

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 Nieuwerkerken^{1,2}, Camiel Doorenweerd¹, Frank R. Stokvis¹, Dick S.J. Groenenberg¹

¹Netherlands Centre for Biodiversity Naturalis, PO Box 9517, 2300 RA Leiden, The Netherlands

²E-mail: nieuwerkerken@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.

Contents

Introduction	1
Material and methods	3
Material	3
Extraction, amplification and sequencing	5
Primers	5
PCR	5
Alignment	5
Neighbor joining trees and distances	5
Diagnostic positions	5
Results	6
Quality material	7
COI versus EF1- α	7

Species recognition	7
The <i>Ectoedemia angulifasciella</i> group	7
The <i>Ectoedemia suberis</i> group	10
The <i>Ectoedemia populella</i> group	10
The <i>Ectoedemia subbimaculella</i> group	11
Discussion	13
One or two genes	13
Barcoding gap	15
Species complexes	15
Introduced species	16
Conclusion	16
Acknowledgements	16
References	17
Appendix	20

Introduction

The Lepidoptera are one of the megadiverse groups of organisms, with currently more than 157,000 described species (van Nieuwerkerken *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 Hesperidae 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), Öunap 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 Palearctic 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* (Szöcs, 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

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 <i>et al.</i> 2004a
LepR1-short	COI	R	TAAACTTCTGGATGTCCAAAAA	Hebert <i>et al.</i> 2004a
T-LepF1-short	COI	F	TAATACGACTCACTATAGGG ATTCAACCAATCATAAAGATAT	new
T-LepR1-short	COI	R	ATTAACCCCTCACTAAAGTAAACTTCTGGATGTCCAAAAA	new
MLepF1* (MF1)	COI	F	GCTTTCCCACGAATAAATAATA	Hajibabaei <i>et al.</i> 2006a
MLepR1* (MH-MR1)	COI	R	CCTGTTCCAGCTCCATTTTC	Hajibabaei <i>et al.</i> 2006a
EF-NepF	EF1- α	F	GCCCCCGGACACAGAGATTTCA	new
EF-NepR	EF1- α	R	CACGACCTACTGGCACTGTTC	new
T-EF-NepF	EF1- α	F	TAATACGACTCACTATAGGG GCCCCCGGACACAGAGATTTCA	new
T-EF-NepR	EF1- α	R	ATTAACCCCTCACTAAAGC CACGACCTACTGGCACTGTTC	new
T-EFcdwF2*	EF1- α	F	TAATACGACTCACTATAGGG CCAGATTYGARGAAATYAAR	new
T-EFcdwR*	EF1- α	R	ATTAACCCCTCACTAAAGG CAACDGCAGCTGGRTTRTA	new

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						
<i>directly in ethanol 96% in minus 80</i>						
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%
<i>older ethanol 70% material</i>						
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						
<i>Abdomens, non destructive</i>						
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%
<i>Hindlegs</i>						
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%

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 rM52.6 (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.

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 μ l. 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.

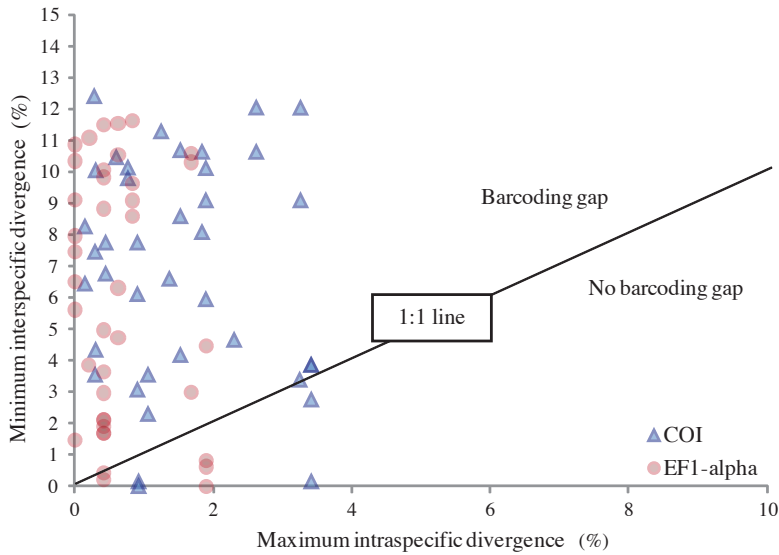


Fig. 1a. Comparison of maximum interspecific divergence versus maximum intraspecific divergence percentages between 39 *Ectoedemia* s. str. species pairs for which multiple sequences for both markers were available. EF1- α reaches the same minimum interspecific distances as COI, but the maximum intraspecific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

