group, where a biogeographic shift is involved for 43% of the speciation events. Biogeographic shifts are less common in the SUPO clade in which there is a shift for 31% of the speciation events and fewest in the POS clade with 25%. Just four species are known to occur in more than one biogeographic region: *E. argyropeza*, *E.intimella*, *E. occultella* and *E. spiraeae*. Of these, DNA barcoding and allozyme studies have shown that *E. argyropeza* in the Nearctic reflects a recent European introduction [37, 66]. However, a large pairwise intraspecific distance of 6.85% in DNA barcodes of *E. spiraeae* and a similarly large distance of 6.5% in DNA barcodes of *E. intimella* [37] suggests that these species have a wide distribution throughout the Palearctic, although we cannot entirely rule out the possibility that different species are involved, as long as specimens in the area between Europe and East Asia not have been genotyped. A 3.0% COI barcode pairwise distance between *E. occultella* specimens from the Nearctic and Europe suggests that also this species exhibits an old Holarctic distribution rather than a recent introduction into the Nearctic. Nonetheless, such vast distribution areas are exceptional, the great majority of species are restricted to a single biogeographic region. The biogeographic origin of the *angulifasciella*, *suberis*, *populella* and *terebinthivora* species groups is most likely the West Palearctic. The *subbimaculella* group ancestor could have lived equally parsimoniously in the West Palearctic or Asia. The *ornatella* group origin is reconstructed to be Asian and the *platanella* group of a Nearctic origin. A West Palearctic origin seems most likely for all higher clades, including the POS, the APOS, the SUPO and the SUPO + *terebinthivora* clade. Only the clade with the*commiphorella* group, with its African origin, causes the parsimony reconstruction to place the origin of the whole subgenus in Africa.

Discussion

Molecules and morphology

Our molecular phylogeny of 92 species represents the most complete phylogeny for *Ectoedemia* to date. Currently, 91 species are described, of which 65 are represented in this phylogeny, the remaining 27 species that we included are putative species. The six molecular markers used in this study were able to resolve most clades with good statistical support. Only the section of the tree with the "*subbimaculella* group satellite taxa" remains unresolved and requires more taxa and/or genes to get fully resolved. The phylogeny allowed us to evaluate published delimitations of species groups proposed on the basis of morphology [38, 40, 45, 47]. All monophyletic species groups that are proposed here already had been named previously, with the exception of the newly erected *E. commiphorella* group. However, most groups include more species than originally conceived, some of which were originally unplaced, and some were placed in other species groups that were not recovered by our phylogenetic analyses. Overall, the molecular results

corroborate the existing phylogenetic—morphology based—relations of the species groups.

The relations among species groups long remained unclear in morphological studies as a result of the many homoplasious characters [40]. We have now been able to resolve these relationships and we assign three overarching clades, named APOS, POS and SUPO, to indicate basal splits in the evolution of the subgenus. The APOS and SUPO clades each contain about half the total number of species. Two small clades are even more basal: the *commiphorella* group containing the African taxa and the monotypic *E. terebinthivora* group. The species-rich APOS clade contains the *angulifasciella*, *platanella*, *ornatella* and *suberis* groups, and within this clade the *angulifasciella* group splits off, leaving the POS clade with the *platanella*, *ornatella* and *suberis* species groups. The second large clade, SUPO, contains two species rich groups, the *subbimaculella* and *populella* species groups, as well as several unplaced taxa that fall between these two groups. The full new classification is provided in S5 Table.

The generally good resolution and support values allowed us to map two characters that are likely drivers of speciation, viz. host plant family shifts and biogeography, onto the obtained phylogeny. The parsimony reconstruction that we used is sensitive to changes in taxon sampling, because a single added taxon with a different host family may have a large effect on the amount of changes required for a large group of taxa. However, as taxon sampling is fairly complete, we can reliably use our data to explore the role of host plant family shifts and biogeography in the evolution of *Ectoedemia*.

Co-evolution?

Although co-evolution is not strictly tested in our analyses, it is evident that the shifts to different plant families never follow host phylogeny [41]. At most it can be said that the four plant families that are consumed by three quarters of the *Ectoedemia* s. str. species fall within the APG fabid (= rosid I) clade, but the remaining species use plants in the order Sapindales [malvid (= rosid II) clade], in the more distantly related order Proteales (basal eudicots) or in the order Cornales (basal in the asterid clade). This indicates that co-speciation or parallel cladogenesis at this phylogenetic level, similar to many groups of herbivorous insects [16], is highly unlikely. Instead, the pattern of changes in host plant family use that we revealed suggests repeated colonisation of different host plant families, congruent with a 'resource archipelago' scenario, and some of these colonisations have been followed by substantial diversification, which may be congruent with the 'intermediate resource hypothesis'.

Diversification following a host plant family shift

Species group boundaries, initially independently defined by morphological characters, commonly coincide with shifts to a different host plant family, indicating that the host plant family shift was accompanied by morphological change. Host plant family conservatism within species-rich clades like we find in *Ectoedemia* is

commonly found among insect herbivores [14, 67, 68], and fits the hypothesis that such a change involves a complex of simultaneous adaptions and is therefore phylogenetically constrained. However, once conceived, it may open a world of resources at intermediate distances. The four host plant families that include the largest radiations of *Ectoedemia*, host 78% of the *Ectoedemia* s. str. species that we included in our phylogeny: Fagaceae (40%), Rosaceae (24%), Salicaceae (8%) and Betulaceae (6%). Of these, Rosaceae and Salicaceae are reconstructed to have been colonised once, Fagaceae two or three times and Betulaceae three to five times. These four plant families all comprise a large diversity of, often ecologically dominant, plant species in the Holarctic that may constitute resources at intermediate distances from the initial host change towards the family. The plant families that do not host substantial *Ectoedemia* radiations are, at least in the Holarctic, relatively species-poor (viz. Anacardiaceae, Juglandaceae, Platanaceae, Cornaceae, Sapindaceae and Ulmaceae). Finally, it should also be noted that the main host families of *Ectoedemia* are also host to many other leaf-mining insects [3, 44], suggesting that these plant families in general have one or more common denominators that have made them suitable for colonisation and speciation. The balance between the odds for colonisation and the possibilities for subsequent diversification are likely not equal between the different host plant families. For example for Betulaceae, the abundance of species utilizing this host plant family can most easily be explained by several independent shifts to this family with little subsequent diversification. There are also indications from other groups that there is a large chance to colonise Betulaceae; a combination of Rosaceae and Betulaceae as hosts as seen in the *E. angulifasciella* group is known for various leafminer groups. Some share these two host families even within one species [e.g. *Lyonetia clerkella* (Lyonetiidae) and *Phyllonorycter corylifoliella* (Gracillariidae); [69, 70].

Allopatric speciation

The borders of the biogeographic regions erected by Wallace [63] that we use here are major barriers to gene flow for most animals. The regions are mostly defined by different climatic conditions in combination with geographic distance. In *Ectoedemia* we observe that the whole subgenus is mostly restricted to the biogeographic regions of the temperate northern hemisphere, and that the vast majority of species is restricted to a single biogeographic region. Restricted intraspecific gene flow over large geographic distances is further corroborated by the COI barcode distances [37], indicating genetic isolation between conspecific specimens from different regions. However, for many species the faunistic knowledge is currently too scarce to know their actual distribution. For those species where we do have rather complete information, the distribution rarely exactly coincides with the boundaries of a biogeographic region but includes only a smaller area. As a consequence, we cannot accurately estimate for the whole subgenus where species distributions may overlap or not and the 33 reconstructed shifts between biogeographic regions provide only a minimum estimate of the speciation events of which we are fairly certain that they involved a geographic component. Focussing on the larger patterns however, it can generally be said that within and between species groups, shifts between biogeographic regions are more common than shifts to another host plant family. To

further unravel the extent of the effect of geographic isolation as a driver of speciation, it will be crucial to examine the population genetic structure of species with wide distributions for isolation by distance patterns.

Different drivers for different clades

Although we find general patterns with regards to host family shifts and biogeography, none appear universally applicable for the whole subgenus. An exception to the general pattern that host plant family is a more conservative character within species groups than biogeographic occurrence is the POS clade, which includes the three smallest polytypic groups. The *E. platanella* group is exclusively Nearctic, the other two groups have representatives either throughout the Palearctic (*suberis* group) or Asia and the East Palearctic (*ornatella* group). Fagaceae host *Ectoedemia* in all biogeographic regions, but the *Ectoedemia* s. str. fauna is especially rich in southern Europe with many representatives from both the *subbimaculella* and *suberis* groups. The lack of a diverse *Ectoedemia* fauna feeding on Fagaceae in the Nearctic is surprising, considering the large radiation of Nearctic *Quercus* (e.g. [71]) as potential resources at intermediate distances. A very different dimension of speciation might be evident in the *E. populella* group. *Ectoedemia populella* is the only *Ectoedemia* s. str. with a clearly different feeding mode, petiole galling, and the biology of the entire group is somewhat divergent. The egg is laid on the petiole or midrib of the leaf where the mine starts and only later enters the leaf, instead of directly starting in the leaf [40]. The leafmining larvae of species in the *populella* group are able to withdraw in the petiole and this adaptation may have been a factor in the emergence and diversification of this species group, possibly strengthened through the combination with an initial adaptation to a different phytochemistry in allopatry. Clearly, neither shifts between biogeographic regions or host plant families alone can explain speciation in *Ectoedemia*. The two factors have possibly acted simultaneously, through differences in the availability of hosts in different parts of the world at different times, and their relative importance for speciation therefore varied for different clades.

Conclusions

1. Our molecular results comprise the most complete overview of the phylogenetics of the subgenus *Ectoedemia* and provide a solid base for the erection of eight monophyletic species groups: *E. commiphorella* group, *E. terebinthivora* group, *E. populella* group, *E. subbimaculella* group, *E. platanella* group, *E. ornatella* group, *E. suberis* group and *E. angulifasciella* group.
2. Our character state mapping analyses provide a coarse view of the role of geographic isolation and host plant shifts during speciation events in *Ectoedemia* and reveals some general patterns as well as exceptions to this pattern. Species groups are generally conserved in their host plant family, more so than their biogeographic regions, although in the POS clade this pattern is reversed.
3. Distant host shifts did not trace plant evolution, and diversification has not always followed from a distant host shift. Diversification has followed from shifts to

Rosaceae, Salicaceae and Fagaceae, but not from shifts to Betulaceae, Platanaceae, Anacardiaceae, Juglandaceae, Cornaceae, Sapindaceae and Ulmaceae.

4. Neither biogeography nor host plant family alone can explain the speciation patterns we find in *Ectoedemia* s. str., instead, a combination of these and other factors has likely been important, and has likely been differently for different clades.

Acknowledgements

The authors would like to thank their colleagues Dick S. J. Groenenberg, Frank R. Stokvis, Kees van den Berg, Renée Glas and Bastian Reijnen for their technical guidance and assistance during this project. We thank Charles Mitter and Kim Mitter for their comments on an earlier version. We kindly thank two anonymous reviewers and a third reviewer, Jim Whitfield, for their constructive comments on the manuscript. We are grateful for the material that we received from numerous people, whether on loan or as a gift: Val Albu (Fresno, USA), Giorgio Baldizzone (Asti, Italy), Bengt Å. Bengtsson (Färjestaden, Sweden), Kaj Berggren (Kristiansand, Norway), Jurate De Prins (Tervuren, Belgium), Rob Edmunds (Downham Market, United Kingdom), Charley Eiseman (MS, USA), Bob (R. J.) Heckford (Plymouth, United Kingdom), Jari Junnilainen (Vantaa, Finland), Lauri Kaila (Helsinki, Finland), Aleš Laštůvka (Prostejov, Czech Republic), Zdeňek Laštůvka (Brno, Czech Republic), Bong Woo Lee (Pocheon, Korea), Anna Mazurkiewicz (Kurów, Poland), Wolfram Mey (Berlin, Germany), Marko Mutanen (Oulu, Finland), Greg R. Pohl (Edmonton, Alberta), Andreas Stübner (Jänschwalde/Ost, Germany), Paolo Triberti (Verona, Italy), David L. Wagner (Storrs, CT, USA), Andreas Werno (Landsweiler-Reden, Germany), Hugo van der Wolf (Nuenen, Netherlands), Kazuo Yamazaki (Osaka, Japan), Shen Horn Yen (Kaohsiung, Taiwan) and Vadim Zolotuhin (Ulyanovsk, Russia). We thank the following persons for arranging collecting material abroad or providing permits: in France and Italy (EDIT ATBI project): Marta de Biaggi (Valdieri, Italy) and Marie-France Leggia (Nice, France); in the USA: Richard Brown and Sangmi Lee (Mississippi State, MS), Paul Super (NP Great Smoky Mountains, Lake Junaluska, NC) and David L. Wagner (Storrs, CT); in Vietnam: the director professor Le Xuan Canh and professor May Phu Quy, Dr. Khuat Dang Long and Dr. Dzuong Huyen (Institute of Ecology and Biological Resources, Hanoi), in China: the late professor Liu You Qiao (Zoological Institute, Academia Sinica, Beijing), in Indonesia, Kalimantan: Tropenbos International Indonesia (Balikpapan, Indonesia), in South Korea: Bong Woo Lee (Korean National Herbarium, Pocheon) and Bong-Kyu Byun (Hannam University, Daejeon) and in Taiwan Shen Horn Yen and students (National Sun Yat-Sen University, Kaohsiung).

References

1. Zhang Z-Q. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 2011;3148: 1–237. doi: 10.11646/zootaxa.3703.1.3
2. Nyman T, Linder HP, Peña C, Malm T, Wahlberg N. Climate-driven diversity dynamics in plants and plant-feeding insects. Ecol Lett. 2012;15(8): 889–98. doi: 10.1111/j.1461-0248.2012.01782.x. pmid:22507539
3. Menken SBJ, Boomsma JJ, van Nieukerken EJ. Large-scale evolutionary patterns of host plant associations in the Lepidoptera. Evolution (N Y). 2010;64(4): 1098–119. doi: 10.1111/j.1558-5646.2009.00889.x
4. Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. Proc Natl Acad Sci U S A. 2009;106(43): 18054–61. doi: 10.1073/pnas.0904106106. pmid:19815508
5. Winkler IS, Mitter C. The phylogenetic dimension of insect-plant interactions: a review of recent evidence. In: Tilmon KJ, editor. The evolutionary biology of herbivorous insects: specialization, speciation and radiation: Universirty of California Press; 2008. p. 240–63.
6. Heath TA, Hedtke SM, Hillis DM. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol. 2008;46(3): 239–57.
7. Ogden TH, Whiting MF, Wheeler WC. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer. Cladistics. 2005;21(3): 295–302. doi: 10.1111/j.1096-0031.2005.00061.x
8. Page RDM. Tangled Trees: phylogeny, cospeciation, and coevolution: The University of Chicago Press; 2003. 378 p.
9. Tilmon KJ. The evolutionary biology of herbivorous insects: specialization, speciation and radiation. Tilmon KJ, editor: University of California Press; 2008. 341 p.
10. McLeish MJ, Miller JT, Mound LA. Delayed colonisation of *Acacia* by thrips and the timing of host-conservatism and behavioural specialisation. BMC Evol Biol. 2013;13(188). doi: 10.1186/1471-2148-13-188. pmid:24010723
11. Vienne D, Refrégier G, López-Villavicencio M, Tellier A, Hood M, Giraud T. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 2013;198(2): 347–85. doi: 10.1111/nph.12150. pmid:23437795
12. Gomez-Zurita J, Hunt T, Kopliku F, Vogler AP. Recalibrated tree of Leaf Beetles (Chrysomelidae) indicates independent diversification of Angiosperms and their insect herbivores. PLoS ONE. 2007;2(4). doi: 10.1371/journal.pone.0000360
13. Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. The genetic basis of a plant-insect coevolutionary key innovation. Proc Natl Acad Sci U S A. 2007;104(51): 20427–31. pmid:18077380 doi: 10.1073/pnas.0706229104
14. Lopez-Vaamonde C, Wikstrom N, Labandeira C, Godfray HCJ, Goodman SJ, Cook JM. Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. Journal of evolutionary biology. 2006;(4): 1314–26. pmid:16780532 doi: 10.1111/j.1420-9101.2005.01070.x
15. Percy DM, Page RD, Cronk QC. Plant-insect interactions: double-dating associated insect and plant lineages reveals asynchronous radiations. Syst Biol. 2004;53(1): 120–7. pmid:14965907 doi: 10.1080/10635150490264996
16. Nyman T, Vikberg V, Smith DR, Boevé J-L. How common is ecological speciation in plant-feeding insects? A 'Higher' Nematinae perspective. BMC Evol Biol. 2010;10(266). doi: 10.1186/1471-2148-10-266. pmid:20807452
17. Nyman T. To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. Biol Rev Camb Philos Soc. 2010;85(2): 393–411. doi: 10.1111/j.1469-185X.2009.00109.x. pmid:20002390

18. Matsubayashi KW, Ohshima I, Nosil P. Ecological speciation in phytophagous insects. Entomol Exp Appl. 2010;134(1): 1–27. doi: 10.1111/j.1570-7458.2009.00916.x
19. Schluter D. Evidence for ecological speciation and its alternative. Science. 2009;323(5915): 737–41. doi: 10.1126/science.1160006. pmid:19197053
20. Schluter D. Ecology and the origin of species. Trends Ecol Evol. 2001;16(7): 372–80. pmid:11403870 doi: 10.1016/s0169-5347(01)02198-x
21. Mayr E. Ecological Factors in Speciation. Evolution. 1947;1(4): 263–88. doi: 10.2307/2405327
22. Mayr E. Animal species and evolution. Cambridge, MA: The Belknap Press of Harvard University Press; 1963. 797 p.
23. Becerra JX, Noge K, Venable DL. Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. Proc Natl Acad Sci U S A. 2009;106(43): 18062–6. doi: 10.1073/pnas.0904456106. pmid:19706441
24. Becerra JX. Synchronous coadaptation in an ancient case of herbivory. Proc Natl Acad Sci U S A. 2003;100(22): 12804–7. pmid:14555762 doi: 10.1073/pnas.2133013100
25. Ohshima I. Host race formation in the leaf-mining moth *Acrocercops transecta* (Lepidoptera: Gracillariidae). Biol J Linn Soc Lond. 2008;93(1): 135–45. doi: 10.1111/j.1095-8312.2007.00949.x
26. Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA. Functional genomics of cactus host shifts in *Drosophila mojavensis*. Mol Ecol. 2006;15(14): 4635–43. pmid:17107489 doi: 10.1111/j.1365-294x.2006.03102.x
27. Ohshima I, Watanabe K, Kawamura T. Distinct parasitoid communities associated with host races of the leaf-mining moth *Acrocercops transecta* on distantly related host plants (Juglandaceae and Ericaceae). J Nat Hist. 2014: 1–14. doi: 10.1080/00222933.2014.953613
28. Fordyce JA. Host shifts and evolutionary radiations of butterflies. Proc R Soc Edinb Biol. 2010;277(1701): 3735–43. doi: 10.1098/rspb.2010.0211. pmid:20610430
29. Wahlberg N. The phylogenetics and biochemistry of host-plant specialization in Melitaeine butterflies (Lepidoptera: Nymphalidae). Evolution (N Y). 2001;55(3): 522–37. doi: 10.1111/j.0014-3820.2001.tb00786.x
30. McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc Natl Acad Sci U S A. 2007;104(12): 4996–5001. pmid:17360391 doi: 10.1073/pnas.0608424104
31. Jeffries M, Lawton J. Enemy free space and the structure of ecological communities. Biol J Linn Soc Lond. 1984;23(4): 269–86. doi: 10.1111/j.1095-8312.1984.tb00145.x
32. Morales-Hojas R, Vieira J. Phylogenetic patterns of geographical and ecological diversification in the subgenus *Drosophila*. PLoS ONE. 2012;7(11). doi: 10.1371/journal.pone.0049552
33. Rota J, Wagner DL. Predator mimicry: metalmark moths mimic their jumping spider predators. PLoS ONE. 2006;1(1). doi: 10.1371/journal.pone.0000045
34. Wahlberg N, Wheat CW, Peña C. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS ONE. 2013;8(11). doi: 10.1371/journal.pone.0 080875
35. Suzuki H, Yasuda K, Ohashi K, Takahashi H, Fukaya M, Yano S, et al. Kanzawa spider mites acquire enemy-free space on a detrimental host plant, oleander. Entomol Exp Appl. 2011;138(3): 212–22. doi: 10.1111/j.1570-7458.2010.01092.x
36. Lopez-Vaamonde C, Charles H, Godfray J, Cook JM. Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. Evolution. 2003;57(8): 1804–21. pmid:14503622 doi: 10.1111/j.0014-3820.2003.tb00588.x
37. van Nieukerken EJ, Doorenweerd C, Stokvis FR, Groenenberg DSJ. DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-α: two are better than one in recognising cryptic species. Contrib Zool. 2012;81(1): 1–24.
38. van Nieukerken EJ, Laštůvka A, Laštůvka Z. Western Palaearctic Ectoedemia (Zimmermannia) Hering and Ectoedemia Busck s. str. (Lepidoptera, Nepticulidae): five new species and new data on distribution, hostplants and recognition. Zookeys. 2010;32: 1–82. doi: 10.3897/zookeys.32.282
39. Diškus A, Puplesis R. Catalogue of the world Nepticuloidea & Tischerioidea. In: Puplesis R, Diškus A, editors. Nepticuloidea ir Tischerioidea (Lepidoptera) pasaulio ir lietuvos faunoje The Nepticuloidea & Tischerioidea (Lepidoptera)—a global review, with strategic regional revisions. Kaunas: Lututе publishers; 2003. p. 318–436.

40. van Nieukerken EJ. A taxonomic revision of the Western Palaearctic species of the subgenerea *Zimmermannia* Hering and *Ectoedemia* Busck s. str. (Lepidoptera, Nepticulidae), with notes on their phylogeny. Tijdschr Entomol. 1985;128(1): 1–164. doi: 10.3897/zookeys.32.282.app.1.ds
41. APG III. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc. 2009;161(2): 105–21. doi: 10.1111/j.1095-8339.2009.00996.x
42. Lewinsohn TM, Novotny V, Basset Y. Insects on plants: diversity of herbivor assemblages revisited. Annu Rev Ecol Evol Syst. 2005;36: 597–620. doi: 10.1146/annurev.ecolsys.36.091704.175520
43. De Prins J, Kawahara AY. Systematics, revisionary taxonomy, and biodiversity of Afrotropical Lithocolletinae (Lepidoptera: Gracillariidae). Zootaxa. 2012; 2012–12–20; 3594: 1–283.
44. Nyman T, Farrell B-D, Zinovjev A-G, Vikberg V. Larval habits, host-plant associations, and speciation in nematine sawflies (Hymenoptera: Tenthredinidae). Evolution (N Y). 2006;60(8): 1622–37. pmid:17017063 doi: 10.1554/05-674.1
45. Puplesis R. The Nepticulidae of eastern Europe and Asia. Western, central and eastern parts. Leiden: Backhuys Publishers; 1994. 290 p.
46. Wilkinson C, Newton PJ. The microlepidopteran genus *Ectoedemia* Busck (Nepticulidae) in North America. Tijdschr Entomol. 1981;124(2&3): 27–92.
47. Wilkinson C. A supplement to the genus *Ectoedemia* Busck (Nepticulidae: Lepidoptera) in North America, dealing with some difficult species and also some new ones. Tijdschr Entomol. 1981;124(3): 93–110.
48. Mey W. Nepticulidae (Lepidoptera, Nepticuloidea).—In: W. Mey (Ed.), The Lepidoptera of the Brandberg Massif in Namibia Part 1. Esperiana Memoir. 2004; 2004–12–30;1: 27–38. doi: 10.4001/003.017.0116
49. Puplesis R, Robinson GS. A review of the Central and South American Nepticulidae (Lepidoptera) with special reference to Belize. Bulletin of the Natural History Museum London Entomology. 2000;69(1): 3–114.
50. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270(1512): 313–21. pmid:12614582 doi: 10.1098/rspb.2002.2218
51. Knölke S, Erlacher S, Hausmann A, Miller MA, Segerer AH. A procedure for combined genitalia dissection and DNA extraction in Lepidoptera. Insect Syst Evol. 2005;35(4): 401–9. doi: 10.1163/187631204788912463
52. Wahlberg N, Wheat CW. Genomic outposts serve the phylogenomic pioneers: Designing novel nuclear markers for genomic DNA extractions of lepidoptera. Syst Biol. 2008;57(2): 231–42. doi: 10.1080/10635150802033006. pmid:18398768
53. Regier JC, Shi D. Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5 ‘ tails. Biotechniques. 2005;38(1): 34–48. pmid:15679081 doi: 10.2144/05381bm02
54. Peña C, Malm T. VoSeq: A Voucher and DNA Sequence Web Application. PLoS ONE. 2012;7(6). doi: 10.1371/journal.pone.0039071
55. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. 4 ed. Sunderland Massachusetts: Sinauer Associates; 2003.
56. Posada D, Crandall K. Findmodel. 2012; Available:http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html. Accessed June 2013.
57. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3): 539–42. doi: 10.1093/sysbio/sys029. pmid:22357727
58. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;18: 1572–4. doi: 10.1093/bioinformatics/btg180
59. Kumar S, Skjaeveland A, Russell JSO, Enger P, Ruden T, Mevik B-H, et al. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinformatics. 2009;10:357. doi: 10.1186/1471-2105-10-357. pmid:19863793

60. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE); 2010; 14–11–2010; New Orleans, LA.
61. Aberer AJ, Krompass D, Stamatakis A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Syst Biol. 2013;62(1): 162–6. doi: 10.1093/sysbio/sys078. pmid:22962004
62. Thompson JN. Coevolutionary Dynamics. In: Thompson JN, editor. Relentless Evolution. Chicago: The University of Chicago Press; 2013. p. 211–33.
63. Wallace AR. The geographical distribution of animals with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface: By Alfred Russel Wallace,…With maps and illustrations. In Two Volumes: Macmillan and Company; 1876.
64. Scoble MJ. A new species of *Ectoedemia* Busck from Botswana with observations on its imaginal skeletal anatomy (Lepidoptera: Nepticulidae). Ann Transvaal Mus. 1979;32(2): 35–54.
65. Scoble MJ. Nepticulidae (Lepidoptera) of southern Africa: the genus *Ectoedemia*Busck. J Entomol Soc South Afr. 1978;41(1): 81–6.
66. Menken SBJ, Wiebosch-Steeman M. Clonal diversity, population-structure, and dispersal in the parthenogenetic moth *Ectoedemia argyropeza*. Entomol Exp Appl. 1988 Nov;49(1–2): 141–52. doi: 10.1111/j.1570-7458.1988.tb02485.x
67. Farrell BD, Sequeira AS. Evolutionary rates in the adaptive radiation of beetles on plants. Evolution (N Y). 2004;58(9): 1984–2001. doi: 10.1554/03-453
68. Leppänen SA, Altenhofer E, Liston AD, Nyman T. Phylogenetics and evolution of host-plant use in leaf-mining sawflies (Hymenoptera: Tenthredinidae: Heterarthrinae). Mol Phylogenet Evol. 2012;64(2): 331–41. doi: 10.1016/j.ympev.2012.04.005. pmid:22531610
69. De Prins J, De Prins W. Global Taxonomic Database of Gracillariidae (Lepidoptera). Belgian Biodiversity Platform; 2011; Available: http://www.gracillariidae.net/. Accessed June 2014.
70. Emmet AM. A field guide to the smaller British Lepidoptera. Second edition, revised and enlarged ed. London: The British Entomological & Naturalis History Society; 1988.
71. Manos PS, Stanford AM. The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int J Plant Sci. 2001;162: S77–S93. doi: 10.1086/323280

Supplemental information

S1 Table. Collecting, identification and voucher information of all the material used in the study.
doi:10.1371/journal.pone.0119586.s001
(XLS)

S2 Table. Primer names, forward and reverse primer sequences and references.
doi:10.1371/journal.pone.0119586.s002
(DOCX)

S3 Table. Support increases for different datasets during Garli runs.
doi:10.1371/journal.pone.0119586.s003
(DOCX)

S4 Table. Identifiers, taxonomy and Genbank accession numbers of the material used in this study.
doi:10.1371/journal.pone.0119586.s004
(DOCX)

S5 Table. New Classification of *Ectoedemia* (*Zimmermannia*) and *Ectoedemia* s. str., including unnamed species.
doi:10.1371/journal.pone.0119586.s005
(XLS)

4

Checklist of fossil Nepticulidae

A revised checklist of Nepticulidae fossils (Lepidoptera) indicates an Early Cretaceous origin

Camiel Doorenweerd[1,2,6], Erik J. Van Nieukerken[1], Jae-Cheon Sohn[3,4] & Conrad C. Labandeira[3-5]

[1] *Department of Terrestrial Zoology, Naturalis Biodiversity Center, Leiden, The Netherlands*
[2] *Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands*
[3] *Department of Paleobiology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA.*
[4] *Department of Entomology, University of Maryland, College Park, USA*
[5] *College of Life Sciences, Capital Normal University, Beijing 100048, China*
[6] *Corresponding author. E-mail: camiel.doorenweerd@naturalis.nl*

Key words: Baltic Amber, Calibration points, Dakota Formation, Evolutionary history, Extinction, Fossil record, Larvae, Leaf mining, Plant hosts, *Stigmella*, *Stigmellites*

Abstract

With phylogenetic knowledge of Lepidoptera rapidly increasing, catalysed by increasingly powerful molecular techniques, the demand for fossil calibration points to estimate an evolutionary timeframe for the order is becoming an increasingly pressing issue. The family Nepticulidae is a species rich, basal branch within the phylogeny of the Lepidoptera, characterized by larval leaf-mining habits, and thereby represents a potentially important lineage whose evolutionary history can be established more thoroughly with the potential use of fossil calibration points. Using our experience with extant global Nepticulidae, we discuss a list of characters that may be used to assign fossil leaf mines to Nepticulidae, and suggest useful methods for classifying relevant fossil material. We present a checklist of 79 records of Nepticulidae representing adult and leaf-mine fossils mentioned in literature, often with multiple exemplars constituting a single record. We provide our interpretation of these fossils. Two species now are included in the collective generic name *Stigmellites*: *Stigmellites resupinata* (Krassilov, 2008) comb. nov. (from *Ophiheliconoma*) and *Stigmellites almeidae* (Martins-Neto, 1989) comb. nov. (from *Nepticula*). Eleven records are for the first time attributed to Nepticulidae. After discarding several dubious records, including one possibly placing the family at a latest Jurassic position, we conclude that the oldest fossils likely attributable to Nepticulidae are several exemplars representing a variety of species from the Dakota Formation (USA). The relevant strata containing these earliest fossils are now dated at 102 Ma (million years ago) in age, corresponding to the latest Albian Stage of the Early Cretaceous. Integration of all records in the checklist shows that a continuous presence of nepticulid-like leaf mines preserved as compression–impression fossils and by amber entombment of adults have a fossil record extending to the latest Early Cretaceous.

Introduction

Numerous molecular phylogenetic studies spanning the entire megadiverse insect order Lepidoptera have been published during the past five years (Mutanen *et al*. 2010; Regier *et al*. 2009; Regier *et al*. 2013; Timmermans *et al*. 2014). Although there is as yet no complete consensus for all phylogenetic relationships, especially among superfamilies (Timmermans *et al*. 2014), the overall topology for the evolution of Lepidoptera presently is clearer than ever. This advancement offers opportunities to study the timeframe during which their evolution took place. One recent study used seven fossil calibration points across all Lepidoptera and has revealed several periods of increased diversification and a plausible, synchronous evolution with angiosperm hosts (Wahlberg *et al*. 2013). Their work became possible by combining the Mutanen *et al*. (2010) phylogenetic dataset with a LepTree project compilation of fossils (Sohn *et al*. 2012; Sohn & Lamas 2013). Molecular dating on phylogenetic trees has been a subject of considerable scientific debate, ranging from pointing out the pitfalls of using poorly supported phylogenetic trees as starting points, to the sensitivity of different Bayesian priors such as mutation rates, the amount of data partitions, or the effects of modelling calibration priors (Wheat & Wahlberg 2013). When such technical issues are taken into account, the remaining, most important factor in constructing a timed phylogenetic tree is the number and reliability of calibration points, including the issue of establishing additional dates (Magallón *et al*. 2013). The reliability of these dates are assured only when such age dates can be assigned to a particular phylogenetic node with a high degree of certainty. In summary, calibration points require reliable identifications.

Nepticulidae commonly are known as pygmy moths and constitute a species rich, basal family within lepidopteran phylogeny, and consequently may offer a series of calibration points of both practical and theoretical importance (see Regier *et al*. 2015). The family comprises some of the smallest adult Lepidoptera known, and is found on all continents except Antarctica. All nepticulid species are herbivores with larvae that feed inside hostplant tissues. There are a variety of larval feeding modes within the family, including gall-formers, fruit-miners, bud-miners, stem-miners and bark-miners, but the vast majority of species are leaf miners. The identification of extant species from larval leaf-mining traces often is reliable, although somewhat dependent on geographic region and highly contingent on correct host identification. Many species are (strict) monophages or oligophages (sensu Menken & Roessingh 1998), and some genera or species groups are specialized on a single plant family. Several studies provide a phylogeny of the family or its subgroups (Scoble 1983, van Nieukerken 1986, Puplesis 1994, Hoare & van Nieukerken 2013, Doorenweerd *et al*. 2015).

Fossils of adult Lepidoptera are rare, but recently fossils of leaf mines frequently have been encountered (Labandeira *et al*. 1994; Sohn *et al*. 2012; Donovan *et al*. 2014). However, identifying leaf-mining taxa from traces, or larval mediated damage on fossilized leaves requires an alternative approach, when compared to

the identification of extant species. The difference in approaches is that there is no independent knowledge of the biogeographical distribution or plant-host specificities of the fossil taxa. In addition, the characters that leaf mines provide are largely behavioural, and because they are influenced by environmental conditions, they also are prone to homoplasy.

It is relatively easy to distinguish a fossil leaf mine from other types of biotic and abiotic foliar damage by focusing on those features that also are relevant for identification. The presence of wound reaction tissue that surrounds oviposition sites is indicative of miner insertion of eggs into inner leaf tissues. The colour and differential contrast in hues among the surrounding leaf tissue, the mine trajectory and mine's contents also provide additional information. The mined areas are thinner and thus paler and of lighter hue than surrounding unaffected tissue, whereas a frass trail consists of concentrated faecal contents and will be darker in hue. Distinguishing between miners feeding on parenchyma and sap-feeding miners consuming epidermal tissues also is important. Sap feeding occurs among extant groups only in several early instar larvae of Gracillariidae and some Agromyzidae (Diptera) (Winkler *et al*. 2010). Only larvae of Phyllocnistinae and Oecophyllembiinae (Gracillariidae) are sap feeding throughout their entire larval feeding period (Davis & Robinson 1998; Hering 1951; Kumata 1998). Parenchyma feeding leaf miners are more common and are found in four insect orders: Diptera, Hymenoptera, Coleoptera and Lepidoptera. Leaf mining is overwhelmingly the most common type of plant mining damage encountered in the fossil record, and, of the different mining types, it provides the most characters for taxonomic identification. A focus on those combinations of characters of taxonomic relevance could result in the identification of a fossil leaf-mine specimen at least to the family level.

Body fossils of adult Nepticulidae are exceedingly scarce, but constitute the most valuable candidates for calibrating genus-level nodes in phylogenies. If sufficient characters are visible, identifications to genus or even species groups are possible. There are 13 adult fossils that have been assigned to Nepticulidae in the literature, some of which have tentative affiliations. Two of these occurrences are compression–impression fossils from the Late Priabonian (Late Eocene) and provide partly visible wing venation. Two others are found in resin-like copal and likely are not older than 150,000 years (Labandeira *et al*. 1994). Seven adult fossils originate from Baltic Amber. Another candidate specimen originates from mid Late Cretaceous Canadian Amber, and an additional specimen lacks a clear stratigraphic provenance. Adult fossils from Baltic Amber are dated from 44.4 to 33.9 Ma, and their occurrences reflect multiple phases of sedimentary recycling of original older amber into successively more recent deposits (Labandeira 2014). The Canadian Amber specimen is estimated to be 72 Ma.

Earlier reviews of Lepidoptera fossils that list Nepticulidae include Skalski (1990a) and Sohn *et al*. (2012), but here we present the most comprehensive and revised checklist to date. Five amber fossils have been described since the latest review in

2012, more than doubling the number of known nepticulid amber fossils (Fischer 2013). Nepticulidae form a Superfamily, Nepticuloidea, together with the Opostegidae, of which the latter lack known fossils. Opostegidae also are herbivores, but, to the extent that larval habits are known, the majority of extant species create stem- or bark-mines, often consuming cambium tissue; very few species construct mines in leaves (Regier *et al*. 2015). Larval traces of mines in bark or cambium are difficult to find and recognize, even in modern live hosts, and the absence of their traces in the fossil record is not surprising. Given the absence of opostegid fossils, the checklist presented here can also be viewed as a checklist for the Superfamily Nepticuloidea. The list is constructed in such a way that it maximizes the potential in calibrating nodes for molecular dating analyses of Nepticulidae, related lineages, and Lepidoptera at large.

Material and methods

Identifying fossil adult Nepticulidae

Identification of fossil adult Nepticulidae relies on the same characters that distinguish extant species (e.g. Johansson *et al*. 1990, van Nieukerken 1986), although commonly only a subset of those characters are preserved or evident in the fossil record. The family as a whole may be tentatively recognized by a combination of their small size; the presence of an enlarged first antennal segment (scape), also known as the eye-cap; erect hair-like scales on the frons and vertex of the head (the frontal tuft); the presence of maxillary and labial palps; usually a short haustellum; and relatively short legs, without a tibial epiphysis. An additional combination of external characters, particularly wing venational features, may lead to a genus level identification. As with many groups of insects, genitalia are the most reliable source of characters for species-level identification. Most amber inclusions of adults, however, are internally hollow, a phenomenon that results from the degradation of internal organs (Labandeira 2014), and genitalia may be absent or partially preserved.

Identifying Nepticulidae from fossil leaf mines

Identification of fossil leaf mines is less exact than identification of adults, mostly because the characters of leaf mines have never been analysed in a phylogenetic context. For many taxa there is a combination of characters that distinguishes the group, but for each individual character there usually are exceptions that obstruct the designation of truly synapomorphic characters. Moreover, many characters are difficult to describe in a quantitative manner, such as the shape of the mine or distribution of faecal pellets (microcoprolites) within the mine's frass trail. In practice, the identifications depend on cumulative evidence. Below, we describe the relevant characters that we have evaluated.

Oviposition

Nepticulidae oviposit on the exposed surfaces of plant tissue, and consequently, there is no scarring around the oviposition site. Female Nepticulidae deposit an

egg-case that covers the entire egg. The egg-case materials consist of a secretion from the collaterial glands resulting in a shiny speck, which typically is black when the larva has hatched (van Nieukerken *et al*. 1990, figures on page 31). However, this structure is lost relatively easily with post-mortem plant-tissue decay, and may be more difficult to recover or recognize in fossils. Notably, an egg-case may also be observed in mines made by the genus *Leucoptera* (Lepidoptera: Lyonetiidae), or alternatively coleopteran leaf-mining groups, such as the genus *Trachys* (Buprestidae) (Ellis 2014; Emmet 1988; Ding *et al*., 2014).

Leaf-mine shape

The shape of the mine is the feature that is of most immediate concern for the description and classification of leaf-mine records. Leaf-mine shape usually constitutes the most conspicuous set of characters. Nepticulidae leaf mines are highly variable, and include linear galleries, blotches or a combination of a gallery and a blotch (see text box "Leaf-mine terminology" and Figs 1–9). Much of this structural variation is present in the most species rich, extant genus, *Stigmella* (Figs 2, 4, 9). Therefore, it is not surprising that most fossil records mention a resemblance to species of extant *Stigmella*. Nevertheless, several other extant genera contain very similar mine types, particularly as the linear mine very commonly occurs in the genera *Acalyptris* (Fig. 1), *Enteucha, Roscidotoga* (Fig. 6), *Pectinivalva* and *Parafomoria*. The second largest genus in the family, *Ectoedemia*, has leaf mines that typically, but not always, start as a thin, usually strongly meandering, gallery mine that abruptly change into a broad blotch, termed an ophistigmatonome (Figs 3, 8). Several fossil mines display an analogy to this mine type, such as *Stigmellites samsonovi* Kozlov, 1988 or *Ectoedemia* sp. (Labandeira *et al*. 1994) (Fig. 18), but such a sequence of mine phases is not an apomorphic trait within the family, and occurs also in other genera (for example, Figs 7, 9 with *Bohemannia* and *Stigmella*, respectively). Nepticulidae larvae generally avoid veins, as they rarely cross primary or otherwise prominent veins, such as midribs of dicotyledonous angiosperms; only later instar larvae have the ability to cross secondary veins. Another relevant feature is the total length of the mine, which from oviposition site to emergence area can range from very short in small mines, likely attributable to early instar Bucculatricidae or Coleoptera that later feed externally, to very long linear mines or large blister-like mines that are more likely made by Lyonetiidae, Gracillariidae, Eriocraniidae, Hymenoptera or some Coleoptera. Care should be taken to only judge the length of completed mines, which may be recognized by the lack of frass at the emergence area. Incomplete mines may contain perished, possibly parasitized, larval remains, particularly mandibles and head capsules. Backtracking, or reversing the larval trajectory of a mined route, combined with initiating a new mining direction, is very rare in Nepticulidae. However, backtracking without starting a new trajectory occasionally occurs, evident from frass trails occurring on both sides of a mine, resulting in a central, frass-free path. Such a condition is seen, for example, in the *Ectoedemia populella* group where larvae retreat into the petiole or midrib during the day. Leaf-mining shapes with backtracking and changes of direction are more typical for *Parectopa* (Gracillariidae) or *Cosmopterix* (Cosmopterigidae), but also occur commonly in many dipteran (Winkler *et al*. 2010) and coleopteran (Ding *et al*. 2014) leaf mines.

Frass

The frass of Nepticulidae consists of granular pellets that are deposited often in a species-specific mode, but do not constitute a distinct, synapomorphic pattern that follows from the movements of the larva in the mine. The range of frass patterns include a central or laterally positioned, thin frass line (Fig. 5), randomly distributed pellets filling the width of the mine (Fig. 6); frass deposited in meniscate arcs (Fig. 2); or distinctive, abrupt changes in the frass pattern following each moult. Some species spread the frass in the final instar along two lateral lines and move in between these trails, examples of which are the *Ectoedemia populella* group (Johansson *et a*l. 1990, Ellis 2014: *e.g.* http://www.bladmineerders.nl/minersf/lepidopteramin/ectoedemia/intimella/intimella.htm), or the oak miners *Stigmella kao* van Nieukerken & Liu and *S. lithocarpella* van Nieukerken & Liu (van Nieukerken and Liu 2000, Figs 95, 96, or http://nepticuloidea.info/stigmella-kao-13; http://nepticuloidea.info/stigmella-lithocarpella-7). In Diptera, frass is generally more fluidized, consists of fewer pellets, and often is deposited in a double-track manner, owing to the larva residing laterally, on its side, within the mine, or occasionally is not visible (Winkler *et al*. 2010). Hymenopteran frass is often arranged into threads or elongate pellets and occasionally is actively removed by the larva from the mine. Coleopteran frass usually comprises granular pellets, or strings of pellets, or elongate pellets (Ding *et al*. 2014). Leaf mines on non-woody herbaceous plants are more commonly made by dipteran leaf miners than any other order of leaf miners (Spencer 1990), although when Lepidoptera, including Nepticulidae, do mine herbaceous plants, they tend to have a more fluidised, dipteran-like frass as well (Fig. 8). The presence, shape, and depositional pattern of the frass within the mine are important characters. Occasionally, the pattern may change abruptly after moulting, and careful examination of the mine may reveal exuviae of earlier moults, including head capsules.

Larva

None of the published fossil records include fossilized larvae. If larvae are present, however, they would provide an important source of characters. Features of the chitinous head capsule frequently are diagnostic to order, usually to family and possibly even to genus, when clearly visible (for Nepticulidae, see Gustafsson & van Nieukerken 1990). Other characters of the larva, such as the presence or reduction of legs and prolegs and the constriction between segments also are informative. Nepticulidae never have prolegs or thoracic legs, and the body is minimally constricted between adjacent segments.

Pupation

Numerous publications claiming to present fossil nepticulid mines mention a "pupation chamber" at the final section or terminus of the mine (e.g. Stephenson 1991), or indicate that the presence of a semi-circular slit at the end of the mine is a reliable character for distinguishing dipteran from lepidopteran mines. With the exception of some species of *Ectoedemia* and *Trifurcula*, all Nepticulidae pupate outside the mine. They create a semi-circular slit to exit the mine (clearly visible in Fig. 7), similar to many Diptera, and usually descend to the soil on a silken thread where they pupate within a silken cocoon. A related feature is the final section of

the mine, which is devoid of frass, as the mine terminus is where the larva resides before vacating the mine. The absence of frass at the mine terminus may be used to indicate the final size of the larva, but should not be interpreted as a pupation chamber. In such mines there are never traces of silk, but some species that do pupate in the mine may construct a cocoon within a silken tunnel that is connected to a previously made slit in the epidermis. Larvae that make such structures include species in the subgenus *Ectoedemia* (*Fomoria*) (Johansson *et al*. 1990) (Fig. 8).

Plant hosts

Identification of the plant host is a crucial step when identifying extant species from leaf mines. However, plant hosts from the fossil record should be viewed in a different perspective. It is likely that many fossil leaf-mining species are extinct, and modern taxa may have recolonized ancient hosts, or that new hosts may have been colonized (Labandeira 2002b). The identification of fossil leaves frequently is not straightforward, and assignment to a certain plant family may be incorrect, especially in treatments from the older literature. With modern and more detailed methods, many plant-host identifications of the older literature subsequently have been revised, and in many cases it is impossible to assign fossil leaves to any particular, modern taxonomic group. One example is the assignments to Proteaceae listed by Berry (1916), of which none probably belong to that family (Dilcher 1973). Dilcher comments that "Many early Tertiary, and certainly many Cretaceous, fossil angiosperm leaves should not be expected to have characters that relate them at the generic level with modern forms." Recently, there has been the tendency to erect extinct genera and families for groups of angiosperm species that are extinct and are devoid of extant representatives (Friis *et al*. 2011). A more practical approach is use of a foliage morphotype system (Johnson 2002) to be followed by an upgrading to a Linnaean binomial once leaf morphology and variation has been extensively documented (Johnson, 1996). We have endeavoured to place the fossil hosts into their families and provide author names using several online sources (including http://fossilworks.org/ and http:// www.theplantlist.org/ for extant plants), but we have not ascertained whether all names are valid.

Some generalizations regarding nepticuloid patterns of host-plant occurrence may be useful to mention. All Nepticulidae feed on angiosperms, mostly on eudicots, except for a few species within *Stigmella* that feed on Poaceae or Cyperaceae. As well, some *Acalyptris* species feed on Cyperaceae and an unclassified species from Brazil has mines found on *Piper* (Piperaceae, Magnoliales) (Kemperman *et al*. 1985; Wilkinson 1979; van Nieukerken, unpublished data). In general, lepidopterous leaf mines on commelinid monocots are more likely created by Elachistidae (Kaila 2011). The nepticulid genus *Roscidotoga* is specialized on Oxalidales and is restricted to Australia (van Nieukerken *et al*. 2011a), although a few other Oxalidales feeders are known outside Australia (EJvN, unpublished data). Most species of the genus *Enteucha* are found on Polygonaceae (van Nieukerken 1986); only one other nepticulid is known to feed on this plant family, an unnamed species of *Acalyptris* that feeds on *Eriogonum* (D.L. Wagner, personal communication). Only species from the

genus *Parafomoria* and a species of *Stigmella*, *S. diniensis* (Klimesch), feed on Cistaceae (Sapindales) (van Nieukerken 1983).

Revision of published records

We reviewed the literature on fossils of Nepticuloidea and documented all references to fossils that have been assigned to Nepticuloidea. Where possible, we re-evaluated the characters used for specimen identification from images or drawings, and when available, the original photographs and amber fossils were obtained on loan for further study. When judging the identifications for material not in our possession, we were aware that the original authors probably had a better view of the material than we have from published images. Many of the records received new identifications based on our collective experience and insight. For records that could not be verified, we judged whether the assigned rank was plausible in terms of geological age and the characters described in the text. If they were plausible, we did not change the identification, but these records have their identification denoted with "[unverifiable]". Uncertain attribution to the identified rank is indicated with cf. (Latin: compare). In general, we were reluctant to assign an extant, genus-level rank to a leaf-mine fossil; most are ranked under Nepticulidae *incertae sedis* or are placed in the ichnogenus *Stigmellites*. The only exceptions allowed were for subfossil leaf mines that resemble more recent extant species, such as *Stigmella ulmivora* (Fologne) (record #14), or mines from host plants and regions where we see a continuous record between the Neogene fauna and the present fauna—for example: *Stigmella* on Californian oaks—in which mines closely resemble extant ones (Opler 1973). Several original descriptions and translations, and when possible, illustrations are placed on the website Nepticuloidea.info (van Nieukerken 2014). We reviewed 79 fossil records, of which many contain multiple exemplars. One correction for several records involved a preservationally exceptional deposit, or Lagerstätte, for Neogene leaf fossils, Willershausen am Harz, in Germany. This deposit was confused by Sohn *et al*. (2012) with another Willershausen locality in the federal state of Hesse (Hessen), which resulted in the confounding addition of the state of Brandenburg. Instead, fossil Lagerstätte of Willershausen am Harz is located in the state of Niedersachsen (Lower Saxony), and belongs to the municipality (Gemeinde) of Kalefeld, in the District (Landkreis) of Northeim (N51.7845° E10.1087°).