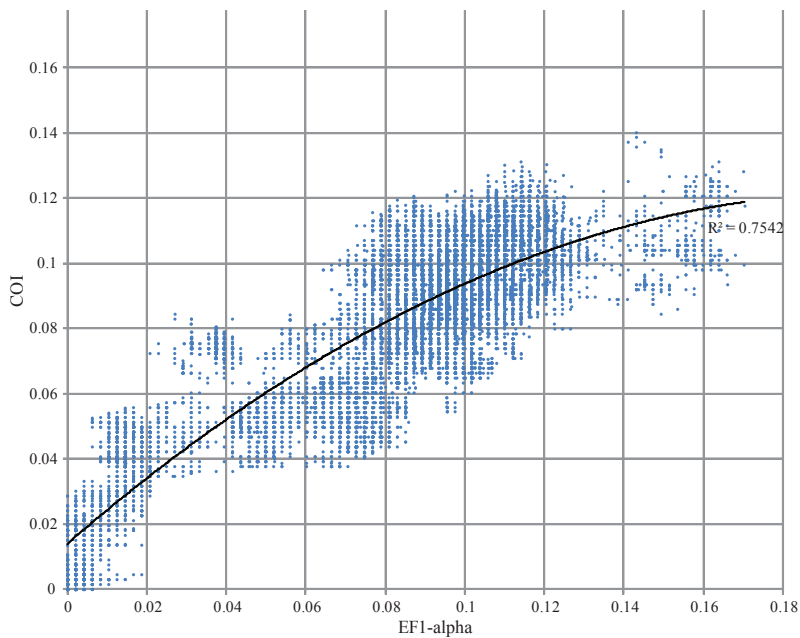


Fig. 1b. Scattergram containing the 21,945 pairwise distance values of COI and EF1- α between all specimens. The polynomial trend line gave the best fit for the data, with an R^2 of 0.75. The data does not show that either marker evolves at a higher rate in general, but closely related specimens show larger distances in COI, where more distantly related specimens are the exact opposite of this and show larger distances in EF1- α .

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 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 data-

points 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. However, the most fitting trend line, polynomial 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 spe-

cies closely related to the aforementioned three, and likely to be found mining *Rubus* sp.

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

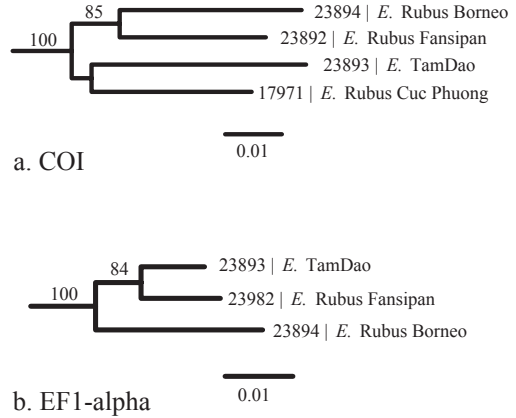


Fig. 2. K2P Neighbor joining trees containing possibly four species in the *Ectoedemia angulifasciella* group. Specimen numbers are RMNH.INS registry numbers. All specimens were collected in Vietnam or Borneo, at light as adult or as larva on *Rubus* spp. 2a: The COI tree. The distances between specimens 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.

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).

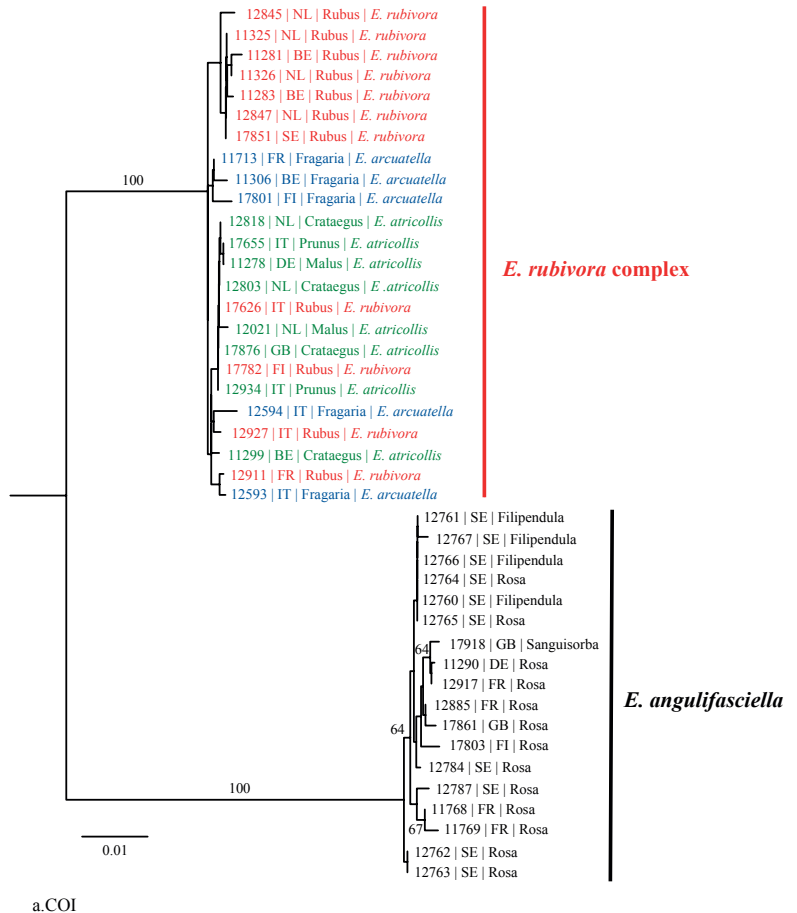


Table 3. Maximum K2P intraspecific pairwise distance of *Ectoedemia* sequences exceeding 2%

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

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.