Each unique combination of geologic stage, locality, host and identification is given a record number. A single record can include multiple exemplars or undescribed taxa. We treat the fossils in the checklist by preservational type and by age: first the adults from the oldest to recent, and then the compressions or impressions of leaf-mines on fossil foliage from oldest to recent. We have numbered the fossil records and present details for each fossil in the following format:

Record number *Genus species* author. This is the revised identification. A chronological list of previous identifications and publications.

- fossil type—Host: host plant—[number of exemplars] Coll: collection
- Loc: the fossil locality
- Stratum: the fossil deposit
- Remarks: revisionary comments and observations

Alphabetical list of collections abbreviations

BMNH Department of Paleontology, Natural History Museum, London, United Kingdom

BPGM Bavarian State Collection for Paleontology and Geology (= Bayerische Staatssammlung für Palaeontologie und Geologie), Munich, Bavaria, Germany

FMUF Florida Museum of Natural History, University of Florida, Gainesville, Florida, U.S.A.

GBIU Department of Geological Sciences and Biology, Indiana University, Bloomington, Indiana, U.S.A.

GDVU Geology Department of Victoria University, Wellington, Victoria, Australia

GPUG Geological-Paleontological Institute, University of Göttingen (= Geologisch-Paläontologisches Institut, Universität Göttingen), Göttingen, Lower Saxony, Germany

HLDG Museum Wiesbaden (= Hessischen Landesmuseums), Darmstadt, Hesse, Germany

IEUH Institute of Evolution, University of Haifa, Israel

IGUSP Institute of Geoscience, University of São Paulo (= Instituto de Geociências, Universidade de São Paulo), São Paulo, Brazil

MCNV Museum of Natural Science in Valencia (= Museo de Ciencias Naturales de Valencia), Valencia, Spain

MPEF Egidio Feruglio Paleontologic Museum (= Museo Paleontólogico Egidio Feruglio), Trelew, Chubut, Argentina

MVVA National Museum of Victoria, Victoria, Australia

NMPC National Museum (= Národní Muzeum or Musei Nationalis Pragae), Prague, Czech Republic

PIRAS Paleontological Institute, Russian Academy of Sciences, Moscow, Russia

PMNH Peabody Museum of Natural History, Yale University, New Haven, Connecticut, U.S.A.

QMSB Queensland Museum, South Brisbane, Queensland, Australia

TBMM Thomas Burke Memorial Museum, University of Washington, Seattle, Washington, U.S.A.

UCMP University of California Museum of Paleontology, Berkeley and Davis, California, U.S.A.

USNM United States National Museum of Natural History, Washington, DC, U.S.A.

Nomenclatural note

We use the original endings for species names that are combined with *Stigmellites*, despite the fact that generic names ending with the suffix "*-ites*" should be considered as masculine (International Commission on Zoological Nomenclature 1999, article 30.1.4.4). All adjectival species names published as *Stigmellites* have feminine endings, showing the intent of the authors. We follow the practice of lepidopterists to retain the original spelling. This practice is in accordance with a resolution adopted by the Societas Europaea Lepidopterologica in 2002 (Sommerer 2002).

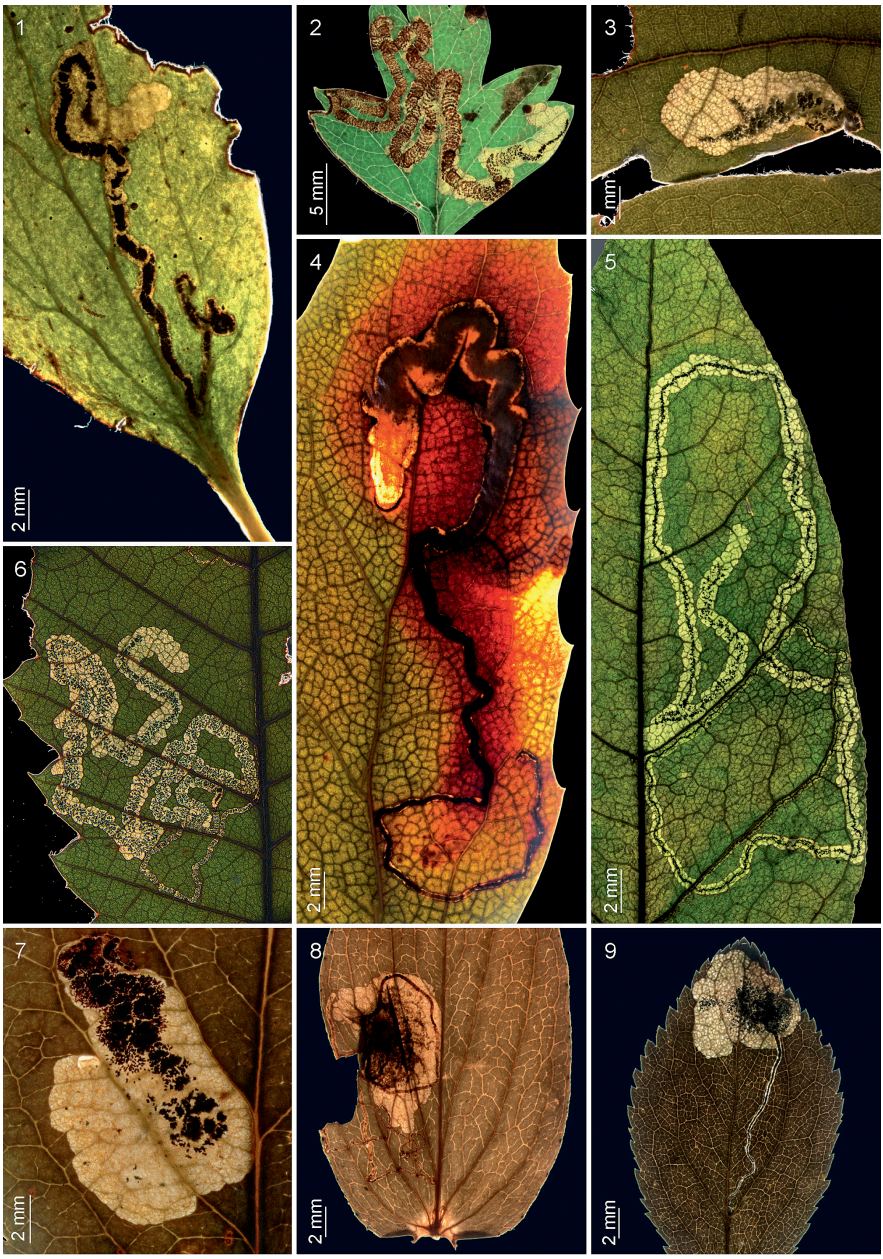
Photography
Photographs of dried or fresh leaf mines were taken using a dark-field illumination setup on either a Zeiss Axioskop H or Zeiss Discovery.V20 microscope with a Zeiss AxioCam MR5 and AxioVision version SE64 4.9 software. Partial photographs were merged using the Adobe Photoshop CS6 photomerge tool and optimized by adjusting levels and curves. Photography of the fossil mines were done on a Canon EOS 500 camera with an EFS 60mm macro lens, accompanied by various combinations of low-angle indirect or direct illumination to accentuate subtle leaf-mine features. Processing of photo images were done by standard Photoshop methods, except for Fig. 10, which is an archival black-and-white photograph processed by standard darkroom techniques. Adobe Indesign CC was used to assemble individual images into a plate.

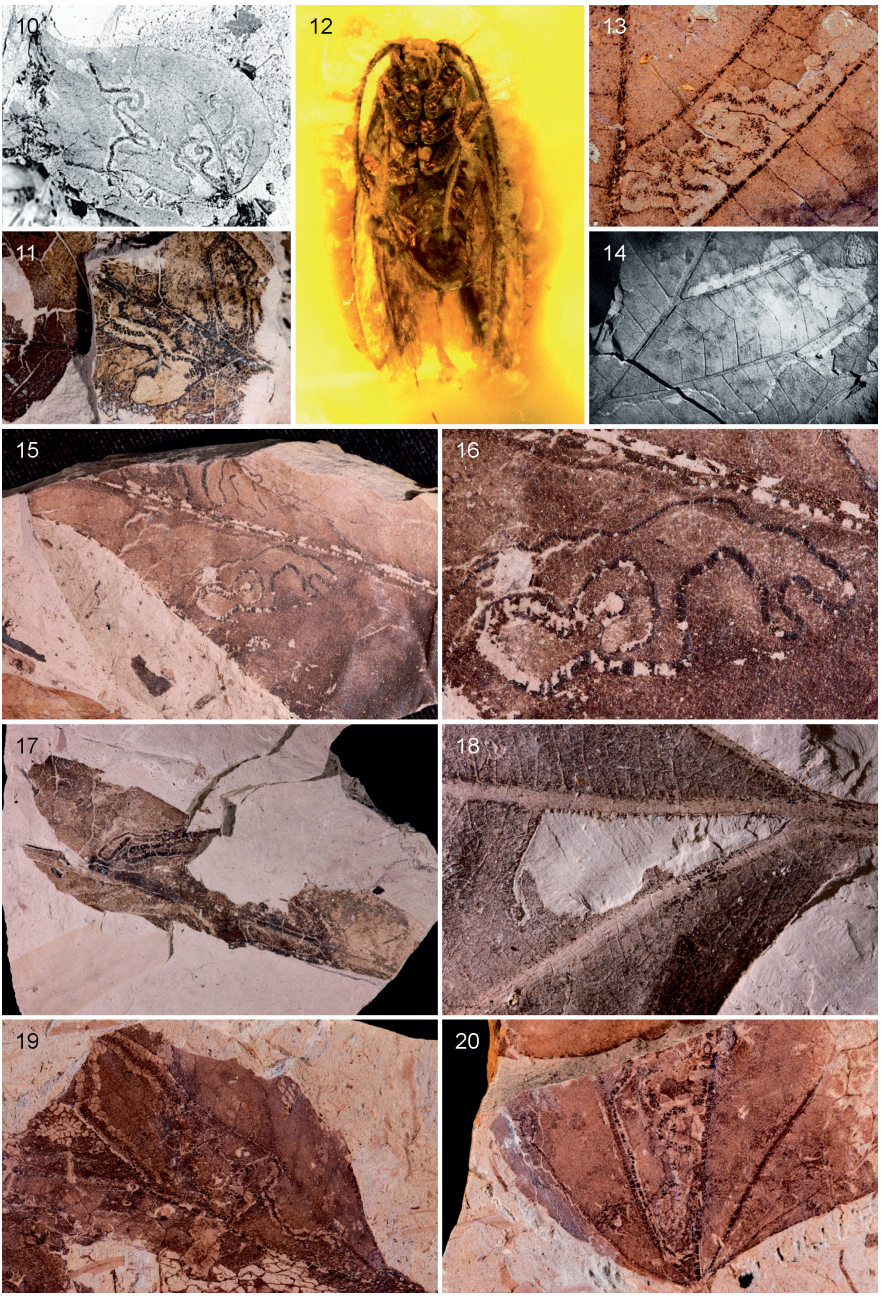
Checklist

Adult body fossils

 # 1 *Ectoedemia* sp. [unverifiable]
Ectoedemia sp.; Skalski 1976: 199.
Ectoedemia sp.; Skalski 1990a: 127
Ectoedemia; Sohn *et al*. 2012: 22
- Adult in amber—[1 ex.] Coll: not stated
- Loc: Baltic Region (Baltic Amber)
- Stratum: Prussian Fm.; Lutetian Stage, middle Eocene
- Remarks: It is unclear as to what basis this fossil was assigned to *Ectoedemia*. There exists no description or images of the fossil. From his publications it is clear that Skalski used wing venation as one of the principal characters. If the venation had indeed been completely visible, assignment to *Ectoedemia* s. l. could be correct.

Figs 1–9. Leafmines of selected extant Nepticulidae from the Naturalis collection. 1. *Acalyptris loranthella* (Klimesch) on *Loranthus europaeus* (Loranthaceae) near Foloi, Greece (EvN2011367). 2. *Stigmella oxyacanthella* (Stainton) on *Crataegus monogyna* (Rosaceae) near Wassenaar, The Netherlands (EvN2007118, RMNH.INS.12638). 3. *Ectoedemia caradjai* (Groschke) on *Quercus pubescens* (Fagaceae) near Exokhori, Greece (EvN2011380–1). 4. Undescribed *Stigmella* on *Berberis nervosa* (Berberidaceae) in Gifford Pinchot National Forest, Washington state, USA (EvN2010027). 5. Undescribed *Trifurcula (Glaucolepis)* on *Wisteria* (Fabaceae) near Aomori, Japan (CD13060). 6. *Roscidotoga callicomae* Hoare on *Callicoma serratifolia* (Cunoniaceae) in Lamington National Park, Queensland, Australia (EvN2004078). 7. *Bohemannia pulverosella* (Stainton) on *Malus sylvestris* (Rosaceae) in Barendrecht, The Netherlands (Koster nr 2582). 8. *Ectoedemia* (*Fomoria*) *septembrella* (Stainton) on *Hypericum perforatum* (Hypericaceae) in Losser, The Netherlands (Koster nr 1789). 9. *Stigmella plagicolella* (Stainton) on *Prunus spinosa* (Rosaceae) in Losser, The Netherlands (Koster nr 1984). Photographs by CD and EJvN (6). ▶


1
2 mm
2
5 mm
3
2 mm
4
2 mm
5
2 mm
6
7
2 mm
8
2 mm
9
2 mm


10
12
13
11
14
15
16
17
18
19
20

 # 2 *Stigmellites baltica* Kozlov, 1988
Stigmellites baltica Kozlov, 1988: 30, fig. 4
Stigmellites baltica; Skalski 1990a: 127
Stigmellites balticus; Sohn *et al*. 2012: 24
- Adult in amber—[1 ex.] Coll: Lost: stolen by thieves from the collection of K. M. Sadilenko, Moscow, Russia (HT: no. 15–1–4)
- Loc: Baltic Region (Baltic Amber)
- Stratum: Prussian Fm.; Lutetian Stage, middle Eocene
- Remarks: Unfortunately the holotype could not be studied, as the original material was lost during a robbery (Kozlov pers. comm.). The illustrated and described venation mostly resembles modern *Stigmella*, but other genera cannot be excluded.

 # 3 *Bohemannia aschaueri* Fischer, 2013
Bohemannia aschaueri Fischer, 2013: 88, fig. 3.
- Adult in amber—[1 ex.] Coll: BPGM (HT: SNSB-BSPG 2013 I 94), Fischer no. 5199
- Loc: Russia: Amber mine at Yantarny
- Stratum: Blaue Erde horizon, Prussian Fm.; Lutetian Stage, middle Eocene (Ritzkowski 1997)
- Remarks: The right wing shown in Fischer's Fig. 3d is a hindwing. The drawn short-side vein of vein number 3 (Cu) is, in our opinion, absent, which makes the hindwing venation identical to many modern Nepticulidae, including *Bohemannia*. The abdominal tip is wide, but the specimen is a male; male genitalia could be revealed during a preliminary micro CT scan. The forewing shows purplish scales, much as in modern *Bohemannia*. The venation, size, broad habitus and colour confirm the generic attribution: the species resembles modern *B. quadrimaculella* (Boheman), which is the type species of the genus. [Specimen examined by EJvN].

◀ *Figs 10–20*. Fossil representatives of Nepticulidae. 10. Nepticulidae *incertae sedis* on *Cercidiphillym* in Wyoming, USA, from the early Maastrichtian Stage, Late Cretaceous [#27]. 11. Nepticulidae *incertae sedis* on *Sapindopsis beckeriana* in Kansas, USA, from the late Albian Stage, Early Cretaceous [#14]. 12. *Bohemannia butzmanni* Fischer holotype in Baltic Amber from Russia, from the Lutetian Stage, Middle Eocene (Fischer no. 5058) [#4]. 13. Nepticulidae *incertae sedis* on *Platanus raynoldsii* in Montana, USA, from the Danian Stage, Early Paleocene. [#28]. 14. *Stigmellites kzyldzharica* Kozlov paratype on *Platanus ambicula* in Kazakhstan, from the Turonian Stage, Late Cretaceous [#20]. 15. Nepticulidae *incertae sedis* on *Pandemophyllum kvacekii* in Nebraska, USA, from the late Albian Stage, Early Cretaceous [#13]. 16. Detail of the lower mine on 15. 17. Nepticulidae *incertae sedis* on *Anisodromum upchurchii* in Kansas, USA, from the late Albian Stage, Early Cretaceous [#17] 18. Nepticulidae *incertae sedis* on an undetermined leaf of Platanaceae in Kansas and Nebraska, USA, from the late Albian Stage, Early Cretaceous [#12]. 19. Nepticulidae *incertae sedis* of DT91 on *Zizyphoides flabella* in Montana, USA, from the Danian Stage, Early Paleocene [#30]. 20. Nepticulidae *incertae sedis* of DT41 and DT91 on *Cercidiphyllum genetrix* in Montana, USA, from the Danian Stage, Early Paleocene [#31]. Photographs by CCL, EJvN (12), M. Kozlov (14) and M. Donovan (19, 20).

 # 4 *Bohemannia butzmanni* Fischer, 2013 (Fig. 12)
Bohemannia butzmanni Fischer, 2013: 86, fig. 1.
- Adult in amber—[1 ex.] Coll: BPGM (HT: SNSB-BSPG 2013 I 93) ex coll. Fischer no. 5058
- Loc: Russia: Amber mine at Yantarny
- Stratum: Blaue Erde horizon; Prussian Fm.; Lutetian Stage, middle Eocene (Ritzkowski 1997)
- Remarks: The venation of the holotype is identical to that of modern *Bohemannia*. The specimen has a broad, blunt abdomen, and most likely is a female. The extensions that are termed 'valvae' in the original description are probably protruding scales. The specimen is not significantly different from the holotype of *B. aschaueri* (#3), and they could well be conspecific. [Specimen examined by EJvN].

 # 5 cf. *Stigmella* sp.
Nepticulidae; Fischer 2013: 86, fig. 2
- Adult in amber—[1 ex.] Coll: Fischer collection no. 5217, possibly also 5058
- Loc: Russia: Amber mine at Yantarny
- Stratum: Blaue Erde horizon, Prussian Fm.; Lutetian Stage, middle Eocene (Ritzkowski 1997)
- Remarks: According to the author, this specimen may be conspecific with *Bohemannia butzmanni* (# 4). We disagree with that possibility: the forewing venation clearly differs, with fewer terminal branches of R+M (3 rather than 4 or 5); the antennae are much shorter (20 segments in a complete antenna versus incomplete antenna with at least 35 segments for *Bohemannia*); and the individual flagellomeres are longer, whereas those in *Bohemannia* (and many *Ectoedemia*) are short. This specimen most likely belongs to *Stigmella*. [Specimen examined by EJvN].

 # 6 Nepticulidae: *incertae sedis*
Nepticulidae; Fischer 2013: 88, fig. 4
- Adult in amber—[1 ex.] Coll: Fischer collection no. 5198
- Loc: Russia: Amber mine at Yantarny
- Stratum: Blaue Erde horizon, Prussian Fm.; Lutetian Stage, middle Eocene (Ritzkowski 1997)
- Remarks: Unfortunately the venation cannot be seen, and initial study with micro-CT scans has not yet provided a clear view of the specimen. This specimen could be a *Stigmella*, but an alternative genus also is possible. [Specimen examined by EJvN].

 # 7 Nepticulidae: *incertae sedis*
Nepticulidae; Fischer 2013: 91, fig. 5
- Adult in amber—[1 ex.] Coll: Fischer collection no. 5166
- Loc: Russia: Amber mine at Yantarny
- Stratum: Blaue Erde Horizon, Prussian Fm.; Lutetian Stage, middle Eocene (Ritzkowski 1997)

- Remarks: Re-examination of the fossil has not yet resulted in more precise identification. [Specimen examined by EJvN].

 # 8 cf. *Stigmella* sp.
Incurvariina or Nannolepidoptera species A; Jarzembowski 1980: 270, fig. 50
Stigmellites; Kozlov 1988: 32
Stigmellites; Sohn *et al*. 2012: 25
- Adult impression—[1 ex.] Coll: BMNH (I.9492)
- Loc: United Kingdom: England, Isle of Wight, Bembridge Marls
- Stratum: Bouldnor Fm.; late Priabonian Stage, late Eocene
- Remarks: This is one of two known adult impression fossils. Kozlov was the first to notice that the venation is clearly nepticulid. We think that the venation most closely resembles *Stigmella*, but there are insufficient visible veins to completely exclude *Acalyptris*.

 # 9 Nepticulidae: *incertae sedis*
Tineoidea species C; Jarzembowski 1980: 271, fig. 57
Stigmellites; Kozlov 1988: 32
Stigmellites "Species B"; Sohn *et al*. 2012: 26
- Adult impression—[1 ex.] Coll: BMNH (In.64540)
- Loc: United Kingdom: England, Isle of Wight, Bembridge Marls
- Stratum: Bouldnor Fm.; late Priabonian Stage, late Eocene
- Remarks: This is one of two known adult impression fossils. It is listed as "species B" in Sohn *et al*. (2012), but species B in Jarzembowski (1980) is not depicted and is referred perhaps to *Heliozela*. Kozlov (1988) places Jarzembowski's species A (see record # 8) and species C (this record) in *Stigmellites*. The venation is incomplete, but resembles a nepticulid. The hindwing shows a trifurculine condition of Rs+M, which is characteristic for the genus *Trifurcula*, but we find assignment to that genus premature.

10 *Acalyptris* sp. [unverifiable]
Niepeltia sp.; Skalski 1990a: 127
Acalyptris; Skalski 1990b: 144
Acalyptris; Sohn *et al*. 2012
- Adult in copal—[1 ex.] Coll: not stated
- Loc: Tanzania: Zanzibar Island
- Stratum: East African Copal from unconsolidated sediments (Holocene Stage)
- Remarks: This fossil has been assigned to *Acalyptris*, although the characters to base this identification have not been indicated. Judging from Skalski's other publications he likely used wing venation as a leading character, which is quite characteristic for this genus (van Nieukerken 1986). There are no images or drawings of the fossil, nor is there the option to study the object due to an absence of reference to a collection. We leave this fossil in *Acalyptris* because there are no inconsistencies in characters and it involves an almost modern subfossil specimen.

 # 11 *Enteucha* sp. [unverifiable]
Johanssonia; Skalski 1976: 199
Johanssoniella; Sohn *et al*. 2012: 22
• Adult in copal—[1 ex.] Coll: not stated
• Loc: unclear
• Stratum: unclear, but undoubtedly copal from unconsolidated sediments of Pleistocene or Holocene Age.
• Remarks: The locality is unclear, although it was likely the Baltic region (Skalski 1976). In the publication, it was listed as *Johanssonia* Borkowski in a table with specimens from Baltic Amber: "avec un seule espèce fossile non encore décrite" [with a single undescribed fossil species]. A footnote however indicates that this fossil was embedded in copal instead of Baltic Amber and is thus essentially modern in age, compared to the other specimens in the table. There are no images or drawings, only the mention that it was placed in the genus *Johanssonia*. *Johanssoniella* Koçak, 1981 is a replacement name for *Johanssonia* Borkowski, 1972, but both are subjective junior synonyms of the extant genus *Enteucha* Meyrick, 1915 (synonymised by van Nieukerken 1986). There is no reference to a collection.

Leaf-mine fossils

 # 12 Nepticulidae: *incertae sedis* multiple species (Fig. 18)
Ectoedemia; Labandeira *et al*. 1994: 12279, figs. 1a–d
Ectoedemia; Sohn *et al*. 2012: 21
• Leaf mine—Host: Platanaceae: indeterminate genus—[11 exx.] Coll: FMUF (UF12701; UF7255 etc.)
• Loc: USA: Kansas and Nebraska, Braun Ranch, Hoisington and other localities
• Stratum: Dakota Fm.; late Albian Stage, Early Cretaceous
• Remarks: There are 11 leaf mines on Platanaceae, and one leaf mine on an undesignated host. Kristensen & Skalski (1998) cited this record as the earliest fossil evidence of Nepticulidae and also of the extant genus *Ectoedemia*. The original dating of these fossils was 97 Ma (Labandeira *et al*. 1994). However, by a combination of recent stratigraphic evaluation of the Dakota Formation (Brenner *et al*. 2000), and updates in global geochronological tie-points (Ogg *et al*. 2008), the lower portion of the Dakota Formation is re-dated at 102 Ma. We consider it impossible that these mines are related to modern *Platanus* feeders in the genus *Ectoedemia*, which belong to a much later evolved, subordinate clade (Doorenweerd *et al*. 2015), even though there is a superficial resemblance.

 # 13 Nepticulidae: *incertae sedis* (Figs 15, 16)
Stigmella; Labandeira *et al*. 1994: 12279, 12280, fig. 1e
cf. Stigmella; Sohn *et al*. 2012: 23
• Leaf mine—Host: Laurales: *Pandemophyllum kvacekii* Upchurch and Dilcher, 1990—[1 ex.] Coll: FMUF (UF12712)
• Loc: USA: Nebraska, Rose Creek

- Stratum: Dakota Fm.; late Albian Stage, Early Cretaceous
- Remarks: The mines resemble modern *Stigmella*, but because these cannot be separated from several other genera, we consider them as Nepticulidae *incertae sedis*. Sohn *et al*. (2012) combined this record with the following two, but due to their occurrence on different host plants, we separate them here. There are no records of extant Nepticulidae feeding on Laurales.

 # 14 Nepticulidae: *incertae sedis* (Fig. 11)
Stigmella; Labandeira *et al*. 1994: 12279, 12280, figs. 1f–g
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Platanaceae: *Sapindopsis beckeriana* Wang, 2002—[1 ex.] Coll: FMUF (UF4811)
- Loc: USA: Kansas, Hoisington
- Stratum: Dakota Fm.; late Albian Stage, Early Cretaceous
- Remarks: See record # 13 for remarks, only differing in host plant. *Sapindopsis* is one of the earliest appearing lineages of Platanaceae, confined to the Cretaceous and consists of pinnately-compound leaves. The specimen illustrated in Fig. 11 is a leaf-mined leaflet.

 # 15 Nepticulidae: *incertae sedis*
Stigmella; Labandeira *et al*. 1994: 12279, 12280, fig. 1h
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Sapindales (family uncertain): *Anisodromum wolfei* Upchurch and Dilcer, 1990—[1 ex.] Coll: FMUF (UF12718)
- Loc: USA: Nebraska, Rose Creek—Dakota Fm.
- Stratum: late Albian Stage, Early Cretaceous
- Remarks: see record # 13 for remarks, only differing in host plant. Originally the higher host ranking was indicated as "Rosidae", however, based on the taxonomic framework of Bell *et al*. (2010) and classification of Wang (2002), this plant would be in the (Order) Sapindales.

 # 16 cf. Nepticulidae
Stigmella; Labandeira 1998: 110, fig. 3d
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: Laurales: *?Pabiania*—[1 ex.] Coll: FMUF (UF7252)
- Loc: USA: Kansas, Cloud Co., Braun's Ranch
- Stratum: Dakota Fm.; late Albian Stage, Early Cretaceous
- Remarks: Kristensen and Skalski (1998) cited this record in addition to the following as the earliest fossil evidence of Nepticulidae and also of the extant genus *Stigmella*. The mine is very long for a nepticulid mine and not clearly increasing in width; it perhaps resembles more an epidermal type of mine like those created by species of *Phyllocnistis* (Gracillariidae). There are no records of extant Nepticulidae feeding on Laurales.

 # 17 Nepticulidae: *incertae sedis* (Fig. 17)
Stigmella; Labandeira 1998: 110, fig. 3e
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: Sapindales (family uncertain): *Anisodromum upchurchii* Wang, 2002—[1 ex.] Coll: FMUF (UF16173)
- Loc: USA: Kansas, Cloud Co., Braun's Ranch
- Stratum: Dakota Fm.; late Albian Stage, Early Cretaceous
- Remarks: Kristensen and Skalski (Kristensen & Skalski 1998) cited this record as well as the former occurrence as the earliest fossil evidence of Nepticulidae and the extant genus *Stigmella*. The mine appears very nepticulid-like, but can belong to a number of different genera.

 # 18 Nepticulidae: *incertae sedis* multiple species
Nepticulidae Mine type KLmla, KLm1b, KLm1c, KLm2, KLm3, KLm11; Stephenson 1991: 154–156, 163 *cf. Stigmella*/Nepticulidae; Sohn *et al*. 2012: 23, 26
- Leaf mine—Host: Angiosperms—[32 exx.] Coll: GBIU (IU15706–4811; IU15706–7525; IU15706–7528; IU15709–4818; IU15709–7531; IU15709–7535; IU15706–4539; IU15706–7521; IU15706–7525; IU15706–7527; IU15706–4810; IU15703–3856; IU15703–7523a; IU15706–7255; IU15706–7256; IU15709–3950; IU15709–4819; IU15713–4696; IU15713–4834; IU15713–4936; IU15713–7242; IU15713–7243; IU15713–7244; IU15713–7245; IU15713–7246; IU15723–7247; IU15713–7248; IU15713–7249; IU15713–7324; IU15706–4536; IU15706–7113; IU15714–7250)
- Loc: USA: Kansas and Nebraska, Braun Ranch, Hoisington and other localities (Dakota Fm.)
- Stratum: late Albian Stage, Late Cretaceous
- Remarks: The author compared these fossil mines with dried leaf mines from the Hering collection (BMNH), and suggests recent analogues for different types of leaf mines. Records 12 and 18 were combined by Sohn *et al*. (2012) but actually represent specimens from different time intervals. Record # 18 was also mentioned separately by Sohn *et al*. (2012: 26).

 # 19 *Stigmellites serpentina* Kozlov, 1988
Nepticulidae; Skalski 1979: 64
Stigmellites serpentina Kozlov, 1988: 32, pl. 2: 2 *Stigmellites serpentina*; Skalski 1990a: 127 Nepticulidae; Boucot 1990: 108, fig. 102
Stigmellites serpentina; Sohn *et al*. 2012: 25, 26
- Leaf mine—Host: Cercidiphyllaceae: *Trochodendroides arctica* Heer—[3 exx.] Coll: PIRAS (HT: PIN 2383/205)
- Loc: Kazakhstan: Kyzyl Orda Prov., Chilinsky, northwest spur of Karatau Mountain range, Kyzyl-Dzhar
- Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous
- Remarks: This occurrence was split into two records by Sohn *et al*. (2012), one referring to Skalski and Boucot on page 26 and one to Kozlov on page 25. However, the photograph published in Boucot (1990), referring to Skalski (1979), shows the entire leaf with three mines, of which Kozlov depicted a single

mine and used it to designate the holotype. The image published by Kozlov is a mirror, facsimile version of the original. This mine was listed by Sohn *et al*. (2012) as occurring during the Oxfordian-Kimmeridgian Stages of the Late Jurassic, which would make this record represent by far the oldest nepticulid fossil. Nevertheless, Kozlov (1988) and Boucot (1990) cited the Turonian Stage of the early Late Cretaceous as the age for this occurrence. At Karatau, in the mountains of southernmost Kazakhstan, there are two intervals of strata bearing insects and plants, occurring in vertical succession. The older deposits are, indeed, Late Jurassic, and belong to the Oxfordian and Kimmeridgian Stages. The younger deposit is of early Late Cretaceous Age (Turonian Stage), and also contains fossil plants and insects of similar age. It appears that these two sequences were confused and geochronologically reversed (Friis *et al*. 2011); we consider the Turonian Stage the correct date for this record. Skalski (1979) cited this as a leaf mine from Karatau, "very similar to leaf mines produced by some existing species, e.g. *Nepticula tityrella* Stainton". We agree that this mine very much resembles modern Nepticulidae and is attributable to several constituent genera.

 # 20 *Stigmellites kzyldzharica* Kozlov, 1988 (Fig. 14)
Eriocraniidae; Zherikhin 1978: 79
Nepticulidae; Skalski 1979: 64
Nepticulidae; Zherikhin 1980: 89
Stigmellites kzyldzharica Kozlov, 1988: 32, fig. 5, pl. 2: 1
Stigmellites kzyldzharica; Skalski 1990a: 127
eriocranid; Grimaldi & Engel 2005: 572, fig. 13: 32.
Stigmellites kzyldzharicus; Sohn *et al*. 2012: 25
- Leaf mine—Host: Platanaceae*: Platanus ambicula* Vachr.– [2 exx.] Coll: PIRAS (HT: PIN 2383/206; PT: PIN 2383/214)
- Loc: Kazakhstan: Kyzyl Orda Prov., Chilinsky, northwest spur of Karatau Mountain Range, Kyzyl-Dzhar
- Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous
- Remarks: The mine is very long and narrow, and resembles somewhat modern *Acalyptris* mines. Extant *Platanus* species are host for *Acalyptris platani* (Müller-Rutz) in Europe (van Nieukerken 2007) and for three *Ectoedemia* species in North America (Wilkinson & Newton 1981; Doorenweerd *et al*. 2015).

 # 21 *Stigmellites samsonovi* Kozlov, 1988
Stigmellites samsonovi Kozlov, 1988: 33, pl. 2: 3
Stigmellites samsonovi; Skalski 1990a: 127
Stigmellites samsonovi; Zherikhin 2002: 321, fig. 475
Stigmellites samsonovi; Sohn *et al*. 2012: 25
- Leaf mine—Host: Cercidiphyllaceae: *Trochodendroides arctica* Heer—[1 ex.] Coll: PIRAS (HT: PIN 2383/209)
- Loc: Kazakhstan: Kyzyl Orda Prov., Chilinsky, northwest spur of Karatau Mountain Range, Kyzyl-Dzhar
- Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous

• Remarks: The mine begins as a narrow gallery and abruptly expands into a blotch, as seen with many extant *Ectoedemia* (s. str.) species, but also some leaf-mining species in other genera. This specimen is likely to be a nepticulid mine.

***# 22 Stigmellites sharovi* Kozlov, 1988**
Stigmellites sharovi Kozlov, 1988: 33, pl. 2: 4
Stigmellites sharovi; Skalski 1990a: 127
Stigmellites sharovi; Sohn *et al*. 2012: 25
• Leaf mine—Host: Cercidiphyllaceae: *Trochodendroides arctica* Heer—[1 ex.] Coll: PIRAS (HT: PIN 2383/208)
• Loc: Kazakhstan: Kyzyl Orda Prov., Chilinsky, northwest spur of Karatau Mountain Range, Kyzyl-Dzhar
• Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous
• Remarks: The attribution of this mine to the Nepticulidae seems likely.

***# 23 Stigmellites tyshchenkoi* Kozlov, 1988**
Stigmellites tyshchenkoi Kozlov, 1988: 33, pl. 2: 5
Stigmellites tyshchenkoi; Skalski 1990a: 127
Stigmellites tyshchenkoi; Zherikhin 2002: 321, fig. 475
Stigmellites tyshchenkoi; Sohn *et al*. 2012: 25
• Leaf mine—Host: Platanaceae: *Platanus latior* (Lesquereux) Knowlton—[1 ex.] Coll: PIRAS (HT: PIN 2383/211)
• Loc: Kazakhstan: Kyzyl Orda Prov., Chilinsky, northwest spur of Karatau Mountain Range, Kyzyl-Dzhar
• Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous
• Remarks: The photographs in both publications show rather different views of the same mine. The photograph of the mine in Kozlov (1988) appears to be free of frass, a feature which is also stated in the description ("no excrement line visible"). By contrast, in Zherikhin (2002) there is a clear black frass line with a fine margin from the edges of the mine. The original photographs from Kozlov (1988), now are available on http://nepticuloidea.info/nepticuloidea/stigmellites-tyshchenkoi, reveals that the frass likely is a very shiny and black. This image was overexposed during photography, making the specimen seem devoid of frass. We believe the fossil and photographs actually contain two mines. See also #20, which differs by one mine having a thin frass line.

***# 24 Stigmellites resupinata* (Krassilov, 2008) comb. nov.**
Lepidoptera; Krassilov 2007: 14, fig 1D,E
Ophiheliconoma resupinata Krassilov, 2008b: 100, pl. 34: 1,2
• Leaf mine—Host: Family unknown: *Dewalquea gerofitica* (Dobruskina) Krassilov—[1 ex.] Coll: IEUH, (HT IG1– 139)
• Loc: Israel: southern Negev, Gerofit
• Stratum: mid-Turonian Stage, Late Cretaceous
• Remarks: Krassilov used an ichnotaxonomic ranking that cannot easily be correlated with extant taxonomic ranks. In the species description it is mentioned that the mine is "nepticuliform," and the number of instars (likely 6) appear to

match that of extant *Stigmella*. However, most *Stigmella* have 4 or 5 instars. In Krassilov (2007), the detailed image (fig. 1E) is mentioned to depict an "end-blotch … with a hibernating cocoon." However, in Krassilov (2008), this section was re-interpreted as an intestiniform beginning of the mine, gradually increasing in width, which we also believe as more likely. The ovoidal scar mentioned at the oviposition site might equally be an egg-capsule. The original species description did not assign the species to a higher rank, rendering it a *nomen nudum*. However, we believe that the combined evidence is sufficient to assign this mine to Nepticulidae. The higher taxonomic rank for the host is possibly Myrtales or Rhizophorales (now placed in Malpighiales) (Krassilov 2008b), and the host species was an early angiosperm that had a mangrove-like habit.

 # 25 Nepticulidae: *incertae sedis* [unverifiable]
cf. Stigmella; Donner & Wilkinson 1989: 9
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Coll: Christopher Wilkinson
- Loc: Kazakhstan
- Stratum: Beleuty Fm.; Turonian Stage, Late Cretaceous
- Remarks: Wilkinson had borrowed several specimens that he reports as "exactly similar to *Stigmella* today", from a deposit he claims to be 110 Ma, but also states that it is the same age as Turonian (= 89.8—93.3 Ma). No further details are given. It is possible that this literature record refers to the more recent, Late Cretaceous, fossils from Kazakhstan, as in records # 19–23.

 # 26 Nepticulidae: *incertae sedis*
Stigmella; Labandeira *et al*. 2002: 2062, fig. 1h
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Rosaceae: aff. *Rubus*—[1 ex.] Coll: YPM (6367a)
- Loc: USA: SW North Dakota, Williston Basin, near Marmarth
- Stratum: Hell Creek Fm.; latest Maastrichtian Stage, Late Cretaceous
- Remarks: In Labandeira *et al*. (2002) the host was identified as "Rosaceae"; in Sohn *et al*. (2012), it was listed as "*cf. Rubus*". The leaf with the mine exactly matches leaf morphotype HC80, designated as "aff *Rubus*" in Johnson (2002), from the Hell Creek flora of the Williston Basin. The leaf is morphologically consistent with modern-day *Rubus*, but a Late Cretaceous age seems unlikely for this modern genus.

 # 27 Nepticulidae: *incertae sedis* DT43 (Fig. 10)
Nepticulidae/*Stigmella*; Labandeira 2002a: 49, 252, fig. 2.10e–f
Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: Cercidiphyllaceae: *Cercidiphyllum* sp.—[1 ex.] Coll: USNM
- Loc: USA: Wyoming, Washakie Co., Big Cedar Ridge
- Stratum: Meeteetsee Fm.; early Maastrichtian Stage, Late Cretaceous
- Remarks: Only part of the mine is visible in the figure, but the specimen has the characteristics of a Nepticulidae mine. In the figure caption the mine is mentioned to be nepticulid, but in the appendix of Labandeira (2002a), together with the

supporting information for the specimens that are depicted, the leaf mine is identified as *Stigmella*.

 # 28 Nepticulidae: *incertae sedis Platanus* DT91, DT282 (Fig. 13)
Lepidoptera; Donovan *et al*. 2014: fig. 2a–h, fig. 3e
- Leaf mine—Host: Platanaceae: *Platanus raynoldsii* Newberry—[21 exx.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: The authors identified material by damage type for each insect order. 1073 leaves with varied insect damage were investigated. The figures in the publication show Lepidoptera leaf mines as DT91 and a likely Lepidoptera leaf mine as DT282; both are on *Platanus raynoldsii*. We believe these leaf mines are likely to be nepticulid.

 # 29 Nepticulidae: *incertae sedis Juglandiphyllites* DT91, DT105
Insect-feeding damage; Donovan *et al*. 2014: fig. 3j, fig. 5
- Leaf mine—Host: Family unknown: *Juglandiphyllites glabra* Manchester & Dilcher—[5 exx.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: The frass trail of DT91 mines on *J. glabra* differs from DT91 mines on other hosts from the same publication (# 28–# 35); the frass is spheroidal and spread out in a loose trail or is completely absent in some areas.

The distance between frass and pellet accumulations is greater than for similar mines occurring on other host plants. It is likely there are different species involved in these DT91 mines.

 # 30 Nepticulidae: *incertae sedis Zizyphoides* DT91 (Fig. 19)
Leaf mines; Donovan *et al*. 2014: fig. 8
- Leaf mine—Host: Trochodendraceae: *Zizyphoides flabella* (Newberry) Crane, Manchester & Dilcher—[6 exx.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: DT91 mines on *Z. flabella* originate near a secondary vein and loop around to the adjacent secondary vein, following it until termination. These mines are likely to be Nepticulidae. The host *Zizyphoides* is only known from fossil leaves, but is always found in conjunction with *Nordenskioldia*, which is only known from infructescences. *Zizyphoides* (leaves) and *Nordenskioldia* (fruits) are presumed to be congeneric and placed within Trochondendraceae (Pigg *et al*. 2001).

 # 31 Nepticulidae: *incertae sedis Cercidiphyllum* DT41, DT91 (Fig. 20)
Leaf mines; Donovan *et al*. 2014: fig. 10e–g
- Leaf mine—Host: Cercidiphyllaceae: *Cercidiphyllum genetrix* (Newberry) Hickey—[2 exx.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: DT41 mines (Fig. 10e–k in Donovan *et al*., 2014) mentioned in this publication are found on *Cercidiphyllum* over a period of 6 m.yr. (million years), but likely represent different species based on the different mine morphologies. Mines from the same host, also from Wyoming, have also been recorded from the latest Cretaceous (66–72.1 Ma), as record # 27. This indicates that the association between this host and Nepticulidae minimally spans approximately 13 m.yr.

 # 32 Nepticulidae: *incertae sedis "Populus"* DT91
Insect damage; Donovan *et al*. 2014: fig. 13e–f
- Leaf mine—Host: Cercidiphyllaceae?: "*Populus*" *nebrascensis* Newberry–[1 ex.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: The Late Cretaceous host is not *Populus* (hence its placement in quotes) and likely belongs to a family far removed from the Salicaceae. The host probably is or probably closely related to *Cercidiphyllum* (Cercidiphyllaceae). We believe this leaf-mine record likely to be nepticulid.

 # 33 Nepticulidae: *incertae sedis Browniea* DT91
Insect damage; Donovan *et al*. 2014: fig. 14e
- Leaf mine—Host: Cornaceae: *Browniea serrata* (Newberry) Manchester & Hickey–[1 ex.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.
- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: The host family Cornaceae (including Nyssaceae, to which *Browniea* was originally assigned) is used as a host by some *Ectoedemia* in North-America (Doorenweerd *et al*. 2015) and undescribed *Acalyptris* species in Asia (data EJvN). However, the mines of this record are different from the mines of extant Cornaceae-feeding species. Nonetheless, these mines likely belong to Nepticulidae.

 # 34 Nepticulidae: *incertae sedis Dicot* morphotype 1 DT36, DT91
Insect damage; Donovan *et al*. 2014: fig. 15a–c
- Leaf mine—Host: Dicot leaf morphotype 1—[2 exx.] Coll: USNM (including USNM 560118; USNM 560119; USNM 560120; USNM 560113; USNM 498156; YPM 65939A)
- Loc: USA: Mexican Hat, eastern Montana, Custer Co.

- Stratum: Fort Union Fm.; Danian Stage, early Paleocene
- Remarks: A blotch mine (DT36) and gallery mine (DT91) likely represent two different species. The detailed image of the DT91 mine in figure 15 shows what seems to be an exquisitely preserved egg-capsule. We believe it likely is Nepticulidae.

 # 35 Nepticulidae: *incertae sedis Cercidiphyllum* DT41
Leaf mines; Donovan *et al*. 2014: fig. 10h–k
- Leaf mine—Host: Cercidiphyllaceae: *Cercidiphyllum genetrix* (Newberry) Hickey—[1 ex.] Coll: USNM (USNM 560151, USNM 560152, USNM 560153, USNM 560154)
- Loc: USA: Wyoming, Haz-Mat and Skeleton Coast sites
- Stratum: Fort Union Fm.; Danian Stage, late Paleocene
- Remarks: DT41, see comments under record # 31.

 # 36 *Stigmellites gossi* Jarzembowski, 1989
Mine type 1; Crane & Jarzembowski 1980: 632, figs. 6, 8
Stigmellites gossi Jarzembowski, 1989: 448
Stigmella sp.; Skalski 1990a
Stigmellites gossi; Sohn *et al*. 2012: 24
- Leaf mine—Host: Dicot—[2 exx.] Coll: BMNH (HT: In.64547; PT: In.64548)
- Loc: United Kingdom: Southern England, Berkshire, Newbury, Cold Ash
- Stratum: Reading Fm.; Thanetian Stage, late Paleocene
- Remarks: Crane and Jarzembowski (1980) stated that this mine is similar to an unidentified species of *Stigmella* on *Quercus cerris* L.. The fossil mine was compared with North American *Stigmella pomivorella* (Packard) and *Bucculatrix pomifoliella* (Clemens). We think that the assignment to Nepticulidae is probably correct.

 # 37 *Stigmellites centennis* Jarzembowski, 1989
Mine type 2; Crane & Jarzembowski 1980: 633, fig. 4, 9.
Stigmellites centennis Jarzembowski, 1989: 448
Stigmella sp.; Skalski 1990a
Stigmellites centennis; Sohn *et al*. 2012: 24
- Leaf mine—Host: ?Fabaceae—[1 ex.] Coll: BMNH (HT: In.64549)
- Loc: United Kingdom: S England, Berkshire, Newbury, Cold Ash
- Stratum: Reading Fm.; Thanetian Stage, late Paleocene
- Remarks: We do not find the illustrations sufficiently convincing that we can corroborate Jarzembowksi's suggestion that this specimen resembles modern *Stigmella hemargyrella* (Kollar). However, assignment to Nepticulidae certainly is possible.

 # 38 cf. Nepticulidae: *incertae sedis* multiple exemplars
"Healed wounds on leaf"; Brooks 1955: 4, 6, pl. 1: 5.
Nepticulidae; Opler 1973: 1321
Nepticulidae; Kozlov 1988: 30
Nepticulidae; Sohn *et al*. 2012: 26

- Leaf mine—Host: unknown angiosperm family: *Proteoides wilcoxensis* Berry—[multiple exx.] Coll: USNM
- Loc: USA: Tennessee, Henry Co., SW of Puryear
- Stratum: Claiborne Fm., Wilcox deposits; late Ypresian Stage, early Eocene
- Remarks: The combination of the references by Berry (1916) and Brooks (1955), as provided by Sohn *et al*. (2012), is incorrect. Brooks (1955) referred to Berry's (1916) plates as material that was different from the galleries that he described, and which he did not regard as leaf mines. According to Opler (1973), the specimen does represent a nepticulid, of which we are not completely convinced, but tentatively accept. In addition, the plant host *Proteoides* does not belong to the Proteaceae; its family assignment is unknown (Dilcher 1973).

 # 39 Nepticulidae: *incertae sedis* multiple species [unverifiable]
cf. Stigmella; Wilf *et al*. 2005: 8944
cf. Stigmella; Sohn *et al*. 2012: 24
- Leaf mine—Host: unidentified dicot—[multiple exx.] Coll: MPEF
- Loc: Argentina: Patagonia, Chubut, Laguna del Hunco
- Stratum: Tufolitas Laguna del Hunco Fm.; Ypresian Stage, early Eocene (Genise & Petrulevicius 2001).
- Remarks: The study counted 3599 instances of fossil leaf feeding; however, it was unspecified as to how many of these were nepticulid in affinity.

 # 40 Nepticulidae: *incertae sedis* DT37
leaf mines; Wilf *et al*. 2001: 6222, suppl. fig. 5b
mining DT37; Labandeira *et al*. 2007
- Leaf mine—Host: Fabaceae: "*Caesalpinia*"—[1 ex.] Coll: USNM (USNM 9623)
- Loc: USA: Utah, Uinta Basin
- Stratum: Green River Fm.; Lutetian Stage, middle Eocene
- Remarks: A photographic image of a fossil leaf mine of the type DT37, was provided as an example of a leaf mine. The same image is included is included in # 71. We find it likely that this specimen is a member of the Nepticulidae.