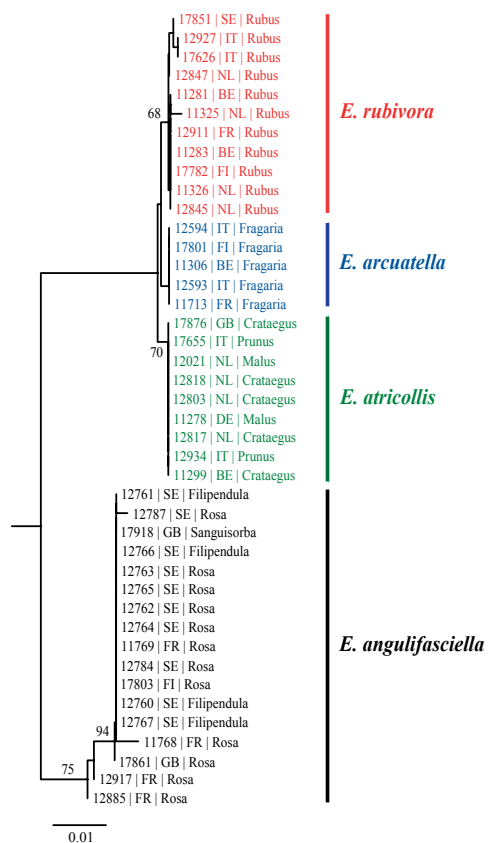
Specimen RMNH.INS.23741 was discovered in

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. 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,



b. EF1-alpha

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
<i>E. atricollis</i>	T	C
<i>E. rubivora</i>	C	T
<i>E. arcuatella</i>	C	C

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- α , 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 Nieuwerkerken 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.

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 Nieuwerkerken 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 morpho-

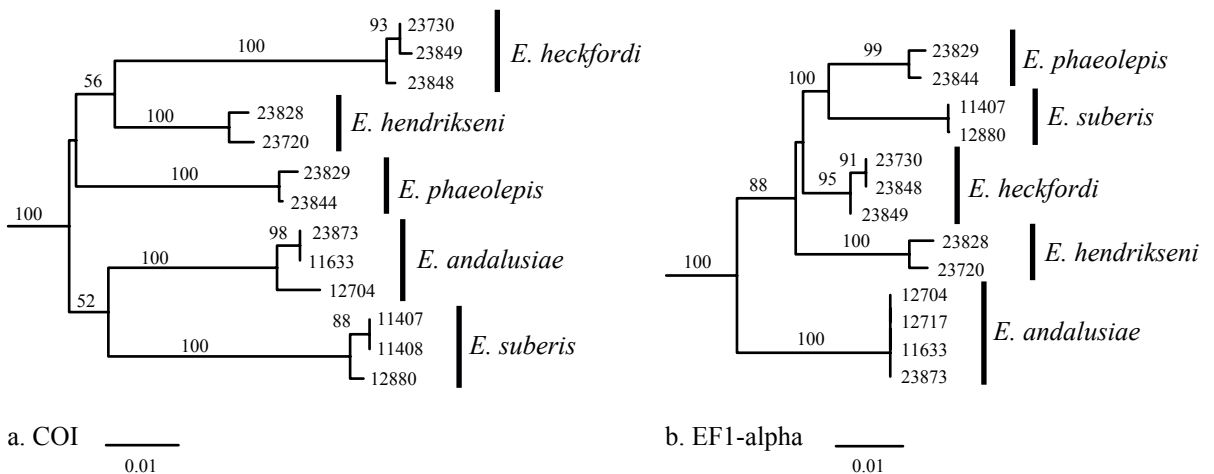


Fig. 4. K2P Neighbor joining trees containing species of the *Ectoedemia suberis* complex. Specimen numbers are RMNH.INS registry numbers. All hosts are *Quercus* spp. and all species are only found in South-West Europe. Outgroup for this analysis was *E. terebinthivora*, bootstrap values represent 10,000 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.

logical 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 Nieuwerkerken *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 Nieuwerkerken, 1985; van Nieuwerkerken *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*.

The Ectoedemia subbimaculella complex. This is the second complex found in this group (van Nieuwerkerken, 1985; van Nieuwerkerken *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 intro-

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
<i>E. albifasciella</i>	A	T	A	C	C	C	C	T	A	T	A	T	T	C	T
<i>E. pubescivora</i>	A	C	A	C	C	C	C	T	A	T	A	T	T	T	C
<i>E. contorta</i>	G	T	A	T	C	C	T	C	A	C	A	T	T	C	T
<i>E. cerris</i>	A	T	G	C	T	T	C	T	G	T	G	C	C	C	T

gression 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 phyllo-tomella* cannot be satisfactorily distinguished in either marker.

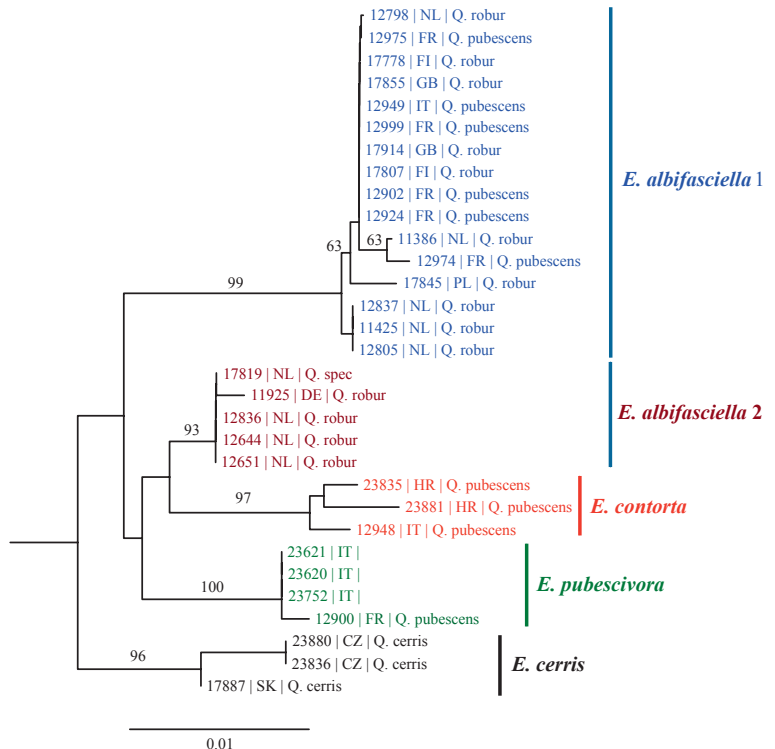
Especially interesting is the clade consisting of two specimens identified as *E. subbimaculella* from Hungary (feeding on *Quercus cerris*) and two specimens identified as *E. heringi* from Greece (found on *Quercus ithaburensis*), further complicating this

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 whole group, although without support. The general picture for this complex is that *E. heringi* and *E. phyllo-tomella* 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

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.



a. COI

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
<i>E. subbimaculella</i>	A	T	T	A	A	G	A	T	T/C	A/G
<i>E. heringi</i>	A	T	T	G	A	A	A	T	A	A
<i>E. liechtensteini</i>	G	T	T	G	A	A	A	T	A	A
<i>E. phyllotomella</i>	A	T	T	G	A	A	A	T	A	A
<i>E. 'subbimaculella' from Q. cerris</i>	A	G	G	G	G	A	G	C	A	T
<i>E. 'heringi' from Q. ithaburensis</i>	A	G	G	G	G	A	G	C	A	T

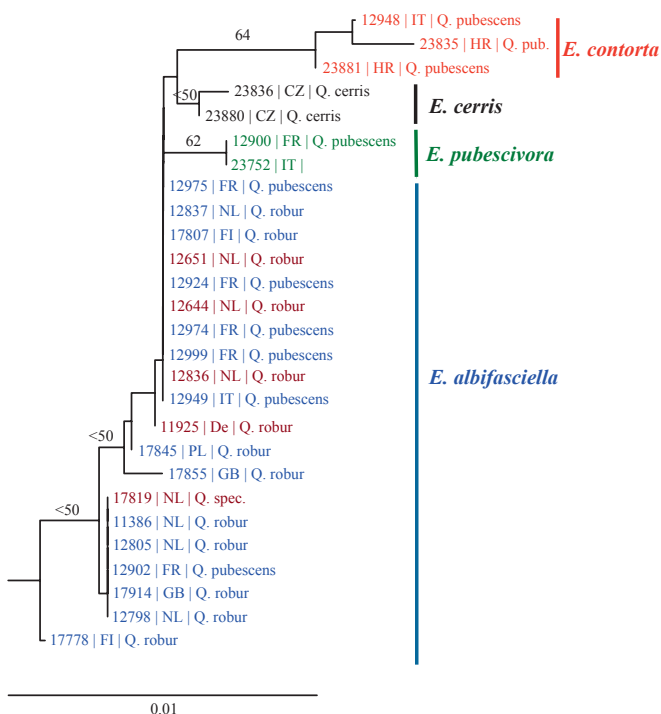
COI distinguishes *E. liechtensteini* (taking into consideration that only three specimens have been sequenced). In EF1- α there is just a single character attribute: *E. subbimaculella* (including the misidentified RMNH.INS.23514) has an A in position 221, where the others have a G. All of these are 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 derivatives 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*