 # 41 Nepticulidae: *incertae sedis*
Stigmella; Labandeira 2002b: 45, figs. 4a–b
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: dicotyledonous angiosperm—[1 ex.] Coll: TBMM (57293a)
- Loc: USA: Washington, Whatcom Co., near Bellingham
- Stratum: Chuckanut Fm.; Lutetian Stage, middle Eocene
- Remarks: Judging from the image, the mine has the general impression of a nepticulid mine.

 # 42 Nepticulidae: *incertae sedis*
Stigmella; Labandeira 2002b: 45, figs. 4e–g
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: Rosaceae: *cf. Sorbus*—[1 ex.] Coll: TBMM (76477)
- Loc: USA: Washington State, Ferry Co., Republic

- Stratum: Klondike Mountain Fm.; Lutetian Stage, middle Eocene
- Remarks: The author stated that the fossil mine is particularly similar to those made by the extant *Sorbus*-feeding *Stigmella nylandriella* (Tengström) and *S. magdalenae* (Klimesch), that construct a thin, threadlike, central frass trail. The fossil mine could have been made by other genera as well; however, we suggest that it likely was made by Nepticulidae. *Sorbus* is an important host for extant Nepticulidae, with species of *Stigmella* and *Ectoedemia* that feed on this genus throughout the Palearctic, but Nepticulidae currently have not been found on this host in the Nearctic. However, it should be considered that there were close biogeographical connections between the Nearctic and the western Palearctic (via the North Atlantic Land Bridge) and between the Nearctic and eastern Asia (via Beringia) during the Paleocene and Eocene. Consequently, it comes as no surprise that the Nearctic Paleogene shares plant hosts and their insect herbivores with the western Palearctic and eastern Asia (Labandeira 2002b).

 # 43 Nepticulidae: *incertae sedis*
cf. *Stigmella*; Kinzelbach 1970: 94, 96, fig. 1
Order uncertain; Kozlov 1988: 54
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: Moraceae—[1 ex.] Coll: HLDG (Me7408)
- Loc: Germany: Hessen, S Frankfurt, near Darmstadt, Messel oil shale-layers
- Stratum: Messel Fm.; early Lutetian Stage, middle Eocene
- Remarks: Kinzelbach (1970) suggested "… mine shape matches the extant genus *Stigmella* …". The fossil is a compression between layers of compacted coal, from a drawing of the leaf and mine that was made. The mine has the general outline of a nepticulid mine. Several extant species of *Stigmella* feed on Moraceae, mostly on various *Ficus* species (Gustafsson 1985; Puplesis 1994; Vári 1963), but also include a species from Japan on *Morus* (Hirano 2010).

 # 44 Nepticulidae *incertae sedis* [unverifiable]
Stigmellites spp.; Jarzembowski 1995: 146
Stigmellites; Sohn *et al*. 2012: 25
- Leaf mine—Host: not mentioned—[multiple exx.] Coll: BMNH
- Loc: United Kingdom: Hampshire, East Dorset, Bournemouth
- Stratum: Branksome Sand Fm.; Lutetian Stage, middle Eocene (McElwaine, 1998)
- Remarks: Jarzembowski (1995) is a publication without illustrations, and represents a checklist of Paleogene insects from Dorset.

 # 45 Nepticulidae: *incertae sedis*
Nepticulidae; Stephenson & Scott 1992: 547, figs. 5: b, d, e, f, h, figs. 6: d, e
Nepticulidae; Lang *et al*. 1995: 159–162, 165–168, 170, figs. 3a, 3b, 3d, 3g, 3h, 4a–g, 4i–k, 4m, 4n, pl. 2: 2, 3, 7, 9, pl. 3: 1–3, 5, 6
?Nepticulidae; Labandeira 2002a: 49, 252, fig. 2.10i–j
Nepticulidae; Sohn *et al*. 2012: 23

- Leaf mine—Host: Angiosperms—[13 exx.] Coll: BMNH (V.45868; V.48524; V.48798; V.49808; V.49905; V.50089; V.50460; V.50622; V.50698; V.50731; V.50733; V.50904; V.50952)
- Loc: United Kingdom: Hampshire, East Dorset, Bournemouth
- Stratum: Branksome Sand Fm.; Lutetian Stage, middle Eocene (McElwaine, 1998).
- Remarks: The authors used analogies to recent leaf mines from the Hering collection (BMNH) to characterize the fossils. They do not identify any specimen beyond that of the family, but often mention similarities to extant species of *Stigmella* on a range of different host plants.

 # 46 cf. *Stigmellites messelensis* Straus, 1976
Worm or larva; Bornhardt 1975: 471
Stigmellites messelensis Straus 1976: 446
Stigmellites messelensis; Kozlov 1988: 32
Stigmellites messelensis; Skalski 1990a: 127
Stigmellites messelensis; Sohn *et al*. 2012: 25
- Leaf mine—Host: Dicot—[1 ex.] Coll: unspecified private
- Loc: Germany: Hessen, S Frankfurt, near Darmstadt, Messel oil-shale layers
- Stratum: Messel Fm.; early Lutetian Stage, middle Eocene
- Remarks: The figure of the mine is rather unclear. The mine could be nepticulid, but given its small size, a bucculatricid affiliation also is a possibility. As identification of this mine awaits further study, we tentatively leave the species assigned to *Stigmellites*.

 # 47 cf. Nepticulidae: *incertae sedis*
Nepticulidae Leaf Mine Form 1–2; Rozefelds 1988a: 2, figs. 2a–d
?Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: unknown dicot—[2 exx.] Coll: MVVA (NMVP183063, NMVP183064)
- Loc: Australia: Victoria, Alcoa Anglesea Coal Mine
- Stratum: Eastern View Fm.; Priabonian Stage, late Eocene
- Remarks: These mines resemble modern nepticulid mines, but the illustrations are inconclusive. The two different types of illustrated mines may represent two species.

 # 48 cf. *Roscidotoga* sp.
Nepticulidae Leaf Mine Form 3; Rozefelds 1988a: 2, figs. 2e–f
?Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: Elaeocarpaceae—[1 ex.] Coll: MVVA (NMVP183065)
- Loc: Australia: Victoria, Alcoa Anglesea Coal Mine
- Stratum: Eastern View Fm.; Priabonian Stage, late Eocene
- Remarks: The single mine resembles very much a modern nepticulid mine. The genus *Roscidotoga* (Fig. 6) is an extant genus endemic to Australian rainforests and is specialized on hosts of Oxalidales, of which two species feed on Elaeocarpaceae (van Nieukerken *et al*. 2011a). The mine morphology fits that of

Roscidotoga, but also several other nepticulid genera, including *Stigmella*. With this in mind, we assign the fossil mine with some doubt to *Roscidotoga*.

 # 49 Nepticulidae: *incertae sedis*
Nepticulidae Leaf Mine Form 4–5; Rozefelds 1988a: 4, figs. 3a–c
Nepticulidae; Labandeira 2002a: 49, 252, fig. 2.10k–l
?Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: Lauraceae—[5 exx.] Coll: MVVA (NMVP183063, NMVP183064, NMVP183065)
- Loc: Australia: Victoria, Alcoa Anglesea Coal Mine
- Stratum: Eastern View Fm.; Priabonian Stage, late Eocene
- Remarks: Rozefelds (1988a) considers both mine types on the same leaf as possibly belonging to two species. Rather, we think that these mine types belong to the same species. The left mine possibly is less developed. We doubt the presence of the large blotch as shown in the reconstruction (fig. 3D). If the host identification is correct, this is another example of an extinct host association, as we do not know of a single, extant nepticulid feeding on Lauraceae.

 # 50 *Stigmellites fossilis* (Heyden, 1862)
Nepticula fossilis Heyden, 1862: 77, pl. 10: 2
unidentified, may be dipterous; Opler 1973: 1321
Stigmellites fossilis; Kozlov 1988: 31
Stigmellites fossilis; Sohn *et al*. 2012: 24
- Leaf mine—Host: Juglandaceae: *Juglans acuminata* A. Braun—[1 ex.] Coll: originally collection of the Senckenberg Nature-Study Society, Frankfurt [not found, probably lost]
- Loc: Germany: Hessen, Bad Salzhausen [the locality in Sohn *et al*. (2012) is incorrect]
- Stratum: [unknown formation]; Chattian Stage, late Oligocene
- Remarks: Although no extant European Nepticulidae feed on Juglandaceae, this family is an important host family for Nepticulidae in North America and Asia that includes several species of *Stigmella* and *Ectoedemia* as leaf miners. The previously known occurrences of these genera would be expected on this European host. The illustrated mine clearly resembles extant nepticulid mines. The suggestion that this mine might belong to Diptera, indicated by Opler (1973,) is implausible, as Juglandaceae seems to be completely absent from the host record for extant leafmining Diptera, and certainly for Agromyzidae (Spencer 1990),. The Agromyzidae is the only dipteran leaf-mining clade known from the fossil record (Winkler *et al*. 2010).

 # 51 *Stigmellites almeidae* (Martins-Neto, 1989) comb. nov.
Nepticula? *almeidae* Martins-Neto, 1989: 381, pl. 1c, Fig. 5a
cf. Stigmella almeidae; Sohn *et al*. 2012: 22
- Leaf mine—Host: Symplocaceae: *cf. Symplocos* sp. A—[1 ex.] Coll: IGUSP (HT: GP/1T-1644)

- Loc: Brazil: São Paulo, Tremembé, along the road that connects Rodovia Presidente Dutra with Campos do Jordão
- Stratum: Tremembé Fm.; Chattian–Aquitanian Stages, late Oligocene–early Miocene boundary interval
- Remarks: The species was described initially as *Nepticula*, a junior synonym of *Stigmella*. We cannot reliably assign this fossil to *Stigmella*. Instead, we place it in the form-genus *Stigmellites*. From the images provided, it is difficult to judge if the specimen actually represents a nepticulid mine, although it is plausible. In eastern Asia (Taiwan), leaf mines of *Acalyptris* have been found on a species of *Symplocos* (EJvN, unpublished data).

 # 52 Nepticulidae: *incertae sedis*
leaf mine; Peñalver 1997: 150, fig. 1
Nepticulidae; Peñalver & Delclòs 2004: 82, fig. 6: 2. pl. 2: 2
Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: Lauraceae: *Laurophyllum*—[1 ex.] Coll: MCNV (MPV RIB-242)
- Loc: Spain: Castellón Prov., near Ribesalbes, "La Rinconada" site
- Stratum: bituminous rhythmites; Aquitanian Stage, early Miocene
- Remarks: The presence of multiple, independent mines on the same leaf occurs frequently in Nepticulidae, such as many *Stigmella* species that are known to feed on certain (sub)tropical plants. (The other alleged mine depicted by these authors is on the host *Celtis* sp. In our opinion, this feature is not a mine, but rather physical damage, where the leaf has been broken along its veinlets.)

 # 53 Nepticulidae: *incertae sedis* DT41
cf. Stigmella; Knor *et al*. 2012: 104, fig 2j
- Leaf mine—Host: Schisandraceae: *?Schisandra*—[multiple exemplars] Coll: NMPC, Bílina Mine Enterprise collections and Senckenberg Naturhistorische Sammlungen Dresden
- Loc: Czech Republic, North Bohemia
- Stratum: Most Fm.; Burdigalian Stage, early Miocene
- Remarks: Knor *et al*. (2012) identified material using the Labandeira *et al*. (2007) guide (# 71), in which one fossil is depicted as a *Stigmella*-like mine. Over 50 fossils are reported as leaf mines at this site, but it is not clear which of those likely are nepticulid or on which hosts the mines occur. A leaf mine classified as DT41 is depicted, which looks plausible for assignment to Nepticulidae. The damage type 41 exemplar shown in Labandeira *et al*. (2007), however, seems less likely a nepticulid, because its long length and pattern of vein crossing is more suggestive of a lyonetiid. If the host record is correct, it is the most basal angiosperm host record for Nepticulidae (see Table 1)

 # 54 *Stigmella* sp.
Stigmella; Liebhold *et al*. 1982: 456, figs. 1–2
Nepticulidae or perhaps Diptera; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23

- Leaf mine—Host: Berberidaceae: *Mahonia reticulata* (MacGinitie) Brown—[1 ex.] Coll: UCMP (8437)
- Loc: USA: Trapper Creek, Southern Idaho
- Stratum: Trapper Creek Fm.; early Langhian Stage, middle Miocene
- Remarks: The authors note the resemblance of the fossil mine to herbarium mines of undescribed *Stigmella* species on "*Mahonia" pinnata* (now *Berberis aquifolium*), also recorded on *Berberis nervosa* and *B. repens* hosts. We illustrate the mine of the extant species here (Fig. 4). No other Nepticulidae are known to feed on Berberidaceae hosts. We think it very likely that the fossil species is closely related to the extant species or possibly is a direct ancestor. The extant species is related to *Stigmella quercipulchella* (Chambers).

 # 55 *Stigmella* sp. [unverifiable]
Nepticula; Opler 1973: 1321
Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Fagaceae: *Quercus hannibali* Dorf—[1 ex.] Coll: UCMP
- Loc: USA: Nevada, Churchill Co., Buffalo Canyon
- Stratum: Buffalo Canyon Fm.; Langhian Stage, middle Miocene
- Remarks: *Quercus hannibali* is the fossil equivalent or very closely related species to extant *Q. chrysolepis* of California. The mine is not depicted. See # 61.

 # 56 Nepticulidae: *incertae sedis*
cf Nepticulidae; Lewis 1969: 1210, fig. 1
Caloptilia; Opler 1973: 1322
eriocraniid; Opler 1974b: 74
Nepticulidae; Kozlov 1988: 30
Stigmella sp.; Skalski 1990a: 127
Nepticula; Lewis *et al*. 1990: 7, fig. 3c
- Leaf mine—Host: Possibly Fagaceae: oak leaf.—[1 ex.] Coll: not stated
- Loc: USA: eastern Washington
- Stratum: Latah Fm.; Aquitanian–Serravallian Stages; early–middle Miocene
- Remarks: Opler (1973) identified the specimen as a *Caloptilia* mine, but later Opler (1974)—probably erroneously—cites it as an eriocraniid mine. There is no collection data provided, but on the figure there is "B16" noted on the slab. In 1990, Lewis *et al*. regarded the mine as *Nepticula*—the junior synonym of *Stigmella*. It is notable that the fossil of a small, apparently oak leaf that seemingly has marginal feeding also has a leaf mine whose trajectory follows the inner margin of the external feeding damage. This leaf mine very likely represents a nepticulid, probably *Stigmella*, but identification as an early phase of a *Caloptilia* mine cannot be excluded.

 # 57 *Stigmella* sp.
Nepticula; Opler 1973: 1321, fig. 1a
Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23

- Leaf mine—Host: Fagaceae: cf. *Quercus virginiana* Mill.—[multiple exx.] Coll: UCMP
- Loc: USA: California, San Luis Obispo Co., Temblor Range
- Stratum: Temblor Fm.; ?Serravallian Stage, middle Miocene
- Remarks: The author mentioned that the fossil leaf mines essentially are identical to those created on *Quercus* by extant Californian leaf miners. The putative host *Q. virginiana* is an extant species of the southeastern United States. Although the depicted mine certainly appears plausibly nepticulid, any comparison with extant species from the same region should be done with care. There is a large variety of leafminer species feeding on *Quercus* worldwide, but particularly in western North America where the host genus is exceptionally diverse. Several *Stigmella* species, of which only one of which is formally named, are known to feed on Californian oaks. Present-day *Ectoedemia* mines have not been observed (coll. Essig Museum, coll. D.L. Wagner). Therefore, we consider it very likely that the fossil nepticulid mines on *Quercus*, as cited by Opler (1973), indeed belong to *Stigmella*.

 # 58 Nepticulidae: *incertae sedis* [unverifiable]
?*Nepticula*; Opler 1973: 1321
Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Fagaceae*: Quercus pseudolyrata* Lesquereux—Coll: UCMP
- Loc: USA: Oregon, Columbia Plateau, Blue Mountains, Stinking Water
- Stratum: Mascall Fm.; Serravallian Stage, middle Miocene
- Remarks: Not illustrated; the identity and affiliation remains uncertain

 # 59 Nepticulidae: *incertae sedis* ?multiple species [unverifiable]
Nepticulidae; Donner & Wilkinson 1989: 9
Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—Host: not mentioned—[multiple exx.] Coll: GDVU
- Loc: New Zealand
- Stratum: [unknown formation]; middle Miocene
- Remarks: The authors mentioned fossil mines occurring in New Zealand that are similar to extant *Stigmella*. However, it is unclear if the fossil leaf-mine material that was referred to still exists and where it is deposited.

 # 60 Nepticulidae: *incertae sedis* [unverifiable]
Nepticulidae; Donner & Wilkinson 1989: 9
Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—[2 exx.] Coll: not stated
- Loc: North America
- Stratum: [unknown formation]; Serravallian Stage, middle Miocene
- Remarks: Mentioned in the same paragraph as records # 25 and # 59. No further details.

 # 61 *Stigmella* sp. [unverifiable]
Nepticula; Opler 1973: 1321
Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23
• Leaf mine—Host: Fagaceae: *Quercus hannibali* Dorf—[1 ex.] Coll: UCMP
• Loc: USA: Nevada, Nye Co., Cedar Mountains, Upper Goldyke
• Stratum: Esmeralda Fm.; Serravallian Stage, middle Miocene
• Remarks: *Quercus hannibali* is the fossil equivalent of extant *Q. chrysolepis*. The leaf mine mentioned in this record has not been illustrated in any publication. There is an undescribed species of *Stigmella* on *Q. chrysolepis*. See # 55.

 # 62 *Stigmella* sp.
Nepticula cf. *variella*; Opler 1973: 1322
Nepticula cf. *variella*; Opler 1974a: 74, pl. 7
Nepticulidae; Kozlov 1988: 30
Stigmella cf. *variella*; Skalski 1990a: 127
cf. Stigmella; Sohn *et al*. 2012: 23
• Leaf mine—Host: Fagaceae: *Quercus wislizenoides* Axelrod—[1 ex.] Coll: UCMP
• Loc: USA: Nevada, Storey Co., Dead Camel Range
• Stratum: Chloropagus Fm.; Serravallian Stage, middle Miocene
• Remarks: The author stated that the leaf mine "is indistinguishable from mines made by living *Nepticula variella* Braun." ("*Nepticula*" is a junior synonym of *Stigmella*). *Stigmella variella* feeds on the evergreen oaks, *Quercus agrifolia* Née (coastal live oak) and *Q. wislizeni* A. DC. (interior live oak) in California; the latter is the extant equivalent of the fossil *Q. wislizenoides* (Opler 1973).

 # 63 *Stigmella* sp. [unverifiable]
Nepticula; Opler 1973: 1321 Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23
• Leaf mine—Host: Fagaceae: *Lithocarpus* or *Quercus simulata* Knowlton—[2 exx.] Coll: UCMP
• Loc: USA: Idaho, Thorn Creek
• Stratum: Payette Fm.; Serravallian–Tortonian Stages, middle to late Miocene
• Remarks: Not illustrated. See # 57. The table in Opler (1973) lists both *Lithocarpus* and *Quercus simulata* in the same row for a single host record. We find it likely that the host of this leaf mine is one of the two, but the identity was impossible to determine with high reliability. Axelrod (1995) also indicates that *Q. simulata* and *Lithocarpus* are regularly confused, and suggests that a number of specimens identified as *Q. simulata* from the Miocene outside the Purple Mountain flora represent actually *Lithocarpus*, which may be tentatively recognized by "coarse secondaries and often with a coarsely serrate margin".

 # 64 *Stigmella* sp.
Stigmella; Kuroko 1987: 119, fig. 1
Stigmella; Kuroko 1990: 1, fig. 1
Stigmella sp.; Skalski 1990a: 127
cf. Stigmella; Sohn *et al*. 2012: 22
- Leaf mine—Host: Betulaceae: cf. *Betula grossa* Siebold & Zucc.—[1 ex.] Coll: Collection of Tachu Koshimizu
- Loc: Japan: central Honshu, at the border between Nagano and Gumma Prefectures
- Stratum: Kabutoiwa Plant Bed; Tortonian–Messinian Stages, late Miocene
- Remarks: A trace of the egg case is recognized in this specimen as a dark, brownish, elliptical spot. The host, *Betula grossa*, is a common extant tree, with a rich nepticulid fauna in Japan, including populations of this host in Nagano Prefecture where this fossil was found. On the basis of the extant fauna and the shape of the mine, we can associate this mine with the genus *Stigmella*, but not with an extant species.

 # 65 *Stigmella* sp. [unverifiable]
Nepticula; Opler 1973: 1321
Nepticulidae; Kozlov 1988: 30
cf. Stigmella; Sohn *et al*. 2012: 23
- Leaf mine—Host: Fagaceae: *Quercus hannibali* Dorf—[1 ex.] Coll: UCMP
- Loc: USA: Nevada, Lyon Co., near Yerington
- Stratum: Aldrich Station Fm.; Zanclean Stage, early Pliocene
- Remarks: *Quercus hannibali* is the fossil equivalent of *Q. chrysolepis*. The mine is not depicted by a photographic image or line drawing. See # 57.

 # 66 *Stigmellites zelkovae* Straus, 1977
Stigmellites zelkovae; Straus 1977: 61, fig. 14
Stigmellites zelkovae; Skalski 1990a: 127
Stigmellites zelkovae; Sohn *et al*. 2012: 25
- Leaf mine—Host: Ulmaceae: *Zelkova*—[1 ex.] Coll: GPUG (HT: 23973)
- Loc: Germany: Niedersachsen, Willershausen am Harz
- Stratum: "Willershausen Shale"; Piacenzian Stage, late Pliocene (Brauckmann *et al*. 2001).
- Remarks: Straus (1977) attributed this fossil to *Stigmellites* because of its similarity to extant nepticulid leaf mines. A drawing of the mine was included, but there were no photographic images. Judging from the drawing, it could be a partial nepticulid mine, or a bucculatricid mine. There are extant *Stigmella* species known from *Zelkova* in Asia, such as *S. zelkoviella* Kemperman & Wilkinson from Japan (Kemperman *et al*. 1985), and an undescribed species from the Caucasus Region (Skala 1941). There is also an extant Bucculatricidae species known to feed on *Zelkova*: *Bucculatrix serratella* Kobayashi *et al*., in Japan (Kobayashi *et al*. 2010).

 # 67 *Stigmellites carpiniorientalis* Straus, 1977 [unverifiable]
Stigmellites carpini-orientalis Straus, 1977: 60, fig. 80, 62
Stigmellites carpini-orientalis; Skalski 1990a: 127
Stigmellites carpiniorientalis; Sohn *et al*. 2012: 24
Leaf mine—Host: Betulaceae: *Carpinus orientalis* Mill.—[2 exx.] Coll: GPUG (HT: 22763; PT: 22134)
- Loc: Germany: Niedersachsen, Willershausen am Harz
- Stratum: "Willershausen Shale"; Piacenzian Stage, late Pliocene (Brauckmann *et al*. 2001)
- Remarks: The host is an extant species. The images are of insufficient quality to re-evaluate the mines. We tentatively leave this occurrence as *Stigmellites*. Currently, *Stigmella microtheriella* (Stainton) and *S. johanssonella* A. & Z. Laštuvka are known to feed on *Carpinus orientalis*. However, eastern Palearctic extant species of *Ectoedemia* also feed on *Carpinus*. We cannot exclude that this genus occurred on the same host in Europe during the Pliocene.

 # 68 cf. *Stigmella ulmivora* Fologne, 1860
Stigmella ulmivora; Kernbach 1967: 106 fig. 5
Stigmella ulmivora; Straus 1977: 61, fig. 12
Stigmella ulmivora; Kozlov 1988: 30
Stigmella ulmivora; Skalski 1990a: 127
cf. Stigmella ulmivora; Brauckmann *et al*. 2001: 33
cf. Stigmella ulmivora; Sohn *et al*. 2012: 22
- Leaf mine—Host: not stated—[1 ex.] Coll: GPUG (596–4–9111)
- Loc: Germany: Niedersachsen, Willershausen am Harz
- Stratum: "Willershausen Shale"; Piacenzian Stage, late Pliocene (Brauckmann *et al*. 2001)
- Remarks: According to Kernbach, this mine was identified by Hering as *S. ulmivora*, but he did not mention the affiliation of the host. Considering the identification of the mine as *S. ulmivora*, the host is likely to be Ulmaceae. The base of the illustrated leaf host has not been preserved, which would be required for identification of the host as ulmaceous. This leaf mine represents the only fossil assigned to an extant species, which we believe is plausible, especially considering the recent age of the fossil. However, the totality of evidence is meagre. Moreover, even extant mines on *Ulmus* cannot be identified with certainty as belonging to either *S. ulmivora* or *S. ulmiphaga* (Preissecker) in the region where both occur (e.g. Laštůvka & Laštůvka 1997). We tentatively attribute this specimen to *Stigmella ulmivora*.

 # 69 *Stigmellites heringi* Kernbach, 1967 [unverifiable]
Stigmellites heringi Kernbach, 1967: 104, fig. 3
Stigmellites heringi; Straus 1977
Stigmellites heringi; Kozlov 1988: 30
Stigmellites heringi: Skalski 1990a: 127
Lepidoptera Suborder uncertain; Carpenter 1992: 380
Family uncertain; Brauckmann *et al*. 2001: 33
Stigmellites heringi; Sohn *et al*. 2012: 24

- Leaf mine—Host: Berberidaceae: *Berberis*—[1 ex.] Coll: GPUG (HT: 596–2–11137)
- Loc: Germany: Niedersachsen, Willershausen am Harz
- Stratum: "Willershausen Shale"; Piacenzian Stage, late Pliocene (Brauckmann *et al*. 2001)
- Remarks: This is the type species of *Stigmellites*. The form of the mine is impossible to judge from the published illustration. Only a single North American species, as yet unnamed, is known for feeding on modern Berberidaceae (see # 38). Brauckmann *et al*. (2001) observed incorrectly that Kernbach's species and generic descriptions are nomenclatorially invalid.

 # 70 *Stigmellites pliotityrella* Kernbach, 1967
Stigmella pliotityrella Kernbach, 1967: 106, fig. 4
Stigmellites pliotityrella; Kozlov 1988: 32
Stigmellites pliotityrella; Skalski 1990a: 127
Family uncertain; Brauckmann *et al*. 2001: 33
Stigmellites pliotityrellus; Sohn *et al*. 2012: 25
- Leaf mine—Host: Fagaceae: *Fagus sylvatica* L.—[1 ex.] Coll: GPUG (HT: 596–3–3050)
- Loc: Germany: Niedersachsen, Willershausen am Harz
- Stratum: "Willershausen Shale"; Piacenzian Stage, late Pliocene (Brauckmann *et al*. 2001)
- Family uncertain; Brauckmann *et al*. 2001: 33
- Remarks: See under species # 69 for the validity of Kernbach's names. Of the two extant European species of Nepticulidae feeding on *Fagus*, the mine resembles more *Stigmella hemargyrella* (Kollar) than that of *S. tityrella* (Stainton). However, we cannot place this mine firmly into *Stigmella*, since *Ectoedemia* species with relatively similar mines also are known to feed on *Fagus* (in Japan) and could have become extinct in Europe.

 # 71 Nepticulidae: *incertae sedis* multiple species (DT's)
Leaf mines; Labandeira *et al*. 2007
- Leaf mine—Host: Various angiosperm hosts—[multiple exx.] Coll: Different collections
- Loc: Numerous localities
- Stratum: Late Cretaceous, Paleogene and Neogene
- Remarks: *The Guide to Insect (and other) Damage Types on Compressed Plant Fossils* (Version 3.0) (Labandeira *et al*. 2007) documents a variety of damage types (DT's), of which several are very similar to extant Nepticulidae. In our opinion, these include DT37 (exemplar also shown in Wilf *et al*. 2001 as supplementary figure 5), DT40, DT43 (also shown in Labandeira *et al*. 2002b), DT45, DT59 (also shown in Labandeira *et al*. 2002b fig 1), DT65, DT90, DT91, DT92, DT93, DT104 and DT105. These exemplars originate from different time intervals, formations, world regions, habitats and host plants. This volume currently is being updated and is scheduled to be published in book form (Version 4.0) in a few years.

Removed from Nepticulidae

 # 72 reference to non-existing fossil
nepticulid; Grimaldi & Engel 2005: 572
putative nepticulids; Fischer 2013: 85
- Adult in amber—[1 ex.] Coll: not stated
- Loc: not stated
- Stratum: [unknown formation]; likely Santonian Stage, Late Cretaceous
- Remarks: This record potentially is the oldest nepticulid amber fossil, but offers a puzzling case. In Grimaldi & Engel (2005) it is mentioned as "a probable adult [nepticulid] in late Cretaceous Siberian Amber", for which the authors refer to the work by Skalski (pers. comm. from Prof. Engel). In reviewing Skalski's work however, including an overview of all the fossils known to the author in 1990, there is no reference to Siberian Amber nepticuloid fossils.

 # 73 Adeloidea
?Nepticulidae; Kristensen & Skalski 1998, 1998: p. 18
putative nepticulids; Fischer 2013: 85
- Adult in amber—[1 ex.] Coll: Ottawa
- Loc: not stated
- Stratum: Foremost Fm.; Campanian Stage, Late Cretaceous
- Remarks: This specimen could be the oldest nepticulid amber fossil, representing an age about twice that of Baltic Amber. However, we believe this specimen is more likely a heliozelid, or at least an adeloid in affinity. The visible veins course to the wing margin, whereas in Nepticulidae veins usually become obsolete before the margin, and the visible valvae of the male genitalia show a structure resembling a stalked pectinifer, characteristic of Adeloidea. The small size and venation suggest it could be heliozelid or closely related to that family. This fossil will be studied with x-ray techniques for further details, which will be published elsewhere. In any case, we remove the specimen herein from the nepticuloid fossil record. The fossil is dated at 72 Ma in manuscript texts that Kristensen *et al*. worked on for the publication of this fossil. [Examined by EJvN].

 # 74 Coleoptera
?Nepticulidae; Rozefelds 1988b: 77, fig. 2
?Nepticulidae; Labandeira *et al*. 1994: 12281
?Nepticulidae; Labandeira 1998: fig 2A
?Nepticulidae; Zherikhin 2002: 320
?Nepticulidae; Sohn *et al*. 2012: 27
- Leaf mine—Host: Umkomasiaceae: *Pachypteris crassa* Townrow—[1 ex.] Coll: QMSB (QMF15346)
- Loc: Australia: North Queensland, Cape York Peninsula, Cape Melville, Clack Island
- Stratum: Battle Camp Fm.; Tithonian–Berriasian Stages, Late Jurassic–Early Cretaceous boundary

• Remarks: This fossil is approximately 38 my older than any other fossil reliably assigned to Nepticulidae. The host is a corystosperm seed fern. The fossil pinnules exhibit five visible mines, of which one is poorly preserved. The mines however do not increase in width as is typical of the Nepticulidae, and are rather long, without clear frass visible. We find it much more likely that this mine is coleopteran rather than lepidopteran (see also Ding *et al*. 2014).

 # 75 Gracillariidae
Tinea araliae Fritsch, 1882: 6, pl. 2: 7
Eriocranioidea; Zherikhin 1978: 74
?Stigmellites araliae; Kozlov 1988: 30
Stigmellites araliae; Skalski 1990a: 127
Stigmellites araliae; Sohn *et al*. 2012: 24
• Leaf mine—Host: Araliaceae—[1 ex.] Coll: not stated

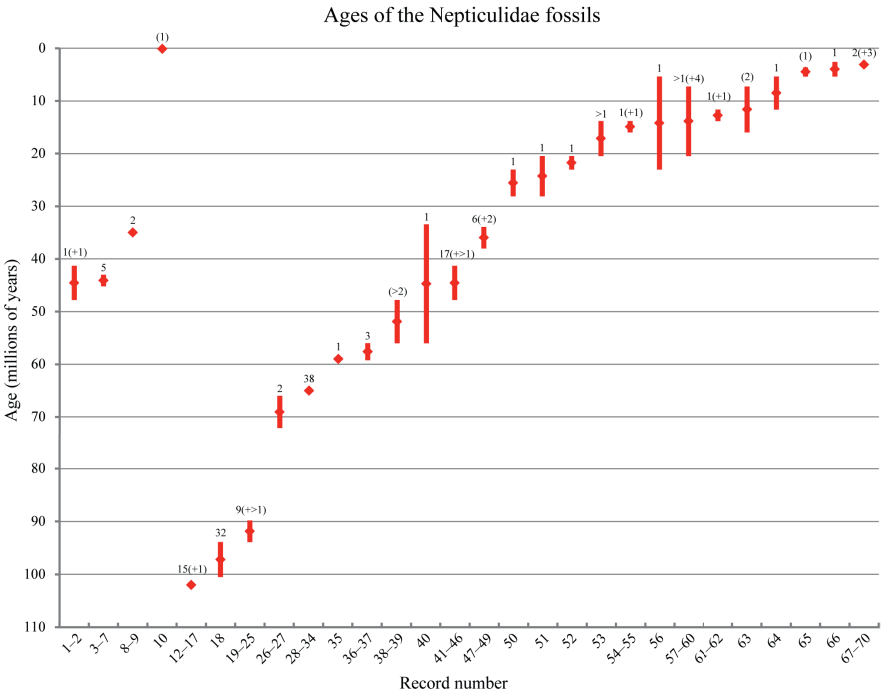


Fig. 21. Geochronologically ranked age intervals corresponding to literature-mentioned ages for the 69 fossil records of Nepticuloidea that have documented geochronologic ages. Entries 1–10, at the upper-left, represent body-fossil occurrences; entries 12–70, forming the linear series from lower-left to upper-right, are leaf miners. Fossil record numbers refer to those in the checklist. The number of exemplars involved for each datapoint is indicated above the range, numbers between parentheses indicate unverified or uncertain identifications. The ">" sign indicates that there was no exact number of exemplars provided. For cases where a subjective indication was given within an epoch, such as Late Priabonian, this is shown spanning the entire Priabonian Stage. Stage-level boundaries are from Ogg *et al*.(2008).

- Loc: Czech Republic: Bohemia, Perucher-Schichten, Vyšerovic; Bohemia, Perucher-Schichten, Lipenz
- Stratum: Perucher Fm.; Cenomanian Stage, Late Cretaceous
- Remarks: This very schematic drawing does not show much else other than a long gallery mine. This specimen is more likely to be a gracillariid mine. Since no nepticulid is known to feed on extant Araliaceae, but several Gracillariidae do feed on this plant-host family, particularly the genus *Eumetriochroa* Kumata, with mines similar to *Phyllocnistis* (Kobayashi *et al*. 2011). We tentatively remove this species from Nepticulidae.

 # 76 Undetermined
Galleries; Berry 1916: 32, pl. 23: 3, pl. 31: 1, 3, pl. 38: 4, pl. 39, pl. 92.
Nepticulidae; Sohn *et al*. 2012: 26
- Leaf mine—[multiple exx.] Coll: not stated
- Loc: Various localities in the Mississippi Embayment.
- Stratum: not reported
- Remarks: This specimen was incorrectly put in the same record as # 38 by Sohn *et al*. (2012). Here, the record is excluded from Nepticulidae. In this one case it is unclear if the material represents leaf mines or other types of damage.

***# 77 Foliofossor cranei* Jarzembowski, 1989**
Mine type 3; Crane & Jarzembowski 1980: 663, fig 10
?Nepticulidae; Kozlov 1988: 30
Foliofossor cranei Jarzembowski, 1989: 448
Stigmella sp.; Skalski 1990a: 127
- Leaf mine—Host: Platanaceae: *Platanus schimperi* (Heer) Saporta & Marrion—Coll: BMNH HT (In. 64550)
- Loc: United Kingdom: Cold Ash, or Newbury, Berkshire (N 51° 22' W 1° 17').
- Stratum: Reading Fm.; late Paleocene, Priabonian Stage
- Remarks: Jarzembowski (1989) doubts the assignment of this leaf mine to a particular insect taxonomic order, and provides an ichnogenus name. Crane & Jarzembowski (1980) consider it to be a dipteran mine, but Kozlov (1988) assigns it to Nepticulidae. There is a certain resemblance of this mine to extant *Stigmellites tyshchenkoi* that also feeds on *Platanus*, but the fossil mine is significantly older by approximately 40 my. We believe that assignment to Nepticulidae is unlikely.

***# 78 Troponoma festunata* Krassilov, 2008**
cf. Stigmella; Krassilov 2008a: 265, fig. 3a, b
Troponoma festunata Krassilov, 2008b: 102, PL XXXVI figs. 1–3
- Leaf mine—Host: Family unknown: *Dewalquea gerofitica* (Dobruskina) Krassilov—[several exx. on one leaf] Coll: IEUH (HT IG1–1001)
- Loc: Israel: southern Negev, Gerofit
- Stratum: Ora Fm.; mid-Turonian Stage, Late Cretaceous
- Remarks: Krassilov (2008b) mentions that "similar mine configurations are known in *Stigmella*, Nepticulidae, although radially spreading festoons (sensu Hering 1951) may indicate a gracillariid miner similar in habit to the digitate

mines of *Parectopa robiniella*". In Krassilov (2008a) this record is referred to as a *Stigmella*-type track. We fail to recognize any nepticulid characteristics in these leaf mine fossils and rather doubt if they actually represent leaf mines.

***# 79 Troponoma curvitracta* Krassilov, 2008**

Troponoma curvitracta Krassilov, 2008b: 101, Pl. VI fig 4b, Pl. XXX fig. 1; Pl. XXXV figs. 1–5

- Leaf mine—Host: Family unknown: *Dewalquea gerofitica* (Dobruskina) Krassilov—[>4 exx.] Coll: IEUH (HT IG1–160)
- Loc: Israel: southern Negev, Gerofit
- Stratum: mid-Turonian Stage, Late Cretaceous
- Remarks: Krassilov (2008b) states that "The mine configurations of th[is] kind are sometimes produced by *Stigmella*. However, egg persistence on mines is a feature typically of coleopteronomes, and comparable looping mines are made by a weevil *Rhamphus pulicarius* (Herbst) on *Betula* (cf. Hering)". From the photographs it is difficult to judge if there is an egg-like structure visible, but if there is, this would be a character also typical for Nepticulidae. As in the case for record # 78, we do not recognize sufficiently relevant characters in these images to attribute this fossil to Nepticulidae.

An overview of the data

A summary of the age intervals of the 69 fossil Nepticulidae records for which age estimates are available is provided in Fig. 21. Records 1–11 are adult fossils, records 12–72 represent leaf mine records. From the late Early Cretaceous onwards, there are episodic occurrences of leaf-mine fossils attributed to Nepticulidae. Based on our checklist, the oldest nepticulid fossils are records #12–17, which were also mentioned 16 years ago by Kristensen and Skalski (1998) as the likely oldest records, currently dated at 102 Ma. Table 1 shows an overview of the variously identified plant-host families from which fossil nepticulid leaf mines have been recorded throughout geological record. Table 1 also provides the extant genera that occur as leaf miners on those fossil host-plant families.

Discussion

The checklist presented herein for fossil Nepticulidae contains 71 records, of which 55 were re-examined by us and the remaining records were judged for plausibility. The records often include multiple exemplars or taxa presented as a single record. The checklist spans publications from 1862 to 2014, a period of time during which the systematics of Nepticulidae changed considerably (van Nieukerken *et al*. 2011b; Hoare 2000; van Nieukerken 1986). The identifications in the checklist have been updated to match recent systematic results and insights. The analogy of many fossil leaf mines with the extant genus *Stigmella* is expected, as *Stigmella* is among the most species-rich genus with the most variable leaf mine types and varied spectrum

Table 1. The geochronologic distribution of host-plant families identified from fossil Nepticulidae leaf mines and the extant nepticulid genera that occur on those host plant families, arranged by systematic plant order (APG III 2009). The Stage level boundaries follow Ogg *et al*. (2008).

Higher Angiosperm group	Order	Host family	Record #'s
Angiosperms			18, 45, 15, 71
Pre-Magnoliid	Austrobaileyales	Schisandraceae	53
Magnoliids	Laurales		16, 13
Magnoliids	Laurales	Lauraceae	52, 49
Eudicots			46, 36, 41, 34, 14, 39, 47, 38
Eudicots	Ranunculales	Berberidaceae	54, 69
Eudicots	Proteales	Platanaceae	12, 20, 23, 28
Eudicots	Trochodendrales	Trochodendraceae	30
Core Eudicots	Saxifragales	Cercidiphyllaceae	19, 21, 22, 27, 31, 35
Rosids	Myrtales or Malpighiales		24
Rosids	Malpighiales	Salicaceae	32
Rosids	Rosales	Rhamnaceae	30
Fabids	Oxalidales	Elaeocarpaceae	48
Fabids	Fabales	Fabaceae	40
Fabids	Fabales	?Fabaceae	37
Fabids	Fagales	Betulaceae	67, 64
Fabids	Fagales	Fagaceae	56, 57, 70, 55, 61, 65, 58, 63, 62
Fabids	Fagales	Juglandaceae	50, 29
Fabids	Rosales	Moraceae	43
Fabids	Rosales	Ulmaceae	66, 68
Fabids	Rosales	Rosaceae	26, 42
Malvids	Sapindales		15, 17
Asterids	Cornales	Cornaceae	33
Asterids	Ericales	Symplocaceae	51

of hosts within the Lepidoptera. Our designations have been conservative; we assign only a few relatively recent leaf-mine fossils on known host plants to the modern genus *Stigmella* and a single Australian mine tentatively to the Australian endemic genus *Roscidotoga*. Leaf mines predominantly represent behavioural characters and, although generally recognized as useful for identification, have never been analysed within a phylogenetic context. Consequently, assignment of fossil leaf mines to lower-level extant groups is precarious. However, when there is use of a combination of characters, assignments to many groups often can be excluded and a fairly certain identification usually is feasible. Such reliably identified assignments of fossil leaf mines constitute an important addition of data,

Early Cretaceous 145.0–100.5 mya	**Late Cretaceous** 100.5–66.0 mya	**Paleocene** 66.0–56.0 mya	**Eocene** 56.0–33.9 mya	**Oligocene** 33.9–23.0 mya	**Miocene** 23.0–5.3 mya	**Pliocene** 5.3–2.6 mya	**Pleistocene** 2.6–0.01 mya	**Extant genera on host family**
+	+		+					
					+			no
+								no
			+		+			no
		+	+					
					+	+		*Stigmella*
+	+	+						*Acalyptris, Ectoedemia*
		+						No, but not searched
	+	+						No, but not searched
	+							*Stigmella, Ectoedemia* on Malpighiales, *Pectinivalva, Acalyptris* on Myrtales
		+						*Stigmella, Ectoedemia*
		+						*Stigmella, Acalyptris, Ectoedemia*
			+					*Roscidotoga*
			+					*Stigmella, Acalyptris, Trifurcula*
		+						*Stigmella, Acalyptris, Trifurcula*
					+	+		*Stigmella, Bohemannia, Ectoedemia*
					+	+		*Stigmella, Ectoedemia*
		+		+				*Stigmella, Ectoedemia*
			+					*Stigmella*
						+		*Stigmella, Ectoedemia*
	+		+					*Stigmella, Areticulata, Trifurcula, Bohemannia, Ectoedemia*
+								*Stigmella, Ectoedemia, Acalyptris, Trifurcula*
		+						*Acalyptris, Ectoedemia*
				+	+			*Acalyptris*

compared to the scarce inventory of adult fossils. As a result, fossil leaf-mine assignments allow for a more reliable perspective regarding the estimated age of the family.

Classifying fossil material

The classification of fossil adult specimens usually is straightforward. Extant species and groups have been erected based on (syn)apomorphic characters of the adults and using those characters that allow identification of fossil material to the lowest taxonomic rank, as provided by the available characters. There often is one or, at most, a few, exemplars that frequently result in each new record as a new

(ichno)species. For fossil insect leaf mines, the best approach for classifying the material frequently is less straightforward. This indecisiveness in assignment is attributable often to dozens to hundreds of specimens that have been examined, and assignment of species and erecting new species for all leaf-mined foliar assemblages becomes a daunting task. Historically, three different approaches linked to different purposes have been used to systematize leaf-mined foliar assemblages, discussed in Scott and Titchener (1999). They are 1), comparative and functional morphology; 2), comparative analogy; and 3), ichnotaxonomy. The Nepticulidae checklist provides examples from all three categories. Certain authors have generally favoured ichnotaxonomic procedures (Kozlov 1988; Krassilov 2008b), and among other authors there appears to be a historical transition from comparative analogy up to the mid-1990's (Opler 1973; Stephenson 1991), towards a comparative and functional morphological approach in more recent studies (Donovan *et al*. 2014; Knor *et al*. 2012).

Comparative and functional morphology is the most conservative approach and prevents unjustified assignment of fossils to extant groups. However, checklists such as the one presented here would be better facilitated if analogies to modern taxa are clearly stated in the primary literature. A single, complete overview of all possible herbivore candidates and their typical characteristics is not yet available (but see Ding *et al*. [2014] for an example from the Coleoptera). Hering's published work (Hering 1951; Hering 1957) has commonly been used as a conventional source on leaf-mine types and for analogies with modern taxa, but it has a strong European bias and is out-of-date. Especially in the subtropical and tropical regions, there are many aberrant and overlapping leaf-mine morphologies for various leaf-mining groups. The European leaf miner website with keys and descriptions to all leaf miners of Europe, replete with abundant images, can be a useful starting point to grasp the diversity of a broad spectrum of leaf-mining groups (Ellis 2014). However, as this site applies to Europe only, it has limitations to extralimital leaf-mine material. The web-documented guide for fossil leaf mine and other insect-caused fossil plant damage types (Labandeira *et al*. 2007) commonly has been used in recent studies, but does not provide links to analogous extant groups. (The next version of the *Damage Guide* will have links to modern analogous groups for each mining and non-mining damage type (DT), many of which will be new.) These links for lepidopteran and other types of distinctive fossil mines will follow a similar format such as those published for dipteran Agromyzidae (Winkler *et al*. 2010) and those presented here for lepidopteran Nepticulidae. Such an improvement hopefully will enable additional commentary regarding the essential characters available for reliable identifications. When publishing fossils in general, references to collections are invaluable in allowing for further study. When stating leaf-mine analogies, the scope of the reference material should be clearly indicated.

In cases where identification to family can be made, we advocate the use of an ichnogenus to describe species from fossil leaf mines. Each leaf-mine ichnogenus also would be represented by a damage-type synonym, or DT number, for the

assessment of herbivory (Labandeira *et al*. 2007), allowing for quantitative analyses (e.g. Labandeira *et al*. 2002; Wilf *et al*. 2005; Donovan *et al*. 2014). When compiling the checklist, it became clear that many fossils have been mentioned or depicted in multiple publications, for which, from our thorough study of the texts and images, it became necessary to link the publications to the same leaf-mine record. In the checklist we have resorted to numbering the records, but allowance for cross-publication links is the most sustainable option for assignment of a species name to a fossil.

It should also be noted that when the fossil cannot be linked to any extant family, the value of ichnotaxonomy quickly deteriorates and devolves to a parallel naming system. This is exemplified by the work of Krassilov and colleagues (Krassilov 2008b). From remarks in the figure captions of this work, we were able to notice that the authors did recognize analogies to extant Nepticulidae; fossils were described as species within genera without higher taxonomic ranking or an indication of affiliation. Consequently, such groups cannot be linked to any modern lineage. We believe that this approach fails to advance any of the purposes for studying nepticulid fossil herbivory, and a preferable approach would be to use either comparative and functional morphology, or alternatively and better yet, use an ichnotaxonomic system such as the damage-type system, with a goal toward recognizing characteristic analogies to better place the fossils in modern groups whenever possible. However, there is one significant exception. In the older part of the geological record where Nepticulidae are absent, such as the later Paleozoic to the mid Mesozoic, all to most of insect-mediated damage, including Triassic and Jurassic leaf mines, may lack links to modern herbivorous taxa, and instead represent extinct groups of herbivores. Under these conditions the best option is to use the parallel ichnotaxonomy of the DT system (Labandeira *et al*. 2007), in which a more functional and morphological perspective is used (e.g., Schachat *et al*. 2014).

Age of Nepticulidae

Although it is difficult to assign some of the leaf-mine fossils with complete certainty to Nepticulidae, it seems unlikely that the overwhelming majority of the records would not be assigned to Nepticulidae. In addition, it is highly probable that nepticulid leaf-mine fossils date to the late Albian of the Early Cretaceous, at 102 Ma (Fig. 21). Molecular dating of Lepidoptera phylogeny by Wahlberg *et al*. (2013) estimated the split between Nepticulidae and Opostegidae between 100 and 130 Ma (95% confidence interval). Their study used seven calibration points throughout Lepidoptera, and also included the nepticulid fossil record # 12. Accidentally, Wahlberg (pers. comm.), used record # 12 to calibrate the split between *Ectoedemia* and *Opostega* at 120 ± 10 Ma, whereas the actual age of the referenced material is estimated at 102 Ma. Without the nepticulid calibration point, and reliance instead only on the remaining six calibration points, the estimated molecular-phylogenetic age range of Nepticulidae is increased to 75–150 Ma. Wilf and Escapa (2014) provide a compelling demonstration of this phenomenon from the fossil records of several land-plant lineages.

Host plant relationships. Table 1 lists the host plants of fossil Nepticulidae that have been identified, according to the approximate taxonomic order of the Angiosperm Phylogeny Group (APG III 2009). After excluding the corystosperm seed fern fossil, which we judged to be non-nepticulid, only angiosperms evidently hosted nepticulids in the fossil record. Of these, six are non-eudicot (Laurales for records # 13, 16, 49, 52; Austrobaileyales for # 53); there are no extant Nepticulidae known to feed on these host orders. Angiosperms have been estimated based on molecularly dated phylogenies to have originated during the Triassic or Jurassic, at 193 Ma, although their empirical fossil record begins in the mid Early Cretaceous (Valanginian Stage), which is probably closer to the accurate date for the true origin of the angiosperms (Friis *et al*. 2011). The principal eudicot radiation began during the mid Early Cretaceous, and intensified to the mid Cretaceous (Magallón *et al*. 2013). Nepticulidae fossils from the Cretaceous have been found on Laurales, Proteales, Saxifragales, and several plant species for which an ordinal placement is uncertain. Nevertheless, these basal angiosperm host lineages indicate that they were the dominant plant groups during the mid Cretaceous, as evidenced by their general abundance in the mid Early Cretaceous to Early Paleogene fossil record (Graham 1999). Platanaceae in particular were a common and diverse group during the Late Cretaceous and Paleogene (Johnson 1996; Graham 1999; Friis *et al*. 2011), and formed a major plant-host family for insect herbivores (Labandeira *et al*. 1994, Labandeira, 1998). This perhaps explains the common occurrence of Nepticulidae leaf mines from the Dakota Formation (102 Ma, records #12–18), which likely represent multiple species and are most abundant on *Platanus* and other Platanaceae. Dakota-age leaf mines also are found on several other, unrelated, non-dicot hosts. Leaf-mine fossils approximately 10 my younger, from Kazakhstan, are also diverse (# 19–23), and include *Platanus* and related genera as hosts (# 20, 23). Additionally important associations include the leaf genera *Trochodendroides* (not to be confused with *Trochodendron* in the Trochodendraceae) (records # 19, 21, 22), and *Cercidiphyllum* in the Cercidiphyllaceae (records # 19, 21, 22, 27, 31, 35), associated with fruits of the genera *Nyssidium* and *Joffrea* (Friis *et al*. 2011). The Cercidiphyllaceae forms a basal clade of the Saxifragales, occurring in floras during the Late Cretaceous and into the Paleogene from Eurasia and North-America. Modern Cercidiphyllaceae consist of two relict species in north-eastern Asia (Stevens 2013), and currently are unknown as host plants for extant Nepticuloidea. However, the modern record also reflects an absence of serious sampling. A few modern nepticulids do occur on other related host-plant lineages of Saxifragales, such as the families Altingiaceae and Hamamelidaceae. The host-plant records from combined Cretaceous occurrences suggest that Nepticulidae at that time were widespread and already had diversified onto a variety of host families across several major angiosperm lineages.

Given the time interval when likely Nepticulidae are first encountered in the fossil record, and taking into account the varied biogeographic dispersal patterns and physiognomic forms of their plant hosts, we find that an Albian to Aptian origin of the Nepticulidae is most likely, ranging from 125 to 100 m.yr. ago. This period of time also is the same interval during which dominant, woody angiosperm lineages

diversified. We expect that much of the deep-time diversity of nepticulid taxa reflected in the fossil record is extinct, as are their plant hosts. One major cause of the extirpation of nepticulid leaf miners was the mass extinction event at the end of the Cretaceous (Labandeira *et al*. 2002), the Cretaceous–Paleogene boundary, which disproportionately affected some plant hosts over others (Johnson 2002). In addition, there is increasing evidence that the diversity of specialized insect herbivores was also heavily and negatively affected (Donovan *et al*. 2014; Heikkilä *et al*. 2012; Labandeira *et al*. 2002). By additional examination of the phylogeny and diversification of Nepticulidae and combining such studies with insights from the fossil record, a clearer picture of the evolutionary history of the family should emerge.

Conclusions

If we would rely on amber body fossils alone, the oldest Nepticulidae representatives would lie between 43 and 45.2 Ma, more than 60 my younger than the estimated age of the family if fossil leaf mines are included. When molecular dating is applied to a phylogeny of Lepidoptera, this young date is extended by a factor of 2.7, to 120 Ma. The identifications of leaf-mine fossils may be less precise, but because of their large numbers and their representation in the older fossil record, they nevertheless represent an undeniably important source of data, especially when classified in a way that allows them, when possible, to be assigned to extant groups. Nepticulidae currently constitute a substantial part of the biodiversity of leaf-mining insects globally, and the checklist provided herein suggests that this likely has been the case for the past tens of millions of years. The potential for assigning adult fossils to extant genera however, makes continued search for amber-entombed Nepticulidae also important (Labandeira 2014). There is promising material in older ambers, such as Cedar Lake (Canadian) Amber (Campanian; 72–83 Ma) or Raritan (New Jersey) Amber (Turonian and Cenomanian; 89–101 Ma), of which rare lepidopteran fossils are known but generally unstudied (Grimaldi & Nascimbene 2010; McKellar & Wolfe 2010).

Acknowledgements

We would like to thank the late Niels Peder Kristensen (Copenhagen, Denmark) and Donald Lafontaine (Ottawa, Canada) for providing the loan of the amber fossil from record # 73 and entrusting to us the resulting drafts of a manuscript regarding this fossil. We would like to thank Thilo Fischer (Munich/Osnabrück, Germany) for loaning us several Baltic Amber fossils with adult Nepticulidae, and we would like to thank Mikhail Kozlov (Turku, Finland) for sharing the original photographs that were used for his publications, allowing us to publish these on-line, and assisting with translations of his original descriptions in Russian. We kindly thank Michael S. Engel (Kansas, USA) for helping us clarifying the status of record # 72, Scott Wing (Washington, D.C., USA) for verifying the host of record # 27, and Niklas Wahlberg (Turku, Finland) for discussing the age of the Nepticulidae based on molecular data. These colleagues greatly assisted us to improve the quality of the manuscript and were available for open discussion on the identifications of the fossils. We would like to thank Niko Korenhof for assembling the colour

plates. We would like to thank Mikhail Kozlov and an anonymous reviewer for their comments, which helped to additionally improve the manuscript. This is contribution 308 of the Evolution of Terrestrial Ecosystems at the National Museum of Natural History in Washington, D.C.

References

APG III. (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, 161, 105–121. http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x

Axelrod, D.I. (1995) *University of California publications in geological sciences. Vol. 139. The Miocene Purple Mountain Flora of Western Nevada*. University of California Press, Berkeley, Los Angeles, London, 98 pp.

Bell, C.D., Soltis, D.E. & Soltis, P.S. (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany*, 97, 1296–1303. http://dx.doi.org/10.3732/ajb.0900346

Berry, E.W. (1916) *The Lower Eocene Floras of Southeastern North America*. US Government Printing Office, Washington, D.C., 481 pp., 117 pls.

Bornhardt, J.F. (1975) Neue fossilfunde aus der Grube Messel und ihre Präparation. *Aufschluss*, 26, 453–473. Boucot, A. (1990) *Evolutionary paleobiology of behavior and coevolution*. Elsevier, Amsterdam, xxiii + 725 pp.

Brauckmann, C., Brauckmann, B. & Groening, E. (2001) Anmerkungen zu den bisher beschriebenen Lepidopteren aus dem Jung-Tertiär (Pliozän) von Willershausen am Harz. *Jahresberichte des Naturwissenschaftlichen Vereins in Wuppertal*, 54, 31–41.

Brenner, R., Ludvigson, G., Witzke, B., Zawistoski, A., Kvale, E., Ravn, R. & Joeckel, R. (2000) Late Albian Kiowa-Skull Creek marine transgression, Lower Dakota Formation, eastern margin of Western Interior Seaway, USA. *Journal of Sedimentary Research*, 70, 868–878. http://dx.doi.org/10.1306/2dc4093e-0e47–11d7–8643000102c1865d

Brooks, H.K. (1955) Healed wounds and galls on fossil leaves from the Wilcox deposits (Eocene) of western Tennessee. *Psyche*, 62, 1–9. http://dx.doi.org/10.1155/1955/76727

Carpenter, F.M. (1992) *Treatise on Invertebrate Paleontology. Part R Arthropoda. Vol. 4*. The Geological Society of America, Inc., Boulder, Colorado & The University of Kansas, Lawrence, Kansas, 377 pp. [pp. 279–655]

Crane, P.R. & Jarzembowski, E.A. (1980) Insect leaf mines from the Palaeocene of southern England. *Journal of Natural History*, 14, 629–639. http://dx.doi.org/10.1080/00222938000770531

Davis, D.R. & Robinson, G.S. (1998) The Tineoidea and Gracillarioidea. *In:* Kristensen, N.P. (Ed.), *Handbuch der Zoology. Vol. IV. (Arthropoda: Insecta), Part 35* (*Lepidoptera, Moths and Butterflies). Vol. 1. (Evolution, systematics and biogeography).* De Gruyter, Berlin/New York, pp. 91–117.

Dilcher, D.L. (1973) A revision of the Eocene flora of southeastern North America. *The Palaeobotanist*, 20, 7–18.

Ding, Q.L., Labandeira, C.C. & Ren, D. (2015) Biology of a leaf miner (Coleoptera) on *Liaoningocladus boii* (Coniferales) from the Early Cretaceous of Northeastern China. *Arthropod Systematics & Phylogeny*, 73, 281–308.

Donner, H. & Wilkinson, C. (1989) Nepticulidae (Insecta: Lepidoptera). *Fauna of New Zealand*, 16, 1–88.

Donovan, M.P., Wilf, P., Labandeira, C.C., Johnson, K.R. & Peppe, D.J. (2014) Novel insect leaf-mining after the end-Cretaceous extinction and the demise of Cretaceous leaf miners, Great Plains, USA. *PLoS ONE*, 9 (7), e103542. http://dx.doi.org/10.1371/journal.pone.0103542

Doorenweerd, C., Nieukerken, E.J. van & Menken, S.B.J. (2015) A global phylogeny of leafmining *Ectoedemia* moths (Lepidoptera: Nepticulidae): exploring host plant family shifts and allopatry as drivers of speciation. *PLoS ONE*, 10 (3), e0119586. http://dx.doi.org/10.1371/journal.pone.0119586

Ellis, W.N. (2014) Bladmineerders van Europa / Leafminers of Europe. Available from: http://www.bladmineerders.nl. (accessed 1 January 2015)

Emmet, A.M. (1988) *A field guide to the smaller British Lepidoptera*. The British Entomological & Natural History Society, London, 288 pp.

Fischer, T.C. (2013) Pygmy moths (Lepidoptera, Nepticulidae) from Baltic Amber (Eocene). *Zitteliana*, 53, 85–92.

Friis, E.M., Crane, P.R. & Pedersen, K.R. (2011) *Early flowers and Angiosperm evolution*. Cambridge University Press, Cambridge, x + 585 pp.

Fritsch, A. (1882) Fossile Arthropoden aus der Steinkohlen- und Kreideformation böhmens. *Beiträge zur Paläontologie Österreich-Ungarns und des Orients*, 2, 1–7.

Graham, A. (1999) *Late Cretaceous and Cenozoic history of North American vegetation, North of Mexico*. Oxford University Press, Oxford, xviii, 350 pp.

Grimaldi, D.A. & Engel, M.S. (2005) *Evolution of the Insects*. Cambridge University Press, New York, xv + 755 pp.

Grimaldi, D.A. & Nascimbene, P.C. (2010) Raritan (New Jersey) Amber. *In:* Penney, D. (Ed.), *Biodiversity of fossils in amber from the major world deposits*. Siri Scientific Press, Manchester, pp. 167–191.

Gustafsson, B. (1985) New species of *Stigmella* from The Gambia (Lepidoptera, Nepticulidae). *Tijdschrift voor Entomologie*, 127, 165–177.

Gustafsson, B. & Nieukerken, E.J. van (1990) Larvae of Nepticulidae. *In:* Johansson, R., Nielsen, E.S., Nieukerken, E.J. van & Gustafsson, B. (Eds.), The Nepticulidae and Opostegidae (Lepidoptera) of NW Europe. *Fauna Entomologica Scandinavica*, 23, 323–356.

Heikkilä, M., Kaila, L., Mutanen, M., Peña, C. & Wahlberg, N. (2012) Cretaceous origin and repeated Tertiary diversification of the redefined butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 279, 1093–1099. http://dx.doi.org/10.1098/rspb.2011.1430

Hering, E.M. (1951) *Biology of the Leaf Miners*. Junk, 's-Gravenhage, iv + 420 pp., 2 pls.

Hering, M. (1957) *Bestimmungstabellen der Blattminnen von Europa. 3 Vols*. Junk, 's-Gravenhage, 1185 pp.

Heyden, C. von (1862) Gliederthiere aus der Braunkohle des Niederrhein's, der Wetterau und der Röhn. *Paleontographica*, 10, 62–82, pl. x.

Hirano, N. (2010) Description of five new *Stigmella* and one *Bohemannia* species from Japan, with a new record of one *Bohemannia* species (Lepidoptera, Nepticulidae). *Japan Heterocerists' Journal*, 256, 124–134.

Hoare, R.J.B. (2000) A new genus of primitive Nepticulidae (Lepidoptera) from eastern Australia, with a revised diagnosis of nepticulid subfamilies. *Zoological Journal of the Linnean Society*, 128, 289–317. http://dx.doi.org/10.1111/j.1096–3642.2000.tb00165.x

Hoare, R.J.B. & Nieukerken, E.J. van (2013) Phylogeny and host-plant relationships of the Australian Myrtaceae leafmining moth genus *Pectinivalva* (Lepidoptera: Nepticulidae), with new subgenera and species. *ZooKeys*, 278, 1–64. http://dx.doi.org/10.3897/zookeys.278.4743.

International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature. Fourth edition. Adopted by the International Union of Biological Sciences*. The International Trust for Zoological Nomenclature, London, xxix, 306 pp.

Jarzembowski, E.A. (1980) Fossil insects from the Bembridge Marls, Palaeogene of the Isle of Wight, southern England. *Bulletin of the British Museum (Natural History). Geology*, 33, 237–293.

Jarzembowski, E.A. (1989) A century plus of fossil insects. *Proceedings of the Geologists' Association*, 100, 433–449. http://dx.doi.org/10.1016/s0016–7878(89)80019–7

Jarzembowski, E.A. (1995) Checklist of Tertiary insects from Dorset. *Proceedings of the Dorset Natural History and Archaeological Society*, 116, 145–146.

Johansson, R., Nielsen, E.S., Nieukerken, E.J. van & Gustafson, B. (1990) The Nepticulidae and Opostegidae (Lepidoptera) of North West Europe 1–7. *Fauna Entomologica Scandinavica*, 23, 1–739. [Brill, Leiden]

Johnson, K.R. (1996) Description of seven common plant megafossils from the Hell Creek Formation (Late Cretaceous: late Maastrichtian), North Dakota, South Dakota, and Montana). *Proceedings of the Denver Museum of Natural History*, 3, 1– 48.

Johnson, K.R. (2002) Megaflora of the Hell Creek and lower Fort Union Formations in the western Dakotas: Vegetational response to climate change, the Cretaceous–Tertiary boundary event, and rapid marine transgression. *Geological Society of America Special Paper*, 361, 329–391. http://dx.doi.org/10.1130/0–8137–2361–2.329

Kaila, L. (2011) Elachistine moths of Australia (Lepidoptera: Gelechioidea: Elachistidae). *Monographs on Australian Lepidoptera,* 11, 1–443. [CSIRO Publishing, Melbourne]

Kemperman, T.C.M., Wilkinson, C., Kuroko, H. & Kumata, T. (1985) Japanese species of the genus *Stigmella* (Nepticulidae: Lepidoptera). *Insecta Matsumurana*, 32, 1–107.

Kernbach, K. (1967) Über die bisher im Pliozän von Willershausen gefundenen Schmetterlings- und Raupenreste. *Bericht der Naturhistorischen Gesellschaft zu Hannover*, 111, 103–108.

Kinzelbach, R.K. (1970) Eine Gangmine aus dem eozänen Ölschiefer von Messel (Insecta: ? Lepidoptera). *Paläontologische Zeitschrift*, 44, 93–96. http://dx.doi.org/10.1007/bf02989798

Knor, S., Prokop, J., Kvaček, Z., Janovský, Z. & Wappler, T. (2012) Plant-arthropod associations from the Early Miocene of the Most Basin in North Bohemia – Palaeoecological and palaeoclimatological implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 321–322, 102–112. http://dx.doi.org/10.1016/j.palaeo.2012.01.023

Kobayashi, S., Hirowatari, T. & Kuroko, H. (2010) A revision of the Japanese species of the family Bucculatricidae (Lepidoptera). *Transactions of the Lepidopterological Society of Japan*, 61, 1–57.

Kobayashi, S., Huang, G.-H. & Hirowatari, T. (2011) Two species of Gracillariidae (Lepidoptera) new to China, and description of the pupal morphology of the genera *Corythoxestis* and *Eumetriochroa*. *Zootaxa*, 26, 25–32.

Kozlov, M.V. (1988) Paleontologiya Cheshuekrylykh i voprosy filogenii otryada Papilionida [Palaeontology of lepidopterans and problems in Papilionida order phylogeny]. *In:* Ponomarenko, A.G. (Ed.), *Melovoj Biotsenoticheskij Krizis i Evolutsiya Nasekomykh*. [*Cretaceous Biocoenotic Crisis and Insect Evolution*.]. Nauka, Moscow, pp. 16–69 & pp. 216–228 (references).

Krassilov, V.A. (2007) Mines and galls on fossil leaves from the Late Cretaceous of southern Negev, Israel. *African Invertebrates*, 48, 13–22.

Krassilov, V. (2008a) Mine and gall predation as top down regulation in the plant–insect systems from the Cretaceous of Negev, Israel. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 261, 261–269. http://dx.doi.org/10.1016/j.palaeo.2008.01.017

Krassilov, V. (2008b) Traumas on fossil leaves from the Cretaceous of Israel. *In:* Krassilov, V. & Rasnitsyn, A. (Eds.), *Plant-Arthropod Interactions in the Early Angiosperm History: Evidence from the Cretaceous of Israel*. Pensoft Publishers & Brill, Sofia & Leiden, pp. 7–187.

Kristensen, N.P. & Skalski, A.W. (1998) Phylogeny and palaeontology. *In:* Kristensen, N.P. (Ed.), *Handbuch der Zoologie. Vol. IV. (Arthropoda: Insecta), Part 35 (Lepidoptera, Moths and Butterflies). Vol. 1. (Evolution, Systematics and Biogeography)*. De Gruyter, Berlin & New York, pp. 7–25.

Kumata, T. (1998) Japanese species of the subfamily Oecophyllembiinae Real et Balachowsky (Lepidoptera: Gracillariidae), with descriptions of a new genus and eight new species. *Insecta Matsumurana*, 54, 77–131.

Kuroko, H. (1987) A fossil leaf mine of Nepticulidae (Lepidoptera) from Japan. *Bulletin of the Sugadaira Montane Research Center*, 8, 119–121.

Kuroko, H. (1990) Lepidopterous leaf miners of Japan (2). *Kita-Kyushu no Konchu*, 37, 1–2, pl. 1.

Labandeira, C.C. (1998) The role of insects in Late Jurassic to Middle Cretaceous ecosystems. *New Mexico Museum of Natural History and Science Bulletin*, 14, 105–123.

Labandeira, C.C. (2002a) The History of associations between plants and animals. *In:* Herrera, C.M. & Pellmyr, O. (Eds.), *Plant-Animal Interactions: An Evolutionary Approach*. Blackwell Science, London, pp. 28–74, 248–261.

Labandeira, C.C. (2002b) Paleobiology of middle Eocene plant-insect associations from the Pacific Northwest: A preliminary report. *Rocky Mountain Geology*, 37, 31–59. http://dx.doi.org/10.2113/gsrocky.37.1.31

Labandeira, C.C. (2014) Amber. *In:* Laflamme, M., Schiffbauer, J.D. & Darroch, S. (Eds.), Reading and writing of the fossil record: preservational pathways to exceptional fossilization. *Paleontological Society Short Course*, 20, 163–216.

Labandeira, C.C., Dilcher, D.L., Davis, D.R. & Wagner, D.L. (1994) Ninety-seven Million years of Angiosperm-Insect association - Paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 12278–12282. http://dx.doi.org/10.1073/pnas.91.25.12278

Labandeira, C.C., Johnson, K.R. & Wilf, P. (2002) Impact of the terminal Cretaceous event on plant-insect associations. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 2061–2066. http://dx.doi.org/10.1073/pnas.042492999

Labandeira, C.C., Wilf, P., Johnson, K.R. & Marsh, F. (2007) *Guide to insect (and other) damage types on compressedplant fossils*, Version 3.0. Smithsonian Institution, Washington, DC., 25 pp. Available from: http://paleobiology.si.edu/pdfs/ insectDamageGuide3.01.pdf (accessed 1 April 2015)

Lang, P.J., Scott, A.C. & Stephenson, J. (1995) Evidence of plant-arthropod interactions from the Eocene Branksome Sand Formation, Bournemouth, England: Introduction and description of leaf mines. *Tertiary Research*, 15, 145–174.

Laštůvka, A. & Laštůvka, Z. (1997) *Nepticulidae Mitteleuropas. Ein illustrierter Begleiter (Lepidoptera).* Konvoj, Brno, 229 pp.

Lewis, S.E. (1969) Lepidopterous larval-mining of an oak leaf from the Latah Formation (Miocene) of Eastern Washington. *Annals of the Entomological Society of America*, 62, 1210–1211. http://dx.doi.org/10.1093/aesa/62.5.1210

Lewis, S.E., Heikes, P.M. & Lewis, K.L. (1990) Entomofauna from the Brick Yard site (Miocene) near Spokane, Washington.

Occasional Papers in Paleobiology St Cloud State University, 4, 1–16.

Liebhold, A.M., Volney, W.J.A. & Schorn, H.E. (1982) An unidentified leaf mine in fossil *Mahonia reticulata* (Berberidaceae).

Canadian Entomologist, 114, 455–456. http://dx.doi.org/10.4039/ent114455–5

Magallón, S., Hilu, K.W. & Quandt, D. (2013) Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, 100, 556–73. http://dx.doi.org/10.3732/ajb.1200416

Martins-Neto, R.G. (1989) Novos insetos terciários do estado de São Paulo. *Revista Brasileira de Geociências*, 19 (3), 375–386.

McKellar, R.C. & Wolfe, A.P. (2010) Canadian Amber. *In:* Penney, D. (Ed.), *Biodiversity of fossils in Amber from the major world deposits*. Siri Scientific Press, Manchester, pp. 149–166.

Menken, S.B.J. & Roessingh, P. (1998) Evolution of insect-plant associations: sensory perception and receptor modifications direct food specialization and host shifts in phytophagous insects. *In:* Howard, D.J. & Berlocher, S.H. (Eds.), *Endless Forms*. Oxford University Press, New York, pp. 145–156.

Mutanen, M., Wahlberg, N. & Kaila, L. (2010) Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2839–2848. http://dx.doi.org/10.1098/rspb.2010.0392

Nieukerken, E.J. van (1983) The Cistaceae-feeding Nepticulidae (Lepidoptera) of the Western Palearctic region. *Systematic Entomology*, 8, 453–478. http://dx.doi.org/10.1111/j.1365-3113.1983.tb00495.x

Nieukerken, E.J. van (1986) Systematics and phylogeny of holarctic genera of Nepticulidae (Lepidoptera, Heteroneura: Monotrysia). *Zoologische Verhandelingen*, 236, 1–88.

Nieukerken, E.J. van (2007) *Acalyptris* Meyrick: revision of the *platani* and *staticis* groups in Europe and the Mediterranean (Lepidoptera : Nepticulidae). *Zootaxa*, 1436, 1–48.

Nieukerken, E.J. van (2014) Nepticulidae and Opostegidae of the world. Vol. 2014. Scratchpads, biodiversity online. Available from: http://nepticuloidea.info (accessed 1 December 2014)

Nieukerken, E.J. van & Liu, Y. (2000) Nepticulidae (Lepidoptera) in China, 1. Introduction and *Stigmella* Schrank feeding on Fagaceae. *Tijdschrift voor Entomologie*, 143, 145–181. http://dx.doi.org/10.1163/22119434-99900042

Nieukerken, E.J. van, Nielsen, E.S., Johansson, R. & Gustafsson, B. (1990) Introduction to the Nepticulidae. *In:* Johansson, R., Nielsen, E.S., Nieukerken van, E.J. & Gustafsson, B. (Eds.), The Nepticulidae and Opostegidae (Lepidoptera) of NW Europe. *Fauna Entomologica Scandinavica*, 23, pp. 11–109.

Nieukerken, E.J. van, Berg, C. van den & Hoare, R.J.B. (2011a) A new species of the endemic Australian genus *Roscidotoga* Hoare from rainforests in southern Queensland (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie*, 154, 193–201. http://dx.doi.org/10.1163/22119434–900000318

Nieukerken, E.J. van, 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, J.C., Kozlov, M., 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. & Zwick, A. (2011b) 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.

Ogg, J.G., Ogg, G. & Gradstein, F.M. (2008) *The concise geologic time scale*. Cambridge University Press, Cambridge, 177 pp. Opler, P.A. (1973) Fossil lepidopterous leaf mines demonstrate the age of some insect-plant relationships. *Science*, 179, 1321–1323. http://dx.doi.org/10.1126/science.179.4080.1321

Opler, P.A. (1974a) Biology, ecology and host specificity of Microlepidoptera associated with *Quercus agrifolia* (Fagaceae). *University of California Publications in Entomology*, 75, 1–83, pls. 1–7.

Opler, P.A. (1974b) Oaks as evolutionary islands for leaf-mining insects. *American Scientist*, 62, 67–73.

Peñalver, E. (1997) Hojas fósiles del Terciaro de Teruel con marcas de herbivorismo debidas a orugas. *Boletín de la Sociedad Entomologica Aragonesa*, 19, 29–33.

Peñalver, E. & Delclòs, X. (2004) Insectos del Mioceno inferior de Ribesalbes (Castellón, España). Interacciones planta-insecto. *Treballs del Museu de Geologia de Barcelona*, 12, 69–95.

Pigg, K.B., Wehr, W.C. & Ickert-Bond, S.M. (2001) *Trochodendron* and *Nordenskioldia* (Trochodendraceae) from the Middle Eocene of Washington State, U.S.A. *International Journal of Plant Sciences*, 162, 1187–1198. http://dx.doi.org/10.1086/321927

Puplesis, R. (1994) *The Nepticulidae of eastern Europe and Asia. Western, Central and Eastern Parts*. Backhuys Publishers, Leiden, 290 pp.

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 Evolutionary Biology*, 9, 280. http://dx.doi.org/10.1186/1471–2148–9–280

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One*, 8, e58568. http://dx.doi.org/10.1371/journal.pone.0058568

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

Ritzkowski, S. (1997) Das Geologische Alter der Bernsteinführenden sedimente in Sambia (Bezirk Kaliningrad), bei Bitterfeld (Sachsen-Anhalt) und bei Helmstedt (SE-Niedersachsen).

In: Kosmowska-Ceranowicz, B. (Ed.), *Baltic Amber and other fossil resins*. The Archeological Museum in Gdansk, Museum of the Earth, Polish Academy of Sciences, Gdansk, pp. 33–38.

Rozefelds, A.C. (1988a) Insect leaf mines from the Eocene Anglesea locality, Victoria, Australia. *Alcheringia*, 12, 1–6. http://dx.doi.org/10.1080/03115518808618992

Rozefelds, A.C. (1988b) Lepidoptera mines in *Pachypteris* leaves (Corystospermaceae: Pteridospermophyta) from the Upper Jurassic/Lower Cretaceous Battle Camp Formation, north Queensland. *Proceedings of the Royal Society of Queensland*, 99, 77–81.

Schachat, S.R., Labandeira, C.C., Gordon, J., Chaney, D., Levi, S., Halthore, M.N. & Alvarez, J. (2014) Plant–insect interactions from Early Permian (Kungurian) Colwell Creek Pond, north-central Texas: the early spread of herbivory in riparian environments. *International Journal of Plant Science*, 175, 855–890. http://dx.doi.org/10.1086/677679

Scoble, M.J. (1983) A revised cladistic classification of the Nepticulidae (Lepidoptera) with descriptions of new taxa mainly from South Africa. *Transvaal Museum Monograph*, 2, 1–105. Available from: http://hdl.handle.net/10499/AJ10150

Scott, A.C. & Titchener, F.R. (1999) Techniques in the study of plant-arthropod interactions. *In:* Jones, T.P. & Rowe, N.P. (Eds.), *Fossil plants and spores: modern techniques*. The Geological Society, London, pp. 310–315.

Skala, H. (1941) Neues über Miner. *Zeitschrift des Wiener Entomologen-Vereines*, 26, 55–57, 77–80, 123–125, pls. 1–4. Skalski, A.W. (1976) Les Lépidoptères fossiles de l'ambre. Etat actuel de nos connaissances (2e partie). *Linnaeana Belgica*, 6, 195–208.

Skalski, A.W. (1979) Records of oldest Lepidoptera. *Nota Lepidopterologica*, 2, 61–66.

Skalski, A.W. (1990a) An annotated review of fossil records of lower Lepidoptera. *Bulletin of the Sugadaira Montane Research Center University of Tsukaba*, 11, 125–128.

Skalski, A.W. (1990b) The Families Nepticulidae and Thyrididae in Baltic amber (Lepidoptera). Proceedings of the Seventh Congress of the Societas Europaea Lepidopterologica, Lunz. *Nota Lepidopterologica Supplement*, 4, 163–164.

Sohn, J.-C. & Lamas, G. (2013) Corrections, additions, and nomenclatural notes to the recently published World catalog of fossil and subfossil Lepidoptera. *Zootaxa*, 3599 (4), 395–399. http://dx.doi.org/10.11646/zootaxa.3599.4.8

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

Sommerer, M. (2002) To agree or not to agree - the question of gender agreement in the International Code of Zoological Nomenclature. *Nota Lepidopterologica*, 25, 191–204.

Spencer, K.A. (1990) *Host specialization in the world Agromyzidae (Diptera)*. Kluwer Academic Publishers, Dordrecht/ Boston/London, xii, 444 pp.

Stephenson, J. (1991) *Evidence of plant/insect interactions in the Late Cretaceous and Early Tertiary*. Ph. D. thesis, University of London, London, 378 pp.

Stephenson, J. & Scott, A.C. (1992) The geological history of insect-related plant damage. *Terra Nova*, 4, 542–552. http://dx.doi.org/10.1111/j.1365–3121.1992.tb00596.x

Stevens, P.F. (2013) *Angiosperm Phylogeny Website Version 13, September 2013* [and more or less continuously updated since]. University of Missouri and Missouri Botanical Garden, St. Louis. Available from: http://www.mobot.org/MOBOT/ Research/APweb/welcome.html (accessed 1 March 2015)

Straus, A. (1976) Eine Gangmine im Eocän von Messel. *Aufschluß*, 27, 445–446.

Straus, A. (1977) Gallen, Minen und andere Frassspuren im Pliokän von Willershausen am Harz. *Verhandlungen des Botanischen Vereins der Provinz Brandenburg*, 113, 43–80.

Timmermans, M.J., Lees, D.C. & Simonsen, T.J. (2014) Towards a mitogenomic phylogeny of Lepidoptera. *Molecular Phylogenetics and Evolution*, 79 C, 169–178. http://dx.doi.org/10.1016/j.ympev.2014.05.031

Upchurch, G.R. & Dilcher, D.L. (1990) Cenomanian angiosperm leaf megafossils, Dakota Formation, Rose Creek locality, Jefferson County, southeastern Nebraska. *United States Geological Survey Bulletin*, 1915, 1–55, pls. 1–30.

Vári, L. (1963) South African Lepidoptera, 3. Descriptions of new Stigmellidae. *Koedoe*, 6, 66–75. http://dx.doi.org/10.4102/koedoe.v6i1.813

Wahlberg, N., Wheat, C.W. & Pena, C. (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS One*, 8 (11), e80875. http://dx.doi.org/10.1371/journal.pone.0080875

Wang, H. (2002) *Diversity of Angiosperm leaf megafossils from the Dakota Formation (Cenomanian, Cretaceous), northwestern interior, USA*. Ph. D. thesis, University of Florida, Gainesville, Florida, 395 pp.

Wheat, C.W. & Wahlberg, N. (2013) Critiquing blind dating: the dangers of over-confident date estimates in comparative genomics. *Trends in Ecology, Systematics and Evolution*, 28, 636–42. http://dx.doi.org/10.1016/j.tree.2013.07.007

Wilf, P. & Escapa, I.H. (2014) Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. *New Phytologist 201*. [published online] http://dx.doi.org/10.1111/nph.13114

Wilf, P., Labandeira, C.C., Johnson, K.R., Coley, P.D. & Cutter, A.D. (2001) Insect herbivory, plant defense, and early Cenozoic climate change. *Proceedings of the National Academy of Sciences*, 98, 6221–6226. http://dx.doi.org/10.1073/pnas.111069498

Wilf, P., Labandeira, C.C., Johnson, K.R. & Cúneo, N.R. (2005) Richness of plant-insect associations in Eocene Patagonia: A legacy for South American biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8944–8948. http://dx.doi.org/10.1073/pnas.0500516102

Wilkinson, C. (1979) A taxonomic study of the micro-lepidopteran genera *Microcalyptris* Braun and *Fomoria* Beirne occurring in the United States of America (Lepidoptera, Nepticulidae). *Tijdschrift voor Entomologie*, 122, 59–90.

Wilkinson, C. & Newton, P.J. (1981) The micro-lepidopteran genus *Ectoedemia* Busck (Nepticulidae) in North America. *Tijdschrift voor Entomologie*, 124, 27–92.

Winkler, I.S., Labandeira, C.C., Wappler, T. & Wilf, P. (2010) Distinguishing Agromyzidae (Diptera) leaf mines in the fossil record: New taxa from the Paleogene of North America and Germany and their evolutionary implications. *Journal of Paleontology*, 84, 935–954. http://dx.doi.org/10.1666/09–163.1

Zherikhin, V.V. (1978) Development and changes of the Cretaceous and Cenozoic faunal assemblages (Tracheata and Chelicerata). *Trudy Paleontologischeskogo Instituta Akademii Nauk, SSSR*, 165, 1–198.

Zherikhin, V.V. (2002) Insect trace fossils. *In:* Rasnitsyn, A.P. & Quicke, D.L.J. (Eds.), *History of insects*. Kluwer, Dordrecht, Boston, London, pp. 303–324.

Text Box: Leaf-mine terminology

Since there is a confusion regarding descriptive terms for leaf mines, and differences exist between North American and European literature, we provide a short list of terms and their synonyms. Hering (1951) provided several Latinized terms, listed below, but these versions are rarely used.

Ophionome – linear mine. A mine in which the larva moves in one forward direction. In British literature such a mine is usually termed a "gallery"; in North American literature an approximate equivalent is a "serpentine mine". The latter usage is confusing (see below). The ophionome mine is the common type in the Nepticulidae.
Heliconome – serpentine mine (auct. Hering). A linear mine that has a sinusoidal, occasionally spiral, trajectory in a leaf, particularly in its earlier stages and often later becomes a more rectilinear mine. Examples in Nepticulidae include *Enteucha acetosae* (Stainton), *Stigmella prunifoliella* (Clemens) and several *Ectoedemia* species.
Visceronome – intestinally coiled mine. A linear mine that turns back and forth in a tight, zigzag pattern, such that the individual coils are adjacent to one another in an intestine-like fashion. Example: *Stigmella viscerella* (Stainton) **Stigmatonome – blotch mine**. A mine in which the larva consumes one or more tissue layers in all or several directions. Stigmatonomes are divided into two types:
Orthogenous stigmatonome. A blotch mine in which the larva consumes tissue in all directions without any preferential feeding pattern. Orthogenous stigmatonomes are rare in Nepticulidae; examples include *Ectoedemia occultella* (Linnaeus) and *Stigmella paradoxa* (Frey).
Ophiogenous stigmatonome. or a false blotch. A blotch mine originating by coalescence of linear mines, such that the larva changes feeding direction and creates the appearance of a blotch mine by an abundance criss-crossing intersections with occasional islands of unmined, often squarish, tissue. Ophiogenous stigmatonomes are more common in Agromyzidae. In Nepticulidae false blotches are often formed when the larva lacks sufficient mining space and is obliged to follow previously made mine tracks.
Ophistigmatonome – linear-blotch mine. A combination of mine types wherein the larva initiates a gallery mine, and frequently after the last molt starts the formation of a blotch, often in the form of a wide, broad gallery. Ophistigmatonomes are common in Nepticulidae, particularly in *Ectoedemia*. If the entire leaf or other foliage organ is mined, Hering (1951) has termed such a construction as a "pantonome".

Additional terms that describe mines that do not occur in Nepticulidae are: *asteronome*, star-shaped mine of radiating mine trails; *physonome*, a blister mine; and *ptychonome*, a tentiform mine. Additional terminology for mines in plant parts other than foliage are: *carponome*, occurring in fruit; *caulonome*, occurring in a stem; and *anthonome*, occurring in a flower.

5

Phylogeny and divergence times of Nepticulidae

Phylogeny, classification and divergence times of pygmy leafmining moths (Lepidoptera: Nepticulidae): the earliest lepidopteran radiation on Angiosperms?

Camiel Doorenweerd[1, 2], Erik J. van Nieukerken[1] and Robert J.B. Hoare[3]

[1] Naturalis Biodiversity Center, Leiden, The Netherlands
[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands
[3] Landcare Research, Auckland, New Zealand

Abstract

Nepticulidae is one of the early diverging Lepidoptera lineages, and currently comprises over 850 described species. The larvae of the vast majority of the species are leafminers on Angiosperms and highly monophagous, which has led to persistent ideas on coevolution with their plant hosts. We here present a molecular phylogeny based on eight gene fragments from 355 species, representing 20 out of 22 extant Nepticulidae genera. Using two fossil calibration points, we performed molecular dating to place the origin of the family in the Early Cretaceous, before the main Angiosperm diversification. Based on our results we propose a new classification, abandoning all ranks between family and genus, as well as subgenera to allow for a stable classification. The position of *Enteucha* within Nepticulidae remains somewhat ambiguous, and the species-rich cosmopolitan genus *Stigmella*, with nearly half of all described Nepticulidae, requires further study. *Ectoedemia*, *Zimmermannia*, *Acalyptris*, *Etainia*, *Parafomoria*, *Muhabbetana* and *Fomoria* appear to have diversified in a relatively short evolutionary period, leading to short branches in the molecular phylogeny and unclear suprageneric relations. Otherwise support values throughout the phylogeny are mostly high and the species groups, genera and higher clades are discussed in respect to their supporting morphological and life history characters. Wing venation characters are confirmed to mostly be reliable and relevant for classification, but some other previously used characters require re-interpretation. The species groups of most genera are recovered, but only partly so in the large genus *Stigmella*. The molecular dating results are compared with existing knowledge on the timing of the Angiosperm radiation and reveal that the diversification of Nepticulidae could largely have been contemporaneous with their hosts, although some of the genera restricted to a single plant family appear to have begun to diversify before their hosts.

Keywords: molecular dating, systematics, Heteroneura, Nepticuloidea, Opostegidae, Cretaceous

Introduction

How the timing and pattern of herbivorous insect diversification relates to Angiosperm plant diversification in the Early Cretaceous remains as one of the great evolutionary questions of today (Wiens *et al*., 2015; Wahlberg *et al*., 2013; Tilmon, 2008; Grimaldi & Engel, 2005). Lepidoptera in particular are striking in their plant-dependent diversity; it is one of the four largest insect orders and the only one that is almost entirely associated with Angiosperms (Wiens *et al*., 2015; Powell *et al*., 1998). To understand the evolutionary history of early Lepidoptera, it is worthwhile to envision the scene of the rapidly evolving mid-Cretaceous environment (Magallón *et al*., 2013; Skelton, 2003; Vakhrameev & Hughes, 1991). Imagine standing on the flood plain of the shallow Western Interior Seaway that covered much of the mid-West United States during the late Albian, ca. 104 Ma (Graham, 1999). The higher country is still covered with forests composed of conifers, cycads and other early seed plants and herds of dinosaurs roam freely. The flood plain, however, is covered with an already diverse pioneer vegetation with hundreds of species of a young group of fast growing plants: the Angiosperms (Lidgard & Crane, 1988). The softer and more palatable leaves of Angiosperms offer a new biome of resources ready to be exploited by the insects. Indeed, as an observer you will soon see that many leaves are damaged in various ways, and some show tracks inside: leaf mines (Labandeira *et al*., 2007; 1994).

Albian fossil leaf mines closely resemble modern day Nepticulidae leaf mines, and the variation in morphology suggests that these insects had already diversified to some extent (Doorenweerd *et al*., 2015a). Higher lepidopteran groups have been identified as well, including the likely presence of the ditrysian leaf blotch mining moths Gracillariidae (Labandeira *et al*., 1994), but butterflies and most larger moths were probably scarce or absent (Wahlberg *et al*., 2013). Fossil evidence provides crucial information on the early evolution of Lepidoptera, but integration with time-calibrated molecular phylogenetic studies will be essential for a full understanding of the timing of diversification of different groups. Initial studies integrating both have mostly focussed on discrepancies (Sohn *et al*., 2015) and many fossil records likely require reinterpretation following modern insights in Lepidoptera classification (Heikkilä *et al*., 2015). However, a revisionary study on Nepticulidae fossils has shown that there is huge potential in the fossil record to provide additional calibration points in studies that employ molecular dating (Doorenweerd *et al*., 2015a).

Knowledge of the phylogeny of Lepidoptera has matured in recent years (Bazinet *et al*., 2016; Heikkilä *et al*., 2015; Regier *et al*., 2015; 2013) and Nepticulidae are consistently placed among the earlier lineages of non-ditrysian moths. Nepticulidae and Opostegidae together form the superfamily Nepticuloidea, a sistergroup relationship that is well supported morphologically and molecularly. A basal division of Heteroneura between Nepticuloidea and all other Lepidoptera is also well supported (Bazinet *et al*., 2016; Regier *et al*., 2015). This means that in the so-called Angiospermivora, the Nepticuloidea evolved after four or five other

clades split off, in all having no more than ca. 700 extant species, of which more than 600 belong to the Hepialidae, a group that consists of often polyphagous root feeders, wood borers and leaf feeders rather than specialist (monophagous or oligophagous) angiosperm-feeders. Nepticulidae are commonly called pygmy moths because the adult moths are amongst the smallest of all Lepidoptera (Fig.

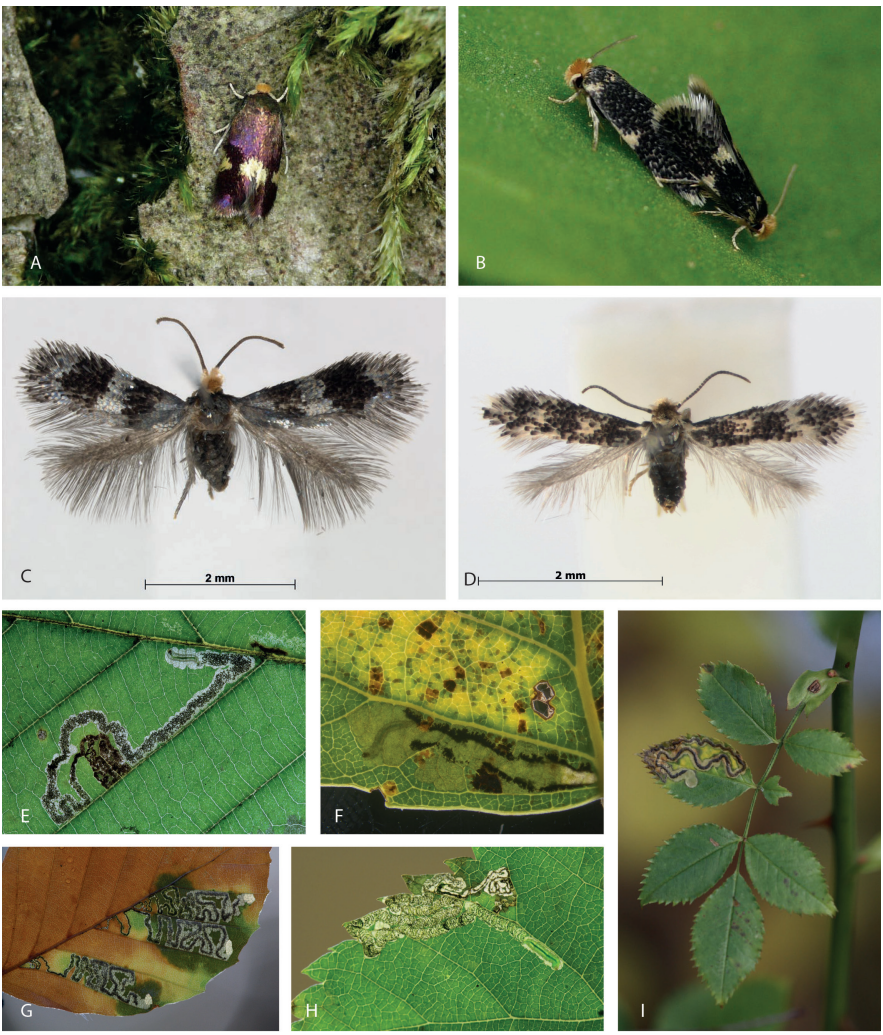


Figure 1: An impression of Nepticulidae adults and larvae. A: *Bohemannia quadrimaculella* adult resting on tree bark, The Netherlands (Photo: George Sinnema). B: *Ectoedemia subbimaculella* copula on a leaf, The Netherlands (Photo: George Sinnema). C: *Enteucha* spread adult female, Japan (4722 Hirano). D: *Simplimorpha promissa* spread adult male, Greece (EvN2011385). E: *Ectoedemia* sp. leafmining larva on *Fagus*, Japan (CD13063 RMNH.INS.29826). F: *Ectoedemia hannoverella* leafmining larva in *Populus* leaf (EvN2008105). G: *Stigmella tityrella* leafmines in *Fagus sylvatica*, The Netherlands. H: *Stigmella sashai* leafmining larva in *Tilia maximowicziana*, Japan (CD13040). 1I: *Stigmella anomalella* leafmine on *Rosa* sp., The Netherlands (CD12028).

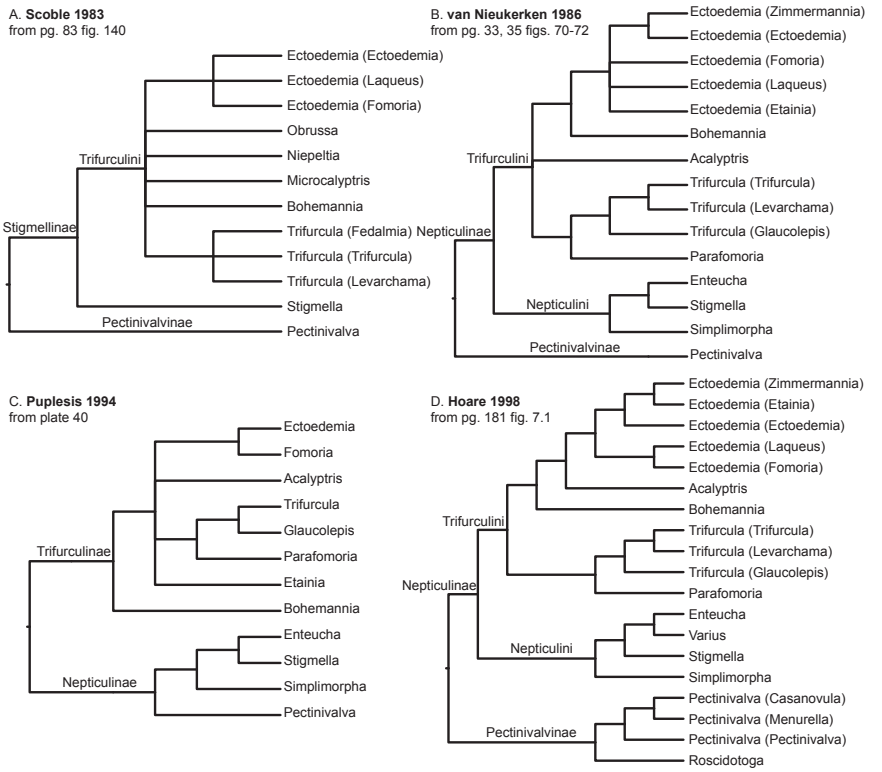


Figure 2: Proposed Nepticulidae phylogenies during the past century based on morphology. The genus names are indicated as originally published and contain synonyms compared to the newly proposed classification. A: Although Beirne (1945) was the first to propose a phylogeny for Nepticulidae that included genitalic characters, Scoble was the first to present a phylogeny strictly based on cladistic principles. B: The expanded phylogeny resulting from the PhD studies of van Nieukerken (1986b). C: the Nepticulidae phylogeny as published in Puplesis (1994), which did not apply the rank of subgenus. D: The phylogeny from the PhD studies of Hoare (1998), which built on the phylogeny of van Nieukerken (1986) with a particular focus on Pectinivalvinae and was the first to use automated algorithms.