0.01

b. EF1-alpha

ciella 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

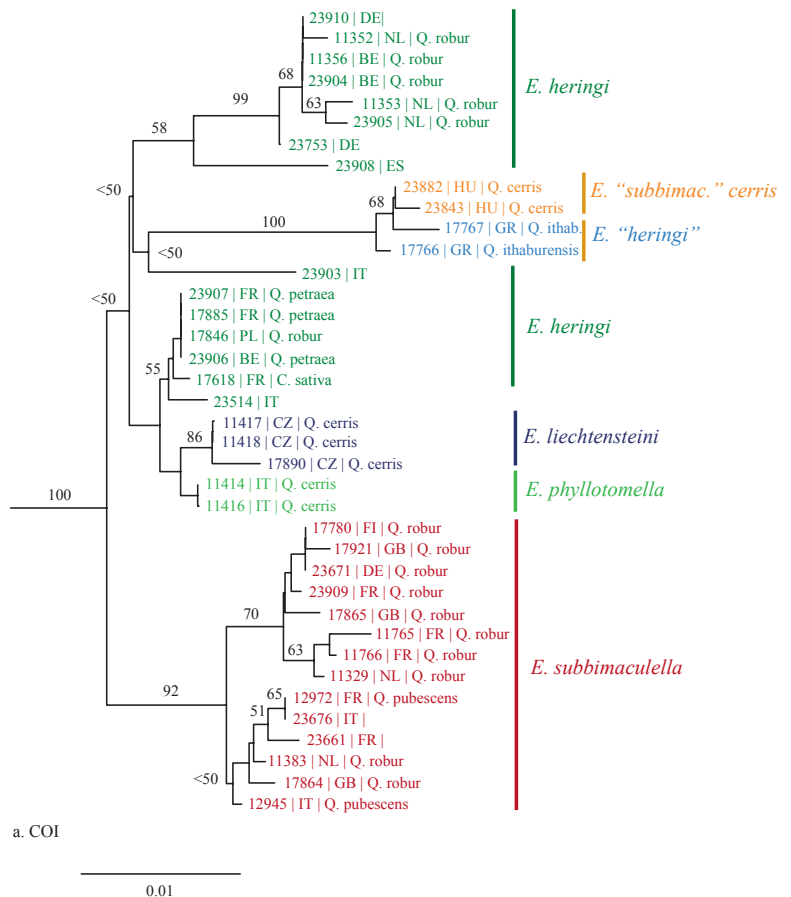
explained by a higher level of saturation in COI, and thus loss of information in this marker at this taxonomic level. However, both genes have their limitations and cannot identify all species on their own. Taken together the resolution becomes much better and almost all species are straightforward 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 Meta-

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.



zoa 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

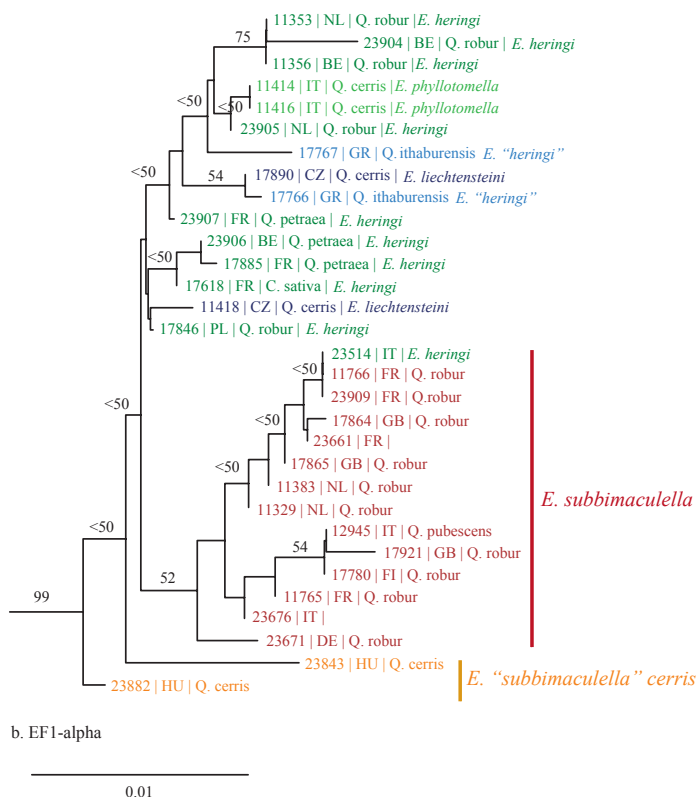
results whether these are widespread species or should be split in two taxa. In the first case we have not enough material to check it by morphology, the single female from Japan is inseparable from European ones, although the barcode would clearly suggest two different species, but we suspect that this 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 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 Nieuwerkerken, 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 Euro-

pean *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 (Lossler, Netherlands), Hiroshi Kuroko (Osaka, Japan), John