1A–D), but in terms of diversity they comprise the largest group of early diverging Lepidoptera with approximately 850 named species and an estimated 2,000–2,500 species globally. They are specialized endophytophagous insects, mostly leafminers (Fig. 1E–I), but a few groups are bark- or stemminers, shoot borers, or feed in green fruits of *Acer*, and a handful are gall makers. Species generally feed in 'core Eudicot' angiosperms (APG III, 2009), with a preference for woody plants.

Ever since the erection of the genus *Nepticula* Heyden, 1843 (Stainton, 1849; Zeller, 1848) - a junior synonym of *Stigmella* Schrank, 1802 (Wilkinson, 1978) - the group has been recognised as a unity, and as the family Nepticulidae since 1854 (Stainton, 1854). Soon after *Nepticula* was erected, two more genera were recognised on the

basis of venation: *Trifurcula* Zeller, 1848 and *Bohemannia* Stainton, 1859 (Stainton, 1859; Zeller, 1848). In the early 20th Century, American authors erected four more genera with venation as leading characters: *Ectoedemia* Busck, 1907 (also based on the galling habit), *Obrussa* Braun, 1915, *Glaucolepis* Braun, 1917 and *Microcalyptris* Braun, 1925 (Braun, 1925, 1917, 1915; Busck, 1907). Gradually the study of genitalia became standard in Lepidoptera and the first of such studies involving nepticulids divided *Nepticula* into a number of species groups (Petersen, 1930). Later, Beirne (1945), when describing the genitalia of the British Lepidoptera, erected five new genera, viz. *Dechtiria* Beirne, 1945, *Levarchama* Beirne, 1945, *Fedalmia* Beirne, 1945, *Etainia* Beirne, 1945 and *Fomoria* Beirne, 1945, and split *Stigmella* into *Stigmella* and *Nepticula*.

The first cladistic analysis and classification resulted from the PhD studies on the South African Nepticulidae fauna of Scoble (1983), here redrawn in Fig. 2A. The family was divided into two subfamilies: the Australian Pectinivalvinae and the global Nepticulinae, the latter subdivided into Nepticulini and Trifurculini. The use of subgenera was introduced in the genus *Ectoedemia*: *Ectoedemia*, *Fomoria* Beirne, 1945 and *Laqueus* Scoble, 1983. In 1986 this classification was refined and extended on the basis of the Holarctic fauna during the PhD studies of van Nieukerken, and included larval and new adult characters in the character matrix (van Nieukerken, 1986b) (Fig. 2B). The main divisions were maintained, and two additional subgenera were included in *Ectoedemia*: *Etainia* Beirne, 1945 and *Zimmermannia* Hering, 1940. Synapomorphies were defined for all clades, except one, the subgenus *Ectoedemia* (*Fomoria*). Both phylogenies were derived manually using Hennigian principles, by the comparison of character states and outgroup argumentation (Hennig, 1966). The next published phylogeny followed in 1994 (Puplesis, 1994) (Fig. 2C), and did not use Hennigian cladistics, but still recognised apomorphies, and divided the family into the subfamilies Nepticulinae and Trifurculinae. *Pectinivalva* Scoble, 1983 was treated within Nepticulinae, and the previously recognized subgenera were raised to full genus, except for *Laqueus*, which was synonymised with *Fomoria*.

Four years later, in his unpublished PhD thesis of Australian Nepticulidae, Hoare prepared the first cladistic analysis of Nepticulidae using maximum parsimony algorithms employed in computer software (Hoare, 1998) (Fig. 2D). The weakness of this analysis was in the low number of species included, but its strength was in the inclusion of characters of all life stages and this resulted in a refined version of the divisions by Scoble (1983) and van Nieukerken (1986b). Additionally from Hoare's PhD studies, analyses of the Australian Pectinivalvinae (Hoare, 2000b; Hoare & van Nieukerken, 2013) resulted in the recognition of the new genus *Roscidotoga* Hoare 2000, and division of the genus *Pectinivalva* Scoble, 1983 into three subgenera: *Pectinivalva*, *Casanovula* Hoare, 2013 and *Menurella* Hoare, 2013. In all cladistic analyses it was clear that some clades were supported by a whole array of characters, whereas others were hardly supported at all, and a bootstrap analysis of Hoare's dataset collapsed most of the branches in the tree into a polytomy and failed to support suprageneric groupings (Hoare, 1998).

Studies on Nepticulidae involving genetic sequence data started in the early 21st century with three gene fragments. The initial family-level results have been presented at several conferences (e.g. van Nieukerken *et al*., 2004a) but were difficult to reconcile with morphological findings and remained unpublished. During the PhD thesis studies of Doorenweerd, there was an opportunity to sequence up to eight genes for a comprehensive set of taxa with representatives from all polytypic nepticulid genera. We here present the resulting molecular phylogeny and suggest a sustainable new classification for Nepticulidae. Furthermore, we use two fossil calibration points to estimate divergence times that provide insight on how one of the earliest lineages of Lepidoptera diversified alongside Angiosperm host plants. Simultaneous with this publication, a revised catalogue for Nepticulidae will be published (van Nieukerken *et al*., 2016a), as well as a publication describing the three new genera and several species that are also included in the present study (van Nieukerken *et al*., 2016b).

Material & Methods

Taxon sampling

DNA barcoding has since 2005 been systematically included in our taxonomic studies of Nepticulidae and DNA analyses in general since 2000. This has thus far yielded over 2,800 DNA extracts and DNA barcodes, with either verified species names, or temporary species names for undescribed material or material that cannot yet be linked to a described species. All DNA barcodes are available through the Barcoding of Life Datasystems (BOLD) (Ratnasingham & Hebert, 2007). From these DNA extracts we made a selection to be analysed for multiple genes, covering all available genera and species groups. As outgroup we selected 11 exemplars of Opostegidae, the family that joins Nepticulidae in Nepticuloidea and is its undisputed sister (Regier *et al*., 2015a; van Nieukerken, 1986b). The final dataset used for the phylogenetic analysis includes 344 ingroup specimens from 20 genera with data from at least three genes. Except for two monotypic South African genera, viz. *Varius* Scoble, 1983 and *Areticulata* Scoble, 1983, all known extant Nepticulidae genera are represented, including three new genera from the Neotropics, that are being published simultaneously (van Nieukerken *et al*., 2016b). Collecting details and photographs of specimens may be found in BOLD dataset DS-NepPhylo [doi: dx.doi.org/10.5883/DS-NEPPHYLO].

DNA extraction, amplification and sequencing

The source specimens for DNA extraction were stored as dried pinned adults, as larvae frozen in ethanol >95% or occasionally as larvae that had been dried inside their leaf mines. For many samples, DNA extraction was performed non-destructively by recovering either the abdomen with genitalia or the larval pelt from the lysis buffer after incubation. The genomic DNA extraction continued from the lysis step using a Macherey-Nagel NucleoMag 96 Tissue magnetic bead kit on a Thermo Fisher KingFisher flex system. Polymerase Chain Reaction (PCR) was used to amplify target DNA sections of eight genes: cytochrome oxydase

subunit I (COI), cytochrome oxidase subunit II (COII), translation elongation factor 1-alpha (EF1-alpha), carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD), isocitrate dehydrogenase (IDH), cytosolic malate dehydrogenase (MDH), histone 3 and 28S ribosomal RNA (28S). Amplification of the genes *Wingless* and GAPDH was attempted using published primers and conditions (Wahlberg and Wheat 2008). For *Wingless* the resulting fragments varied in length from ~350–2,000 bp, judged from gel electrophoreses. Several single copy fragments of <1,000 bp were sequenced, but the results were so riddled with introns that they could not be aligned at all and we did not continue with this gene. For GAPDH, the amplification success was low (19% using 95 test samples). From the successfully sequenced samples we deduced that this was likely due to sequence regions near the primer sites with high similarity to the primers. Also, in GAPDH a single intron was encountered, other introns may have contributed to difficulties during amplification. For the final eight fragments used in this study, PCR chemicals and cycling conditions follow those of Doorenweerd *et al*. (2015b). Two genes that have additionally been used in this study are CAD and MDH, for which the primer pairs CADmidF & CAD1028R and hybMDF & MDHmidR, respectively, were used (Wahlberg & Wheat, 2008). PCR conditions were identical to the other nuclear markers and the annealing temperature was set at 55° C. Bidirectional Sanger sequencing was outsourced to BaseClear, Leiden, The Netherlands. The resulting chromatograms were checked for quality and congruence in Geneious R6.1.8 and the resulting sequences were managed using VoSeq 1.7.4 (Peña & Malm, 2012).

Sequence alignments

The alignment of 28S was prepared with MAFFT 7 (Katoh & Standley, 2013). The sequenced fragments of COI, H3 and MDH contained no insertions or deletions (indels) or introns and were straightforward to align. In the sequenced CAD fragment we found one triplet insertion in *Ectoedemia angulifasciella* (Stainton, 1849), specimen RMNH.INS.12764, in the COII fragment we found a single triplet insertion in an *Ectoedemia quadrinotata* (Braun, 1917) specimen, RMNH. INS.18557. In the IDH fragment there was a single position with up to four triplets inserted, which occurred in many samples throughout *Acalyptris* Meyrick, 1921 (0–2 triplets), *Etainia* (always three triplets) and *Stigmella* (0–4 triplets). In EF1-alpha, several introns were encountered that are presented in the results. All introns were removed prior to phylogenetic analyses, but they are included in the Genbank accessions. COI sequences were available for all specimens in the final dataset, the success rate of the remaining genes was 92% for 28S, 84% for EF1-alpha, 82% for COII, 66% for CAD, H3 and IDH, and 35% for MDH. A list of the available sequences per specimen and Genbank accession numbers can be found in S1. The final aligned length of the dataset used for analyses was 4,557 bp.

Phylogenetic analysis

Maximum likelihood and Bayesian approaches were used for phylogenetic inference on the concatenated dataset. The appropriate substitution model and optimal partitioning were determined using PartitionFinder 1.1 (Lanfear *et al*.,

2012). For all markers individually, as well as the combined dataset, the GTR + Gamma model proved most suitable according to the Bayesian information criterion. We initially reconstructed maximum likelihood trees using the PhyML plugin in Geneious (Guindon & Gascuel, 2003) for each genetic marker individually and assessed those for contamination issues or conflicting signal, we then repeated that approach for the mitochondrial markers combined, versus the nuclear markers combined. Although strongly differing in resolution, there was no incongruence between the phylogenetic signal of different datasets. In all subsequent analyses the dataset was analysed following the partitioning from PartitionFinder: eight partitions mostly following the division of gene fragments, except that IDH and MDH were combined, as well as the second and third codon positions of COI and COII, and the first codon position of COI and COII. Bayesian analyses were run with the Linux MPI version of Exabayes 1.4.1 (Aberer *et al*., 2014), ML analyses were done using Garli 2.01 (Zwickl, 2006). Exabayes was set to run until 1.5% convergence between four sets of four heated chains was reached, after which the sampled trees were examined with Tracer 1.6 (Rambaut *et al*., 2014) and revealed stable convergence and sufficient sampling (all ESS values >> 200). Eight best trees were searched with Garli to see if a single best topology could be found consistently with the used settings, and subsequently four independent runs with 100 bootstrap replicates were performed and averaged to obtain bootstrap supports. Complete consistency in the best ML topology could not be reached, which is included in the interpretation of the results. When interpreting the results, we considered branches with posterior probabilities over 0.95 and bootstrap values over 60 as well supported.

Divergence time estimation

Two fossil calibration points were selected to estimate timing of divergences with the software package BEAST 2.3.2 (Bouckaert *et al*., 2014). There is a regular occurrence of Nepticulidae-like leaf mine fossils in the fossil record since the earliest finds, representing multiple species, in the Dakota formation of the early Cretaceous, dated at 102 Ma (Doorenweerd *et al*., 2015a). We used this to calibrate the Nepticulidae crown with a log normal distribution. The second calibration is from adult Baltic Amber entombments identified as two species of *Bohemannia* (Doorenweerd *et al*., 2015a; Fischer, 2013), from a formation dated between 43 and 45.2 Ma. We used this to calibrate the crown *Bohemannia* clade, with a log normal distribution. The site models were divided in eight partitions. The clock models were separated for 28S, the protein coding nuclear genes and the mitochondrial genes. For each clock set the substitution rate was constrained for a single partition within the set according to values that are likely to approach realistic biological values (Papadopoulou *et al*., 2010): for 28S at 6E-4, for the nuclear genes 0.0017 and for the mitochondrial genes 0.0168. With multiple partitions in one set only one gene was constrained, allowing the results to be tested for convergence from these estimates. The length of the single chain Bayesian analysis was set to 120 million and results were checked for convergence and sufficient sampling using Tracer 1.6. The resulting trees were combined using TreeAnnotator (included in the BEAST package) and visualised with FigTree 1.4.2 (Rambaut, 2014). The BEAST runs

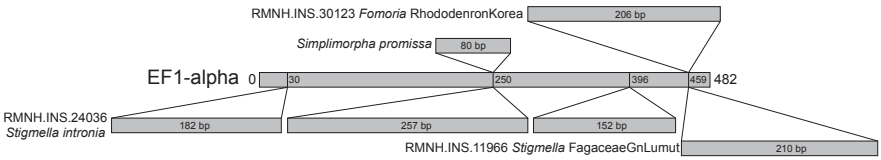


Figure 3: The different positions and sizes, drawn to scale, of introns encountered in Elongation Factor 1-alpha relative to the exon of 482 bp, indicated in the centre. Specimen RMNH.INS.11966 was not included in the phylogenetic analyses but sequence details are included in the BOLD dataset and S1.

were repeated multiple times to verify a consistent outcome, and the reliability of the calibration points was assessed by repeating the analyses with each point left out. This increased the confidence intervals of the age estimates of the left-out calibration, but there was no incongruence between the mean of the two calibration points. When discussing the results, we mostly indicate the 95% highest probability density (HPD) as the estimated age-range, to prevent overconfident conclusions that may result from focussing on the average estimated age.

Morphology

Morphological terms largely follow our earlier treatments (Hoare, 2000b; van Nieukerken, 1986b; van Nieukerken *et al*., 1990a), but we replace the previously used term aedeagus for the male intromittent organ with phallus, following the general trend in lepidopterological literature as suggested by Kristensen (2003). In the discussion we largely refer to published morphological information, but we also include unpublished information from RJBH's PhD thesis (Hoare, 1998) and of EJvN's study, and we also checked several characters on our own material when working on this manuscript.

Results

In the 482 base pair (bp) fragment of EF1-alpha that we amplified we encountered four positions with introns (Fig. 3). *Stigmella intronia* van Nieukerken & Nishida, 2016 from Costa Rica (RMNH.INS.24036) (van Nieukerken *et al*., 2016b) contains three positions with introns, on the second position in that specimen (pos. 250), all specimens of *Simplimorpha promissa* (Staudinger, 1870) have a much shorter intron. A fourth intron position was encountered in larvae of an undescribed *Fomoria* (*Fomoria* RhododendronKorea, RMNH.INS.30123) and of an undescribed *Stigmella* (*Stigmella* FagaceaeGnLumut, RMNH.INS.11966) with introns of different length and composition. After removing the introns, the amino acid translations closely matched those of other Nepticulidae, indicating that there were no pseudogenes involved.

The results of the phylogenetic analyses of the 355 taxa data set are summarized in Fig. 4, which presents the eight-gene Bayesian topology of the 344 ingroup-taxa

along with ML bootstrap support values for all species groups and higher branching (see Fig. S2 for the full BEAST Bayesian tree with all taxon names). Although we did not sample the Opostegidae as extensively as the Nepticulidae, with only three out of seven genera, our results show an interesting, albeit poorly supported sister group relationship between *Opostega* Zeller, 1839 and *Opostegoides* Kozlov, 1985, never seen in the earlier morphologically based phylogenies (Davis & Stonis, 2007; Davis, 1989). Nepticulidae are always monophyletic with highest support (PB 1, BS 100). Twenty monophyletic clades are interpreted as full genera. The thorough sampling allowed us to confirm many existing species groups within many of the

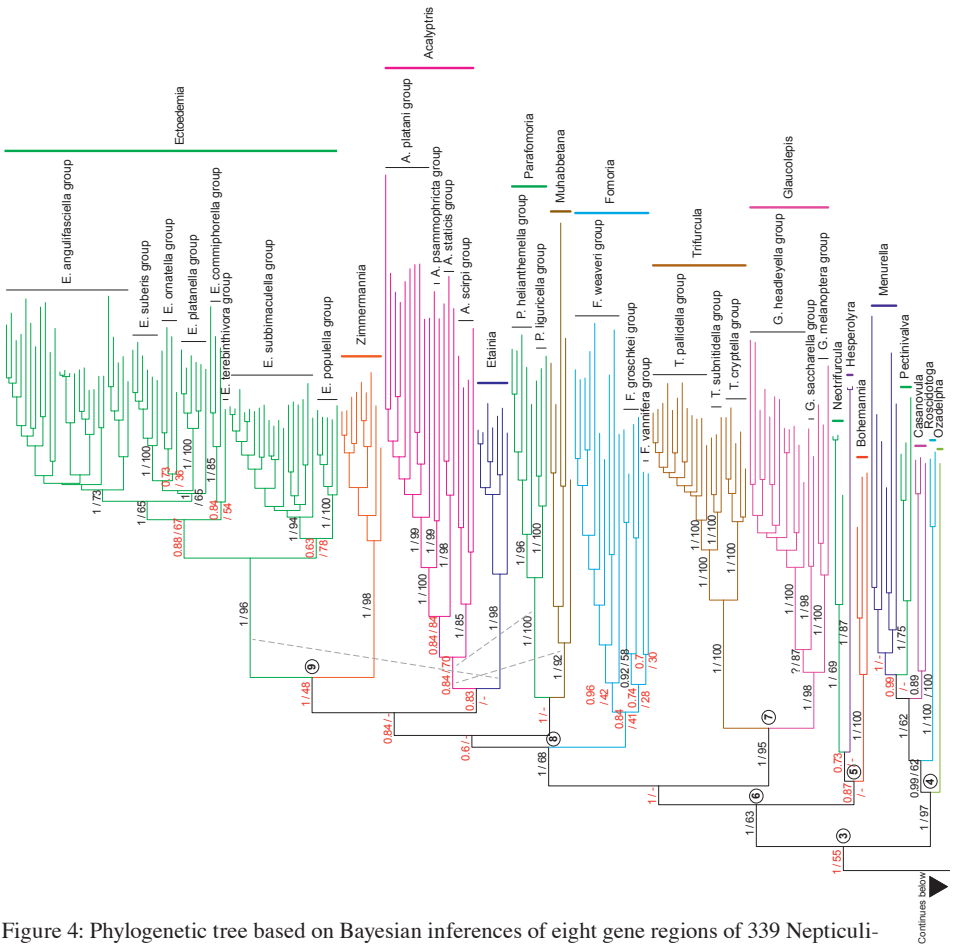
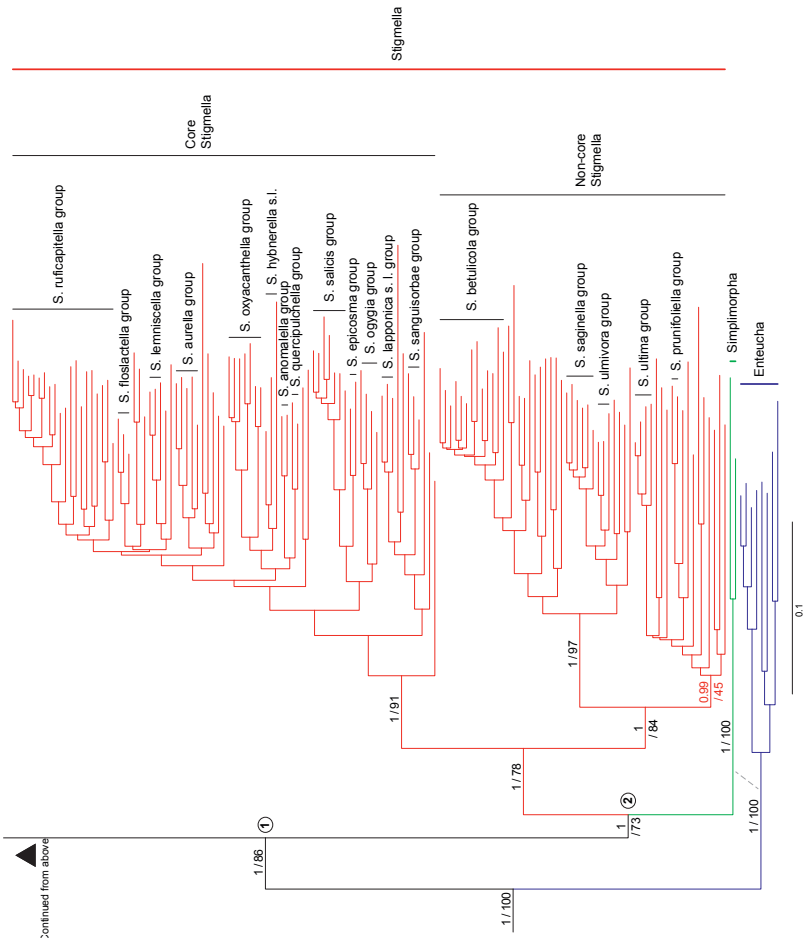


Figure 4: Phylogenetic tree based on Bayesian inferences of eight gene regions of 339 Nepticulidae exemplars, with clades collapsed according to the newly proposed classification. Support for species groups and all higher clades are indicated on the respective branches. The values on branches are Bayesian posterior probabilities, followed by maximum likelihood bootstrap support values. Support values in green are interpreted as well supported clades, support values in red as unsupported and are discussed in the text. Nodes above genus rank are numbered (in circles to the right of the node) for the purposes of discussion. Dashed grey lines indicate alternative topologies. The scale bar indicates nucleotide substitutions per site.

genera and designate some new ones. The relationships between genera are only partly well supported, but close examination of the topologies of intermediate as well as the final phylogenetic analyses allowed us to distinguish a limited number of possible distinct topologies for most cases, which we discuss and are indicated with grey dotted lines in Fig. 4.

The first division of the Nepticulidae is between the genus *Enteucha* Meyrick, 1915 and all remaining Nepticulidae (Fig. 4 node 1). Although the position of *Enteucha* is well supported (PB 1, BS 86), upon examining the results of individual analyses we encountered several bootstrap ML outcomes, as well as a BEAST Bayesian analysis that converged on a suboptimal posterior probability, where *Enteucha* is



joined on a clade with *Stigmella* and *Simplimorpha* Scoble, 1983 (Fig. 4 node 2), which would be the previously morphologically recognized group 'Nepticulini' (Fig. 1B, D). *Stigmella* and *Simplimorpha* together receives good support (PB 1, BS 75), lack of full support is due to the variable position of *Enteucha*. *Stigmella* is by

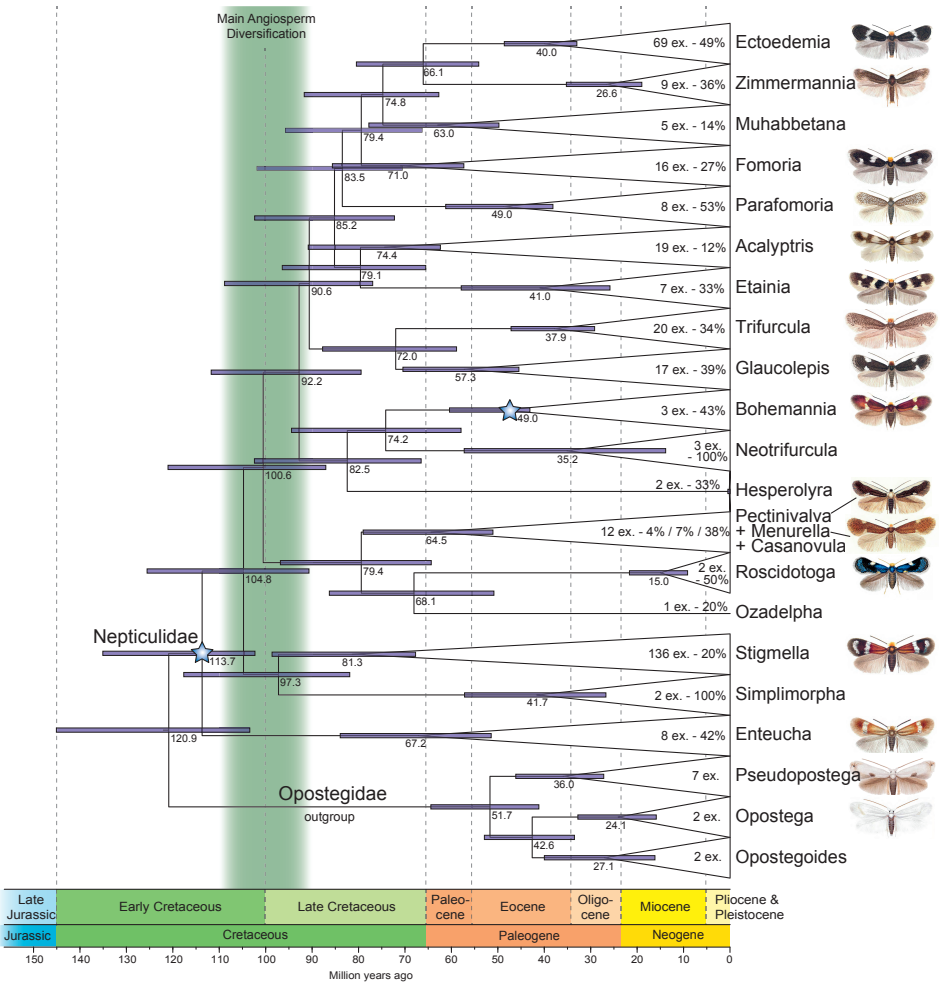


Figure 5: Divergence time estimations generated with BEAST on the 350 taxon, eight-gene dataset, using two fossil calibration nodes indicated with stars. The blue bars on the nodes indicate 95% Highest Posterior Density (HPD) confidence intervals. By the end of the Cretaceous most of the generic diversity was most likely already present. Genera have been collapsed into triangles, the tip of the triangle indicates the start of the crown clade, based on our taxon sampling. The genera *Pectinivalva*, *Menurella* and *Casanovula* were repeatedly not monophyletic in the analyses and are therefore grouped together here. The numbers in the triangles indicate how many species have been included, followed by the estimated percentage this represents from the known diversity. The known diversity includes described species, as well as undescribed species known from collections that are known to the authors. Where available, a water colour by Roland Johansson of a representative is placed next to the genus. The full tree is available in S2.

far the largest genus within Nepticulidae, with almost half of all species. From our phylogeny it is possible to subdivide *Stigmella* in two or three clades, but morphological evidence for this is incomplete at this stage. We therefore tentatively name two clades 'Core *Stigmella*', and 'Non-core *Stigmella*'. There are currently many species groups within *Stigmella*, some of which are partially or wholly recognized in our phylogeny. The well-supported groups are indicated in Fig 4.

The next group that splits off in the remaining part of the tree (Fig. 4 node 3) is comprised of the new Neotropic genus *Ozadelpha* van Nieukerken, 2016 and the Australian genera: *Roscidotoga*, *Casanovula*, *Pectinivalva* and *Menurella* (Fig. 4 node 4). In most analyses *Ozadelpha* is sister to the remaining genera, occasionally *Ozadelpha* is grouped with *Roscidotoga*. The division between *Roscidotoga* and the remaining Australian genera is well supported, but the remaining suprageneric supports are low. Although the Exabayes Bayesian analysis supports three monophyletic genera, in the ML bootstrap analyses as well as repeated BEAST analyses, *Casanovula minotaurus* (Hoare, 2013) is placed among *Menurella*, making *Menurella* paraphyletic.

The sistergroup relation between a clade of Australian genera + *Ozadelpha* (Fig 4. node 4) and the trifurculine genera (Fig. 4 node 6) is always recovered. The phylogenetic placement of the two newly discovered Neotropical genera *Hesperolyra* van Nieukerken, 2016 and *Neotrifurcula* van Nieukerken, 2016 remains uncertain. In the current results they are grouped together, but supports are lacking, likely due to a combination of long branch attraction and poor taxon sampling. Also their phylogenetic origin relative to *Bohemannia* remains uncertain (Fig. 4 node 5). The position of *Bohemannia* relative to all other genera has been stable in all analyses. *Glaucolepis* and *Trifurcula* (Fig. 4 node 7), both formerly in *Trifurcula*, form a monophylum with high support, and the supports for both genera and the species groups within these genera are high. The previously recognized subgenus *Levarchama* is now redefined as the *Trifurcula cryptella* species group.

Suprageneric relationships in the remaining part of the tree (Fig. 4 node 8) are uncertain from our current data. However, in any case, there is no support for a monophyletic *Ectoedemia* sensu lato with subgenera: *Ectoedemia, Zimmermannia, Etainia, Muhabbetana* Koçak & Kemal, 2007 (= *Laqueus*) and *Fomoria* (van Nieukerken, 1986b, Hoare 1998, Fig. 2B, D). In the presented phylogeny as well as all examined bootstrap topologies, an *Ectoedemia* sensu lato is always paraphyletic with regards to *Acalyptris* and *Parafomoria* Borkowski, 1975. The redefined genera are all recovered individually with high support, except for *Fomoria*, where the analysis possibly suffered from a relatively large amount of missing data for the sampled taxa (Fig. S1). *Ectoedemia* and *Zimmermannia* are joined without high support (PP 0.95, BS 45) (Fig 4. node 9), the only observed alternative topology groups *Ectoedemia* with *Etainia* as sister. *Acalyptris* is in the presented phylogeny joined with *Etainia*, but in intermediate results it has also been grouped with *Parafomoria* or *Muhabbetana*.

Using the recently revised fossil record for Nepticulidae (Doorenweerd *et al*., 2015a), we were able to reconstruct a time-calibrated phylogeny based on two fossil calibration points (Fig. 5). The estimated age of origin of Nepticulidae falls completely within the Early Cretaceous. Almost all of the currently recognized genera are estimated to have originated before the end of the Cretaceous, 65 Ma. In an analysis in which the BEAST run converged at a topology with suboptimal posterior probability, with *Enteucha* grouped with *Stigmella* and *Simplimorpha*, this only affected the age estimates for *Enteucha* and *Simplimorpha*, dating the split between the two at an average of 78 Ma (data not shown). There are large differences in the number of known species per genus, our sampling has been aimed to cover the diversity across the genera. The percentage of species included per genus ranges from 4% to 100%, mean 38% (Fig. 5). However, it should be noted that our estimates of diversity are minimum values, even though they include undescribed diversity that is known from collections. Especially large regions such as the Neotropics, Tropical Africa and Australia are heavily understudied. Nonetheless, our current data suggest that *Acalyptris*, *Fomoria*, *Stigmella* and possibly *Enteucha* (the latter somewhat dependent on its systematic position) had likely already begun to diversify in the Cretaceous.

Discussion

Introns in Nepticulidae

During our DNA sequencing efforts we encountered more introns than would have been expected from previous Lepidoptera-wide sequencing efforts (Mutanen *et al*., 2010). Introns in EF1-alpha have previously not been reported for Lepidoptera, but are present in most other insect orders (Djernæs & Damgaard, 2006). Recently, phylogenomic datamining has revealed that there are likely two copies of EF1-alpha in Lepidoptera, where one copy is highly fragmented by introns (Niklas Wahlberg, in litt.). The phylogenetic signal we found in EF1-alpha for Nepticulidae is congruent with other genes, leading us to believe that we either have consistently been targeting only one of the two copies with our primers, or that the two copies evolve in parallel. Either case should have no negative affect on our results, but care should be taken when our data is re-used in other studies. In the vast majority of specimens (98.7%) that we successfully sequenced for EF1-alpha, we did not encounter introns, but part of the 16% that failed to amplify may have failed due to introns, rather than primer site mismatches. During the exploration of suitable gene fragments to be sequenced, we also found an intron in GAPDH, and many introns in different positions in *Wingless*. For these genes there are no indications that there are multiple copies. The reasons for a higher abundance of introns in Nepticulidae remain unclear. The introns were found in different unrelated genera and appear to contain no phylogenetic information.

Plausibility of the divergence time estimations

There are many studies that urge caution with interpreting molecular dating results (Wilf & Escapa, 2014; Wheat & Wahlberg, 2013; Warnock *et al*., 2012), and show

that the results largely depend on the number and quality of the calibration points (Magallón *et al*., 2013). In our analysis we used only two calibration points, but we are fairly confident of their reliability and placing. With two calibrated nodes it is no surprise that the 95% HPD often exceeds 20 myr, we feel that this provides a realistic view of the current knowledge. This view holds when we compare our results with previously published molecular dating studies of Lepidoptera. Wahlberg *et al*. (2013) used the Mutanen *et al*. (2010) eight-gene genetic dataset covering all Lepidoptera, and used seven calibration points. Within Nepticuloidea they included one *Ectoedemia occultella* (Linnaeus, 1767), and one *Opostega salaciella* (Treitschke, 1833), and estimated the split between the two between 101–136 Ma. This corresponds to the split between Nepticulidae and Opostegidae, which in our analysis is estimated between 103–145 Ma. Their study also included the Dakota Formation fossil leafmines (Labandeira *et al*., 1994) as one of the calibrations. However, instead of using it to calibrate the crown Nepticulidae like in our study, it was used to calibrate the stem. A second, mitigating, issue is that the date of the Dakota Formation fossils was mistaken as 120 Ma. The Dakota Formation was at the time of the publication of the fossils dated at 99 Ma (Labandeira *et al*., 1994), but has now been readjusted to 102 Ma (see Doorenweerd *et al*., 2015a). The combination of both issues has led to comparable and similar results. There are currently no other molecular dating studies that have included Nepticuloidea, but from two other recent studies that included Lepidoptera it is clear that we are in the early stages of understanding the timing of Lepidoptera diversification. In an insect-wide transcriptome based study (Misof *et al*., 2014), genetic data from 1478 single copy nuclear genes of 144 taxa, of which ten Lepidoptera, were calibrated with 37 fossils. The crown age for Lepidoptera was estimated between 118–180 Ma, significantly younger than estimated in the study by Wahlberg, where the crown age for Lepidoptera was estimated between 200–230 Ma. In another recent study (Condamine *et al*., 2016), 874 taxa, of which 114 Lepidoptera, eight genes and 89 fossils were used and the Lepidoptera crown age was estimated even earlier, at ca. 250 Ma. Increasing the reliability of the fossil calibration points will most likely bring important progress in the coming years and will place these findings into perspective. Nonetheless, our findings for Nepticulidae appear plausible within range of estimates of the current studies.

On the use of systematic ranks

Our analyses recover the previously defined subfamily Pectinivalvinae and tribus Trifurculini as monophyletic clades (as sister groups), but we never recover a monophyletic Nepticulinae as in previous classifications (Fig. 2) and rarely a monophyletic Nepticulini. Based on our current knowledge we therefore discontinue the use of subfamily and tribe within Nepticulidae (and Opostegidae) for the time being. We also abandon the use of subgenera, which were introduced in the 1980's (van Nieukerken, 1986b; Scoble, 1983), mainly because the former large genus *Ectoedemia* (*Ectoedemia* + *Zimmermannia* + *Etainia* + *Muhabbetana* + *Fomoria*) is shown to be polyphyletic, which can only be addressed by either raising all subgenera to full genus, or by including even more genera in this large polytypic entity. We reject the latter solution because that would leave no reliable or practical apomorphies. Instead, the

newly assigned genera are almost all – except *Fomoria* – readily defined by apomorphies and well recognisable in adult and larval morphology and biology. Because the use of subgenera is often considered awkward by many users and impractical in many databases, we decided to extend this policy throughout the family and also abandon the use of subgenera in *Trifurcula* (*Trifurcula* + *Glaucolepis*) and *Pectinivalva* (*Pectinivalva* + *Casanovula* + *Menurella*). We discussed the new classification and the arguments extensively with other lepidopterists, who overall endorsed our views.

A new Nepticulidae classification

In this section, we review the agreement and disagreement of our molecular results with previous hypotheses based on morphology. We proceed sequentially through the tree in Fig. 4 using named genera or node numbers for higher clades, discuss how our findings correspond to the new classification – which is formally established in van Nieukerken *et al*. (2016a; 2016b) – and include discussion on the molecular dating results presented in Fig. 5.

Nepticulidae

Nepticulidae are morphologically well supported by at least nine apomorphies (van Nieukerken, 1986b; Scoble, 1983), including the unique sensillum type: sensillum vesiculocladum (van Nieukerken & Dop, 1987), and in the larvae the reduction of abdominal setae to six pairs per segment and the larval antenna with only two basiconic sensilla: an apomorphy not previously published (listed by Hoare, 1998). It is therefore no surprise that the monophyly of Nepticulidae is also molecularly well supported. Overall it is striking that clades that were previously well supported in the morphological cladograms, typically by at least five apomorphies, are also supported in molecular analyses and most differences are in previously poorly supported clades (Hoare, 1998; van Nieukerken, 1986b; Scoble, 1983). The reliable morphological characters often include wing venation characters, confirming their relevance for classification. However, the thickening of vein 1+2A in the forewing was formerly considered an apomorphy of Nepticulinae (*Enteucha* + nodes 2 + 6), with the Pectinivalvinae (node 4 minus *Ozadelpha*) showing the plesiomorphic, normal condition, but the character must now be considered homoplasious. The cathrema, a usually striate thickening around the opening of the ductus ejaculatorius into the vesica of the phallus in the male genitalia was previously considered as a synapomorphy for the Nepticulinae, and considered absent in the Pectinivalvinae (Hoare *et al*., 1997; van Nieukerken, 1986b) and in *Enteucha* considered to be replaced by a smooth thickening (van Nieukerken, 1986b). More recent studies had proved the cathrema to be present in Pectinivalvinae as well (Hoare and van Nieukerken, 2013; Hoare, 200b) and close examination of genitalia slides of *Enteucha* has convinced us that a striate thickening is in fact always or nearly always present, although it is usually narrow and only weakly striate, and is often obscured by cornuti. The thickening itself should be termed the cathrema, whilst the interconnected sclerites are considered to be associated structures and not part of the cathrema. The presence of a cathrema is thus an additional apomorphy for the Nepticulidae. It has been suggested (John Dugdale, in litt.) that the cathrema may represent a modified or internalised bulbus ejaculatorius, a structure that is absent in

Nepticulidae. A major previous argument for a sister group relationship between Nepticulinae and Pectinivalvinae has been the reduction of the number of antennal segments in the larva: whereas Pectinivalvinae have a 2- or 3-segmented antenna, all Nepticulinae share the reduced antenna with a single segment (Hoare & van Nieukerken, 2013; Hoare, 2000b; van Nieukerken, 1986b). The current phylogeny suggests that either this reduction happened more than once, or that segments were regained in node 4. Unfortunately we do not know yet the condition in *Ozadelpha*.

Enteucha

Enteucha is a small genus with currently 11 species described and about eight undescribed species known from collections. We were able to include eight species of *Enteucha* in our phylogeny, about 42% of the known diversity. In all morphology-based classifications, *Enteucha* has been placed in a monophyletic group with *Stigmella* and *Simplimorpha* (Fig. 2). In our analyses this relationship is not recovered, instead, *Enteucha* is the first clade to split off. In a study on non-ditrysian Lepidoptera based on data from 19 genes (Regier *et al*., 2015), three representatives of Opostegidae and seven representatives of Nepticulidae were included. In their results, as in our results, *Ectoedemia* and *Trifurcula* group together, but unlike our results, they showed high support for a clade with *Enteucha*, *Pectinivalva* and *Stigmella*, which was also the result from the morphological treatment by Puplesis (1994; Fig. 2B). Although the most likely outcomes from our data suggest otherwise, we did find indications in the bootstrap trees as well as suboptimal Bayesian topologies that there is some support in the molecular data for grouping *Enteucha* with *Stigmella* and *Simplimorpha*. The condition of the collar was previously used to characterise this grouping (former Nepticulini): comprising lamellar scales versus the plesiomorphic condition piliform scales. Lamellar scales are only known from all species in the genera *Enteucha*, *Stigmella* [although sometimes less clearly so] and have now also been found in *Ozadelpha* and in subgroups of *Bohemannia* and *Acalyptris*. It is still a useful character to recognize genera and subgroups, but does not define any larger clades in a cladistics sense. Other morphological synapomorpies that support this grouping include the larval labrum without lateral setae, pupa with only a single row of spines per segment (van Nieukerken, 1986b), and the presence of a subdorsal retinaculum, which is paralleled in *Acalyptris* (Hoare, 1998). It will probably require a combination of both many more genes as well as taxa to fully resolve this issue. By abandoning intermediate ranks between family and genus it currently does not affect the classification. *Enteucha* receives the oldest estimated stem age, between 102–135 Ma, and a crown age between 51–84 Ma. This age is puzzling in the light of the single host plant family: the Polygonaceae are considered a relatively young family (34–41 Ma: Forest & Chase, 2009). It seems unlikely that the various species colonized the Polygonaceae independently. Either the species evolved on the ancestor or extinct stem group species of Polygonaceae, or the age of the host family requires further study. Within *Enteucha*, *E. basidactyla* (Davis, 1978) from Florida is sometimes classified in a separate genus *Manoneura* Davis, 1979 (Puplesis & Robinson, 2000), but it consistently groups with the other species from Florida, *E. gilvafascia* (Davis, 1978) (both feeding on sea grape, *Coccoloba uvifera*), and both are always

subordinate in a larger clade with unnamed Asian species and the European *Enteucha acetosae* (Stainton, 1854). This supports our view that *Manoneura* has correctly been synonymised (van Nieukerken, 1986b). The rather unique characters of the male genitalia of *E. basidactyla* should therefore be considered as highly autapomorphic, and without value for generic classification. The monotypic *Varius* Scoble, 1983 from South Africa, feeding on Ochnaceae, is potentially a synonym of *Enteucha* based on morphology, but no recent material is available for DNA analyses. Without molecular evidence and because of its isolated occurrence and aberrant host-plant choice, we prefer to keep it separate for the time being. Morphologically *Enteucha* can be recognised by the reduced venation (absence of R2+3), the absence of a transverse bar in the transtilla (paralleled in *Pectinivalva* sensu lato, *Glaucolepis*, part of *Acalyptris*) and the anterior apophyses lacking anterior apodemes (van Nieukerken, 1986b). For the condition of the cathrema see above. Van Nieukerken (1986b) also listed unmelanised patches in the anterior sclerotisation of tergum 2, but this is doubtful as the amount of melanisation can vary and an insufficient number of species of Nepticulidae have has been screened for this character.

Node 2: Stigmella *and* Simplimorpha
In all analyses *Simplimorpha* is consistently placed as sister to *Stigmella*, with an estimated divergence time between the two in the Cretaceous, between 82–118 Ma. *Stigmella* and *Simplimorpha* share the venation, with curved main trunk of R+M usually with four terminal branches and a separate Cu. Otherwise there are no obvious morphological apomorphies, unless *Enteucha* would be included in this clade (van Nieukerken, 1986b).

Simplimorpha
There are only two species of *Simplimorpha* known, with clear allopatric distributions, and both have been sampled in our phylogeny. There are no undescribed species known from collections, but given the poor knowledge on the African fauna there is potential for undiscovered diversity. Both known species are oligophagous on Anacardiaceae. *Simplimorpha promissa* feeds on *Pistacia* species, *Rhus* species and *Cotinus coggygria* in the Mediterranean area, and *S. lanceifoliella* (Vári, 1955) feeds on many *Searsia* species (formerly in *Rhus*), *Protorhus longifolia* and introduced *Schinus molle* in Southern Africa (Scoble, 1983). All specimens of *S. promissa* that we sequenced for EF1-alpha contained an 80 bp intron (Fig. 3), which was not present in *S. lanceifoliella*. The estimated divergence time between the two species is in the Paleogene, between 27–57 Ma. This is the oldest estimated divergence time between two sister-species in the phylogeny (see Fig. S1).

Stigmella
Stigmella is by far the largest nepticulid genus, containing nearly half of all the known species (currently 420) and can be found on all continents except Antarctica. From our current sampling and estimations, the crown *Stigmella* age is estimated to be Late Cretaceous, 68–99 Ma, making it one of the earliest diverging genera. For such a large and well-supported genus, it has remarkably few morphological

apomorphies: van Nieukerken (1986b) only listed two: uncus bilobed (bifid) and larval antenna with sensilla placed cross-wise. To these we can add: collar with lamellar scales, although sometimes appearing hairy, and this is paralleled in *Enteucha*, *Ozadelpha*, part of *Bohemannia* and *Acalyptris*. There are several exceptions to the bilobed uncus, e.g. *Stigmella naturnella* (Klimesch, 1936). The previously used character "gnathos with single posterior process versus two processes", e.g. to separate the *lapponica* group as basal from the rest, appears to be very homoplasious; the previously considered plesiomorphic condition, a single process, similar to most other Nepticulidae, occurs in many groups throughout the genus and is possibly not always the plesiomorphic condition, but may have originated from fusion of the two processes. This is supported by New Zealand species, which exhibit a range of degrees of fusion, with *S. tricentra* (Meyrick, 1889) showing an intermediate state with processes fused basally and contiguous apically (Donner & Wilkinson, 1989: fig. 97, an accurate representation). From the molecular phylogeny it is clear that *Stigmella* can be subdivided in two [or three] large clades, which may warrant generic status. However, without sufficient morphological evidence to support this, it seems impractical to do so at this point and instead we have termed them here as 'Core *Stigmella*' and 'Non-core *Stigmella*'. One character that seems to follow the division in two clades is the feeding position of the larva: only larvae of Core *Stigmella* feed with their dorsum upwards, whereas all other Nepticulidae feed with their venter upwards. This is probably a strong apomorphy for the Core group, but impractical for classification purposes when larvae are unknown. Our results recover the following species groups with sufficient support: in Non-core *Stigmella*: the *S. prunifoliella*, *S. ultima*, *S. ulmivora*, *S. saginella* and *S. betulicola* groups, and in Core *Stigmella*: the *S. sanguisorbae*, *S. lapponica* sensu lato, *S. ogygia*, *S. epicosma*, *S. salicis*, *S. quercipulchella*, *S. anomalella*, *S. hybnerella* sensu lato, *S. oxyacanthella*, *S. aurella*, *S. lemniscella*, *S. floslactella* and *S. ruficapitella* groups (Fig. 4). In the genitalia, Non-core *Stigmella* males usually have a wide uncus, sometimes shallowly bilobed, a juxta is often present, a manica never; the phallus often has few cornuti. In Core *Stigmella* the uncus is either bifurcate or sometimes deeply split into four lobes, and a juxta is rarely present. In a large clade including the *S. aurella* and *S. ruficapitella* groups there is often a manica (phallocrypt) around the phallus, which usually has many cornuti. In Non-core *Stigmella*, female genitalia often have spiny signa, paired (e.g. *S. betulicola* group, some in *S. ultima* group) or unpaired (*S. saginella* group, some species in the *S. paliurella* group), accessory sacs are rare and when present rather small. Core *Stigmella* females usually have a strong accessory sac, that in part of the *S. ruficapitella* group clearly takes over the function of the reduced corpus bursae (van Nieukerken & Johansson, 2003). Several species groups in both clades are specialised on host plant families such as Fagaceae, Betulaceae, Rosaceae and Rhamnaceae. Most tropical species are placed in Non-Core *Stigmella* and feed on such host families as Fabaceae, Moraceae, Euphorbiaceae, Phyllanthaceae, Meliaceae, Rutaceae, Dipterocarpaceae and Malvaceae.

Notable is the well supported grouping of all examined New Zealand species (the *S. ogygia* group), the South American members formerly placed in an extended *S.*

salicis group (Puplesis & Robinson, 2000), for which we propose the name '*S. epicosma* group' and a strict *S. salicis* group (those feeding on Salicaceae). Overall the males in this clade have rather similar genitalia, and females have a non-coiled ductus spermathecae as potential synapomorphy. This clade splits first between the New-Zealand *S. ogygia* group and the *S. epicosma* + *S. salicis* groups, that could be suggested to have a Neotropic origin, since the Neotropic *Salix* feeder *Stigmella molinensis* van Nieukerken & Snyers, 2016 appears to be sister to all other *S. salicis* group members. This geographic distribution can only be explained by long distance dispersal, since at the average estimated age of this split (Eocene, 41.5 Ma), New Zealand and Latin America were even further apart than today. It is also notable that all Asteraceae feeding *Stigmella* species are in this clade: both in New Zealand and Latin America, even though both groups also use several other host plant families (Puplesis & Robinson, 2000; Donner & Wilkinson, 1989).

Other new findings include an expanded *S. lapponica* group, containing several Rosaceae feeders, including *S. malella* (Stainton, 1854) and *S. slingerlandella* (Kearfott, 1908) (Fig. S1) next to the originally included Betulaceae feeders, and another assemblage of groups feeding on Rosaceae: the *S. hybnerella*, *S. paradoxa* and *S. irregularis* groups, that we here combine as the *S. hybnerella* sensu lato group, even though genitalia characters are rather diverse. Several other species groups that have previously been recognized were not or only partly recovered or we sampled only one species or none at all. Where applicable, they are further discussed in the catalogue (van Nieukerken *et al*., 2016a). Future studies on *Stigmella* that include more taxa globally, particularly from tropical regions, will be needed to get a better grip on the evolutionary history of this genus.

Node 3
This new grouping represents the former Pectinivalvinae and Trifurculini, and is better supported in the Bayesian analysis than in the ML analysis (PP 1, BS 55). As earlier studies did not recognise this grouping, no morphological apomorphies have been noted previously, nor can we easily point out any now.

Node 4: Ozadelpha *and the Australian genera*
Ozadelpha has been newly described from the Neotropics (van Nieukerken *et al*., 2016b) and at present contains at least four species, of which one is included in our phylogeny: *Ozadelpha* specimen EvN4680, which is closely related to the type species *Ozadelpha conostegiae* van Nieukerken and Nishida, 2016. Both species feeds on *Conostegia* (Melastomataceae) in Costa Rica. Another species feeds on Myrtaceae, and has been recombined by us as *Ozadelpha guajavae* (Puplesis & Diškus) (van Nieukerken *et al*., 2016b). The stem age for *Ozadelpha* is estimated at 51–86 Ma. We find it likely that more species of this genus are to be discovered in the Neotropics. Currently all known species have two fasciae on the forewing, and a collar comprising lamellar scales. In the venation *Ozadelpha* resembles *Stigmella*, but Cu is usually very long, and since also the condition of the collar is similar to *Stigmella* an external diagnosis is difficult. The long Cu resembles that in *Roscidotoga* and also in the genitalia both genera share several characters: the

large vinculum, bilobed uncus, and broadened anterior apophyses. However, *Ozadelpha* does not share any of the listed adult or pupal synapomorphies of the Australian genera (Hoare, 2000b). Larvae have not yet been studied. The grouping of *Ozadelpha* with all Australian nepticulid genera: *Roscidotoga*, *Pectinivalva*, *Menurella* and *Casanovula* is well supported (PP 1, BS 97). If the hosts for the *Ozadelpha* species we currently know are representative for the genus, the majority of species in node 4 are associated with Myrtales.

A major character that gave its name to the genus *Pectinivalva*, including *Casanovula* and *Menurella*, is the valval pecten that was considered homologous to the stalked pectens in Opostegidae and various Adeloidea (van Nieukerken, 1986b; Scoble, 1983). In the light of our findings it is more likely that the pecten in most *Casanovula*, *Menurella* and *Pectinivalva* species is a neoformation, similar to a type of pecten found in several *Acalyptris* species (see also: Regier *et al*., 2015a). The morphological characters supporting a grouping of *Pectinivalva*, *Menurella*, *Casanovula* and *Roscidotoga* and for each of these genera are detailed by Hoare (2000b) and Hoare and van Nieukerken (2013). Although morphological characters appear reliable and support a generic status for all groups, instead of the previous subgeneric status, monophyly for all genera is not always recovered in the molecular results. This is likely a sampling issue: the estimated sampled diversity for these genera is between 4 and 38%, and especially low for *Pectinivalva* (4%) and *Menurella* (7%). Both are expected to have a diversity in the order of 70–80 species based on counts in the Australian National Insect Collection (Hoare, 1998). In the molecular dating analyses, one species of *Casanovula* is consistently recovered among *Menurella*, and *Roscidotoga* is grouped with *Ozadelpha*. The age estimates for these clades are therefore combined: the stem age for *Pectinivalva* + *Menurella* + *Casanovula* is estimated between 64–97 Ma. The three genera likely diverged after the K-Pg boundary, in the Paleogene, somewhat later than most other genera, which originated in the Cretaceous. Hoare and van Nieukerken (2013) stressed the fact that some species of *Menurella* and *Casanovula* are still specialised on rain forest dwelling Myrtaceae, whereas all *Pectinivalva* species, where known, feed on drought-resistant *Eucalyptus*, a genus that diversified particularly after Australia's Miocene aridification (5–24 Ma). Still, we see that splits between these genera in our estimate predate this drying out, with the estimate for the stem-group of *Pectinivalva* between 27–50 Ma. For *Roscidotoga* the stem age is here estimated between 51–86 Ma, but there is some added uncertainty due to its variable position relative to *Ozadelpha*. While Hoare (2000b) still suggested a split between *Pectinivalva* sensu lato and *Roscidotoga* in the Early Cretaceous, later Hoare and van Nieukerken (2013) assumed that this split dates closer to the Miocene aridification; the latter is supported by our data.

Node 6: trifurculine genera

The remaining twelve genera form a well-supported monophyletic group (PP 1, BS 63). Previously this grouping has been classified as the Trifurculini or Trifurculinae (Fig. 2) and is well supported morphologically (van Nieukerken, 1986b), amongst others by the character "Veins Rs and M of forewing separate basally, forming

closed cell" (van Nieukerken, 1986b). This character is also valid for the clade *Hesperolyra* + *Neotrifurcula* + *Bohemannia*, where it is apparent in *Neotrifurcula*, but reduced in the other two genera. The character "paired reticulate signa in female bursa copulatrix" is another character that is almost always present in this node, but probably secondarily reduced in groups with a reduced bursa (viz. *Hesperolyra*, *Parafomoria* and the *Acalyptris staticis* group). The origin of this group is securely estimated in the Cretaceous.

Node 5

The grouping of *Bohemannia* with the two new Neotropic genera *Hesperolyra* and *Neotrifurcula* is not well supported, and also morphologically these genera have little in common (van Nieukerken *et al*., 2016b). The position of the new genera remains unclear, and these taxa often acted as 'rogue taxa' in our analyses and could be recovered in very different parts of the tree. Both genera probably have many more species in the Neotropics. Age estimates for these remain uncertain, but they are likely of Cretaceous origin.

Bohemannia

This small Palearctic genus comprises one leafmining species (*Bohemannia pulverosella* (Stainton, 1849), on *Malus*), whereas the remainder of the total seven species are probably shoot- or budminers, although rearing records are mostly absent (van Nieukerken & Johansson, 1990). Hosts are known through association and at least one reared specimen from Betulaceae, without knowledge of the larval feeding (*Alnus*, *Betula*) (van Nieukerken, 1986a). Recently, two fossil species were recognised from Baltic Amber (Fischer, 2013), which allowed us to use these as a calibration point for divergence time estimations. The stem age is estimated as 58–94 Ma, but this is also somewhat uncertain due to the problematic placing of *Hesperolyra* and *Neotrifurcula*. The phylogenetic placement of *Bohemannia* relative to the other genera in node 6 as found here is similar to that in Puplesis (1994), whereas van Nieukerken (1986b) regarded it as sistergroup to *Ectoedemia* sensu lato on the basis of two larval characters: viz. the shape of frontoclypeus and length of tentorial arms. Given that *Ectoedemia* is no longer recovered as monophyletic in a sensu lato composition, these characters are likely unreliable for classification.

Node 7: Trifurcula *and* Glaucolepis

The new classification matches that of Puplesis (1994; Fig. 2C) in assigning generic status to *Trifurcula* and *Glaucolepis*. Other classifications have treated a larger *Trifurcula*, with subgenera *Trifurcula*, *Glaucolepis* and *Levarchama* (Fig. 2B, D). The previously recognised subgenus *Levarchama* is not raised to full genus, but treated here as the *T. cryptella* species group. Phylogenetically this entails no change, and the group is strongly supported by the combination of three probably interrelated characters: R+M with three branches in hindwing, male abdomen with paired tufts on T6–8 and the underside of the male hindwing with a velvet patch of androconial scales.

Glaucolepis

The genus *Glaucolepis* comprises currently ca. 40 named species, of which the majority is known from the Mediterranean region. Most are leaf or stem-miners in various trees and shrubs, a few feed in herbs and probably some make galls. Three apomorphies were given by van Nieukerken (1986b), and a fourth was suggested by van Nieukerken and Puplesis (1991), while adding some doubt to two of the previous apomorphies in the male genitalia. Our results recognise three well supported species groups: the *G. raikhonae* group comprises three Palearctic species that are probably all gall-makers on Rosaceae, although this is only documented for *G. oishiella* (Matsumura, 1931) (=*Sinopticula sinica* (Yang, 1989)), that feeds on *Prunus* (van Nieukerken & Puplesis, 1991; Yang, 1989). The group has a wider distribution, as we already know unnamed species in this group from Australia and North America. The *G. saccharella* group contains the type species *G. saccharella* (Braun, 1912) from Eastern North America, a leafminer of *Acer* (Sapindaceae) and an unnamed species from Japan that makes leafmines on various woody Fabaceae. The *G. headleyella* group is the most diverse and is confined to Europe, the Mediterranean and adjacent areas. Many species in this group mine in more than one leaf and continue to the next leaf via the petiole and stem, or sometimes only mine in the stem. Most of these feed on Lamiaceae, several on Plantaginaceae (*Globularia*) and Apiaceae (*Bupleurum*) and a few other families (Laštůvka *et al*., 2013; Ivinskis *et al*., 2012; Laštůvka & Laštůvka, 2007; 2000). The stem age for *Glaucolepis* is estimated as 59–88 Ma, the crown age 45–70 Ma. For the *G. raikhonae* group (three species) the crown age is 16–40 Ma, for the *G. saccharella* group (two species) 19–39 Ma and for the *G. headleyella* group (12 species) 25–40 Ma. The proliferation of the last group in the Mediterranean region, where species specialised on small drought resistant shrubs, now common in Mediterranean habitats such as maquis and garrigue, is probably partly explained by the Oligocene–Miocene aridification of the region (Dong *et al*., 2013). We were unable to study substantial DNA data of the two Neotropic *Glaucolepis* species, but morphologically they are so different from other *Glaucolepis* that it is very unlikely they are placed correctly. At least *G. argentosa* Puplesis & Robinson, 2000, of which we examined one male morphologically and have a partial DNA barcode, lacks the three apomorphies of *Trifurcula* + *Glaucolepis* and genitalia only show some superficial similarities. It is not unlikely that these species represent another new genus.

Trifurcula

The genus *Trifurcula* as defined here is largely a western Palearctic genus, with 36 named species and several unnamed species. The genus and its three species groups are all recovered with highest support, congruent with earlier morphological analyses (van Nieukerken, 2007; 1990; 1986b). The best morphological apomorphy is the loss of the connection between R2+3 and R4+5 in the forewing and another apomorphy is the hostplant family: Fabaceae, for which no exceptions are known. The *T. cryptella* group (seven species) comprises leaf miners on herbaceous and shrubby Fabaceae: Loteae, all in Europe and North Africa, including Macaronesia. The group was revised and its phylogeny analysed by van Nieukerken (2007). The *T. subnitidella* group (eight species) comprises European and Mediterranean species

that make stem mines, occasionally starting in a leaf, also in herbaceous and shrubby Fabaceae, tribes Loteae and Hedysareae (van Nieukerken, 1990). The remaining ca. 20 named species and several unnamed species belong to the *T. pallidella* group, and all are stemminers of Fabaceae belonging to the tribe Genisteae in the Mediterranean region (van Nieukerken *et al*., 2010; Laštůvka & Laštůvka, 2005; van Nieukerken *et al*., 2004c; Laštůvka & Laštůvka, 1994). Our phylogeny supports the evolutionary scenario sketched by van Nieukerken (2007): the ancestor of *Trifurcula* in the present sense was a leafminer on Loteae, and in the clade of the *T. subnitidella* + *pallidella* groups the stem mining habit evolved; in the *T. pallidella* group a shift to Genisteae as hostplant occurred. Scoble (1980) added two South-African species to *Trifurcula*, of which the position is unknown. We doubt whether they are genuine *Trifurcula*, but a further molecular analysis is required to answer this doubt. The stem age of *Trifurcula* is 59–88 Ma, the crown age 29–47 Ma. As in *Glaucolepis*, the proliferation of species in this genus can possibly be partly attributed to the Oligocene-Miocene aridification of the region (Dong *et al*., 2013).

Node 8: seven genera with unclear suprageneric relationships

This clade comprises the genera *Parafomoria*, *Acalyptris* and the former subgenera of *Ectoedemia*: *Fomoria, Muhabbetana, Etainia, Zimmermannia* and *Ectoedemia*. The estimated origin is Late Cretaceous (72–102 Ma), and they likely diverged in a relatively short time, leading to short branches and lacking supports. Alternative topology hypotheses are indicated with grey dotted lines in Fig. 4. Most of the genera are well defined from morphology and life history characters, but no morphological apomorphies can be indicated for the whole clade.

Fomoria

Fomoria is the genus that stands out as being not well defined either molecularly or morphologically, an issue noted previously (van Nieukerken, 1986b; Scoble, 1983). The absence of distinct apomorphies has probably also lead to the previous inclusion of species that are now in *Hesperolyra*, on grounds of some similarity in male genitalia (Puplesis & Robinson, 2000). In total 48 named species belong to *Fomoria*, with the largest number, 22 species, in Africa. In our analysis, we incorporated three species groups, although all receive poor support. The *F. weaveri* group with 14 named and several unnamed species comprises mainly leafminers on *Hypericum*, Ericaceae and Rutaceae, and most species pupate inside the leafmine. The species are characterised by a number of apomorphic genitalia characters (van Nieukerken, 2008; Hoare, 2000a). The *Hypericum* and Ericaceae feeding species form a clade, sister to the Rutaceae feeding species. As sister to these groups we see an Australian species reared from *Scolopia* (Salicaceae), the very species that one of us discussed as the first potential Southern hemisphere member of the *F. weaveri* group, based on a single specimen, before the mine and larva were discovered by us at the same locality (Hoare, 2000a). The *F. groschkei* group (three named species from Europe and Africa, several unnamed Asian ones) comes as poorly supported sister in some of our analyses or as paraphyletic. It comprises several species feeding on woody Lamiaceae that were previously placed in Verbenaceae (*Vitex, Callicarpa*), but we also found one species on

Bignoniaceae (*Radermachera*) in Taiwan. Unfortunately we were only able to sequence one species of the morphologically and biologically well supported *F. vannifera* group (Hoare, 2000a): *F. vannifera* (Meyrick, 1914) itself. They are leafminers on Brassicaceae: *Capparis* and relatives. Here it groups, without support, with the isolated Apiaceae feeding species *F. viridissimella* (Caradja, 1920). The stem age for *Fomoria* is estimated as 66–96 Ma, the crown age: 57–86 Ma, showing overlapping estimates. The genus originated already in the Cretaceous.

Muhabbetana

Muhabbetana, in most literature known as *Ectoedemia* (*Laqueus*), is an essentially African clade with 32 named species, with a small group of species occurring in the Mediterranean region and Macaronesia. The Afrotropical species feed on Ebenaceae and Celastraceae, whereas the Mediterranean species feed on Euphorbiaceae (genus *Euphorbia*) and Apocynaceae. The genus was previously synonymised with *Fomoria* (van Nieukerken *et al*., 2004b; Puplesis, 1994), but in our analyses the species ascribed to it always group separately. Morphologically it is characterised by the anal loop in the forewing and the larval stipes with two setae (van Nieukerken, 1986b; Scoble, 1983), but the last character has only been checked for Mediterranean species. Unfortunately, we have not been able to include the type species *M. grandinosa* (Meyrick, 1911) or a close relative in our analyses, but we include one unidentified African species (RMNH.INS.24076) and African material studied after finishing these analyses (not included here) also confirms this placement. The stem age is estimated as 63–92 Ma, the crown age: 50–78 Ma.

Parafomoria

The genus *Parafomoria* is a small group with eight named, and ca. seven unnamed species, occurring only in the Mediterranean region with one species going north into Central Europe. Most species occur in the Iberian Peninsula. All species make leafmines on shrubby Cistaceae, often in winter. Morphologically the genus can be recognised by the reduced venation (loss of R3), expansion of the lateral arms of the vinculum, reduction of female bursa and development of male hair pencil. There is also an apomorphy in the 28S gene: a large gap in the D2 region occurs in all species and can be ascribed to the shortening of one loop in the secondary structure. The two groups recognised in the molecular analysis conform to the morphological division into two groups (van Nieukerken, 1983). The age for crown-group Cistaceae is around (18.5–)14.2(–10.2) Ma (Guzmán & Vargas, 2009; Vargas *et al*., 2014). Our estimates for the crown group *Parafomoria*, are much older, between 38 and 61 Ma. Whether this is due to unknown host relationships in *Parafomoria* with ancestors of the Cistaceae and its sistergroup Dipterocarpaceae, or to inaccuracy in the estimated age of Cistaceae is unclear. The age estimate for the stem of *Parafomoria* is 71–102 Ma.

Etainia

The genus *Etainia* with 16 named and five unnamed species, is both molecularly and morphologically well supported (van Nieukerken & Laštůvka, 2002; Puplesis & Diškus, 1996). As far as known none of the species are leaf miners, but feed in

various ways in other organs than leaves. Several Holarctic species are associated with *Acer* (Sapindaceae) and either feed in the winged fruits and seeds (in the summer generation) and in the buds and shoots (winter generation or univoltine generation) (Emmet, 1984; Johnson, 1982; Kulman, 1967), but *E. albibimaculella* (Larsen, 1927) and an unnamed North American species feed on Ericaceae, and the possibility that *E. obtusa* Puplesis & Diškus, 1996 feeds on *Fraxinus* (Oleaceae) cannot be excluded (van Nieukerken & Laštůvka, 2002). Hosts for the African species are unknown, but *Acer* does not occur there. The included African species is sister to the clade of Holarctic species, at some distance, suggesting an African origin. Estimated stem age: 66–96 Ma, crown age: 26–58 Ma.

Acalyptris

The genus *Acalyptris* is the second largest genus after *Stigmella*, with 93 named and more than 60 unnamed species. It is widespread especially in the (sub)tropical and desert regions of the world, but in North America species can be found at higher latitudes in bogs and wetlands, feeding on Cyperaceae (EJvN, personal observations). Most species are leafminers, some mine stems, and hosts are varied. There are records from at least 21 plant families, mostly Eudicots, but also the monocot family Cyperaceae. There is one morphological apomorphy for the genus: Closed cell in forewing shifted towards base, vestigial. The 19 species in our analysis group in four relatively well supported species groups: the New World *A. scirpi* group, the West Palearctic *A. staticis* group, specialised on Plumbaginaceae, the Palearctic *A. psammophricta* group – formerly also known as the *repeteki* group – that occurs primarily in desert and steppe areas of Central Asia, the Middle East and North Africa and the *A. platani* species group that is confined to the Old World. The estimated stem age is: 66–96 Ma, crown 62–91 Ma. Like in *Fomoria*, also in *Acalyptris* there is a large overlap in the estimates of the crown and stem groups, suggesting an early diverging genus.

Zimmermannia

The small genus *Zimmermannia* is restricted to the Holarctic region with 17 named species and ca. eight unnamed ones, of which nine species are included in the analysis. The species are bark miners where known, with most species associated with Fagaceae, but a few with Ulmaceae, Salicaceae and possibly Betulaceae (EJvN, unpublished information). Whereas old mines are usually easily seen, finding and collecting larvae is a challenge and most species are only known from light collected adults. Morphologically the sister group relation with *Ectoedemia* is well supported (van Nieukerken, 1986b), which is also the best supported molecular scenario. The estimated stem age is 54–80 Ma, the crown age 19–35 Ma.

Ectoedemia

Ectoedemia is the third largest genus in Nepticulidae with 90 named and more than 50 unnamed species, especially in East Asia. It has a wide distribution in the northern Hemisphere, and is particularly diverse in the Palearctic region. The phylogeny of *Ectoedemia* has already been treated in detail in a previous study (Doorenweerd *et al*., 2015b) and will not be repeated here. Different in the present

analysis from the published phylogeny is that the clade with the African species, the *E. commiphorella* species group, is not the first to split off, but instead splits off after the clade with the *E. subbimaculella* and *E. populella* species groups. Considering the low support for this, its real position remains uncertain. The stem age for *Ectoedemia* is estimated to be between 54–80 Ma, either just before or just after the K-Pg boundary. The crown age is more securely estimated, between 33–49 Ma, Eocene. Because the centre of diversity for *Ectoedemia* is in the Palearctic and we have a thorough sampling coverage of all species groups, we believe this is one of the most reliable estimates in our dataset. All species in the *E. populella* group feed on Salicaceae, and started diversifying between 14.3–24.4 Ma, Miocene (see Fig. S1). The crown age for Salicaceae is generally estimated to be much older, with the most recent estimates in the Cretaceous, between 73–87 Ma (Xi *et al*., 2012). The crown age for the *E. angulifasciella* group, for which the ancestral host was most likely a Rosaceae (Doorenweerd *et al*., 2015b) is estimated between 25.1–36.5 Ma. The estimated crown age for Rosaceae is just after the K-Pg boundary, around 61 Ma (Hohmann *et al*., 2015), or in the Cretaceous, around 88 Ma (Chin *et al*., 2014): in either case, long before the *E. angulifasciella* group started to diversify. The other species groups in *Ectoedemia* receive similar crown age estimates, ranging from 12–38 Ma, covering the late Eocene, Oligocene and Miocene.

Biogeographic Notes

Most genera and higher clades are distributed throughout several biogeographical regions without clear evolutionary patterns, but there are some splits that may correlate with the shifting of tectonic plates (Fig. 5). In particular the strongly supported relationship between *Ozadelpha*, which appears restricted to the Neotropics, and all Australian genera (Fig. 4 node 4) suggests ancient vicariance. The split between these two is estimated to have occurred in the Late Cretaceous, where all southern continents were practically merged and slowly broke up. However, other splits in the phylogeny indicate that long distance dispersal has likely occurred multiple times throughout the evolutionary history of Nepticulidae and cannot not be ruled out for the *Ozadelpha*-Australian relationships. For example, the estimated age of the monophyletic clade of *Stigmella* species from New Zealand (*S. ogygia* group) is estimated at 29–46 Ma, whereas New Zealand has been isolated for ca. 80 myr (Waters & Craw, 2006). To further understand the role of Gondwanan vicariance it would be important to target future sampling on southern Hemisphere areas such as New Caledonia, which has an extremely diverse flora including many ancient angiosperm groups. Other general biogeographic trends observed in the phylogeny include the repeated apparent dispersal from Africa to different parts of the Northern Hemisphere: in *Simplimorpha*, *Muhabbetana*, *Etainia* and *Ectoedemia*. To further understand this pattern it will be vital to include further sampling of African taxa. In terms of species numbers, two of the three largest genera, *Stigmella* and *Ectoedemia*, representing ca. 2/3 of the Nepticulidae diversity, appear to have diversified primarily in the temperate regions. Within *Stigmella*, the predominantly tropical clades include less species and older splits, in contrast to the clades mostly distributed in the temperate region. Overall the biogeographical region spurring the most spectacular finds at present is

the Neotropics, with just three new genera already included in this paper from a very fragmented and incidental sampling, leaving undoubtedly more to be found. Other groups of leafminers have also been shown to have a large undiscovered Neotropical diversity, including at generic levels (Lees *et al*., 2014; Gilson de Moreira, in litt.).

Contemporaneous evolution with Angiosperms

The molecular chronogram presented in Fig. 5 allows us to make a first comparison of the timing of diversification of Nepticulidae with respect to that of the Angiosperms. The main Angiosperm radiation has been estimated to have occurred between 90–110 Ma, during which most plant lineages split into groups we now mostly refer to as families (Silvestro *et al*., 2015; Magallón *et al*., 2013; Schneider *et al*., 2004), and has been suggested as an important driver of speciation in Lepidoptera before (Wahlberg *et al*., 2013). Our molecular dating results show that it is likely that Nepticulidae were already present before this time, meaning that there was ample opportunity for coevolution in a broad sense during the main Angiosperm diversification in the Cretaceous and thereafter. The pattern of host use we observe in Nepticulidae does not mirror the phylogeny of Angiosperms in any way. At most we can say that the majority of species feed on plants in the fabid (Eurosid I) clade, but there are many exceptions to that rule. We do find three genera that are restricted to a single host plant family, viz. *Enteucha* on Polygonaceae, *Trifurcula* on Fabaceae and *Parafomoria* on Cistaceae. With two of those, viz. *Enteucha* and *Parafomoria*, the estimated stem age of the leafmining genus is older than the estimated crown age of the host family, indicating that there has likely been a different ancestral host, or that either the dating estimates of the insects or the hosts are inaccurate. Recent studies on Angiosperms have focussed more on the diversification, rather than just stem and crown ages (Bouchenak-Khelladi *et al*., 2015; Xing *et al*., 2014). This reveals that in Fagales and Rosales, predominant host orders for Nepticulidae, diversification has intensified during the Miocene. Combined with the observation that most nepticulid species specialize on just one or several closely related host species, it is likely that many of the host relations we observe at species or species group level have been established mostly during the Miocene. It will be particularly interesting in future studies to include fine scale host use data and compare this with the diversification estimates of the hosts.

Acknowledgements

First and foremost, we are much indebted to everyone who has contributed specimens for our analyses, including (in alphabetical order): Leif Aarvik, James Adams, David Agassiz, Val Albu, Kees Alders, Giorgio Baldizzone, Bengt Å. Bengtsson, Kees van den Berg, Oleksiy V. Bidzilya, Willy Biesenbaum, Richard Brown, Sifra Corver, Don Davis, Jurate and Willy De Prins, Charley Eiseman, Willem Ellis, Cees Gielis, Robert Heckford, Douglas Hilton, Martin Honey, Michael Hull, Povilas Ivinskis, Jari Junnilainen, Lauri Kaila, Ole Karsholt, Sjaak Koster, Mikhail Kozlov, John Langmaid, Aleš and Zdeněk Laštůvka, Houhun Li, Anna Mazurkiewicz, Wolfram Mey, Marko Mutanen, Kenji Nishida, Greg Pohl, Lynn Raulerson, Jadranka Rota, Rudi Seliger, Chris Snyers, Kazuhiro Sugisima, Reinhard Sutter, Rumen Tomov, Paolo Triberti, Hector Vargas,

David L. Wagner, Andreas Werno, Hugo van der Wolf, K. Yazamaki, Shen Horn Yen, Michael Zerafa and Vadim Zolotuhin. We would like to thank our colleagues in the molecular laboratory of the Naturalis Biodiversity Center: Dick Groenenberg, Frank Stokvis, Kevin Beentjes and Marcel Eurlings for their help and hints. We are indebted to Kees van den Berg for the many ways he assisted us, e.g. with rearing, mounting of moths, as well as optimizing the methods for preparing genitalia and larvae. We very much appreciate that Roland Johansson and the Swedish Species Information Centre permitted us to use the accurate water colour drawings of Nepticulidae for Fig. 5, and we appreciate George Sinnema allowing us to use his excellent photographs of live Nepticulidae in Fig. 1. We would like to thank Niklas Wahlberg for making us aware of the copy of the elongation factor 1-alpha gene and discussing this matter, and we thank Marie Djernaes for discussing the relevance of introns in elongation factor 1-alpha and her help with determining the exact position of these. We thank John Dugdale for sharing his insightful ideas on the cathrema and Gilson Moreira for sharing his information on newly discovered Gracillariidae genera in the Neotropics. Hugo van Duijn and Atze de Vries of the IT department of Naturalis were very helpful with compiling multithreaded versions of phylogenetic analysis software in an OpenStack cloud computing environment, which allowed us to perform computationally demanding analyses. Our work was made possible by financial support from the Dutch government Economic Restructuring Fund (FES) as well as multiple collecting expeditions funded in part by the Uyttenboogaart Eliasen Foundation. A Naturalis Temminck Fellowship to RJBH in 2009, to finalise publications out of his PhD thesis, lead the authors to their discussions on the nepticulid phylogeny that eventually lead to this collaborative work.

References

Aberer, A.J., Kobert, K., & Stamatakis, A. (2014) ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

APG III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, **161**, 105–121.

Bazinet, A.L., Cummings, M.P., Mitter, K.T., & Mitter, C.W. (2013) Can RNA-Seq Resolve the Rapid Radiation of Advanced Moths and Butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An Exploratory Study. *PLoS ONE*, **8**, e82615.

Bazinet, A.L., Mitter, K.T., Davis, D.R., van Nieukerken, E.J., Cummings, M.P., & Mitter, C.W. (2016) Phylotranscriptomics resolves ancient divergences in the Lepidoptera. *Systematic Entomology*, **accepted**.

Beirne, B.P. (1945) The male genitalia of the British Stigmellidae (Nepticulidae) (Lep.). *Proceedings of the Royal Irish Academy. Section B*, **50**, 191–218.

Bouchenak-Khelladi, Y., Onstein, R.E., Xing, Y., Schwery, O., & Linder, H.P. (2015) On the complexity of triggering evolutionary radiations. *New Phytologist*, **207**, 313–26.

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 Computational Biology*, **10**, e1003537.

Braun, A.F. (1915) New genera and species of Tineina. *Canadian Entomologist*, **47**, 188–197.

Braun, A.F. (1917) Nepticulidae of North America. *Transactions of the American Entomological Society*, **43**, 155–209, pls. 5–8.

Braun, A.F. (1925) Microlepidoptera of Northern Utah. *Transactions of the American Entomological Society*, **51**, 183–226.

Busck, A. (1907) New american Tineina. *Proceedings of the Entomological Society of Washington*, **8**, 86–99.

Chin, S.-W., Shaw, J., Haberle, R., Wen, J., & Potter, D. (2014) Diversification of almonds, peaches, plums and cherries–Molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Molecular Phylogenetics and Evolution*, **76**, 34–48.

Condamine, F.L., Clapham, M.E., & Kergoat, G.J. (2016) Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Nature Scientific Reports*, **6**, 19208.

Davis, D.R. (1989) Generic revision of the Opostegidae, with a synoptic catalogue of the world's species (Lepidoptera: Nepticuloidea). *Smithsonian Contributions to Zoology*, **478**, 1–97.

Davis, D.R. & Stonis, J.R. (2007) A revision of the New World plant-mining moths of the family Opostegidae (Lepidoptera: Nepticuloidea). *Smithsonian Contributions to Zoology*, **625**, i–v, 1–212.

Djernæs, M. & Damgaard, J. (2006) Exon-intron structure, paralogy and sequenced regions of elongation factor-1 alpha in Hexapoda. *Arthropod Systematics & Phylogeny*, **64**, 45–52.

Dong, X., Ding, Z., Yang, S., Luo, P., Wang, X., & Ji, J. (2013) Synchronous drying and cooling in central Asia during late Oligocene. *Chinese Science Bulletin*, **58**, 3119–3124.

Donner, H. & Wilkinson, C. (1989) Fauna of New Zealand. Nepticulidae (Insecta: Lepidoptera). *Fauna of New Zealand*, **16**, 1–88.

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

Doorenweerd, C., van Nieukerken, E.J., & Menken, S.B.J. (2015b) A global phylogeny of leafmining *Ectoedemia* moths (Lepidoptera: Nepticulidae): exploring host plant family shifts and allopatry as drivers of speciation. *PLoS ONE*, **10**, e0119586.

Emmet, A.M. (1984) The feeding habits of overwintering larvae of *Etainia sphendamni* (Hering, 1937) (Lep.: Nepticulidae). *Entomologist's Record and Journal of Variation*, **96**, 22–26.

Fischer, T.C. (2013) Pygmy moths (Lepidoptera, Nepticulidae) from Baltic Amber (Eocene). *Zitteliana A*, **53**, 85–92.

Forest, F. & Chase, M.W. (2009) Eudicots. In The timetree of life (ed. by S.B. Hedges & S. Kumar), pp. xxi, 551. Oxford University Press, Oxford.

Graham, A. (1999) Late Cretaceous and Cenozoic history of North American vegetation, north of Mexico. pp xviii, 350. Oxford University Press, Oxford.

Grimaldi, D.A. & Engel, M.S. (2005) Evolution of the insects, pp 755. Cambridge University Press, New York.

Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Guzmán, B. & Vargas, P. (2009) Historical biogeography and character evolution of Cistaceae (Malvales) based on analysis of plastid rbcL and trnL-trnF sequences. *Organisms Diversity & Evolution*, **9**, 83–99.

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**, 260.

Hennig, W. (1966) Phylogenetic systematics, English translation., pp 263. University of Illinois Press, Urbana.

Hoare, R.J.B. (1998) Systematics of Australian Nepticulid moths (Lepidoptera: Nepticulidae), Thesis Australian National University, Canberra.

Hoare, R.J.B. (2000a) Gondwanan Nepticulidae (Lepidoptera)? Systematics and biology of the *Ectoedemia (Fomoria) vannifera* group. *Tijdschrift voor Entomologie*, **142**, 299–316.

Hoare, R.J.B. (2000b) A new genus of primitive Nepticulidae (Lepidoptera) from eastern Australia, with a revised diagnosis of nepticulid subfamilies. *Zoological Journal of the Linnean Society*, **128**, 289–317.

Hoare, R.J.B., Johansson, R., van Nieukerken, E.J., & Nielsen, E.S. (1997) Australian Nepticulidae (Lepidoptera): redescription of the named species. *Entomologica Scandinavica*, **28**, 1–26.

Hoare, R.J.B. & van Nieukerken, E.J. (2013) Phylogeny and host-plant relationships of the Australian Myrtaceae leafmining moth genus *Pectinivalva* (Lepidoptera, Nepticulidae), with new subgenera and species. *ZooKeys*, **278**, 1–64.

Hohmann, N., Wolf, E.M., Lysak, M.A., & Koch, M.A. (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *The Plant Cell*, **27**, 2770–2784.

Ivinskis, P., van Nieukerken, E.J., & Rimsaite, J. (2012) *Trifurcula (Glaucolepis) lituanica* sp. n., an unexpected new stem-miner on *Salvia pratensis* occurring in eastern Europe (Lepidoptera: Nepticulidae). *Zootaxa*, **3570**, 41–55.

Johnson, P.J. (1982) A possible discovery of the first generation larva of *Etainia decentella* (Herrich-Schaffer, 1855). (Lepidoptera: Nepticulidae). *Entomologist's Record and Journal of Variation*, **94**, 245.

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

Kristensen, N.P. (2003). Skeleton and muscles: adults. In Lepidoptera, Moths and Butterflies, 2. Morphology, physiology and development (ed. by N.P. Kristensen), Handbuch der Zoologie/ Handbook of Zoology Vol. 4, pp. 39–131. De Gruyter, Berlin, New York.

Kulman, H.M. (1967) Biology of the hard maple bud miner, *Obrussa ochrefasciella*, and notes on its damage (Lepidoptera:Nepticulidae). *Annals of the Entomological Society of America*, **60**, 387–391.

Labandeira, C.C., Dilcher, D.L., Davis, D.R., & Wagner, D.L. (1994) 97-Million years of Angiosperm-Insect association – Paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 12278–12282.

Labandeira, C.C., Wilf, P., Johnson, K.R., & Marsh, F. (2007) Guide to Insect (and other) Damage Types on Compressed Plant Fossils, Version 3.0. Smithsonian Institution, Washington, DC. Available from http://paleobiology.si.edu/pdfs/InsectDamageGuide3.01.pdf

Lanfear, R., Calcott, B., Ho, S.Y., & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Laštůvka, A. & Laštůvka, Z. (2000) Zwei neue *Globularia* minierende *Trifurcula*-Arten (Lepidoptera, Nepticulidae). *Acta Musei Moraviae, Scientiae biologicae*, **85**, 289–296.

Laštůvka, A. & Laštůvka, Z. (2005) Four new *Trifurcula* species and additional faunal data on Nepticulidae from Italy (Lepidoptera: Nepticulidae). *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, **53**, 7–14.

Laštůvka, Z. & Laštůvka, A. (1994) Drei neue Arten der *Trifurcula pallidella*-Gruppe aus Mitteleuropa (Lepidoptera: Nepticulidae). *Entomologia Generalis*, **18**, 201–212.

Laštůvka, Z. & Laštůvka, A. (2007) Seven new species of the subgenus *Glaucolepis* Braun from southern Europe (Lepidoptera: Nepticulidae, *Trifurcula*). *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, **55**, 101–110.

Laštůvka, Z., Laštůvka, A., & van Nieukerken, E.J. (2013) The *Bupleurum* (Apiaceae) feeding species of *Trifurcula* (*Glaucolepis*): new species, biology and distribution (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie*, **156**, 191–210.

Lees, D.C., Kawahara, A.Y., Rougerie, R., Ohshima, I., Kawakita, A., Bouteleux, O., Prins, J., & Lopez-Vaamonde, C. (2014) DNA barcoding reveals a largely unknown fauna of Gracillariidae leaf-mining moths in the Neotropics. *Molecular Ecology Resources*, **14**, 286–296.

Lidgard, S. & Crane, P.R. (1988) Quantitative analyses of the early angiosperm radiation. *Nature*, **331**, 344–346.

Magallón, S., Hilu, K.W., & Quandt, D. (2013) Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, **100**, 556–73.

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., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, 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., Schütte, 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*, **346**, 763–767.

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

van Nieukerken, E.J. (1983) The Cistaceae-feeding Nepticulidae (Lepidoptera) of the western Palaearctic region. *Systematic Entomology*, **8**, 453–478.

van Nieukerken, E.J. (1986a) A second British specimen of *Bohemannia auriciliella* (Joannis) (=*bradfordi* Emmet) (Lepidoptera: Nepticulidae) and its possible host. *Entomologist's Gazette*, **37**, 207–208.

van Nieukerken, E.J. (1986b) Systematics and phylogeny of Holarctic genera of Nepticulidae (Lepidoptera, Heteroneura: Monotrysia). *Zoologische Verhandelingen*, **236**, 1–93.

van Nieukerken, E.J. (1990) The *Trifurcula subnitidella* group (Lepidoptera: Nepticulidae): taxonomy, distribution and biology. *Tijdschrift voor Entomologie*, **133**, 205–238.

van Nieukerken, E.J. (2007) Review of the subgenus *Trifurcula* (*Levarchama*), with two new species (Lepidoptera: Nepticulidae). *Acta Zoologica Academiae Scientiarum Hungaricae*, **53**, 101–140.

van Nieukerken, E.J. (2008) Two new species in the *Ectoedemia* (*Fomoria*) *weaveri*-group from Asia (Lepidoptera: Nepticulidae). *Zoologische Mededelingen (Leiden)*, **82**, 113–130.

van Nieukerken, E.J. & Dop, H. (1987) Antennal sensory structures in Nepticulidae (Lepidoptera) and their phylogenetic implications. *Zeitschrift für Zoologische Systematik und Evolutionsforschung*, **25**, 104–126.

van Nieukerken, E.J., Nielsen, E.S., Johansson, R., & Gustafsson, B. (1990). Introduction to the Nepticulidae. In The Nepticulidae and Opostegidae (Lepidoptera) of NW Europe (ed. by R. Johansson, E.S. Nielsen, E.J. van Nieukerken & B. Gustafsson), Fauna Entomologica Scandinavica Vol. 23, pp. 11–109.

van Nieukerken, E.J. & Johansson, R. (1990). Tribus Trifurculini. In The Nepticulidae and Opostegidae (Lepidoptera) of NW Europe (ed. by R. Johansson, E.S. Nielsen, E.J. van Nieukerken & B. Gustafsson), Fauna Entomologica Scandinavica Vol. 23, pp. 239–321.

van Nieukerken, E.J. & Puplesis, R. (1991) Taxonomy and distribution of the *Trifurcula (Glaucolepis) raikhonae* group (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie*, **134**, 201–210.

van Nieukerken, E.J. & Laštůvka, A. (2002) *Ectoedemia (Etainia) obtusa* Puplesis & Diškus, new for Europe: taxonomy, distribution and biology (Lepidoptera: Nepticulidae). *Nota Lepidopterologica*, **25**, 87–95.

van Nieukerken, E.J. & Johansson, R. (2003) The *Quercus* feeding *Stigmella* species of the West Palaearctic: new species, key and distribution (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie*, **146**, 307–370.

van Nieukerken, E.J., Hoare, R.J.B., & Groenenberg, D.S.J. (2004a) Evolution of Nepticulidae & host plant use. In International Congress of Entomology XXII, CD rom, Brisbane.

van Nieukerken, E.J., Laštůvka, A., & Laštůvka, Z. (2004b) Annotated catalogue of the Nepticulidae and Opostegidae (Lepidoptera: Nepticuloidea) of the Iberian Peninsula. *SHILAP Revista de Lepidopterologia*, **32**, 211–260.

van Nieukerken, E.J., Mazurkiewicz, A., & Pałka, K. (2004c) *Trifurcula pallidella* (Duponchel, 1843) (Nepticulidae): distribution, biology and immature stages, particularly in Poland. *Nota Lepidopterologica*, **27**, 159–178.

van Nieukerken, E.J., Biesenbaum, W., & Wittland, W. (2010) Die Gattung *Trifurcula* Zeller, 1848 in Deutschland mit zwei Erstnachweisen für die deutsche Fauna (Lep., Nepticulidae). The genus *Trifurcula* Zeller, 1848 in Germany with two first records for Germany (Lep., Nepticulidae). *Melanargia*, **22**, 1–26.

van Nieukerken, E.J., Doorenweerd, C., Hoare, R.J.B., & Davis, D.R. (2016a) Revised classification and catalogue of global Nepticulidae and Opostegidae (Lepidoptera: Nepticuloidea). *ZooKeys*, **accepted**..

van Nieukerken, E.J., Doorenweerd, C., Nishida, K. & Snyers, C. (2016b) New taxa, including two new genera show uniqueness of Neotropic Nepticulidae (Lepidoptera). *ZooKeys*, **accepted**.

Papadopoulou, A., Anastasiou, I., & Vogler, A.P. (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Bioliology and Evolution*, **27**, 1659–72.

Peña, C. & Malm, T. (2012) VoSeq: A Voucher and DNA Sequence Web Application. *PLoS ONE*, **7**, e39071.

Petersen, W. (1930) Die Blattminierer-Gattungen *Lithocolletis* und *Nepticula* (Lep.). Teil II: *Nepticula* Z. *Stettiner Entomologische Zeitung*, **91**, 1–82, 3 pls.

Powell, J.A., Mitter, C., & Farrell, B. (1998). Evolution of larval food preferences in Lepidoptera. In Lepidoptera, Moths and Butterflies, 1. Evolution, systematics and biogeography (ed. by N.P. Kristensen), Handbuch der Zoologie/ Handbook of Zoology Vol. 4, pp. 403–422. De Gruyter, Berlin, New York.

Puplesis, R. (1994) The Nepticulidae of eastern Europe and Asia. Western, central and eastern parts, pp 290. Backhuys Publishers, Leiden.

Puplesis, R. & Diškus, A. (1996) First record of the genus *Etainia* Beirne from Central Asia with descriptions of two new species and some provisional notes on the world fauna (Lepidoptera: Nepticulidae). *Phegea*, **24**, 41–48.

Puplesis, R. & Robinson, G.S. (2000) A review of the Central and South American Nepticulidae (Lepidoptera) with special reference to Belize. *Bulletin of the Natural History Museum London. Entomology*, **69**, 3–114.

Rambaut, A. (2014) FigTree v1.4.2, Available from http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C., & Mitter, K.T. (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS ONE*, **8**, e58568.

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

Schneider, H., Schuettpelz, E., Pryer, K.M., Cranfill, R., Magallon, S., & Lupia, R. (2004) Ferns diversified in the shadow of angiosperms. *Nature*, **428**, 553–557.

Scoble, M.J. (1980) *Trifurcula* Zeller: a critical analysis of the genus, cladistic relationships and descriptions of two new species from South Africa (Lepidoptera: Nepticulidae). *Journal of the Entomological Society of Southern Africa*, **43**, 139–150.

Scoble, M.J. (1983) A revised cladistic classification of the Nepticulidae (Lepidoptera) with descriptions of new taxa mainly from South Africa. *Transvaal Museum Monograph*, **2**, 1–105.

Silvestro, D., Cascales-Minana, B., Bacon, C.D., & Antonelli, A. (2015) Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytologist*, **207**, 425–36.

Skelton, P.W. (2003) The Cretaceous World, pp 360. Cambridge University Press.

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

Stainton, H.T. (1849) An attempt at a systematic catalogue of the British Tineidae & Pterophoridae, pp iv + 32. John van Voorst, London.

Stainton, H.T. (1854) Insecta Brittanica. Lepidoptera: Tineina, pp viii, 313. Lovell Reeve, London., London.

Stainton, H.T. (1859) Manual of British butterflies and moths, 2, pp xi, 480. Cooke & Son, London.

Tilmon, K.J. (2008) The evolutionary biology of herbivorous insects: specialization, speciation and radiation, pp 341. University of California Press.

Vakhrameev, V.A. & Hughes, N.F. (1991) Jurassic and Cretaceous floras and climates of the Earth, pp 318. Cambridge University Press.

Vargas, P., Valente, L.M., Blanco-Pastor, J.L., Liberal, I., Guzmán, B., Cano, E., Forrest, A., &

Fernández-Mazuecos, M. (2014) Testing the biogeographical congruence of palaeofloras using molecular phylogenetics: snapdragons and the Madrean–Tethyan flora. *Journal of Biogeography*, **41**, 932–943.
Wahlberg, N. & Wheat, C.W. (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**, 231–242.
Wahlberg, N., Wheat, C.W., & Pena, C. (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS ONE*, **8**, e80875.
Warnock, R.C.M., Yang, Z., & Donoghue, P.C.J. (2012) Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, **8**, 156–159.
Waters, J.M. & Craw, D. (2006) Goodbye Gondwana? New Zealand biogeography, geology, and the problem of circularity. *Systematic Biology*, **55**, 351–356.
Wheat, C.W. & Wahlberg, N. (2013) Critiquing blind dating: the dangers of over-confident date estimates in comparative genomics. *Trends in Ecology, Systematics and Evolution*, **28**, 636–642.
Wiens, J.J., Lapoint, R.T., & Whiteman, N.K. (2015) Herbivory increases diversification across insect clades. *Nature Communications*, **6**, 8370.
Wilf, P. & Escapa, I.H. (2014) Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. *New Phytologist*, **207**, 283–290.
Wilkinson, C. (1978) On the *Stigmella – Nepticula* controversy (Lepidoptera). *Tijdschrift voor Entomologie*, **121**, 13–22.
Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., & Mathews, S. (2012) Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 17519–17524.
Xing, Y., Onstein, R.E., Carter, R.J., Stadler, T., & Peter Linder, H. (2014) Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. *Evolution*, **68**, 2821–2832.
Yang, C.-K. (1989) *Sinopticula sinica* (Lepidoptera: Nepticulidae) a new genus and species from China. *Entomotaxonomia*, **11**, 79–82.
Zeller, P.C. (1848) Die Gattungen der mit Augendeckeln versehenen blattminirenden Schaben. *Linnaea Entomologica*, **3**, 248–344, pl. 2.
Zwickl, D.J. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.

Supplemental Files

Supplemental File 1 (S1): List of samples and Genbank accession numbers. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/XxzY3k5RhERYSV5

Supplemental File 2 (S2): Full BEAST tree with Bayesian branch support values and estimated times of divergence for all clades. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/4FNcYQGC0ItHhkV

6

Increased diversification rates across lineages of leaf-mining moths

Increased diversification rates across lineages of leaf-mining moths after entering the new adaptive zone of the temperate region

C. Doorenweerd[1, 2], E.J. van Nieukerken[1], C. Lopez-Vaamonde[3], S.B.J. Menken[2]

Keywords: biodiversity, herbivory, speciation, Lepidoptera, Nepticulidae, Gracillariidae, Lithocolletinae

[1] Biodiversity Discovery group, Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA, Leiden, The Netherlands
[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE, Amsterdam, The Netherlands
[3] Unité de Recherche en Zoologie Forestière, Institut National de la Recherche Agronomique, 45075 Orléans Cedex 2, France
Corresponding author: C. Doorenweerd, camiel.doorenweerd@naturalis.nl

Abstract

Despite over 150 years of evolutionary research since Darwin, the factors that drive speciation still remain largely elusive. Herbivorous insects are often extremely species-rich groups, which allow testing of a central prediction from two classic evolutionary theories, the escape-and-radiate theory and the adaptive radiation theory. Both suggest that colonization of a novel resource zone (e.g. a novel host plant family or biogeographical region) results in increased rates of speciation. We studied the phylogenetic diversification patterns and possible impact of biogeographical and host plant family shifts of two groups of leaf-mining moths: pygmy leaf-mining moths (Nepticulidae) and leaf-blotch mining moths (Gracillariidae: Lithocolletinae). We gathered a large molecular dataset by sequencing up to eight markers for 335 species of Lithocolletinae and 645 species of Nepticulidae, which encompasses approximately 50% of the known diversity in both groups. We found that there is no linear correlation between clade age and diversity; some of the most species-rich clades are relatively young. There is a repeated pattern where clades experience a rate-shift with increased diversification rates, which after this initial burst slows down again, for which we discuss several possible explanations, such as density-dependent-selection and the co-evolution of pathogens with their hosts. Our results show that the diversification of Lithocolletinae and Nepticulidae has in parts likely been contemporaneous with increased rates of diversification of the hosts. We found no evidence that a novel host plant family in itself would constitute a new adaptive zone, which would fit the escape-and-radiate theory but we conclude that the diversity of the hosts is a prerequisite for diversification. The diversification pattern we find in Nepticulidae and Lithocolletinae fits the adaptive radiation theory, and their adaptive zone can best be described as the temperate region. Our results also highlight that "entering" a new zone requires evolutionary adaptation in order to be able to diversify: the clades that experienced increased rates of diversification appear to be uniquely characterised by a pre-adult winter diapause, which may have been a "key innovation" to successfully adapt to temperate climate.