Langmaid (Southsea, United Kingdom), Aleš Laštůvka (Prostějov, 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 Maritime 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.0905845106](https://doi.org/10.1073/pnas.0905845106).
- Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesankurthi 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 fuligator*. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14812-14817. doi: [10.1073/pnas.0406166101](https://doi.org/10.1073/pnas.0406166101).
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004b. Identification of Birds through DNA Barcodes. *PLoS Biol* 2: e312. doi: [10.1371/journal.pbio.0020312](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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, Thi-aucourt 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](https://doi.org/10.1111/j.1755-0998.2009.02628.x).
- Kailla L, Ståhls G. 2006. DNA barcodes: Evaluating the potential of COI to differentiate 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 Nieuwerkerken 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](https://doi.org/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.
- Nieuwerkerken EJ van. 1985. A taxonomic revision of the western Palearctic 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>.
- Nieuwerkerken EJ van. 1986. Systematics and phylogeny of Holarctic genera of Nepticulidae (Lepidoptera, Heteroneura: Monotrypsia). *Zoologische Verhandlungen* 236: 1-93. <http://www.repository.naturalis.nl/record/317599>.
- Nieuwerkerken EJ van, Laštůvka A, Laštůvka Z. 2010. Western Palearctic *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](https://doi.org/10.3897/zookeys.32.282).
- Nieuwerkerken EJ van, Berggren K. 2011. *Ectoedemia rosae*, a new species from the French Alps and Norway (Lepidoptera: Nepticulidae). *Tijdschrift voor Entomologie* 154: 181-191.
- Nieuwerkerken EJ van, Kailla 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.
- Öunap 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](https://doi.org/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.1471-8286.2007.01678.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: 10.1080/106635150802033006.
- 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, Nieuwerkerken 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: 10.1163/002829683X000093.
- 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: Hesperidae). *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.

Received: 13 July 2011

Revised and accepted: 11 November 2011

Published online: 19 January 2012

Editor: J.A. Miller

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

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

neighbours (K2P model) and the name of the nearest 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
	–	–	<i>E. liebwerdella</i>	4.85	0					
	–	–	<i>E. longicaudella</i>	4.85	1		–	–	<i>E. admiranda</i>	1.21
	0.15	0.15	<i>E. liebwerdella</i>	8.6	1		–	–	<i>E. liebwerdella</i>	8.12
	–	–	<i>E. longicaudella</i>	6.77	0					
	0	0	<i>E. caradjai</i>	6.27	2		0	0	<i>E. preisseckeri</i>	6.71
	1.6	3.25	<i>E. arcuatella</i>	8.09	5		0.17	0.42	<i>E. aegilopidella</i>	7.13
	6.82	6.82	<i>E. agrimoniae</i>	5.79	2	2	2.47	2.47	<i>E. agrimoniae</i>	5.01
	0.82	1.25	<i>E. spiraeae</i>	5.79	2		0	0	<i>E. spiraeae</i>	5.01
	0.99	1.52	<i>E. ivinskisi</i>	7.46	3		0.42	0.42	<i>E. aegilopidella</i>	6.69
	0.37	0.77	<i>E. arcuatella</i>	7.27	17		0.14	0.84	<i>E. rubivora</i>	2.99
	0.07	0.31	<i>E. rubivora</i>	0	9		0	0	<i>E. rubivora</i>	0.42
	0.58	0.91	<i>E. atricollis</i>	0.15	6		0	0	<i>E. rubivora</i>	0.21
	0.35	0.93	<i>E. atricollis</i>	0	11		0.09	0.42	<i>E. arcuatella</i>	0.21
	0.74	1.89	<i>E. mahalebella</i>	5.95	5		0.67	1.68	<i>E. mahalebella</i>	2.98
	0.15	0.3	<i>E. spinosella</i>	5.95	5		0.08	0.21	<i>E. spinosella</i>	2.98
	0.15	0.3	<i>E. caradjai</i>	5.64	5		0.08	0.21	<i>E. minimella</i>	4.97
	0.1	0.15	<i>E. caradjai</i>	6.44	3		0.28	0.42	<i>E. occultella</i>	4.97
1	–	–	<i>E. andalusiae</i>	5.36	1	1	–	–	<i>E. heckfordi</i>	3.69
	1.01	1.37	<i>E. chasanella</i>	4.65	3		0.28	0.42	<i>E. chasanella</i>	3.19
	0.3	0.45	<i>E. caradjai</i>	5.3	2		0	0	<i>E. heckfordi</i>	2.76
	0.6	0.6	<i>E. andalusiae</i>	4.98	2		0.63	0.63	<i>E. phaeolepis</i>	3.62
	0.61	0.91	<i>E. hendrikseni</i>	4.98	4		0	0	<i>E. heckfordi</i>	3.86
	0.3	0.45	<i>E. hendrikseni</i>	5.63	3		0.14	0.21	<i>E. phaeolepis</i>	2.33
	0.3	0.3	<i>E. hendrikseni</i>	5.14	2		0.42	0.42	<i>E. heckfordi</i>	2.33
	2.47	7.1	<i>E. klimeschi</i>	6.11	6	1	0.4	1.2	<i>E. canutus</i>	4.57
	0.91	1.53	<i>E. klimeschi</i>	4.18	5		0.25	0.63	<i>E. turbidella</i>	4.29
	0.15	0.3	<i>E. argyropeza</i>	3.55	4		0	0	<i>E. argyropeza</i>	1.47
	0.15	0.15	<i>E. argyropeza</i>	2.3	3		0.28	0.42	<i>E. turbidella</i>	1.9
	0.35	1.06	<i>E. klimeschi</i>	2.3	7		0	0	<i>E. turbidella</i>	1.47
	0	0	<i>E. phyllotomella</i>	5.94	3	2	0	0	<i>E. alnifoliae</i>	3.69
	0.1	0.15	<i>E. coscoja</i>	3.64	3		0	0	<i>E. algeriensis</i>	2.97
	0.15	0.15	<i>E. gilvipennella</i>	3.71	3	1	0.54	0.8	<i>E. coscoja</i>	2.75
	–	–	<i>E. quinquella</i>	3.64	2		0	0	<i>E. algeriensis</i>	2.75
	0.3	0.3	<i>E. algeriensis</i>	3.71	2		0.63	0.63	<i>E. algeriensis</i>	3.85
	0.1	0.15	<i>E. subbimaculella</i>	4.49	3		0.42	0.63	<i>E. haraldi</i>	5.41