Significance statement

If and how speciation occurs in nature besides from geographic isolation barriers is only partially understood. By comparing the timing and extent of diversification in large-scale species-level phylogenies, macroevolutionary patterns emerge that indicate the importance of different, theoretically conceived, candidate speciation factors. Insect herbivores in particular are an interesting group to apply this method to because the importance of the host plants for diversification is a long-standing debate. We show that for two groups of highly specialised herbivorous insects, adaptation to a temperate climate by surviving the winter as larva or pupa in a suspended state has been an important factor.

Introduction

Darwin's work (Darwin, 1859) marked a revolution in our thinking on evolution. Despite over 150 years of evolutionary research since, the factors that drive the key process, speciation - the splitting of one species into two - still remain largely elusive (Futuyma, 2013). There is a growing number of theories, each suggesting different candidate drivers (Howard & Berlocher, 1998; Tilmon, 2008) and some of which have become well established, such as geographic isolation, limiting genetic exchange and leading to allopatric speciation (Mayr, 1947; 1963). For others, their importance is less clear, but it appears evident that vicariance alone cannot explain many of the diversity patterns we observe today (Condamine *et al*., 2016; Wiens *et al*., 2015). Other drivers that have been proposed include changes in climate (Nyman *et al*., 2012), possibly related to tectonic shifts (Chaboureau *et al*., 2014), genetic rearrangements (Edger *et al*., 2015) or ecological opportunity, in particular the availability of hosts (McKenna *et al*., 2009; Becerra & Venable, 1999). The latter applies only to trophic levels above primary producers, but includes an important part of all biodiversity, many of which are insects. Insects compose approximately 62% of all animal biodiversity (Mitter *et al*., 1988), and the largest radiations commonly include many herbivores (Wiens *et al*., 2015; Althoff *et al*., 2014). Furthermore, herbivory has been found to have a positive effect on the diversification of clades of insects (Wiens *et al*., 2015), suggesting that host plants have contributed to speciation in these lineages (Winkler & Mitter, 2008).

There are striking recurrent patterns of diversification and host use in groups of herbivorous insects that are specialised with respect to feeding substrate, i.e. having narrow diet breadths. Often there is little to no overlap in host use between closely related species, but species groups tend to feed on a single host plant family (Doorenweerd *et al*., 2015a; Nyman *et al*., 2006). Studies using time-calibrated phylogenies of herbivores and hosts have repeatedly shown that their diversification has mostly not been contemporaneous, with significant lags in the diversification of the insects (Lopez-Vaamonde *et al*., 2006; Percy *et al*., 2004). Host shifts are thus often involved in speciation events, but speciation of the hosts does not appear to initiate the speciation process in the insects. Climate change has been identified as

a driver for the diversification of leaf-mining flies (Winkler *et al*., 2009), and genetic rearrangements as a driver in butterflies (Edger *et al*., 2015), but many more candidate drivers and mechanisms have been proposed, but none have been analysed comparatively across multiple groups (Althoff *et al*., 2014).

A central prediction in two classic evolutionary theories, the escape-and-radiate theory (Ehrlich & Raven, 1964) and the adaptive radiation theory (Simpson, 1953) is that colonization of a novel resource zone would result in increased rates of speciation. Novel resources are abstractly defined, and may have physical, biotic or behavioural boundaries. Specialized, monophagous and oligophagous, insects always feed on a single or a few closely related host plants. For such groups, a different host plant family may likely constitute a novel resource zone, under the assumption that evolutionary similarity highly correlates with phytochemical and physiognomical similarity (Nyman, 2010). Leafminers are a guild of insects that feed enclosed within plant tissues. They are always highly specialised and have proliferated into significant radiations across all holometabolous insect orders (Forister *et al*., 2015; Wiens *et al*., 2015). Within the moths and butterflies (Lepidoptera), the leaf-mining habit is found in multiple families across the Lepidoptera tree of life (Regier *et al*., 2015). The high degree of specialisation of leaf-mining insects on angiosperm plant families and multiple colonisations of novel host families, make them a suitable group for studying potentially host-driven diversification dynamics.

Here we study the impact of biogeographical and host plant family shifts on the diversification of two lineages of moths, pygmy leaf-mining moths (Nepticulidae) and leaf-blotch mining moths (Gracillariidae: Lithocolletinae). Both groups are leaf-miners, but with distinct leaf-mine types. Pygmy leaf-mining moths feed only on plant tissue, whereas leaf-blotch mining moths start with sap-feeding larval stages and in later instars continue feeding on plant tissue. The tissue-feeding instars of leaf-blotch mining moths often use silk to create a fold in the leaf for a more spacious leaf-mine, pygmy leaf-mining moths do not develop spinning glands before their final larval instar and use it only to construct a cocoon (Gustafsson & Nieukerken, 1990; Emmet *et al*., 1985). Both groups are globally distributed and display the typical non-random patterns of host associations described above. There is a striking overlap in favoured host plant families between the two groups of leaf-miners: dominant plant families include oaks and beeches (Fagaceae), legumes (Fabaceae), the rose family (Rosaceae), the birch family (Betulaceae) and the poplars and willows (Salicaceae) (Doorenweerd *et al*., 2016; De Prins & Kawahara, 2012). Equally intriguing, however, are the differences in the use of particular host plants and species richness across their geographic distribution. By comparing the diversification patterns of both leaf-mining groups in relation to their distribution and biology, we aimed to identify common drivers of speciation. More specifically, we tested for shifts in diversification rate and assessed the effects of two candidate drivers: 1) host plant family shifts and, 2) shifts between biogeographic regions, and discuss a theoretic framework with intrinsic drivers, such as functional adaptations, host shifts and population dynamics, and extrinsic

drivers, such as climate change, shifting host distributions, pathogens or geological events, that may have enabled or triggered diversification rate shifts.

Material and Methods

Data collection

We gathered sequence data from as many species of Nepticulidae and Lithocolletinae as possible, and verified all identifications using morphological characters and/or life history and distribution data. For detailed information on all specimens see BOLD. When available, material has been vouchered with pinned adult specimens, genitalia slides or preparations of larval pelts, and dried leaf mines from which larvae were taken. Voucher material is stored, along with the corresponding DNA extracts, either in the RMNH (Naturalis, Leiden, The Netherlands), INRA (Orléans, France), the Canadian Centre for Biodiversity (Guelph, Canada) or several other collections, to allow for future study. Previously published sequences have been re-used (Doorenweerd *et al*., 2016; De Prins *et al*., 2013; De Prins & Kawahara, 2012; Davis & De Prins, 2011; Lopez-Vaamonde *et al*., 2006; Lopez-Vaamonde *et al*., 2003), but only after verifying the identifications. DNA extractions were mostly performed at Naturalis (Leiden, The Netherlands), some were performed at the Canadian Centre for Biodiversity or the Smithsonian Institution (Washington, USA), in all cases using silica membrane based extraction methods. Up to eight genes were sequenced for a selection of the material, aiming to include multiple genes for representatives of all species groups in order to produce a robust phylogeny. For Nepticulidae, we sequenced fragments of mitochondrially encoded cytochrome c oxidase I (COI), mitochrondrially encoded cytochrome c oxidase II (COII), 28S ribosomal DNA, elongation factor 1 alpha (EF1-alpha), carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (CAD), isocitrate dehydrogenase (IDH), malate dehydrogenase (MDH) and histone 3 (H3) (Doorenweerd *et al*., 2016). For Lithocolletinae, we used the same selection, except that we sequenced a fragment of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) instead of H3, and used different primers for EF1-alpha, possibly targeting a different copy of the gene (see Doorenweerd *et al*., 2016). The primer sequences and amplification conditions for EF1-alpha, using the 'Starsky & Monica' primer-set, and for GAPDH are described in Wahlberg and Wheat (2008). Because more variable and more conserved markers are included, the combination of these markers provides resolution for the basal branches in the phylogenies as well as towards the tips.

Phylogenetic analysis

The workflow is schematized in Fig. 1. All software was used in an OpenStack cloud computing environment at Naturalis using the PhylOStack v1.4 instructions (Doorenweerd, 2016). Input data were combined from different resources. The RMNH VoSeq (Peña & Malm, 2012) leaf-miner sequence database and BOLD (Ratnasingham & Hebert, 2007) were mined for DNA sequences. The reference dataset for Nepticulidae with minimally sequence data for five out of eight genes

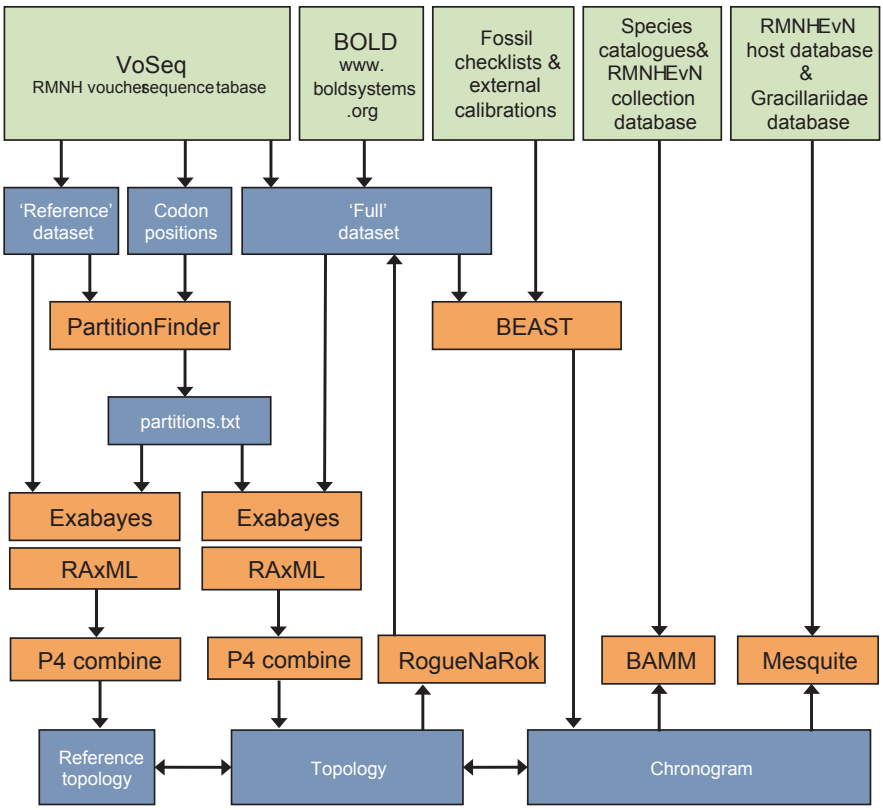


Figure 1: Schematic representation of the phylogenetic workflow, displaying where the source sequence data was obtained from and how this data was processed through different software packages. For details see the material and methods section.

includes 374 ingroup species. The reference dataset for Lithocolletinae includes 131 ingroup species. NCBI Genbank accession numbers can be found in Table S1. The most appropriate partitioning scheme was determined using PartitionFinder v2 pre-release 13 (Lanfear *et al*., 2012), initially allowing all codon positions to be separate subsets, but requiring at least 300 nucleotides per subset to ensure sufficient information for likelihood calculations, and rcluster searching using the corrected Akaike information criterion to combine subsets. Bayesian and Maximum Likelihood (ML) analyses were done using, respectively, ExaBayes v1.4.1 (Aberer *et al*., 2014) and RAxML v8 (Stamatakis, 2006). For the Lithocolletinae dataset, this resulted in seven subsets (pos = codon position): [COII_pos1, 28S, COI_pos1] [CAD_pos3, EF1-alpha_pos3, MDH_pos3] [COI_pos2, CAD_pos1] [GAPDH_pos1, GAPDH_pos2, CAD2_pos2, COII_pos2] [COII_pos3, COI_pos3, IDH_pos3, GAPDH_pos3] [IDH_pos1, IDH_pos2, EF1-alpha_pos1, MDH_pos1] [EF1-alpha_pos2, MDH_pos2]. For Nepticulidae, this also resulted in seven subsets, of slightly different composition, but also favouring the combination of the same

codon positions: [EF1-alpha_pos1, CAD_pos1, 28S] [CAD_pos3, MDH_pos3, COII_pos3, COI_pos3, H3_pos3] [IDH_pos1, CAD_pos2] [EF1-alpha_pos3, IDH_pos3, COI_pos1] [H3_pos2, COI_pos2, EF1-alpha_pos2, MDH_pos2] [COII_pos1, MDH_pos1] [COII_pos2, H3_pos1, IDH_pos2]. For all subsets, a GTR+G substitution model was applied for subsequent analyses, except with the Bayesian analysis in BEAST (Bouckaert *et al*., 2014), where a HKY+G model was applied to prevent the over-parametrisation of priors. Bayesian analyses with ExaBayes were performed using two independent replicates with four heated chains and a 2% convergence cut-off value. RAxML was instructed to search for the best tree, followed by a multiparametric bootstrapping analysis, and the support values were plotted on the best tree. The support values obtained with the Bayesian and ML approach were then both superimposed on the ML best tree using the python combine script from the P4 v1.2 package (Foster, 2004).

On the one hand, it is known that phylogenetic (topologic) uncertainty can have strong effects on diversification analyses (Pena & Espeland, 2015), but on the other hand, non-random missing taxa can produce skewed patterns as well (Moen & Morlon, 2014; Cusimano & Renner, 2010). The phylogenetic analyses were therefore iterated with different sets of taxa, where we attempted to find a balance between a stable topology, the support for clades and the number of species included. Initially, a dataset was constructed with a single representative per species with sequence data for at least five out of eight genes, which functioned as a reference topology with good support (Bayesian posterior probability >0.95 and bootstrap support >60) for most clades in both the Lithocolletinae and Nepticulidae topology. A second dataset, termed 'full dataset' in Fig. 1, was constructed for each group with species that had sequence data of at least one marker, often just a DNA barcode. This dataset followed the same phylogenetic approach outlined above, with the addition that after the Bayesian and ML analyses the bootstrapped trees were tested for 'rogue' taxa with RogueNaRok v1 (Aberer *et al*., 2013), which are taxa that cannot be placed in the tree with any confidence and distort the topology compared to the reference topology. The trees were visually compared using the extended dendrogram package 'dendextend' in R (Galili, 2015). Problematic taxa were removed from the dataset and this process was repeated until a topology with the maximum number of species included was found that was congruent with the reference topology. The congruence was tested statistically with a Baker's Gamma Index permutation test, implemented in the dendextend R package (Fig. S2) (Galili, 2015). The resulting 'full' datasets consisted out of 645 species of Nepticulidae, 53% of the known diversity (Fig. S3), and 335 species of Lithocolletinae, 52% of the known diversity, where the known diversity is defined as the described species plus undescribed species known from collections (Fig. S3).

Molecular dating

The software package BEAST v2.3.2 (Bouckaert *et al*., 2014) was used to perform divergence time estimations on both datasets. For Nepticulidae we used the same two calibration points as in Doorenweerd *et al*. (Doorenweerd *et al*., 2016). Fossils of Lithocolletinae have not yet been extensively studied [but see summarized lists

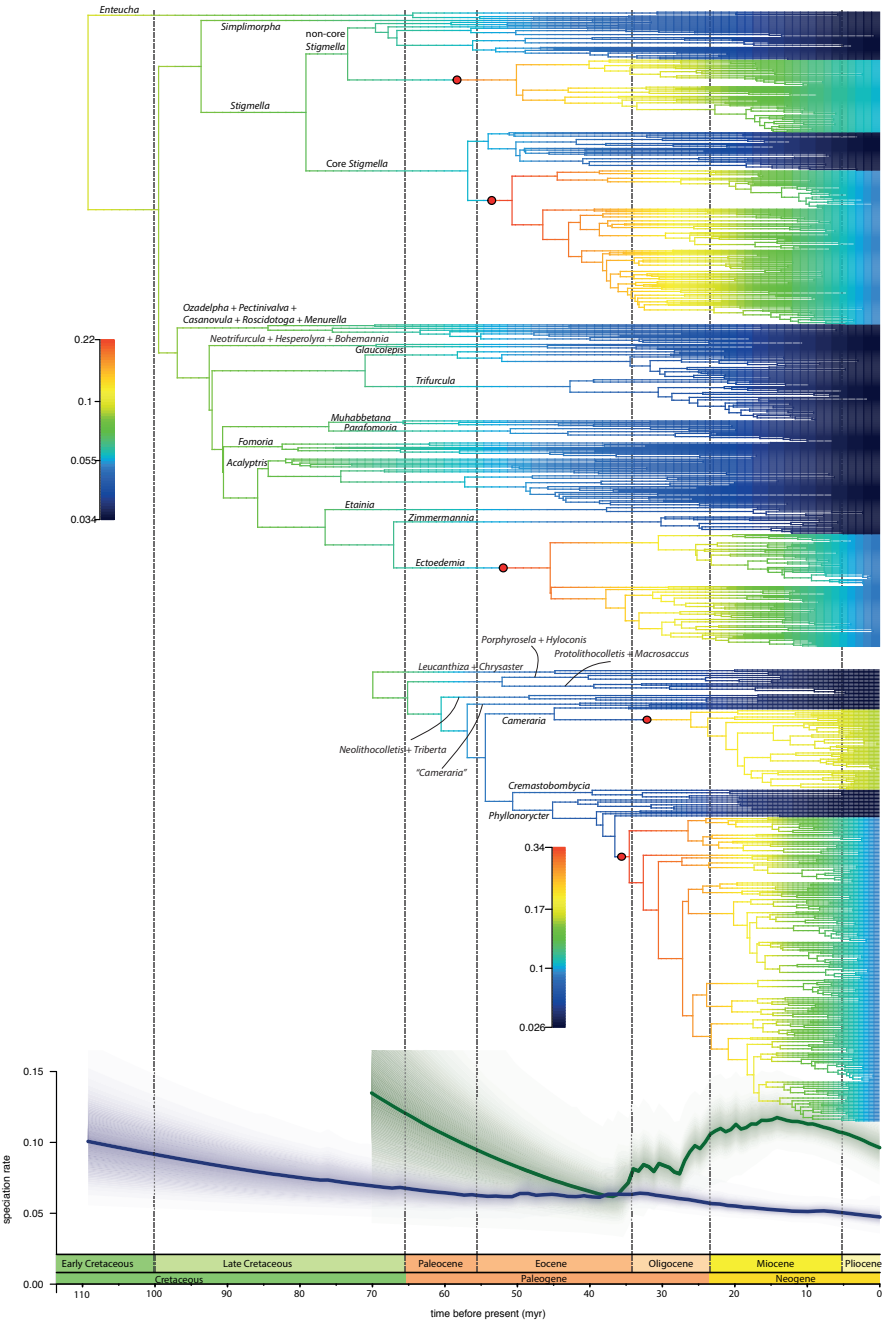


Figure 2: Molecular chronograms for Nepticulidae and Lithocolletinae with together nearly 1,000 species. Branches are coloured following a temperature gradient of net diversification rates with legends next to the respective tree and speciation rate graphs for both groups depicted below. Everything is scaled to a single age axis to allow for temporal comparisons. The red dots on the branches indicate the most supported position for a shift in diversification rate.

in Sohn *et al*. (2012) and Lopez-Vaamonde *et al*. (2006)] and the records are either too recent to be of use for calibrating nodes or we have doubts on the identifications and have therefore not applied them to our data. Instead, we used an external calibration to date the split between *Gracillaria* and *Phyllocnistis* + *Phyllonorycter* with 74–104 myr 5% and 95% boundaries of a normal distribution, matching the estimates from Wahlberg *et al*. (2013). The Bayesian Monte-Carlo Markov-Chain was ran for 200 million generations on both datasets and results were checked for sufficient sampling of all priors using Tracer v1.6 (Rambaut *et al*., 2014); topologies were checked for congruence with the topologies obtained with Exabayes and RAxML analyses on the full and reference datasets. The full beast trees including 95% highest posterior densities for the age estimates of each node are in Fig. S4. In the results and discussion sections we indicate the minimum and maximum 95% highest posterior density values between square brackets. Following standard practice in Geology, we use the abbreviation Ma for million years ago and myr for millions of years.

Ancestral state reconstruction

To ascertain the importance of host plant family shifts and biogeographic distribution for diversification rates, we mapped these data onto the trees. Host plant family, distribution and winter survival mode characters were obtained from EJvN's leafminer host database (partly unpublished, partly published in Menken *et al*. [2010]), the De Prins & De Prins (2016) Global Gracillariidae database and numerous other publications with species descriptions that included data on life history. The characters were plotted on the chronograms that resulted from the BEAST analysis using the ancestral state reconstruction in Mesquite v3.04 (Maddison & Maddison, 2015). The maximum parsimony reconstruction method with unordered states was applied, which allows multiple character states per taxon. Resulting trees are in Fig. S5. A host shift here implies a process of initially incorporating an additional host in the diet breadth, followed by specialisation and diverging lineages where one of the lineages continues on the new host, rather than a single lineage leaving one host for another. The discussed timing of host shifts or colonisation refers to the minimum age for when a clade became associated with that host.

Diversification analysis

The BEAST-produced chronograms were used as input for speciation-extinction analyses in BAMM v2.5 (Rabosky, 2014), after removing the outgroups, which were not randomly sampled. BAMM was run for 5 million generations, mostly following the BAMM project website instructions for a speciation-extinction analyses (Rabosky, 2016). We based the sampling estimates on the 'known diversity': a combination of described species and undescribed species known from collections to the authors. The diversity of Lithocolletinae is mostly estimated from the Global Gracillariidae database (De Prins & De Prins, 2016); for the described Nepticulidae diversity records see Nieukerken *et al*. (2016). Our estimates of the known diversity are in Fig. S3. We used two different approaches regarding the sampling estimates: 1) an average global sampling estimate [53% for

Nepticulidae, 52% for Lithocolletinae] and 2) genus-specific sampling estimates, based on the known diversity (Fig. S3). The results were highly similar (results not shown), indicating that although the sampling coverage differs per genus, sampling has effectively been random. We tested for shifts in diversification rates across lineages and the scenario with the highest posterior probability is presented in Fig. 2. Net diversification rates are indicated with a temperature colour gradient on the tree using the 'Jenks natural breaks' binning mode. Alternatively supported scenarios are in Fig. S6 and treated in the results section.

Results

The average speciation rates in Lithocolletinae have always been higher than those in Nepticulidae, except for a relatively short period around 36 Ma (Fig. 2). We detect five shifts with increased diversification rates: three in Nepticulidae and two in Lithocolletinae. The Bayesian speciation-extinction analysis in BAMM allows testing of distinct diversification scenarios. The rate shift scenario with the highest posterior probability for Lithocolletinae was supported by 76% of the data, and identified a shift during the Eocene and one during the Oligocene. The second best option was only supported by 7% of the data, and identified shifts in almost identical clades (Fig. S6). Within Nepticulidae, there was much more ambiguity in the data and no single scenario was found to stand out as much better than the others (Fig. S6). The highest posterior probability scenario supports three shifts, one during the Paleocene and two during the Eocene (Fig. 2). The next best scenarios support four shifts, but always including similar clades to the ones in the best shift configuration scenario. In those scenarios, the fourth shift was estimated around 15 Ma, mid-Miocene, in the *Stigmella salicis* group, which is specialised on Salicaceae. Because this group has been intensively studied, particularly in Europe, the sampled fraction is likely higher than the average estimated 54% for all *Stigmella*. Indeed, after repeating the analysis assuming a 90% sampled fraction of the *S. salicis* group (results not shown), there was no longer support for a shift in diversification rate in that clade. In Lithocolletinae, the shifts in diversification have a pronounced effect on the net diversification curve, whereas for Nepticulidae the effects are smaller and at most weaken the decline.

The Nepticulidae crown group is estimated to have originated during the Early Cretaceous [see also (Doorenweerd *et al*., 2015b)]. Two rate shifts are found in the species-rich genus *Stigmella*, and one in the stem lineage of *Ectoedemia*. *Stigmella* splits in two larger clades (Doorenweerd *et al*., 2016), so-called core and non-core *Stigmella*, and they display a similar pattern where one clade branches off and has lower diversification rates and a second clade with higher diversification rates. The clades with higher diversification rates are mostly specialised on deciduous hosts in temperate regions (Fig S5). The earliest diversification rate shift has been estimated to be in the Paleocene, around 58 Ma, in part of the non-core *Stigmella* clade. The second rate shift involves part of the core *Stigmella* clade, and is estimated to have occurred in the early Eocene, around 53 Ma. The rate shift

involving *Ectoedemia* is estimated to have occurred about one-and-a-half million years later. In some scenarios, the genera *Etainia* and *Zimmermannia* are also included in this higher diversification regime, which would estimate the rate shift as far back as the Late Cretaceous. The three genera have different biologies - *Etainia* are mining buds, shoots and sometimes fruits, *Zimmermannia* are barkminers in trees and most *Ectoedemia* are leafminers (a few are gallers) - but they are all predominantly Holarctic, whereas the sister clade *Acalyptris* has a more tropical distribution (Fig. S5).

Lithocolletinae are a younger group than Nepticulidae with crown group age estimates just before the K-Pg boundary, around 70 Ma, but with a large confidence interval [52–88 95% HPD] (Fig. S4). The two detected diversification rate shifts in Lithocolletinae have occurred after the shifts in Nepticulidae. The first, in *Phyllonorycter*, is estimated around 36 Ma (late Eocene), the second, in *Cameraria*, several million years later, 32 Ma (early Oligocene). Not surprisingly, the shifts involve the two largest genera, but they do not encompass each genus entirely. In *Cameraria,* the generic boundaries are somewhat unclear and likely the tropical taxa require revision (De Prins & Kawahara, 2012; Kumata, 1993); *Cameraria* is also paraphyletic in our analyses. In a clade that can be distinguished as a monophyletic *Cameraria*, there is first a clade with more tropically distributed species that splits off before the rate shift occurs and a Holarctic, and particularly Nearctic radiation ensues. In *Phyllonorycter* there is an initial clade with Malvaceae feeders found in the Afrotropical and Australian region, before the rate shift occurs and a predominantly Holarctic clade begins to diversify. After the initial rate increases, rates are slowly decreasing, more so in *Phyllonorycter* than in *Cameraria*, visible from the 'cooler' colours on the branches of the tree (Fig. 2). This pattern of an initial increase in speciation rates followed by a gradual slowdown is also observed in Nepticulidae.

The reconstructed association with a particular host plant family typically does not predate the Miocene, with several notable exceptions (Fig. S5). The oldest host association that could be reconstructed is the ancestral relationship of Lithocolletinae with Fabaceae. Because most of the smaller genera, which are confidently placed in the topology, are feeding on this plant family this is a firmly established scenario. The estimated crown age for Fabaceae is strikingly similar to Lithocolletinae, around 70 Ma (Naumann *et al*., 2013; Tank *et al*., 2015). Fabaceae were colonized once again in a European clade of *Phyllonorycter* around 18 Ma [13–23], followed by a substantial diversification (Lastuvka *et al*., 2013). The Fabaceae-feeding *Phyllonorycter* use typical Mediterranean Fabaceae: brooms (Genisteae) and trifoliate herbs (Trifolieae), whereas the smaller lithocolletine genera mostly feed on tropical Fabaceae, in different subgroups. In Nepticulidae, there are multiple shifts to Fabaceae. The two shifts involving the largest numbers of species are in the stem lineage of *Trifurcula*, 43 Ma [35–51], and in *Stigmella*, 53 Ma [45–62]. Also here there is discrimination between Mediterranean Fabaceae (in *Trifurcula*) and earlier colonized tropical Fabaceae (in *Stigmella*), where the tropical radiations, as far as known, involve fewer species. The oldest reconstructed host association in

Nepticulidae is in the genus *Enteucha* with the plant family Polygonaceae, estimated to date back 64 Ma [49–79]. The crown age of Polygonaceae is estimated to be substantially younger, which casts some doubt on the age of this association (see also discussion in Doorenweerd *et al*., 2016). The *Phyllonorycter* species that are not included in the higher diversification regime all feed on Malvaceae. The Malvaceae were likely colonized 42 Ma [30–52], during the Eocene in the Aftrotropical or Australian region, followed by a modest southern hemisphere radiation. It is possible that the association with Malvaceae dates back 50 Ma [37–64], but this is likely the result of phylogenetic uncertainty in the placement of *Phyllonorycter grewiaephyllos*, which in this tree groups with *Cremastobombycia*. All *Cremastobombycia* feed on Asteraceae, an association that is estimated to date back 33 myr [20–43].

All of the dominant host plant families, viz. Fagaceae, Rosaceae, Betulaceae, Fabaceae and Salicaceae, have been colonized more recently, with the exception of all but the non-tropical Fabaceae. During the Oligocene, Lithocolletinae have colonized the Fagaceae at least twice. The first shift to Fagaceae is estimated at 31 Ma [22–38], in *Phyllonorycter*, which has led to multiple regional radiations: a mostly East-Palaearctic radiation, a mostly West-Palearctic radiation and a Nearctic radiation (Fig S5). The second shift to Fagaceae occurred 23 Ma [17–30] in *Cameraria*, in the Nearctic, which now holds a large diversity of *Cameraria* species associated with almost all extant North American oaks. In Nepticulidae, the shifts to Fagaceae with subsequent radiations have been more common and have less strong regional aspects, although in *Stigmella* the distinction between Nearctic and Palearctic radiations is partly visible. Two shifts to Fagaceae are reconstructed to have occurred in *Ectoedemia*, one in *Zimmermannia*, and five in *Stigmella*, of which two have led to a substantial radiation. These shifts are all estimated to have occurred during the late Oligocene or Neogene. There have been four shifts to Rosaceae in Lithocolletinae, during the Neogene, with only one substantial radiation which includes species that are morphologically difficult to distinguish (Triberti, 2007). In Nepticulidae, we find at least five shifts to Rosaceae: one in *Ectoedemia* around 33 Ma [27–36], and at least four in *Stigmella*, the oldest of which may date back 43 Ma [37–48], Eocene, others are at the earliest of Miocene origin. Salicaceae have been colonized twice in Lithocolletinae, both in *Phyllonorycter*. The first occurred early in the Miocene, 24 Ma [17–31], the second was a shift from Fabaceae to Salicaceae later during the Miocene, 10 Ma [6–15]. Both have resulted in groups with a large number of species. In Nepticulidae, there have been at least four shifts to Salicaceae: one in *Ectoedemia*, one in *Zimmermannia*, one in *Fomoria* and two in *Stigmella*, all during the Miocene or more recent. It is noteworthy that in contrast to *Ectoedemia*, *Zimmermannia* and *Stigmella*, species of *Fomoria* do not feed on willows and poplars, but on different (tropical) subgroups of Salicaceae, which were previously placed in Flacourtiaceae (APG III, 2009). Furthermore, in Lithocolletinae we count at least six independent colonization events of Betulaceae, in Nepticulidae at least ten. All shifts to Betulaceae are estimated to have occurred during the Miocene or more recent, with the possible exception of the association of stem lineage of *Bohemannia* with this family, but

this is based on few data and therefore uncertain. The three independent colonizations of Betulaceae by *Phyllonorycter* are associated with radiations extending across different biogeographic regions. In Nepticulidae, there has been one large Holarctic radiation on Betulaceae in *Stigmella*, and many smaller radiations on Betulaceae in *Ectoedemia*, *Bohemannia*, and *Stigmella*.

Discussion

Robustness of phylogenetic and dating estimates

We note several potential sources of error in our study, but none that should overturn our major conclusions. Firstly, due to the high number of species included with partially missing sequence data, there is phylogenetic uncertainty in the analysed trees, which is not taken into account in the diversification analysis. We addressed this issue by comparing the topology of well-supported trees from datasets with fewer taxa and few missing sequence data with trees resulting from datasets with more taxa and more missing sequence data. We found that the support values suffer from the increase in missing data and rogue taxa, but the topologies do not deviate significantly from each other according to Baker's Gamma Index permutation test (Fig. S2). Dated molecular phylogenies for Lepidoptera are only partly congruent with estimates from the fossil record and there are considerable differences between different studies (Condamine *et al*., 2016; Doorenweerd *et al*., 2016; Sohn *et al*., 2015). Using the revised checklist of nepticulid fossils (Doorenweerd *et al*., 2015b), the estimates for Nepticulidae are likely more reliable than those for Lithocolletinae, where we had to rely on an external calibration. In Lopez-Vaamonde *et al*. (2006), the crown age for *Phyllonorycter* was estimated at 62 Ma [50–76 50% confidence interval]. In our results this is significantly more recent: 45 Ma [33–57 95% HPD]. The difference is due to calibration in Lopez-Vaamonde *et al*. with a fossil assigned to *Phyllocnistis* (Labandeira *et al*., 1994), that should more likely be assigned to the clade with Phyllocnistinae + Marmarinae + Oecophyllembinae (Kawahara *et al*., 2016). This would have resulted in younger age estimates for *Phyllonorycter*, and in turn would make the time lag between the diversification of *Phyllonorycter* and the crown ages of the host plant families even bigger (20). Overall, our age estimates for Nepticulidae and Lithocolletinae appear plausible within the range of confidence values of other studies. The confidence intervals for the age estimates (Fig. S4) are not included in the diversification analyses, but we include them in the relevant parts of the discussion. Sampling intensity can have strong effects on diversification estimates; low levels of sampling and non-random sampling are known to sometimes result in false slowdowns in diversification rates (Moen & Morlon, 2014; Cusimano & Renner, 2010) and some authors have suggested that at least 80% of the diversity needs to be included for reliable results (Toussaint *et al*., 2015). We sampled between 8% and 100% of the species in nepticulid genera and between 11% and 100% for lithocolletine genera. However, as BAMM includes clade-specific sampling estimates, our results should be supported as long as the numbers of described and undescribed species in each clade are roughly

proportional to our estimates (Fig. S3). We have further tested the potential effects of sampling issues by estimating the diversity per genus and comparing this with the outcome of a general sampling estimate, as well as by repeating the diversification analyses on the reference datasets with less than half the number of taxa from the larger trees (Fig. S2). The high similarity of the results between these different approaches suggests that our sampling has effectively been random and the results are reliable.

Clade-specific diversification dynamics

Some large-scale macroevolutionary studies have indicated potential clock-like diversification with regular intervals of splitting lineages (Hedges *et al*., 2015). Our results, however, show that, at family or subfamily resolution, different clades have strongly differing diversification dynamics. This is more in line with other studies with increased resolution for a selection of clades (Condamine *et al*., 2016; Wahlberg *et al*., 2013). The differing results are likely a result of the 'law of large numbers': because the shifts in diversification rates are effectively randomly distributed through time, they succumb when large amounts of data are combined in a single speciation rate graph. There is thus no linear correlation between clade age and diversity; some of the most species-rich clades are relatively young. There is a repeated pattern where clades experience a rate-shift with increased diversification rates, which after this initial burst slows down again. This pattern has also been observed in many other groups of organisms (Ricklefs, 2010). One of the explanations involves density-dependent selection, where the availability of resources is a limiting factor and as more species start using them, competition increases and speciation decreases (Zou *et al*., 2016). Unfortunately, due to a lack of wide-scale studies, we do not know much about the importance of intra- or interspecific competition in leaf-miners. Some studies suggested that, based mostly on insect life tables, competition was not an important factor in population dynamics of leaf-miners (Faeth, 1991; Lawton & MacGarvin, 1986). Other studies suggest that there may be complex ecological interactions, possibly also involving parasitoid predation pressures, where competition is in fact one of the important community shaping factors that result in a maximum number of species that can utilize a single host (Boomsma *et al*., 1987; Morris *et al*., 2004). Another explanation for rapid initial diversification followed by decreasing rates of diversification suggests pathogen coevolution as a selective pressure (Ricklefs, 2010). If there is pathogen-host co-evolution subsequent to the rapid diversification of the leafminers, this may reduce leafminer population sizes and restrict geographic distribution or host plant usage, resulting in a slowdown of diversification. The presence of endosymbiotic bacteria has been shown to have evolutionary implications in the morphologically cryptic *Phyllonorycter blancardella* group (Gutzwiller *et al*., 2015; Triberti, 2007), but presence nor role of bacteria and viruses has not been studied widely in leafminers.

Host family shifts and adaptive radiations

Species belonging to Lithocolletinae and Nepticulidae are highly host specific, most commonly monophagous, occasionally oligophagous and exceptionally

polyphagous, such as *Phyllonorycter messaniella*, which has been recorded from seven host families (De Prins & De Prins 2016). Early ideas on the evolution of herbivorous insects envisaged cospeciation between herbivorous insects and their hosts, but there are other explanations for the observation that related insects feed on related plants (Suchan & Alvarez, 2015; Vienne *et al*., 2013). Phylogenetic studies that examined the role of host plants in the evolution of insect herbivores revealed that patterns are often complex. Some indicate that the colonization of a new host plant family has been key for some bursts of diversification, e.g. in the leafmining flies (Agromyzidae) that colonized the daisy family (Asteraceae) (Winkler *et al*., 2009), others suggest large lags between the diversification of the hosts and the insects (e.g. in thrips [McLeish *et al*., 2013]) and *Phyllonorycter* [Lopez-Vaamonde *et al*., 2006]), and again others suggest a combination [e.g. in weevils (McKenna *et al*., 2009) and sawflies [Nyman *et al*. 2006]). Many of these studies relied on early molecular dating estimates of the Angiosperms, which have not been without controversy (Wilf & Escapa, 2014). They mostly focussed on the period of the main Angiosperm diversification, around 110-90 Ma, when the Angiosperms diversified into groups we now recognize as families (Silvestro *et al*., 2015; Wilf & Escapa, 2014). The relationships of the leafminers with their hosts are often more narrow, involving a few host genera or species. Recent studies on the diversification of Angiosperms indicate intensified diversification at generic and species-level during the Miocene for important hosts for Lithocolletinae and Nepticulidae [e.g. Fagaceae and Betulaceae (Bouchenak-Khelladi *et al*., 2015) and *Salix* (Wu *et al*., 2015)]. Looking at our data from that perspective, we can indeed conclude that with the exception of a few associations (viz. Lithocolletinae with Fabaceae, possibly *Enteucha* with Polygonaceae), the leafmining moths have mostly colonized their host plant families millions of years after they started to diversify (Lopez-Vaamonde *et al*., 2006). However, the timing of the intensified diversification of the leaf miners and hosts is growing closer than previously thought and may likely have been contemporaneous in parts of the evolutionary history (Fig S4, S5). At least for Fagaceae and Betulaceae the diversification has likely been contemporaneous during the Miocene with multiple leafmining species groups (Bouchenak-Khelladi *et al*., 2015). However, none of the host family shifts in Nepticulidae and Lithocolletinae coincide with diversification rate shifts. It thus appears that host plant family shifts have not triggered adaptive radiations, but they have been an important enabling factor for diversification. In other words, host plants are necessary for increased rates of diversification but they are not sufficient. Host plant family shifts do roughly coincide with the timing when hosts become ecologically dominant (Lopez-Vaamonde *et al*., 2006), suggesting that family shifts are mainly driven by opportunity and involved in processes such as host recognition during oviposition.

Intrinsic or extrinsic drivers

Drivers of speciation can be divided in intrinsic drivers, such as functional adaptations, host shifts and population dynamics, and extrinsic drivers, such as climate change, shifting host distributions, pathogens or geological events, that all may have enabled or triggered diversification rate shifts. Scholars who emphasize

intrinsic drivers suggest that evolution is mainly driven by "key innovations" (Simpson 1953), others suggest that the balance between intrinsic and extrinsic drivers is more important (Winkler and Mitter 2009). In our data, we primarily see a role for climate change as an enabling factor, sensu the first condition involved in an adaptive radiation into a new resource zone, i.e. Simpson's "physical access" (Simpson, 1953 p. 207). Since the Cretaceous, the climate has gradually been cooling, increasingly so in periods of the Eocene and Miocene, which resulted in stronger seasonality and latitudinal climate stratification (Zachos *et al*., 2001). The diversification rate shifts we find in our data all coincide, more than anything else, with clades that are primarily found in temperate regions. Vice versa this is not always the case: not all clades with typically temperate species have diversified substantially, even if they have similar age estimates. For these, it may be argued that they have not sufficiently adapted to temperate climate, sensu the second condition for an adaptive radiation, i.e. Simpson's "evolutionary access" (Simpson, 1953 p. 207). Although we have not tested for winter survival strategies rigorously because it is not known for all species, based on our knowledge of the groups it appears that species in clades with higher diversification rates all have evolved ways to survive the winter with a pre-adult stage diapause, whereas the other clades have not. Intrinsic adaptation to winter survival may thus be a "key innovation". Winter has before been identified as a key driver of individual performance, community composition and ecological interactions in terrestrial habitats (Williams *et al*. 2015). In the advent of increased climate change, the potential evolutionary repercussions of this warrant further study.

Conclusion

In summary, we found that diversification patterns in two large groups of leaf-mining moths fit the adaptive radiation scenario of Simpson (1953), where increased speciation rates are the consequence of entering a new adaptive zone through adaptation. We also found that a novel host plant family in itself does not constitute a new adaptive zone, which would fit the escape-and-radiate theory of Ehrlich and Raven (1964). The adaptive zone that would fit all the leaf-mining moths with increased rates of diversification can best be described as the temperate region. However, not all clades that are found in the temperate region have seen increased diversification rates and our results therefore also highlight that "entering" a new zone implies more than just being able to survive. It requires evolutionary adaptation to thrive. For Nepticulidae and Lithocolletinae, we suggest that the key innovation that enabled their diversification in the temperate region was likely adaptation to winter climate by pre-adult diapause. This development was the result of intrinsic evolutionary processes of the moths and resulted in diversification rate shifts at different points in time. Lastly, we should note that the support we find for the adaptive radiation scenario does not necessarily exclude the validity of other theories on speciation, and they each deserve their due attention and testing before we will fully understand which factors drive speciation in the different groups of life.

Acknowledgements

We are indebted to many people who have contributed to the project, by providing specimens and ideas or otherwise helped, whom we unfortunately cannot all name here but they have our utmost gratitude. CD and most of the molecular work has been funded by a grant from the Economic Restructuring Fund (FES) of the Dutch Ministry of Finance; several of the field trips undertaken by CD and EvN that yielded source material have been supported financially by the Uyttenboogaart-Eliasen Foundation. We are very grateful to staff at the Biodiversity Institute of Ontario for their assistance with DNA barcoding.

References

Aberer, A.J., Kobert, K., & Stamatakis, A. (2014) ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

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

Althoff, D.M., Segraves, K.A., & Johnson, M.T.J. (2014) Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology & Evolution*, **29**, 82–89.

APGIII: Bremer, B., Bremer, K., Chase, M.W., Fay, M.F., Reveal, J.L., Soltis, D.E., Soltis, P.S., Stevens, P.F., Anderberg, A.A., Moore, M.J., Olmstead, R.G., Rudall, P.J., Sytsma, K.J., Tank, D.C., Wurdack, K., Xiang, J.Q.Y., & Zmarzty, S. (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, **161**, 105–121.

Becerra, J.X. & Venable, D.L. (1999) Macroevolution of insect-plant associations: The relevance of host biogeography to host affiliation. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 12626–12631.

Boomsma, J., Timmermans, H., Cörvers, C., & Kabout, J. (1987) Monophagous leaf-mining larvae of Stigmella (Lepidoptera: Nepticulidae) on birch: patterns and differentiation in exploitation of the host. *Ecography*, **10**, 206–218.

Bouchenak-Khelladi, Y., Onstein, R.E., Xing, Y., Schwery, O., & Linder, H.P. (2015) On the complexity of triggering evolutionary radiations. *New Phytologist*, **207**, 313–26.

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 Computational Biology*, **10**, e1003537.

Chaboureau, A.C., Sepulchre, P., Donnadieu, Y., & Franc, A. (2014) Tectonic-driven climate change and the diversification of angiosperms. *Proceedings of the National Academy of Science of the United States of America*, **111**, 14066–70.

Condamine, F.L., Clapham, M.E., & Kergoat, G.J. (2016) Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Nature Scientific Reports*, **6**, 19208.

Cusimano, N. & Renner, S.S. (2010) Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology*, **59**, 458–64.

Darwin, C. (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *Nature*, **5**, pp. 502.

Davis, D.R. & De Prins, J. (2011) Systematics and biology of the new genus *Macrosaccus* with descriptions of two new species (Lepidoptera, Gracillariidae). *ZooKeys*, **98**, 29–82.

De Prins, J., Davis, D.R., De Coninck, E., Sohn, J.-C., & Triberti, P. (2013) Systematics, phylogeny and biology of a new genus of Lithocolletinae (Lepidoptera: Gracillariidae) associated with Cistaceae. *Zootaxa*, **3741**, 201.

De Prins, J. & De Prins, W. (2016) *Global Taxonomic Database of Gracillariidae (Lepidoptera)*. Belgian Biodiversity Platform. URL: http://www.gracillariidae.net [accessed on: 10 July 2016].

De Prins, J. & Kawahara, A.Y. (2012) Systematics, revisionary taxonomy, and biodiversity of Afrotropical Lithocolletinae (Lepidoptera: Gracillariidae). *Zootaxa*, **3594**, 1–283.

Doorenweerd, C. (2016) PhylOStack: a phylogenetic analyses pipeline in Ubuntu 14.04 using OpenStack cloud computing: http://cdoorenweerd.github.io/PhylOStack/ [accessed on: 10 July 2016]

Doorenweerd, C., van Nieukerken, E.J., & Hoare, R.J.B. (2016) Phylogeny, classification and divergence times of pygmy leafmining moths (Lepidoptera: Nepticulidae): the earliest lepidopteran radiation on Angiosperms? *Systematic Entomology*, **accepted**.

Doorenweerd, C., van Nieukerken, E.J., & Menken, S.B. (2015a) A global phylogeny of leafmining *Ectoedemia* moths (Lepidoptera: Nepticulidae): exploring host plant family shifts and allopatry as drivers of speciation. *PLoS ONE*, **10**, e0119586.

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

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glockner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., Tang, M., Hofberger, J.A., Smithson, A., Hall, J.C., Blanchette, M., Bureau, T.E., Wright, S.I., dePamphilis, C.W., Eric Schranz, M., Barker, M.S., Conant, G.C., Wahlberg, N., Vogel, H., Pires, J.C., & Wheat, C.W. (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Science of the United States of America*, **112**, 8362–3386.

Ehrlich, P.R. & Raven, P.H. (1964) Butterflies and plants - a study in coevolution. *Evolution*, **18**, 586–608.

Emmet, A.M., Watkinson, I.A., & Wilson, M.R. (1985). Gracillariidae. In: The moths and butterflies of Great Britain and Ireland, 2. Cossidae - Heliodinidae (ed. by J. Heath, A.M. Emmet & J. Heath), The moths and butterflies of Great Britain and Ireland Vol. 2, pp. 244–363, Harley Books, Colchester.

Faeth, S.H. (1991) Effect of Oak Leaf Size on Abundance, Dispersion, and Survival of the Leafminer *Cameraria* sp. (Lepidoptera: Gracillariidae). *Environmental Entomology*, **20**, 196–204.

Forister, M.L., Novotny, V., Panorska, A.K., Baje, L., Basset, Y., Butterill, P.T., Cizek, L., Coley, P.D., Dem, F., Diniz, I.R., Drozd, P., Fox, M., Glassmire, A.E., Hazen, R., Hrcek, J., Jahner, J.P., Kaman, O., Kozubowski, T.J., Kursar, T.A., Lewis, O.T., Lill, J., Marquis, R.J., Miller, S.E., Morais, H.C., Murakami, M., Nickel, H., Pardikes, N.A., Ricklefs, R.E., Singer, M.S., Smilanich, A.M., Stireman, J.O., Villamarin-Cortez, S., Vodka, S., Volf, M., Wagner, D.L., Walla, T., Weiblen, G.D., & Dyer, L.A. (2015) The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Science of the United States of America*, **112**, 442–447.

Foster, P. (2004) Modeling Compositional Heterogeneity. *Systematic Biology*, **53**, 485–495.

Futuyma, D.J. (2013) Evolution, third edition, pp 656. State University of New York, Stony Brook.

Galili, T. (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**, 3718–3720.

Gustafsson, B. & Nieukerken, E.J.v. (1990). Larvae of Nepticulidae. In The Nepticulidae and Opostegidae (Lepidoptera) of NW Europe (ed. by R. Johansson, E.S. Nielsen, E.J. van Nieukerken & B. Gustafsson), Fauna Entomologica Scandinavica Vol. 23, 323–356. E.J. Brill, Leiden.

Gutzwiller, F., Dedeine, F., Kaiser, W., Giron, D., & Lopez-Vaamonde, C. (2015) Correlation between the green-island phenotype and *Wolbachia* infections during the evolutionary diversification of Gracillariidae leaf-mining moths. *Ecology and Evolution*, **5**, 4049–4062.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, **32**, 835-845.

Howard, D.J. & Berlocher, S.H. (1998) Endless Forms. Oxford University Press, New York, Oxford.

Kawahara, A.Y., Plotkin, D., Ohshima, I., Lopez-Vaamonde, C., Houlihan, P.R., Breinholt, J.W., Kawahara, A.Y., Xiao, L., Regier, J.C., Davis, D.R., Kumata, T., Sohn, J.C., De Prins, J., & Mitter, C. (2016) A molecular phylogeny and revised higher-level classification for the

leaf-mining moth family Gracillariidae and its implications for larval host use evolution. *Systematic Entomology*, **in press.**

Kumata, T. (1993) A contribution to the knowledge of the Malaysian Lithocolletinae (Gracillariidae, Lepidoptera), with a revision of Indian *Cameraria* associated with Leguminosae. *Insecta Matsumurana*, **48**, 1–85.

Labandeira, C.C., Dilcher, D.L., Davis, D.R., & Wagner, D.L. (1994) 97-Million years of Angiosperm-Insect association - Paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 12278–12282.

Lanfear, R., Calcott, B., Ho, S.Y., & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lastuvka, Z., Lastuvka, A., & Lopez-Vaamonde, C. (2013) A revision of the *Phyllonorycter ulicicolella* species group with description of a new species (Lepidoptera: Gracillariidae). *SHILAP Revista de Lepidopterologica*, **41**, 251–265.

Lawton, J.H. & MacGarvin, M. (1986). The organization of herbivore communities. In: Community Ecology: Pattern and process (ed. by J. Kikkawa & D.J. Anderson), pp. 163–186. Blackwell, Oxford.

Lopez-Vaamonde, C., Charles, H., Godfray, J., & Cook, J.M. (2003) Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution*, **57**, 1804–1821.

Lopez-Vaamonde, C., Wikstrom, N., Labandeira, C., Godfray, H.C.J., Goodman, S.J., & Cook, J.M. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, **19**, 1314–1326.

Maddison, W.P. & Maddison, D.R. (2015) Mesquite: a modular system for evolutionary analysis. Version 3.04. http://mesquiteproject.org.

Mayr, E. (1947) Ecological Factors in Speciation. *Evolution*, **1**, 263–288.

Mayr, E. (1963) Animal species and evolution, pp 797. The Belknap Press of Harvard University Press, Cambridge, MA.

McKenna, D.D., Sequeira, A.S., Marvaldi, A.E., & Farrell, B.D. (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7083–7088.

McLeish, M.J., Miller, J.T., & Mound, L.A. (2013) Delayed colonisation of *Acacia* by thrips and the timing of host-conservatism and behavioural specialisation. *BMC Evolutionary Biology*, **13**, 188.

Menken, S.B.J., Boomsma, J.J., & Van Nieukerken, E.J. (2010) Large-scale evolutionary patterns of host plant associations in the Lepidoptera. *Evolution*, **64**, 1098–1119.

Mitter, C., Farrell, B., & Wiegmann, B. (1988) The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist*, **132**, 107–128.

Moen, D. & Morlon, H. (2014) Why does diversification slow down? *Trends in Ecology & Evolution*, **29**, 190–197.

Morris, R.J., Lewis, O.T., & Godfray, H.C.J. (2004) Experimental evidence for apparent competition in a tropical forest food web. *Nature*, **428**, 310–313.

Naumann, J., Salomo, K., Der, J.P., Wafula, E.K., Bolin, J.F., Maass, E., Frenzke, L., Samain, M.S., Neinhuis, C., dePamphilis, C.W., & Wanke, S. (2013) Single-copy nuclear genes place haustorial Hydnoraceae within piperales and reveal a cretaceous origin of multiple parasitic angiosperm lineages. *PLoS ONE*, **8**, e79204.

van Nieukerken, E.J., Doorenweerd, C., & Hoare, R.J.B. (2016) Revised classification and catalogue of global Nepticulidae and Opostegidae (Lepidoptera: Nepticuloidea). *ZooKeys*, **accepted**.

Nyman, T. (2010) To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. *Biological Reviews of the Cambridge Philosophical Society*, **85**, 393–411.

Nyman, T., Farrell, B.-D., Zinovjev, A.-G., & Vikberg, V. (2006) Larval habits, host-plant associations, and speciation in nematine sawflies (Hymenoptera: Tenthredinidae). *Evolution*, **60**, 1622–1637.

Nyman, T., Linder, H.P., Peña, C., Malm, T., & Wahlberg, N. (2012) Climate-driven diversity dynamics in plants and plant-feeding insects. *Ecology Letters*, **15**, 889–898.

Pena, C. & Espeland, M. (2015) Diversity dynamics in Nymphalidae butterflies: effect of phylogenetic uncertainty on diversification rate shift estimates. *PLoS ONE*, **10**, e0120928.

Peña, C. & Malm, T. (2012) VoSeq: A Voucher and DNA Sequence Web Application. *PLoS ONE*, **7**, e39071.

Percy, D.M., Page, R.D.M., & Cronk, Q.C.B. (2004) Plant-insect interactions: double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology*, **53**, 120–127.

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) Bayesian Analysis of Macroevolutionary Mixtures project website. http://bamm-project.org. [accessed 10 July 2016]

Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. (2014) Tracer v1.6. *available from* http://beast.bio.ed.ac.uk/Tracer.

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

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

Ricklefs, R.E. (2010) Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proceedings of the National Academy of Science of the United States of America*, **107**, 1265–1272.

Silvestro, D., Cascales-Minana, B., Bacon, C.D., & Antonelli, A. (2015) Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytologist*, **207**, 425–436.

Simpson, G.G. (1953) The Major Features of Evolution, pp 434. Columbia University Press.

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

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

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Suchan, T. & Alvarez, N. (2015) Fifty years after Ehrlich and Raven, is there support for plant-insect coevolution as a major driver of species diversification? *Entomologia Experimentalis Et Applicata*, **157**, 98–112.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., & Harmon, L.J. (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist*, **207**, 454–67.

Tilmon, K.J. (2008) The evolutionary biology of herbivorous insects: specialization, speciation and radiation, pp 341. University of California Press.

Toussaint, E.F., Condamine, F.L., Hawlitschek, O., Watts, C.H., Porch, N., Hendrich, L., & Balke, M. (2015) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, **64**, 3–24.

Triberti, P. (2007) The *Phyllonorycter* species from Palaearctic Region feeding on Rosaceae (Lepidoptera, Gracillariidae). *Bollettino del Museo Civico di Storia Naturale di Verona*, **31**, 147–221.

Vienne, D.M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M.E., & Giraud, T. (2013) Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, **198**, 347–385.

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

Wahlberg, N., Wheat, C.W., & Pena, C. (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS ONE*, **8**, e80875.

Wiens, J.J., Lapoint, R.T., & Whiteman, N.K. (2015) Herbivory increases diversification across insect clades. *Nature Communications*, **6**, 8370.

Wilf, P. & Escapa, I.H. (2014) Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. *New Phytologist*, **207**, 283-290.

Winkler, I.S. & Mitter, C. (2008). The phylogenetic dimension of insect-plant interactions: a review of recent evidence. In The evolutionary biology of herbivorous insects: specialization, speciation and radiation. (ed. by K.J. Tilmon), pp. 240–263. Universirty of California Press.

Winkler, I.S., Mitter, C., & Scheffer, S.J. (2009) Repeated climate-linked host shifts have promoted diversification in a temperate clade of leaf-mining flies. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18103–18108.

Wu, J., Nyman, T., Wang, D.C., Argus, G.W., Yang, Y.P., & Chen, J.H. (2015) Phylogeny of Salix subgenus Salix s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evolutionary Biology*, **15**, 31.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–93.

Zou, Y., Sang, W., Hausmann, A., & Axmacher, J.C. (2016) High phylogenetic diversity is preserved in species-poor high-elevation temperate moth assemblages. *Scientific Reports*, **6**, 23045.

Supplemental Information

Table S1: Genbank accessions, Sequences available per specimen. Available from https://cdoorenweerd.stackstorage.com/index.php/s/AHeNsGMYLo8Ors0

Fig. S1: Tanglegrams displaying the similarity between the topology of the reference dataset and the topology obtained with the larger dataset. Dotted clades indicate branches that are not present in both trees, the lines connected the two trees indicate the differing placement of taxa. There are differences in both datasets, but not in places in the tree that affect our diversification estimates. Below each tanglegram is the graphical representation of the Baker's gamma index permutation test. The dotted red line refers to the topology of the reference dataset, which functioned as H0 with value 1. The dotted blue line refers to the topology of the full dataset and the test showed there was no significant deviation from H0 for both Lithocolletinae and Nepticulidae. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/KP2fnoD21RidsP8

Fig. S2: Fractions of the known and sampled diversity of each genus relative to the whole dataset. For both groups we managed to sample about half of the known diversity, although some genera are relatively oversampled, others relatively undersampled. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/8iQVD2ivPCcbwph

Fig. S3a: BEAST time-calibrated tree resulting from the full Lithocoletinae dataset, with taxa labels. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/4ADW7iwmtwnXLoY

Fig. S3b: BEAST time-calibrated tree resulting from the full Nepticulidae dataset, with taxa labels. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/Gc5YgNoXKdUPpvC

Fig. S4a: Ancestral state reconstruction trees of host family and distribution for Lithocolletinae. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/h23UWom3aTpfHRJ

Fig. S4b: Ancestral state reconstruction trees of host family and distribution for Nepticulidae. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/sUieYCUVl5WtVHB

Fig. S5a: Alternative diversification rate shift scenarios (credible shift set) for Lithocolletinae. The F values indicate the relative probability of each scenario. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/bH2zYvS14la7RzA

Fig. S5b: Alternative diversification rate shift scenarios (credible shift set) for Nepticulidae. The F values indicate the relative probability of each scenario. Available from: https://cdoorenweerd.stackstorage.com/index.php/s/lkU0mI6h8ueDDq5

7

Epilogue

Epilogue

21st Century systematics

Traditional systematics relied on morphological characters to establish relationships between species, genera and higher systematic ranks. During the 1980's, molecular methods started to become available, first with allozyme markers, later with sequencing of DNA fragments. These methods were slowly adopted in systematics, but early results based on small data sets were often contradictory and this opened a debate between those emphasizing the value of morphological characters versus those claiming molecular data to be more reliable (Patterson, 1987). Furthermore, it was difficult to reconcile or congregate results between different studies because different scientists used disparate methods and DNA markers. A significant development in the beginning of the 21st century was the introduction of DNA barcodes, a universally 'agreed upon' fragment of DNA that can be used to recognize and delimit (animal) species (Hebert *et al*., 2003). Parallel developments allowed scientists to generate increasing amounts of molecular genetic data for phylogenetic reconstructions. In this thesis we were able to synergistically combine the information from comparative morphology, e.g. external morphology, wing venation, genitalia, leaf-mine characters etc., with molecular data; DNA barcodes and up to seven additional Sanger-sequenced markers.

Using DNA barcodes the right way

DNA barcoding using a fragment of the COI gene has not been without controversy. Various authors have argued that the method is often applied improperly or is even principally unsuitable for species delimitation and identification (Collins & Cruickshank, 2012; Rubinoff *et al*., 2006). In chapter 2 of this thesis, DNA barcoding was applied on *Ectoedemia*, a genus that was relatively well-known taxonomically (van Nieukerken, 1985) and addressed several potential caveats. Some studies using DNA barcoding focused on a 'barcode gap'; a pairwise distance cut-off value between sequences that could be used to differentiate species. Some of the automated species delimitation software still rely on such a value (e.g. Puillandre *et al*., 2012). Our study on *Ectoedemia* (chapter 2) included a thorough global sampling of the genus and revealed that it is treacherous to attempt to define a general barcode gap cut-off value. Although it may properly delimit a large percentage of species in the group, there will be unpredictable exceptions with larger intraspecific variation. Such variation may be the result of the current or historical geographic distribution, including refugia, or a deviating mitochondrial genealogical history. We tested a way to address the latter by using a 'second opinion' nuclear DNA barcode marker and found that the use of a secondary barcoding marker can improve the reliability of the species recognition and delimitation, particularly in groups of closely related species where morphological characters are difficult to interpret. We advocated that instead of focusing on a barcoding gap, a more reliable way to employ DNA barcodes is by using a monophyly criterion; the individuals of a species should have a single common ancestor (Mutanen *et al*., 2016). The remaining difficulty is when different monophyletic clades can be recognized and it is unclear which one corresponds to

a species. This is often true for species with a large (e.g. Palaearctic) distribution or when allopatric populations have been analysed. From this, it is evident that DNA barcodes can only be used for species delimitation in conjunction with comparative morphology and life history information. Several chapters in this thesis have exemplified that DNA barcoding cannot and should not attempt to replace conventional morphology based taxonomy, but DNA barcodes are a valuable source of additional data, provided that they are appropriately interpreted (chapters 2, 5, 6).

Phylogenetics: enough is enough?

Next Generation Sequencing (NGS) methods are a collection of fairly recently developed methodologies that have in common that they sequence DNA by other means than Sanger dideoxy cycle sequencing and were actively being developed around the time the work on this thesis started (e.g. Regier *et al*., 2010). NGS methods produce orders of magnitudes more genetic data than can be obtained with Sanger sequencing, but even though the costs for these methods have dropped considerably over the past eight years, there is still a tipping point where Sanger sequenced data is cheaper. The method of choice depends on how much data is required to be able to answer the research question. In chapter 1, I argued that more data is always better when it comes to phylogenetics. Resolving more (all) nodes in a phylogeny with high statistical support requires exponential increases of data. The challenge in this thesis was balancing the taxon sampling, where we attempted to include as many species as possible, with sufficiently sequenced data for each species. The final assembled dataset for Lithocolletinae and Nepticulidae included in total almost 1,000 species, which is about 50% of the known diversity of each group (chapter 6). For such a large amount of species it would have been uneconomic and high risk to apply NGS, because bioinformatic capabilities for processing NGS data were at that time only partially available. Instead, the approach taken in this thesis is that we used morphology, distribution and life history characters to make a selection of material for DNA extraction, and sequenced DNA barcodes for all extracts. From there the species delimitations were established or confirmed and a representative of each species was sequenced for seven additional markers; resulting in alignments with 4,000 to 5,000 base pairs. The phylogenetic trees that we could infer from this data generally have good statistical support for the nodes and were sufficient for the research questions (chapter 6), but future inclusion of NGS data will allow us to be more confident about the exact timing of divergences and some of the relationships, such as the monophyly and relationships of the species groups in *Stigmella*, the Fabaceae-feeding *Phyllonorycter* or the status of the Oriental *Cameraria*.

Timing is crucial

In multiple chapters of this thesis (chapters 4, 5, 6), we have addressed issues with molecular calibration points in Lepidoptera, and primarily the lack thereof. Reliable estimates for divergence timing of both the herbivorous insects and the plant hosts are a first step for testing co-evolutionary theories and multiple studies have shown that the reliability of divergence time estimations depends largely on the number of

calibration points (Magallón *et al*., 2013). For Nepticulidae, we were able to thoroughly study and revise all the literature that potentially involved fossil Nepticulidae (chapter 4). Although this arguably makes Nepticulidae at present the lepidopteran family with the best-studied fossil record, even for this group it is possible that even older nepticulid fossils will present themselves when older paleontological material is investigated, which will push back its estimated age of origin. For Gracillariidae, the few published fossil findings require reinterpretation; the current best age estimates are based on timed phylogenies of all Lepidoptera (chapter 6). All molecular divergence-timing estimates published in this thesis should be interpreted as the best estimations possible with the current knowledge. Increasing the number of properly interpreted fossil calibration points in the Lepidoptera tree of life in the coming years will lead undoubtedly to significant new insights and will be an important advance in the study of macroevolutionary patterns.

If the theoretical shoe fits...

Evolutionary theories predict drivers of speciation, which can be empirically tested from macroevolutionary datasets. However, often there are multiple theories that fit the data because they make overlapping predictions or the data only allows partial testing of the theory. From the phylogenies of Lithocolletinae and Nepticulidae with almost 1,000 species included (chapter 6), we discussed our findings with regard to two influential evolutionary theories: the escape-and-radiate theory (Ehrlich & Raven, 1964) and the adaptive radiation theory (Simpson, 1953). We concluded that there is no support for the escape-and-radiate theory, which predicts that a host plant family shift would result in increased diversification rates. This agrees with findings from other studies involving herbivorous insects from different orders [for reviews see Janz (2011) and Suchan & Alvarez (2015)] and highlights that we should try to understand the role of the plants in the evolution of the herbivorous insects from a different perspective. In this thesis, we suggest that the hosts, in terms of diversity and availability, have primarily played an enabling role by constituting a variety of resources which may be specialised upon. Overall, the diversification patterns that we found in Lithocolletinae and Nepticulidae fit the adaptive radiation theory well, where increased rates of diversification follow from the adaptation to a new resource zone. We found that there have been lineages with increased rates of diversification, that these increases have been at different moments in time even across lineages of moths with similar life history characteristics and that they are likely correlated with functional adaptations, or, "key innovations". However, there are many more theories regarding speciation that should be tested to further understand the mechanisms and drivers involved in speciation (see chapter 1). For example, other than the potential drivers evaluated in this thesis, they may also involve tritrophic relationships with microbionts (Gutzwiller *et al*., 2015), parasitoids (Morris *et al*., 2004) or pathogens (Ricklefs, 2010). The findings from this thesis will hopefully function as a basis for further research in these directions.

References

Collins, R.A. & Cruickshank, R.H. (2012) The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, **13**, 969-975.

Ehrlich, P.R. & Raven, P.H. (1964) Butterflies and plants - a study in coevolution. *Evolution*, **18**, 586–608.

Gutzwiller, F., Dedeine, F., Kaiser, W., Giron, D., & Lopez-Vaamonde, C. (2015) Correlation between the green-island phenotype and Wolbachia infections during the evolutionary diversification of Gracillariidae leaf-mining moths. *Ecology and Evolution*, **5**, 4049–62.

Hebert, P.D.N., Cywinska, A., Ball, S.L., & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Janz, N. (2011) Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 71–89.

Magallón, S., Hilu, K.W., & Quandt, D. (2013) Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, **100**, 556–73.

Morris, R.J., Lewis, O.T., & Godfray, H.C.J. (2004) Experimental evidence for apparent competition in a tropical forest food web. *Nature*, **428**, 310–313.

Mutanen, M., Kivela, S.M., Vos, R.A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., Huemer, P., Dinca, V., van Nieukerken, E.J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaens, T., Efetov, K.A., Hebert, P.D., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., Segerer, A., Tarmann, G., Zahiri, R., & Godfray, H.C. (2016) Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. *Systematic Biology*, early online.

Patterson, C. (1987) Molecules and morphology in evolution: conflict or compromise? pp 240. Cambridge University Press.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, **21**, 1864–1877.

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., & Cunningham, C.W. (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, *463*, 1079–83.

Ricklefs, R.E. (2010) Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 1265–72.

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

Simpson, G.G. (1953) The Major Features of Evolution, pp 434. Columbia University Press.

Suchan, T. & Alvarez, N. (2015) Fifty years after Ehrlich and Raven, is there support for plant–insect coevolution as a major driver of species diversification? *Entomologia Experimentalis Et Applicata*, **157**, 98–112.

8

Summary

Summary

In this thesis, I studied the role of host plants in the diversification of two groups of leaf-mining moths, in order to understand if there had been common drivers of speciation. In chapter 1, I gave a general introduction into the studies of macroevolutionary patterns, introduced some of the theories that have been developed for describing the process of speciation and the potential role for host plants, how leaf mining moths are a suitable group to study the role of host plants and other candidate drivers of speciation and outlined the technical developments that enable diversification studies. I provided a short introduction on the two groups of leaf-miners that have been studied in this thesis, Nepticulidae and Lithocolletinae.

In chapter 2, we investigated how DNA barcoding can be used to delimit species of leafmining moths, in order to establish reliable species definitions that can function as the basis for diversification studies. We used the nepticulid genus *Ectoedemia*, of which we had a good systematic reference and good sample coverage. The results showed that both COI and EF1-α gene fragments proved reliable as an alternative to conventional, i.e. morphological, species identification for the majority of species, and the combination of both markers increased the reliability of species delimitation and validation. This approach has been applied to all other groups that feature in the different thesis chapters. For all groups, we attempted to obtain COI sequences for as many species as possible, and we contributed all these sequences to the international BOLD database.

After supplementing the DNA barcoding data from chapter 2 with additional genetic markers, we reconstructed a phylogeny for *Ectoedemia* in chapter 3. Six nuclear and mitochondrial DNA markers with a total aligned length of 3,692 base pairs were used to infer phylogenetic relationships among 92 species belonging to the subgenus *Ectoedemia* of the genus *Ectoedemia*, representing a thorough taxon sampling with a global coverage. The results support monophyletic species groups that are congruent with published findings based on morphology. We used the obtained phylogeny to explore host plant family association and geographical distribution to investigate if host shifts and allopatry have been instrumental in the speciation of these leaf-mining insects. We found that, even though most species within species groups commonly feed on plants from one family, shifts to a distantly related host family have occasionally occurred throughout the phylogeny and such shifts are most commonly observed towards Betulaceae. The largest radiations have occurred within species groups that feed on Fagaceae, Rosaceae, and Salicaceae. Most species are restricted to one of the seven global biogeographic regions, but within species groups representatives are commonly found in different biogeographic regions. Although we found general patterns with regard to host use and biogeography, there were differences between clades that suggested that different drivers of speciation, probably drivers that we did not examine, have shaped diversity patterns in different clades.

In chapter 4, we studied the fossil record of Nepticulidae in order to obtain reliable calibration points for the family phylogeny. Using our experience with extant global Nepticulidae, we discussed a list of characters that may be used to assign fossil leaf mines to Nepticulidae, and suggested useful methods for classifying relevant fossil material. We presented a checklist of 79 records of Nepticulidae representing adult and leaf-mine fossils mentioned in literature, often with multiple exemplars constituting a single record. We provided our interpretation of these fossils. Eleven records have for the first time been attributed to Nepticulidae. After discarding several dubious records, including one possibly placing the existence of the family at a latest Jurassic position, we conclude that the oldest fossils likely attributable to Nepticulidae are several exemplars representing a variety of species from the Dakota Formation (USA). The relevant strata containing these earliest fossils are now dated at 102 million years ago in age, corresponding to the latest Albian Stage of the Early Cretaceous. Integration of all records in the checklist showed that a continuous presence of nepticulid-like leaf mines preserved as compression–impression fossils and as amber entombment of adults have a fossil record extending to the latest Early Cretaceous.

In chapter 5, a molecular phylogeny for Nepticulidae is presented where the fossil calibrations resulting from chapter 4 were used to calibrate the phylogeny in time. We were able to include 355 species, representing 20 out of 22 extant Nepticulidae genera, and sequenced up to eight gene fragments for all species. Two solid fossil calibration points were used to place the origin of the family in the Early Cretaceous, before the main Angiosperm diversification. Based on the results we proposed a new classification, abandoning all ranks between family and genus, as well as subgenera, to allow for a stable classification. The position of *Enteucha* within Nepticulidae remained somewhat ambiguous, and the species-rich cosmopolitan genus *Stigmella*, with nearly half of all described Nepticulidae, requires further study with more taxa and genes. *Ectoedemia*, *Zimmermannia*, *Acalyptris*, *Etainia*, *Parafomoria*, *Muhabbetana* and *Fomoria* appeared to have diversified in a relatively short evolutionary period, leading to short branches in the molecular phylogeny and unclear suprageneric relations. Otherwise, support values throughout the phylogeny were mostly high and the species groups, genera and higher clades were discussed in respect to their supporting morphological and life history characters. Wing venation characters were confirmed to mostly be reliable and relevant for classification, but some other previously used characters required reinterpretation. The species groups of most genera were recovered, but only partly so in the large genus *Stigmella*. The molecular dating results were compared with existing knowledge on the timing of the Angiosperm radiation and revealed that the diversification of Nepticulidae could largely have been contemporaneous with their hosts, although some of the genera restricted to a single plant family appeared to have begun to diversify before their hosts.

Chapter 6 combines all the data from the previous chapters with many new data and largely forms the synthesis of my thesis work. We tested the central prediction from two classic evolutionary theories, the escape-and-radiate theory and the

adaptive radiation theory, that colonization of a novel resource zone (e.g. a novel host plant family or novel geographic area) would result in increased rates of speciation. In chapter 6, we compared the phylogenetic diversification patterns and possible impact of biogeographical and host plant family shifts of two large groups of leaf-mining moths: pygmy leaf-mining moths (Nepticulidae) and leaf-blotch mining moths (Gracillariidae: Lithocolletinae). We gathered a large molecular dataset by sequencing up to eight markers for 335 species of Lithocolletinae and 645 species of Nepticulidae, for each group encompassing approximately 50% of the known diversity. We found that there is no linear correlation between clade age and diversity; some of the most species-rich clades are relatively young. Additionally, at different points in time there is a repeated pattern where clades experience a rate-shift with increased diversification rates, which after this initial burst slows down again, for which we discussed several possible explanations, such as density-dependent-selection and co-evolution of pathogens. The results showed that the diversification of Lithocolletinae and Nepticulidae has in parts likely been contemporaneous with increased rates of diversification of the hosts. We found no evidence that a novel host plant family in itself would constitute a new adaptive zone, which would fit the escape-and-radiate theory of Ehrlich and Raven, but we do conclude that the diversity of the hosts is a prerequisite for diversification. The diversification patterns we find in Nepticulidae and Lithocolletinae fit the adaptive radiation theory, and their adaptive zone can best be described as the temperate region. Our results also highlighted that "entering" a new zone requires evolutionary adaptation in order to be able to diversify. We suggested that the clades that experienced increased rates of diversification are uniquely characterised by a pre-adult winter diapause, which may have been a "key innovation", a prerequisite to successfully adapt to the temperate climate.

In chapter 7, I discuss the consequences of the findings in this thesis for the field of research. We found more than 100 undescribed species with our studies, which will be described elsewhere. For both Nepticulidae and Lithocolletinae, the phylogenies presented in this thesis are the first with high species-level coverage, which enables the reliable testing of evolutionary theories regarding speciation, as well as hypotheses in other fields of study, such as functional morphology. By comparing the diversification patterns of two large groups of leaf-miners in time we were able to, for the first time, estimate to what extent their radiation has been contemporaneous - with respect to each other as well as with their plant hosts. We found that host plants have not been a principle driver in the diversification of these leaf-miners, but host plant diversity has been a prerequisite for leaf-miner diversification. Under favourable changes in climate, the diversification of the host plants, particularly Fagaceae, Salicaceae, Betulaceae, Rosaceae and Fabaceae in temperate regions likely generated a novel resource space for Nepticulidae and Lithocolletinae. Several lineages of these leaf-miners were able to successfully adapt to survival in a temperate climate, likely primarily by adaptation to winter climate, after which their diversification rates initially increased and eventually led to the diverse groups of leaf-miners we find today in temperate regions.

Samenvatting

In dit proefschrift heb ik de rol van waardplanten bij de diversificatie van twee groepen van bladminerende vlinders onderzocht, om te begrijpen of er overeenkomende sturende krachten voor soortvorming zijn. In hoofdstuk 1 gaf ik een algemene inleiding tot macro-evolutionair onderzoek, introduceerde een aantal theoriën over het proces van soortvorming en de potentiële rol van waardplanten, waarom bladmineerders een geschikte groep zijn voor onderzoek naar de rol van waardplanten bij soortsvorming en andere mogelijke sturende krachten voor soortvorming, en ik heb de technische ontwikkelingen omschreven die dit onderzoek mogelijk maken. Ik gaf een korte inleiding tot de twee bladminerende groepen die ik heb onderzocht: Nepticulidae en Lithocolletinae.

In hoofdstuk 2 onderzochten we hoe met DNA barcoding de soortsgrenzen van bladminerende vlinders kunnen herkennen, zodat we betrouwbare soortsdefinities op kunnen stellen die de basis vormen voor onderzoek naar diversificatie. Hiervoor gebruikten we het geslacht *Ectoedemia*, waarvoor veel systematische kennis gepubliceerd was en we veel materiaal beschikbaar hadden. De resultaten toonden aan dat zowel fragmenten van het gen COI als het gen EF1- α in de meeste gevallen betrouwbaar waren als alternatief voor conventionele, morfologische, herkenning van soorten, en voor het bepalen van wat een soort is. De combinatie van beide merkers verhoogde de betrouwbaarheid van soortsherkenning. Deze aanpak werd verder gebruikt voor de alle andere groepen in de verschillende hoofdstukken van dit proefschrift. Voor alle soorten probeerden we COI fragment sequenties te bepalen en deze DNA sequenties zijn opgeslagen in de internationale BOLD database.

Nadat de DNA barcode dataset van hoofdstuk 2 aangevuld was met extra genetische merkers, werd de fylogenie (stamboom) van *Ectoedemia* gereconstrueerd in hoofdstuk 3. Hiervoor werden zes kerngenoom markers en mitochondrieel genoom markers gecombineerd tot een dataset met 3,692 nucleotiden en gebruikt om de fylogenetische relaties tussen 92 soorten van het ondergeslacht *Ectoedemia* te bepalen, die een goede vertegenwoordiging vormen van het totaal aantal soorten wereldwijd. De resultaten ondersteunen soortgroepen die op basis van morfologie gedefinieerd waren. We gebruikten de verkregen fylogenie om te kijken naar waardplant familie assocaties en geografische verspreiding, om te zien of deze van belang zijn geweest bij de soortvorming van deze bladminerende insecten. We ontdekten dat, ondanks dat soorten binnen een soortgroep meestal op planten zitten van één familie, overstappen naar andere waardplant families soms voor zijn gekomen op verschillende plekken in de fylogenie en de overstappen zijn dan het vaakst naar de berkenfamilie (Betulaceae). De grootste radiaties vinden we in groepen die eten van eiken en beuken (Fagaceae), roosachtigen (Rosaceae) en wilgen en populieren (Salicaceae). De verspreiding van de meeste soorten is beperkt tot een van de zeven biogeografische regio's, maar binnen soortgroepen zijn er regelmatig overstappen naar een andere regio geweest. Ondanks de algemene patronen die we herkennen in het gebruik van waardplanten en biogeografie, vonden we ook verschillen tussen

verschillende takken in de fylogenie die suggereren dat er voor verschillende groepen verschillende drijvende krachten van soortvorming zijn geweest, mogelijk krachten die we niet onderzocht hebben.

In hoofdstuk vier hebben we alle vermeldingen van fossielen van Nepticulidae bestudeerd om betrouwbare calibratiepunten voor de fylogenie vast te stellen. Gebruik makende van onze kennis van Nepticulidae wereldwijd, stelden we een lijst van kenmerken die gebruikt kunnen worden om fossiele bladmijnen toe te schrijven aan Nepticulidae, en we stellen methoden voor om fossiel materiaal te classificeren. De checklist bevat 79 vermeldingen van fossiele adulten en bladmijnen van Nepticulidae, vaak met meerdere exemplaren per vermelding. We geven onze interpretatie van alle vermeldingen. Elf vermeldingen worden in de checklist voor het eerst toegeschreven aan Nepticulidae. Na het verwerpen van enkele dubieuze vermeldingen, inclusief een die het bestaan van de familie mogelijk in het laat Jura zou plaatsen, concludeerden we dat de oudste fossielen die aan Nepticulidae toegeschreven kunnen worden enkele fossiele bladmijnen, waarschijnlijk van meerdere soorten, uit de Dakota formatie (USA) zijn. Deze formatie wordt op dit moment geschat op 102 miljoen jaar oud, wat overeen komt met de laatste periode van het Vroege Krijt. Door alle vermeldingen te combineren konden we concluderen dat er sinds de oudste fossielen een min of meer continue aanwezigheid van bladmijn- of barnsteenfossielen is die aan Nepticulidae toe te schrijven zijn.

In hoofdstuk 5 presenteren we een moleculaire fylogenie voor Nepticulidae, waarbij we de fossiele calibratiepunten van hoofdstuk 4 gebruikt hebben om de fylogenie in de tijd te calibereren. Op dit punt van het onderzoek konden we aan de hand van acht genetische merkers, 355 soorten in de fylogenie plaatsen, die 20 van de 22 geslachten van Nepticulidae vertegenwoordigen. Twee solide calibratiepunten werden gebruikt om het ontstaan van de familie in het Vroege Krijt te plaatsen, voor de voornaamste diversificatie van de bloemvormende planten (Angiospermae). Op basis van de resultaten wordt een nieuwe classificatie van de familie voorgesteld, waarbij de systematische rangen tussen familie en geslacht niet meer gebruikt worden, alsook de ondergeslachten, om zo tot een duurzame classificatie te komen. De positie van *Enteucaha* binnen Nepticulidae blijft wat onduidelijk, en het soortenrijke wereldwijde geslacht *Stigmella*, met meer dan de helft van alle beschreven Nepticulidae, heeft nog meer onderzoek nodig met meer genen en soorten. De geslachten *Ectoedemia*, *Zimmermannia*, *Acalyptris*, *Etainia*, *Parafomoria*, *Muhabbetana* en *Fomoria* lijken in korte tijd van elkaar te zijn afgesplitst, wat tot korte takken in de fylogenie leidt en onduidelijke intergeslachtelijke verwantschappen. Verder is de statistische ondersteuning voor de takken in de fylogenie over het algemeen hoog en de verwantschappen tussen soortgroepen, geslachten en hogere groepen zijn bediscussieerd in relatie tot hun morfologische en biologische kenmerken. Vleugeladering kenmerken werden bevestigd als betrouwbaar en belangrijk voor het indelen van groepen, maar sommige andere van de eerder gebruikte kenmerken moeten opnieuw geïnterpreteerd worden. De soortgroepen binnen de meeste geslachten werden teruggevonden in de fylogenie, maar slechts deels in het grote geslacht *Stigmella*. De resultaten van de moleculaire

datering werden vergeleken met de bestaande kennis van de diversificatie van de bloemvormende planten en daaruit bleek dat de diversificatie van Nepticulidae voor een groot deel in dezelfde periode plaats heeft kunnen vinden als die van de waardplanten, hoewel de diversificatie van sommige van de geslachten die maar op één waardplant familie voorkomen ouder lijkt te zijn dan de waarplant familie zelf.

Hoofdstuk 6 combineert alle gegevens van de voorgaande hoofdstukken met veel nieuwe gegevens en vormt de synthese van het promotiewerk. We testte de voorspelling van twee klassieke evolutionaire theoriën, de 'escape-and-radiate' theorie en de 'adaptive radiation' theorie, dat de kolonisatie van nieuwe 'resource zones' (bijvoorbeeld een nieuwe waardplantfamilie of een nieuw geografisch gebied) zou resulteren in versnelde soortvorming. We vergeleken de fylogenetische diversificatie patronen en bestudeerden de mogelijke gevolgen van waardplantfamilie overstappen in twee grote groepen van bladminerende vlinders: de dwergmineermotten (Nepticulidae) en vouwmotten (Lithocolletinae). We verzamelden een grote genetische dataset door van acht merkers de genetische code te bepalen voor 355 soorten Lithocolletinae en 645 soorten Nepticuldiae. Voor elke groep is dit ongeveer 50% van alle bekende soorten. De resultaten lieten zien dat er geen linaire relatie is tussen de leeftijd van een groep en het aantal soorten; sommige van de meest soortenrijke groepen zijn relatief jong. Daarbij, op verschillende momenten in de tijd is er versnelde soortvorming gedetecteerd voor verschillende groepen, en na een kort moment van versnelling zwakt dit weer af. Hiervoor bespreken we enkele mogelijke verklaringen, bijvoorbeeld 'density-dependent-selection' en de co-evolutie van pathogenen. De resultaten toonden verder aan dat de diversificatie van Lithocolletinae en Nepticulidae gedeeltelijk gelijktijdig heeft plaatsgevonden met versnelde soortvorming bij de belangrijkste waardplanten. We vonden geen bewijs dat een nieuwe waardplant familie op zichzelf een nieuwe 'adapative resource zone' is, de voorspelling van de 'escape-and-radiate' theorie van Ehrlich en Raven, maar we concluderen dat de diversiteit van de waardplanten wel een voorwaarde is voor de diversificatie van de motten. De diversificatiepatronen die we vinden in Nepticulidae en Lithocolletinae passen bij de voorspellingen van de 'adaptive radiation' theorie, en de adaptieve zone kan het best omschreven worden als de geografische zone met een gematigd klimaat. Onze resultaten laten verder zien dat het binnenkomen van een adaptieve zone gepaard moet gaan met de ontwikkeling van functionele aanpassingen, om tot versnelde soortvorming te komen. Mogelijk was deze aanpassing het overleven van de winter in een pre-adult stadium, wat dan een zogenaamde 'key innovation' genoemd kan worden.

In hoofdstuk 7 bediscussieer, en kijk ik terug op, de resultaten van dit proefschrift vanuit een breder perspectief. We ontdekten meer dan 100 nieuwe soorten voor de wetenschap, die in andere artikelen beschreven zullen worden. Voor zowel Nepticulidae als Lithocolletinae, zijn de fylogeniën die in dit proefschrift gepresenteerd worden de eersten met een dergelijke soorten-dekking, wat het mogelijk maakt om evolutionaire theoriën te testen, en theoriën uit andere vakgebieden, zoals bijvoorbeeld functionele morfologie. Door de diversificatie patronen van twee grote groepen bladmineerders te vergelijken konden we voor het

eerst inschatten in hoeverre hun evolutie gelijktijdig plaats heeft gevonden, met elkaar en ten opzichte van de waardplanten. De resultaten lieten zien dat de waardplanten niet de voornaamste drijvende kracht zijn geweest voor de diversificatie van de bladmineerders, maar de diversiteit van de waardplanten was wel een voorwaarde voor de diversificatie van de bladmineerders. Onder gunstige klimatologische omstandigheden sinds het Krijt konden de belangrijkste waardplant families, Fagaceae, Salicaceae, Betulaceae, Rosaceae en Fabaceae, zich geografisch uitbreiden en ontstonden er veel nieuwe soorten in gebieden met een gematigd klimaat, waardoor er een nieuwe adaptieve zone voor Nepticulidae en Lithocolletinae ontstond. Verschillende bladmineerders wisten zich succesvol aan te passen aan het overleven in een gematigd klimaat, waarschijnlijk voornamelijk door nieuwe manieren om de winter te overleven, en daarna volgden perioden met versnelde soortvorming, wat uiteindelijk leidde tot de soortenrijke groepen bladmineerders die we tegenwoordig terugvinden in gebieden met een gematigd klimaat.

9

Word of thanks

Word of thanks

Throughout the work that finally led to this thesis I have been fortunate to have physical and mental support from many people. At Naturalis, the many colleagues from different departments have been great to work with and have helped me in myriad ways. The list of helpful colleagues is not limited to Naturalis; studies like the one presented are only possible through large collaborative efforts of specialists and naturalists from around the world. I am grateful to everyone who supplied specimens, either from their collections or freshly collected material that we could include in our studies and made the project what it has become.

From Naturalis, I would like to particularly thank our small Lepidoptera working group; the guidance from copromotor and office companion Erik van Nieukerken, striving for perfected methods with Kees van den Berg, and the inspiration from our students and guest researchers, in particular Sjaak Koster. The fieldwork has probably been the most enjoyable part of the production of this thesis, but working in the molecular lab always felt very comforting in a pleasant environment with great people where we discussed both the heavier and lighter subjects of life with equal ease. I thank all my colleagues at the Sylvius laboratory for that. Steph Menken, thank you for being my promotor and guiding me, and I appreciate the ease with which you could spare time and ideas and the structured discussions that followed.

Hard work requires one to also relax hard. Luckily there were always sufficient people surrounding me to take my mind of tantalizing scientific questions, with the aid of special quality beers, pizza, movies, mothing at the Klip, fitness sessions, renovating houses and of late, climbing. I will not attempt to mention everyone, but you know who you are. A special thanks to Nienke, both your support and confusion on my professional passion has helped me a great deal.

That the enthusiasm for the natural world is beyond those who have studied biology or work in a natural history museum became clear to me when I joined the microlepidoptera section “Snellen” of the Dutch Entomological Society. People from all different backgrounds unite in this group to share their findings regarding small Lepidoptera, and besides enjoyable Saturdays with interesting discussions this has led to several joint publications. Thank you all and I look forward to many more years of meeting each other. My membership of the Societas Europaeas Lepidopterologica (SEL) has been likewise enjoyable. The biyearly conferences have proven to be a great place to meet like-minded people and start exciting research projects. Everyone I spoke with was very open and supported me however possible, for which I am thankful.

Last, but certainly not least, I would like to thank my family for helping me get to the point where I could start working on a thesis, and actually finish one. Peter, Marijke, Nathalie and Yoram, thank you for bearing with me these years and I hope that you appreciate the outcome as much as I do.

CV and list of publications

CV and list of publications

Camiel Doorenweerd is born on the 3rd of May 1984, in Middelburg, in the sunniest province of The Netherlands: Zeeland. After obtaining his VWO diploma from the Christelijke Scholengemeenschap Walcheren in 2003, he starts the study of Medical Biology at the University of Nijmegen, currently known as the Radboud University. Into his first year he quickly realises that his interests in biology go much beyond medical biology, and completes the bachelors program as a student of Biology. For his third year bachelors internship he studies the genetic diversity of the endangered water soldier (*Stratiotes aloides* L.) under supervision of Dr. Joop Ouborg, where he is first exposed to the potential of molecular methods for understanding evolutionary processes.

He starts his Masters program at the Radboud University in 2006, aiming to specialise in ecology and evolution using molecular methods. His first nine-month internship, again with Dr. Joop Ouborg, focussed on the effect of the seasonal flooding of the river Waal on the population structure and distribution of yellow waterlily (*Nuphar lutea* L.). At this point his interest in larger evolutionary patterns had began to grow, as well as the awe for the beauty and diversity of insects and spiders. Combining these two interests, he sought a second nine-month internship at the Natural History Museum Naturalis, in Leiden. Dr. Erik van Nieukerken answered this call, and offered a study on the DNA barcoding and phylogenetics of the leafmining moth genus *Ectoedemia* Busck. With some time to spare before the start of this internship, Camiel worked for one year as the student-assessor in the Faculty Council of the Faculty of Natural Sciences and Informatics at the Radboud University, where he voiced the student opinions and learned about management. Returning to science with his internship at Naturalis, he found his subject not only satisfactory on theoretical aspects, but also in how it linked field work with the challenge to portray the genetic patterns that, following meticulous work in the molecular laboratory, appear on a computer screen to real-world scenarios, as well as the practical challenges of handling small moths and curating an entomological collection. Camiel finished his masters bene meritum in 2009.

Following his graduation Camiel works several temporary contracts with curating the Lepidoptera collection at Naturalis, until in 2010 Naturalis offered him a job to create the molecular laboratory infrastructure for a four year project that contributed to a successful international project generating so-called DNA barcode sequences for all biodiversity on earth. In the meantime, he worked together with Dr. Erik van Nieukerken and Prof. Dr. Steph B.J. Menken on a proposal for a PhD project, which was granted in 2012.

Over the years fieldwork was carried out at inspiring locations all over the world, including France, Greece, several south-eastern states of the USA, and during the PhD period in Taiwan, Japan and even Hawai'i, the results of which are included in the various chapters of this thesis.

Publications in peer review journals

Doorenweerd, C., van Nieukerken, E.J., & Hoare, R.J.B. (2016) Phylogeny, classification and divergence times of pygmy leafmining moths (Lepidoptera: Nepticulidae): the earliest lepidopteran radiation on Angiosperms? *Systematic Entomology,* Accepted.

Nieukerken, E.J.v., **Doorenweerd, C.**, Nishida, K., & Snyers, C. (2016) New taxa, including two new genera show uniqueness of Neotropic Nepticulidae (Lepidoptera). *ZooKeys*, Accepted.

Nieukerken, E.J.v., **Doorenweerd, C.**, & Hoare, R.J.B. (2016) Revised classification and catalogue of global Nepticulidae and Opostegidae (Lepidoptera: Nepticuloidea). *ZooKeys*, Accepted.

Breeschoten, T., **Doorenweerd, C.**, Tarasov, S., & Vogler, A.P. (2016) Phylogenetics and biogeography of the dung beetle genus *Onthophagus* inferred from mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 105, 86-95.

Mutanen, M., Kivela, S.M., Vos, R.A., **Doorenweerd, C.**, Ratnasingham, S., Hausmann, A., Huemer, P., Dinca, V., van Nieukerken, E.J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaens, T., Efetov, K.A., Hebert, P.D., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., Segerer, A., Tarmann, G., Zahiri, R., & Godfray, H.C. (2016) Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. *Systematic Biology,* early online 20 june 2016

Jordan, M.P., Langmaid, J.R., & **Doorenweerd, C.** (2016) Morphological difference between upperside and underside leaf-mining larvae of *Phyllocnistis unipunctella* (Stephens, 1834) (Lep.:Gracillariidae) and its changing phenology. *Entomologist's Record and Journal of Variation*, **128**, 121-127.

van Roosmalen, J.A.M. & **Doorenweerd, C.** (2015) *Coleophora gryphipennella* (Hübner, 1796) (Lepidoptera, Coleophoridae) on Fragaria vesca L. (Rosaceae), a novel host, in the coastal dunes of The Netherlands. *Nota Lepidopterologica*, **38**, 147-155.

Doorenweerd, C., van Nieukerken, E.J., & Menken, S.B. (2015) A global phylogeny of leafmining *Ectoedemia* moths (Lepidoptera: Nepticulidae): exploring host plant family shifts and allopatry as drivers of speciation. *PLoS ONE*, **10**, e0119586.

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

Doorenweerd, C., van Haren, M.M., Schermer, M., Pieterse, S., & van Nieukerken, E.J. (2014) A Linnaeus NG (TM) interactive key to the Lithocolletinae of North-West Europe aimed at accelerating the accumulation of reliable biodiversity data (Lepidoptera, Gracillariidae). *ZooKeys*, **422**, 87-101.

Trimbos, K.B., **Doorenweerd, C.**, Kraaijeveld, K., Musters, C.J.M., Groen, N.M., de Knijff, P., Piersma, T., & de Snoo, G.R. (2014) Patterns in Nuclear and Mitochondrial DNA Reveal Historical and Recent Isolation in the Black-Tailed Godwit (*Limosa limosa*). *PLoS ONE*, **9**, e83949.

Doorenweerd, C., As, B.v., & Scheffers, J. (2014) The explosive expansion of the lime leaf miner in Europe: invading the Netherlands? *Entomologische Berichten,* **74**, 111-114.

Groenen, F., Huisman, K.J.H., & **Doorenweerd, C.** (2013) *Phalonidia manniana*, a complex of two species: *Ph. manniana* en *Ph. udana* (Lepidoptera: Tortricidae). *Entomologische Berichten*, **73**, 191-196.

van Nieukerken, E.J., **Doorenweerd, C.**, Stokvis, F.R., & Groenenberg, D.S.J. (2012) DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-α: two are better than one in recognising cryptic species. *Contributions to Zoology*, **81**, 1-24.

van Nieukerken, E.J., **Doorenweerd, C.**, Ellis, W.N., Huisman, K.J., Koster, J.C., Mey, W., Muus, T.S.T., & Schreurs, A. (2012) *Bucculatrix ainsliella* Murtfeldt, a new North American invader already widespread on northern red oaks (*Quercus rubra*) in Western Europe (Bucculatricidae). *Nota Lepidopterologica*, **35**, 135-159.

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

van Nieukerken, E.J., Mutanen, M., & **Doorenweerd, C.** (2012) DNA barcoding resolves species complexes in *Stigmella salicis* and *S. aurella* species groups and shows additional cryptic speciation in *S. salicis* (Lepidoptera: Nepticulidae). *Entomologisk Tidskrift*, **132**, 235-255.

Thesis co-authors and contribution to the thesis

Chapter 2
Erik J. van Nieukerken and Camiel Doorenweerd conceived the idea, Camiel Doorenweerd, Dick S.J. Groenenberg and Frank R. Stokvis designed and performed the experiments, Camiel Doorenweerd analysed the data, Erik J. van Nieukerken and Camiel Doorenweerd wrote the paper.

Chapter 3
Camiel Doorenweerd Erik J. van Nieukerken Steph B.J. Menken conceived and designed the experiments. Camiel Doorenweerd performed the experiments and analyzed the data. Camiel Doorenweerd Erik J. van Nieukerken Steph B.J. Menken wrote the paper.

Chapter 4
Camiel Doorenweerd conceived the idea, Camiel Doorenweerd, Erik J. van Nieukerken, Jay-Cheon Sohn and Conrad C. Labandeira gathered the data, Camiel Doorenweerd and Erik J. van Nieukerken analysed the data, Camiel Doorenweerd, Erik J. van Nieukerken, Jay-Cheon Sohn and Conrad C. Labandeira wrote the paper.

Chapter 5
Camiel Doorenweerd, Erik J. van Nieukerken and Robert J.B. Hoare conceived the ideas, Erik J. van Nieukerken and Robert J.B. Hoare gathered data and did morphological studies, Camiel Doorenweerd performed the molecular experiments, Camiel Doorenweerd, Erik J. van Nieukerken and Robert J.B. Hoare wrote the paper.

Chapter 6
Camiel Doorenweerd, Erik J. van Nieukerken, Carlos Lopez-Vaamonde and Steph B.J. Menken conceived the ideas, Camiel Doorenweerd and Carlos Lopez-Vaamonde performed the experiments, Camiel Doorenweerd analysed the data and Camiel Doorenweerd, Erik J. van Nieukerken, Carlos Lopez-Vaamonde and Steph B.J. Menken wrote the paper. The interpretation of the results has been discussed with attending scientists of the International Gracillariidae Symposium in Hawai'i, 2016.