List of *Ectoedemia* species (continued)

			COI
Species	Hostplant	Region	# sequences
<i>Ectoedemia haraldi</i> (Soffner, 1942)	<i>Quercus</i> evergreen	West Palearctic	4
<i>Ectoedemia ilicis</i> (Mendes, 1910)	<i>Quercus</i> evergreen	West Palearctic	4
<i>Ectoedemia pseudoilicis</i> Laštůvka, Z. & A., 1998	<i>Quercus</i> evergreen	West Palearctic	3
<i>Ectoedemia heringella</i> (Mariani, 1939)	<i>Quercus</i> evergreen	West Palearctic	8
<i>Ectoedemia alnifoliae</i> Van Nieukerken, 1985	<i>Quercus</i> evergreen	West Palearctic	3
<i>Ectoedemia rufifrontella</i> (Caradja, 1920)	<i>Quercus</i> deciduous	West Palearctic	4
<i>Ectoedemia albifasciella</i> (Heinemann, 1871)	<i>Quercus</i> deciduous	West Palearctic	22
<i>Ectoedemia cerris</i> (Zimmermann, 1944)	<i>Quercus cerris</i>	West Palearctic	3
<i>Ectoedemia pubescivora</i> (Weber, 1937)	<i>Quercus pubescens</i>	West Palearctic	4
<i>Ectoedemia contorta</i> Van Nieukerken, 1985	<i>Quercus</i> deciduous	West Palearctic	3
<i>Ectoedemia subbimaculella</i> (Haworth, 1828)	<i>Quercus</i> deciduous	West Palearctic	16
<i>Ectoedemia heringi</i> (Toll, 1934)	<i>Quercus</i> deciduous, Castanea	West Palearctic	17
<i>Ectoedemia liechtensteini</i> (Zimmermann, 1944)	<i>Quercus cerris</i>	West Palearctic	3
<i>Ectoedemia phyllotomella</i> (Klimesch, 1946)	<i>Quercus cerris</i>	West Palearctic	2
Species from other regions			
Subgenus <i>Zimmermannia</i> Hering, 1940			
<i>Ectoedemia admiranda</i> Puplesis, 1984		East Palearctic	
Subgenus <i>Ectoedemia</i> Busck, 1907			
<i>angulifasciella</i> group			
<i>Ectoedemia</i> <i>Rubus</i> Borneo	<i>Rubus</i>	Oriental	1
<i>Ectoedemia</i> <i>Rubus</i> Cuc Phuong	<i>Rubus</i>	Oriental	1
<i>Ectoedemia</i> <i>Rubus</i> Fansipan	<i>Rubus</i>	Oriental	1
<i>Ectoedemia</i> TamDao		Oriental	1
<i>Ectoedemia picturata</i> Puplesis, 1985	<i>Rosa</i>	East Palearctic	4
<i>suberis</i> group			
<i>Ectoedemia ivinskisi</i> Puplesis, 1984	<i>Quercus deciduous</i>	East Palearctic	2
<i>Ectoedemia ornatella</i> Puplesis, 1984	<i>Quercus deciduous</i>	East Palearctic	3
<i>Ectoedemia chasanella</i> (Puplesis, 1984)		East Palearctic	1
<i>populella</i> group			
<i>Ectoedemia populella</i> (Busck, 1907)	<i>Populus</i>	Nearctic	2
<i>Ectoedemia canutus</i> Wilkinson & Scoble, 1979	<i>Populus</i>	Nearctic	1
<i>Ectoedemia</i> California		Nearctic	1
<i>subbimaculella</i> group			
<i>Ectoedemia Quercus gilva</i>	<i>Quercus gilva</i>	East Palearctic	3
not assigned to group			
<i>Ectoedemia quadrinotata</i> (Braun, 1917)	Betulaceae	Nearctic	2
<i>Ectoedemia arisi</i> Puplesis, 1984		East Palearctic	1
<i>Ectoedemia christophori</i> Puplesis, 1985		East Palearctic	
<i>Ectoedemia Annomocarya</i> Vietnam	<i>Annomocarya</i>	Oriental	1
<i>Ectoedemia Carpinus</i> Vietnam	<i>Carpinus</i>	Oriental	1
<i>Ectoedemia expeditionis</i> Mey, 2004		Afrotropical	1
<i>Ectoedemia tersiisi</i> Mey, 2004		Afrotropical	1
<i>Ectoedemia</i> 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
1	3.16	4.22	<i>E. pseudoilicis</i>	3.54	4	2	0.69	2.53	<i>E. pseudoilicis</i>	2.02
	0.3	0.61	<i>E. heringella</i>	3.07	4		0.17	0.42	<i>E. heringella</i>	0.8
	2.21	3.24	<i>E. heringella</i>	2.77	2		0.21	0.21	<i>E. heringella</i>	1.68
	0.5	0.91	<i>E. pseudoilicis</i>	2.77	6	1	0.16	0.4	<i>E. ilicis</i>	0.8
	2.05	2.62	<i>E. pseudoilicis</i>	3.86	3		0.14	0.21	<i>E. haraldi</i>	3.28
	1.02	1.84	<i>E. phyllotomella</i>	3.64	3	1	0	0	<i>E. albifasciella</i>	1.61
	0.8	2.3	<i>E. pubescivora</i>	1.22	22		0.12	0.42	<i>E. cerris</i>	0
	0	0	<i>E. albifasciella</i>	1.84	2	1	0	0	<i>E. albifasciella</i>	0
	0.08	0.15	<i>E. albifasciella</i>	1.22	2		0	0	<i>E. albifasciella</i>	0.21
	0.61	0.76	<i>E. albifasciella</i>	1.37	3	1	0.2	0.4	<i>E. cerris</i>	0.4
	1.2	3.4	<i>E. heringi</i>	0.3	15	1	0.62	1.62	<i>E. heringi</i>	0.63 (0)
	1.45	3.41	<i>E. phyllotomella</i>	0.15	12	1	0.66	1.89	<i>E. liechtensteini/phyllotomella</i>	0
	0.2	0.31	<i>E. phyllotomella</i>	0.3	2	1	0.4	0.4	<i>E. heringi</i>	0
0	0	<i>E. heringi</i>	0.15	2		0	0	<i>E. heringi</i>	0	
					1	1	–	–	<i>E. liebwerdella</i>	1.21
4	0.38	0.65	<i>E. Rubus Fansipan</i>	5.78	1		–	–	<i>E. Rubus Fansipan</i>	4.08
			<i>E. Rubus Fansipan</i>	6.43			–	–		
			<i>E. Rubus Borneo</i>	6.25	1		–	–	<i>E. TamDao</i>	2.11
			<i>E. Rubus Cuc Phuong</i>	7.23	1		–	–	<i>E. Rubus Fansipan</i>	2.11
			<i>E. spiraeae</i>	7.46	2	2	0	0	<i>E. aegilopidella</i>	5.83
1	–	–	<i>E. caradjai</i>	5.78	1		–	–	<i>E. aegilopidella</i>	7.61
			<i>E. chasanella</i>	7.26	2	2	0	0	<i>E. caradjai</i>	4.99
			<i>E. caradjai</i>	4.65	1		–	–	<i>E. caradjai</i>	3.19
1	–	–	<i>E. canutus</i>	7.11	1	1	–	–	<i>E. canutus</i>	4.54
			<i>E. klimeschi</i>	4.66	1		–	–	<i>E. populella</i>	4.54
			<i>E. klimeschi</i>	4.08	1		–	–	<i>E. turbidella</i>	2.11
	0.2	0.3	<i>E. heringi</i>	4.42	3		0	0	<i>E. albifasciella</i>	4.07
1	–	–	<i>E. heringi</i>	6.6	2	2	1.21	1.21	<i>E. christophori</i>	6.71
			<i>E. heringi</i>	5.65	1	1	–	–	<i>E. christophori</i>	3.28
			-	-	1				<i>E. arisi</i>	3.28
			<i>E. cerris</i>	5.42	1		–	–	<i>E. preisseckeri</i>	5.01
			<i>E. cerris</i>	4.18						
			<i>E. agrimoniae</i>	7.85	1		–	–	<i>E. tersiusi</i>	7.92
			<i>E. agrimoniae</i>	7.85	1		–	–	<i>E. expeditionis</i>	7.92
			<i>E. spiraeae</i>	7						

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.

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.

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]’.